Mapping studies on the Platelet-Derived Growth Factor A-chain (*PDGFA*) Gene.

S.J. Louise Smith

PhD

The University of Edinburgh

1995

Declaration

I declare that I composed this thesis myself and that the work presented is my own unless otherwise stated,

S.J. Louise Smith
July 1995

Acknowledgements

I would like to thank all the people who have helped me throughout this project.

Firstly, my supervisors Dr. David Bonthron and Professor David Brock.

All of the members of the lab, past and present, especially Kathryn Sinclair and Jon Warner for their endless technical advice and support. Jon also kindly donated his copy of the ICI YAC library. Lisa Strain who gave me endless samples, probes and helpful information throughout this project. Annette Gilfillan who synthesised many of the oligonucleotide primers and Tommy Johnstone for his expertise in FISH. Along with Jon, Lisa and Annette, Lilias Barron kindly lent me DNA samples. Finally I thank Anna Weir, Bruce Hayward, Lindsay Liddle and Sophie Intody for their friendship and company.

From the MRC Human Genetics Unit I thank John Brown and Kate Dry for the gift of the ICI YAC library and all their help, Judy Fantes who analysed the FISH experiments. Alan Wright and Karen Wakefield kindly lent me all of their CEPH DNA samples without which the linkage analysis would not have been possible. Most importantly I owe many thanks to Micheala Aldred who helped endlessly with the computer databases and linkage programs to analyse my data.

All of the diagrams throughout this thesis were drawn by Jason McColl.

Finally to my friends, especially Jason and my parents, who have supported me throughout this project: a mere 'thank you' feels inadequate.

Abstract

As the major mitogen in serum, platelet-derived growth factor (PDGF) has been implicated in many fundamental proliferative processes such as growth, development, wound healing, atherosclerosis and neoplasia. While earlier studies had mapped PDGFA to 7p22.1-pter, the location of the gene relative to other loci, cytogenetic landmarks such as the chromosome telomere was not known. Deletions of chromosome 7p are associated with craniosynostosis which has been mapped to two regions: 7p13-15 and 7p21-22. At the start of this project it was not known whether the location of PDGFA on chromosome 7p overlapped with the more distal of these two regions. This project undertakes genetic and physical mapping studies on the A-chain gene of PDGF to refine the location of the gene on chromosome 7p.

To further describe the structure of the gene and to develop polymorphic markers for linkage analysis, 7.3kb of gDNA from PDGFA was sequenced. This sequencing project identified polymorphic loci including a minisatellite and two dimorphic markers. This minisatellite, which is in intron 3 of the gene, lies less than 300 bp upstream from another minisatellite, in intron 4, which has been described previously (Bonthron 1992). The intron 3 minisatellite described in this project has a heterozygosity of about 56% however some alleles are refractory to PCR amplification making it unsuitable for the large scale typing of data required for the linkage analysis. For this, the two dimorphisms were used. The linkage analysis placed *PDGFA* as the most distal locus in chromosome 7p, about 10 to 16 cM distal to the nearest proximal markers. Two maps using different sets of markers are presented.

Pulsed-field restriction mapping identified a CpG island at the 5' end of PDGFA but excluded close physical linkage to the nearest linked locus, MS31 (*D7S21*). This PFGE restriction map has been extended as far as possible with the probes available.

YAC and P1 clone libraries were exhaustively screened for clones spanning PDGFA however, no positive clones were identified. The paucity of clones may reflect the difficulty of cloning subtelomeric regions. Candidate telomere YACs for chromosome 7p were screened for PDGFA but were also negative. End cloning of these clones showed that they contain repetitive sequence around their cloning site, no single copy probes are available from these YACs at present.

RARE (RecA-assisted restriction endonuclease) cleavage experiments showed that PDGFA lies about 630 to 679 kb from the telomere. Analysis using a probe from exon 7 of PDGFA, with four different individuals, indicated that there may be size variation in this distance. Polymorphic length variation in the subtelomeric regions has also been described for other chromosome telomeres. A telomeric location for PDGFA, which is a growth factor gene, is interesting since the loss of chromosome telomeres is associated with both ageing and malignancy. Loss of telomere repeats results in instability of the chromosome ends and may explain the increase in chromosome rearrangements seen in malignant cells. Such telomere rearrangements can have dramatic effects on the expression of nearby genes; for example truncation of the 16p telomere can silence expression from the nearby α-globin cluster. Expression of the gene has been shown to be silenced even when the gene itself is intact. The consequence of similar rearrangements involving chromosome 7p and the expression of PDGFA are open to speculation.

Abbreviations used in this report

ADP	adenosine 5'-diphosphate
ATP-γ-S	adenosine 5'[γ-thio] triphosphate
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary DNA
c^7dGTP	7-deaza-2'deoxyguanosine 5'-triphosphate
°C	degrees Centigrade
CEPH	Centre d'Étude du polymorphisme Humain
cM	centiMorgan
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
TTP	deoxythymidine triphosphate
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E.Coli	*Escherichia Coli*
EDTA	ethylenediaminetetra-acetic acid disodium salt
EGTA	[ethylenebis (oxyethylenenitrilo)]-tetraacetic acid
FISH	fluorescent in-situ hybridisation
gDNA	genomic DNA
ICRF	Imperial Cancer Research Fund
IPA	isopropanol
IPTG	isopropyl β-D-Thiogalactopyranoside
kb	kilobase
kDa	kiloDaltons
kV	kiloVolts
L	litres
LMP	low melting point
LOD	logarithm of the odds
mA	milliAmps
Mb	megabases
ml	millilitres
μg	micrograms
mg	milligrams
mm	millimetres
M	molar
mM	millimolar
ng	nanograms
nm	nanometres

PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDGF	platelet-derived growth factor
PDGFA	platelet-derived growth factor A-chain gene
PDGFB	platelet-derived growth factor B-chain gene
pdgfrα	platelet-derived growth factor α receptor
pdgfrβ	platelet-derived growth factor β receptor
PEG	polyethylene glycol
PFGE	pulsed field gel electrophoresis
PMSF	phenylmethylsulfonylfluoride
RARE	recA-assisted restriction endonuclease cleavage
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rpm	revolutions per minute
SAM	S-adenosylmethionine
S.cerev	*Saccharomyces cerevisiae*
SDS	sodium dodecyl sulphate
SSV	simian sarcoma virus
STS	sequence tagged site
TE	10 mM TrisHCl, 1 mM EDTA pH 8.0
TEMED	N,N,N'N'-tetramethylethylenediamine
Tris	Tris(Hydroxymethyl)aminomethane
tYAC	telomere YAC
V	volts
Xgal	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside
YAC	yeast artificial chromosome

TABLE OF CONTENTS

Declaration ---------- ii

Acknowledgements ---------- iii

Abstract ---------- iv

Abbreviations used in this report ---------- vi

TABLE OF CONTENTS ---------- viii

TABLE OF DIAGRAMS AND TABLES. ---------- xvii

CHAPTER 1: INTRODUCTION ---------- 1

Aims of this project ---------- 2

Platelet-derived growth factor (PDGF) ---------- 2

Expression of PDGF ---------- 3

Homology between the PDGF A and B-chain genes ---------- 4

PDGF A-chain gene: sequence and structure ---------- 6

Polymorphisms within PDGFA ---------- 7

The minisatellite in intron 4 ---------- 7

An uncharacterised StyI RFLP ---------- 9

Transcriptional control of PDGFA ---------- 9

Alternative splicing of PDGFA ---------- 10

Chromosomal localisation of PDGFA ---------- 11

The PDGF B-chain gene ---------- 14

The PDGF α and β receptors ---------- 14

The *Patch* mutant mouse ---------- 15

The function of PDGF ---------- 16

Mitogenic activity ---------- 16

1.8.2 Chemotaxis 17
1.8.3 Vasoconstriction 17

1.9 The physiological role of PDGF 18
1.9.1 Development 18
1.9.2 Wound repair 18
1.9.3 Atherosclerosis 19
1.9.4 Neoplasia 21

1.10 Abnormalities associated with chromosome 7p 21
1.10.1 The Craniosynostosis Syndromes 22
1.10.2 Greig Syndrome 23
1.10.3 Saethre-Chotzen syndrome 23
1.10.4 Ring Chromosome 7 26

1.11 Approaches to gene mapping 27

1.12 Linkage analysis 28
1.12.1 Maximum likelihood calculations and lod scores. 29
1.12.2 Statistical significance of linkage 30
1.12.3 Mapping functions and genetic distance 30
1.12.4 The relationship between genetic and physical distance 31
1.12.5 Linkage maps available for Chromosome 7p 32

1.13 Physical mapping 34
1.13.1 Long range restriction mapping 34
1.13.2 Restriction enzymes used for pulsed-field mapping 35
1.13.3 The YAC cloning vector 36
1.13.4 Telomere YAC vectors 38
1.13.5 The P1 cloning vector 39

1.14 Physical mapping resources 40
1.14.1 The ICI YAC Library 41
1.14.2 The CEPH 'megaYAC' Library 41
1.14.3 The Washington University total human DNA YAC library 42
1.14.4 Chromosome 7 YAC libraries 42
1.14.4.1 The Toronto human chromosome 7 library 42
1.14.4.2 The Washington University human chromosome 7 YAC library 44
1.14.5 The ICRF human gDNA P1 Library 46

Objectives ---------- 47

CHAPTER 2: MATERIALS AND METHODS ---------- 48

Suppliers of chemicals and enzymes ---------- 49

Bacterial cell culture: Media and additives ---------- 49

Plasmid cloning vectors ---------- 50

Yeast cell culture ---------- 51

Transformation and propagation of plasmid DNA ---------- 52

Preparation of competent cells ---------- 52

Transformation of calcium competent cells ---------- 52

Preparation of electro-competent cells ---------- 52

Electro-transformation of competent cells ---------- 53

DNA extraction and preparation ---------- 53

Small scale preparation of plasmid DNA ---------- 53

Alkali lysis extraction of plasmid DNA ---------- 53

Promega Magic Minipreps ---------- 54

Phenol/chloroform extraction of small scale DNA preparations ---------- 54

Large scale preparation of plasmid DNA ---------- 55

Total yeast DNA extract ---------- 55

P1 clones DNA extraction ---------- 56

Geneclean ---------- 56

Extraction of gDNA from whole blood ---------- 57

Preparation of DNA agarose blocks for PFGE ---------- 57

Preparation of yeast and YAC chromosome DNA for PFGE ---------- 58

Restriction endonuclease digestion ---------- 59

Restriction endonuclease digestion of gDNA for PFGE ---------- 59

Separation of DNA molecules by electrophoresis ---------- 59

Solutions and buffers ---------- 59

Loading buffers ---------- 60

Size markers ---------- 60

Conventional agarose gel electrophoresis ---------- 61

Preparative agarose gel electrophoresis ---------- 61

Pulsed field Gel Electrophoresis (PFGE) ---- 61
Polyacrylamide gel electrophoresis (PAGE) ---- 62

Southern Blotting ---- 63
Transfer of DNA from bacterial colonies to Hybond N^+ membranes ---- 63

Hybridisation of radiolabelled probes to Hybond N^+ filters ---- 64
Radiolabelling DNA probes ---- 64
Checking incorporation of radio-isotope ---- 64
Hybridisation of radiolabelled DNA ---- 65
Autoradiography ---- 65
Removal of Hybridisation signal from Hybond N+ filters ---- 65

ExonucleaseIII unidirectional deletion clones ---- 66

Double stranded DNA Sequencing ---- 67
Oligonucleotide sequencing primers ---- 68

Polymerase chain reaction (PCR) ---- 70
Random PCR amplification of DNA ---- 70
Cloning PCR products ---- 73

End-cloning YACS by vectorette PCR ---- 74

RecA assisted restriction endonuclease cleavage (RARE) ---- 76
PAGE purification of oligonucleotide primer ---- 76
RecA-assisted restriction endonuclease (RARE) cleavage ---- 77

Fluorescent in situ hybridisation ---- 78
Biotin labelling of DNA probes ---- 78
Hybridisation of DNA probes to metaphase chromosome spreads ---- 79
Detecting the biotin labelled probe ---- 80

CHAPTER 3: SEQUENCE AND STRUCTURAL STUDIES ON PDGFA -- 81

Introduction ---- 82

Plasmid clones from PDGFA ---- 83
Exon 2 to exon 4: plasmid clone ACY7 ---- 83
Exon and intron 1: plasmid clone ACZ ---- 83

A series of exonucleaseIII unidirectional deleted clones ---------- 85
Size analysis of the deleted clones from pACY7 ---------- 88

ACY7 sequence data assembly ---------- 90
Completion of the seqence data ---------- 90

Sequence data from intron 1 ---------- 91

Analysis of the sequence data ---------- 96
Database searches ---------- 96
Analysis of the splice donor and acceptor sites ---------- 96

Polymorphic loci in PDGFA ---------- 97
Intron 1: an interrupted GT repeat ---------- 97
Dimorphic markers in PDGFA suitable for linkage analysis ---------- 97
The exon 3 dimorphism ---------- 98
The intron 4 dimorphism ---------- 100

Analysis of pACY7 for repetitive sequences ---------- 100

Analysis of the minisatellite in intron 3 ---------- 102
Database searches for homologous sequences ---------- 104
PCR analysis of the intron 3 minisatellite ---------- 104
Sequence structure of the minisatellite alleles ---------- 106
Allelic repeat number variation ---------- 107
Subrepeat sequence variation ---------- 109

Analysis of the inheritance of the minisatellite alleles ---------- 112
PCR analysis. ---------- 113
Southern blot analysis ---------- 115
Heterozygosity of the minisatellite ---------- 116

Conclusion and discussion ---------- 117

CHAPTER 4: LINKAGE ANALYSIS OF PDGFA ---------- 121

Localisation of PDGFA by genetic linkage ---------- 122

The CEPH panel of families ---------- 122

Polymorphic markers used in linkage analysis--**124**
Polymorphic loci within PDGFA--124
Markers along chromosome 7p---125

PDGFA haplotype frequencies---**128**

Two point linkage analysis.---**130**

Multipoint linkage analysis--**132**

Cri-Map analysis.---**134**

Eurogem Consortium.--**136**

FISH analysis of a cytogenetic breakpoint in chromosome 7p22.-----------------------------**138**

Conclusion and discussion--**139**

CHAPTER 5: PULSED FIELD RESTRICTION MAPPING ------------------142

PFGE restriction map of the PDGFA locus --**143**

Probes used in the PFGE restriction mapping---**143**
Probe ACY7--144
A downstream probe: AC41HR---144
An upstream probe: ACIXSR --146
The minisatellite MS31 (*D7S21*)---146

The pulsed-field electrophoresis gels --**147**

Restriction enzymes used in PFGE---**148**

The physical distance between MS31 and *PDGFA* --**150**

A long-range restriction map for PDGFA ---**161**
NotI restriction sites--161
NruI restriction sites --161
MluI restriction sites --162
XhoI restriction sites --163
SfiI restriction sites--164
SacII restriction sites --170

BspDI restriction sites -- 170
BstZI restriction sites -- 170

The long range restriction map -- 171

Extension of the restriction map --- 173

Discussion and conclusion -- 174

CHAPTER 6: ANALYSIS OF THE HUMAN YAC LIBRARIES. -------------176

Introduction -- 177

PCR analysis of the YAC libraries -- 177
A Positive Control: A PCR assay for the Factor VIII Gene ------------------------------------- 181

Filter hybridisation analysis of the YAC libraries -- 181

Results of the YAC library screening -- 182
The positive control: the Factor VIII gene --- 182
Screening for PDGFA --- 184

Conclusion and discussion -- 184

CHAPTER 7: ANALYSIS OF THE ICRF HUMAN gDNA P1 LIBRARY.---189

The P1 cloning vector -- 190

Probes used to screen the ICRF P1 clone library -- 190
Results of the P1 library filter hybridisations -- 193
Identification of the positive clones -- 193

Analysis of the P1 clones --- 195
Fluorescent *in situ* hybridisation. --- 195
Restriction digests of P1 clones. --- 198
PCR Screening of the P1 clones. -- 200

Conclusion and discussion. --- 202

CHAPTER 8: TELOMERE YACS ---205

Chromosome telomere YAC clones ---206

Telomere YAC vectors---206

The human telomere YAC Library---207

Analysis of the telomere YAC clones for PDGFA---209

PCR analysis of the telomere YACs---209

Analysis of the telomere YACs by filter hybridisation---209

Screening the whole telomere YAC library---213

Further analysis of the telomere YACs ---213

Fluorescent *in situ* hybridisation ---214

Screening for CA repeats in the YAC clones---216

Isolation of endclones from YACs---216

Vectorette PCR end cloning---217

Development of PCR assays for the YAC endclones---220

Discussion and conclusion---222

CHAPTER 9: RECA-ASSISTED RESTRICTION ENDONUCLEASE (RARE) CLEAVAGE. ---226

Introduction---227

RecA assisted restriction endonuclease (RARE) cleavage---227

Application of the RARE cleavage technique---228

A suitable EcoRI site in intron 4 of PDGFA ---230

The RARE experiments ---232

Results of the RARE cleavage experiments ---234

Polymorphic length variation in the sub-telomeric region ---236

9.8 Discussion ---------- 240

CHAPTER 10: CONCLUSION AND DISCUSSION.---------- 243

10.1 A telomeric location for PDGFA ---------- 244

10.2 The telomere sequence and structure---------- 244

10.2.1 Minisatellite sequences in subtelomeric regions---------- 245

10.3 Development of maps of telomeric regions ---------- 246

10.3.1 Linkage analysis---------- 246

10.3.2 Long range restriction maps---------- 247

10.3.3 YAC clone contig maps---------- 247

10.4 Representation of loci in YAC clone libraries ---------- 248

10.4.1 Development of chromosome telomere clones ---------- 248

10.5 The implications of a telomere location for PDGFA---------- 249

10.5.1 Chromosome telomeres in malignancy ---------- 250

10.5.2 Telomere length and ageing---------- 250

10.5.3 Ring Chromosome 7 and PDGFA ---------- 251

10.6 Future work---------- 252

REFERENCES ---------- 254

APPENDIX I SUPPLIERS OF CHEMICALS AND ENZYMES---------- 268

PUBLICATION ARISING FROM DATA PRESENTED IN THIS THESIS 272

TABLE OF DIAGRAMS AND TABLES.

CHAPTER 1: INTRODUCTION

Diagram 1.1: Homology between PDGFA, PDGFB and *v-sis* genes. 5

Diagram 1.2: The PDGFA minisatellite in intron 4. 8

Diagram 1.3: Localising PDGFA. 13

Diagram 1.4: Linkage analysis in Saethre-Chotzen craniosynostosis 25

Diagram 1.5: Summary of genetic linkage maps for chromosome 7p15-7pter. 33

Diagram 1.6: The YAC cloning vector pYAC4 37

Diagram 1.7: The Toronto chromosome 7 YAC library 43

Diagram 1.8: The Washington University chromosome 7 YAC library 45

CHAPTER 2: MATERIALS AND METHODS

Table 2.1: Oligonucleotide sequencing primers 69

Table 2.2: Oligonucleotide primers and reaction conditions used in PCR assays. 71-72

Table 2.3: Oligonucleotide primers used for 'Vectorette PCR'. 75

CHAPTER 3: SEQUENCE AND STRUCTURAL STUDIES ON PDGFA

Diagram 3.1: The plasmid subclones pACY7 and pACZ. 84

Diagram 3.2: ExonucleaseIII nested deletion clones 86

Diagram 3.3: The exonucleaseIII deleted subclones from pACY7 87

Diagram 3.4: PvuI and PvuII digests to size the deleted subclones from pACY7 89

Diagram 3.5: The gDNA sequence of PDGFA from exon 1 to exon 4.---------------------- 92-95

Diagram 3.6: The dimorphic loci in exon 3 and intron 4 of PDGFA.--------------------------- 99

Diagram 3.7: Identification of a repetitive sequence by sequence comparison --------------101

Diagram 3.8: Sequence of the repeats in the intron 3 minisatellite ---------------------------103

Diagram 3.9: PCR amplification of the intron 3 minisatellite ----------------------------------105

Table 3.10: Minisatellite repeat variation---108

Diagram 3.11: Subrepeat sequence variation in the minisatellite alleles. --------------------110

Diagram 3.12: Sequence variation in the repeats of allele 7337/1-----------------------------111

Diagram 3.13: Analysis of the inheritance of the intron 3 minisatellite alleles--------------114

Diagram 3.14: The relative positions of the two minisatellites in PDGFA -------------------119

CHAPTER 4: LINKAGE ANALYSIS OF PDGFA

Table 4.1. The CEPH panel of families--123

Diagram 4.2: The PDGFA dimorphic markers used in the linkage analysis ----------------126

Table 4.3: The markers used in the linkage analysis. --127

Table 4.4. PDGFA Haplotype frequencies. ---129

Table 4.5: Pairwise linkage anlaysis using the *PDGFA* haplotype data.----------------------131

Diagram 4.6: Multipoint linkage analysis---133

Diagram 4.8: The Eurogem Cri-Map linkage analysis of chromosome 7p. -----------------135

CHAPTER 5: PULSED FIELD RESTRICTION MAPPING

Diagram 5.1: Single copy probes from PDGFA.--145

Table 5.2: Restriction enzymes used for PFGE restriction mapping ------------------------149

Diagram 5.3: Partial EcoRI digests.--151

Table 5.4: Measuring the physical distance between *PDGFA* and *D7S21* (MS31).---------153

Diagram 5.5: Pulsed field restriction analysis: experiment 1---------------------------------155

Diagram 5.6: Pulsed field restriction analysis: experiment 2---------------------------------156

Diagram 5.7: Pulsed field restriction analysis: experiment 3 --------------------------------157

Diagram 5.8: Pulsed field restriction analysis: experiment 4---------------------------------158

Diagram 5.9: Pulsed field restriction analysis: experiment 5---------------------------------159

Diagram 5.10: Pulsed field restriction analysis: experiment 6 ------------------------------160

Diagram 5.11: Pulsed field restriction analysis: experiment 7-------------------------------165

Diagram 5.12: Pulsed field restriction analysis: experiment 8 ------------------------------166

Diagram 5.13: Restriction sites for XhoI on hybridisation with ACY7. --------------------167

Diagram 5.14: Restriction sites for SfiI on hybridisation with ACY7. ----------------------168

Diagram 5.15: Localisation of the restriction sites for BstZI --------------------------------169

Diagram 5.16: The long range restriction map---172

CHAPTER 6: ANALYSIS OF THE HUMAN YAC LIBRARIES

Diagram 6.1: PCR assays and probes used to screen the YAC libraries for clones spanning the PDGFA gene ---178

Table 6.2: PCR analysis of the YAC libraries--180

Diagram 6.3: Analysis of the ICI YAC library for Factor VIII.------------------------------183

Diagram 6.4: PCR analysis of the ICI library using primer pair 430/731 ------------ 185-186

CHAPTER 7: ANALYSIS OF THE ICRF HUMAN P1 LIBRARY

Diagram 7.1: Filter hybridisation of the P1 library using PDGFA probe ACY7 ----------192

Diagram 7.2: Potential positive clones identified by hybridisation with ACIXSR---------194

Table 7.3: The positive clones requested from ICRF. ---------------------------------------196

Diagram 7.4: Screening single colonies from the P1 clones. ----------------------------------197

Diagram 7.5: Fluorescent *in situ* hybridisation analysis of the P1 clones--------------------199

Diagram 7.6: Sequence data from the ACIXSR probe ---201

Diagram 7.7: PCR analysis of the P1 clones. ---203

CHAPTER 8: TELOMERE YACs

Table 8.1: Candidate telomere YACs for 7p.---208

Diagram 8.2: PCR analysis of the telomere YAC clones for PDGFA -----------------------210

Diagram 8.3: Filter hybridisation analysis of the telomere YAC libraries------------------212

Diagram 8.4: Fluorescent *in situ* hybridisation using y2185 and y2185e --------------------215

Diagram 8.5: The vectorette PCR products digested with EcoRI----------------------------219

Diagram 8.6: Sequence from the Vectorette endclones of the tYACs -----------------------222

Diagram 8.7: PCR analysis of somatic cell hybrids using primer pair sv-1-2178F/R ------223

CHAPTER 9: RARE CLEAVAGE ANALYSIS

Diagram 9.1: RecA-assisted restriction endonuclease (RARE) cleavage --------------------229

Diagram 9.2: Sequence data from intron 4 ---234

Diagram 9.3: EcoRI methylation of genomic DNA--263

1. Chapter 1: Introduction

1.1 Aims of this project

The aims of this project were to develop genetic and physical maps for the platelet-derived growth factor A-chain (PDGFA) gene. Mapping projects were undertaken at all scales: to provide more structural and sequence data from the gene, a linkage map and a physical map in the form of a clone contig. To put into context the work in this project and to explain the reasons for studying this gene, the protein (PDGF), its constituent chains (A and B genes), its receptors and physiological role are discussed. The work in this project provided further information about the A-chain gene sequence and structure, this is presented along with a discussion of the implications of these findings. Importantly, this work has shown that the PDGF A-chain gene is located in a subtelomeric position on chromosome 7p. The implications of such a location for a growth factor gene are discussed.

1.2 Platelet-derived growth factor (PDGF)

As the major mitogen in serum and a potent growth factor secreted by many cell types, the platelet-derived growth factor protein has been extensively studied. PDGF has been implicated in many important processes such as growth, development and wound healing as well as in the development of pathologies such as neoplasia and atherosclerosis.

The existence of PDGF was realised with the observation that material released from platelets was the principal source of mitogens present in serum and was responsible for the growth in culture of many serum dependent cells (Kohler 1974, Ross 1974). PDGF was initially purified from the α-granules of human platelets (Heldin 1979, Antoniades 1979), although it is now known to be secreted by a range of cell types which are discussed below (section 1.3).

PDGF from human platelets is a 30 kDa cationic glycoprotein comprised of two disulphide-bonded polypeptide chains. The PDGF heterodimer is comprised of two chains: A and B which are encoded by two separate genes which are described in sections 1.4 to 1.6 below. All three possible isoforms (AA, AB and BB) have been described and are functionally active for cells carrying the appropriate α and/or β receptors. This project concentrates on the A-chain gene, with specific interest in the detailed mapping of its sequence, structure and chromosomal location.

1.3 Expression of PDGF

PDGF is secreted by platelets and by a variety of different cells such as vascular endothelial cells (Bowen-Pope 1984, Collins 1985, 1987), vascular smooth muscle cells (Seifert 1984, Walker 1986, Majesky 1988), uterine and placental smooth muscle cells (Taylor 1988, Mendoza 1990, Holmgren 1991), neurons and glial cells (Sasahara 1991, Yeh 1991), macrophages (Shimokado 1985) and transformed cells such as osteosarcomas (Heldin 1986), embryonal carcinoma cells (Rizzino 1985), SSV and SV40-transformed cells (Bowen-Pope 1984).

The PDGF-AB and -BB isoforms were found to constitute about 70 and 30% respectively of PDGF purified from human platelets (Hammacher 1988) although the relative proportions of each isoform appears to depend on the isolation procedure used (Hart 1990). The homodimer PDGF-AA is produced by normal cells such as fibroblasts (Paulsson 1987), as well as by tumour cell lines such as osteosarcomas (Heldin 1986). The secretion of PDGF isoforms by a cell line stably transfected with PDGF-A and B implied that the formation of each isoform of PDGF is dictated by the relative abundance of each transcript and is not directly controlled by the cell. It has been proposed that there may be a difference in the secretory behaviours of the PDGF isoforms. PDGF-BB remains associated with the producer cell and is not secreted (Ostman 1988), this retention of PDGF-BB is based on the basic amino

acids in the C-terminal of the protein (Ostman 1991). The A-chain is produced as two splice variants (see section 1.5.3). It has been proposed that the longer variant, which carries a motif similar to the C-terminus of the B chain remains bound to the cell or the extracellular matrix while the shorter variant is secreted (Raines 1992). However, a major control over the functional activity of PDGF is clearly the expression of the PDGF α and β receptors by the target cells (discussed in section 1.7).

Normal human plasma contains undetectable levels of PDGF. Experiments based on intravenous injection of PDGF into baboons to measure the half-life of the protein found it to be less than two minutes (Bowen-Pope 1984). The binding of PDGF by heparan sulphate glycoproteins in the extra-cellular matrix has been demonstrated and the presence of plasma has been shown to reduce the ability of PDGF to bind to cultured cell receptors which implies that the serum contains proteins which bind PDGF.

Homology between the PDGF A and B-chain genes

The PDGF heterodimer comprises of two chains, A and B, which are encoded by two separate genes. The A-chain gene has been mapped to human chromosome 7 while the B-chain gene is located on chromosome 22. The PDGF B-chain shows homology to the simian sarcoma virus gene *v-sis*. Characterisation of the two genes has shown them to be very similar in their structure; both genes span 23 - 24 kb and have 7 exons. The exon/intron boundaries correspond to similar domains in the polypeptide chains and are comparable between both human genes and the *v-sis* gene (see diagram 1.1). Exon 1, in addition to an unusually long untranslated region, encodes the signal peptide. Exons 2 and 3 encode the N-terminal propeptide, exons 4 and 5 the mature PDGF subunit. Exon 6 in the A-chain gene may by either

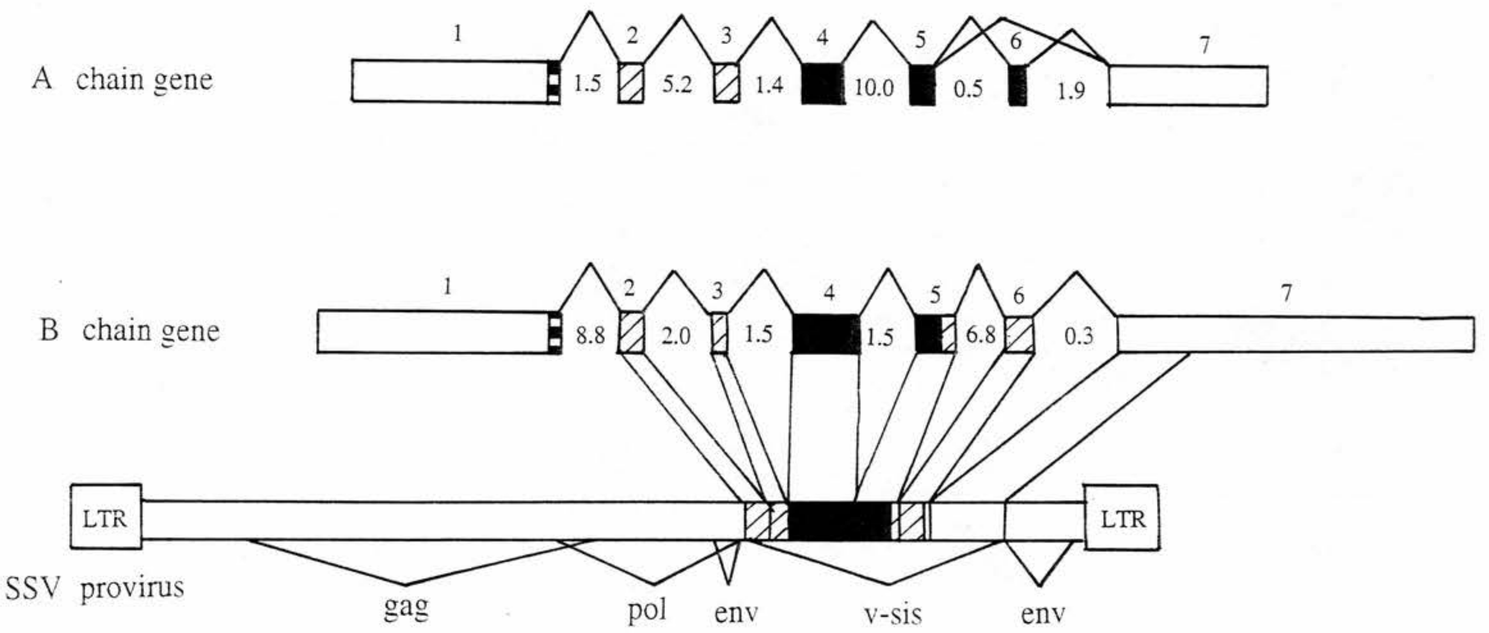


Diagram 1.1: Homology between PDGFA, PDGFB and *v-sis* genes.

Schematic representation of the platelet-derived growth factor A-chain, B-chain and *v-sis* genes, comparing the exon/intron structures of the three genes. Boxes in the PDGF genes represent exons, open boxes: untranslated sequences, checked boxes: leader sequences, hatched boxes: sequences encoding propeptides; filled boxes: mature PDGF. LTR: long terminal repeat. The figures over each exon represent its number whereas those between the exons represent the sizes of the introns in kilobases. From Westermark (1990a).

included or skipped by alternate splicing generating proteins with different C-termini. (section 1.5.3). This exon shows a much lower level of homology to the PDGF B exon 6 than other exon pairs. Exon 7, which is very large, encodes almost entirely untranslated sequences. The amino acid sequence of mature proteins from the two genes show about 56% homology (Betsholtz 1986). This high level of homology between the two genes suggests that they are both derived from a common source.

1.5 PDGF A-chain gene: sequence and structure

Complementary DNA (cDNA) clones encoding PDGFA were first isolated from human glioblastoma cDNA libraries (Betsholtz 1986). Three transcripts of 1.9, 2.3 and 2.8 kb were identified, of which the 2.3 kb form was the most abundant. Subsequent isolation of human genomic clones, DNA sequencing and characterisation of its structure showed that the gene has 7 exons and spans 24 kb (Bonthron 1988). The chromosomal localisation of PDGFA to 7p21-22 is discussed below in section 1.5.4. As described above in section 1.4, PDGFA shows strong homology to the PDGF B-chain gene.

Sequencing of PDGFA identified a TATAA consensus promoter element in a GC rich promoter region and a AATAAA polyadenylation consensus signal in exon 7 (see section 1.5.2). The exon/intron boundaries were identified and all splice junctions were found to conform to the GT/AG rule (Bonthron 1988). Alternative splicing, omitting exon 6, has been described and is discussed below in section 1.5.3.

Sequencing of PDGFA has revealed a number of unusual features. At the start of intron 3 a 41 bp sequence is directly repeated while intron 4 begins with a minisatellite repeat. This structure starts within exon 4 and each repeat includes the splice donor site, it is described below in section 1.5.1. Analysis of the 3' region of

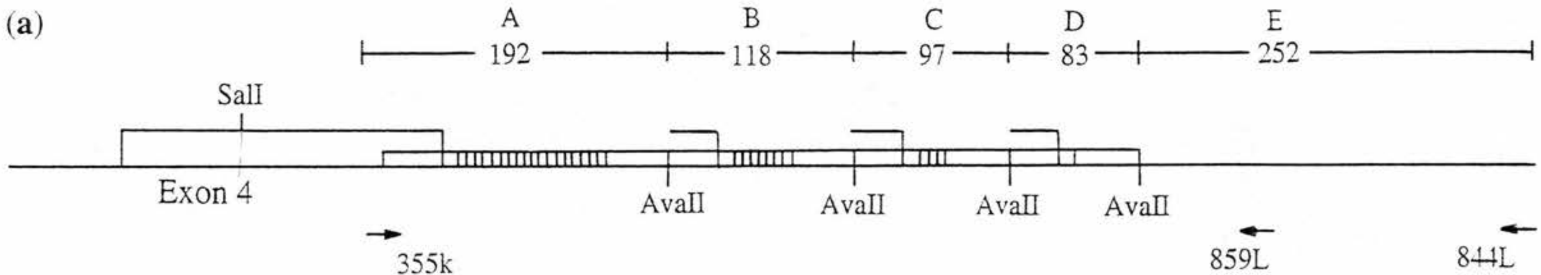
the gene found that it is poorly conserved outside the protein coding regions, however a 32 bp sequence near the middle of intron 6 is perfectly conserved. Another well conserved region in intron 6 is a 100 bp sequence which is repeated in human DNA but not in the mouse. The functional significance of these elements is not known (Bonthron 1992).

1.5.1 Polymorphisms within PDGFA

To achieve one of the major aims of this project, that is to map the PDGFA locus by linkage analysis, polymorphic markers were required. Two polymorphic loci which had been identified within PDGFA prior to the start of this project are described below. Unfortunately neither of these were suitable as markers for linkage analysis and the search for useful markers is described in Chapter 3.

1.5.1.1 The minisatellite in intron 4

The minisatellite mentioned above is shown in diagram 1.2. As the diagram shows, the minisatellite comprises an 80 bp monomeric repeat which is repeated three times (Bonthron 1992). The first repeat includes the last 29 bases of exon 4, as a result there four copies of the splice donor site for exon 4. It is not known whether this repetition of the splice donor site has an effect on splicing patterns. At the centre of the first three repeat elements are, respectively, 18, 7 or 3 copies of the pentamer 'CCTCC'. Within the first repeat element which contains 18 copies of the pentamer there is some subunit sequence variation and some repeats have the sequence 'TCTCC'. A polymorphic variant which carries only 16 pentamers in the first repeat element has been described, however this allele is rare and the polymorphism is not sufficiently heterozygous to be a useful marker (Bonthron 1992).



(b)

```
                                                                                      GTCG
ACCCCACGTCCGCCAACTTCCTGATCTGGCCCCCGTGCGTGGAGGTGAAACGCTGCACCGGCTGCTGCAACACGAGCAGTGTCAAGTGCCAGC

CCTCCCGCGTCCACCACCGCAGCGTCAAG/GTGAGCCCT CC(CCTCC)18 CCACCCAAGCCC GGGGCCGCCCCTACCAAGCAGCCAGGGTC
CCTCCTGCATCCACCACCACAGCGTCAAG/GTGAGCCCTGCC(CCTCC)07 CCACCCGAGCCCAAGGGCCGCCCCTACCTAGCAGCCAGGGTC
CCTCCCGCGTCCACCACCGCAGCGTCAAG/GTGAGCCCTGCC(CCTCC)03 CCACCCGAGCCC GGGGCCGCCCCTACCAAGCAGCCAGGGTC
CCTCACGCATCCAACACTGCAGCATCCAG/GTGAGCCCTGCC          CCACGCAGGCCTGAGGGTCGCCCCTACTGGGTTGCCAGGGTC
CCTCC

   TGGGAGGCCACAAGGTGCCCCTGGTGTGGCTGGTGGGGGTGCACGGCTGCGCTCCCATCTCCCTGGCAGGTACACTGAGGCCCAGGGA
GGAGAAGGAGCACCTGGGGGTGCGGGAAAGGCAGGGGCGCCGGGGCTTCT
```

Diagram 1.2: The PDGFA minisatellite in intron 4.

a) Schematic representation of the components of the repeated sequence. AvaII sites which separate each repeat block are shown with the corresponding fragment sizes A-E in base pairs. b) Alignment of the sequences of the individual 80 bp repeat units. The positions of the intron 4 splice donor are shown '/ ', the exon 4 sequence is shown in italics. From Bonthron 1992.

1.5.1.2 An uncharacterised StyI RFLP

Ferns (1990) described a StyI RFLP detected by a PDGFA cDNA probe. Southern blot analysis gave a complex pattern of hybridisation with one band appearing to be multiallelic. The molecular basis of this RFLP was not known at the start of this project, however the distribution of StyI sites in the genomic sequence suggested that the polymorphic band was due to hybridisation to exon 4 (Bonthron 1992). Unfortunately identification of this reported polymorphism using Southern blotting proved difficult to replicate. Thus, in this form, the polymorphism was not a useful marker and needed further analysis to develop a useful marker.

1.5.2 Transcriptional control of PDGFA

The transcriptional regulation of the PDGF A-chain gene is independent of that of PDGFB. The 5' region of the A-chain gene is very GC rich (78.6%), thus at the sequence level this region fits the criteria for a hypomethylated CpG island (Bonthron 1988). The nature and extent of this putative CpG island is investigated in Chapter 5: Pulsed-field restriction mapping. A single transcriptional start was identified by primer extension analysis and S1 nuclease mapping and is located about 36 bp downstream of a TATAA consensus promoter element (Bonthron 1988). A more recent report identified a second putative promoter downstream of the original promoter site however, it lacks a TATA-box and seems to be comprised of several GC-elements (Rorsman 1992).

The promoter region of the A-chain gene contains multiple copies of the consensus binding sequence for the transcription factor Sp1 (GGGCGG), several copies of the putative binding site for AP2 (CCCCAGGC) and three putative binding sites for the zinc finger binding protein WT1 (Gashler 1992). Mutations in the WT1 binding

protein gene lead to the development of Wilm's tumours. These kidney tumours, which develop in childhood have been shown to over express PDGFA. For this reason the role of WT1 in the transcriptional control of PDGFA was investigated. It was found to repress transcription (Gashler 1992).

The alternative splicing of exon 6 which is discussed below (section 1.5.3) does not cause a large enough variation in transcript size to explain the three different sized transcripts which have been described. It is thought that these different sized mRNAs reflect the use of alternative poly(A) sites which have been identified in the 3' end of the gene (Bonthron 1988, Rorsman 1992). Of these only one, the first described (Bonthron 1988), matches the consensus sequence for polyadenylation signal sites (AATAAA). Transcripts isolated from a human U-2 OS cell line were analysed and at least one used an atypical polyadenylation signal with the sequence AATTAAA (Hoppe 1987). This signal site was located 470 bp downstream from the usual consensus signal site. It is thought that this product corresponds to the largest (2.8 kb) product and that the smallest transcript of PDGFA uses a third, upstream, polyadenylation signal. These three sites may explain the different transcript sizes however, the functional significance of these alternative adenylation sites is unknown.

1.5.3 Alternative splicing of PDGFA

Alternative splicing of the PDGF A-chain cDNA clones in human glioma and endothelial cell lines produces mRNA transcripts which differ by 15 amino acids at the C-terminus (Collins 1987, Tong 1987, Rorsman 1988). The two species are identical in their sequence before and after the insert which corresponds to the 69 bp exon 6 (Bonthron 1988, Rorsman 1988). The omission of exon 6 creates a structure which is 15 amino acids shorter, removing a highly basic carboxy-terminal

region. This alternative splicing has been shown to be conserved through evolution which implies that the two forms have an important functional role (Mercola 1988).

Initially it was speculated that the longer mRNA form, which includes exon 6, is expressed only by malignant cells (Collins 1987), however the differently spliced mRNA products have been shown to be normal cellular proteins (Matoskova 1989, Young 1990). It has been proposed that the exon 6 sequence and the corresponding sequence in PDGF-B allow the protein to be bound by the cell and that the shorter protein may be secreted. However, PCR analysis detected similar levels of the two splice variants in a wide range of mouse tissues (Young 1990). These findings do not support the hypothesis that exon 6 encodes a targeting sequence.

A third PDGFA transcript which carries an extra 110 bp insert at its C-terminus has been described in rabbit vascular smooth muscle cells. Expression of this transcript is induced by angiotensinII (Nakahara 1992). The insert has been found to be a sequence in intron 6 which is well conserved in human gDNA (D.T. Bonthron pers. comm.). It is not clear why this sequence is so well conserved or whether it has any functional role in human PDGFA expression.

1.5.4 Chromosomal localisation of PDGFA

PDGFA was initially assigned to the region 7q22-7pter, using a mouse-human somatic cell hybrid panel (Betsholtz 1986). Subsequently, conflicting assignments of PDGFA to 7p21-p22 (Bonthron 1988) and to 7q11.23 (Stenman 1988) were made using *in situ* hybridisation studies. More recently additional somatic cell hybrids containing fragments of chromosome 7 were analysed using *in situ* hybridisation,

these results confirm the assignment of PDGFA to the tip of chromosome 7p (Bonthron 1992). The first two *in situ* hybridisation studies used cDNA clones differing only in the presence or absence of the alternately spliced exon 6. The third study used a genomic probe consisting largely of intronic sequences, thus the results of this third study may be viewed as independent of those obtained using the cDNA probes.

The panel of somatic cell hybrids used in this third study (Bonthron 1992) were also analysed for the presence of PDGFA by PCR. The expected PCR product was seen using those cells known to contain the distal portion of chromosome 7. These hybrids include one (0044 Rag 1-15) which contains only a very small distal portion of chromosome 7 (7p22.1-7pter), which further refines the localisation of PDGFA on chromosome 7p. The conflicting assignment of PDGFA to 7q11.23 was subsequently retracted by the authors (Stenman 1992) and there is no evidence for a second locus for the PDGFA gene.

Speleman (1989) reported a dysmorphic patient DSE who had been shown to carry a *de novo*, and apparently terminal, deletion of 7p22.1-pter. Unexpectedly, *in situ* hybridisation studies showed that DSE had retained the PDGFA locus at the tip of chromosome 7p. The location of the PDGFA is therefore confined to two relatively small cytogenetic regions: (i) between the breakpoint in 0044 Rag 1-15 and the proximal breakpoint in DSE or (ii) between the distal breakpoint in DSE and the chromosome 7p telomere. These two possible locations for the gene are summarised in diagram 1.3. The second possibility would imply that the deletion in DSE is not truly a terminal deletion and that some interstitial DNA has been retained from the tip of the chromosome which has been healed with its own telomere. This would give a subtelomeric location for PDGFA. More refined mapping required linkage studies and physical mapping at the molecular level, this refined mapping of the PDGFA locus formed a major aim of this thesis.

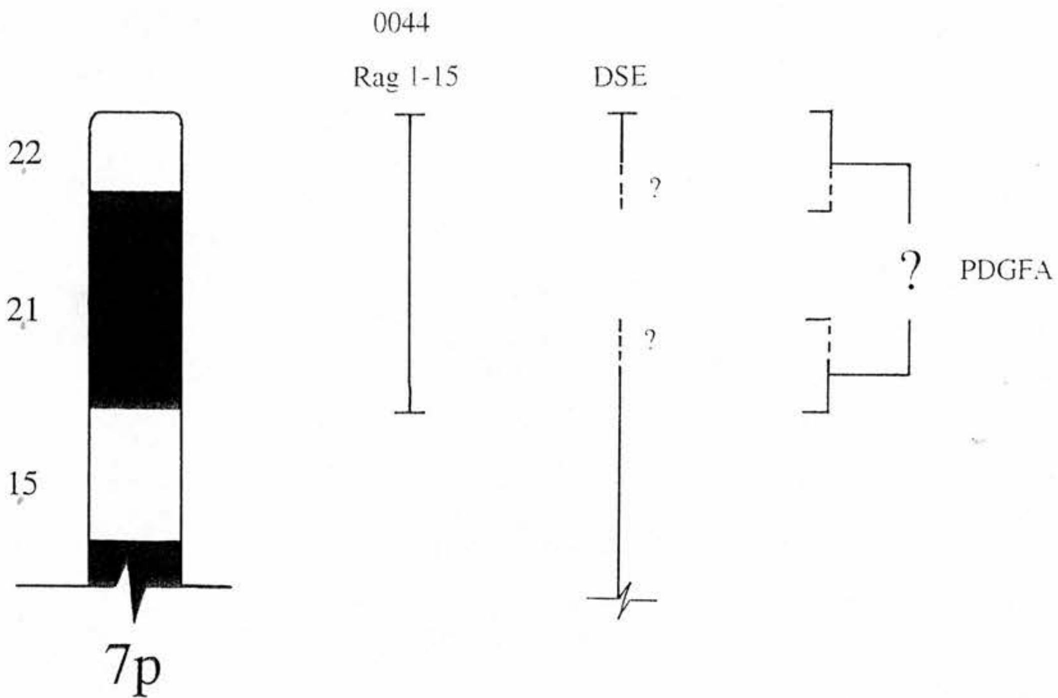


Diagram 1.3: Localising PDGFA.

Two possible locations for PDGFA at the start of this project were inferred from the information available from cytogenetic landmarks. The hybrid cell line 0044 Rag 1-15 indicated that PDGFA was located in 7p21-7pter, while the apparently terminal deletion in the patient DSE narrowed the location of PDGFA to 7p21 or the subtelomeric tip of 7p.

1.6 The PDGF B-chain gene

The PDGF B-chain was isolated from platelets and initially sequenced at the amino acid level. This partial sequencing showed that PDGFB is strongly homologous to $p28^{sis}$, the transforming protein of the simian sarcoma virus (Waterfield 1983, Doolittle 1983). The strong homology between the two proteins suggested that the virus had acquired the cellular sequence which encodes PDGFB. This was the first demonstration of a link between an oncogene (*v-sis*) and a cellular growth factor (c-sis). Comparison of the DNA sequence shows that *v-sis* encompasses all of exons 2 to 6 and the 5' part of exon 7. The signal peptide encoded by exon 1 in c-sis is lost, but it is compensated for by the 5' flanking *env* gene in the viral genome (reviewed in Westermark 1990a). The simian sarcoma virus only transforms cells which are responsive to PDGF, by stimulating cell growth via PDGF using an autocrine feedback mechanism (Westermark 1990b).

Subsequent sequencing of PDGFB at the DNA level showed that the gene has seven exons and includes an unusually long 5' untranslated region which has been shown to inhibit translation (Collins 1985, Rao 1986). The exon/intron structure of the gene is reviewed in comparison to the PDGFA gene in section 1.5. The transcriptional control of PDGFB is independent to that for PDGFA. The PDGFB gene has been mapped to human chromosome 22q12.3-13.2 using a panel of cell lines with stable translocations involving chromosome 22 (Zhang 1990).

1.7 The PDGF α and β receptors

Two PDGF receptor proteins have been described, the α-receptor (pdgfrα) binds all three isoforms with high and equal affinities. The β-receptor binds only PDGF-BB with high affinity, PDGF-AB is bound to the β-receptor at a lower affinity and

PDGF-AA is not bound at all. Like PDGF, the receptor is only functional as a dimer and it is thought to dimerise on binding of the ligand (Heldin 1989).

Pdgfrα has homology with the β receptor, c-kit, colony stimulating factor I receptor (CSF-IR) and fibroblast growth factor receptor (FGFR). Together these proteins form a subfamily of receptor kinases (Yarden 1986). Each receptor consists of an extracellular section with five immunoglobin-like domains and an intracellular part containing a kinase domain with a characteristic inserted sequence which is unique to this subfamily of receptors (reviewed in Heldin 1992). Pdgfrα and c-kit have been localised to a common 630 kb fragment on mouse chromosome 5 (Smith 1991) and to the same cytogenetic band on human chromosome 4q11-12 (Stenman 1989, Gronwald 1990). Interestingly pdgfrβ and CSF-IR are syntenic on mouse chromosome 18 (Wang 1988) and human chromosome 5q31-32 (Yarden 1986). On human chromosome 5 the two genes are found in a head-to-tail arrangement separated by only 500 bp (Roberts 1988).

1.7.1 The *Patch* mutant mouse

Deletion of the pdgfrα gene has been shown to produce the *Patch* mutant mouse model which has been used to evaluate the developmental consequences of this mutation (Smith 1991, Stephenson 1991). *Patch* heterozygotes are relatively normal although they show an abnormal increase in the width of the prefrontal bone and an absence of functional melanocytes from much of their torso. This abnormality of melanocytes causes the white patching of their coat from which their name is derived.

Patch homozygotes are much more severely affected and do not survive to birth. Retardation of proliferation of mesenchymal cells is seen throughout the body of the embryo: they fail to develop cartilage to form a skeleton, the dermis layer of the

skin is missing and organs such as the oesophagus fail to develop their smooth muscle layer. A second set of defects is seen in tissues which develop from the non-neurogenic subset of neural crest cells, thus the embryos fail to develop thyroid or thymus tissue and the septation of the cardiac outflow tract into aorta and pulmonary artery fails to occur. Pdgfrα is also expressed by derivatives of the ectoderm, thus development of the choroid plexus and the optic lens is abnormal (reviewed in Bowen-Pope 1991). It is not clear whether these defects are due to lack of expression of pdgfrα in the defective tissues and reflect an ongoing requirement for PDGF or are the secondary consequences of earlier defects.

1.8 The function of PDGF

1.8.1 Mitogenic activity

PDGF was identified as the element in serum which is required for serum dependent cells in culture to divide (Ross 1974, Kohler 1974). A major function of PDGF is to induce mitosis in target cells such as fibroblasts, arterial smooth muscle cells and endothelial cells. All three isoforms of PDGF have been shown to exert equal mitogenic activity on 3T3 cells and human dermal fibroblasts in culture (Hosang 1989).

An ordered sequence of events is required before a quiescent cell will undergo mitosis. PDGF alone does not promote optimal stimulation for cell division or the synthesis of DNA, rather it appears to induce modulators which make the cell competent to progress into the second stage of division which has been shown to be mediated by substances found in platelet poor plasma (Ross 1974, Pledger 1978). Cells have been shown to respond to stimulation by PDGF with an increase in the levels of c-fos and c-myc mRNA levels. This increased transcription has been

described in high passage senescent fibroblasts which are terminally growth arrested, confirming that the control of the progression of cells into mitosis is controlled by factors other than PDGF (Paulsson 1987).

1.8.2 Chemotaxis

Chemotaxis is the directed movement of cells along a chemical gradient. Unusually for a growth factor, PDGF is a potent chemoattractant for cultured fibroblasts, vascular smooth muscle cells monocytes and neutrophils (Seppa 1982, Grotendorst 1982, Deuel 1982, Senior 1985). PDGF has been shown to exert its chemotactic ability at physiological levels (Westermark 1990b) and all three isoforms have been shown to exert a stimulus (Hosang 1989, Hammacher 1989). The chemotactic role of PDGF is of obvious importance for a role in recruiting inflammatory cells to the site of injuries to promote wound healing (discussed below in section. 1.9.2).

1.8.3 Vasoconstriction

PDGF has also been shown to cause vasoconstriction. In vitro studies using strips of rat aorta smooth muscle found PDGF to be more potent than angiotensinII (Berk 1986). It has been suggested that vasoconstriction is an important factor in atherosclerosis which may suggest that the role of PDGF in atherosclerosis is extensive (Ross 1986, discussed below in section 1.9.3).

1.9 The physiological role of PDGF

The dual role of chemotactic recruitment and mitogenic stimulation of cells such as fibroblasts suggests that PDGF may play an important role in normal proliferative responses such as development and wound repair, and also in pathological processes such as atherosclerosis and neoplasia.

1.9.1 Development

The mitogenic role of PDGF suggests that it would be involved in growth and development. The two PDGF chains and the α and β receptors show distinct and specific differences in their expression during development. The PDGF-AA isoform and the α receptor are expressed very early in the embryo, before implantation. The importance of the expression of the α receptor is shown by the developmental abnormalities seen in the *patch* mouse. By comparison PDGFB and the β receptor are not expressed in the mouse embryo until day 8 of development.

1.9.2 Wound repair

Following injury an inflammatory response is seen in the affected tissue. Neutrophils and platelets migrate to the wound area, followed by macrophages and fibroblasts. As the fibroblasts proliferate and lay down an extracellular matrix, new capillaries start to develop. This provisional or granulation tissue is formed over 2 -3 weeks. Remodelling of the wound, forming permanent scar tissue continues for up to a year. The ability of PDGF to induce migration and proliferation of connective tissue cells would suggest that it has an important role in wound repair.

PDGFB and PDGFβ receptor have been shown by *in situ* hybridisation to be expressed by epithelial cell and fibroblasts in wounded tissue 1-2 days after injury. The expression levels in the wounded site were highest during the initial stages of wound repair and then declined by 5 - 9 days after the injury (Antoniades 1991). These findings were surprising since normally epithelial cells do not express PDGF or its receptor. They provide strong evidence for a role for PDGF in normal wound repair.

In vivo studies in rats have shown that local application of PDGF-BB to the site of injury can enhance would healing by stimulating an earlier influx of cell into the wound area. However the beneficial effects of PDGF-BB are not seen beyond two weeks and longer term treatment with PDGF-BB is no more effective (Grotendorst 1985, Lawrence 1986, Pierce 1988, 1991). Impaired wound healing is seen in conditions such as diabetes where a marked delay in cell infiltration is seen. However once the cells arrive in the area normal development of granulation tissue is seen. The application of PDGF to wounds in diabetic rats restored the rate of wound repair to that of normal animals (Grotendorst 1985, Greenhalgh 1990). The results of a human clinical trial also showed a stimulatory effect of recombinant PDGF-BB on topical application to chronic pressure ulcers (Robson 1992).

1.9.3 Atherosclerosis

Atherosclerosis is characterised by focal plaques in the intima of the artery wall. The lesions contain a core of deposited lipids surrounded by an abnormal proliferation of arterial smooth muscle, macrophages and an excessive deposition of connective tissue. It is thought that the lesions arise as the result of an abnormal response to minor injuries to the vascular endothelium, simulating processes similar to those seen in wound repair.

PDGF is the principal mitogen implicated in the proliferation seen in atherosclerosis. It has been shown that under normal conditions PDGF is produced by vascular smooth muscle (Seifert 1984, Majesky 1988), fibroblasts (Paulsson 1987), macrophages (Shimokado 1985) and vascular endothelium (Bowen-Pope 1984, Collins 1985,1987) all of which are involved in the development of atherosclerotic plaques. Thus PDGF has the potential to be involved in this pathological process. Confirming this, increased secretion of PDGF by plaque tissue has been demonstrated (Libby 1988, Barrett 1988).

The pattern of production of PDGF by the cells in the atherosclerotic plaque is complicated. Evidence from *in situ* hybridisation studies and dissection of the plaques tissues suggests that increased PDGFB is transcribed by macrophages (Ross 1990) and vascular endothelial cells (Barrett 1988, Wilcox 1988) while smooth muscle cells increase PDGFA transcription (Barrett 1988, Majesky 1990). This complicated pattern of expression of the PDGF isoforms is further confused by the different expression of PDGF receptor genes seen in each tissue (Majesky 1990). Each receptor isoform will bind the different isoforms of PDGF with different levels of affinity and thus the effect of secretion of PDGF on the secreting cell and the surrounding tissues will depend on the combination of ligand and receptor isoforms.

The effect of PDGF as a vasoconstrictor has also been mentioned, it has been suggested that arterial spasms occur predominantly at the site of atherosclerotic lesions and that atherosclerotic lesions increase the basal tone of the artery walls (reviewed in Berk 1986). This would suggest an even greater role for PDGF in atherosclerosis although its effect remains to be clearly established.

1.9.4 Neoplasia

The finding of homology between PDGFB and the SSV *v-sis* gene suggested that the abnormal expression of PDGF may lead to malignant transformation (Doolittle 1983, Waterfield 1983). The *v-sis* oncogene has been shown to be required for the transformation of human fibroblasts by SSV and a wide range of human malignant cell lines express one or both PDGF genes. It is thought that naturally occurring human tumour cells express both the c-sis gene and the β receptor allowing an autocrine stimulation of cell growth. Transfection studies using the c-sis gene have produced inconsistent results, thus it is not clear where the role of PDGF in the development of human malignancy lies. It has been suggested that PDGF may be involved in the later events of malignancy such as the development of growth autonomy, local tumour expansion and perhaps metastasis (reviewed in Silver 1992). Thus the role of PDGF in the development and propagation of human malignancies needs further investigation.

1.10 Abnormalities associated with chromosome 7p

One of the major aims of this project was to develop genetic and physical maps for the region surrounding PDGFA. As discussed above (section 1.5.4), PDGFA had been mapped to 7p22.1-7pter prior to the start of this work. The function of a gene and its location within a chromosomal region may be indicated by the finding of chromosomal rearrangements such as deletions or translocations in dysmorphic patients.

At least 32 cases of deletions of chromosome 7p have been identified and characterised. Of these, 14 were associated with craniosynostosis. The association between chromosome 7p deletions and craniosynostosis is discussed below. The familial craniosynostosis syndromes Greig cephalopolysyndactyly and Saethre-

Chotzen syndrome have also been linked to chromosome 7p (see sections 1.10.2 - 1.10.3). A number of patients with ring chromosome 7 have been described. Ring chromosome arises from fusion of the two chromosome termini to each other. Since one of the two possible locations for PDGFA at the start of this project was in the subtelomeric region of chromosome 7p, this syndrome is also discussed below (section 1.10.4).

1.10.1 The Craniosynostosis Syndromes

There are a wide range of syndromes which affect the development of the bones of the skull (craniosynostosis). The mutations causing a number of these defects have now been described. Four of these syndromes, Pfeiffer, Crouzon, Jackson-Weiss and Apert syndromes, have been found to be due to mutations in the fibroblast growth factor receptor 2 gene (FGFR2) (Jabs 1994, Reardon 1994, Wilkie 1995, Rutland 1995, Lajeunie 1995). In addition, mutations causing Pfeiffer syndrome have also been found in the FGFR1 gene (Muenk 1994). The description of these different mutations is an important reminder that the categories of dysmorphology syndromes are often based on clinical findings and that distinct clinical syndromes may arise from different mutations in the same gene.

For chromosome 7p, at least 32 cases of varying deletions have been identified and characterised. Of these, 14 are associated with craniosynostosis although there is considerable variation on the size and location of the deleted segment. Inspection of the deletions in these patients has identified two regions, 7p13-15 and 7p21-22, which appear to be consistently associated with craniosynostosis (Schömig-Spingler 1986, Kikkawa 1993, Chotai 1994). At the start of this study the more distal of these two regions, 7p21-p22, overlapped with the assigned location for PDGFA. Molecular studies have been conducted using DNA from a subset of 6 patients with

chromosome 7p deletions and craniosynostosis (Chotai 1994). Unfortunately this study did not include PDGFA, however all of the patients studied were found to be heterozygous for MS31 (*D7S21*). The importance of this finding in relation to the location of PDGFA is discussed in the light of the results of the linkage analysis in Chapter 4.

1.10.2 Greig Syndrome

Greig cephalopolysyndactyly (GCPS) is an autosomal dominant disorder which affects limb and craniofacial development. The genetic locus for this syndrome was identified as 7p13 by balanced translocations associated with GCPS (Tommerup 1983, Krüger 1989, Drabkin 1989) and by linkage analysis (Brueton 1988). From the *in situ* hybridisation data the location of PDGFA on chromosome 7p appeared to be more distal in 7p22-7pter (Bonthron 1988), thus PDGFA was not a good candidate gene for GCPS. Around the time of the start of this project this impression was confirmed when two of the three translocations described above were shown to interrupt the zinc-finger gene GL13 (Vortkamp 1991). This suggests that mutations arising in the GL13 gene account for the development of GCPS which is compatible with the role of zinc-finger binding proteins in the regulation of transcription during development.

1.10.3 Saethre-Chotzen syndrome

The Saethre-Chotzen syndrome shows a wide variety in symptoms and severity. The main presenting feature is craniosynostosis although this may be mild. Other symptoms typical of the syndrome include facial asymmetry, ptosis, hypertelorism, hearing impairment and vertebral column defects. Partial syndactyly is seen, usually

affecting the second interdigital space. The hallux is often abnormally broad and may show a valgus deformity (Reardon 1994).

Two reports have described Saethre-Chotzen syndrome in association with translocations of chromosome 7p. Reardon (1993) reported a breakpoint at 7p21.2 while Reid (1993) found the breakpoint in their patient to be at 7p22. This discrepancy between the two reports may due to the limits of cytogenetic resolution or it may be reflect an underlying genetic heterogeneity. The finding of a translocation breakpoint as far distal as 7p22 meant that this syndrome mapped to the same region as PDGFA.

An early linkage analysis study linked Saethre-Chotzen syndrome to the region between *D7S10* and *D7S370.* Importantly in this study there was no evidence for linkage to MS31 (*D7S21*) (Brueton 1992). More recently this assignment has been refined, narrowing the candidate region to between *D7S513* and *D7S516* (Lewanda 1994a) although two further linkage studies have defined separate regions within this interval (Lewanda 1994b, van Herwerden 1994). For clarity, the results from these later, comparable linkage studies are summarised in diagram 1.4. Importantly both linked regions lie proximal to *D7S513.*

It is interesting to note that the patient DSE who has been shown to have retained PDGFA (see section 1.5.4) shows no signs of craniosynostosis. At the start of this project *PDGFA* had not been placed on a linkage map so its location relative to these markers was not known. Loci *D7S21* and *D7S513* are both included in the linkage analysis in Chapter 4, the relationship of that data to the search for candidate genes for Saethre-Chotzen craniosynostosis syndrome is discussed.

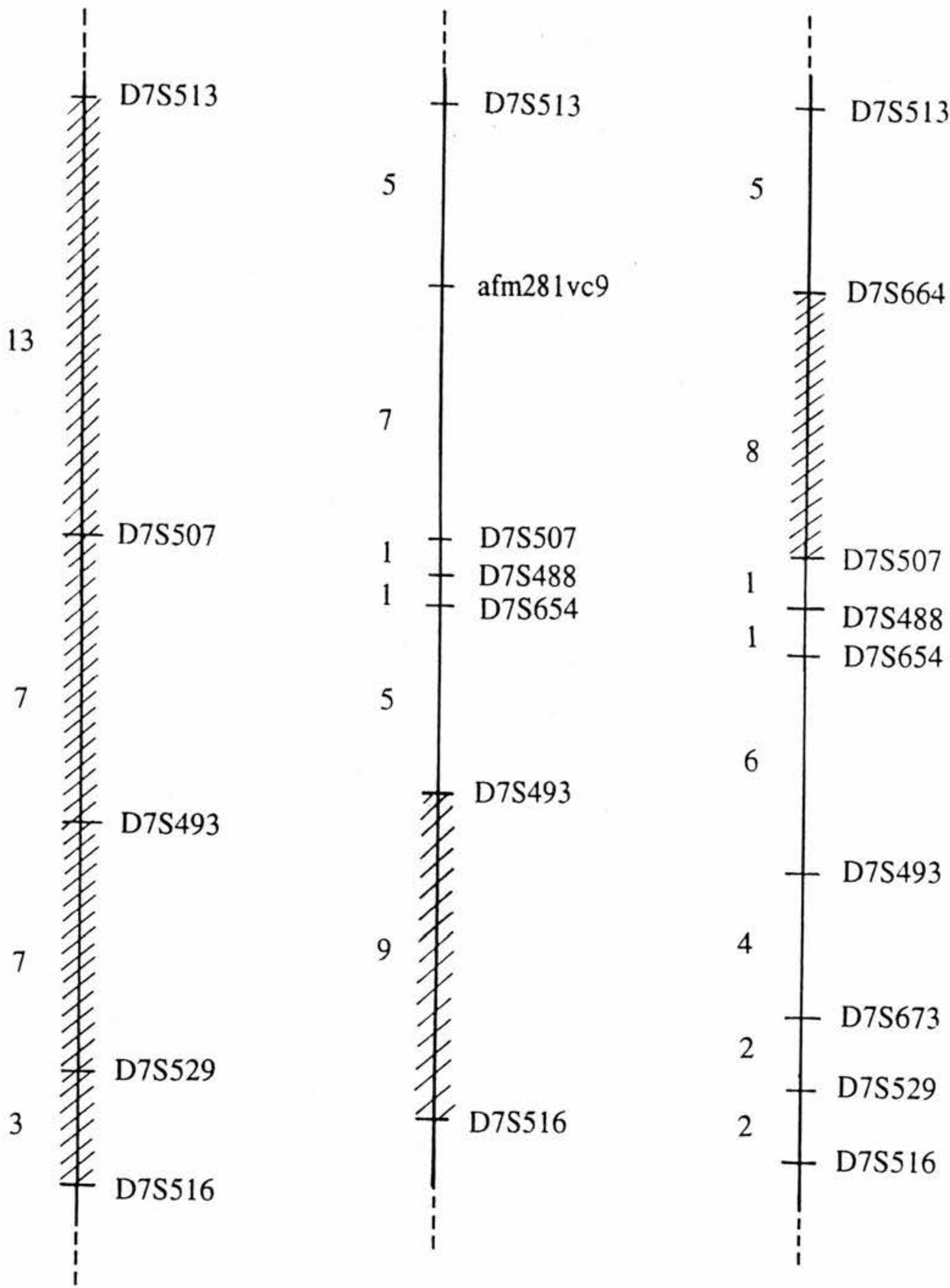


Diagram 1.4: Linkage analysis in Saethre-Chotzen craniosynostosis

These linkage maps are summarised from the linkage analysis in Lewanda (1994a, 1994b) and van Herwerden (1994), distances are in centiMorgans (cM). The hatched lines indicate the chromosomal regions to which the Saethre-Chotzen syndrome was mapped in each of these studies. These regions are not overlapping but they do all lie proximal to the marker *D7S513*.

1.10.4 Ring Chromosome 7

As discussed in section 1.5.4, the retention of PDGFA in the apparently terminal deletion in DSE gives two possible locations for the gene: in the subtelomeric region of 7p or between the breakpoint in DSE and cell line 004 Rag 1-15. If the location of PDGFA were subtelomeric than it would be expected that it may be deleted in a ring chromosome 7.

Ring chromosomes result from the deletion of varying quantities of the telomeric regions of both the long and short arms of the same chromosome followed by fusion of the two arms to each other. To date, 10 cases of ring chromosome 7 have been reported (DeLozier 1992, Tsukamoto 1993). The phenotype of affected patients is very variable but all show growth retardation and have a wide range of skin lesions including cafe-au-lait spots, haemangiomas and pigmented naevi. In one girl a malignant melanoma developed from one of these lesions (Vollenweider 1993). The presence of such skin lesions is interesting since the *patch* mutant mouse (section 1.7.1) also shows abnormalities of melanocytes in its coat colouring. One report of a patient with ring chromosome 7 was a severely affected boy who died at 20 months (Tsukamoto 1993). A neuropathological autopsy reported abnormalities including fusion of the anterior cerebral hemispheres, agenesis of the olfactory bulbs and tracts, absence of pigmentation within brainstem pigmented neurones and severe hypomyelination of the whole brain. Such developmental abnormalities are interesting when considering the chromosomal location of PDGFA since PDGF has been shown to be expressed during neuronal development. This may mean that the PDGFA gene was deleted in this patient however such conclusions can only be speculative given the unfortunate absence of any molecular studies in this report.

1.11 Approaches to gene mapping

Classically chromosomal rearrangements identified in dysmorphic patients were used to map genes to cytogenetic regions. This approach has been immensely useful for mapping many genes, however its usefulness is limited by its dependence on identifying individuals carrying suitable chromosomal rearrangements. Rearrangements of chromosome 7p and their association with disease are discussed above in section 1.10, unfortunately these are not useful for mapping PDGFA. Thus further mapping of PDGFA required approaches which would provide information at a much higher resolution.

Linkage mapping can be used to map genes both to chromosomes and relative to other loci within a defined chromosomal region. Linkage mapping is discussed below in section 1.12, it was used in this project to map PDGFA (see Chapter 4). This approach to mapping has the advantage that it can provide information at a much higher resolution than cytogenetic mapping however linkage mapping cannot provide physical data or clones from the region.

To relate the information from a linkage map to the physical landmarks seen on the chromosome and to allow a more detailed study of the region, physical mapping must be undertaken. Physical mapping is the cloning of a chromosomal region by the collection of a series of stable, easily manipulated, ordered and overlapping fragments. A number of cloning systems containing different sizes of fragments can be used. Yeast artificial chromosomes (YACs) are now a widely used cloning system. These clones can contain large fragments of human DNA allowing a large area to be mapped. The cloning vector is described below in section 1.13.2 and screening of YAC libraries for clones spanning PDGFA is described in Chapter 6.

While YAC clones provide a huge advantage for the construction of physical maps some regions of the genome appear to be unclonable in YACs. Regions which are unclonable in YACs may be cloned using other systems such as P1 clones which are

discussed below in section 1.13.3 and are used in Chapter 7. Other approaches such as BACs (bacterial artificial chromosome) and MACs (mammalian artificial chromosomes) are being developed. It is hoped that these systems will overcome some of the problems associated with YAC clones. Unfortunately these alternative systems are not yet available.

Since the DNA inserts in clones such as YACs may be rearranged it is important to compare the information from the clone contig back to genomic DNA. This is often done by comparing the pattern of restriction maps of rare cutting enzymes such as NotI derived from the YAC clone contigs and from human gDNA. This approach requires long range restriction mapping by pulsed field gel electrophoresis (PFGE) which is discussed below and undertaken for the region surrounding PDGFA in Chapter 5.

Each of these approaches to mapping genes is discussed below, along with the resources and information about chromosome 7p21-pter and PDGFA which was available at the start of this project.

1.12 Linkage analysis

Linkage analysis measures the natural occurrence of recombination or 'cross-overs' between a pair of chromosomes during meiosis. When traced through generations of the same family using a suitable marker, these recombinations events can be used to statistically infer the distance between two sites: the further apart two sites are the greater number of cross-overs will occur between them. Linkage analysis therefore has two major requirements: (i) DNA must be available from suitable families to allow the events of meiosis to be traced, (ii) Polymorphisms or markers which can

be used to identify sites along the chromosome of interest and to distinguish between the two chromosomes are required.

Classically, linkage analysis has been used to identify regions of the genome which may contain disease causing, mutated genes using DNA collected from families suffering from the disease in question. However, large numbers of random polymorphic markers are now available and these are being used to produce linkage maps for each chromosome. This approach uses linkage analysis to determine the relative order of the markers on the chromosome and to measure the genetic distance between each marker. Since these mapping projects are not attempting to identify linkage between a marker and a disease locus, they can be undertaken using normal families such as the CEPH panel of families (Dausset 1990). The panel of CEPH families which were used for the linkage analysis in this project are described in Chapter 4: Linkage analysis of PDGFA.

1.12.1 Maximum likelihood calculations and lod scores.

Two loci on different chromosomes will cosegregate, on average 50% of the time whereas two loci physically close together on the same chromosome will be inherited together more frequently. The frequency with which two loci do not segregate together is termed the recombination fraction (denoted θ). The selected polymorphic markers are used to type all the individuals from the selected families. Using this data, the likelihood of linkage between each pair of markers is calculated using a range of values for θ from 0.5 (no linkage) to 0 (no recombination). Frequently the phase of the haplotypes in crucial individuals cannot be determined and thus which individuals are recombinant will be unknown. This means that the likelihood calculation has to be made for each of the possible phase orientations of the markers. Since the likelihood value would depend on the size of the family being studied, a likelihood ratio is calculated comparing each possible value for θ against

the null hypothesis that there is no linkage (θ). This ratio is the *odds for linkage*, it indicates whether the two markers are more likely to be linked or unlinked. For convenience the logarithm of this odds ratio is used. The *log of the odds* or *lod score* for each family in the study can be added together. The mathematical and statistical details of linkage calculations are described fully in Ott (1991).

In very simple situations the maximum lod score may be calculated by hand, however the increasing complexity of a larger number of families and complicating factors such as missing information, unknown phase and unequal male and female recombination frequencies usually requires the aid of a computer for the calculations.

1.12.2 Statistical significance of linkage

The maximum lod score gives a measure of the statistical significance of the result. For autosomal loci a value greater than +3 is usually accepted as demonstrating that linkage is present. This corresponds to odds for linkage of at least 1000:1. As mentioned above, linkage analysis is increasingly being used to develop multipoint genetic linkage maps for whole chromosomes. When comparing alternative orders for two loci odds of greater than 1000:1 are generally used before the loci are placed on the map.

1.12.3 Mapping functions and genetic distance

In order to construct a genetic map, distances as measured by the recombination fraction (θ) are converted, using a mapping function, into map distances which are measured in centiMorgans (cM). Recombination events do not happen truly randomly since the occurrence of one-cross will inhibit a second cross-over occurring nearby. This phenomenon, termed interference, means that maps derived from

nearby. This phenomenon, termed interference, means that maps derived from functions which do not account for it tend to overestimate genetic distance. The most commonly used mapping function, Kosambi's mapping function, is designed to account for interference (Ott 1991).

1.12.4 The relationship between genetic and physical distance

Recombination fractions and genetic distance do not bear a direct relationship with physical distance since the frequency of recombination differs between male and female and varies with location in the genome. However, the order of the loci is the same on genetic and physical maps. Recombination is generally higher in females than in males so linkage maps for autosomal loci are usually calculated separately. As a result the female linkage map is generally longer than the male although in the telomeric regions of the chromosomes the reverse is seen. Recombination frequency has been shown to increase dramatically in some regions, termed recombination 'hot-spots', and also at the telomeres of the chromosomes (e.g. Allito 1990, Burmeister 1991).

To allow rough estimations of the correlation between genetic and physical distance, the concept of a sex averaged map of the genome was introduced. Historically the length of such a map of the human genome was calculated to be 33 Morgans or 3,300 cM. As a rough guide, the haploid genome is estimated to be 3×10^9 bp, 1cM is therefore equivalent to 1 Mb. However this comparison between physical and genetic distance is only a rough guide and may be misleading in some regions such as the chromosome telomeres (discussed further in Chapter 4).

1.12.5 Linkage maps available for Chromosome 7p

At the time of writing this thesis a number of linkage maps of chromosome 7p were available, details from four of these maps are summarised in diagram 1.5. At the start of this project, the first of these maps, Mishra (1992), was the most recently published. Four loci from this map (*D7S21*, *D7S108*, *D7S89* and *D7S103*) were used in the linkage analysis in Chapter 4. The second and third maps, Mishra (1994) and Hill (1994), were both published as 'provisional framework maps' in the report from the 1993 International Workshop on Chromosome 7 (Grzeschik 1994). However one of these maps (Hill 1994) is based on unchecked data which may result in some inaccuracies. These three maps were all constructed using the program Cri-Map. The fourth map (Gyapay 1994) is from the genome wide 1994 Généthon human genome linkage map. It was constructed using the LINKAGE program.

It is difficult to determine whether published maps use data which has been independently obtained since all of these maps use the CEPH panel of families for which databases are available. This is of interest especially since some of the maps are based on data from a small subset of the CEPH families, in the maps published by Gyapay (1994) only 8 out of the 40 families had been typed. More recently, a further two sets of linkage maps of the human genome have been published (Buetow 1994, Matise 1994). These maps are also based on collated data and do not provide any further information for the chromosome 7p region.

The two loci, *D7S21* and *D7S531*, at the telomeric end of the 7p map were found to be of particular interest in the mapping analysis in this project. Of the six linkage maps referred to above, five included both loci (Mishra 1994, Hill 1994, Gyapay 1994, Buetow 1994, Matise 1994). Two of these maps placed *D7S531* telomeric to *D7S21* (Hill 1994, Buetow 1994) while the other three could not place the two loci relative to each other (Mishra 1994, Gyapay 1994, Matise 1994). Thus it is not possible to place *D7S531* relative to *D7S21* with certainty using the

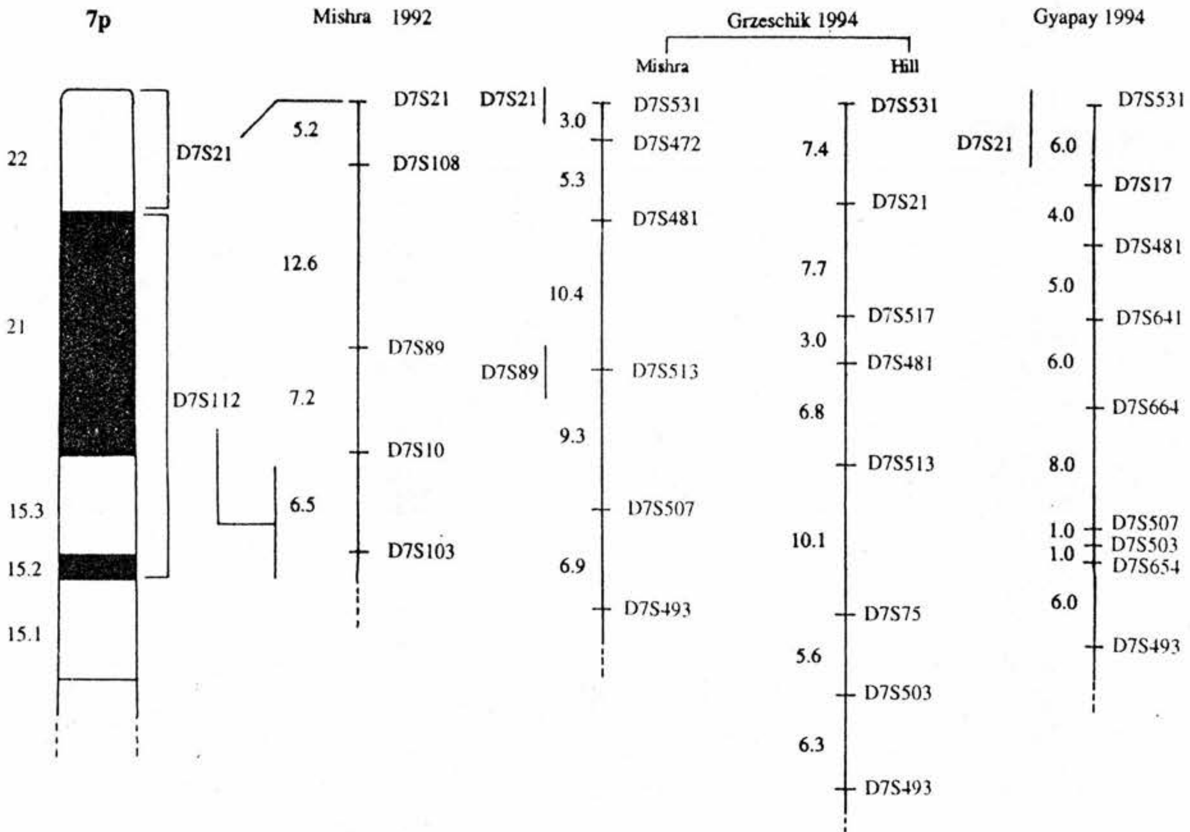


Diagram 1.5: Summary of genetic linkage maps for chromosome 7p15-7pter.

The data for the distal region of 7p is summarised from the linkage maps by Mishra 1992, Grzeschik 1994 (Mishra 1993, Hill 1993) and Gyapay 1994, distances are in cemtiMorgans (cM). The data is discussed further in section 1.12.5.

data which is available. The relative position of these two loci was found to be important in the linkage analysis of PDGFA, this is discussed further in Chapter 4.

Locus *D7S513* is also of interest since it was used in the linkage analysis of the craniosynostosis syndromes which is described above in section 1.10. The two regions which were identified as candidates regions for Saethre-Chotzen syndrome were both proximal to *D7S513*. Two of the four maps in diagram 1.5 include *D7S513*. This locus is also included in the Eurogem chromosome 7p linkage map (diagram 4.8, Badbanchi 1994). Thus the location of PDGFA is mapped relative to the candidate regions for Saethre-Chotzen craniosynostosis in this project.

1.13 Physical mapping

The different approaches to physical mapping which were used to map PDGFA are listed above in section 1.11, the approaches of PFGE restriction mapping, the YAC cloning vector and library screening and P1 clones are described below.

1.13.1 Long range restriction mapping

Long range restriction mapping uses 'rare-cutting' restriction enzymes to produce a map. The average fragment sizes produced by such 'rare-cutting' enzymes are generally in the range 100 - 1000 kb. Such fragments are too large to be resolved by conventional agarose gel electrophoresis which uses a static electrical field and can separate fragments up to 20 - 30 kb in size. Periodically changing the direction of the electrical field in pulses during electrophoresis has been found to allow the resolution of much larger fragments of DNA (Schwartz 1984). Changing the direction of the electric field during electrophoresis forces the DNA molecule to change its conformation and reorientate in the direction of the new electrical field before it can migrate in the new direction. The time required for the molecule to

reorientate is dependant on the size of the molecule, thus separation of DNA molecules on the basis of their size is possible.

The PFGE work described in this project was produced using the CHEF (contour clamped homogeneous electric field) system which separates the DNA in straight lanes in contrast to the curved lanes produced by earlier systems (Chu 1986). This is achieved by dividing the voltage from the power supply between 24 electrodes in an hexagonal array, creating a voltage which is constant across the gel. Earlier apparatus, such as that originally used by Schwartz and Cantor (1984), used a gradient electric field which caused the lanes of DNA to run at an angle. It has been found that a 90° reorientation angle results in a large increase in the mobility of the DNA molecules causing a decrease in their resolution. The hexagonal array of electrodes in the CHEF apparatus is selected to create a 120° reorientation angle.

1.13.2 Restriction enzymes used for pulsed-field mapping

The majority of 'rare-cutting' restriction enzymes have recognition sites that contain the dinucleotide CpG. The dinucleotide CpG is strikingly rare. It is present in the genome approximately 5 times less frequently than would be expected on a random basis. This rarity of CpG dinucleotides is thought to be a consequence of methylation: cytosine is prone to deamination and when methylated this gives rise to thymine. DNA repair mechanisms appear to fail to recognise and correct this mutation. About 60 -90% of CpG dinucleotides are methylated and this high level of methylation in vertebrate DNA seems to have led to the evolution of a genome which is depleted for this dinucleotide (reviewed in Bird 1985, 1986).

The pattern of methylation of CpG dinucleotide in the genome is not uniform. A specific site may be only partially methylated and different regions of the genome are

methylated to different degrees. An unusually low frequency of methylation is often found in the region at the start or 5' end of many genes. These hypomethylated areas are referred to as CpG islands and are defined as having a high G and C content (>50%). In contrast to the rest of the genome, the dinucleotide CpG is found to be present at a level which would be expected on a random basis. It is thought that the low level of methylation in these regions is responsible for the high CpG content by avoiding the mutation of cytosine to thymine. CpG islands are found at the 5' end of many gene including all 'housekeeping' genes. The region is often large (500 - 2000 bp) and may include the first few exons of the gene as well as upstream sequences (Bird 1986, Larsen 1992).

The functional effect of these islands on restriction mapping projects is two fold. These regions may be useful, an area with a high frequency of unmethylated CpG dinucleotides may indicate the start of an unknown gene (reviewed in Bird 1987). However their presence may also frustrate restriction mapping projects around known genes. If the CpG island contains restriction sites for all enzymes in a cluster at the 5' end of the gene it may not be possible to extend the restriction map. For this a probe which lies beyond the island will be required.

1.13.3 The YAC cloning vector

The yeast artificial chromosome (YAC) vector comprises the basic functional units of yeast chromosomes: a centromere, telomeres and a replication origin (ARS) (Burke 1987). The most widely used vector is pYAC4 (diagram 1.6). Itwhich allows the direct cloning of fragments produced by EcoRI partial digests into the cloning site in the *sup4* gene (Burke 1987). Restriction of the vector with EcoRI and BamHI generates three fragments, two of these fragments form the arms of the YAC vector. The BamHI ends are recognised by *S. cerevisiae* and are 'healed' into functional telomeres by the addition of simple sequences (reviewed in Burke 1987, Riley 1992).

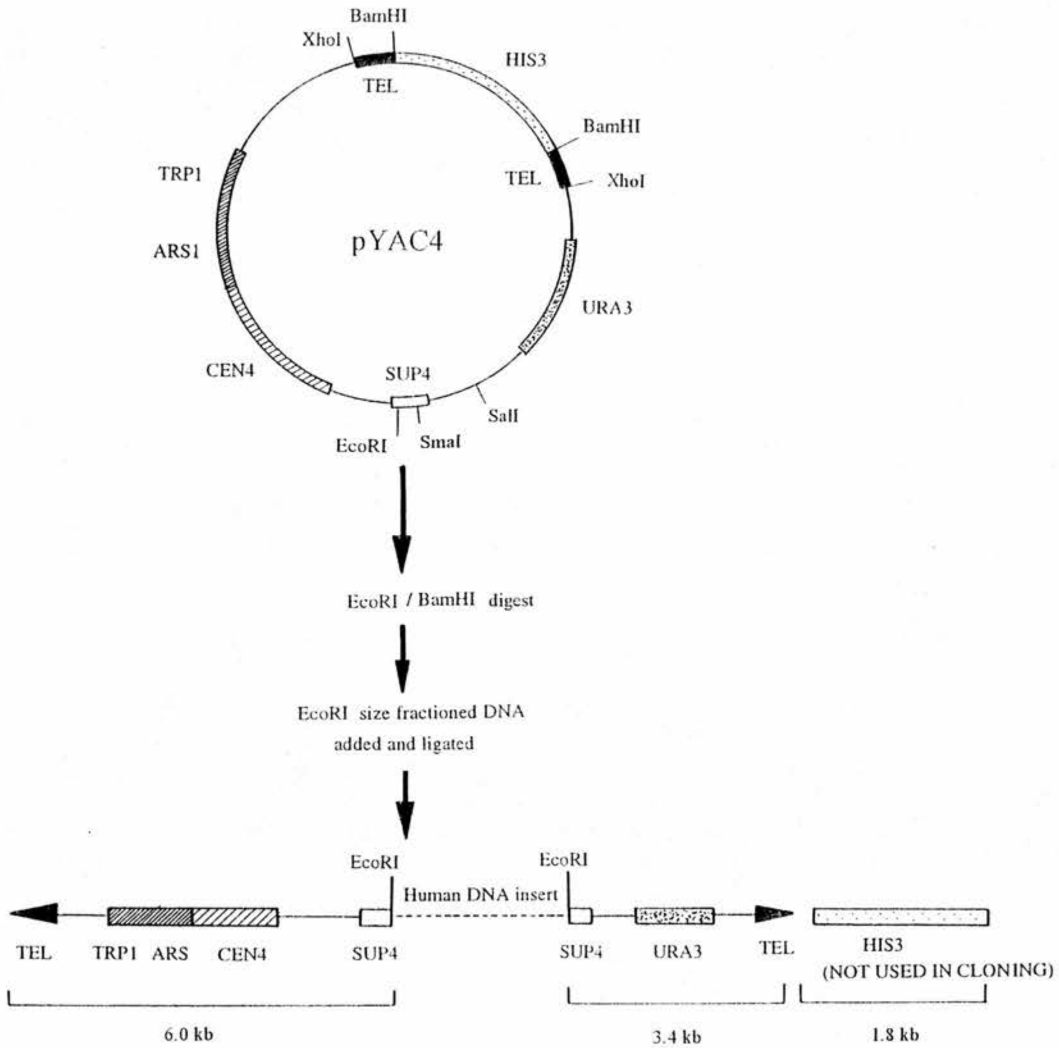


Diagram 1.6: The YAC cloning vector pYAC4

The yeast artificial cloning vector pYAC4. Restriction sites used for the construction of YACs and the structure of the final artificial chromosome are shown. Singles lines represent pBR322-dervived sequence while the boxed regions are yeast-derived. Adapted from Riley (1992).

The recombinant vector is transformed into yeast cells where it is maintained as a single copy artificial chromosome. Large DNA fragments, up to 1 Mb, can be cloned using YACs. This major advantage means that they are now the most widely used vector for large scale mapping projects. The large clone sizes mean that YAC vectors have been used to develop a series contigs spanning the entire length of human chromosome 21 (Chumakov 1992). However the isolation of vector DNA from the yeast genome is difficult in comparison to the easy isolation of cloned DNA of plasmids This makes the subsequent analysis of the cloned fragments much more difficult.

YAC clones have been used to develop several human genome clone libraries, these are described below in section 1.14. To identify YAC clones which span PDGFA, these libraries were screened using probes and PCR assays from the gene. The YAC libraries and the screening experiments are described in Chapter 6 of this thesis.

1.13.4 Telomere YAC vectors

Several modified vectors have been developed incorporating different cloning sites and selectable markers. One of these vectors, pTYAC1 has been developed to clone telomere specific fragments (Riethman 1989). This telomere YAC vector has a single yeast telomere and cannot function as a linear chromosome in the yeast host cell until it is complemented with a second telomere acquired during the cloning procedure. The telomere YAC vector, pTYAC1, consists of one vector arm which carries the essential elements for cloning and replication such as selectable marker *ura3* , the ARS, a telomere and a functional centromere. The second arm of the recombinant artificial chromosome is provided by the cloned DNA insert. The aim of the cloning procedure is that this DNA insert will consist of the telomere and subtelomeric sequence of the genome being cloned.

However, yeast is capable of 'healing' chromosomes by the addition of telomeric repeat sequences. This means that sequences which are not telomeric in origin can be cloned, healed and function as telomeres in the yeast cell. Thus it is important to screen colonies for the presence of the human telomere repeat (TTAGGG)n before the YAC can be confirmed as a telomere clone. Typically 1-5 % of transformants will hybridise to human telomere probe and contain DNA which may be of telomeric origin (reviewed in Dobson 1992).

Following the results of the linkage analysis and from screening the YAC libraries, YAC clones for the chromosome 7p telomere were sought. The telomere YAC library and the analysis of the candidate clones for 7p is described in Chapter 8: Telomere YACs.

1.13.5 The P1 cloning vector

The cosmid cloning vector has a high cloning efficiency and the cloned DNA is easy to isolate. However it is limited by the small size of DNA inserts (up to 40 kb) which the vector can carry. The development of yeast artificial chromosomes (YACs) allowed DNA inserts up to 1 Mb to be cloned (Burke 1987), however the YAC cloning system suffers from a number of inherent problems: its cloning efficiency is low, the cloned DNA is more difficult to isolate and many clones contain chimeric cloning artefacts. The bacteriophage P1 cloning system can carry DNA inserts of up to 95 kb (Sternberg 1990) and can be analysed using conventional techniques avoiding the laborious techniques required for YAC clones, thus it complements these two widely used cloning systems.

The original P1 cloning vector was pAd10, which was developed by Sternberg (1990). However it is the second generation vector, pAd10-sacBII, which is the vector of choice (Pierce 1992). It carries the normal vector features such as a

selectable kanamycin marker to allow positive selection of recombinants. The ICRF human P1 clone library is described below. It was screened for P1 clones spanning the PDGFA locus (described in Chapter 7).

1.14 Physical mapping resources

At the start of this project there was very little physical mapping information available for the short arm of chromosome 7. The majority of mapping projects had concentrated on 7q which carries CFTR, the gene responsible for cystic fibrosis.

A somatic cell hybrid panel of cell lines containing fragments of chromosome 7 was available (Vortkamp 1991). A subset of cell lines from this hybrid panel had been used to map PDGFA to 7p21-pter as discussed above in section 1.5.4. The most distal breakpoint known in chromosome 7p is the breakpoint of the X;7 translocation: 46,XX,t(X;7)(q21;p22) in the cell line GM 1696. A series of 148 probes were mapped to a panel of translocated chromosomes including GM 1696, of these probes only 2 mapped to the distal region defined by this breakpoint (Vortkamp 1991). This cell line was not included in the panel of hybrids which were screened for PDGFA (Bonthron 1992) so at the start of this project it was not known whether PDGFA was retained on chromosome 7 in GM 1696.

A number of whole genome YAC clone libraries have been developed and are available for screening. Those libraries which were screened for PDGFA in this project are described below. One YAC library, the CEPH YAC library has been used to develop a genome wide physical map of YAC contigs (Cohen 1993). This mapping project found particularly poor coverage of YAC clones in the telomeric regions of the chromosomes and unfortunately this map does not cover the region 7p21-7pter to which PDGFA has been localised. Thus this library was not screened for PDGFA.

1.14.1 The ICI YAC Library

The ICI YAC library (Anand 1990) contains about 35 000 YAC recombinants whose average insert size is 350 kb. This gives a theoretical coverage of 3.5 copies of the human genome. The presence of chimeric clones has been estimated to be about 10%. The library is available both gridded onto nylon filters for screening by hybridisation and as a series of pooled clones which can by screened by PCR. The gridded filters have the individual clones laid out on 40 nylon filters of 12 x 8 cm. In contrast the PCR screening approach uses 40 master or primary pools which contain 864 clones each.

The gridded filters were supplied as a gift from the Human Genome Mapping Project (HGMP) Resource Centre, London. Two sets of primary pools for PCR screening were also obtained, one set were a kind gift from K. Dry and A. Wright (MRC Human Genetics Unit, Edinburgh), the other set were a gift from J. Warner (Edinburgh University Human Genetics Unit, Edinburgh).

1.14.2 The CEPH 'megaYAC' Library

The CEPH 'megaYAC' library contains 23 808 recombinants with a much larger average insert size of 918 kb. It has a significantly higher level of chimeric recombinants at 30 - 40 % although this level is thought to be regionally dependent. This library was supplied as a set of pooled clones for screening by PCR as a kind gift by M. Jones (Cambridge).

1.14.3 The Washington University total human DNA YAC library

This human total genomic DNA YAC library was described by Brownstein (1989). It contains an estimated 60 000 clones with an average insert size of 275 kb. This is smaller than the average insert size of other libraries such as the ICI library (Anand 1990). The frequency of chimeric clones is higher at 40 -50 %. Primer pairs for STS's in the PDGFA gene were sent by E. Green (National Centre for Human Genome Research, NIH, Bethesda, USA) who screened this library.

1.14.4 Chromosome 7 YAC libraries

The first report of the international workshop on human chromosome 7 in 1993 reported the development of two human chromosome 7 specific YAC libraries (Grzeschik 1994). Obviously these were of great interest in the attempts to identify clones for the region surrounding PDGFA.

1.14.4.1 The Toronto human chromosome 7 library

The Toronto chromosome 7 YAC library was initially developed using a chromosome 7 somatic cell hybrid cell line (Scherer 1992, 1993). However in the course of the development of the YAC library the human chromosome 7 in the cell line was found to have been rearranged. This involved the duplication of a segment of 7q which had been inserted into the short arm of the same chromosome. As a result the chromosome was deleted for part of the short arm of chromosome 7. A second chromosome 7 hybrid cell line was used in an attempt to generate more YAC clones from 7p, however this cell line was also found to have lost a portion of chromosome 7p (Kunz 1994). Thus this library is relatively depleted for the distal region of short arm of chromosome 7 (see diagram 1.7).

Diagram 1.7: The Toronto chromosome 7 YAC library

A summary of the localisation of 725 YACs from the Toronto human chromosome 7 YAC library which have been mapped to specific chromosomal regions by FISH. From Kunz 1994.

This is not ideal for screening for clones of PDGFA since this depleted region is the region of the chromosome to which the gene has been mapped. The library contains over 1000 clones with an average insert size of about 475 kb. Over 750 of these YAC clones have been mapped to chromosome 7 using fluorescent *in-situ* hybridisation (Kunz 1994).

1.14.4.2 The Washington University human chromosome 7 YAC library

The second chromosome 7 library which was screened for PDGFA was constructed using YACs derived from a human chromosome 7 somatic hybrid cell line and chromosome 7 enriched YACs from the CEPH 'megaYAC' library with the addition of a small number of other YACs known to be from chromosome 7 (Green 1994). The library has an estimated coverage of chromosome 7 of 6.5 - 7.5 times. Using this library, 117 genetic markers have identified 650 YACs which are contained in 80 contigs which are estimated to cover 60 - 85 % of chromosome 7. A integrated map has been constructed which compares the 1993 Généthon linkage map of chromosome 7 (Gyapay 1994) to contigs constructed from YACs in this library (Green 1994) this is shown in diagram 1.8. As can be seen in the diagram, the YAC clone coverage of chromosome 7 is heavily concentrated on the long arm of the chromosome. The distal region of the short arm which contains the PDGFA gene is much less fully mapped.

Diagram 1.8: YAC clone contigs from the Washington University YAC library

The 1993 Genethon linkage map of human chromosome 7 is depicted (Gyapay *et al*), the dashed line brackets indicate where markers on this map are placed with <100:1 odds. On the far right the vertical brackets indicate the YAC contigs which have been developed from the Washington university chromosome 7 YAC library. The thin brackets indicate where a single YAC has been shown to link two or more STSs.

1.14.5 The ICRF human gDNA P1 Library

The ICRF P1 library was constructed using genomic DNA from the human lymphoblastoid cell line GM1416B (Human Genetic Cell Repository, Camden, N.J.), karyotype: 48XXXX, using the vector pAd10SacBII which is described above in section 1.13.5. (Francis 1994). The library was constructed and supplied as a gift by F. Francis (ICRF, London). As for many YAC libraries, the human gDNA used for cloning was prepared and digested (using MboI) in agarose and size-fractionated by pulsed-field electrophoresis. Handling the DNA in agarose reduced the risk of damage from shear forces, optimising the size of cloned inserts. The average size of the cloned inserts in the library is about 75 kb.

The library contains about 47 000 clones, giving a theoretical genome coverage of at least one times. It is gridded onto two sets of duplicate filters, thus four filters of 21 by 21 cm were supplied. The two filters which contained duplicate copies of the clones were not gridded in the same order so matches between the two sets of filters could only be identified by the scientists at ICRF. Each filter contained an array of 144 x 144 clones arranged in 48 x 48 boxes each of which contained 9 P1 clones in a 3 x 3 array. The screening of the P1 library for clones spanning PDGFA is described in Chapter 7.

1.15 Objectives

At the start of this project a number of objectives were established. They were:

1) To further sequence the PDGFA gene looking for useful polymorphic markers and information on the structure of the gene in the light of the unusual features which have been described.

2) To map PDGFA by linkage analysis using new polymorphic markers identified in the sequencing project. A location for PDGFA relative to other markers in the distal regions of chromosome 7p may provide further information about the disease syndromes associated with this region and the function of PDGFA in processes such as neoplasia.

3) To develop a long range restriction map of the region spanning PDGFA using pulsed field gel electrophoresis. Such a map would be important for the development of a YAC clone contig and would provide further information about the region surrounding the gene.

4) To screen human YAC chromosome and P1 libraries for clones spanning PDGFA and to use these to develop a clone contig. An aim of such a physical mapping project would be to physically link PDGFA to other landmarks in the region such as flanking loci or the telomere. The development of a physical map was also of interest since such information about this region is sparse in comparison to, for example, the long arm of chromosome 7.

5) In the light of developments as this project proceeded, a further aim became increasingly important. That was to provide a measure of the physical distance between PDGFA and the chromosome 7p telomere using Rec-A assisted restriction endonuclease (RARE) cleavage technique.

2. Chapter 2: Materials and methods

2.1 Suppliers of chemicals and enzymes

The companies which were routinely used for supplies of chemicals and enzymes are detailed in appendix I.

2.2 Bacterial cell culture: Media and additives

Luria Bertani (LB) medium.

LB medium consists of 5 g/L Bacto yeast extract, 10 g/L Bacto tryptone, 0.17 M NaOH adjusted to pH 7.0. It was used routinely for the culture of plasmid containing *E.coli*. In addition, LB agar plates contained 15 g/L agar which was added to the media prior to autoclaving.

Terrific Broth.

Terrific broth (24 g/L Bacto yeast extract, 12 g/L Bacto tryptone, 4 g/L glycerol, 10 mM KH_2PO_4, 72 mM K_2HPO_4) is a richer medium which was used to give high yields of plasmids from overnight culture for large scale isolation of plasmid DNA. The 'terrific' broth and the potassium phosphate buffer were prepared and autoclaved separately, the buffer was added to the broth immediately before use

TYM Broth

TYM broth contained 5 g/L Bacto yeast extract, 12 g/L Bacto tryptone, 100 mM NaCl and 10 mM $MgCl_2$. It was used for cell culture during the preparation of calcium competent cells.

SOC medium

SOC medium was used to resuspend electro-competent cells after transformation with plasmids. It contains 5 g/L Bacto yeast extract, 20 g/L Bacto tryptone,

8.5mM NaCl, 2.5 mM KCl and 20 mM glucose, adjusted to pH 7.0. The media and glucose were autoclaved separately and mixed prior to use.

2YT medium

2YT medium (10 g/L Bacto yeast extract, 16 g/L Bacto yeast tryptone, 85 mM NaCl pH 7.0) was used to culture the P1 clones described in Chapter 7.

Media was sterilised by autoclaving for 20 minutes, at 121°C. Where appropriate, ampicillin, at 100 μg/ml or kanamycin at 50 μg/ml was added to media immediately before use. *E.Coli* were grown at 37°C, on a shaker for aeration of liquid cultures. Stocks were stored at -80°C in 15% (v/v) glycerol.

2.2.1 Plasmid cloning vectors

pBluescript II KS(+/-) (Stratagene)

This 2961 bp phagemid derived from pUC19 was used to produce the plasmid pACY7 which was used in the sequencing project described in Chapter 3. The KS designation indicates that the polylinker in this plasmid is orientated such that lacZ transcription proceeds from the KpnI restriction site to the SacI site.

pPCRII

The vector supplied in the TA cloning kit (Invitrogen) carries single 3' T overhangs. The procedure used in this kit for cloning PCR products takes advantage of the non-template dependant activity of *Taq* polymerase which adds a single deoxyadenosine to the 3' end of all duplex DNA molecules produced in the PCR reaction. This kit was used to clone the PCR products from the intron 3 minisatellite repeat described in Chapter 3.

PCR-Script

This cloning vector, which is supplied in the PCR cloning kit (Stratagene), is based on pBluescript II SK(+). The SK multicloning site has been modified to contain an SrfI restriction endonuclease target sequence. This kit was used to clone telomere YAC ends rescued by Vectorette PCR.

2.2.2 Yeast cell culture

YPD medium

YPD medium (10 g/L Bacto yeast extract, 20 g/L Bacto tryptone, 20 g/L glucose) was used to culture *S. cerevisiae* strain YP148. This strain of cells was a gift from Dr F. Richards, Cambridge. It was used as a size marker in the pulsed field gel electrophoresis experiments since it carries an additional chromosome of 75 kb which provides a useful additional size marker.

Ura^- medium

A medium which lacked uracil was used for the selective growth of YAC recombinants. It contained 6.7 g/L Bacto yeast nitrogen base, 20 g/L glucose, 55 mg/L adenine, 14 g/L casaminoacids. The tyrosine, glucose, adenine and casaminoacids were autoclaved separately and mixed with the YNB (yeast nitrogen base) after the liquid had been allowed to cool.

Agar plates also contained 15 g/L agar which was added to the medium prior to autoclaving. All yeast were grown at 30°C, with aeration on a shaking platform for liquid cultures.

2.3 Transformation and propagation of plasmid DNA

2.3.1 Preparation of competent cells

To prepare calcium competent cells for transformation, 100 ml of fresh TYM was inoculated with 1 ml of an overnight culture of DH5α cells and incubated at 37°C, 300 rpm for 2-3 hours, until the optical density at 550 nm was between 0.5-1.0. The cells were recovered and then resuspended in 30 ml of TfbI (30 mM $KC_2H_3O_2$, 50 mM $MnCl_2$, 100 mM KCl, 10 mM $CaCl_2$, 15% (v/v) glycerol, pH 5.8). This was left on ice for 10 minutes, centrifuged to recover the cells and resuspended in 4 ml of TfbII (10 mM MOPS pH 7.0, 75 mM $CaCl_2$, 10 mM KCl, 15% (v/v) glycerol) divided into 200 μl aliquots and stored at -80°C.

2.3.2 Transformation of calcium competent cells

The 200 μl aliquots of competent cells were thawed on ice with up to 100 ng of plasmid DNA. The plasmid/cell mixture was left on ice for 30 minutes and then placed in a 37°C water bath for 120 seconds to heat shock the cells. To this, 800 μl of pre-warmed TYM were added, mixed gently and the cells were incubated at 37°C for 45 to 60 minutes. The bacteria were then plated out on LB agar plates containing ampicillin.

2.3.3 Preparation of electro-competent cells

Electro-competent cells were prepared using a method adapted from Dower (1988). A litre of LB was inoculated with 10 ml of a fresh overnight culture of DH5α cells. This culture was incubated at 37°C with vigorous shaking until the optical density at

600 nm was in the range 0.5 to 0.7. The culture was chilled on ice, the cells were recovered and then resuspended in 1 litre of 10% (v/v) glycerol at 4°C. The cells were washed in successively decreasing volumes of 500 ml, 60 ml and finally 2 to 3 ml of 10% glycerol. Aliquots of 40 μl were dispensed, frozen in a dry-ice/methanol freezing bath and stored at -70 °C.

2.3.4 Electro-transformation of competent cells

The 40 μl aliquots were thawed on ice and mixed with about 1 ng of DNA. They were transferred to a pre-cooled cuvette (0.2 cm) and placed between the electrodes in the Gene Pulser (BioRad). Pulses were delivered at 2.5 kV using a 25 μF capacitor and a 200 Ω resistor. Immediately after pulsing, 1 ml of SOC was added. The cell suspension was incubated at 37°C, 200 rpm for an hour and then spread onto LB/ampicillin agar plates.

2.4 DNA extraction and preparation

2.4.1 Small scale preparation of plasmid DNA

Three different methods were used for the small scale isolation of plasmid DNA from LB medium cultures:

2.4.1.1 Alkali lysis extraction of plasmid DNA

Small scale preparations of plasmid DNA were extracted using the alkali lysis method described in Sambrook (1989). The cells were recovered from 2 ml of an overnight

culture and resuspended in 100 μl Solution I (50 mM glucose, 25 mM TrisHCl, 10 mM EDTA, pH 8.0). To this, 200 μl of Solution II (0.2 M NaOH, 1% SDS) were added and the tube was mixed well. On ice, 150 μl Solution III (3 M potassium, 5 M acetate) were added, the sample was mixed and left on ice for 5 minutes. The lysate was centrifuged, the supernatant was recovered and the DNA was extracted with phenol/chloroform and precipitated in an equal volume of isopropanol (IPA). The DNA pellet was resuspended in TE (10 mM TrisHCl, 1 mM EDTA pH 8.0) containing RNase at 20 μg/ml.

2.4.1.2 Promega Magic Minipreps

A Promega 'magic minipreps' kit was used as an alternative miniprep method on occasions. DNA extraction using this kit is based on the alkali lysis method above to lyse the bacterial cells. The DNA is extracted from the lysate by binding it to a resin column, purified in a salt/ethanol wash while bound to the column and then eluted in dH_2O or TE.

2.4.1.3 Phenol/chloroform extraction of small scale DNA preparations

Cells from 1.5 ml of an overnight culture were suspended in 100 μl TE. To this, 100 μl buffered phenol/chloroform in a 1:1 ratio were added and the mixture was vortexed at high speed to lyse the cells. The phenol was buffered in 0.1 M TrisHCl (pH 8.0) with the addition of 0.1% hydroxyquinoline. After centrifugation, the aqueous layer was recovered and precipitated in 0.25 volumes 10 M ammonium acetate and 2 volumes of ethanol. The DNA pellet was then suspended in 50 μl TE/RNase (20 mg/ml).

2.4.2 Large scale preparation of plasmid DNA

For large scale cultures, the plasmid DNA was incubated overnight at 37°C in 500ml 'Terrific' broth. The DNA was extracted by alkaline lysis (section 2.4.1) using 10 ml Solution I, 20 ml Solution II and 15 ml Solution III per 500 ml starting culture.

The plasmid DNA was purified further with LiCl and PEG (polyethylene glycol). To remove high molecular weight RNA, 0.33 volumes 5 M LiCl at 4°C was added to the DNA and the sample was centrifuged at 14 000 rpm for 10 minutes at 4°C. The supernatant was transferred to a fresh tube, the DNA was precipitated in an equal volume of IPA, resuspended in TE/RNase (20 mg/ml) and then mixed with an equal volume of 1.6 M NaCl containing 13% PEG. The DNA was recovered by centrifugation, resuspended in TE and extracted with phenol/chloroform.

2.4.3 Total yeast DNA extract

An extract of total yeast DNA including the YAC was prepared from the recombinant clones in liquid form for analyses in which large DNA fragment size was not critical.

A streak of yeast colonies were used to inoculate Ura^- medium and were incubated at 30°C on a shaking platform until the OD_{600} was in the range 0.8 to 1.4. The cells were harvested, resuspended in 0.9 M sorbitol, 0.1 M EDTA, 0.4 mg/ml Zymolyase (20T) and then incubated at 100 rpm, 37°C for 1 hour. The cells were pelleted, washed by resuspending in lysis buffer (0.9 M sorbitol, 50 mM TrisHCl pH 7.4, 50 mM EDTA) and recentrifuged. The cell pellet was suspended in lysis buffer with 1% SDS, 0.13 mg/ml Proteinase K, and incubated at 37°C overnight. The total yeast DNA was extracted from the lysed cell suspension by phenol/chloroform extraction.

2.4.4 P1 clones DNA extraction

The P1 clones were supplied as a gift by F. Francis (ICRF, London). A single colony was used to inoculate 25 ml 2YT, 50 μg/ml kanamycin and was incubated at 200 rpm, 37°C overnight. This was used to inoculate 800 ml 2YT/Kanamycin and the incubation was continued for 1 hour. IPTG was added to a final concentration of 1 mM and the cells were incubated for a further 4 hours. The cells were harvested and the DNA was extracted by alkali lysis (see section 2.4). The P1 lytic replicon in the plasmid is regulated by the lac operon promoter which is repressed in cells containing the lac Iq repressor. The addition of IPTG releases this repression and allows the copy number of the plasmid to increase. This method yielded 115 to 150 μg of DNA from 1 litre of medium.

2.4.5 Geneclean

To remove PCR primers and unincorporated nucleotides from PCR products and to extract DNA from agarose the 'Geneclean' method was used (Vogelstein 1979). DNA in agarose was diluted in 3 to 4 volumes of 6 M NaI and incubated at 55°C to melt the agarose. Usually 5 μl of 'Glassmilk' was added to the suspension, mixed well and left on ice for 20 minutes to allow the DNA to bind to the silica matrix. The glassmilk/DNA mixture was pelleted by centrifuging and then washed three times by resuspension in 0.2 M NaCl, 20 mM TrisHCl pH 7.2, 2 mM EDTA, 50% ethanol, finally the DNA was eluted in dH_2O.

2.4.6 Extraction of gDNA from whole blood

The unclotted blood samples were mixed with 3-4 volumes of 'blood buffer': 155 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA (pH 8.0) and left at 4°C for at least an hour to lyse the red blood cells. The lymphocytes were collected and washed by resuspending in BB. If the pellet was red in colour, the cell lysis was repeated. The cell pellet was then washed in 75 mM NaCl, 25 mM EDTA (pH 8.0), pelleted and suspended in 0.2 volumes of 10 mM TrisHCl pH 8.2, 0.4 M NaCl, 2 mM EDTA with 0.3 mg/ml Proteinase K and 0.67% SDS. This was incubated overnight at 37°C.

The sample was then mixed well with 0.33 volumes of saturated (~ 6 M) NaCl and centrifuged at 5 000 rpm for 15 minutes. The supernatant was collected and mixed gently with two volumes of ethanol. As the sample was mixed the DNA precipitated at the interface and was spooled out using a sealed glass pipette.

2.4.7 Preparation of DNA agarose blocks for PFGE

Generally when undertaking a restriction digest the size of starting DNA should be 3 to 5 times larger than the desired restriction fragment. Pulsed-field gel electrophoresis (PFGE) generally aims to separate DNA fragments sized from 50 kb to 5 Mb. However, conventional methods such as pipetting and vortexing introduce shear forces which damage the DNA. These methods produce DNA molecules which are on average 200 to 300 kb in length, thus DNA prepared in solution is not suitable for PFGE. To protect DNA from mechanical breakage it is prepared by embedding the cells in agarose followed by enzymatic digestion of the non-DNA cell components.

A pellet of lymphocytes was prepared in 'blood buffer' from a 50 ml fresh blood

sample (see section 2.4.6), and then suspended in 5 ml PBS (phosphate buffered saline). The cell concentration was estimated by counting in a haemocytometer and adjusted to give a final concentration of 2.5 x 10^7 cells/ml. The PBS suspension of lymphocytes was warmed to 37°C, mixed with an equal volume of pre-warmed 2% molten LMP agarose in PBS and pipetted into the mould wells. The agarose blocks were allowed to set and were then pushed into 0.5 M EDTA, 1% SDS, 1 mg/ml Proteinase K. The blocks were incubated at 55°C for at least 48 hours, the buffer was changed after 24 hours. This digests virtually all the unwanted cell components leaving the DNA suspended in the agarose block.

To remove all traces of the SDS and Proteinase K, the agarose blocks were washed at least 10 times in 10 volumes of TE (pH 8.0) for 30 minutes at room temperature. Since the blocks were washed extensively it was not found to be necessary to use PMSF to inactivate the Proteinase K. The washed agarose blocks were stored for several weeks in 10 to 20 volumes of TE at 4°C. For longer term storage the blocks were stored in 0.5 M EDTA, before these blocks were used they were washed 10 times in 10 volumes of TE.

2.4.8 Preparation of yeast and YAC chromosome DNA for PFGE

The yeast chromosomal DNA was prepared according to the same principles as for human gDNA, however before the yeast cells were suspended in agarose they were treated with Zymolyase to remove the cell wall which cannot be digested by Proteinase K. A 100 ml culture was incubated until the optical density at 600 nm was in the range 0.8 to 1.4. The cells were harvested, suspended in 5 ml 1 M sorbitol, 2 mg/ml Zymolyase (20T) and incubated for 1 hour at 37°C, 100 rpm. The cells were collected by centrifuging (2 500 rpm for 10 minutes), and resuspended in 1 M sorbitol, 100 mM EDTA (pH 8.0) to a final concentration of 1 x 10^9 cells/ml. This cell suspension was used to make agarose blocks (section 2.4.7).

2.5 Restriction endonuclease digestion

Digestion of DNA in solution with restriction endonucleases was conducted using the buffer and temperature recommended by the supplier. Plasmid DNA and PCR amplified DNA were incubated for 1 to 4 hours using 2 to 6 units/μg DNA, gDNA was digested overnight using 4 to 8 units of enzyme/μg DNA. The reactions were stopped by the addition of 0.1 volume of 10 x BPB loading buffer (see section 2.6.2).

2.5.1 Restriction endonuclease digestion of gDNA for PFGE

Agarose blocks for PFGE were 80 μl in volume, containing about 10 μg of gDNA. Before digesting the DNA the blocks were equilibrated at room temperature by 3 washes of 45 minutes in 2 ml of the appropriate buffer (at 1x concentration). Restriction digests were incubated overnight in 200 μl of 1 x buffer using 80 to 100 units of restriction enzyme. For each buffer that was used a negative control, containing no enzyme, was set up.

2.6 Separation of DNA molecules by electrophoresis

2.6.1 Solutions and buffers

Tris acetate EDTA (TAE) Buffer: The majority of agarose gels were prepared and run in 1x TAE buffer (0.04 M Tris acetate, 0.02 M EDTA pH 8.0).

Tris borate EDTA (TBE) Buffer: TBE buffer (1x: 0.09 M Tris borate, 0.02 M

EDTA pH 8.0) was used at 1x concentration to run polyacrylamide sequencing gels and at 0.5x concentration for pulsed field gel electrophoresis.

2.6.2 Loading buffers

10 x Bromophenol blue (BPB): The majority of DNA samples which were separated in agarose were loaded into the wells of the gel using a simple bromophenol blue loading buffer. A 10x concentrate contained 0.25% bromophenol blue and 30% glycerol. On occasions the marker dye xylene cyanol was added, also to a final concentration of 0.25%.

Ficoll orange: Bromophenol blue dye absorbs UV light and may mask the presence of DNA fragments with which it comigrates, thus when very small DNA fragments were expected, 0.25% orange G in 15% Ficoll was used as a loading buffer.

2.6.3 Size markers

1 kb ladder (Gibco-BRL): This size marker was used routinely for the majority of agarose gels. It contains 23 bands which range in size from 75 bp to 12.2 kb.

S. cerevisiae strain YP148: This cell line was a kind gift from Dr F Richards. Whole chromosomes suspended in agarose were used as size markers in pulsed field gels. There were 17 bands ranging in size from 75 kb to 1 900 kb.

λ ladder (New England Biolabs): Concatemers of λ DNA were also used as a size marker. The ladder spans the size range 48 kb to 1 018 kb, each band increasing in size by 48 kb. The marker is supplied suspended in 1% agarose.

2.6.4 Conventional agarose gel electrophoresis

Agarose concentrations of 0.8 to 3.5% in 1x TAE, 0.5 μg/ml of ethidium bromide were used for conventional, horizontal agarose gel electrophoresis to separate DNA fragments. The electrophoresis of small DNA fragments was usually run at 80 to 120 V, gDNA samples which were used for Southern blots were separated more slowly, at 20 to 50 V, usually overnight. Gels were viewed on a UV transilluminator and photographed with a Polaroid MP4 Land Camera using Polaroid 667 coatless instant film.

2.6.5 Preparative agarose gel electrophoresis

Probes for hybridisation to Hybond N^+ filters were labelled either as whole plasmids or the DNA insert was cut from the plasmid vector using restriction endonucleases. DNA fragments which were to be used as hybridisation probes were separated in LMP agarose at 20 to 60 V, the band of DNA was cut out of the gel, diluted with about 2 volumes of dH_2O and melted at 65°C.

2.6.6 Pulsed field Gel Electrophoresis (PFGE)

Pulsed field gel electrophoresis was conducted using the CHEF (contour clamped homogenous electric field) DR-II apparatus (BioRad). The 1 % agarose gels were made in 0.5x TBE and run at 200 V for 20 to 24 hours at 12-15°C using pulse times between 1 to 120 seconds. The gels were run without ethidium bromide since this slows DNA migration through the gel and changes the pulse conditions needed for separation (Southern 1987). The gels were stained in 500 ml of 0.5x TBE, 100μg/ml ethidium bromide for 30 minutes before viewing.

2.6.7 Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide gel electrophoresis was used to visualise or purify small DNA fragments such as oligonucleotides, to separate fragments which were very close together in size and for sequencing reactions. The 30% acrylamide stock was made up by adding 500 ml dH_2O to the pre-made 190g acrylamide and 10g bis-acrylamide mixture (supplied by Merck). The gel mixture was polymerised using 1.25 µl/ml each of 25% ammonium persulphate and TEMED (N,N,N'N'-tetramethylethylenediamine).

PCR products were separated in non-denaturing 10 % acrylamide, 1x TBE using 1mm spacers and combs. Gels were run at 80 mA for 3 to 4 hours. A denaturing polyacrylamide gel containing 8 M urea was used to purify the 60 base primer, RARE1, used in the RARE analysis. The primer was run out in 15% acrylamide using 3 mm spacers at 20 to 40 mA for 5 hours. The gel was not stained, the shadow of the primer created by the DNA absorbency was viewed under UV light (240 nM) using an intensifying screen. To purify the DNA the band was cut out of the acrylamide gel and treated as described in section 2.13.1.

Sequencing reaction products were run out at 70 W in a 6% acrylamide, 8 M urea gel in 1x TBE for up to 5 hours depending on the size of DNA fragments to be resolved. The sequencing reactions, in a formamide loading buffer (see section 2.10), were denatured at 95°C and then placed on ice, 3 µl of each reaction were loaded into a well. Initially, after the gel had been run it was 'fixed' in 10 % acetic acid, 10 % methanol to remove urea. This step was found to be unnecessary and was omitted in later experiments. The gel was placed on 3MM chromatography paper and dried using a gel dryer attached a vacuum pump. Once dry the gel was exposed overnight, at room temperature, to X-ray film (Curix RPI, Agfa).

2.7 Southern Blotting

DNA which had been separated in an agarose gel was transferred onto Hybond N^+ membrane by Southern blotting. Following electrophoresis the gel was photographed under UV light and then soaked in 0.4 M NaOH for 30 minutes. To ensure efficient transfer of large (>20 kb) DNA fragments from pulsed field gels the gels were exposed to UV light for 90 seconds.

A standard capillary blot was constructed using 3MM chromatography paper as wicks placed into a reservoir of 0.4 M NaOH The agarose gel was blotted overnight or for 24 hours for pulsed field gels. When the blot was dismantled, the membrane was removed from the blot apparatus, rinsed in 0.4 M NaOH, neutralised in 100 mM sodium phosphate, and allowed to dry at room temperature.

2.7.1 Transfer of DNA from bacterial colonies to Hybond N^+ membranes

Single colonies were picked from an LB agar plate and streaked onto a gridded Hybond N^+ membrane on a fresh agar plate. A second agar plate, marked with the same grid was then streaked with the same sample. This was used as a master plate to store a copy of each colony growing directly on the agar surface. The two plates were incubated at 37°C for 6-16 hours.

To lyse the cells and fix the DNA, the Hybond N^+ membrane was placed (colony side up) on 3MM paper, soaked in 10% SDS, for 3 minutes. The membrane was transferred to paper soaked in 0.5 M NaOH, 1.5 M NaCl for 5 minutes and then a 5 minute soak on 1.5 M NaCl, 0.5 M TrisHCl (pH 7.4) was used to neutralise the membrane. Finally the membrane was rinsed by placing it on 2x SSC (0.3 M NaCl, 0.03 M trisodium citrate, pH 7.0) for 5 minutes. The filter was then allowed to dry at room temperature.

To remove debris from large colonies of cells, the Hybond N^+ filters were prewashed for 30 minutes at 50 °C in 5x SSC, 0.5% SDS, 1 mM EDTA (pH 8.0). The soaked membrane was then wiped with a tissue to remove the cell debris and transferred immediately to the prehybridisation buffer.

2.8 Hybridisation of radiolabelled probes to Hybond N^+ filters

2.8.1 Radiolabelling DNA probes

About 25 to 50 ng of probe DNA was labelled with [α-^{32}P] dCTP and Klenow DNA polymeraseI by random priming from hexadeoxyribonucleotides. For these labelling reactions a kit supplied by Boehringer Mannheim was used according to the method described by Feinberg and Vogelstein (1983, 1984). The DNA to be labelled was denatured by boiling and immediately added to the labelling reaction. The reaction was incubated at room temperature for 20 minutes and then at 37°C for a further 20 to 40 minutes. The labelling reaction was stopped by the addition of 100 μl of 'Stop' (20 mM NaCl, 20 mM TrisHCl, 2 mM EDTA, 0.25% SDS). The probe was denatured by boiling and immediately used to hybridise to a Hybond N^+ filter. Frequently the probe being labelled contained LMP agarose. It was found that it was not necessary to remove this prior to hybridisation.

2.8.2 Checking incorporation of radio-isotope

The percentage of radioactive label which was incorporated into the DNA probe was estimated by precipitation in TCA (trichloroacetic acid) using a method adapted from Sambrook (1989). 10 μl of a 1:100 dilution of the labelling reaction was applied to a nitrocellulose filter to use as an unprecipitated control. A second 10 μl aliquot was

precipitated with 10 μg tRNA, 1 ml 10% TCA. This was passed through a nitrocellulose filter using vacuum suction and rinsed with 2 - 3 ml of 10% TCA. Both filters were placed scintillation fluid and counted using a scintillation counter. The level of incorporation of the label in the precipitated sample was calculated as a percentage of the unprecipitated sample. This was usually >65%. If a reaction was found to have incorporated < 50% of the radiolabel it was not used and the probe was relabelled.

2.8.3 Hybridisation of radiolabelled DNA

Hybond N^+ filters with bound DNA were prehybridised in a rotating oven (Hybaid) at 65°C for at least an hour in 0.5M NaCl, 1% SDS, 100 mM sodium phosphate. Hybridisation of the denatured radiolabelled probe to the filters was conducted at 65°C overnight in 10 ml of the hybridisation buffer with the addition of 200 μg/ml denatured sonicated herring testes DNA. Non-specific binding of radioactivity to the filter was removed by washing the filters 3 or 4 times for 30 minutes at 65°C in 1% SDS, 20 mM sodium phosphate.

2.8.4 Autoradiography

Filters were wrapped in cling film to prevent them from drying out and were then exposed against X-ray film in cassettes with intensifying screens at -70°C overnight or for longer if required. The films were developed in an automatic X-ray film processor.

2.8.5 Removal of Hybridisation signal from Hybond N+ filters

Radioactive label was removed from the Hybond N^+ filters in boiling 0.5% SDS. The filters were placed in the SDS and left to cool to room temperature. They were

then rinsed well in 100 mM sodium phosphate and allowed to dry at room temperature.

2.9 ExonucleaseIII unidirectional deletion clones

A series of unidirectional deletion clones were produced using exonucleaseIII as described in Henikoff (1984) and Hoheisel (1986). These were used in the sequencing project described in Chapter 3. This method is based on two useful features of exonucleaseIII: it progressively digests DNA at a uniform rate and it cannot initiate digestion at DNA ends with 4 base 3' protrusions. About 50 μg of plasmid DNA was digested with 600 units KpnI, it was recovered by precipitation with sodium acetate and ethanol and then digested with 600 units of EcoRI. The restriction digests were undertaken separately to ensure that digestion was complete. The DNA was purified by phenol/chloroform extraction, precipitated in 0.1 volume 3M sodium acetate, 2 volumes ethanol and then resuspended in dH_2O to a final concentration of 0.8 μg/μl.

ExonucleaseIII digestion was carried out in 75 mM TrisHCl pH 8.0, 3.3 mM $MgCl_2$ using 16 units enzyme/μg DNA. This was calibrated to digest the plasmid DNA at a rate of about 200 bases per minute. The 3μl aliquots were removed at 2 minute intervals, diluted in an equal volume of water and the reaction was stopped by heat inactivation at 70 °C for 10 minutes. The aliquots were then treated with 18 units S1 nuclease in 15 μl S1 buffer (16 mM $NaC_2H_3O_2$ pH 4.6, 400 mM NaCl, 1.6 mM $ZnSO_4$, 8% (v/v) glycerol) to trim single-stranded tails from the DNA fragments. The reaction was stopped by adding 5 μl S1 'stop buffer' (800 mM TrisHCl pH 8.0, 20 mM EDTA pH 8.0, 80 mM $MgCl_2$).

The samples were run out in 0.7% LMP agarose (see diagram 3.3) and the bands were cut out from the agarose gel and diluted in 3 volumes of dH_2O. The molten

agarose/DNA sample was treated with 3 units Klenow DNA polymeraseI in 60 mM NaCl, 10 mM TrisHCl pH 8.0, 10 mM $MgCl_2$ and 50 μM dNTP's. A 20 μl aliquot of the Klenow treated DNA was ligated overnight as described in section 2.12. The ligated products were used to transfect competent cells (see section 2.3), plated out on LB agar plates, and incubated at incubated at 37°C overnight.

2.10 Double stranded DNA Sequencing

Double stranded DNA sequencing was conducted using a chain-termination method with [α-^{35}S] dATP and T7 polymerase. About 2 to 4 μg of the plasmid DNA was denatured in 0.4 M NaOH, 0.4 mM EDTA (pH 8.0), precipitated in 0.1 volume 3M sodium acetate, 2 volumes ethanol and then resuspended in 6 μl dH_2O. The annealing reactions were incubated at 65°C for 2 minutes in 40 mM TrisHCl (pH 7.5), 20 mM $MgCl_2$ and 50 mM NaCl with 2 pmoles of primer in a volume of 10 μl and allowed to cool to room temperature.

To this annealing reaction 1μl 0.1M DTT, 2 μl 'Labelling mix' (7.5 mM each dCTP, dGTP and dATP) 0.5μl (0.5 μCi) [α-^{35S}]dATP and 8 units T7 polymerase were added. The extension reaction was allowed to continue at room temperature for 2 to 5 minutes. 3.5 μl of this extension mix was added to 2.5μl of each Termination mix (80 μM each dATP, dCTP, dTTP and dGTP, 8μM each ddATP, ddCTP, ddGTP or ddTTP as appropriate) which had been preheated to 37°C. After incubation for 5 minutes the reaction was stopped by adding 4μl of 'STOP' buffer (95% deionized formamide, 10 mM EDTA (pH 8.0), 0.025% bromophenol blue, 0.025% xylene cyanol) which was used as a loading buffer.

For regions which had been shown to contain secondary structure which interfered with the sequencing reactions or the resolution of the reaction products, 7-deaza dGTP was used. The labelling mix was made by substituting dGTP with 7-deaza

dGTP at the same concentration of 7.5 mM. Termination mixes A, C and T contained 80 μM 7-deaza dGTP instead of normal dGTP while the G termination mix contained 160 μM 7-deaza dGTP. The concentrations of the other components in the sequencing reactions were the same as before.

2.10.1 Oligonucleotide sequencing primers

The majority of PCR and sequencing oligonucleotide primers were synthesised on the Pharmacia 'Gene Assembler'. The oligonucleotide primers were cleared from the support cassette by an overnight incubation at 55°C in ammonia. Initially the primers were purified by passing the crude mixture down a Sephadex column, this removed the ammonia and small molecular weight material and eluted the primer in dH_2O. Later primers were extracted from the ammonia by drying in a centrifuge evaporator, the resultant pellet was simply resuspended in 500 μl dH_2O. The concentration of the primer was calculated by measuring the optical density of the solution at 260 nm.

A number of primers, usually 18 bases in length, were synthesised for sequencing. The majority of these primers were used in the large scale sequencing project described in Chapter 3. Sequencing was also undertaken in the upstream probe ACIXSR (described in Chapter 7), downstream in intron 4 of PDGFA (Chapter 9) and on the end-clones derived from the telomere YACs rescued by Vectorette PCR (see Chapter 8 and section 2.12). The sequencing primers are listed in table 2.1.

Primer	Sequence	Used for:
Universal reverse	dAACAGCTATGACCATG	pUC vectors
T3	dATTAACCCTCACTAAAG	pUC vectors
PEXON3	dGCCACTGGGGAACAGGCA	ACY7
P70501	dCGATAAGGATCAGCCCAG	ACY7
P70110	dCCGGTGCCTGCAGGCTGG	ACY7
P70110B	dTCCTACAGCACTTGCTGC	ACY7
P70708	dTTACCGAAAATGACTCAA	ACY7
P70621	dCCGACAAGACAGGGCGCT	ACY7
P71209	dCCAGGCCGAGTGCTCATT	ACY7
950L	dAAGGAGAGGAGTGCTTG	Intron 4
844REV	dTCTGGCGACGAGCCTGGC	Intron 4
PACIXSR2	dTGTTCCCTGCACCCAAGG	ACIXSR
SP6	dGATTTAGGTGACACTATAG	ACIXSR
7397	dCAGGCCCTCCCAGGTGACATT	Intron 3

Table 2.1: Oligonucleotide sequencing primers

Sequencing of the plasmid pACY7 is described in Chapter 3, 'vectorette-PCR' product sequencing from the telomere YACs is described in Chapter 8, clone ACIXSR was sequenced during the analysis of the P1 clones in Chapter 7 and intron 4 was sequenced for the RARE cleavage analysis in Chapter 9.

2.11 Polymerase chain reaction (PCR)

The polymerase chain reaction was used extensively throughout this project. The primers which were used along with their application are listed in table 2.2. A standard reaction buffer of final concentration: 1.5 mM $MgCl_2$, 50 mM KCl, 10mM TrisHCl (pH 8.3), 0.1% gelatin was used in all reactions unless otherwise stated in the table. PCR reactions were usually 50 µl in volume, containing 1 µmol each dNTP, 50 pmol oligonucleotide primers, 1 unit *Taq* DNA polymerase and 100 to 500 ng DNA. The reactions were overlaid with mineral oil. On occasions a 'hot start' of 5 minutes at 94°C was used to denature the DNA before adding the *Taq* DNA polymerase. The reaction cycles were then started without allowing the mixture to cool (see table 2.2). Once the PCR reaction cycles had been completed 10µl of the reactions were run out in a agarose gel to check for the presence of PCR products.

Bacterial colonies were screened by PCR using the primers T3 and T7. The colonies growing on the LB agar plates were touched and transferred into the PCR mix using a toothpick or a Gilson pipette yellow tip. The PCR reactions were denatured for 5 minutes at 98°C prior to adding the *Taq* DNA polymerase and commencing the cycling reactions.

2.11.1 Random PCR amplification of DNA

Random PCR primers were used to amplify the telomere YAC DNA for the FISH experiments (see section 2.14) using a method from Dr J. Warner. The YACs were separated from the host cell genome by pulsed field gel electrophoresis using 30 second pulses. PFGE is described in section 2.6.6. The YAC bands were cut from the gel, purified by 'Geneclean' (see section 2.4.5) and resuspended in 20 µl dH_2O.

Primer	Sequence	Reagents	Cycling conditions	Product size	Comments
920601A 631	dCTGAGAGCTCACGGGGTGCA dGCCAGGCCTATTTTAGGGAG	Standard 0.01% triton	94° 1 minutes 60° 1” 72° 2” X 30 cycles	241 bp	PDGFA exon 3 dimorphism digest with ApaII: 241/223 bp.
ApaLI 844	dAGGGTCCCTCCTGGGAGGC dAGACAGCTTCCGGCTAAGCAGCT	Standard	94° 1” 68° 1” 72° 2” X 30 cycles	256 bp	PDGFA IVS4 dimorphism digest with ApaLI: 256/202 bp.
631 632	dGCCAGGCCTATTTTAGGGAG dCTAAGCATGTGCCCGAGAAG	Standard 0.01% triton	94° 1” 60° 1” 72° 2” X 30 cycles	216 bp	PDGFA exon 3 STS. YAC library screening.
F8INT19A F8INT19B	dGGCGAGCATCTACATGCTGGGATGAGC dGTCCAGAAGCCATTCCCAGGGGAGTCT	DMSO	94° 2.5” 60° 6” X 1 cycle 94° 1” 60° 30 sec X 30 cycles 60° 10 min X 1 cycle	~600 bp	Factor VIII intron 19. Positive control for YAC library screening.
430F 731R	dGGGGCTGGGAGTGCATAAGACACCCAATCC dTGAAGAACTGAAGCAACAAATACCTACGGC	Standard DMSO	94° 1” 68° 1” 72° 2” x 30 cycles	331 bp	PDGFA intron 5 STS. YAC library screening.
2660 2790	dGGCTGGATTTGGGCGTCTTTAGT dCGGGCCGAACCCGTGGAGAGGT	25 mM $MgCl_2$ Triton	94° 1” 60° 1” 72° 2” x 30 cycles	152 bp	PDGFA CA repeat intron 1. Spurious band at 500 bp.
ACIXSR91 ACIXSR260	dACGGAGCTGGGGTGGGTGTG dGCTGTGAAGTGGGGCTGATT	10 mM $MgCl_2$ Hot start	94° 5minutes 94° 1” 60° 1” 72° 2” x 35 cycles	189 bp	STS from upstream of PDGFA. Screening P1 clones.

sv-1-2178F sv-1-2185R	dGCTTTCGGGTACAGCTATTTAG dCTCTGACCCAGATCCTAATCC	Triton Hot start	92° 3 minutes 92° 1” 55° 2” 72° 2” x 35 cycles 72° 5 minutes	98 bp	Telomere YAC 2178.
sv-1-2185F sv-1-2185R	dCCTGGGTGTGCCGTAAAC dTCACTGAGCTTGCTGGCTTA	Triton Hot start	92° 2 minutes 92° 1” 65° 2” 72° 2” x 35 cycles 72° 5 minutes	357 bp	Telomere YAC 2185.
7355 7970	dCCTTAAAGTGTCATGTGTGGCT dGAGGAATCTCGTAAATGACCGT	DMSO Hot start 750 mM KCl Deaza dGTP	94° 5 minutes 94° 1” 55° 2” 72° 2” x 35 cycles 72° 10 minutes	637 bp	PDGFA IVS3 minisatellite, allele size depends on no.repeats.
T3 T7	dATTAACCCTCACTAAAG dAATACGACTCACTATAG	Standard Hot start	98° 5 minutes 94° 30 sec 50° 30 ” 72° 1 minute x 30 cycles 72° 10 minutes	depends on size of cloned insert	Screening colonies for cloned insert in plasmid vector.
D-580 PBR-4177	dCGAATCGTAACCGTTCGTACGAGAATCGCT dTTGGAAGCATTTATCAGGGTTATTGTCTCAT	10 mM $MgCl_2$ Hot start	95° 10 minutes 92° 2” 60° 2” 72° 3” x 38 cycles	variable	Vectorette end-rescue of telomere YAC clones.

Table 2.2: Oligonucleotide primers and reaction conditions used in PCR assays.

The PCR amplification was conducted using a 10:1 molar ration of the primers TailL dGGG GAG ACT ACG ATA CAG AGG and DegenL dGGG GAG ACT ACG ATA CAG AGG NPuN PuNPu NPuN PuN. The sequence of TailL was selected to minimise the formation of dimers and loops. It was checked to avoid common repeats, restriction sites and palindromes. An 11 base degenerate 3' end is added to primer TailL to make primer DegenL. The alternative bases were limited to purines to minimise hydrogen bonding, reducing primer-primer binding.

The PCR samples were amplified in a standard 1.5 mM $MgCl_2$ buffer using 2 μM and 0.2 μM primers, 2 μl DNA and 1.5 units *Taq* DNApolymerase in a 50 μl volume. The DNA was denatured at 94°C for 5 minutes and then 5 precycling cycles were used. These denatured the DNA at 94 °C for 1 minute and annealed the primers at 27°C for 2 minutes 30 seconds. The temperature was then increased slowly to 72°C over 3 minutes and then left at 72°C for a further 2 minutes. These rather unusual precycles are designed to bind and amplify from the degenerate primer. Following these precycles, the DNA was denatured at 94°C for 5 minutes and then 25 cycles of 94°C for 1 minute, 53°C for 1 minute 30 seconds and 72°C for 2 minutes 30 seconds were used. Finally the PCR products were extended at 72°C for 10 minutes. These samples were checked in an agarose gel and then reamplified using the same reaction components. The DNA was denatured at 94°C for 5 minutes and then 25 cycles of 94°C for 1 minute, 53°C for 1 minute and 72°C for 2 minutes were used. The PCR products of 5 times 50 μl of PCR product were pooled and the 250 μl were used as a probe for FISH (see section 2.14).

2.11.2 Cloning PCR products

PCR products were cloned using either the 'PCR-script' cloning kit supplied by Stratagene or the 'TA' cloning kit supplied by Invitrogen (described in section 2.2.1). Both kits were used according to the instructions supplied with the product.

2.12 End-cloning YACS by vectorette PCR

'Vectorette' end-cloning uses PCR to amplify the end fragments of YAC DNA clones (Riley 1990). The YAC clone is digested with a restriction enzyme and a common linker is ligated to the DNA fragments. Terminal sequences are then amplified using a vector specific primer and a linker specific primer. The linker, or 'vectorette cassette' which is ligated to the digested DNA contains an area of mismatch. The linker specific primer (D-580) is the same as the lower sequence of the cassette. It cannot therefore be used to seed the PCR reaction until its complementary strand has been synthesised from the vector primer. This ensures that PCR products can only be amplified from DNA molecules which have both the vector and cassette sequences.

Total YAC DNA was isolated as described in section 2.4.8. About 2.5 μg of DNA was digested by incubation at 37°C for 2 to 3 hours using 12 unit/μg of Sau3AI or RsaI. The reaction was stopped by incubation at 65°C for 10 minutes. The DNA was then precipitated in sodium acetate/ethanol and resuspended in 10 μl of dH_2O.

The Vectorette cassette primers are shown in table 2.3. Primers Q292, Q293, A294 and D-580 which were a kind gift from Dr J. Brown (MRC Human Genetics Unit, Edinburgh) correspond to the primer sequences given in Riley (1990) The primer PBR-4177 was designed for the vector sequence of the telomere YACs. When annealed, primers Q292 and A294 give a linker cassette which is blunt ended. Primers Q292 and Q293 give a cassette with a four base 5' overhang of the sequence GATC. This cassette is used to clone fragments produced by Sau3AI digestion which leaves a 3' overhang. The primers were mixed to a final concentration of 1 μM, heated to 94°C for 10 minutes and then slowly cooled to 20°C, over 20 minutes, to allow the primers to anneal.

Primer	Sequence
Q293	dCTC TCC CTT CTC GAA TCG TAA CCG TTC GTA CGA GAA TCG CTG TCC TCT CCT TG
Q292	dGAT CCA AGG AGA GGA CGC TGT CTG TCG AAG GTA AGG AAC GGA CGA GAG AAG GGA GAG
A294	dCTC TCC CTT CTC GAA TCG TAA CCG TTC GTA CGA GAA TCG CTG TCC TCT CCT TGG ATC
D-580	dCGAATCGTAACCGTTCGTACGAGAATCGCT
PBR-4177	dTTGGAGCATTTATCAGGGTTATTGTCTCAT

Table 2.3: Oligonucleotide primers used for 'Vectorette PCR'.

The primers sequences for Q293, Q292, A294 and D-580, are from Riley (1990). Primer PBR-4177 was designed specifically for the YAC cloning vector used to construct the telomere YACs (analysed in Chapter 8).

About 1µg of the restricted yeast DNA was ligated to 6 pmole of the vectorette primer cassette in 50 mM TrisHCl pH 7.5, 7 mM $MgCl_2$, 1mM DTT, 1 mM rATP, using 2 units T_4 ligase in 20 µl. The ligation reaction was incubated at 16°C overnight, heated to 65°C for 10 minutes to inactivate the ligase and then stored on ice prior to use. The ligated DNA fragment was amplified by PCR using the primers D-580 and PBR-4177, as described in table 2.2 using 100 µM each dNTP and 100 pmole of each primer. A 10µl sample of the resultant PCR products was checked by digestion with EcoRI since only genuine YAC end-clones should contain an EcoRI site at the junction between the vector and the cloned insert. PCR products which were shown to cut with EcoRI were cloned using the Stratagene PCR cloning kit.

2.13 RecA assisted restriction endonuclease cleavage (RARE)

2.13.1 PAGE purification of oligonucleotide primer

The 60 base primer RARE1 (5'-GCCTGCCTTCTTCAGGCTGCTTTAGCAGAATTCC ATCCACTAGGAGCTGATTAGCAACAA-3') was used in the RecA-assisted restriction endonuclease (RARE) cleavage experiments described below and in Chapter 9.

To purify the primer it was resolved in a denaturing polyacrylamide gel as described above in section 2.6.7. The DNA was cut from the gel giving blocks of acrylamide of about 300 to 400 µl, which were crumbled in an Eppendorf tube using a pipette tip. At least two volumes (1 ml) of elution buffer (0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA pH 8.0, 1% SDS) was added and the sample was left overnight at 37 °C on a shaking platform. The elution buffer was collected and the acrylamide was rewashed in a further 500 µl elution buffer. The DNA primer was precipitated with ethanol, the pellet was resuspended in dH_2O and

reprecipitated using sodium acetate and ethanol. The purified primer was quantitated and compared to the original primer in a polyacrylamide gel.

Since the primer could not be seen under ethidium bromide staining the gel was silver stained using a method base on that described in Budowle (1991). The polyacrylamide gel was soaked in 10% ethanol for 5 minutes followed by 1% nitric acid and then rinsed in dH_2O. The gel was soaked for 20 minutes in 0.12 M silver nitrate, rinsed again in dH_2O and then developed using 0.28 M sodium carbonate, 0.019% formaldehyde. Once the DNA had stained sufficiently the reaction was stopped in 10% acetic acid for 2 minutes.

2.13.2 RecA-assisted restriction endonuclease (RARE) cleavage

Preparation of agarose DNA blocks used for pulsed field gel electrophoresis was described in section 2.4.7. These blocks, which contain about 10 µg of gDNA were equilibrated by at least three washes of 30 to 45 minutes in 2 ml 1x RARE buffer (25 mM Tris-acetate pH 7.85, 4 mM $MgCl_2$, 0.5 mM spermidine, 0.4 mM DTT). The reaction reagents were then added, in the following order, dH_2O to make the final volume of the reaction 200 µl after methylase addition: 24 µl 5x RARE buffer, 30µg RecA protein, 2 µl 10 mM EGTA, 20 µl 11 mM ADP, 20 µl 3 mM ATP-γ-S, 360 to 1200 ng oligonucleotide primer (RARE1), the 80 µl DNA agarose block and 10 µl 2 mg/ml acetylated BSA. The reaction was incubated at 37°C for 30 minutes to allow the RecA protein to bind the oligonucleotide primer, 600 units of EcoRI methylase and 10 µl of 2.4 mM SAM (S-adenosylmethionine) were added and the reaction was incubated at 37°C for 1 hour to methylate the gDNA. The reaction was stopped by the addition of 200 µl 2% SDS and incubation at 37°C for 20-30 minutes.

To remove the SDS before digestion with EcoRI the agarose blocks were rinsed well in TE and then washed at least four times in 2 ml TE for 30 to 45 minutes on a rotating wheel at room temperature. It was essential to remove all traces of the detergent to ensure that the subsequent restriction digests would be complete. The blocks were then equilibrated into the restriction endonuclease buffer for EcoRI as described in section 2.5.1. The EcoRI digest was incubated overnight at 37°C using 200 μl of buffer and 600 units of enzyme. The subsequent running of the pulsed field gel, blotting and filter hybridisation were conducted as described above (sections 2.6.6, 2.7 and 2.8).

2.14 Fluorescent in situ hybridisation

Fluorescent *in situ* hybridisation (FISH) was conducted using biotin labelled DNA probes with metaphase chromosome spreads from a normal male essentially as described in Fantes (1992).

2.14.1 Biotin labelling of DNA probes

DNA probes were labelled with biotin by nick translation using biotin-16-dUTP. The reaction labelled about 0.5 μg DNA in 50 mM TrisHCl pH 7.5, 10 mM $MgSO_4$, 100 μM DTT, 50 μg/ml BSA, 62.5 μM each dATP, dGTP, dCTP and biotin-16-dUTP using 0.2 unit/ml DNase and 500 units/ml DNA polymeraseI in a total volume of 20 μl. The reaction was incubated at 15°C for 90 minutes and was stopped by precipitating the DNA in 0.1 volume 3 M sodium acetate, 2 volume ethanol, 1μl 0.5M EDTA using 1 μl (20 μg) glycogen as a carrier for the DNA. The pellet was allowed to dry and then resuspended in 25 μl of TE.

2.14.2 Hybridisation of DNA probes to metaphase chromosome spreads

Prior to hybridisation to the metaphase chromosome spreads, 5 μl of the biotin labelled probe was precipitated with 5-12 μg Cot1 DNA and 50 μg salmon sperm DNA in 2 volumes of ethanol. The ethanol was removed from the precipitated DNA using an evaporator centrifuge. Once the DNA pellet was completely dry it was resuspended in 10μl hybridisation buffer (50% formamide, 2x SSC, 1% Tween 20 and 10% dextran sulphate) by incubation at 37°C for an hour. The probe was transferred to 70°C for 5 minutes to denature the DNA and then returned 37°C to preanneal for 15 minutes.

The metaphase chromosome spreads on microscope slides were prepared by T. Johnstone. They were incubated in RNase (100 μg/ml in 2x SSC) for one hour at 37°C. The slides were rinsed in 2x SSC and then washed for 2 minutes each in 70%, 90% and then 100% ethanol to dehydrate the DNA, blotted and dried under vacuum for 10 minutes. The slides were then warmed to 70°C for 5 minutes, denatured in 70% formamide, 2x SSC for 3 minutes at 70°C and then dehydrated by washes in 70%, 90% and 100% ethanol. Prior to use the slides were returned to the vacuum dryer for a further 10 minutes.

To hybridise the probe to the metaphase chromosomes the biotin labelled DNA suspended in hybridisation buffer was applied to the microscope slide over the chromosome spread and sealed under a glass cover slip using vulcanising solution. The slides were incubated for 16 to 20 hours at 37°C in a humid atmosphere. The rubber seal around the cover slides was removed and the slides were washed four times in 2x SSC at 45°C. The slides were then washed four times in 0.1x SSC at 60°C and finally were transferred to 4x SSC, 0.1% Tween 20.

2.14.3 Detecting the biotin labelled probe

Hybridised probe was detected by incubation with avidin FITC followed by biotinylated antiavidin and an additional layer of avidin FITC. A 1:500 dilution of avidin FITC was prepared in 4x SSC, 5% skimmed milk giving a working concentration of 4 μg/ml. This dilution was centrifuged for 15 minutes and the slides were incubated for 5 minutes at room temperature with 40 μl of blocking buffer (4x SSC, 5% skimmed milk). A 40μl aliquot of avidin FITC in blocking buffer from the top of the centrifuged sample was applied to the chromosome spreads which were incubated at 37°C in a box lined with paper soaked in 4x SSC, 0.1% Tween 20 for 30 minutes. The slides were washed three times for 2 minutes in 4x SSC, 0.1% Tween 20 at 45°C.

Biotinylated antiavidin was diluted 1:100 to a final concentration of 0.5 mg/ml in 4x SSC, 5% skimmed milk. This dilution was centrifuged for 15 minutes and then 40μl were applied to each slide. The slides were incubated at 37°C for 30 minutes and then washed three times for 2 minutes in 4x SSC, 0.1% Tween 20 at 45°C. To this a further 40 μl of diluted avidin FITC was added and again the slides were incubated for 30 minutes at 37 °C and then washed three times for 2 minutes in 4x SSC, 0.1 % Tween 20 at 45°C.

The slides were mounted in antifadent ('Vectormount') containing 10 μg/ml DAPI, the cover slips were replaced, sealed with vulcanising rubber and then examined using a Leitz fluorescent microscope and CCD camera.

3. Chapter 3: Sequence and structural studies on PDGFA

3.1 Introduction

The introduction to this thesis describes the PDGFA gene, compares it to the PDGFB gene and discusses a number of unusual features. At the start of this project the sequence of the exons of PDGFA was known (Bonthron 1988). In addition to this, the short introns 5 and 6 had been sequenced along with the first 750 bases of intron 4 (Bonthron 1992). However introns 1 to 3 and most of intron 4 were unsequenced. Thus a sequencing project was undertaken to sequence 8 kb of genomic sequence spanning from exon 1 to exon 4 of PDGFA. This would leave only the large (10 kb) intron 4 partially unsequenced.

This sequencing project was undertaken for a number of reasons: of pressing importance was the need for polymorphic markers from within the PDGFA gene for linkage analysis. At the start of this project two polymorphic loci had been identified in PDGFA. These were the minisatellite in intron 4 and the StyI RFLP which are described in section 1.5. Neither of these provided useful polymorphic markers for linkage analysis. The minisatellite was not sufficiently polymorphic to be useful and, while the StyI RFLP looked more promising, the probe which was available for Southern blotting proved difficult to use. The molecular basis for this RFLP was not known although the distribution of StyI sites in the genomic sequence suggested that the polymorphic band resulted from hybridisation to exon 4 (Bonthron 1992). It was hoped that this sequencing project would answer these questions, provide a useful probe or PCR to type the StyI polymorphism and discover other useful polymorphic markers in PDGFA.

Alternative splicing of the PDGFA transcript involving exon 6 in human DNA and the existence of a previously unknown exon from within intron 6 in rabbits have been described. In addition, the omission of exon 2 and exon 6 has been found in transcripts isolated from a human osteosarcoma cell line (section 1.5.3). These known alternative splicing patterns raised the possibility that other exons which have

not yet been identified exist within the introns of the gene. This sequencing project was expected to cover about 8 kb of the genomic DNA sequence fully sequencing introns 1 to 3. The large size of the introns meant the presence of an unknown exon was possible. In the factor VIII gene, sequencing of intronic sequence has even revealed the presence of a new gene sitting within an intron of another gene (Levinson 1990).

3.2 Plasmid clones from PDGFA

Four bacteriophage clones covering 45 kb and spanning PDGFA were available (Bonthron 1988). Three of these are shown in diagram 5.1. One of these clones, λAC23, was used to make plasmid subclones for this sequencing project.

3.2.1 Exon 2 to exon 4: plasmid clone ACY7

The bacteriophage clone, λAC23, was digested with restriction endonuclease SalI and XbaI producing a 7.2 kb fragment spanning from the XbaI site in intron 1 to the SalI site within exon 4. This fragment was subcloned into pBluescript II KS(+/-)and the resultant colonies were screened by PCR amplification by D.T Bonthron. The PCR screening used either the T3 or T7 primer with primer 144P. The primer 144P lies within exon 4 and thus the PCR assay could identify the orientation of the cloned insert. From this, the clone pACY7 was selected with its cloned DNA insert in the direction shown in diagram 3.1.

3.2.2 Exon and intron 1: plasmid clone ACZ

The plasmid clone pACX, a subclone of λAC23, which spans from the SalI site upstream of exon 1 to the XbaI site in intron 1 was available at the start of this

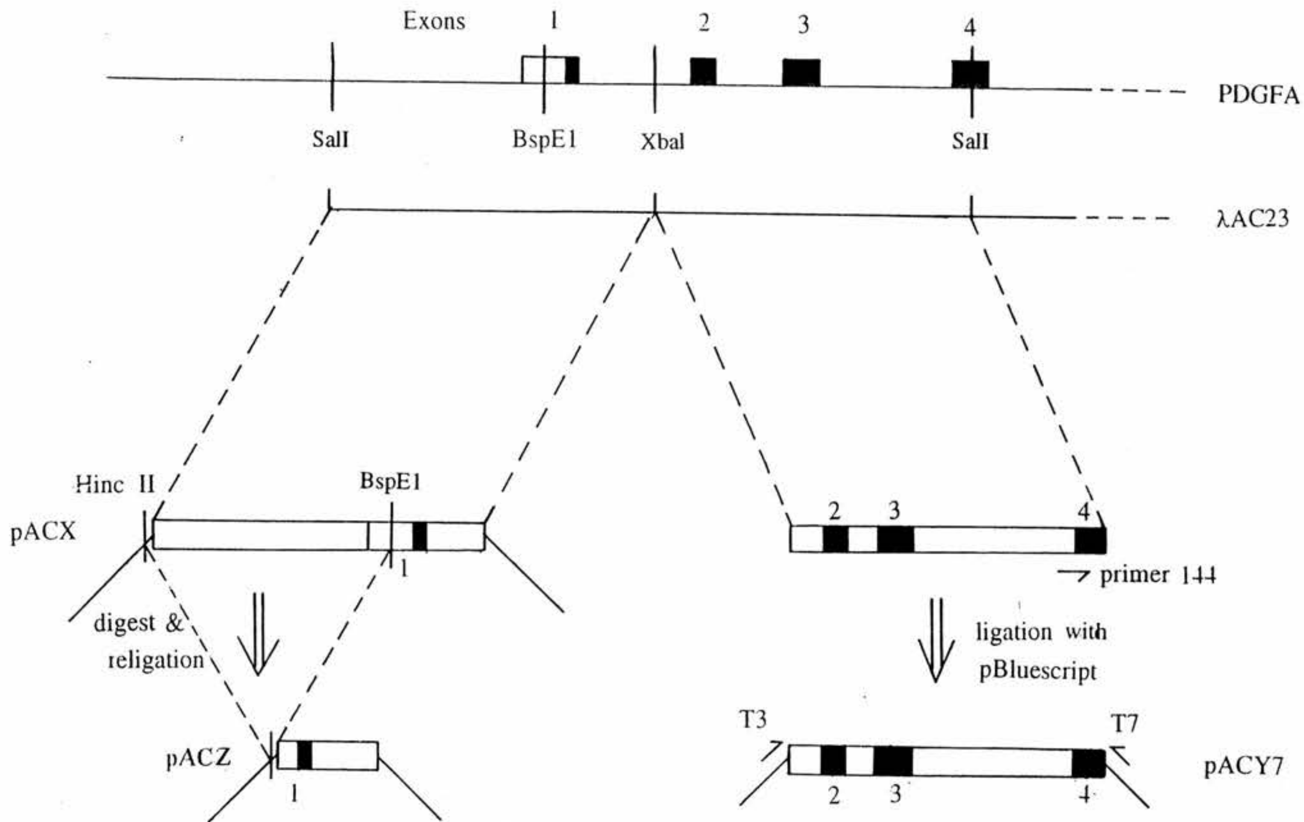


Diagram 3.1: The plasmid subclones pACY7 and pACZ

The clone pACZ, which spans from the BspEI site in exon 1 to the XbaI site in intron 1 was subcloned from pACX using a BspEI/HincII digest. pACX was subcloned from λAC23 by D.T Bonthron prior to the start of this project. Clone pACY7 which spans from intron 1 to exon 4 of PDGFA was subcloned directly from λAC23 using an XbaI/SalI restriction digest. The primer 144 was used in PCR reactions with T3 and T7 to check the orientation of the cloned insert.

project. A BspEI/HincII restriction endonuclease digest was used to remove the promoter region and 5' end of exon 1 from this clone. On religation this produced a plasmid, pACZ, with a 1.6 kb insert. This insert spans from a BspEI site in exon 1 to the XbaI site in intron 1, pACZ is shown in diagram 3.1.

3.2.3 A series of exonucleaseIII unidirectional deleted clones

A series of unidirectional deleted clones was prepared from the two plasmid clones pACY7 and pACZ using exonucleaseIII as described in section 2.9. As diagram 3.2 shows, the plasmid clones were restricted in the polylinker region with KpnI and EcoRI producing a linear plasmid with overhanging ends at both the 3' KpnI site and the 5' EcoRI site. Since exonucleaseIII cannot initiate digestion at DNA ends with four base 3' protrusions (the KpnI site) digestion was unidirectional from the EcoRI site.

Aliquots of the digestion reaction were removed at 2 minute time intervals. These were optimised for a decrease in size of about 200 - 300 bp between each sample creating a series of clones with progressively larger deletions from one end of the plasmid insert. The panel of plasmids were checked by separation in an agarose gel (diagram 3.3) and then religated and transformed into DH5α cells. Using these deleted clones for the sequencing project allowed all of the clones to be sequenced using the vector primer T3.

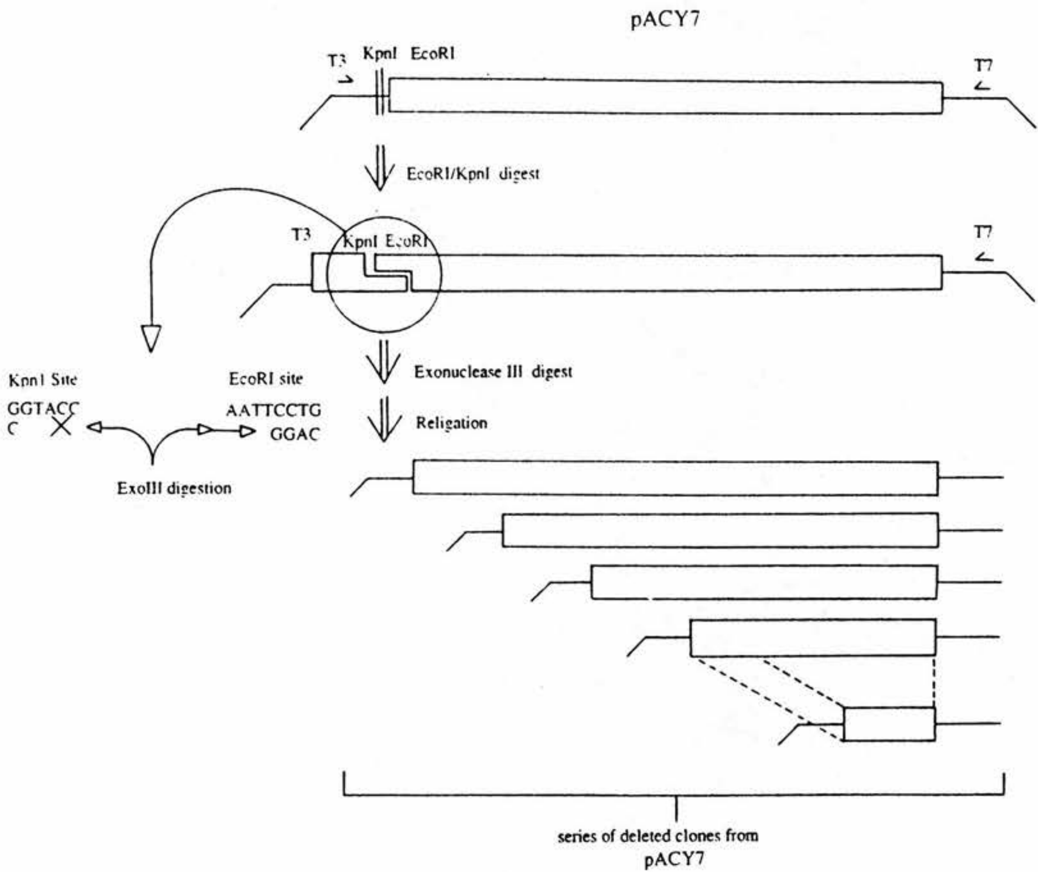


Diagram 3.2: ExonucleaseIII nested deletion clones

Schematic diagram illustrating the use of exonucleaseIII to generate a series of nested deletion clones from the plasmid pACY7. ExonucleaseIII digestion progresses only from the EcoRI site.

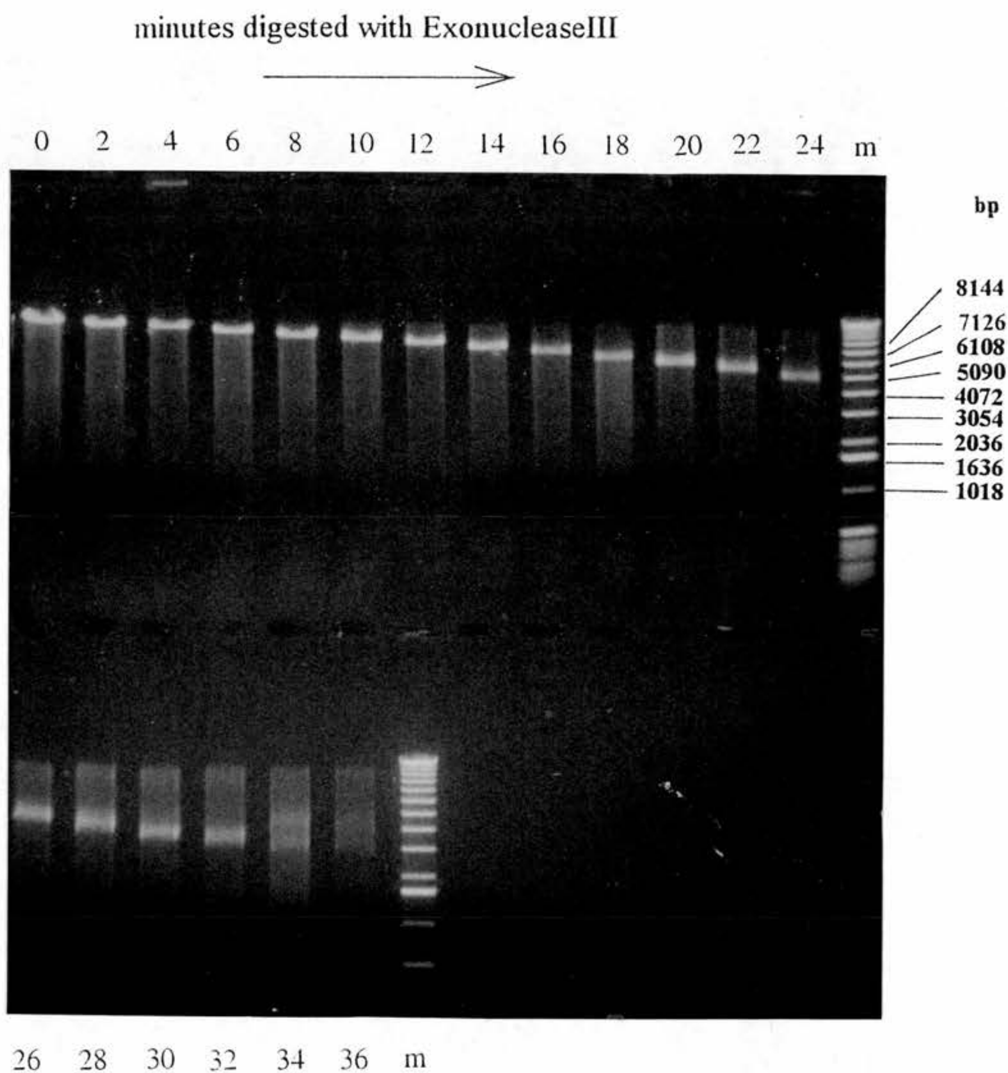


Diagram 3.3: The exonucleaseIII deleted subclones from pACY7

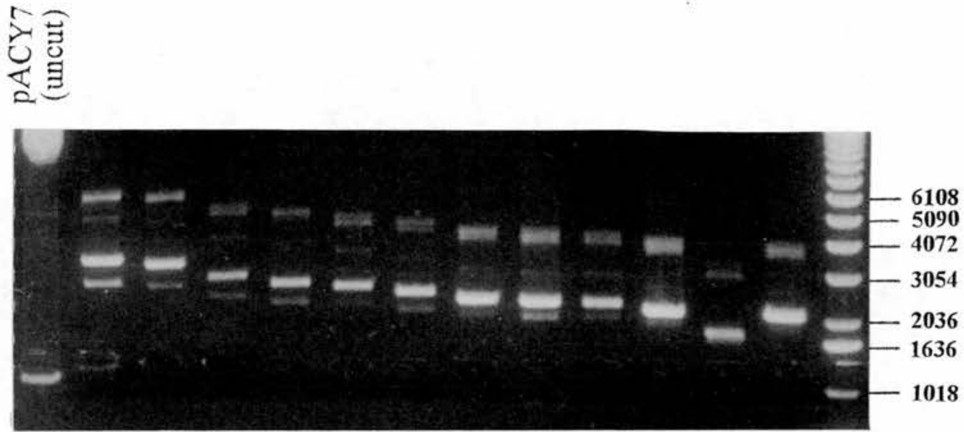
The deleted subclones are separated in 1% agarose prior to religation and transformation. The clones show a progressive decrease in size with increasing time points (shown in minutes).

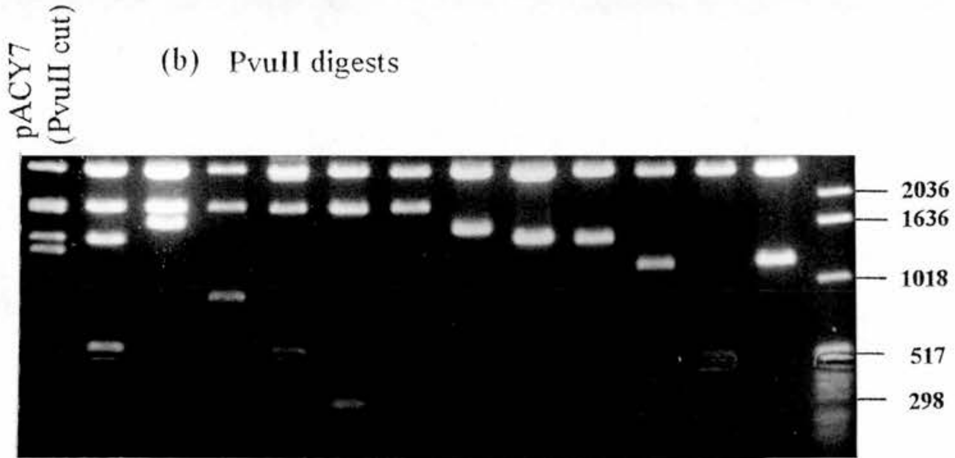
3.2.4 Size analysis of the deleted clones from pACY7

The size of the deletion in each clone derived from the same aliquot from the exonucleaseIII digest should, in theory, be the same. However the cloned inserts were checked before the sequencing was started to avoid redundant sequencing caused by large overlaps between the clone inserts. For this a small scale DNA extraction using phenol/chloroform (see section 2.4) was prepared from a random selection of the deletion clones. The plasmid contained two restriction sites for the enzyme PvuI. This separated the cloned insert from the vector, producing two bands and allowing the size of the insert to be measured directly. As an example, a series of clones with smaller inserts cut with PvuI are shown in diagram 3.4(a). This photograph clearly shows the sequential decrease in insert size in the deleted clones. However in a standard agarose gel it was difficult to adequately resolve small differences in size (<1 kb) between the clones which contained the larger (~8 kb) sized inserts.

The majority of the clones were digested with PvuII. This restriction enzyme has several sites within the cloned insert, producing a complicated pattern of bands in the undeleted original clone pACY7. Inspection of the pattern of the bands, the order in which they decreased in size and then disappeared allowed a provisional PvuII restriction map of pACY7 to be drawn (diagram 3.4(c)). The second photograph in diagram 3.4 (b) shows PvuII digests of the same clones as in the PvuI digests. The clones which have been selected for this illustration contain relatively small inserts, thus several of the DNA fragments which are marked on the restriction map are already missing from these plasmids. The undeleted whole plasmid pACY7 is shown in lane 1. Unfortunately the photograph is not well exposed and the fragments 0.2 kb, 0.4 kb and 0.5 kb are not easy to see. Several of the bands can be seen to reduce in size and then disappear with successive clones.

(a) PvuI digests





(c)

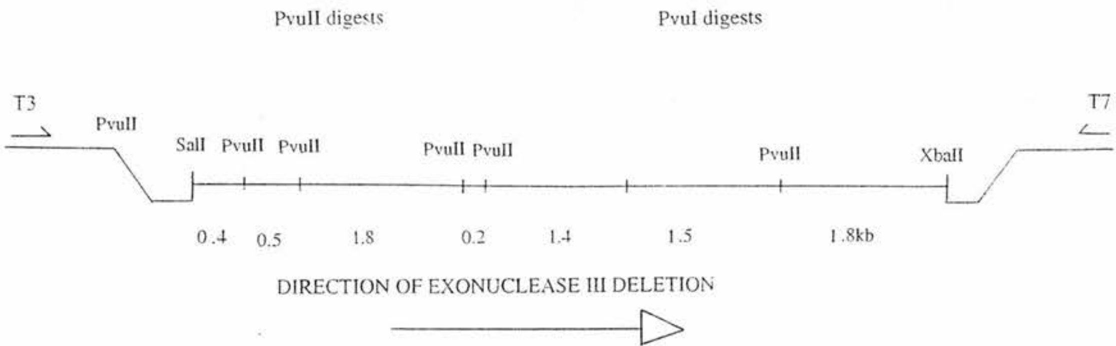


Diagram 3.4: PvuI and PvuII digests to size the deleted subclones from pACY7. PvuI digests (a) show the decrease in size of the cloned inserts. The order in which the bands disappear in the PvuII digests (b) correspond to the restriction map shown (c). The original undeleted plasmid pACY7 is shown in lane 1 of each photograph.

Sizing the inserts in the panel of clones found that there were substantial differences in size (>300 bp) between some clones derived from the same time point. Thus, the relative order of size of the cloned inserts in the deleted clones derived from these restriction digests was used to direct the order of the sequencing. Clones of all sizes were selected and sequenced in the first round of sequencing giving an even coverage of sequence across ACY7. This increased the likelihood of overlaps in the sequence data from each subsequent round of sequencing. The sequencing reactions and the running of the sequencing gels is described in Chapter 2, sections 2.6.7 and 2.10.

3.3 ACY7 sequence data assembly

The sequence data, which was read manually from the 'autorads', was assembled using the Seqaid and 'GCG' computer programs. The sequence data was checked against a database of vector sequences prior to being entered into the fragment contigs at points of overlapping sequence. Discrepancies in the overlapping sequence were checked against the original autorads. If they could not be resolved from this original data those plasmids were resequenced. Where appropriate, 7-deaza dGTP was used in the reaction mixes to remove compressions and 'stops' in the sequencing reactions (see section 2.10).

3.3.1 Completion of the seqence data

Once the majority of the sequence had been assembled, the remaining gaps were filled by synthesising sequencing primers and sequencing using the original pACY7 clone as the DNA template. In total 7 short, 18 base, primers were synthesised: P71209, P70501, P70708, P7EXO3, P70621, P70110 and P70110B. Their sequence is shown in table 2.1. The use of these primers filled the gaps between the aligned sequence contigs, completing the sequence of the 8 kb pACY7. The full

sequence is shown in diagram 3.5, the sequencing primers which were used are marked.

3.4 Sequence data from intron 1

The sequencing of plasmid pACZ was started using the T7 primer from the plasmid sequence. This generated about 250 bases of sequence from the downstream end of the cloned insert, that is the region immediately upstream from the XbaI site in intron 1 which is marked on diagram 3.5 (position 1110). Searching the GenEmbl databases using this sequence found the sequence for intron 1 which had already been submitted (Takimoto 1993: Accession number S62078). This sequence for the 3' end of PDGFA from the GenEmbl database extended to exon 2 of the gene. Comparison of this sequence to the data generated by this project revealed a region of overlap of about 650 bp. Since the sequence data in the region of overlap was found to be identical, sequencing of pACZ was not pursued any further. The additional sequence for the rest of intron 1 is included in the sequence analysis described below.

Exon 1
TGCTGCTCCTCGGCTGCGGATACCTCGCCCATGTTCTGGCCGAGGTTGGTGCCGCCCCCGCGCCCCGTCC 70

PstI
CTCGGCGGCTCCTCCGGCGCACACCCCCCGCCGGCTGGGGCGCCACGGGCT**CTGCAG**AGAGCGTTTTGGT 140

TCCTGCCAGGGTGGTCGATGTTTAGATTTTCCCAACTGGGTTCTATTTTCTGCCATGCCTAAGTGTGTGT 210

GGGGAGTTTCCAGCGTGTTGCGTCGGTCGCTCTGCCTCGGGGGGACCGGGGAGCAGCGAGGTGGACGAGT 280

PstI
GGGGAGAAACTTCGGGGGGCGCAGTGTCCTCTCCGGGGAAGGGGAGCTGGGGCCAGGCGGGCG**CTGCAG**C 350

AGCGGGGTGCGGGAGCGGGGAAGGGGGTGTGTGTGAGCCGGCGGAGCCTGGGGAGGAGAGAAGGGCGGCG 420

PstI
GATAAATAGGTAGGGGCGGCTGCTGCGGAACGCCCTGGCCGGGCAGGGTGGGCAGCGCAAG**CTGCAG**CCC 490

TCGGTTCCCAGTCCCCAAGCCCGGCGGCTCAGGGCCGGCGCAGTCCCGCCCGCCCGGGCTGCTTACAGGT 560

GGCTGAGCGAGGCAGCTGGGCTGGGCCAGGCGTCCGACCCTCCCACCGCGGCGTCCCCGCCCCCCATCCC 630

CATCCCCATCCCCATCCCCGGGAGCTTTGCTGTGATGAATGACTTTCCACAGCCCCCGTCGTGGGAACTT 700

AGTGCTTCAGGCTCCCCAGCGACCCCCGACTCTCCTTAAAAATCGGGGAGGGGGAGTGGGGGAGGCCAGA 770

GGTGAGGGTGGTCTCCGCAGGGCACGGAGAGCTGCGCGAGGTGGAAGTTTAACCCTCGCTTTCCCTGGAG 840

2260
GCTTCTTCTCCCGGCCGCCGGCTGGATTTGGGCGTCTTAGTCGGCGCTAGGGCCCGGCTGGGCGAGCCGG 910

GGTGGGGCGAGTGTGTGAGTGAGATGTGTGTGCGCGTGTGTGTGTCTGCGCGCGCGTGCTGTGTTTGTGA 980

2790
GACGGAGAGACCTCTCCACGGGTTCGGCCCGGGATTCCGGGCGTGAAATTTAAGCCCTGCCTGGATTAAA 1050

XbaI
GTGTTAGGGCTGCCGTGACTCTCGCCGACCCGTGACGCATTTTAACTTGATCTT**TCTAGA**TGTTGATCTC 1120

ACTATTAAATCAGAAACTTTCGGCCCCCTCGCTGCCCCCTAGACCTTCTGTAACTGACACCCCGGGTGGA 1190

GGGGCCGGAGTGCGGGGGCTTTTGTCTGAAAATATAAAGGGGGGGGGGCTTTGATTTTGCAAAGGGCGGA 1260

ACATGGATGCGGGTGGAGAATTTGGGGTCTCTTCCAGCCCCACCCCCGGACTTCTGGTTCCAGGGCCGGG 1330

GCGAGGGCTCCGCGGGGGCTGGGGCGGGGACTCGGCCGGCGCTCGCGGGGTCTCCTCGGGCCCGGGGCGG 1400

CGGCGCTCCTGGCGGCCGGGGGCGCGTGCCGGCGGGGTCGGGCCGCGGGCGCTGACCGTGTGGCCTCTGC 1470

Exon2
TTGCAGGAAGCCGAGATCCCCCGCGAGGTGATCGAGAGGCTGGCCCGCAGTCAGATCCACAGCATCCGGG 1540

ACCTCCAGCGACTCCTGGAGATAGACTCCGTAGGTAAATCGCGCCCCTTCCCTCCGCGCGCGGGGAGGGC 1610

GCGGGCGGGCGGCCTGTGCGCCCGAGCTGGGGCCGGGCGGGGCGCCGCCAGGAGCGCACCGAAAGGTCAG 1680

AATCCAATCCCCAGCCATAAGCGCCCGGGCGGATGAGTCAGGAGTTTCTCTTCCAGTCTCCCACCTTTCT 1750

CTCCCAGCGGCGGCCGAGGGCTGCTCGCTCGCCCCGGGATTGTTTTCGCCTTTAAGGAAAACGCCGCGTG 1820

TGTCACACAAAGTTCAGCCGCGCCTCCGCCAGGTGTGCGGGCCCCGCGCCTGCCCTCGGGCCGGGGGAGC 1890

GGCAGGATCCGGCCCAGACCCTGGGCGCAGAGCCCTGGCCCGTGCGCGGGGGCGCAGCCCGCGGAGGGAG 1960

StyI
GC**CCAAGG**GCAGGCGCGGCTCCCGGTCCGACGGGCGGGCGGGCGGAGGCGGCCCCGGGTCAGCCCGGCCC 2030

PstI
CCTCTGTAATCCCTGCGAGCTGGTGCCGCGTTCTTGCGCAGGGCTGGGCCTCTGTGG**CTGCAG**AAGCCGG 2100

StyI
TGTCTTTCTGTTTTCACGACCCAATGAAAAG**CCTTGG**ACGCATCTGCTGCCTGGGGGCTATTTTTGCGGC 2170

TCACCCTCGGCCCTGGGCCGGGCCACTAGAGAACACAGCTTGTTTGCAAGCACAGAGTTTTGCTGCGGGT 2240

TGGCAAGGGTTAAAGGGGCCTCTCACCCCGGGGGACCCCGGTGGACCACGGGTGGGTGGGAGGGGCGAGG 2310

Diagram 3.5: The gDNA sequence of PDGFA from exon 1 to exon 4.

GTGTAGGGTGGCCGCCCGCCCGCTCAGGGCTAGAGAGCCCCTTCCTAGAAGGCGAGGTCTCTCTTCCAGG 2380

P71209

CGTGCGTGGCGCTCTTTTCAAAATGAGCACTCGGCCTGGGAACTCCAGTCGGAGCCTTAACCTGGTTTAA 2450

AAAGGGGCTCTGCGATAAACTTGCTAAAAAGAGTAACTAAAGGCAGCAGAATTGGAAGAAGTTACGTGAC 2520

ATGAAGATTCTTCTCTCCCTGCAAAGGACTCTGAAAAAGGGGGGCATGCGGGGATTAAGTGAAATCACA 2590

TGGTCTAGTCATTTGGAAATTTGTACCTAATTATGGTGGTGGAGAGGCCTGAACTATTTTCGTTTCATGC 2660

P70501

CAAAACCATTTTCTAACAAGGCCGATGGTTTCTGTAGTTTCTTGACCCAGGAAATCCCTGGGCTGATCCT 2730

TATCGGGGGCGGAGGGTGGGGGTGCGGTAGAGTTTCACTCCCACAGCACAGCTGGGCGGGGGATTTTGA 2800

CCGGATCCCTCAGTGGATAACTCTAGGGAAAAGGATCTGGCTCTGGTTTGTGGGGGGGAATCCCTTGCCT 2870

GAATCTCGCTGGTCCTGGATGGGTTTGGAAAAGTGGGGGGGGGGGGCTGCTTCCTGGAGAAGGTTTCAG 2940

GGCTTTTTCTGGGGGGGGGGTGCTAGCAGAAGGGGAGGCTTGTGGAGAGCTGGGCCTCCTCGTGGGAACT 3010

PstI

GGCTTTGGAAG**CTGCAG**GGAAATGGGCCATTCACTTTCTGCTGGGACTATTCTTATTCTGGTCTTGGCGC 3080

CCCGCCTCCTCCACCCCCACCCATTATCGCCTTTAGCATTGCAGCTTCGTGGGGTATGCTGATGGCATTC 3150

CTCACCCGGGGGCTTCTAGGTCGTAGGGTCCGAACAGAAAAAAACAAGGATGCCCCTAAGCTAGTCCTGG 3220

PstI

CCGGTTTCAAATTCCCT**CTGCAG**TCATGTCCCAGGAGCTCTCTCCTCCTGCCCCAGGGAATGCAGAGTGG 3290

ATCTCAGCCCTCCTGTGTCCGTGAGAGGTTTTGGGGTCTGAATAGGATTCTCCTAGTGTCTGGGAGGCTT 3360

TGTGTGTGAGAATATGGGGGTGGCATCTACCTCCTCTCTTGCCCTGGTGGGGGACCTTTTGTAACATAGG 3430

StyI

GGGGCCT**CCTTGG**TCCTGTCCAGTCCCCTCCCCAGGTGTGGCTGAAAACAGAACCAGAGAAACCCACAGT 3500

CTCTCGCCTGAGGACAGGAGAGGCACCGGCTGCTCCTCCAGCTTGTAGGGGGTGTCTTTTCTTCAACTCC 3570

CGTTCTCTGTCGTAGGGAGGGTTGTAAACTGGTTGAGGTCAGTTCTCATTGGGTAAGCATCCTCCGGGTC 3640

CTGGACTCCCTTTTCAATCCAGAGGTGCTTCTGGTCTGTTAGGGGCAGGGGACAGTGCAGGGGCTCATGA 3710

GGAAGGGGCTGGGTGCTGGGAGAAACCCAACGCAACGAACTTTCCCTCGGAGTCCCTTTATAACCCTTCC 3780

AGGCTGATGTTCTTGGCATTAGGTGCCTCTAAGGCCCACGCTGTAGGAAGCAGAAAAGAGGCTTTGAGGA 3850

ATCCCAGGGGTGCGCCCCAGCTATGCTGACTCCTCCTCCAGGTGGGTGGGGCTTTGTGACTGGCATTTGG 3920

CTCCTCCCAGCATGCAGGCTGCGTCCCACTGTACAGTTGGGGAGACTGAGCCAGGCACCGTCTTGGCCGC 3990

TCCTTGACTGTGGCTGGAGGAGCACGGGTAGTGGGATTCCTCACCATCCTCCCGGAGCCCAAGCGTTTCA 4060

TTCGGAGGGTGTGCTGGAGTGAAGAGTGAAGGCTTTTCACCTGTGGCTTTTAGATCCAGAGGCCAGCATG 4130

TGGTCTTTTAATGACCGGCAGCCCGCGTTTAGGGCCTGTTTGTCCTGTGTGCAGTGGCAGCTGGGGAGCT 4200

TTGTTCCCCAAAGTGGTCTGGCCTGGCGCCCGTGGTTCATTGAGCGGCTGATAACATAAATAGCCCCATC 4270

StyI

TCGGGCCAGTAATGCCAGACCAGGACTGATGGTTGGGAGATTCTCCCTGTCCTGGAGTTGT**CCTTGG**ACA 4340

TCTGTGCCTAAAAGACAAAAGGGAAAGCCTCCTGGTTAACTGTCCTGGGAGCAGTTTTAGGGGTGATTCT 4410

TGCCGGCCAGGCCACCTGGGCCTGGGCTGCTCAGAGCTGTGTTTGGTGAGCTGTGGGCGGCTTTGGTGAG 4480

CGTGGTCCACTCTCGGCATCTTGAGCTGGGAGCAAAGGGTGGGCTTGGTAGACTGGGAGCTATTCCCCCT 4550

CTGTCCCTTGTCTCAAGAAATTTGCCAGCTCAGGCTTTCTGGGGGTCAATGAGCTAGCACGGCTGCCGTC 4620

TTCTCTGTTTATTAAGCCAGGGCTAGAGGGTGGCGGGGCACGGGGACTGTGCTCTGTGGACGCAGACCAC 4690

Diagram 3.5: The gDNA sequence of PDGFA from exon 1 to exon 4.

ATGAGTCTCCTGCGAGCCACTTCTGGTCCTGTTATATTTCTCTGGGCCTCAGTTTCCTCGGATGTGAAGA 4760

AGTGATAGCACCCACCGCACACTCAGAAGCTAAGTCATTGGTAAAGCAGCTAATGCAAACGTGGCTATCA 4830

CAAGGCATCCGAAAATGGGAGCTGGTGTTATTGAATTTTCTCCTACGAGTGGAGAACGGACCCGTTACTC 4900

StyI
TTTGGTTTGG**CCATGG**CAAGATGGCCACCACCAGGATGTGATTCATAAGTGTCACTTTTTCATGAAAAGA 4970

AATGCTGCTTGTAACAGTTGGTTGATGGGAAATTCTTACTGCATGAGTAGTTCTGTTCTGATCTTATTTT 5040

P70708
GGGTCTCAAAAAAATTAGTGTCATTGGAACATTGAGTCATTTTCGGTAAAAAGTTGGGCTCCGGGCGAGC 5110

TTCGTCTTTGCCTGACACCTTCTGTGAGGTTTGCGGCTTCATTTTAAATCCGCAGGCTCCTCGGGGGTG 5180

CGGAGGAAGCTGAGGTTTCTCCGGTTGGTTTGTGGATAGTCACGGATGCCTCGGTGGGCTGGGGAGAGAC 5250

ACTTTGGGCTGCTAGTCCTGAACCCAGGAGGTGCAGAGACCCATTAGCCGTCTGGGCTGGTGACCGGGAC 5320

TTCCCAAGCTTCCCAGGAGGGGGCAGGGAGATCTCTGTGCCTCTGGAGCATCCTGGCTGTGAGGCTCACT 5390

StyI
GTCTCCGCGCCTGTTCAC**CCTTGG**TTATATGGTGGGGGGGGGGCAGGATGGAGGGAGATGGGAAGAGACC 5460

StyI
CCCCAGCTGGGACCAGGGG**CCAAGG**CAAGGAGGCCCAGAGCCACCCGCGGTGCCCAGGTGCGGGCAGTGT 5530

StyI
CCATGGGCAGGACTGTGGGTCCTCTCCCTACTGCGGGATTTGGAGAGACCTGAGACTCCCCCAGCCTGTC 5600

TGCTGTAGCCGTCAGCGCCTGTGAAGGGGTGGTGTGTGCAGCTGGGTGGGGTGGCGGGGATAGGTTCTCA 5670

GTTCTGCTCCCCGGTCAGTGCTGGTCTTGGCTCTGGCCACTGCCCCACCTGCACCTGGCCGCCTGGCTGT 5740

AAACACGGCTCGCAGATGCCAGCTTCTCTCAACCCCTCTCCGCAGGCCGCCTCCCTCTGTCAGGCTGCGG 5810

GAACTCCTCCGTGTAGCCTCCACGAGAGGCCACGTGCCCTCCTGCCTCTCCTCATGCCGCTGGCCCTAGA 5880

GTAGACTCATTTCAGGAATGAGTGGGGTCTGGCCCGCAGCCCAGTCTGGCGGCTTCCAGTGCCCCTCAGA 5950

PstI
GGTTAGGCCACTTCTTCCAGGAGTGCTTTAGTGAGGGGGAGGGAGGGAGACGTACAGGCAC**CTGCAG**AAA 6020

TCTCAGCGGCCTCTGGCTTCTCCCTCGCTCCAGCCCCCAGACAGAGGTGGGGCCTGGCTTAGCCCTGTGC 6090

CCCTCGGGGAGGTGCACGGAGCTGGGGTCTCAGGCCCCTCCTGCCCCGGGGCAGCCCTCACCTGGGCTAC 6190

CTTGGAGGAGGCACCGTAGGGGAGACCACCTGCAACCTCTTCCCCGAAGGGCCCTGCGCTTCGTCACGGG 6230

CCCCTGAATGTGGGGACCTCTGTGCTAACGGGGCGTTTTCTGCGGTTTCCAGTGTTTCATGCCCTGCATC 6300

TCTGGAAAGCCCTGGTGGGAATCGAGGTTCTCCTGGAGGGACACCACATAATGGCGGGAGCTGTGAGGGG 6370

P7EXO3
CTGTGCCGCCAGGTGCCTGTTCCCCAGTGGCTCCCAAAGCTGGTCTGTGGGAAGTGCGGCTGGACAGGCC 6440

PstI Exon3
CAGGGCACAGCGCACGGGCCTGGGGCATTCACGGTGTTCTCCTTCCGCC**CTGCAG**GGAGTGAGGATTCTTT 6510

920601A 632
GGACACCAGCCTGAGAGCTCACGGGGTCCACGCCACTAAGCATGTGCCCGAGAAGCGGCCCCTGCCCATT 6580

CGGAGGAAGAGAAGCATCGGTGAGTCCAGGAGGCCGCGATGGGCAGGGCAGGGCCGGGTCGGGGTGAGTC 6650

CAGGAGGCCGCGATGGGCAGGGCAGGGCCGGGTGGGGAGGAGCTGCCCGCTCTCCCAGCGCAGTGGCCTC 6720

631
AGGCAAGCCACCCGTTCCCTCCTCCCTAAAATAGGCCTGGCCCTGGTCGCTCTGGCTCCCTGTGCTCCCG 6790

AGGCCTCCCTCCTTCGAGGGTGTCTCCTGCCCCAGCCCGCCTGCTCCTCCTGGAGTCCTGGTCTGGGCCA 6860

P70621
AGAGAGCTTCCAGGAGAGAAGAGCCATCATAAATTCCTTAAGCGCCCTGTCTTGTCGGGGCCTGTCCCCA 6930

GGAGCCTTGCCGGAAACGTGGTCGCGCCTAATTTCCAATTCGCAGTGTGGGCATCGTGCTCCTGCGCCGG 7000

CTGCTGAACTGGTGGGGACGCGCGTGCTGTGTCCGGCGGCTGCCCCCCGGCACCCTCACCACACCCGTGT 7070

Diagram 3.5: The gDNA sequence of PDGFA from exon 1 to exon 4.

```
ACCCCGCCCCACTCACACCTGGCTCCAGGGCTCCACCCCCGCGTCCATTTGGGGGATATTTTGTCCCTTC  7140
               StyI
TTACAGGACAGCCCCATGGCCCTCCAGCCCCCTGAGGAGGAACTAGCAGGTTTGGGGAGTTCAGGGGCCA  7210

GTCCCTGCTTTCTACCCAAGTCCGCGTCCTCGGGGTCCCACCTGATGGGTTTGAGCCCTGGTGCTGCTCT  7280

GCTCTGGGGCCCTGAGCAGTGACCCCCCTCTCTAGCTGTGGTTTCCTCCATGAAGATGGATTTTGTATCT  7350
                    7355                                          7397
TCCACCTTAAAGTGTCATGTGTGGCTGGCTAGTCCCAGCTGAGCAGCAGGCCCTCCCAGGTGACATTGCT  7420
                                                      P70110B
GTGGGTCTGCATTTGCACCCTTCAGACCACTTTTGAAGGCTAGAGCATGCAGCAAGTGCTGTAGGATTCA  7490

CTGTTAAGGCCCCCTCCTCCCCCAGCCACTGGCCCCACCTGGACTAGGTTCCCCTCTGCCCCTCCAGCCA  7560

CTGGCCGCACCTGGACTAAGTTCCCCTCTGCCCCTGCCACCCCCAGTCTCTGGCCCCACCTGGACTAAGT  7630

TCCCCTCTCCCCACCCCCAGCCACTGGCCCCACCTGGACTAAGTTCCCCTCCCCTCCGACCACTGGCCCC  7700
              P70110    PstI
ACCTGTCCAGGTCCCTCCAGCCTGCAGGCACCGGGCTGGGCTCCTCTGGATGTGTGGTGCTCCTGCCCAC  7770

CCAGCCCCTGAGCCTCTGCTCCCAGCTCAGCCTCTGCTGCCTGGGAGGAATCCTGGCCTGTGGCTTACCC  7840

TGGTTGCCCCCCAGGCCCAGCTGGAGCCGCCTCAGCCCTGGGGTGGGGCCGTGGTCGCAGAGGCCGGTCC  7910
                             Exon 4                               7970
CCGCTCACTGTGCCCCCGCCGTTGCAGAGGAAGCTGTCCCCGCTGTCTGCAAGACCAGGACGGTCATTTA  7980
                      SalI
CGAGATTCCTCGGAGTCAGGTCGAC  8005
```

Diagram 3.5: The gDNA sequence of PDGFA from exon 1 to exon 4.

The full sequence of PDGFA from intron 1 to the SalI site in exon 4. Primers 2260 and 2790 were used to amplify the GT repeat in intron 1 which is double underlined (positions 80 to 1010), it is described in section 3.6.1. The XbaI site at 1105 marks the start of the insert in plasmid ACY7. The seven primers which were used to complete the sequence data are underlined (P71209, P70501, P70708, P7EXO3, P70621, P70110 and P70110B). The primers 920601A and 631 in exon and intron 3 were used to amplify the exon 3 dimorphism which was used in the linkage analysis in Chapter 4, it is described in section 3.6.2. Primer pair 632 and 631 were used to screen the YAC libraries for PDGFA in Chapter 6. In intron 3 the minisatellite, described in section 3.8, is shown between positions 7400 and 7730 in bold italics. It was amplified using the primers 7355 and 7970 which are shown underlined. In addition to these, primer 7397 was used to sequence the cloned PCR products from the minisatellite. StyI and PstI restriction sites are shown. The StyI RFLP described by Ferns et al (1990) corresponded to a fragment from the StyI site in position 7155 to a site 340 bp downstream from the end of this sequence, in intron 4. A PstI fragment from ACY7 was used to hybridise the Southern blots shown in diagram 3.13, this was derived from the two sites in positions 6490 and 7730.

3.5 Analysis of the sequence data

Diagram 3.5 shows the genomic sequence of PDGFA from exon 1 to the SalI site in exon 4. A number of features in these sequence are annotated and these are discussed below in relation to the analysis of the sequence.

3.5.1 Database searches

The Genbank, Embl, and EST (expressed sequence tag) databases were searched for homologous sequences using the entire 8 kb of sequence. No sequences other than those for the PDGFA genes showed significant homology to any region of the sequence.

3.5.2 Analysis of the splice donor and acceptor sites

Since the presence of alternative splicing in human PDGFA and the existence of another, previously unknown exon within intron 6 of the rabbit PDGFA gene have been described (see section 1.3) an analysis of the intron/exon boundaries of this sequence was conducted. Two different computer programs were used to analyse the sequence. The programs, 'Netgene' (Brunak 1991) and 'Grail: genes recognition and analysis internet link' (Uberbacher 1988), both identified exons 2 and 3 of PDGFA. The exons 1 and 4 were not easily recognised by the analysis programs because the sequence for these exons marks the beginning and end of the data which was analysed. No other (unknown) exons were identified in this region of PDGFA.

3.6 Polymorphic loci in PDGFA

3.6.1 Intron 1: an interrupted GT repeat

An interrupted GT repeat in intron 1 was found about 120 bp upstream from the XbaI site (position 1105). It is double underlined in diagram 3.4. Comparison of the sequence of the GT repeat generated by this project with the sequence previously submitted to the GenEmbl database (Accession number S62078) found no difference in the sequence or the number of repeats.

Despite this, since dinucleotide repeats are frequently polymorphic, a PCR assay was developed using the primers 2260 and 2790 which are shown on the sequence. The reaction and cycling conditions for this PCR assay are described in section 2.10. PCR analysis of 20 unrelated individuals showed that this repeat is not polymorphic. This was not entirely surprising since the comparison of the two versions of sequence had found them to be identical. In addition, this GT repeat sequence is not perfectly repeated and such interruptions in repeats are associated with stability. A stable sequence would not be expected to be polymorphic.

3.6.2 Dimorphic markers in PDGFA suitable for linkage analysis

An easy method for finding polymorphic loci is to compare sequence data from different sources. If sequence data derived from different individuals is compared it may be possible to identify differences. Frequently these are simple variations such as single base deletions or substitutions. Comparison of the published cDNA and genomic sequence for PDGFA (Betsholtz 1986, Bonthron 1988, Sánchez 1991) was conducted by D.T. Bonthron. This identified a dimorphic locus in exon 3. A second dimorphic locus, in intron 4, was identified by sequencing cloned PCR products encompassing the intron 4 minisatellite (Bonthron 1992). PCR assays for

both of these dimorphisms were developed. The two markers are described below and are shown in diagram 3.6.

3.6.2.1 The exon 3 dimorphism

The exon 3 dimorphism is a silent C to T transition which is amplified using primers 920601A and 631K (see section 2.11 for PCR conditions). As shown in diagram 3.6, primer 920601A is a mismatched primer, the sequence mismatch creates an ApaLI restriction endonuclease recognition site when the C nucleotide is present in the target DNA. The dimorphic C to T transition abolishes this ApaLI site. Thus restriction of the PCR product provides an easy assay to determine whether the C or T base is present. Conveniently, this C to T transition creates a BbuI restriction endonuclease recognition site which can be used to distinguish between heterozygosity and partial digestion of the PCR product.

The amplified PCR product is 241 bp, on digestion, in the presence of an ApaLI site (a 'C' allele) a fragment 223 bases is generated. The two alleles 'T' (241 bp) and 'C' (223 bp) have frequencies of 0.74 and 0.26 respectively. These frequencies are derived from the analysis of 284 chromosomes comprising 64 unrelated British individuals and 78 parents from the CEPH families. This gives a calculated heterozygosity of 38.5%. Codominant inheritance was confirmed over 36 meioses in two families before analysis of the CEPH panel of families for the linkage analysis (described in Chapter 4).

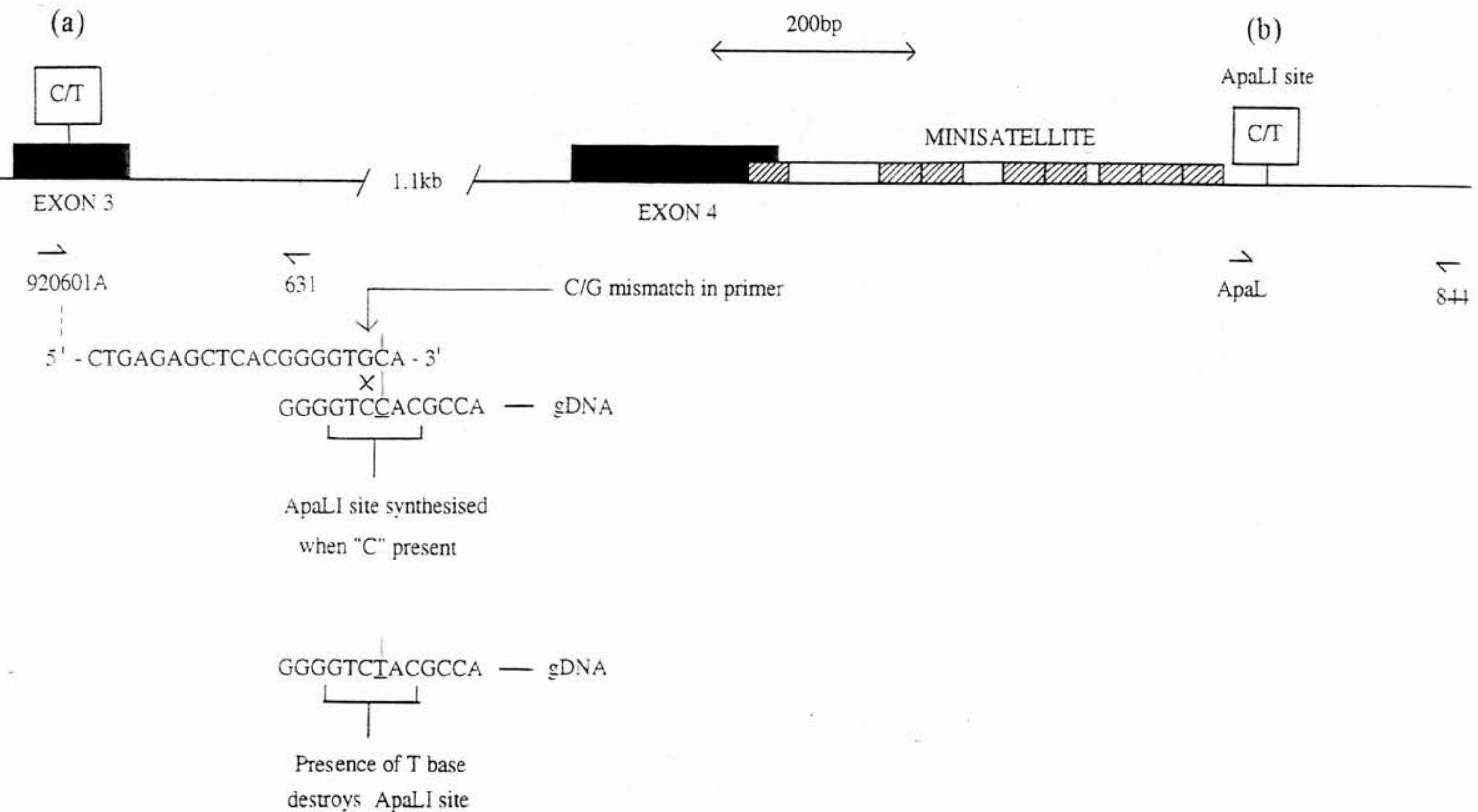


Diagram 3.6: The dimorphic loci in exon 3 and intron 4 of PDGFA.

(a) The dimorphic locus in exon 3 was analysed using primer pair 920601A and 631. The upstream primer creates the ApaLI site which is destroyed by the C to T transition. (b) The C to T transition in intron 4 which destroys a natural ApaLI site, was amplified with primers ApaL and 844.

3.6.2.2 The intron 4 dimorphism

The intron 4 dimorphism is also due to a C to T transition, it lies 489 bp downstream of the intron 4 splice donor site. It abolishes a natural ApaLI site as shown in diagram 3.6. The locus was amplified by the primers 844 and ApaL, generating a 256 bp product. Digestion of the PCR product in the presence of the ApaLI site ('+') generates fragments of 202 bases. The two alleles '–' (T residue: 256 bp) and '+' (C residue: 202 bp) have frequencies of 0.30 and 0.70 respectively; calculated from the same 284 chromosomes described above. Thus the calculated heterozygosity of this dimorphic locus is 42%.

The level of heterozygosity for each these two loci does not compare well to highly polymorphic loci such as the minisatellite MS31 which was used in the linkage analysis in Chapter 4 and has a heterozygosity of > 98%. However, the two dimorphic loci are easy to analyse using PCR, restriction digests and separation in agarose or polyacrylamide gels. Both of these dimorphic loci were used for typing the CEPH families in the linkage analysis described in Chapter 4. The PCR products, digested with ApaLI and separated in agarose and polyacrylamide gels for an example family (CEPH family 102) are shown in diagram 4.2.

3.7 Analysis of pACY7 for repetitive sequences

The sequence of PDGFA from exon 1 to exon 4 (in diagram 3.5) was compared to itself using the 'compare' program in 'GCG'. This program will mark sequences which match thus easily identifying regions which contain repetitive sequences. The graphical output or 'dotplot' of the analysis revealed a sequence repeat in intron 3 between the approximate positions 7500 to 7800. The 'dotplot' output, which is shown in diagram 3.7, suggests that this sequence is repeated five times.

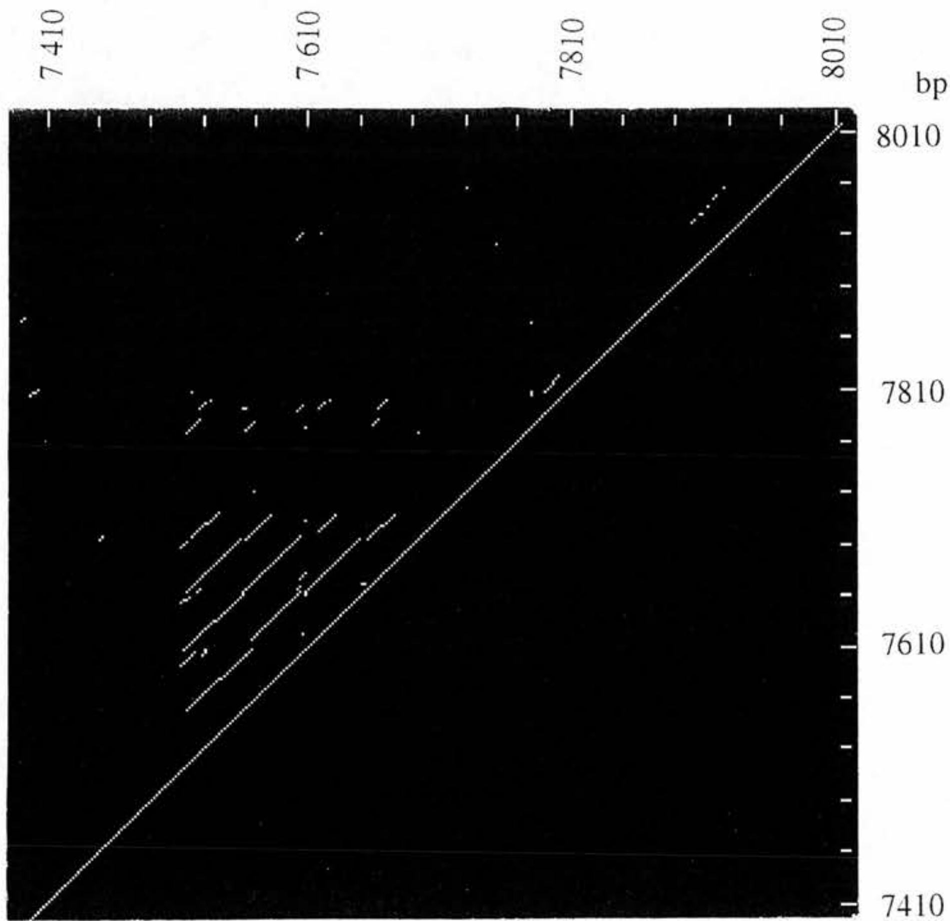


Diagram 3.7: Identification of a repetitive sequence by sequence comparison
A 'dotplot' output of the comparison of the entire sequence data in diagram 3.5 against itself. This diagram shows a closer view of the sequence between base 7400 to the end of the sequence (8000). A repetitive region containing up to five repeat motifs is identified between position 7500 and 7800, which is intron 3 of PDGFA.

Inspection of the sequence in intron 3 revealed a region about 250 bp upstream from exon 4 which contained a repetitive sequence. The repetitive sequence which spans from position 7397 to 7722 is marked on the sequence in diagram 3.5. An alignment of the repeats in this minisatellite is shown in diagram 3.8. There are six repeat elements which have been named A to F. Each repeat is based on a 49 bp sequence which is complete in repeat C. Gaps, shown by underlining, have been introduced into the other five repeats to allow the best fit alignment of this sequence. The size of each repeat is very variable with repeat D containing 44 bases while repeat F is only 14 bases in length. There are a number of base mismatches between the repeats which are shown underlined in bold in diagram 3.8. This alignment of the repeat sequences was determined retrospectively following the sequence analysis of the different sized alleles which is described in section 3.8.3. It reflects the sequence of the sections which were found to be missing from the smaller sized alleles (see below). There are no useful markers such as restriction sites which could be used to delineate each repeat sequence and a number of other alignments of this sequence are possible.

3.8 Analysis of the minisatellite in intron 3

A review of the restriction endonuclease recognition sites in the sequence of this region showed that this minisatellite repeat and exon 4 are both within the same StyI fragment. The upstream StyI site, at position 7155, is marked in the sequence shown in diagram 3.5. The downstream StyI site is positioned 340 bp downstream from the SalI site in exon 4 which marks the end of the sequence data shown in diagram 3.5. Minisatellites are commonly multiallelic. The balance of evidence from this sequencing analysis and earlier restriction mapping and hybridisation studies (see section 1.4) strongly suggests that this repetitive sequence accounts for the previously described StyI RFLP (Ferns 1990).

REPEAT

A AG**G**G_**C**CCCCT_____CCT____CCCCCAGCCACTGGCC

B CCACCTGGACTA**G**GTTCCCCTCTGCCCCT_______CCAGCCACTGGCC

C GCACCTGGACTAAGTTCCCCTCTGCCCCTGCCACCCCCAG**T**C**T**CTGGCC

D CCACCTGGACTAAGTTCCCCTCT_CCCC____ACCCCCAGCCACTGGCC

E CCACCTGGACTAAGTTCCCCT___CCCCT_______CC**GA**CCACTGGCC

F CCACCTG**TC**C_A**G**G_TCCC_T_______________CCAGCC

Diagram 3.8: Sequence of the repeats in the intron 3 minisatellite

The 49 bp repeat sequence in the minisatellite in intron 3 of PDGFA is repeated 5 times. The repeats have been named A to F. This sequence was derived from pACY7. Bold underlined bases indicate sequence mismatches between the repeats, single lines indicate where gaps have been introduced into the sequence to allow alignment of the repeats.

3.8.1 Database searches for homologous sequences

The sequence from this minisatellite was used to search the Genbank, Embl, and EST (expressed sequence tag) databases. As would be expected, the 40 bp repeat matched other repeats bases on the pentamer 'CCCCT'. However the level of homology between these sequences was not high (less than 40%). No significant sequence matches to this minisatellite were found.

3.8.2 PCR analysis of the intron 3 minisatellite

A PCR assay which spanned the intron 3 minisatellite was developed using the primers 7355 and 7970 (marked on diagram 3.5). This PCR assay was used to analyse this minisatellite to determine whether the sequence was polymorphic. The PCR products were cloned and sequenced to establish the repeat structure of different sized alleles. The PCR reaction conditions and the primer sequences are described in section 2.11 (table 2.2). It is worth noting that the conditions which were used are unusual and included the use of 750 mM KCl, DMSO and 7-deaza dGTP nucleotides. The region did not amplify well even under these optimised conditions and required twice the normal amount of PCR product to be loaded into the agarose gel to visualise the fragments. It seems likely that the problems which were associated with the PCR assay across this region are due to its very high G and C content.

PCR analysis of 62 unrelated individuals identified 4 different sized alleles. The PCR products from six of these unrelated individuals separated in an agarose gel are shown in diagram 3.9. This photograph clearly shows the four different sized PCR product bands or alleles which have been named 1 to 4, as shown.

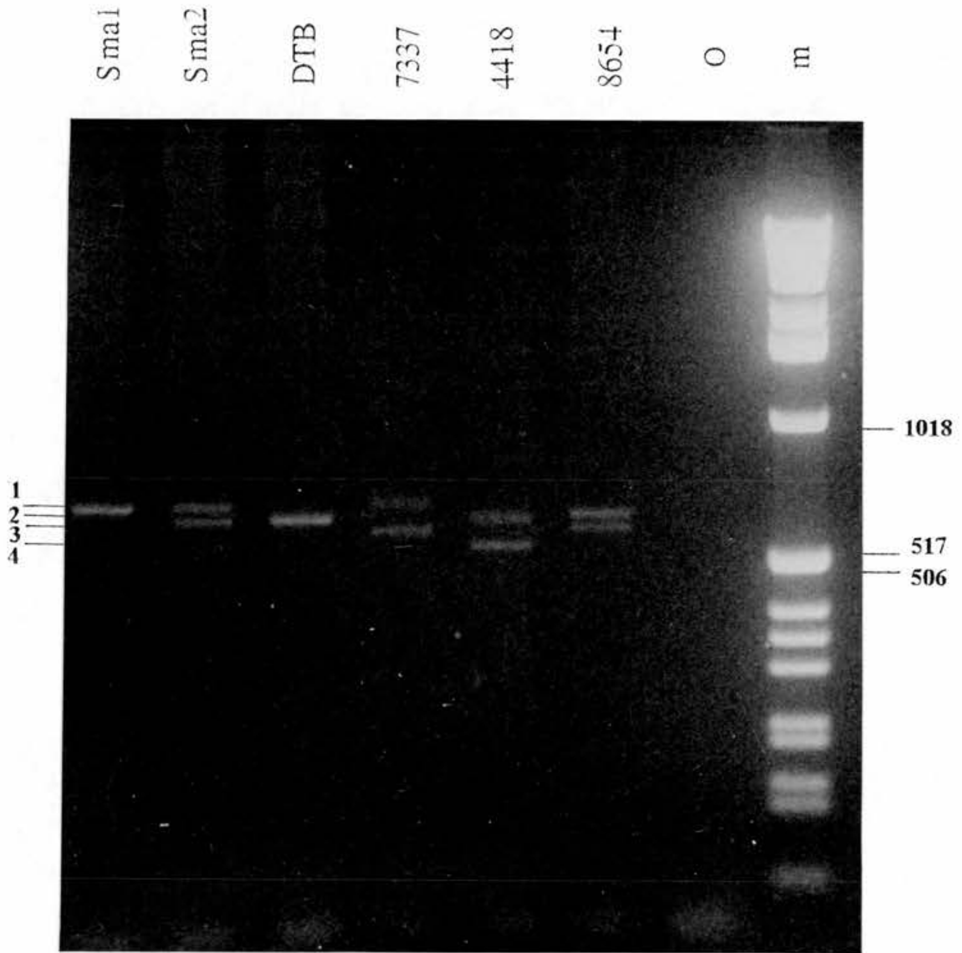


Diagram 3.9: PCR amplification of the intron 3 minisatellite

PCR products using the primer pair 7355 and 7970 amplified from 6 unrelated individuals showing four different sized alleles named 1 to 4. The DNA fragments from the alleles in the heterozygous samples were cloned and sequenced as described in section 3.8.

PCR analysis of 62 unrelated individuals found that the frequencies of these alleles were '1': 30.6%, '2': 59.7%, '3': 8.9% and '4': 0.8%, giving a calculated heterozygosity of 54.2%. The reported heterozygosity of the StyI RFLP on the basis of analysing 40 individuals was 40%. The frequency and size of each allele was not reported. However, despite this discrepancy it seems highly probable that this minisatellite is the basis of the StyI RFLP.

A four allele marker system with a heterozygosity of 54% such as this minisatellite would be ideal for the linkage analysis described in Chapter 4. In preparation for using this marker for linkage analysis and to confirm a Mendalian inheritance of these alleles, DNA samples from a number of families were tested by PCR. The results of these analyses revealed some inconsistencies. This is discussed fully in section 3.9.

3.8.3 Sequence structure of the minisatellite alleles

To analyse the repeat number and sequence of the different alleles amplified by the PCR assay described above, the PCR products from both alleles from the four heterozygous individuals shown in diagram 3.9 were sequenced. For this, the PCR product bands were cut from the agarose gel, extracted using 'Geneclean' and cloned into the 'PCRII' cloning vector as described in section 2.2. Two individuals, Sma1 and DTB, are homozygous, these samples were not analysed further. Thus 8 alleles were selected for sequencing giving three copies of allele size '2', two copies of allele sizes '1' and '3' but only one copy of allele size '4'. Allele 4 is the rarest allele with a frequency of 0.8%.

The resultant colonies were screened for the presence of an appropriately sized insert by PCR using the vector primers T3 and T7. Unfortunately the PCR product inserts were refractory to PCR using standard conditions and the colonies were eventually screened by filter hybridisation as described in section 2.7. This identified

successfully cloned inserts in 7 of the 8 alleles. One allele (Sma2/1) could not be cloned despite several attempts. It seems likely that this failure was due to technical problems due to the poor quality and amount of the PCR products. Allele 7337/1 appears to be the same size and presumably carries the same number of repeats was cloned successfully.

Sequencing of the cloned PCR products was conducted using the primers 7355 which was used to amplify the original PCR products, primer P70110 which was originally used in the sequencing of ACY7 and primer 7397 which is located immediately upstream of the minisatellite. These primers are indicated on the sequence shown in diagram 3.5. The sequence data for the 7 alleles was compared using the 'GCG' program 'pileup'.

3.8.3.1 Allelic repeat number variation

The 7 alleles which were cloned and sequenced were found to contain combinations of the same repeats A to F which were described in pACY7 and are shown in diagram 3.8. Comparison of the sequence data from these alleles to that from pACY7 did not identify any other basic repeat sequences. The size differences between pACY7 and the majority of the alleles was found to be due to the complete absence of some repeats from the smaller alleles. Alleles containing 4, 5, 6 and 9 repeats were found. These results are summarised in table 3.10. As the table shows, repeats A, C, E and F were common to all of the sequenced alleles. Repeat B was missing from repeats 8654/3 and 7337/3. In allele 4418/4 both repeats B and D were missing. By contrast, the largest allele 7337/1 which contained 9 repeats, was found to include a second copy of repeat B and two additional copies of repeat C.

Allele	Repeats present	No. of repeats
pACY7	A _ B _ _ C D E F	6
4418/4	A _ _ _ _ C _ E F	4
4418/2	A _ B _ _ C D E F	6
8654/2	A _ B _ _ C D E F	6
8654/3	A _ _ _ _ C D E F	5
7337/1	A C B C B C D E F	9
7337/3	A _ _ _ _ C D E F	5
Sma2/2	A _ B _ _ C D E F	6

Table 3.10: Minisatellite repeat variation

A summary of the repeat variation found in the sequenced alleles for the intron 3 minisatellite. The original sequence shown in diagram 3.8 corresponds to sample pACY7, it contains 6 repeats, as do alleles 4418/2 and 8654/2. Alleles 8654/3 and 7337/3 carry 5 repeats while allele 4418/4 contains only 4 repeats. Allele 7337/1, which is larger, breaks the pattern with 9 repeats which include two copies of repeat B and three copies of repeat C.

3.8.3.2 Subrepeat sequence variation

In addition to the variation in allele repeat number described above, a number of subrepeat sequence differences were found. These sequence variations are shown in diagrams 3.11 and 3.12 where the sequence for each repeat in each of the sequenced alleles is compared to the sequence from pACY7.

These diagrams show the aligned sequence from each sequenced allele for each repeat, a dashed line '-----' indicates where the sequence was identical, a dotted line '.....' indicates that the repeat was not found in that allele. Diagram 3.11 shows the sequence data for 7 of the 8 alleles which were sequenced, with those alleles of the same length, such as 8654/3 and 7337/3, adjacent. Diagram 3.12 shows the larger allele 7337/1 which was shown to contain extra copies of repeats B and C, as shown in table 3.10.

Comparison of the same repeat (A to F) between the different alleles (1 to 4) found that the base changes are relatively minor, single base substitutions. The clear differences between repeats A to F such as their length, which were described in diagram 3.8, are retained. There is not sufficient change in sequence in the repeat sequences to describe a new repeat. Base changes in the sequence were found largely in repeats C and D, while no sequence variation was found in repeat F. These base variations meant that only two of the seven alleles which were sequenced were found to have exactly the same sequence: they were alleles ACY7 and 8654/2.

A more complex pattern of repeat order and sequence variation was seen in allele 7337/1 which was the largest allele to be sequenced. As described above, this allele contained two copies of repeat B and three copies of repeat C in the pattern 'A C B C B C D E F'. These additional copies of repeats B and C showed a higher level of sequence variation (diagram 3.12).

```
REPEAT A

ACY7        AGGCCCCCTCCTCCCCCAGCCACTGGCC
SMA2/2      ----------------------------
4418/2      ----------------------------
8654/2      ----------------------------
8654/3      ----------------------------
7337/3      ----------------------------
4418/4      ----------------------------

REPEAT B

ACY7        CCACCTGGACTAGGTTCCCCTCTGCCCCTCCAGCCACTGGCC
SMA2/2      --------------------------------A---------
4418/2      ------------------------------------------
8654/2      ------------------------------------------
8654/3      ..........................................
7337/3      ..........................................
4418/4      ..........................................


REPEAT C

ACY7        GCACCTGGACTAAGTTCCCCTCTGCCCCTGCCACCCCCAGTCTCTGGCC
SMA2/2      -------------------------------------------------
4418/2      -------------------------------------------------
8654/2      -------------------------------------------------
8654/3      ----------------------------C--------------------
7337/3      ----------------------------C--------------------
4418/4      ----------------------------C--------------------

REPEAT D

ACY7        CCACCTGGACTAAGTTCCCCTCTCCCCACCCCAGCCACTGGCC
SMA2/2      -------------------------------------------
4418/2      T------------------------------------------
8654/2      -------------------------------------------
8654/3      T-----------G------------------------------
7337/3      ------------G------------------------------
4418/4      ...........................................

REPEAT E

ACY7        CCACCTGGACTAAGTTCCCCTCCCCTCCGACGCACTGGCC
SMA2/2      ----------------------------------------
4418/2      ----------------------------------------
8654/2      ----------------------------------------
8654/3      ----------------------------------------
7337/3      ----------------------------------------
4418/4      ------------G---------------------------

REPEAT F

ACY7        CCACCTGTCCAGGTCCCTCCAGCCTGC
SMA2/2      ---------------------------
4418/2      ---------------------------
8654/2      ---------------------------
8654/3      ---------------------------
7337/3      ---------------------------
4418/4      ---------------------------
```

Diagram 3.11: Subrepeat sequence variation in the minisatellite alleles.

The dashed line '---' indicates that the sequence was the same, a dotted line '...' indicates that the allele did not contain this repeat. Each repeat sequence is compared to the sequence of pACY7.

```
REPEAT A

pACY7          AGGCCCCCTCCTCCCCCAGCCACTGGCC
7337/1         ----------------*-----------

REPEAT C1

pACY7          GCACCTGGACTAAGTTCCCCTCTGCCCCTGCCACCCCCAGTCTCTGGCC
7337/1         ----------------------------C--------------------

REPEAT B1

pACY7          CCACCTGGACTAGGTTCCCCTCTGCCCCTCCAGCCACTGGCC
7337/1         ---------------------------------T--------

REPEAT C2

pACY7          GCACCTGGACTAAGTTCCCCTCTGCCCCTGCCACCCCCAGTCTCTGGCC
7337/1         ------------G------------------------*--------*--

REPEAT B2

pACY7          CACCTGGACTAGGTTCCCCTCTGCCCCTCCAGCCACTGGCC
7337/1         ---*--*-----**------------------T--------

REPEAT C3

pACY7          GCACCTGGACTAAGTTCCCCTCTGCCCCTGCCACCCCCAGTCTCTGGCC
7337/1         ----------------------------C--------------------

REPEAT D

pACY7          CCACCTGGACTAGGTTCCCCTCTCCCCACCCCAGCCACTGGCC
7337/1         ---------------------G---------------------

REPEAT E

pACY7          CCACCTGGACTAAGTTCCCCTCCCCTCCGACGCACTGGCC
7337/1         ------------------------------GC--------

REPEAT F

pACY7          CCACCTGTCCAGGTCCCTCCAGCCTGC
7337/1         ---------------------------
```

Diagram 3.12: Sequence variation in the repeats of allele 7337/1

The sequence of the minisatellite repeats of allele 7337/1 which had the repeat pattern 'A C B C B C D E F' are compared to the sequence of the basic repeats from pACY7. Note that pACY7 has the repeat structure A B C D E F. The dashed line '---' indicates that the sequence of 7337/1 was the same as pACY7 while the symbol '*' indicates where a base in pACY7 is deleted from 7337/1.

In the second copies of both repeats B and C there were some bases missing. These deleted bases have been marked (*) in the diagram. No deleted bases were identified in the other sequenced alleles shown in diagram 3.11. Despite this higher level of variation in allele 7337/1 the sequence and structure of the repeats were essentially the same as those described in pACY7. The sequence data suggests that repeats B and C have been copied rather than the presence of a novel repeat sequence.

3.9 Analysis of the inheritance of the minisatellite alleles

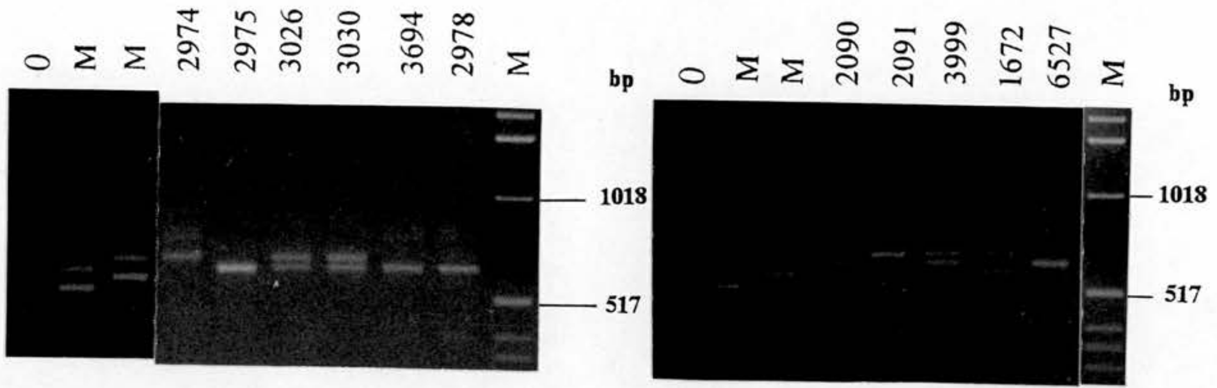
Inspection of the results of the PCR analysis of the 62 unrelated individuals described above in section 3.8.2 suggested some worrying features in the data. In a few samples of what were apparently homozygous individuals a trace smear could be seen above the amplified PCR products when they were separated in an agarose gel. This smear was easier to see early in the gel run before the PCR products had been fully resolved. As the gel was run further it became much fainter suggesting the presence of DNA fragments of variable size. The appearance of this faint smear aroused suspicion. Inspection of the data also suggested an abnormally high number of individuals homozygous for the rarer allele '3' although this was not found to be statistically significant.

Since it seemed likely that the faint smear which was seen in some homozygous individuals was caused by the PCR assay failing to amplify the second, possibly larger allele in these samples, PCR analysis was used to trace the inheritance of the minisatellite alleles in a panel of 10 nuclear families. The DNA samples for these families were kindly supplied by A. Gilfillan. The families which were selected all contained two parents and sibships of three or four children.

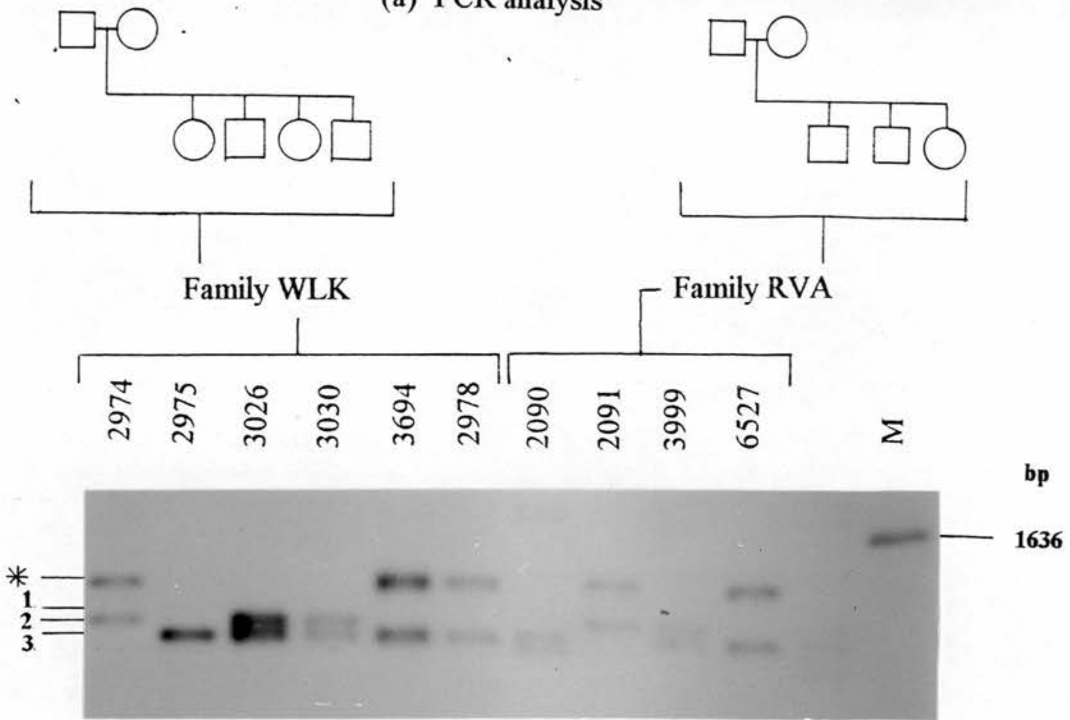
3.9.1 PCR analysis.

The DNA from the members of the 10 selected families was amplified using the primer pair 7355 and 7970 which are described above in section 3.8.2. Individuals 4418 (genotype 2/4) and 7337 (genotype 1/3) which are shown in diagram 3.8. were used to provide size markers for the amplified products. Seven of these families showed normal inheritance of the PDGFA intron 3 minisatellite within the limitations imposed by the informativeness of the parental alleles. These PCR assays amplified the same four alleles previously described. However, the PCR results from three families were unusual. In each of these families one parent appeared to be homozygous for an allele not present in the other parent. Inspection of the pattern of inheritance of the minisatellite alleles in the children in each of these families revealed children who appeared not to have inherited an allele from the homozygous parent. The other children in each sibship carried both maternal and paternal alleles.

The PCR results from two of these families ('WLK' and 'RVA') are shown in diagram 3.13(a). As shown in the diagram, the parents in family WLK are samples 2974 (father) and 2975 (mother). The genotype of 2974 appears to be 1/1 while 2975 has the genotype 2/2. The children, in samples 3026 and 3030, are both heterozygous with the genotype 1/2. However children 3694 and 2978 appear to have the genotype 2/2 with no apparent paternal contribution. A similar pattern is seen in family 'RVA' with the parents being samples 2090 (genotype 2/3) and 2091 (genotype apparently 1/1). The children are samples 3999 (1/2), 1672 (1/3) and 6527 (2/2). Thus, in this family individual 6527 does not appear to carry a maternal allele for the intron 3 minisatellite.



(a) PCR analysis



(b) Southern blot analysis

Diagram 3.13: Analysis of the inheritance of the intron 3 minisatellite alleles
The analysis of the inheritance of the intron 3 minisatellite alleles for two families (WLK and RVA) by PCR (a) and Southern blot hybridisation (b). In both families one parent and at least one child where found to carry a new larger allele (marked '*') which failed to amplify by PCR but was clearly visible on hybridisation of a radiolabelled probe to a Southern blot of PstI digests of gDNA.

3.9.2 Southern blot analysis

The samples which were used for this PCR analysis came from a bank of families which have been extensively studied. This meant that explanations for this result such as incorrect samples, while theoretically possible, seemed unlikely. The most likely reason for the observed data was that the PCR assay which amplified the minisatellite had failed to amplify some alleles. A failed PCR assay, or preferential amplification of a smaller allele, would explain both the patterns seen in the families analysed and the smears seen in the PCR products from some individuals.

To investigate whether additional minisatellite alleles which fail to amplify by PCR exist, DNA samples from the three families with abnormal PCR results were analysed by filter hybridisation to Southern blots of genomic digests. For this, a single copy probe was generated using PstI sites at positions 6490 and 7730 of the sequence in diagram 3.5. This 1.23 kb probe, named ACY7-PstI, spans from exon 3 of PDGFA to the end of the minisatellite repeat in intron 3. A PstI fragment was selected because it gave a suitably sized fragment which could be easily separated from the other fragments produced by the restriction digest of pACY7. PstI gDNA digests of the individuals in the three identified families were separated slowly in a 1.5% agarose gel which was blotted onto Hybond N^+ using standard techniques as described in Chapter 2. The resultant filter was probed with the radiolabelled probe ACY7-PstI. The results of this hybridisation for two families are shown in diagram 3.12. The results from the third family which was analysed are fully consistent with the results which are presented and add no further information.

The Southern blot analysis of these genomic digests clearly reveals another larger allele, which is indicated with an asterisk. This larger allele is seen in the parents 2974 and 2091 and in the children 3694, 2978 and 6527 all of whom had appeared to be homozygous in the PCR analysis described above. In each family the children who had appeared not to have inherited an allele from the 'homozygous' parent are

shown to carry this same larger allele. Interestingly all three families appear to carry the same larger sized additional allele. This fragment appears to be three size 'steps' larger than the nearest smaller allele '1' which contains 9 repeats. Thus this new larger allele may contain as many as 12 repeats. Without the ability to PCR amplify and clone this allele it is difficult to determine its sequence.

Since alleles with 4, 5, 6 and 9 repeats have been identified, it seems likely that there will also be individuals with alleles containing 7 and 8 repeats although these have not been identified. Given the Southern blot analyses it is possible that further alleles containing 10 and 11 repeats also exist but have not been identified. It is surprising that all three of these families carried the same larger sized allele rather than different sized larger alleles. It is also surprising that this allele is considerably bigger than allele '1' which was the largest allele detected by PCR. Without the option of PCR analysis for these larger alleles they would have to be sought by Southern blotting.

3.9.3 Heterozygosity of the minisatellite

This new larger allele, named allele '5' has been identified in at least three of the 62 unrelated individuals analysed in section 3.8.2. This gives it a frequency of 2.4%. Although this figure is based on PCR analysis only for the majority of the individuals which were tested. The adjusted frequencies for the alleles '1' and '2' are 29.8% and 58.1% respectively. The frequencies for alleles '3' and '4' remain the same at 8.9% and 0.8%. These figures, which can only be regarded as estimates, give a calculated heterozygosity of 56.5% for this minisatellite in intron 3.

A marker system such as this with a heterozygosity of 56.5% which could be amplified by PCR and analysed simply by separation in an agarose gel would be ideal for linkage analysis. However the problems described above caused by the failure of

the assay to amplify at least one allele mean that this loci cannot be reliably typed by PCR. Southern blot analysis of the minisatellite would be required to ensure that individuals were correctly typed. Processing a large number of samples by hybridisation would be time consuming, require larger volumes of DNA and the handling of a large amount of radioactivity. These problems mean that this marker is not ideal for linkage and it was not used to map PDGFA in the analysis described in Chapter 4.

3.10 Conclusion and discussion

The major objective of this sequencing project, to find polymorphic loci has been achieved and three potentially suitable markers have been described. Of these, the two dimorphic sites are used for the linkage analysis described in Chapter 4. The minisatellite in intron 3 explains the molecular basis of the StyI RFLP described by Ferns (1990). It is more informative than the two dimorphic loci however, as described above, PCR analysis of this minisatellite is not reliable and it was not used in the linkage analysis.

Analysis of the sequence data shown in diagram 3.5 excluded the presence of any further exons in introns 1 to 3 of PDGFA despite the previously described alternative splice forms of the gene. Following this sequencing project, only some of the very large intron 4 of PDGFA remained unsequenced. The sequence of this intron was extended from that published by Bonthron (1992) during the Rec-A assisted restriction endonuclease (RARE) cleavage experiments described in Chapter 9.

The position of the minisatellite repeat in intron 3 which ends about 215 bp upstream from the start of exon 4 is particularly interesting since a minisatellite overlapping exon and intron 4 has already been described (Bonthron 1992, see section 1.3). Diagram 3.14 shows, to scale, the positions of these two minisatellite sequences relative to exon 4. It is unusual to find two minisatellite repeats so closely flanking an exon and the functional effects on expression of an exon wedged between two minisatellites are open to speculation. The repetition of the exon 4 splice donor sequence also raises further questions about the functional effect of this sequence structure on splicing of PDGFA. The potential effect of these repeated sequences on the expression of the gene is not known and it would be interesting to investigate this further.

While both these minisatellites are based on a 'CCCT' repeat, their sequence and structure do not show a high homology suggesting that the two repeat sequences have separate origins. At least two other pairs of minisatellite sequences separated by only 1 to 2 kb have been described (Royle 1988). As with the two minisatellites within PDGFA, characterisation of these minisatellite pairs showed that they appeared to have different origins with different core sequences, heterozygosity and in one case opposite orientations. It is not clear whether these closely linked minisatellites have arisen systematically or by coincidence.

The sequencing of the different sized alleles of the intron 3 minisatellites in section 3.8 included allele 7337/1 which contained extra copies of repeats B and C. Allele Sma1/1 which looks the similar size 7337/1 failed to clone. Inspection of the photograph in diagram 3.9 reveals a qualitative difference between the two PCR

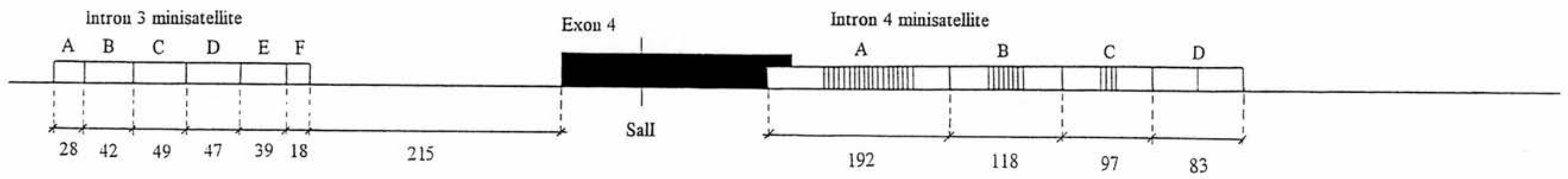


Diagram 3.14: The relative positions of the two minisatellites in PDGFA

A schematic diagram showing the two minisatellites in intron 3 (described above) and intron 4 (Bonthron 1992) of PDGFA which flank exon 4 of the gene. The intron 3 minisatellite contains 6 repeats which are marked A to F. It lies only 215 bp upstream from the start of the exon. The second minisatellite sequence starts within the exon, repeating the splice donor sequence within each of its repeats (marked A to E, from Bonthron 1992).

products with allele 7337/1 producing a more smeared band which is possibly slightly larger than Sma1/1. Given the sequence structure of 7337/1 and the subsequent identification of alleles which are refractory to PCR amplification, it seems possible that these two alleles may be different. Furthermore alleles containing 7, 8, 10 and 11 repeats have not been identified. It would be interesting to investigate further the existence and sequence of these larger alleles. However the problems associated with the PCR assays of the minisatellite make such an analysis difficult and limit its usefulness.

The subsequent work described in this thesis concentrates on mapping the PDGFA gene, firstly by using the dimorphic markers described above for linkage analysis (see Chapter 4), followed by physical mapping of the region. Physical mapping of the region was approached by pulsed field restriction mapping (described in Chapter 5) and the screening of YAC libraries which are described in Chapter 6.

4. Chapter 4: Linkage Analysis Of PDGFA

4.1 Localisation of PDGFA by genetic linkage

As described in the introduction (section 1.5.4), human PDGFA has been mapped to chromosome 7p22.1 - pter using *in situ* hybridisation and a panel of mouse-human hybrid cell lines containing fragments of human chromosome 7. Two cytogenetic landmarks, the translocation in patient DSE and the cell line 004 Rag 1-15, narrow the possible locations of PDGFA to within 7p21 or to the subtelomeric tip of 7p (shown in diagram 1.3). However further mapping required approaches which would provide information at a much higher resolution. Thus one of the aims of this project was to map PDGFA by linkage analysis. Placing PDGFA relative to other markers on a genetic linkage map was of particular interest given the various disease syndromes associated with deletions of chromosome 7p and the putative role of PDGFA in the various pathological processes discussed in Chapter 1.

Linkage analysis (described in section 1.12) requires (i) suitable polymorphic markers from the region of interest and (ii) DNA from suitable families. The search for polymorphic markers within the PDGFA gene is described as part of the sequence analysis described in Chapter 3. Other markers from chromosome 7p and the family samples which were used are described below.

4.2 The CEPH panel of families

The linkage analysis in this project uses the CEPH panel of families (Dausset 1990). The family structures in the CEPH panel are selected to give the maximum possible number of informative meioses for the number of individuals studied. Most families consist of two sets of grandparents, two parents and, most importantly, a large sibship of children (average size 8.3). Table 4.1 summarises the family structure and origin of each of the 39 families which were used in this study.

Family	Source	Sibship size	No. grandparents	No. males
2	F	7	0	2
12	F	7	3	2
17	F	8	0	7
21	F	6	0	3
23	F	6	0	2
28	F	7	0	3
35	F	8	0	3
37	F	6	0	4
45	F	5	0	3
102	V	14	0	6
104	V	10	4	3
884	A	12	4	5
1331	U	9	4	4
1332	U	10	4	5
1333	U	8	4	7
1334	U	7	4	6
1340	U	6	4	4
1341	U	8	4	2
1344	U	9	4	4
1345	U	7	4	5
1346	U	8	4	5
1347	U	9	4	6
1349	U	8	4	3
1350	U	7	4	2
1362	U	10	4	5
1375	U	7	4	3
1377	U	7	4	7
1408	U	7	4	3
1413	U	15	2	11
1416	U	8	4	4
1418	U	8	4	2
1420	U	6	4	0
1421	U	8	2	3
1423	U	8	4	4
1424	U	8	4	6
13291	U	7	4	4
13292	U	6	4	3
13293	U	7	4	5
13294	U	6	4	5

Table 4.1. The CEPH panel of families

DNA samples from this panel of families was used in the linkage analysis. Source of families: A: Amish pedigree, F: France, U: Utah, V: Venezuela.

The CEPH panel of families are ideal for linkage analysis. Their use also allows the use of data from other markers in the same chromosomal region. This data is made available on databases. The DNA samples were a kind gift from Dr. A. Wright and K. Wakefield, MRC Human Genetics Unit, Edinburgh.

4.3 Polymorphic markers used in linkage analysis

4.3.1 Polymorphic loci within PDGFA

Three polymorphic loci which were potentially suitable for linkage analysis were identified in PDGFA: the two dimorphic markers in exon 3 and intron 4 and the multiallelic minisatellite in intron 3 (described in Chapter 3). All three of these markers can be amplified by PCR and would have been suitable for typing the CEPH families for linkage analysis. However the intron 3 minisatellite PCR is a less reliable assay since the largest of the alleles which have been identified is refractory to amplification by PCR and identification of this allele requires a Southern blot of genomic DNA. This requires far more DNA, handling radioactivity and is a much lengthier process than PCR analysis. Since a large number of samples needed to be analysed it was decided to use the two PCR assays which detect the two dimorphic markers in exon 3 and intron 4 of PDGFA.

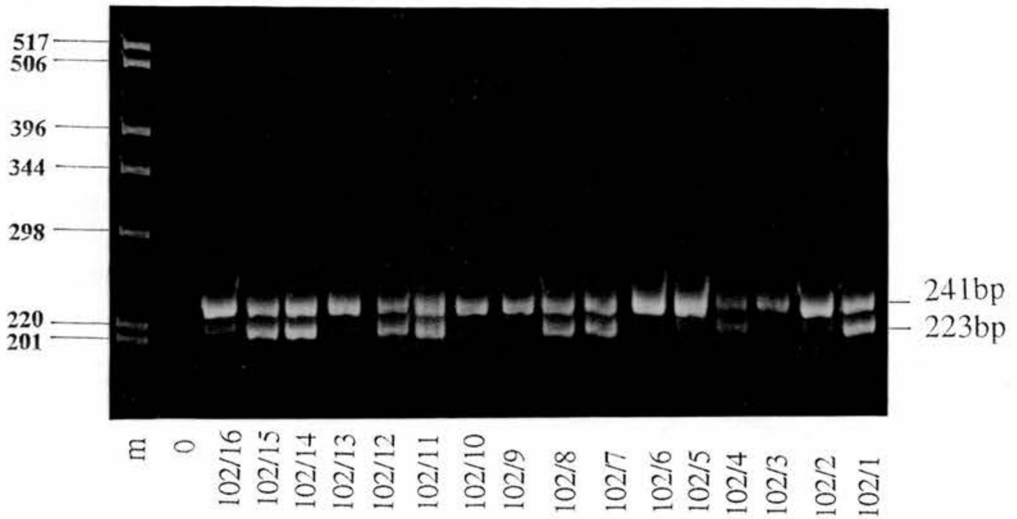
The potential disadvantage of using the two dimorphic markers was that both required restriction of the PCR products with ApaLI. Thus it was possible that failed or incomplete digestion of the PCR products could lead to the wrong assignment of alleles for individuals. All results were assigned and checked independently by both myself and D.T Bonthron. Any dubious or conflicting results were reanalysed. In addition, the C to T transition in the exon 3 dimorphism which destroys the artificial ApaLI site also creates a BbuI restriction site. This could be used to check results although this was seldom found to be necessary.

Thus the two dimorphic markers in exon 3 and intron 4 of PDGFA were used to type the CEPH family samples. Initially the parents of each family were typed. Families in which at least one parent was heterozygous for a marker were analysed fully for that marker. The PCR conditions which were used are described in Chapter 2, sections 2.11 and table 2.2. Following amplification, the PCR products were digested with ApaLI (or its isoschizomer Alw44I). The exon 3 PCR products were run out on a 10% polyacrylamide gel, the intron 4 PCR products were separated on a 3.5% agarose gel. The restricted and separated PCR products for an example family (CEPH family 102) for each marker are shown in diagram 4.2.

4.3.2 Markers along chromosome 7p

The linkage analysis was conducted using the combined haplotype data for *PDGFA* (see below) along with data for the markers MS31 (*D7S21*), CRI-S202 (*D7S108*), CRI-S127 (*D7S89*) and CRI-S193 (*D7S103*). The data for these markers was obtained from the CEPH database (version 6.0). These markers were the most distal loci used in the first chromosome 7p linkage map shown in diagram 1.5 (Mishra 1992) which was the most recently available linkage map for chromosome 7p at the start of this study. Table 4.3 lists the heterozygosity of each of these markers. The marker, *D7S10*, which lies between *D7S89* and *D7S103*, was omitted from the analysis. Using marker *D7S103* instead of *D7S10* extended the analysis towards the centromere, covering a larger area and including a marker mapped to 7p21. If *PDGFA* had appeared to map to the region around *D7S10*, this marker would have been included in the subsequent analyses however this was not found to be necessary.

(a) Exon3 dimorphism



(b) Intron 4 dimorphism

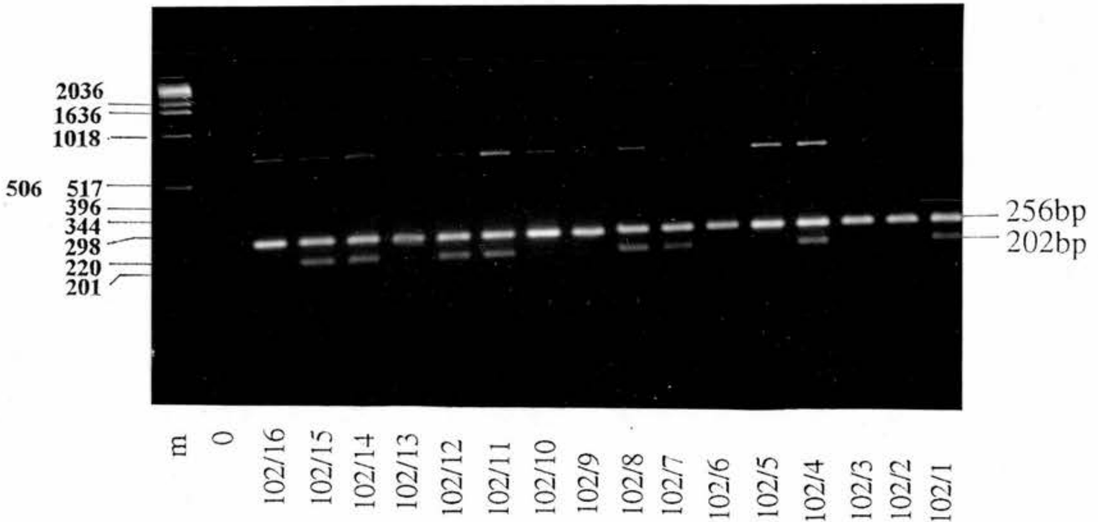


Diagram 4.2: The PDGFA dimorphic markers used in the linkage analysis
PCR analysis of the PDGFA dimorphisms illustrated using DNA samples from CEPH family 102. Individuals 1 and 2 are the mother and father while samples 3 to 16 are the fourteen siblings. The exon 3 PCR products (a) are amplified using the primer pair 90601A/631. The intron 4 PCR products (b) use primers ApaLI and 844. All the PCR products have been digested with ApaLI.

Locus		Probe / PCR assay	Heterozygosity
PDGFA	exon 3	PCR	0.38
	IVS 4	PCR	0.42
	haplotype	combined data	0.73
D7S21		LMS31	>0.98
D7S108		CRI-S202	0.31
D7S89		CRI-S127	0.61
D7S103		CRI-S193	0.59
D7S531		PCR	0.77

Table 4.3: The markers used in the linkage analysis.

The chromosome 7p markers whch were used in the linkage analysis combined the PDGFA markers (described in Chapter 3) with five other loci from the region. Most of these markers were ideal for linkage analysis with heterozygosities greater than 50%. The PDGFA haplotye heterozygosity was derived from the CEPH family data.

A more recently described marker, *D7S531*, was not used in the Mishra linkage map. However the more recent linkage maps which are shown in diagram 1.5 do include *D7S531* (Grzeschik 1994, Gyapay 1994). As discussed in section 1.12, these linkage maps place *D7S531* distal to *D7S21* or cannot place the two markers relative to each other. However, the data on which some of these maps are based includes very few families or is data which has not been checked. Following publication of these linkage maps, the two-point linkage and Cri-Map linkage analysis in this study was repeated using *D7S531*.

4.4 PDGFA haplotype frequencies

The two dimorphic sites in PDGFA are separated by only about 420 bp (see diagram 3.14), however they lie either side of the two minisatellites in introns 3 and 4, described in Chapter 3. Minisatellites have been shown to be 'hot-spots' for recombination, thus there is a possibility that recombination events may occur between the two markers in PDGFA. Bearing this in mind, each locus was typed independently for each individual in the CEPH families. Inspection of the data, both by hand and using pairwise linkage analysis using the MLINK program, showed that there were no recombinants between the two markers in PDGFA in the individuals studied in this analysis. Thus, the data for the two markers were combined into one haplotype for each individual. As table 4.3 shows, the calculated combined heterozygosity of the two PDGFA markers was 0.73 which provides a highly informative marker system, ideal for linkage analysis. All the linkage analysis presented in this report was performed using the combined haplotype data.

A chi-squared analysis was applied to the haplotype data. Table 4.4 shows the observed frequency of each haplotype and the calculated expected frequencies for each of these haplotypes (calculated from the allele frequencies given in Chapter 3).

Haplotype	Observed no.	Expected no.	χ^2	p
+C	51	41	2.44	ns
+T	110	117	0.42	ns
–C	4	18	10.89	<0.005
–T	61	50	2.42	ns
Total	226	226	n/a	n/a

Table 4.4. PDGFA Haplotype frequencies.

The two dimorphic markers span the two minisatellites in intron 3 and intron 4 of PDGFA. Despite this, the two loci, which are only 420 bp apart, are in linkage disequilibrium for the haplotype '-C' which is observed with a much lower than expected frequency. The expected frequencies were calculated from the allele frequencies: '–' = 0.3, ' + ' = 0.7 and C = 0.26, T = 0.74.

A striking difference can be seen for the '–C' haplotype which was found in only 4 chromosomes in contrast to an expected 18 chromosomes (χ^2=10.89, 1 df, p = 0.001). The other three haplotypes were found at the expected frequencies and the observed heterozygosity was 72%, which is close to the expected 73%. The finding of linkage disequilibrium also suggests that these two loci do not span a recombination 'hot-spot'. However it is important to remember that only a relatively small number of meioses are being studied in this analysis.

4.5 Two point linkage analysis.

Data for the loci *D7S21*, *D7S108*, *D7S89*, *D7S103* and *D7S531* was combined with the data for *PDGFA* to produce an appropriately formatted file using the Genbase program (version 1.1). Following this, two point linkage analysis was carried out using the MLINK program from the LINKAGE package, version 5.2 for ultrix (Lathrop 1984) using equal recombination fractions for males and females. Table 4.5 shows the pairwise lod scores for linkage between *PDGFA* (haplotype data) and the previously described markers on chromosome 7p.

A maximum pairwise lod score of 39.49 (θ = 0.14) was obtained between *PDGFA* and MS31 (*D7S21*), *D7S21* was the most telomerically placed marker on the 7p map published by Mishra (1992, see diagram 1.5) and is a highly informative marker (heterozygosity > 98%). *D7S531* shows a peak lod of 4.11 at θ = 0.264, these maximum two-point lod scores indicate that the genetic distance between *PDGFA* and *D7S531* (28.3 cM) is much larger than between *PDGFA* and *D7S21* (14.0 cM). The third marker from chromosome 7p which shows significant linkage to *PDGFA* in these two point analyses is *D7S89* with a lod score of 9.86 at θ = 0.15. The other markers are less informative; *D7S103* reaches a just significant lod score of 3.96 at θ = 0.21; *D7S108*, which is a much less informative marker with a heterozygosity of 0.31, does not show significant linkage to *PDGFA*.

Locus	$\theta = 0.5$	0	0.01	0.05	0.1	0.2	0.3	0.4	$\hat{z}$	$\hat{\theta}$
D7S21	$-\infty$	0	-2.08	30.03	38.33	37.18	27.38	13.15	39.49	0.14
D7S108	$-\infty$	0	-10.09	-3.56	-1.16	0.54	0.85	0.52	0.85	0.29
D7S89	$-\infty$	0	-3.63	6.48	9.29	9.45	6.99	3.29	9.86	0.15
D7S103	$-\infty$	0	-12.85	-1.58	2.13	3.95	3.28	1.58	3.96	0.21
D7S531	$-\infty$	0	-21.94	-5.80	-0.03	3.73	3.92	2.33	4.11	0.264

Table 4.5: Pairwise linkage anlaysis using the *PDGFA* haplotype data.

Pairwise linkage analysis for *PDGFA* against chromosome 7p markers *D7S21*, *D7S89*, *D7S108*, *D7S103* and *D7S531*.

$\hat{z}$ and $\hat{\theta}$ show the maximun lod scores obtained and the recombination frequency for that value.

4.6 Multipoint linkage analysis

A multipoint analysis using the haplotype data and the three chromosome 7p markers *D7S21*, *D7S108* and *D7S89* was undertaken using the LINKMAP program from the LINKAGE package (Lathrop 1984). The results are shown in diagram 4.6. For these calculations, the distances between adjacent markers were set equal to the maximum θ derived from two-point intermarker analyses on the same data set. Map distances were calculated using Kosambi's mapping function.

This multipoint analysis assigns *PDGFA* a sex averaged location 14 cM telomeric to MS31 (*D7S21*) with a maximum lod score of 41.53 at $\theta = 0.13$. The next highest lod score peak is seen beyond *D7S89* moving towards the centromere in the position $\theta = 0.32$. This peak is 9 lod scores lower than the highest peak, giving odds for a telomeric location for *PDGFA* against this location of $1 : 10^9$. Thus *PDGFA* is assigned a telomeric location on chromosome 7p, extending the previously available linkage maps.

The marker *D7S531* was not included in this multipoint analysis since the data available cannot place this marker in an unambiguous position on the chromosome 7p map. Unambiguous placing of each marker other than the test locus (in this case *PDGFA*) is required prior to running the LINKMAP program. However a multipoint linkage analysis was conducted using just *D7S21* and *D7S531* as fixed loci with *PDGFA* as the test locus. This multipoint linkage analysis clearly placed *PDGFA* on the opposite side of *D7S21* from *D7S531*, with odds of $1{:}10^9$ against a location for *PDGFA* closer to *D7S531*. This is consistent with the results of the two-point linkage analysis presented above (table 5.4).

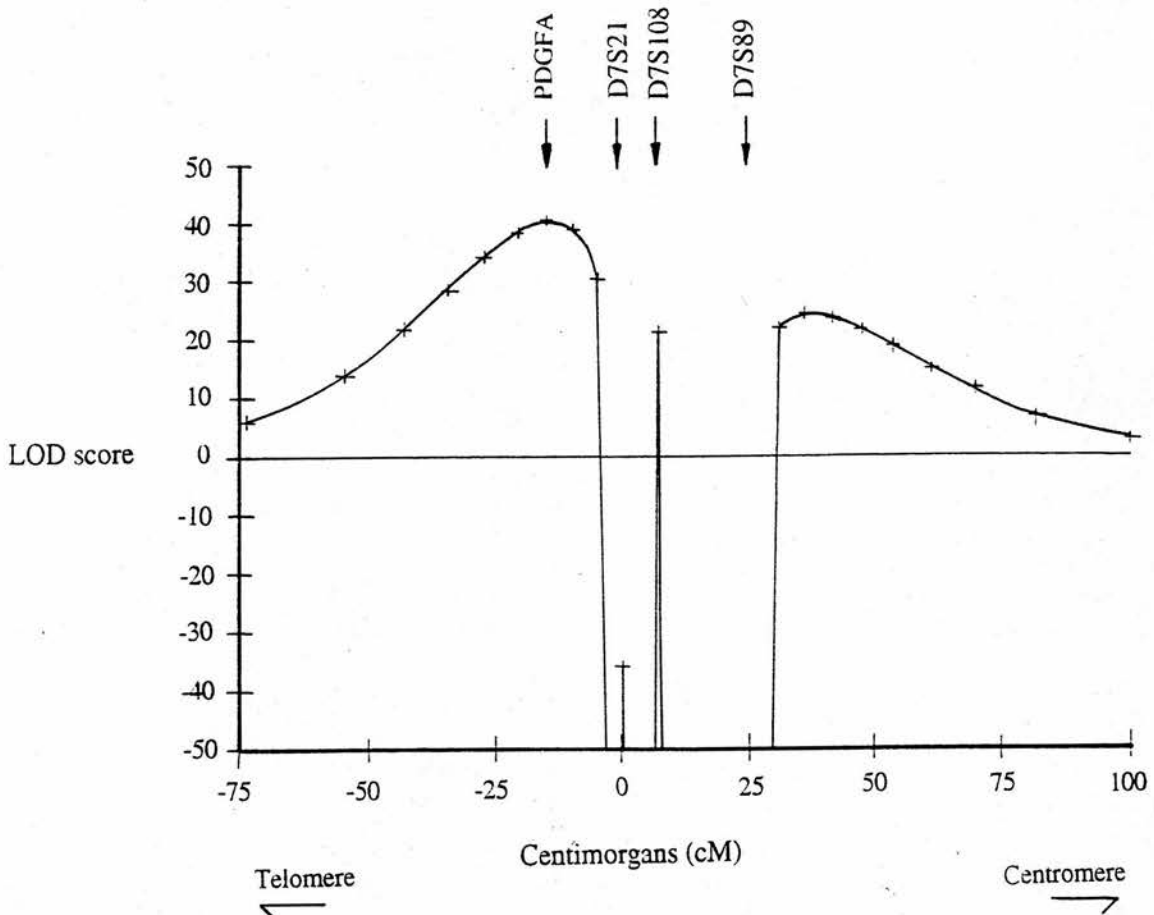


Diagram 4.6: Multipoint linkage analysis

Multipoint linkage analysis of PDGFA against the loci *D7S21*, *D7S108* and *D7S89* using the LINKMAP program.

4.7 Cri-Map analysis.

Since LINKMAP is not designed for the analysis of large amounts of data across large distances, an alternative computer program Cri-Map (version 2.4), was also used. Diagram 4.7 shows the sex specific and sex averaged chromosome 7p genetic linkage maps derived from using the Cri-Map program and the chromosome 7p loci *PDGFA*, *D7S21*, *D7S108*, *D7S89*, *D7S103* and *D7S531*. Map distances were calculated using Kosambi's mapping function.

The Cri-Map sex-averaged analysis places *PDGFA* 16 cM telomeric to *D7S21* although a large difference of 6.5 cM can be seen between the male and female maps. This distal position for *PDGFA* is favoured against inversion of *PDGFA* and *D7S21* by odds of 1: 10^{16}.

This analysis could not place *D7S531* relative to the other loci used in this map. The most likely position for *D7S531* is between *D7S21* and *D7S89*. The odds against a location for *D7S531* distal to *D7S21* are 1:1000. The *D7531* locus was of particular interest since it is used in other linkage maps of chromosome 7p (see diagram 1.5) and was potentially distal to *D7S21*. This would have meant that *D7S531* would have been the closest marker to *PDGFA*, however this Cri-Map analysis suggests that *D7S531* is proximal to *D7S21* when compared to *PDGFA*. This confirms the results from the pairwise and multipoint analyses, presented above, which both suggested that PDGFA is further from *D7S531* than from *D7S21*.

The Cri-Map analysis places PDGFA as the most distal marker on chromosome 7p. No other marker has been placed closer to *PDGFA* than *D7S21* and no flanking marker has been identified.

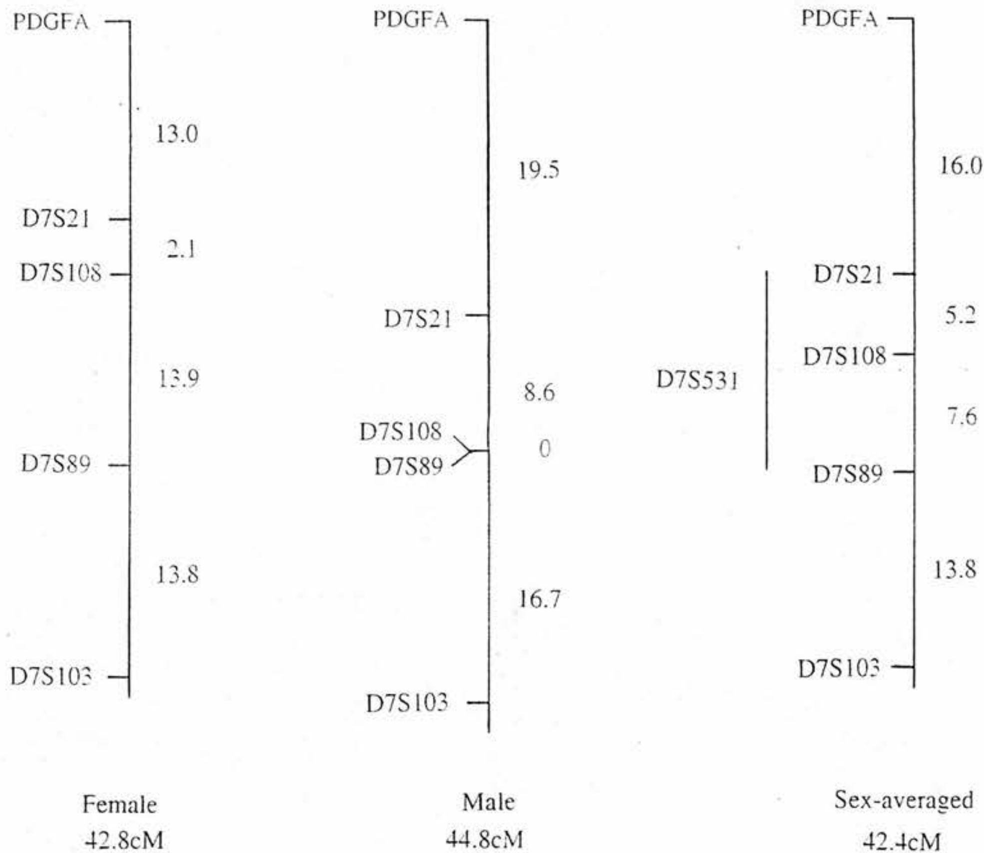


Diagram 4.7: Cri-Map analysis of chromosome 7p.

Cri-Map analysis of the chromosome 7p loci *PDGFA*, *D7S21*, *D7S108*, *D7S89*, *D7S103* and *D7S531*. *PDGFA* is placed at the telomeric end of the chromosome 7 map, 16 cM from *D7S21*. There is large difference of 6.5 cM between the male and female maps. *D7S531* is located proximal to *D7S21* although it could not be placed relative to *D7S108* and *D7S89*.

4.8 Eurogem Consortium.

The haplotype data for *PDGFA* was submitted to the Eurogem consortium collaboration and was incorporated into their data. The Cri-Map program was used to build whole chromosome maps for each chromosome in the human genome using the collated data generated using the CEPH panel of families. The chromosome 7 linkage map was compiled by F. Badbanchi in Marburg, Germany (Badbanchi 1994).

As can be seen in diagram 4.8, this map uses a different set of markers across chromosome 7p to those used in the analysis described above. Unfortunately these two maps have no markers in common.

Like the earlier analysis, the Eurogem linkage map places *PDGFA* as the most distal marker on the chromosome 7p linkage map. In the sex-averaged map from this analysis, *PDGFA* is placed 10.0 cM telomeric to the nearest proximal marker which is *D7S596*. The difference between the male and female linkage maps, of 3.8 cM is less marked than the earlier analysis in diagram 4.7.

The locus *D7S531* was initially used in the analysis for this Eurogem linkage map. However it was not included in the final published map because it could not be placed in an unambiguous position on the linkage map, that is with odds of 1:1000 or greater. *D7S531* was mapped to the region between *D7S462* and *D7S513* which is clearly proximal to *PDGFA* and its nearest linked marker, *D7S596*.

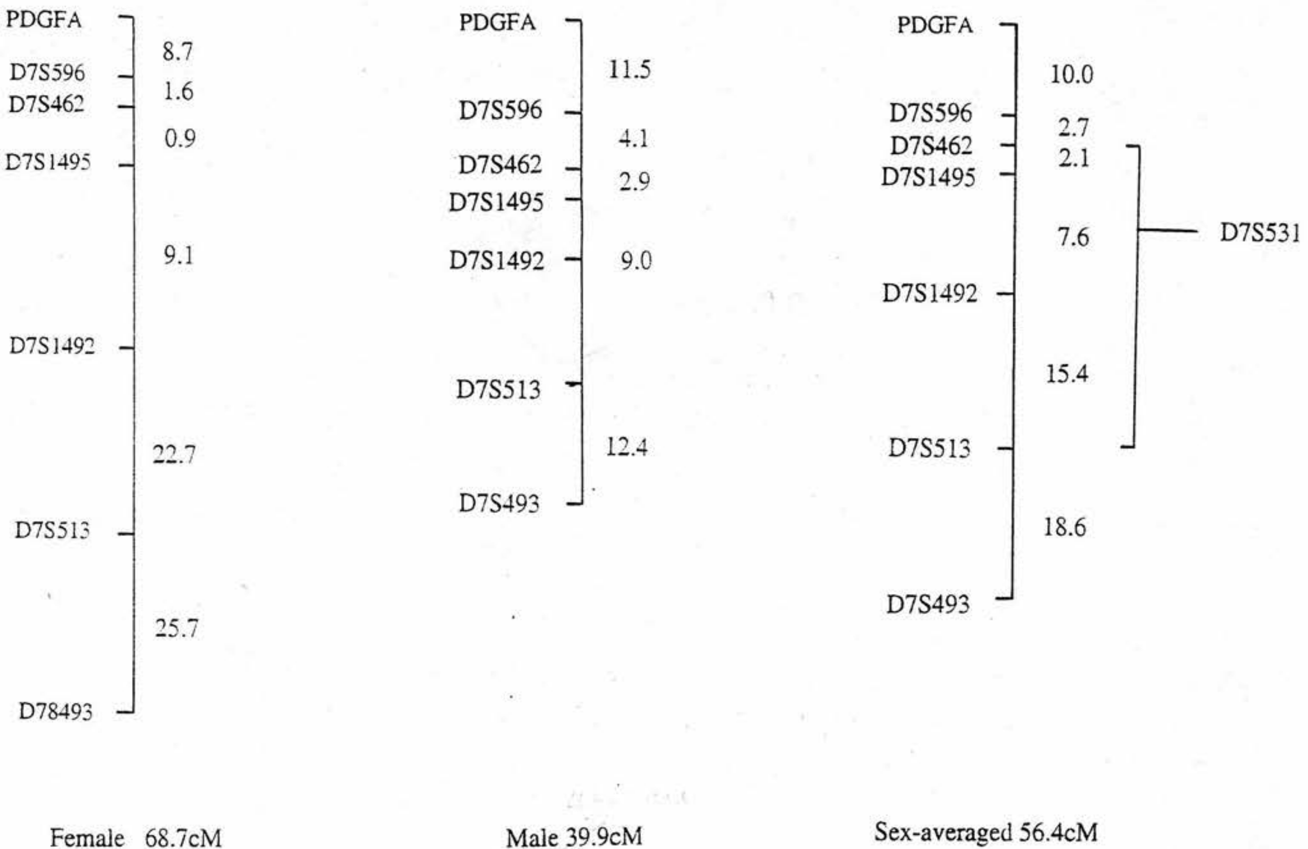


Diagram 4.8: The Eurogem Cri-Map linkage analysis of chromosome 7p.

Cri-Map linkage analysis of chromosome 7p using the Eurogem panel of markers including PDGFA

From Badbanchi (1994), see appendix II

4.9 FISH analysis of a cytogenetic breakpoint in chromosome 7p22.

A somatic cell hybrid panel had been used previously to map *PDGFA* to chromosome 7p21-pter as described in section 1.3 (Bonthron 1992). The most distal known cytogenetic landmark on chromosome 7p is the breakpoint of the X;7 translocation: 46,XX,t(X;7)(q21;p22) in the GM 1696 cell line (Vortkamp 1991). Following the results of the linkage analysis above, this cell line was analysed by fluorescent *in situ* hybridisation with PDGFA and MS31 (*D7S21*). This was done by N. van Roy and Dr F Speleman (Department of Medical Genetics, University Hospital, Gent, Belgium). Two PDGFA clones, pAC23/4 and pAC41/12 which span exon and intron 4 and from intron 4 to intron 6 respectively, were used. The clones are more fully described in Bonthron (1992). For *D7S21* the plasmid pMS31 which contains a Sau3AI genomic fragment cloned into the BamHI site of pUC13 was used. The pMS31 plasmid was kindly provided by Professor A. Jeffreys (Department of Genetics, University of Leicester).

These experiments showed that both MS31 (*D7S21*) and *PDGFA* are located on the extreme distal fragment of chromosome 7 which is translocated to the derivative X chromosome (not shown). Unfortunately the signals obtained with MS31 were somewhat weak, with background signals generally observed elsewhere (>20 metaphases examined). However, the MS31 signal was consistently present in most metaphases on the tip of the normal 7p as well as on the derivative X chromosome. Therefore, both *PDGFA* and *D7S21* lie distal to the GM 1696 7p22 translocation breakpoint and the results of this FISH experiment do not provide any further information about the relative positions of *PDGFA* and *D7S21*.

4.10 Conclusion and discussion

The linkage analyses conducted for this project and for the Eurogem collaboration assign *PDGFA* to a location beyond the previous telomeric limit of the 7p linkage map. This distal location for both *PDGFA* and *D7S21* is confirmed by the FISH analysis of the GM 1696 cell line. There is no known marker flanking *PDGFA* and no telomere STS or probe for the chromosome 7p telomere is available at present.

D7S531 was included in the analysis because some linkage maps have placed it distal to *D7S21*, as the most telomeric marker on chromosome 7p. The ambiguities which arise in the mapping of *D7S531* in these linkage maps may be due to the very low number of families in the CEPH panel which have been typed. The *D7S531* locus maps to the proximal, or centromeric side of the nearest marker to *PDGFA* in both of the analyses presented in this report. Thus, *D7S531* lies proximal to *D7S21* in the first analysis and proximal to *D7S596* in the Eurogem analysis. The relative order of *D7S21* and *D7S596* is not known.

The linkage maps, especially the Eurogem map, show an increase in the distance between *PDGFA* and its nearest proximal marker in comparison to the distances between other markers on the map. This may reflect a larger physical distance between the two loci but is seems more likely that this increase reflects the increase in recombination which is seen in telomeric regions.

The increase in recombination frequencies towards the chromosome telomeres is generally greater in males than females. Thus the male genetic map is frequently longer towards the telomere than the female map. These differences are in contrast to most areas of the genome where female recombination is usually greater than male. The distance between, for example, *D7S21* and *PDGFA* also shows a large difference between the male and female maps: 19.5 cM to 13 cM respectively. A similar pattern is seen between *PDGFA* and *D7S596* in the Eurogem analysis. The

observation that the male map is longer than the female map in this region also suggests that *PDGFA* is in a subtelomeric location on chromosome 7p.

There is no marker available for chromosome 7p which is known to be telomeric to *PDGFA*, so the distance between *PDGFA* and the telomere cannot be measured on the basis of this linkage analysis. The physical distance between *PDGFA* and *D7S21* is also difficult to estimate from this linkage data. The rough conversion of 1 cM = 1 Mb (section 1.10.3) would suggest that *PDGFA* lies 12.5 - 16 Mb away from *D7S21*. This seems unlikely given the available information about this region of chromosome 7p. The cell line GM 1696 which was used for the FISH analysis in section 4.9 was used in a physical mapping study of chromosome 7 (Vortkamp 1991). This study found that only 2 probes out of 148 mapped to the region distal to the breakpoint in this cell line while the FISH experiments reported above showed that both *PDGFA* and *D7S21* map to this small distal region. Assuming a random distribution of the probes, and estimating the size of chromosome 7p to be 60 Mb, the size of this region of the chromosome could be less than 1 Mb. However, this estimate assumes that these probes were distributed randomly which may not be the case. An alternative source of information for estimating the size of this region of chromosome 7p comes from the CEPH/NIH collaborative study linkage map (Mishra 1992). This map was estimated to cover >97% of the physical map of chromosome 7 between the 7q telomere and *D7S21*, suggesting that the distance from *PDGFA* to the telomere may well be less than 2 Mb.

As described in section 1.10, some forms of craniosynostosis have been mapped to chromosome 7p. At the start of this study the more distally identified region for craniosynostosis in the panel of patients with chromosome 7 deletions was 7p21-p22. This overlapped with the assigned location for *PDGFA*. Molecular studies had been conducted using DNA from a subset of patients with non-familial deletions. However these did not include *PDGFA*. These studies found that all of the patients studied

were heterozygous for MS31 (*D7S21*). As this linkage analysis shows, *PDGFA* maps distal to *D7S21* and thus is excluded as a candidate gene for craniosynostosis in these patients.

The linkage analysis of Saethre-Chotzen craniosynostosis syndrome (see diagram 1.4) included the locus *D7S513*. The regions identified as candidates for Saethre-Chotzen syndromes all lay proximal to this locus. The Eurogem linkage analysis presented in this report (diagram 4.8) includes *D7S513* which is mapped some distance proximal to *PDGFA*. Thus, *PDGFA* clearly lies at some distance from even the most distal region linked to craniosynostosis and therefore is excluded as a candidate gene. Recently the homeobox gene *MOX2* has been mapped to 7p22.1-p21.3 making it a more likely candidate for the Saethre-Chotzen craniosynostosis syndrome (Grigoriou 1995).

To obtain further information about the location of *PDGFA*, physical mapping of the region by pulsed field restriction mapping (described in Chapter 5) and screening of libraries for YAC clones (Chapter 6) was undertaken. A major aim of the physical mapping project was to physically link *PDGFA* to other loci such as *D7S21* and to landmarks such the telomere through the development of a clone contig. In addition, markers distal to *PDGFA* were required. Ideally such markers would include the chromosome 7p telomere and thus would allow the distance between the telomere and *PDGFA* to be measured. For this, telomeric YAC clones which were candidates for chromosome 7p were analysed in Chapter 8.

5. Chapter 5: Pulsed Field Restriction Mapping

5.1 PFGE restriction map of the PDGFA locus

To develop a physical map of the region two approaches were used: a long-range restriction map was constructed using pulsed-field gel electrophoresis (described below) and YAC and P1 clone libraries were screened for large clones spanning the region (Chapter 6). The methods and principles behind pulsed-field restriction mapping are described in the introduction (section 1.13.1). Since the sequence of DNA in clones may be rearranged during the cloning process it is important to compare contigs to the genomic DNA sequence from which they derive. This is usually done by comparing long range restriction maps derived from both cloned DNA and genomic DNA. With this in mind a pulsed-field restriction map of PDGFA was undertaken at the same time as the screening of the YAC clone libraries

The linkage analysis described in Chapter 4 mapped *PDGFA* as the most telomeric marker on chromosome 7p. The most closely linked proximal locus was *D7S21* (MS31). The linkage maps show a large sex averaged genetic distance of 16 cM between these two loci. However, as discussed in section 4.10, the increase in recombination in subtelomeric regions means that the physical distance between these two loci may be considerably less than that implied by the genetic distance. It is impossible to accurately estimate the physical distance between *PDGFA* and *D7S21* and the two loci may even be very close together. The possibility that the two loci are very close was investigated by pulsed-field restriction analysis since if two markers are close they may be found to map to the same restriction fragment.

5.2 Probes used in the PFGE restriction mapping

Prior to the start of this project PDGFA had been cloned in bacteriophage clones which spanned about 45 kb (Bonthron 1988). These clones were used to identify suitable single copy probes from PDGFA which could be used for the restriction

mapping. Four potentially suitable probes from PDGFA were identified (see diagram 5.1), two of these: ACY7 and ACIXSR were selected, they are described below. The third probe which was used for the pulsed-field mapping was MS31 (*D7S21*), a highly polymorphic minisatellite probe which was found to be the closest linked locus to *PDGFA* in the linkage analysis.

5.2.1 Probe ACY7

Plasmid clone ACY7 spans 8 kb from the XbaI site in intron 1 to the SalI site in exon 4. It was subcloned from λAC23 (see diagram 3.1) and is more fully described in Chapter 3, where its full sequence is presented. The insert was isolated from its plasmid vector by an XbaI/SalI digest and the probe was hybridised to genomic DNA digests on a Southern blot which confirmed that the insert is a suitable, single copy probe. Since ACY7 spans 8 kb of the gene including three exons, it was the first probe to be used in the PFGE analyses.

5.2.2 A downstream probe: AC41HR

As shown in diagram 5.1, AC41HR spans 1.7 kb in exon 7 of PDGFA. It was isolated using a HincII/EcoRI digest of λAC41. Like ACY7 it was hybridised to genomic DNA restriction digests in Southern blots to confirm that it did not contain repetitive sequences. However AC41HR is only about 10 kb downstream from ACY7. The initial mapping data (described below) showed that AC41HR lies well within the boundaries of the restriction map produced with ACY7. Thus, AC41HR would not add further information to the restriction map and was not used in the PFGE restriction mapping.

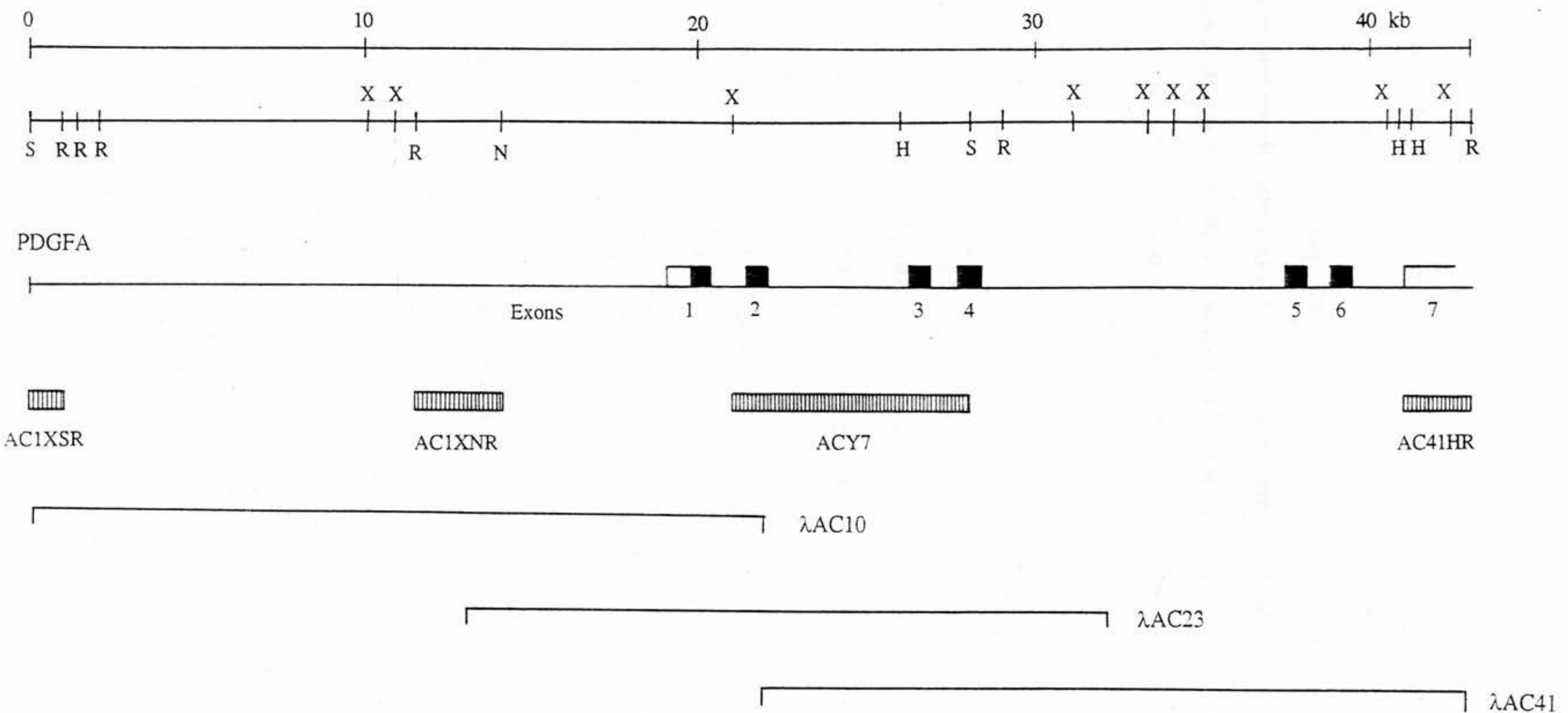


Diagram 5.1: Single copy probes from PDGFA.

The bacteriophage clones λAC10, λAC23 and λAC41 which span PDGFA and a restriction map for the enzymes EcoRI (R), SalI (S), NotI (N), XbaI (X) and HindIII (H) are shown (from Bonthron 1988). The single copy probes, ACY7, AC41HR, ACIXNR and ACIXSR were developed. Of these, the probes ACY7 and ACIXSR were used for the pulsed-field restriction mapping.

5.2.3 An upstream probe: ACIXSR

A probe which lay upstream from PDGFA was sought following the results from the restriction mapping using ACY7. A number of different fragments from a subclone of λAC10 were tested by hybridisation to gDNA Southern blots before a single copy probe was found. Repetitive DNA was found using probes from the region between the first EcoRI site and the second XbaI site which are about 0.8 kb and 10.7 kb respectively from the 3' end of λAC10 (see diagram 5.1). This eliminated over 10 kb of the upstream region of PDGFA. However two potentially useful probes were identified, they were ACIXNR and ACIXSR both of which are shown in diagram 5.1.

The probe ACIXNR spans from an EcoRI site to the NotI site upstream of PDGFA. The position of this NotI site was found to be critical in the development of the pulsed-field restriction map of the region (discussed below). Thus a probe further upstream from this site was preferred in an attempt to extend the restriction map as far as possible given the available clones. The second probe, ACIXSR, is a 1.6 kb SalI/EcoRI fragment from about 20 kb upstream from PDGFA. It is also about 12 kb upstream from ACIXNR and therefore was selected to be used in the PFGE mapping.

All three of the probes from PDGFA: ACY7, AC41HR and ACIXSR were used to screen the YAC and P1 libraries (described in Chapters 6 and 7).

5.2.4 The minisatellite MS31 (*D7S21*)

The third probe which was used for the PFGE restriction mapping was MS31. It identifies the highly polymorphic minisatellite at *D7S21* which has previously been mapped to chromosome 7p22-pter (Royle 1988). The probe is a 3 kb

Sau3AI genomic fragment cloned into the BamHI site of pUC13. The pMS31 plasmid was kindly provided by Professor A. Jeffreys (Department of Genetics, University of Leicester).

MS31 (*D7S21*) was included as one of the markers in the linkage analysis which placed *PDGFA* as the most distal marker on chromosome 7p. MS31 was found to be the nearest proximal marker, separated from PDGFA by 16 cM. As discussed in Chapter 4 and above (section 5.1) it is difficult to estimate the physical distance between *PDGFA* and MS31. However, given the increase in recombination which is seen in subtelomeric regions, it seems unlikely that the two loci are as far apart as the large genetic distance suggests. Thus, to investigate the possibility that two loci may be close together, the probe MS31 was used to probe the pulsed-field Southern blots.

It is possible to measure the maximum distance between two markers if they are shown to hybridise to the same band from gDNA digests on Southern blots. For this, several different enzyme digests should be shown to produce a common band with both probes to eliminate the possibility that the two probes identify two different DNA fragments which are comigrating. The apparent size of fragments in PFGE may be inaccurate, for example overloading gel lanes can cause DNA fragments to migrate at apparently larger sizes. Thus, the size of bands from different probes should only be compared when each probe has been hybridised to the same Southern blot.

5.3 The pulsed-field electrophoresis gels

Throughout the restriction mapping study, DNA was prepared from a single normal male to avoid problems in interpreting the data arising from polymorphisms. DNA

from lymphocytes was prepared in agarose blocks from fresh blood samples. Standard methods were used for the preparation of agarose blocks containing high molecular weight DNA, preparation of yeast chromosome DNA as size markers, restriction digests, alkaline blotting and the subsequent hybridisation of Hybond N^+ filters (described in Chapter 2). Fresh blood samples were used because cultured cells have been shown to methylate genes whose functions are superfluous for growth in a tissue culture flask. As a consequence, up to half all CpG sites will be methylated in cells grown *in vitro* (Antequera 1990).

Smaller fragments reorientate faster when the direction of the electrical field is changed, as a consequence they spend a larger proportion of each pulse interval migrating through the gel. The size of the fragments resolved is directly related to the length of the pulse times. The pulse times were altered to allow different DNA fragment sizes to be resolved. The pulse times which were used to run each gel are included in the title of each diagram. Otherwise, all of the pulsed field gels were run at 12 to 15°C using a 1% agarose gel in 0.5x TBE, at 200 V for 24 hours. To ensure that the sizes of bands could be directly compared, the same pulsed-field gel was probed with all three probes ACY7, MS31 and ACIXSR. 'Autorads' of the Southern blots, hybridised with each of these probes are shown in diagrams 5.5 to 5.12.

5.4 Restriction enzymes used in PFGE

The restriction endonuclease enzymes used in this pulsed field work are listed in table 5.2. These restriction enzymes, which generate large DNA fragments, often recognise eight base cleavage sites and/or contain at least one CpG in their recognition sites. Most of these rare cutting restriction enzymes cannot cleave methylated CpG sites. As the table shows, the enzymes were selected to cover a range of average sized fragments.

Restriction Enzyme	Recognition site	Appx. size (kb) of average fragment
NotI	GCGGCCGC	1 000
NruI	TCGCGA	500
MluI	ACGCGT	500
SfiI	GGCC(N_5)GGCC	200
BspDI (ClaI)	ATCGAT	100
BstZI (EagI)	CGGCCG	100
SalI	GTCGAC	100
SacII	CCGCGG	100
XhoI	CTCGAG	100

Table 5.2: Restriction enzymes used for PFGE restriction mapping
These restriction enzymes, or their isoschizomers (shown in brackets) were used for the pulsed-field restriction mapping. Their recognition sites and the approximate average sizes of gDNA following digestion are listed (From Birren and Lai 1993)

A negative control, without restriction enzyme, was run with each different buffer every time that buffer was used to ensure that no nuclease activity was present. This is essential since even low levels of nuclease activity may degrade DNA sufficiently to interfere with the PFGE mapping. It is often impossible to distinguish low levels DNA degradation from restriction digests. Where appropriate double digests, using two different enzymes simultaneously, were used to position restriction sites more accurately.

Prior to commencing the restriction digests for the pulsed-field gels, a control titration using increasing amounts of the restriction enzyme EcoRI was used conducted. This was used to measure the amount of enzyme which would be required to fully digest the DNA samples. It is shown in diagram 5.3, full digestion is seen using between 80 and 160 units of EcoRI. For subsequent digestions, 100 to 140 units of enzyme were used.

5.5 The physical distance between MS31 and *PDGFA*

The patterns of hybridisation using the PDGFA probes, ACY7 and ACIXSR, were compared to the band sizes from MS31. The results are shown in diagrams 5.5 to 5.10, they are summarised in table 5.4 which is a guide to the following photographs. The nine restriction enzymes which are listed in table 5.2 were all found to generate different sized fragments when hybridised with MS31 on comparison to hybrisisation with ACY7 or ACIXSR.

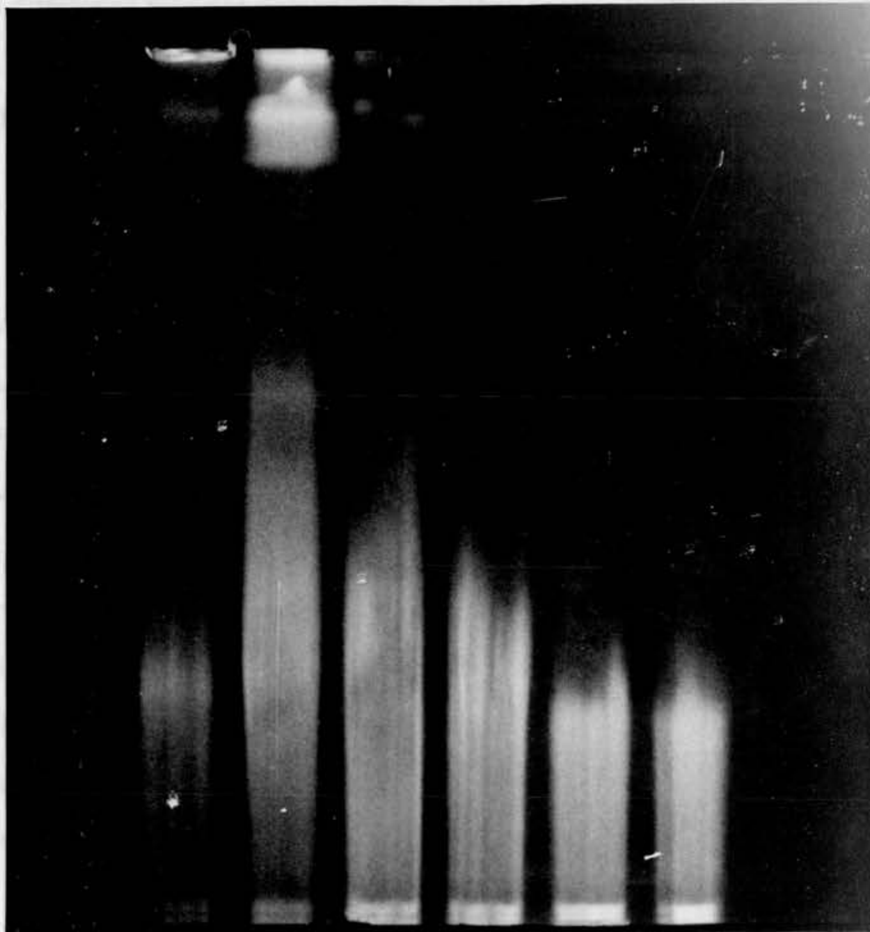


Diagram 5.3: Partial EcoRI digests.

EcoRI digests using 0 to 320 units were used to titrate the amount of restriction enzyme needed to fully digest the genomic DNA. Full digestion is seen between 80 and 160 units of enzyme.

The restriction digests using the enzymes NruI, SalI and BspDI all showed clearly different band sizes when comparing ACY7 and MS31 (see table 5.4 for diagram numbers). By contrast, the band sizes hybridised by these two probes with other enzymes initially looked similar in size. Size differences between the fragments hybridised by ACY7 and MS31 could be seen if the two autorads from each digest were superimposed. Doing this confirmed that the fragments hybridised in the SfiI digests (diagram 5.5) and the SacII digests (diagram 5.5) were different sizes, however this difference was not convincing for other enzymes. For example, the XhoI digests in diagram 5.5. The XhoI digests were repeated and are shown in diagram 5.10, where a clear size difference between the hybridised fragments can be seen.

The pulse times in the gel in diagram 5.10 were changed to optimise the resolution of smaller DNA fragments. In the gel in diagram 5.5 the pulse times separate fragments sized from 48 kb to 533 kb, the gel in diagram 5.10 does not resolve fragments larger than 242 kb. Pulsed -field gels have been shown to provide optimal separation of DNA fragments in the central area of the gel with areas of compression both at the top in the region of limiting mobility and at the bottom of the gel. It seems likely that the apparent similarity in size between the fragments hybridised by ACY7 and MS31 in the XhoI digests in diagram 5.5 was due to this compression effect. Changing the pulse times of the gel run has confirmed that these fragments are not the same size.

Comparing the hybridisation patterns of ACY7 and MS31 shows a similar effect for NotI (diagrams 5.6 and 5.9), BstZI (diagrams 5.7 and 5.9) and MluI (diagrams 5.6 and 5.10).

Enzyme	Diagram	Comparing ACY7 and MS31
NotI	5.6	Not well resolved
	5.9	Clear size difference
NruI	5.6	Clear size difference
BspDI	5.7	Clear size difference
BstZI	5.7	Not well resolved
	5.9	Clear size difference
SalI	5.9	Clear size difference
XhoI	5.5	Not well resolved
	5.10	Size difference when autorads superimposed
MluI	5.6	Not well resolved.
	5.10	Size difference when autorads superimposed
SfiI	5.5	Size difference when autorads superimposed
SacII	5.5	Size difference when autorads superimposed

Enzyme	Diagram	Comparing ACIXSR and MS31
NotI	5.9	Clear size difference
NruI	5.9	MS31 band just below LM, not well resolved
BspDI	5.7	Clear size difference
BstZI	5.7	Not well resolved
	5.9	Size difference when autorads superimposed.
SalI	5.9	Clear size difference
XhoI	5.5	Not well resolved
	5.8, 5.10	Clear size difference
MluI	5.8, 5.10	Clear size difference
SfiI	5.8	Clear size difference
SacII	5.5	Size difference when autorads superimposed
	5.8	Clear size difference

Table 5.4: Measuring the physical distance between *PDGFA* and *D7S21* (MS31). No common sized fragments were found on comparing the hybridisation patterns between MS31 and ACY7 or ACIXSR, thus the physical distance between the two loci cannot be estimated.

If the PDGFA gene is sited within a CpG island then the average fragment sizes from using ACY7 would be very small. The cluster of restriction sites at the 5' end of the gene would make it impossible to identify physical linkage to any upstream site. At this time during the project, the orientation of PDGFA on chromosome 7p and thus the side on which MS31 lay was not known. For this reason, the hybridisation patterns seen using probes ACIXSR and MS31 were compared. The probe ACIXSR lies about 20 kb upstream from PDGFA, and would be far more likely to identify common restriction fragments if MS31 lay on this side of PDGFA. Obviously if MS31 lay downstream of PDGFA a downstream probe would potentially have been more useful. The probe AC41HR was available and mapped to downstream of PDGFA. However the development of a long range restriction map around PDGFA, which is discussed below, found that AC41HR lies within the majority of restriction fragments identified by ACY7 and thus using AC41HR would not supply any further information. The orientation of PDGFA on chromosome was subsequently determined by Rec-A assisted restriction endonuclease (RARE) cleavage experiments (discussed in Chapter 10).

Comparison of the hybridisation patterns produced by the probes ACIXSR and MS31 did not identify any common sized fragments. As for ACY7, some of the digests (BstZI, XhoI and SacII) were repeated with changed pulse times to confirm that the hybridised bands were of different sizes. In the BstZI digest shown in diagram 5.9 ACIXSR hybridises very faintly to a small, 20 kb band. When the two autorads are superimposed to compare the hybridisation patterns of MS31 and ACIXSR the bands identified by the two probes are found to be different sizes. The diagrams which show each of these experiments are summarised in table 5.4.

Since no common restriction fragments were hybridised, the distance between PDGFA and MS31 (*D7S21*) cannot be estimated from these pulsed-field restriction mapping experiments.

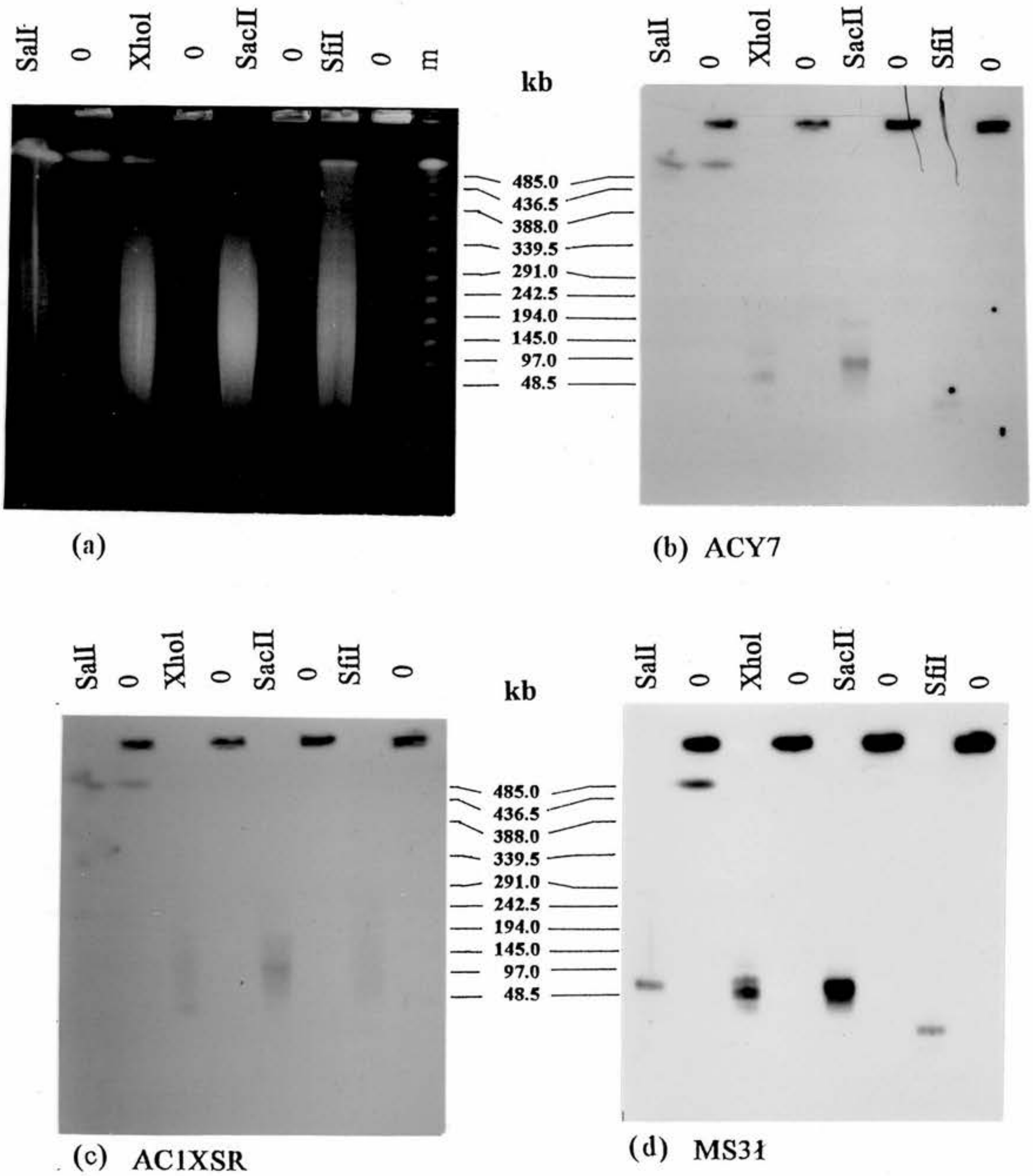
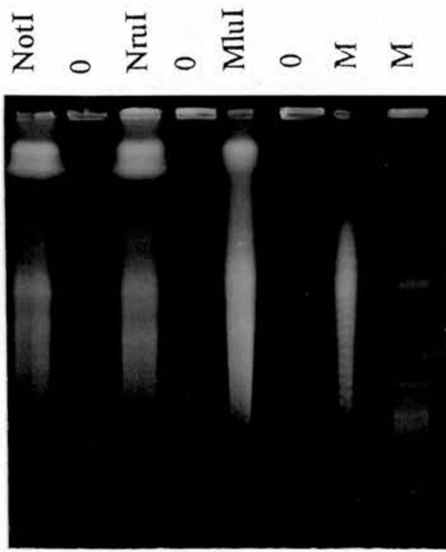
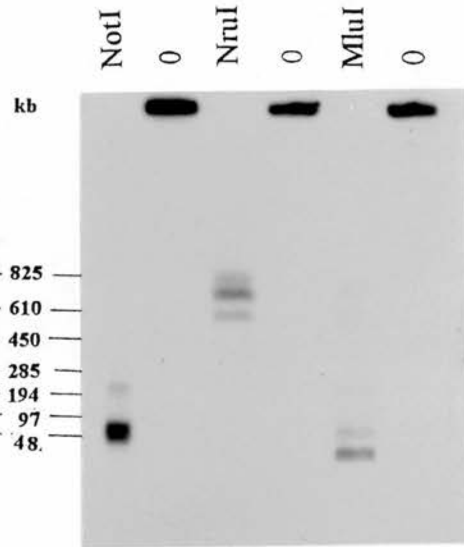


Diagram 5.5: Pulsed field restriction analysis: experiment 1

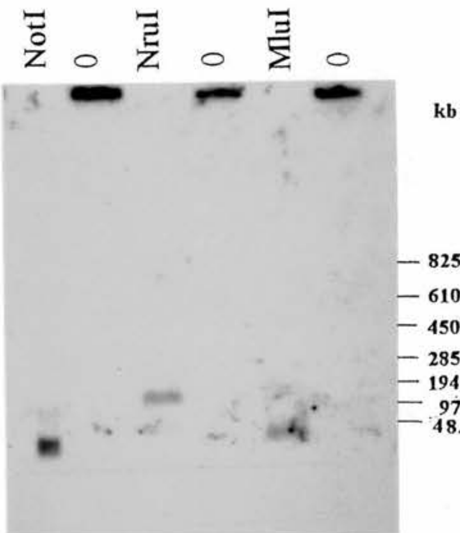
(a) Pulsed field analysis using the restriction enzymes SalI, XhoI, SacII and SfiI, probed with (b) ACY7, (c) ACIXSR and (d) MS31. The gDNA samples were separated using 30 second pulses.



(a)



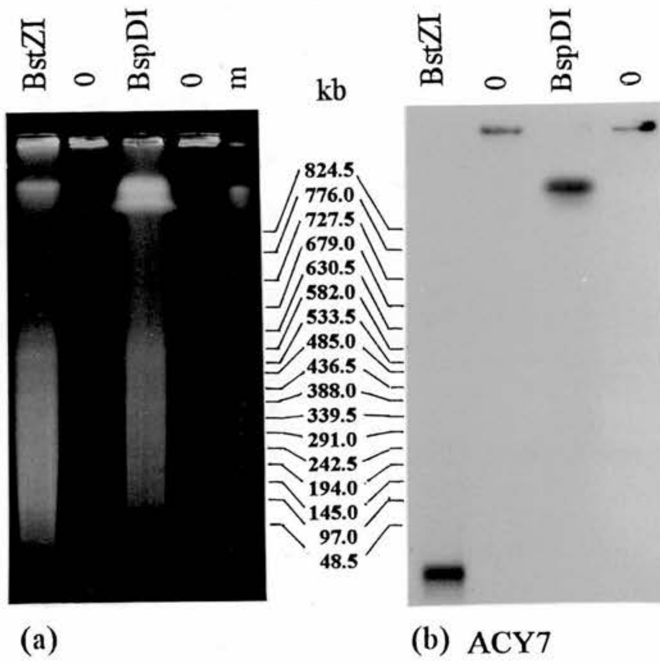
(b) ACY7



(c) MS31

Diagram 5.6: Pulsed field restriction analysis: experiment 2

(a) Pulsed field analysis using the restriction enzymes MluI, NruI and NotI, probed with (b) ACY7, (c) MS31. The gDNA samples were separated using 90 second pulses.



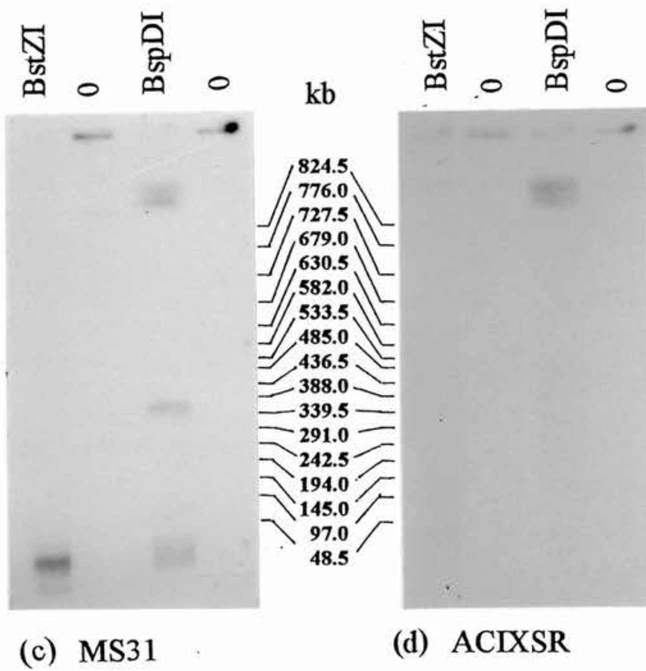
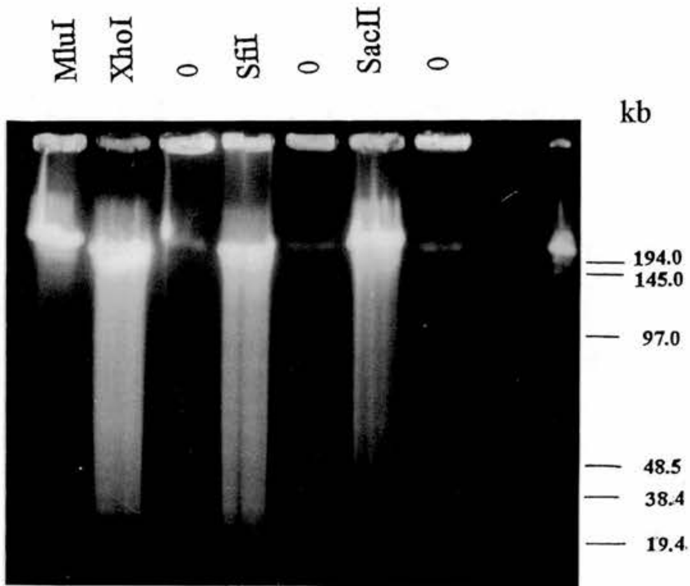
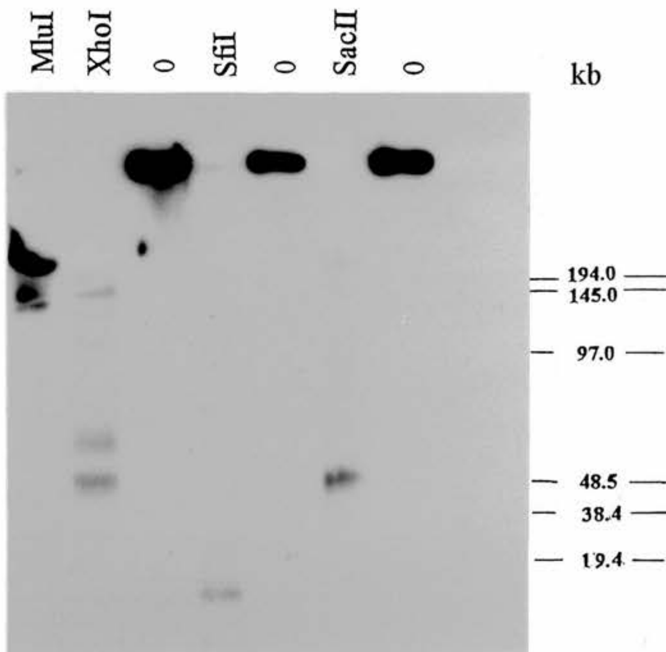


Diagram 5.7: Pulsed field restriction analysis: experiment 3

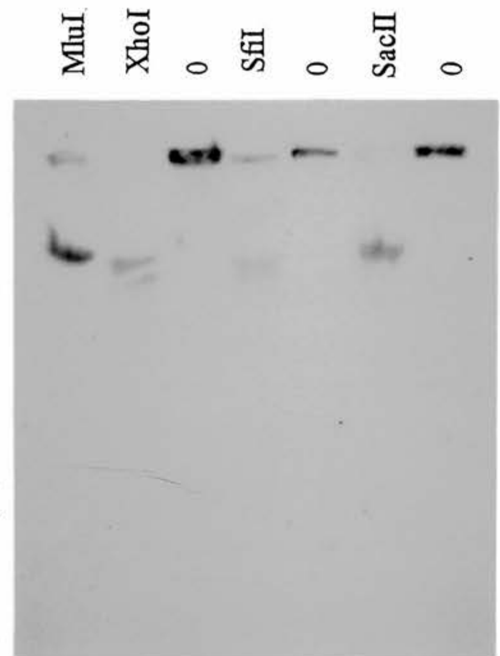
(a) Pulsed field analysis using the restriction enzymes BstZI (EagI) and BspDI (ClaI), probed with (b) ACY7, (c) MS31 and (d) ACIXSR. The gDNA samples were separated using 60 second pulses



(a)



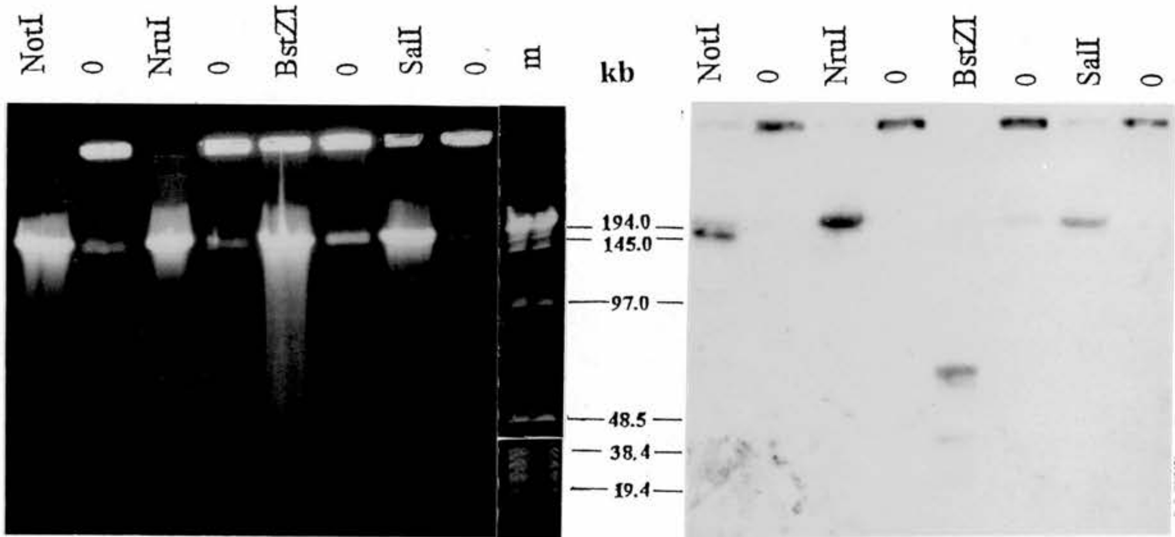
(b) MS31



(c) ACIXSR

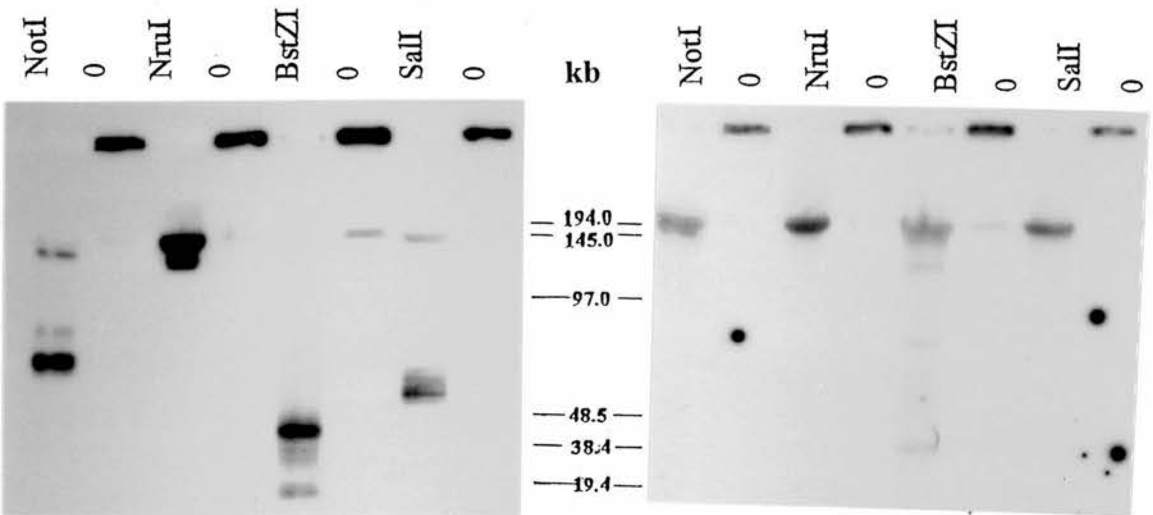
Diagram 5. 8: Pulsed field restriction analysis: experiment 4

(a) Pulsed field analysis using 5 second time pulses with the restriction enzymes MluI, XhoI, SfiI and SacII, probed with (b) MS31 and (c) ACIXSR.



(a)

(b) ACY7

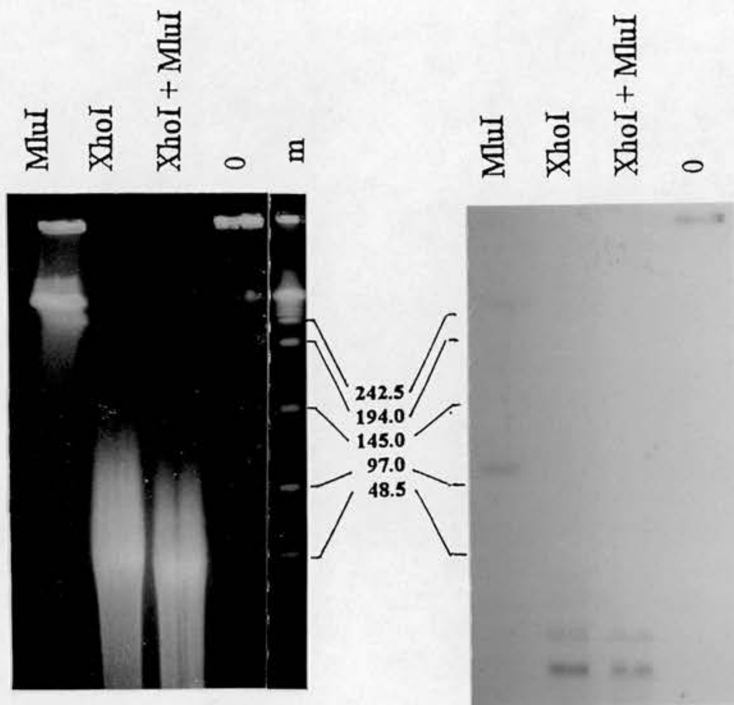


(c) MS31

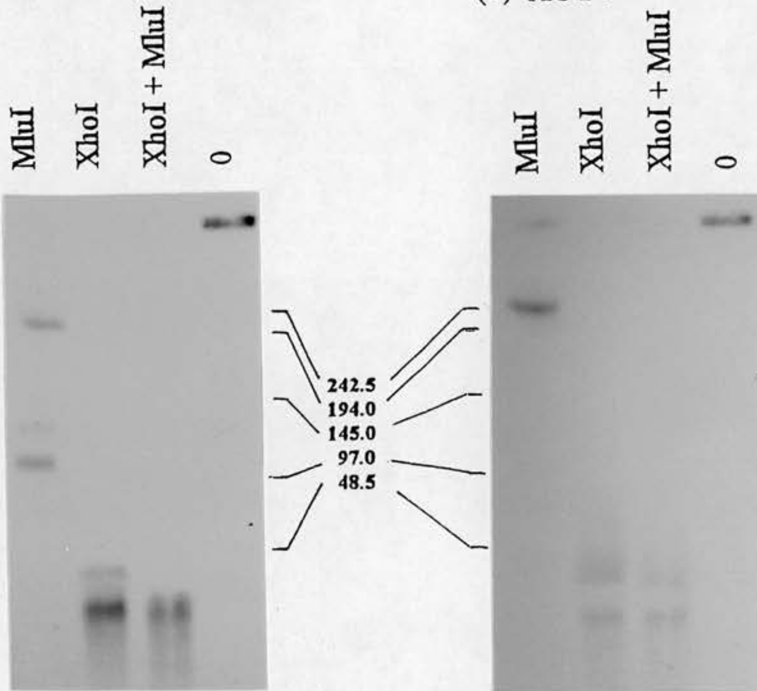
(d) ACIXSR

Diagram 5.9: Pulsed field restriction analysis: experiment 5

(a) Pulsed field analysis using 5 second time pulses with the restriction enzymes NotI, NruI, BstZI and SalI, probed with (b) ACY7, (c) MS31 and (d) ACIXSR.



(a) (b) ACY7



(c) MS31 (d) ACIXSR

Diagram 5.10: Pulsed field restriction analysis: experiment 6

(a) Pulsed field analysis using the restriction enzymes MluI, and XhoI, probed with (b) ACY7, (c) MS31 and (d) ACIXSR. The gDNA samples were separated using 10 second pulses.

5.6 A long-range restriction map for PDGFA

Single and double restriction digests were used to construct a long range restriction map around the PDGFA locus. The results from hybridising the Southern blots of the pulsed-field gels with ACY7 and ACIXSR are shown in diagrams 5.5 to 5.12, in diagrams 5.13 and 5.14 Southern blots from conventional electrophoresis agarose gels are shown, the restriction map derived from these results is shown in diagram 5.16.

5.6.1 NotI restriction sites

The NotI site which lies about 7 kb upstream from the 5' end of ACY7 proved to be critical in the development of this restriction map. This NotI site had been mapped prior to that start of this project (Bonthron 1988) and it is shown in the restriction map in diagram 5.1. A NotI digest probed with ACY7 is shown in diagram 5.11, the probe hybridises to a 145 kb sized fragment. A single band is hybridised indicating that this site cuts fully, this was found to be extremely useful in the analysis of the other restriction enzymes and this NotI site is assigned position 0 on the restriction map. Probing the NotI restriction digests with ACIXSR gave a 150kb band, thus three NotI restriction sites can be placed on the restriction map in diagram 5.16, they are in positions -150, 0 and +145.

5.6.2 NruI restriction sites

Probing an NruI digest with ACY7 identified a DNA fragment of around 600 kb (diagram 5.6). Probing with ACIXSR in diagram 5.9 found that the probe hybridises to the large limiting mobility band which marks the limits of resolution under the pulse times used in this gel. Therefore the NruI fragment must be larger than 194 kb

and it is possible that both probes hybridise to the same fragment. A double digest using NruI and NotI is shown in diagram 5.11. For both probes ACY7 and ACIXSR, the bands hybridised in the double digest are the same size as those in the NotI only digest. Thus the restriction sites for NruI lie outside the NotI sites, and cannot be placed on the restriction map in diagram 5.16.

5.6.3 MluI restriction sites

A single MluI digest produced a 100 kb fragment when probed with ACY7 (diagram 5.10). A double digest of MluI with NotI is shown in diagram 5.12, hybridisation with ACY7 produced an interesting pattern. The MluI band remains the same size in the double digest indicating that two MluI sites lie within the NotI sites. However closer inspection of the MluI single digest in diagram 5.12 shows that there is a second, larger band of about 200 kb. Hybridisation to this band is much fainter suggesting that it arises from a partial digest which may arise as a result of incomplete methylation of a restriction site. This band disappears in the MluI/NotI double digest. These results suggest that three MluI sites can be mapped from hybridisation with ACY7, the first lies close to the NotI at position 0, with the other two sites being at + 100 and + 170 in the restriction map in diagram 5.16.

The alternative explanation for the two fragments seen in the MluI single digest is there is an MluI restriction site within ACY7. If this were the case, the double digest with NotI would not produce a band corresponding to the NotI single digest. It would produce two fragments: one the same size as one of the MluI fragments, the second corresponding to the distance between the NotI site and the MluI site in ACY7.

Probing MluI and MluI/NotI digests with ACIXSR confirmed the first hypothesis proposed above. An MluI site within ACY7 would mean that one of the two bands hybridised by ACY7 should also be seen, at the same size, when probed with ACIXSR, no such band was seen. The single MluI digest hybridised with ACIXSR produced a fragment sized over 436 kb, within the region of limiting mobility in the gel in diagram 5.12. The MluI/NotI digest produced the same pattern as the NotI digest confirming that this MluI site is beyond the upstream NotI site and cannot be marked on the restriction map in diagram 5.16.

5.6.4 XhoI restriction sites

In contrast to the very large distances covered using enzymes such as NruI and MluI, XhoI produced small DNA fragments. Diagram 5.10 shows an XhoI digest probed with ACY7, the smallest fragment is considerably smaller than the smallest band in the λ ladder marker which is 48 kb in size. The sizes of the XhoI digests were investigated further in a standard agarose gel, using double digests of XhoI with HindIII and EcoRI (see diagram 5.13). There is a HindIII site within ACY7, about 12 kb downstream from the NotI site in position 0 on the restriction map in diagram 5.13 (shown in diagram 5.1). On hybridisation with ACY7, the HindIII/XhoI double digest produced bands of 4.5 and 7 kb, giving two possible locations for the XhoI site on each side of ACY7.

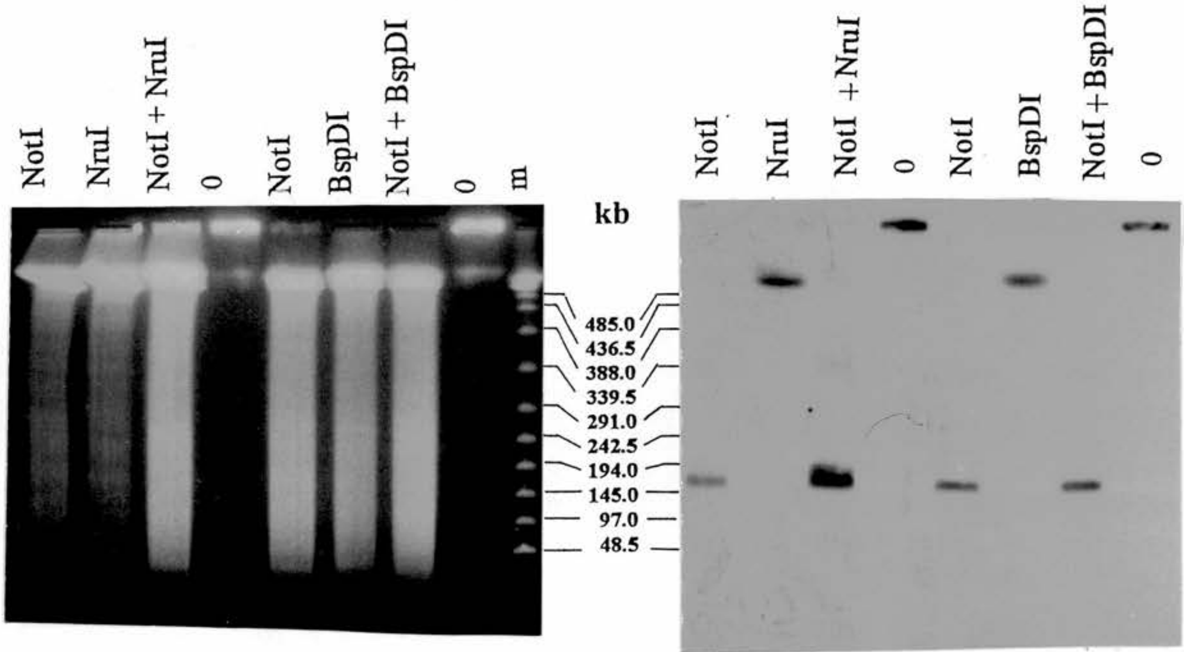
The sequence of the promoter region of PDGFA was known to contain an XhoI site (D.T Bonthron, pers comm.). This known site placed the upstream XhoI site in position +5 on the restriction map in diagram 5.16, thus the downstream site corresponds to position +16.5. These positions for the XhoI sites were confirmed by the XhoI/EcoRI digest shown in diagram 5.2 which produced just one band of 10.5 kb in size.

The upstream XhoI fragments, hybridised by ACIXSR are shown in diagram 5.10. Two bands sized about 20 and 50 kb can be seen. These have been marked on the restriction map in diagram 5.16 in positions -20 and -50.

5.6.5 SfiI restriction sites

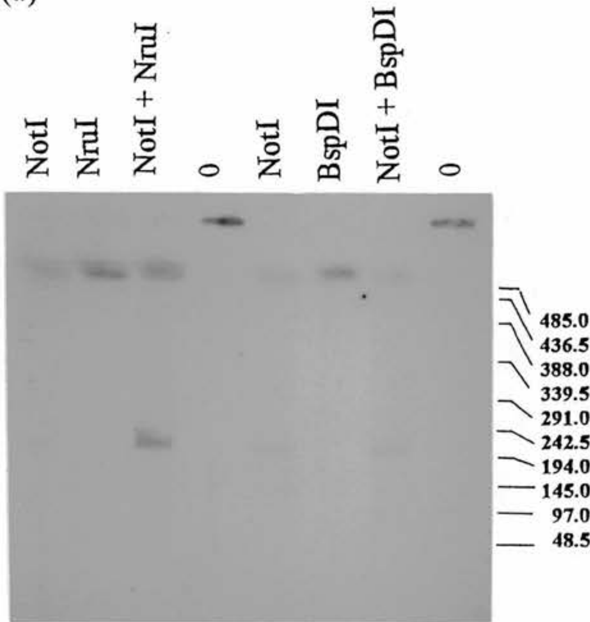
The restriction enzyme SfiI also produced very small fragments, in diagram 5.5 ACY7 hybridises to a band smaller than the 48 kb size marker. Thus the restriction sites for this enzyme were mapped using conventional electrophoresis. A single SfiI digest is shown in diagram 5.14 where two bands sized 10 kb and about 15 kb are hybridised. The two bands are of equal intensity, implying that there is an SfiI site which cuts within the ACY7 probe. The presence of an SfiI restriction site within ACY7 (in position 4427) was confirmed by inspection of the sequence.

The SfiI/NotI double digest in diagram 5.5 also produced two bands, sized about 8kb and 10.5 kb. Since the NotI site in position 0 cuts fully, one of these bands must correspond to the NotI-SfiI fragment while the other is an SfiI-SfiI fragment. This would mean that the 8 kb band is a NotI-SfiI fragment, placing SfiI sites in positions + 8 (within ACY7) and +18.5 on the restriction map. The third site which accounts for the two fragments seen in the single SfiI digest must lie upstream from the NotI site. An SfiI site at -7 would give rise to the two fragments of 15 kb and 10.5 kb in the SfiI single digest.



(a)

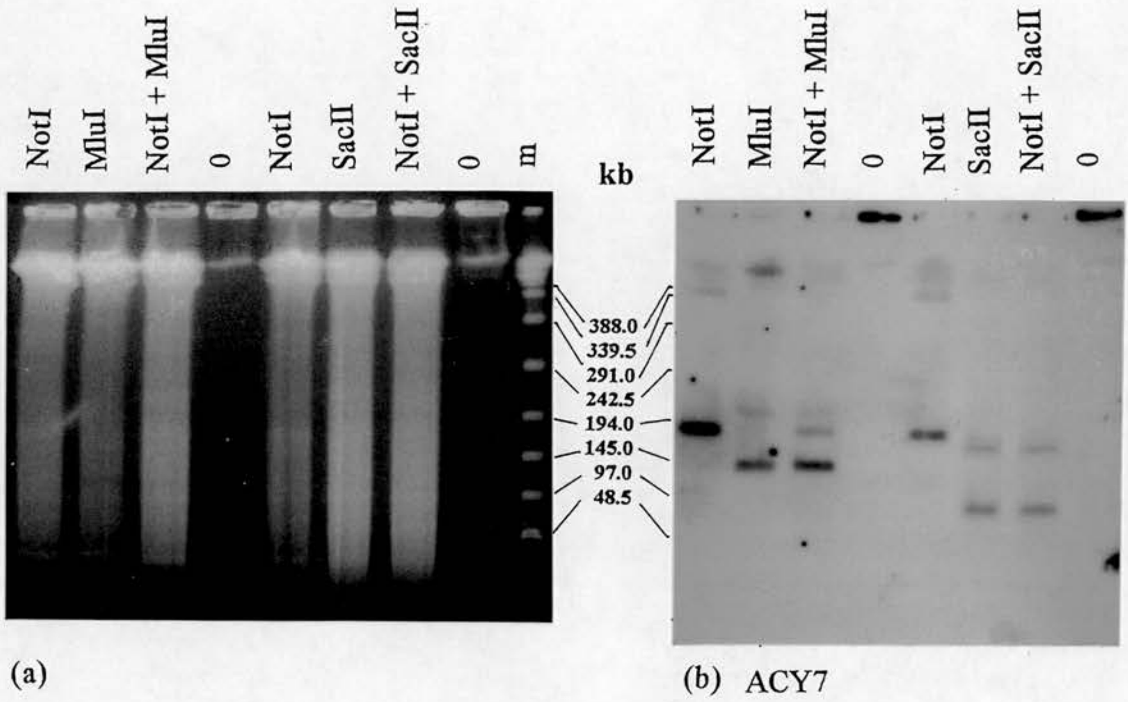
(b) ACY7



(c) AC1XSR

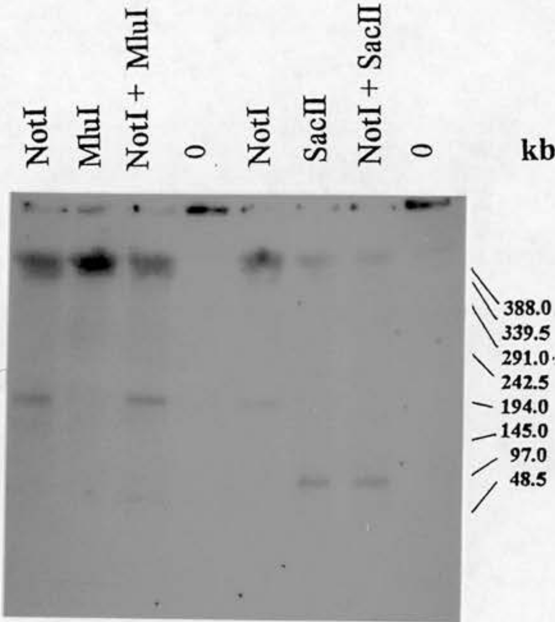
Diagram 5.11: Pulsed field restriction analysis: experiment 7

(a) Pulsed field analysis using the restriction enzymes NotI, NruI and BspDI, probed with (b) ACY7 and (c) ACIXSR. The gDNA samples were separated using 30 second pulses.



(a)

(b) ACY7



(c) ACIXSR

Diagram 5.12: Pulsed field restriction analysis: experiment 8

(a) Pulsed field analysis using the restriction enzymes NotI, MluI and SacII, probed with (b) ACY7 and (c) ACIXSR. The gDNA samples were separated using 20 second pulses.

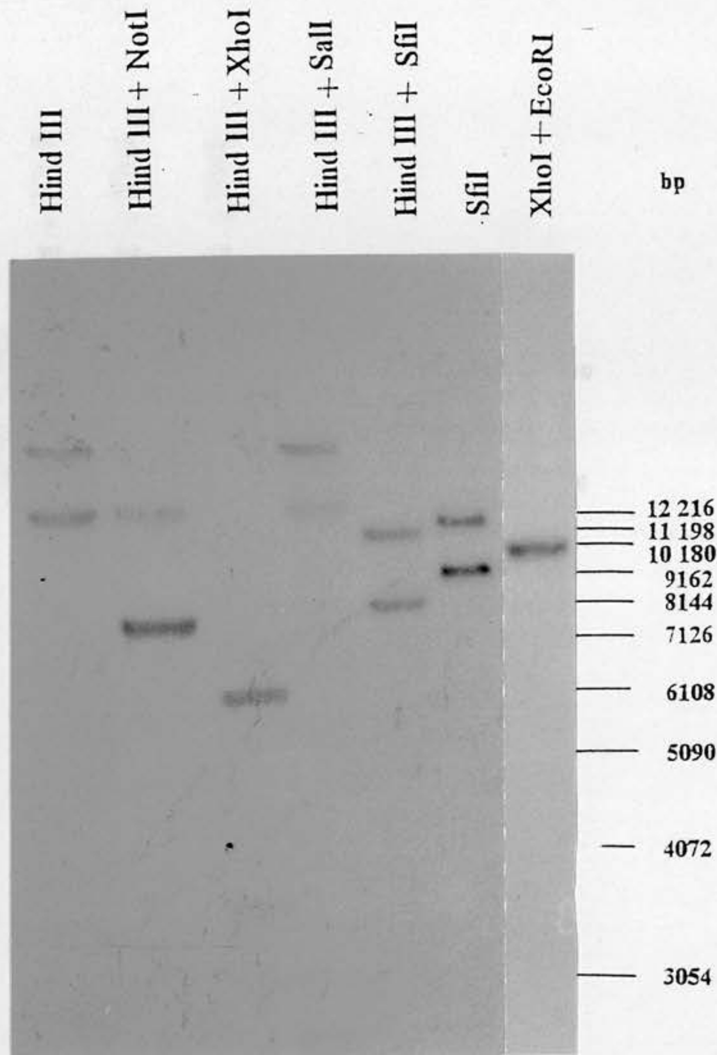


Diagram 5.13: Restriction sites for XhoI on hybridisation with ACY7.

Conventional gel electrophoresis analysis placed sites for XhoI 4.5 kb downstream and 7 kb upstream from the HindIII site in ACY7 (discussed in section 5.6.4). The 4.5 kb band in the HindIII/XhoI digest was very faint and is not shown in this diagram. The HindIII/SalI double digests produced the same pattern as HindIII alone, however the 15.6 kb band is fainter in the double digest suggesting that the SalI site which marks the 3' end of ACY7 does cut to a very small extent, however SalI was not included in the restriction map in diagram 5.16. A gDNA sample from the same individual used for the pulsed field gels was used for these digests.

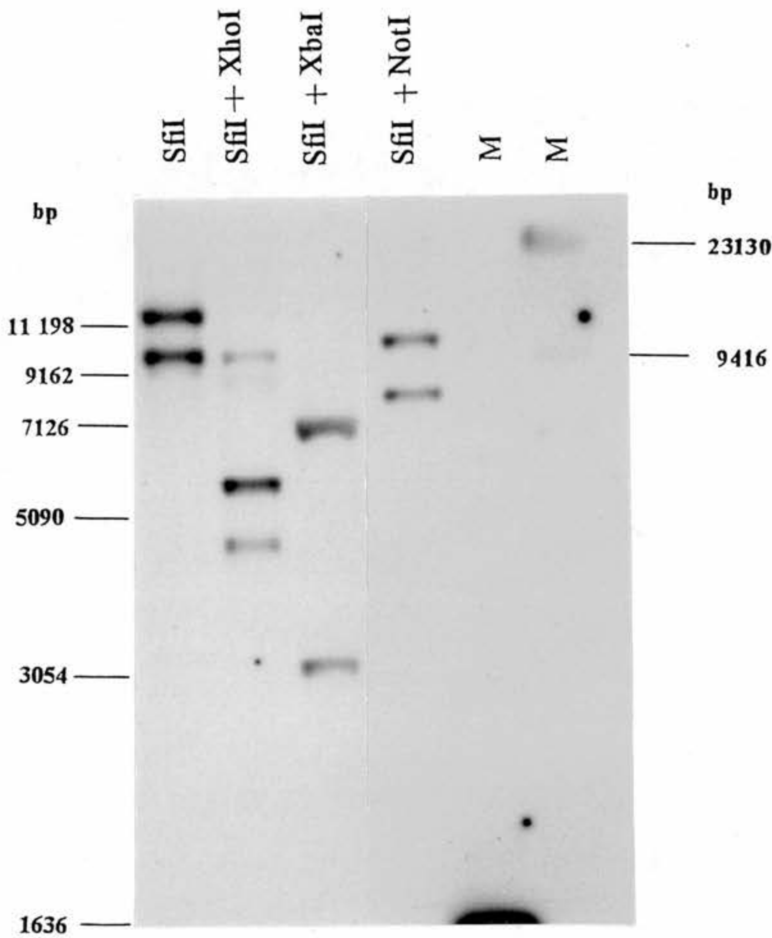


Diagram 5.14: Restriction sites for SfiI on hybridisation with ACY7.
Conventional gel electrophoresis produced two fragments for SfiI of about 10 kb and 15 kb. The SfiI/NotI double digest placed these sites in postions + 8 and + 18.5 in the restriction map in diagram 5.16. These positions were confirmed by the SfiI/XhoI and SfiI/XbaI digests which are shown. The known restriction sites for XbaI are shown in diagram 5.1. A gDNA sample from the same individual used for the pulsed field gels was used for these digests.

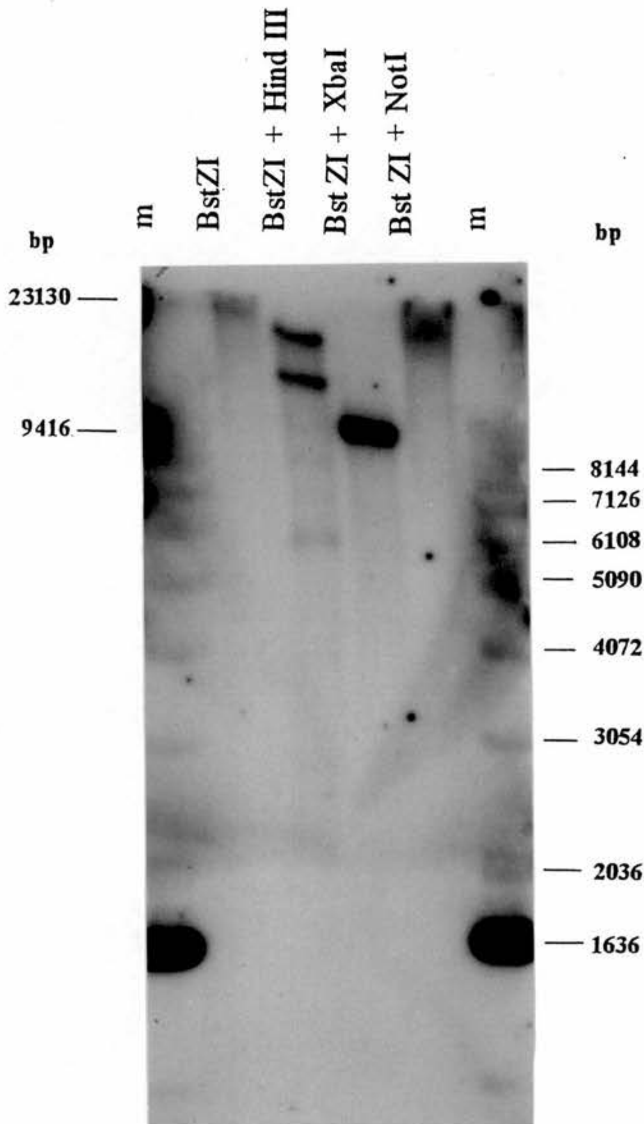


Diagram 5.15: Localisation of the restriction sites for BstZI
Conventional gel electrophoresis was used to analyse the position of restriction sites for the enzyme BstZI. Genomic DNA from the same individual as used in the pulsed field gel was used. The blot was hybridised with the probe ACY7

5.6.6 SacII restriction sites

The SacII single digest in diagram 5.12 shows two bands of 50 kb and 130 kb when hybridised with ACY7, this pattern remains the same in the SacII/NotI double digest. This suggests that the SacII sites lie within, or very close to, the NotI sites. The two bands hybridised with ACY7 in the single SacII digest must to be due to a partial digest. The alternative explanation for two bands, that a SacII restriction site exists within ACY7, cannot be correct because the SacII sites do not span the NotI site 7kb upstream from ACY7 and no SacII sites were identified in the sequence of ACY7. Thus, SacII sites have been marked on the restriction map in diagram 5.16 in positions 0, +50 and +130. Probing the same blot in diagram 5.12 with ACIXSR produced a 50 kb fragment in the single SacII digest. This pattern does not change in the SacII/NotI digest, thus a SacII site lies at position -50 in the restriction map.

5.6.7 BspDI restriction sites

Very large DNA fragments of > 776 kb were produced on probing BspDI digests with both ACY7 and ACIXSR (shown in diagram 5.11). Hybridisation to double digests with BspDI and NotI identified the same fragment sizes as the NotI single digest, thus the nearest BspDI sites lie outwith the NotI sites and cannot be marked on the restriction map in diagram 5.16.

5.6.8 BstZI restriction sites

Like the enzymes XhoI and SfiI, BstZI was found to give very small DNA fragments in single digests (see diagram 5.7). In diagram 5.9, ACY7 hybridises to a 38 kb fragment, thus the restriction sites for BstZI were mapped using conventional electrophoresis (see diagram 5.15). A BstZI/HindIII double digest gave fragments

of 14 kb and 21 kb. Since the position of the HindIII site is known this gives two possible locations for the two BstZI sites, those are positions -5 and +30 or +2 and +37. In the BstZI/NotI double digest, ACY7 hybridises to a band which is clearly smaller than that in the BstZI single digests. Since the NotI site is known to be in position 0 and to cut completely, the two locations for the BstZI sites must be -5 and + 30, these are marked on the restriction map in diagram 5.16. ACIXSR hybridises to a 38 kb band in the single BstZI digest (see diagram 5.9) locating an upstream restriction site to position - 45.

5.7 The long range restriction map

The restriction map produced by these experiments is shown in diagram 5.16, it shows a cluster of restriction sites around the exons of PDGFA. Sites are present for all the restriction enzymes which were studied, with the exception of NruI and BspDI. The enzymes XhoI, SfiI and BstZI all have several restriction sites within the area spanned by the exons of PDGFA. For other enzymes such as NotI, MluI and SacII the sizes of fragments hybridising to ACY7 are also relatively small, the majority being less than 100 kb. This pattern clearly demonstrates a CpG island (Bird 1986). The apparent extent of this CpG island on the basis of the restriction mapping is marked in the diagram. This cluster of restriction sites for rare cutting enzymes is most marked around position 0 on the restriction map which corresponds to the promoter region for PDGFA. The identification of a CpG island around the start of the gene is not surprising. This cluster of restriction sites covers about 25 kb which means that it spans the majority of the exons of the PDGFA gene.

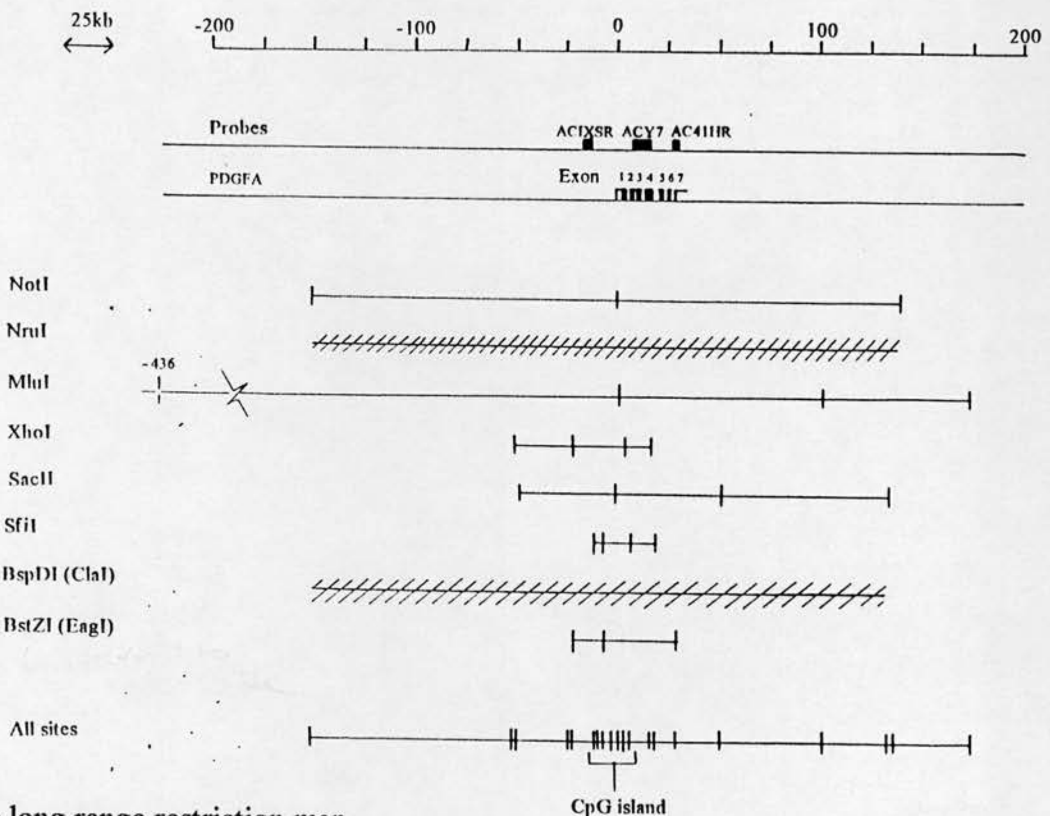


Diagram 5.16: The long range restriction map

The long range restriction map derived from the pulsed field restriction digests using the two probes ACIXSR and ACY7. The exons of PDGFA are shown for comparison to the positions of the the restriction sites. The bottom line of the diagram includes a mark for every site which has been identified for each of the restriction enzymes which are shown. On the basis of the restriction site distribution, the CpG island at the 5' end of PDGFA spans the first three exons of the gene

A dramatic increase in fragment size for enzymes such as MluI and SfiI is seen on hybridisation with ACIXSR. This second probe lies about 20 kb upstream from exon 1 of PDGFA. As the restriction map shows, this is upstream from the CpG island. Thus, this probe would be expected to identify much larger fragments.

5.8 Extension of the restriction map

Extending the long range restriction map as far as possible is obviously of great interest if the map is to be useful for a long range physical mapping project. The downstream probe AC41HR which maps to exon 7 of PDGFA has been described above (see diagram 5.1). This probe is from the extreme 3' end of the cloned region around the gene and is therefore potentially useful for extending the restriction map. However, as described above, the majority of the enzymes which were used for the restriction map have been mapped to sites beyond AC41HR and this probe would provide very little further information.

Extension of the restriction map described above was not possible given the resources which were available at this time during the project. The probes ACIXSR and AC41HR mark the limits of the cloned region around PDGFA thus extending the restriction map in this direction requires new clones to be identified. Towards this aim, YAC and P1 clone libraries were screened (described in Chapters 6 and 7). It was hoped that this library screening would extend the cloned region and provide further probes. Restriction sites for the enzymes BspDI and NruI have not been found within the mapped region. The large fragments produced by these enzymes could be useful for extending the restriction map if more distant probes were identified.

5.9 Discussion and conclusion

The physical distance between *PDGFA* and its nearest proximal marker MS31 (*D7S21*) cannot be measured in this pulsed-field mapping project given the probes which are available. Thus the distance between *PDGFA* and MS31 remains difficult to estimate. As discussed, above it is likely that the distance between the two loci is not as large as the linkage data suggests. Unfortunately this PFGE analysis cannot confirm or refute this hypothesis.

The presence of a CpG island at the 5' end of PDGFA is not surprising since sequence data had already shown that the region had a high CG content (Bonthron 1988). CpG islands are described in section 1.13, they are defined on the basis of the CpG content of the sequence of a region. Thus the presence of such a CpG island around PDGFA in this project is inferred from the cluster of restriction sites for rare-cutting enzymes. CpG islands have been described at the 5' end of most genes which, like PDGFA, are widely expressed. They are less commonly associated with genes that have a restricted, tissue specific pattern of expression. The size of CpG islands appears to be similar regardless of the size of the associated gene. Consequently a wide variation in the area covered by CpG island has been described, with some genes the CpG island is found to be tightly restricted to the 5' promoter region while other genes such as the α-globin gene have been found to have CpG islands which span their entire length (Bird 1987b). As discussed above, this CpG island across PDGFA spans most of the gene. No functional significance has been determined for this variation in the extent of the CpG islands.

The major aim of this project is to map PDGFA within the region of 7p22, thus the possibility of other loci which could be physically mapped to PDGFA should be

considered. Unfortunately there are few other probes available for the telomeric region of chromosome 7p and most of the loci which were used in the linkage analysis in Chapter 4 were based on PCR assays for which there are no probes available. The second set of markers which were analysed in the Eurogem linkage analysis placed the marker *D7S596* as the nearest proximal marker to *PDGFA*. Unfortunately the linkage analyses do not place the relative order of MS31 and D7S596 in the same linkage map, thus there is no information available about the distance between MS31 and *D7S596*. Also, there are no probes available for *D7S596* and the possibility of establishing physical linkage between this locus and *PDGFA* has not been pursued. Thus, at present there are no candidates other than MS31 which could be used to establish the distance between PDGFA and proximal loci.

The linkage analysis presented in Chapter 4 suggests that PDGFA may have a subtelomeric location. Thus a second goal for this physical mapping project is to physically link the gene to the telomere of 7p. Attempts have been made, both in this project and by others, to try to isolate a suitable single copy probe using candidate chromosome 7p telomere YACs (see Chapter 8). Isolating single copy probes from the telomeric regions of chromosomes is complicated by the presence of a high number of repeated sequences which may be present on several chromosome telomeres and unfortunately there are no telomere probes available at present. The extension of the restriction mapping project towards the telomere of PDGFA cannot be undertaken until suitable probes are available.

6. Chapter 6: Analysis of the human YAC libraries.

6.1 Introduction

To allow the construction of a clone contig, human YAC clone libraries were screened for clones spanning the PDGFA gene. Traditionally, the screening of libraries has been undertaken using filter hybridisation of radiolabelled probes to gridded filters. The gridded filters have the individual clones laid out on nylon filters. When a potentially positive clone is detected, the original YAC clone can immediately be identified from its position on the grid. More recently most libraries have been constructed as a series of pooled samples which contain many clones. These pools can be analysed by PCR screening the master or primary pools. For those pools which screen positive, that is that they contain a clone carrying the target DNA, the secondary pools and subsequently the tertiary pools are screened. Each round of testing screens progressively fewer YACS until the original clone can be identified (Green 1990). In this project both filter hybridisation and PCR analysis were used to screen the YAC libraries for the PDGFA gene.

A number of YAC libraries are now available. Both total human genomic DNA libraries and chromosome 7 specific libraries were screened. The libraries which were screened for this project are described in section 1.14.

6.2 PCR analysis of the YAC libraries

Three PCR assays from the PDGFA gene were used to screen by PCR, these identified loci in exon/intron 3 (primers 631/632), intron 4 (primers ApaL/844) and intron 5 (430/731). The primer sequences and conditions used for each PCR assay are described in Chapter 2 (table 2.2). The location of each STS along the PDGFA gene is shown in diagram 6.1.

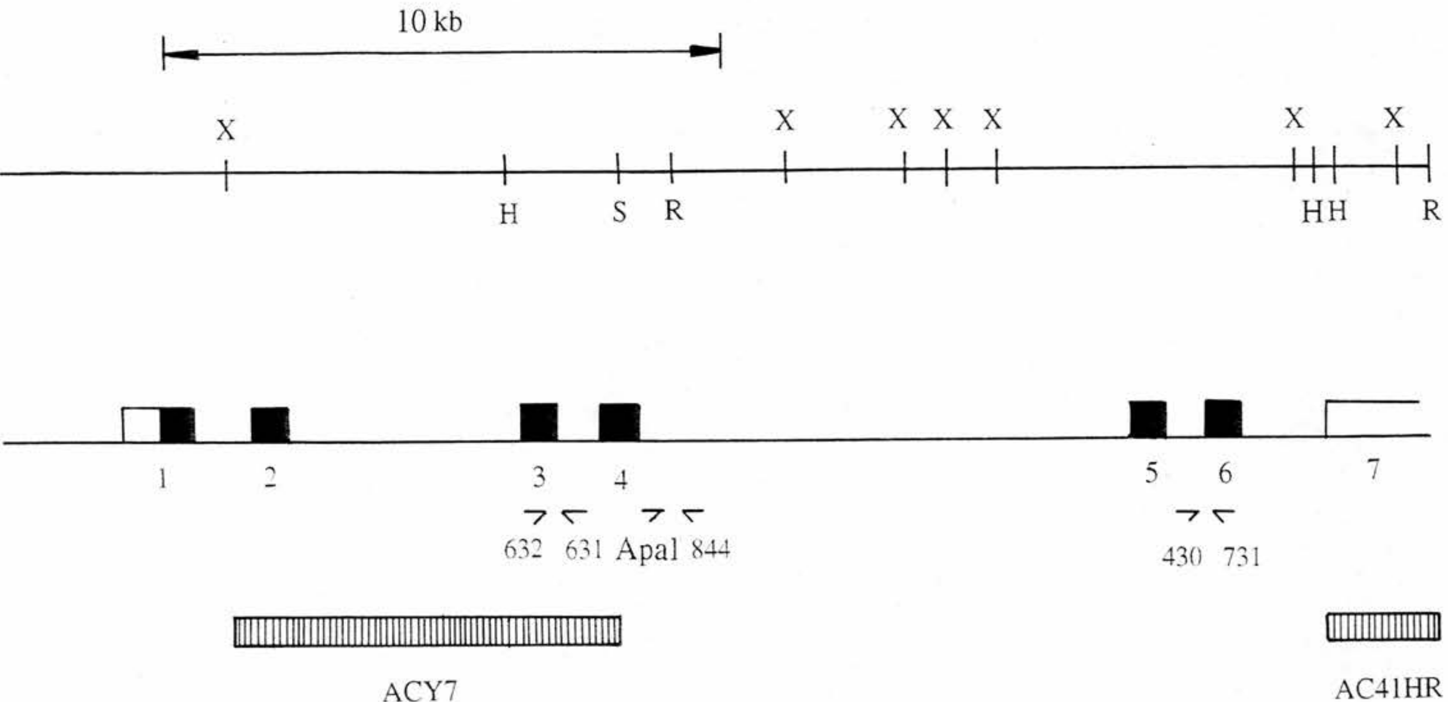


Diagram 6.1: The PCR assays and probes used to screen the YAC libraries for clones spanning the PDGFA gene

The primer pair ApaLI/844 was used in the linkage analysis in Chapter 4. Primer pair 632/631 include the downstream primer 631 which was used to analyse the exon 3 dimorphism in the linkage analysis. Primer pair 430/731 are longer 30 base primers which were designed specifically for the library screening. The probes ACY7 and AC41HR are described in the pulsed-field analysis in Chapter 5, ACY7 was also fully sequenced, see Chapter 3.

The intron 4 primer pair, ApaL/844, identify a dimorphic site which is described in Chapter 3 (see diagram 3.6) and was used in the linkage analysis in Chapter 4. The other two primer pairs amplify sites which are not polymorphic. The primer pair, 631/632, uses the downstream primer, 631, which was used to amplify the dimorphic site in exon 3 for the linkage analysis. However the upstream primer used to amplify the dimorphic site, 920601A, contains a mismatched base in its sequence (discussed in section 3.6). It has been substituted with the primer 632 which lies downstream of the variant base. This provided a more 'robust' PCR assay. The intron 6 primer pair (430/731) are larger oligonucleotides of 30 bases, this STS was developed specifically for the YAC library screening.

Table 6.2 summarises the YAC libraries which were screened, their source and the PCR assays which were used. The majority of the resolved PCR products were blotted onto Hybond N^+ membrane and then probed with the radiolabelled PCR product. This was done to ensure that there was no PCR product in the reaction. A poor amplification of the target DNA may not have been visible by eye but would have been detected by this approach. Where the PCR products were blotted and probed this is noted in table 6.2.

The Washington total genomic library and both the chromosome 7 libraries (see section 1.14) are not made widely available. All three sets of the PCR primers shown in diagram 6.1 were sent to E. Green (NIH, Bethesda, USA) and to S. Scherer in L-C Tsui's laboratory in Toronto, Canada. They both kindly screened their chromosome 7 specific libraries using the primer pairs noted in table 6.2. In addition, E. Green also screened the ICI total genomic DNA library, the Washington total genomic DNA library and the telomere YAC library from H. Riethman. The telomere YAC library from H. Riethman is discussed further in Chapter 8: Telomere YACs.

Library	Source	Primer pair	? probed	Comments
ICI	K.Dry	ApaL/844	yes	no positives
		631/632	yes	no positives
		430/731	no	no positives
		Factor VIII	no	positive control see diagram 6.5
	J.Warner	430/731	yes - pool 28 positive	2° pool screened - no positives (diagram 6.6)
	E. Green	ApaL/844	not available	screened by E. Green - no positives
		631/632	not available	
		430/731	not available	
CEPH	M. Jones	ApaL/844	no	no positives
		631/632	yes	no positives see diagram 6.7
		430/731	yes- pool 27 positive	2° pool screened - no positives
Washington total genomic	E. Green	ApaL/844	not available	screened by E.Green no positives
		631/632	not available	no positives
		430/731	not available	no positives
Washington chromosome 7	E. Green	ApaL/844	not available	no positives
		631/632	not available	no positives
		430/731	not available	no positives
Toronto chromosome 7	L-C Tsui	430/731	not available	screened by S.Scherer

Table 6.2: PCR analysis of the YAC libraries

A summary of the PCR screening of the human total genomic and chromosome 7 specific YAC clone libraries. Whenever possible the agarose electrophoresis separated PCR products were Southern blotted and probed using radiolabelled PCR product from a gDNA PCR control, this is indicated in column 4.

6.2.1 A Positive Control: A PCR assay for the Factor VIII Gene

To ensure that the aliquots of the YAC library primary pools and the general method PCR screening were reliable, a PCR for a different gene and chromosomal location was used as a positive control. Thus, the library was screened using a PCR assay which identified an STS from intron 19 of the Factor VIII gene (Graham 1990). This assay was selected since it was known that the Factor VIII gene was represented in the library (pers. comm. J. Warner). The PCR primers and the reaction conditions were kindly supplied by L. Strain (Human Genetics Unit, Edinburgh University) and are described in table 2.2. It is acknowledged that this control would not identify failures to find positive YACs which were due to problems specific to the PDGFA PCR assays.

6.3 Filter hybridisation analysis of the YAC libraries

A PCR reaction may fail if it is particularly sensitive to the presence of agarose or to the low levels of target DNA which are present for each individual YAC in the primary pools. An alternative approach which avoids these problems is to screen the YAC libraries by filter hybridisation using membranes which have been gridded with an array of individual YACs. The ICI library, which is described in section 1.14.1, was kindly supplied as a set of 40 gridded filters by R. Elaswarapu (HGMP Resource Centre, London).

Two single copy probes from PDGFA were used to probe the ICI YAC library filters. These two probes, ACY7 and AC41HR, are shown in diagram 6.1 and they are described in section 5.2. Subsequent to this a third probe, ACIXSR, which maps to a region about 20 kb upstream from PDGFA was developed (see section 5.2). Since this probe is also single copy it would potentially have been useful in the

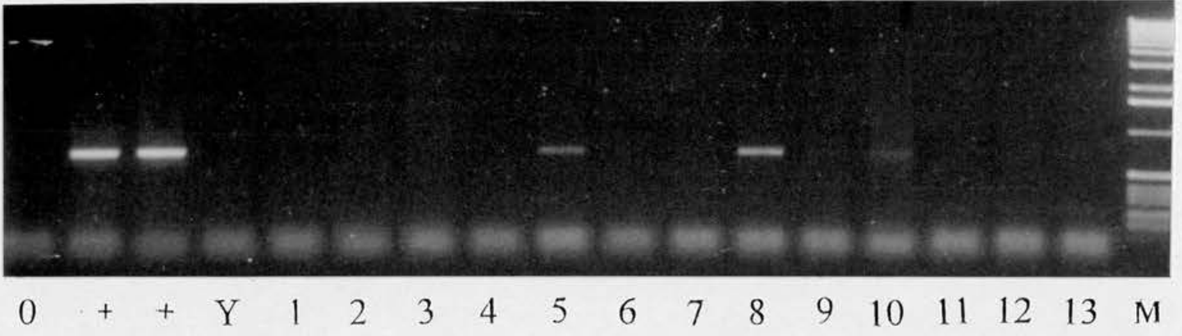
screening of the YAC libraries. Using this probe to screen the library would have extended the area around the gene for which screening had been undertaken. However the results from using this probe to screen the ICRF P1 library, which are described in Chapter 7, led to the decision not to use this probe to screen the ICI YAC library.

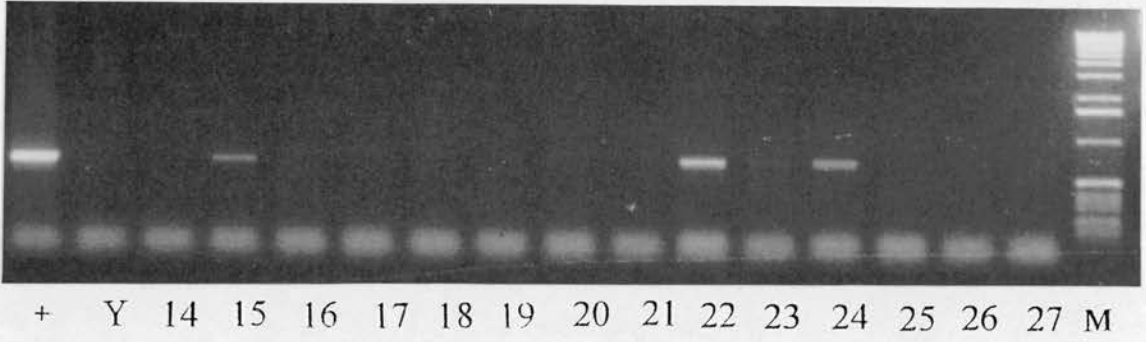
6.4 Results of the YAC library screening

6.4.1 The positive control: the Factor VIII gene

The Factor VIII gene PCR assay was used to screen the ICI YAC library using the primary pools from K. Dry. The results of screening the ICI YAC library are shown in diagram 6.3. In addition to the 0 control (marked '0') and the genomic DNA positive control (marked '+'), a yeast gDNA sample was included (marked 'Y') to ensure that the PCR primers do not amplify the yeast DNA background. Three strong (pools 8, 21 and 24) and three fainter (pools 5, 15 and 26) PCR products were seen in the primary pools. This number of potential positives corresponds to the expected number for this library which is estimated to have a 3 to 4 times coverage of the genome. Since this screen was used simply as a positive control to check the integrity of the samples and the screening procedure, the secondary pools for these potential positives were not screened.

This positive control was not used on the second set of samples of the ICI YAC library from J. Warner or the CEPH 'megaYAC' library since there was a limited amount of each sample available. The results of this positive control suggest that the YAC library primary pools and the PCR screening method are reliable.





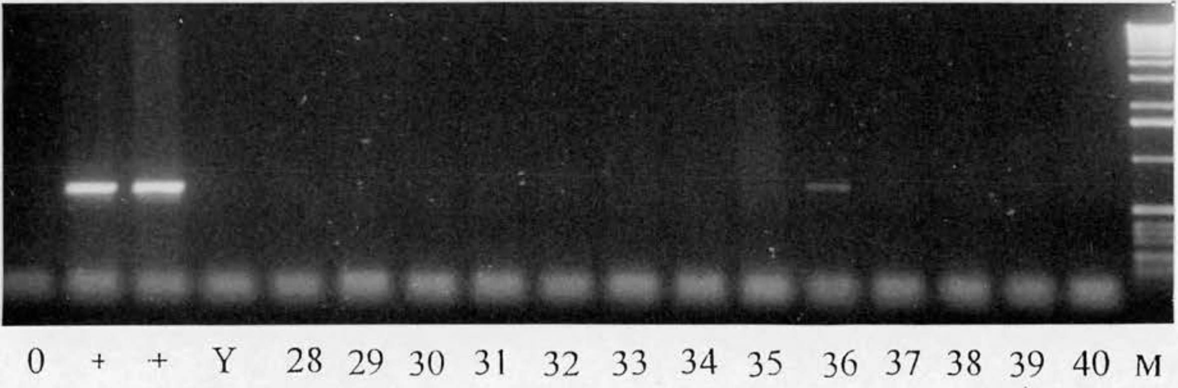


Diagram 6.3: Analysis of the ICI YAC library for Factor VIII.

The ICI YAC library primary pools screened with the PCR assay for intron 19 of the Factor VIII gene. Strong positive PCR products are seen in pools 8, 22 and 24, fainter PCR products are seen in pools 5, 15 and 36.

6.4.2 Screening for PDGFA

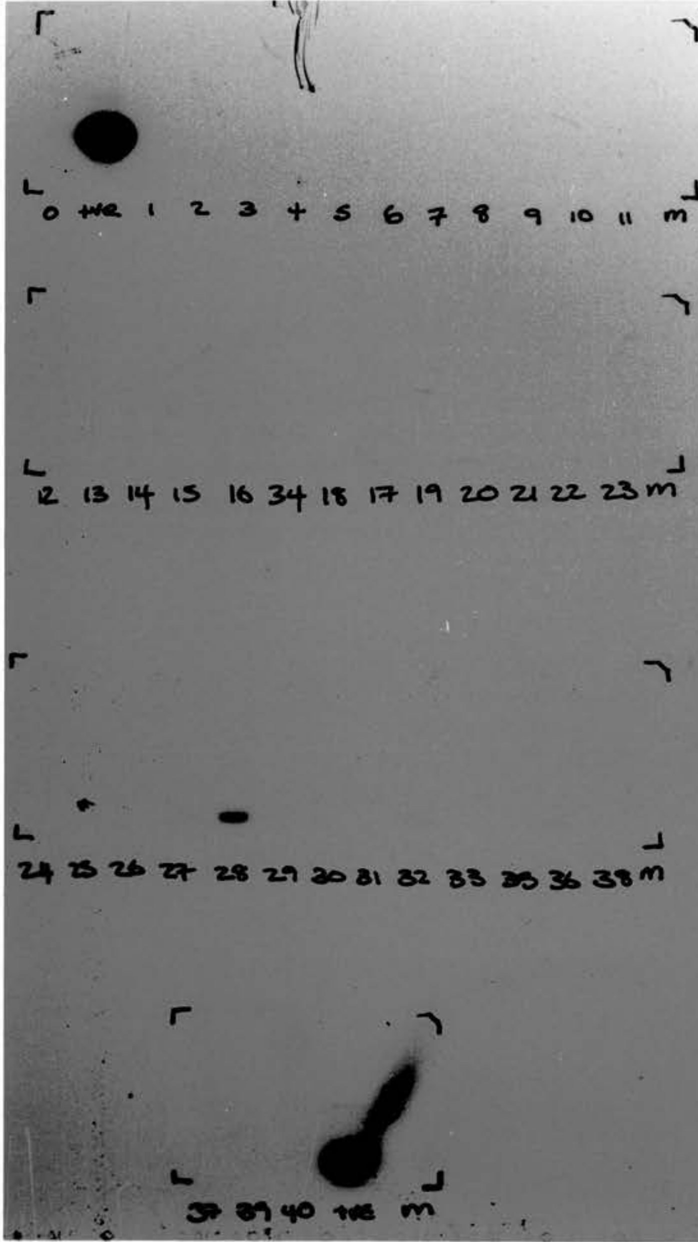
The total genomic and chromosome 7 YAC libraries listed in table 6.2 were screened extensively using three different PCR assays. In addition, the ICI YAC library was screened by filter hybridisation using two single copy probes from PDGFA. Despite this exhaustive screening project, no positive clones for PDGFA were identified. An example of the negative results from both the ICI YAC library is shown in diagram 6.4. The PCR products were blotted and probed, identifying one apparently positive pool, number 28. However screening of the secondary pools for pool 28 did not identify any positive clones suggesting that this weak positive signal was due to contamination of the first PCR reaction.

6.5 Conclusion and discussion

The failure to find any positive clones for PDGFA from the YAC libraries despite an exhaustive search is interesting. Since it is possible that any individual library may not contain clones spanning PDGFA, five different libraries have been screened. The absence of positive clones from all of these YAC libraries suggests that this region is genuinely not represented in these libraries and that clones have not been accidentally missed.

However the lack of positive clones could be due to a failure to find clones from PDGFA which are present in the libraries, so a number of steps have been taken to avoid sources of error. It is theoretically possible that the aliquots of the primary pools for a library from an individual source contained faults which led to a positive clone being missed. To eliminate this possibility, three different sources of the ICI YAC library and two separate sets of aliquots of the CEPH 'megaYAC' library were screened.

A



B

Errors arising from using the PCR approach to screening the libraries have been countered by screening by hybridisation of radiolabelled probes to gridded filters. In addition, three different PCR assays for PDGFA were used. The positive control PCR assay for the Factor VIII gene, used with the ICI library, has demonstrated that other PCR assays are effective. Also, since agarose in the samples may chelate magnesium, inhibiting the PCR reaction, the PCR screening was repeated using $MgCl_2$ concentrations increased to 2.5 mM. The possibility that the PCR reactions had failed to amplify sufficient product to be seen in the agarose gels was investigated by blotting and probing the PCR 'products'.

Finally, the possibility of operator errors leading to the failure to find positive clones has been countered by the independent screening undertaken by the two other laboratories in Toronto and Washington.

It is possible that the failure to find clones may be due to chance. However, the likelihood of any locus being present in a YAC clone library with a genome coverage of 3 to 4 times is about 95 %. For this project three total genomic libraries were screened for PDGFA, giving a combined probability of > 99 % for finding a clone spanning the gene. In addition, two chromosome 7 specific libraries were screened. The screening of several genomic and chromosome 7 specific YAC libraries lowers the likelihood that the failure to find any clones was due to chance.

It is possible that the PDGFA gene is not represented in these YAC libraries because it is unclonable in YACs. The use of the eukaryotic yeast as host for the YAC cloning system was expected to overcome the problems associated with cloning some regions in *E.Coli*. For many regions of the human genome the YAC system is a great improvement, however the cloning of some sequences into YACs leads to structural instability of the recombinant clones. This instability may cause breakdown of the YAC clone producing several YAC bands of different sizes or the cloned DNA insert may be found to be rearranged without a change in the clone size. If the clone is particularly unstable then it may be completely deleted and thus lost

from the library. Very little information has been published about DNA regions which are unclonable or unstable in YACs. A review on this topic which depended largely on personal communication reports rather than published data concluded that highly repetitive regions of the human genome such as rDNA and alphoid satellites were the major causes of instability in YAC clones (Vilageliu 1992). The Y chromosome repeat sequences have also been shown to be unstable in YAC clones (Neil 1990). From the data which is described above it is not possible to determine whether YAC clones spanning PDGFA are not represented as a consequence of clone instability.

The search for clones spanning PDGFA in an attempt to build a physical map of the region was continued using two separate approaches. Firstly, since the region may well be unstable in YACs, an alternative vector system was sought. For this the ICRF P1 clone library was screened, the screening of this library is described in Chapter 7. An alternative possibility, given the results of the linkage analysis described in Chapter 4, is that the PDGFA gene lies very close to the telomere. Such regions are difficult to clone in the standard YAC vectors but are clonable using a telomere YAC vector system. Thus a telomere YAC library was sought and analysed, this is described in Chapter 8.

7. Chapter 7: Analysis of the ICRF human gDNA P1 library.

7.1 The P1 cloning vector

This project aimed to generate a physical map of the PDGFA locus, instrumental to this was the construction of a contig of clones from the region. Since P1 clones carry much smaller DNA inserts than YACs they are not the system of choice for a mapping project such as this. However, following the screening of the YAC clone libraries (described in Chapter 6) which did not find any YAC clones which spanned *PDGFA*, an alternative cloning vector system was sought.

The P1 cloning system, which is described in section 1.13.5, can clone inserts up to 95 kb in size which is considerably larger than the insert size possible in cosmid clones (up to 40 kb). The P1 cloning vector system has been shown to be useful for cloning regions which are unstable in YACs and where cosmid walks have failed (Halford 1993). During the construction of the *S. pombe* whole-genome mapping project, the P1 cloning vector was found to be more successful than cosmids in spanning rDNA repeat regions (Hoheisel 1993, Francis 1994). It is possible that instability of the YAC clones, perhaps caused by repetitive regions of sequence, is the reason why no YAC clones of PDGFA have been identified. More importantly, the P1 cloning system has been used to make a human total genomic DNA clone library which is available to the scientific community for screening (Francis 1994). These factors made the P1 cloning system an obvious choice for a second round of library screening in an attempt to find clones spanning PDGFA.

7.2 Probes used to screen the ICRF P1 clone library

Three different single copies from PDGFA were used to screen the library. The probes ACY7, AC41HR and ACIXSR are described in Chapter 5, see diagram 5.2. Both ACY7 and ACIXSR were used extensively in the pulsed-field restriction mapping experiments. The probes, ACY7 and AC41HR, were also used to screen

the ICI YAC library, as described in Chapter 6. A genomic DNA probe for the yeast asparaginase gene, ASP1 (Sinclair 1994) was included in the hybridisations to hybridise to the yeast background to allow the position in the grid of any positive colonies to be identified. This probe, which was supplied by D.T. Bonthron, was labelled with [α-^{35}S] dCTP to give a weaker signal than the [α-^{35}P] dCTP labelled human gDNA probes.

The ICRF P1 clone library was sent as a gift by F. Francis (ICRF, London). The library is supplied as two sets of two duplicates filters. In each case, all four filters of the P1 clone library were screened with each probe. The duplication of the filters gave some degree of internal control which was used to asses whether potentially positive colonies were genuine. The pattern of the background grid suggested that the filters are gridded in 6 blocks whose position is varied but whose internal order remains the same. Thus, although the clones were not gridded in the same order on the two duplicate filters, the order of the clones within each segment of the grid was the same. Thus any positive clones within the same segment would have the same spatial orientation to each other on the separate filters. This is discussed further in section 7.4, below.

The level of background DNA hybridised by the yeast ASP1 gDNA probe was found to be very variable. The grid shown in diagram 7.1 is an extremely good example, other filters were found to have much more patchy background. Thus, a major advantage of using the two filters was that it gave two opportunities to detect a positive clone.

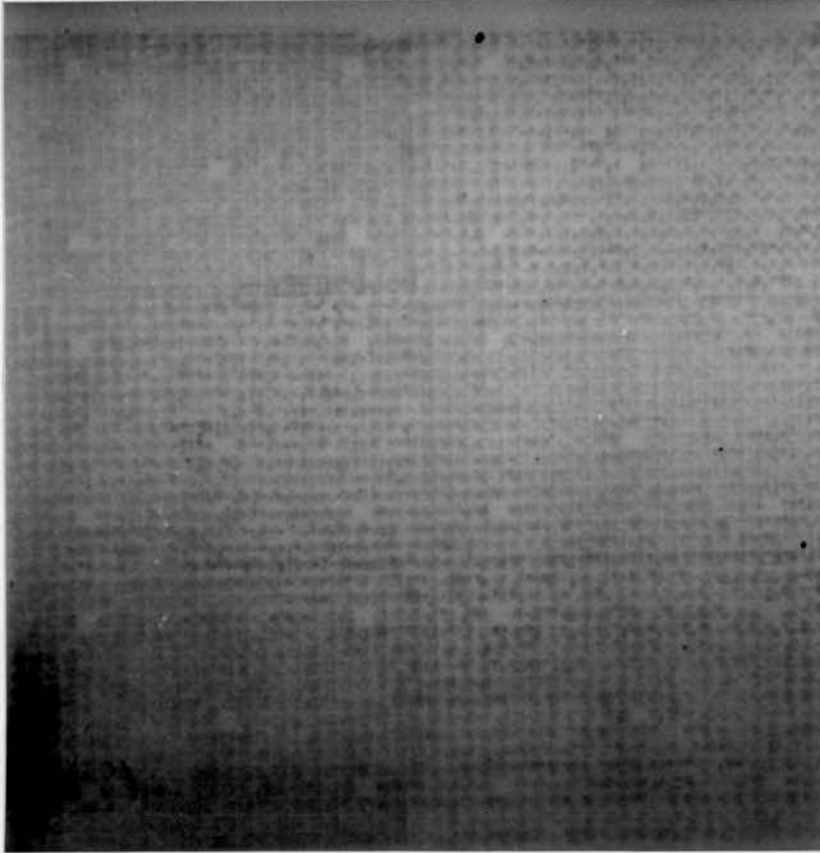


Diagram 7.1: Filter hybridisation of the P1 library using PDGFA probe ACY7

Screening the ICRF P1 library by filter hybridisation using the PDGFA probe ACY7. The background filter grid is hybridised using a[α-^{35}S] dCTP labelled yeast ASP1 gene probe. No positive colonies are identified.

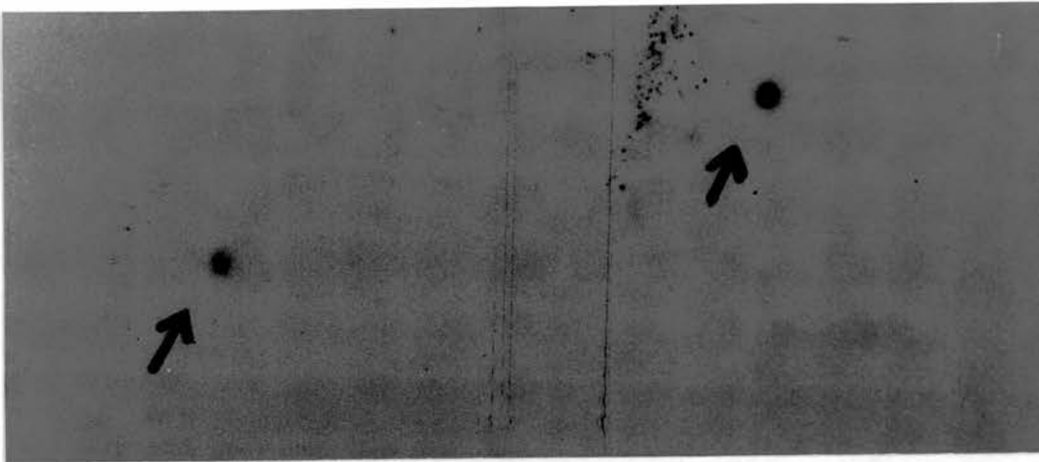
7.3 Results of the P1 library filter hybridisations

Screening the library with the probes ACY7 and AC41HR did not detect any positive colonies. An example filter (100.5.4) probed with ACY7 and the ASP1 probe is shown in diagram 7.1. There were no positive colonies identified in this hybridisation.

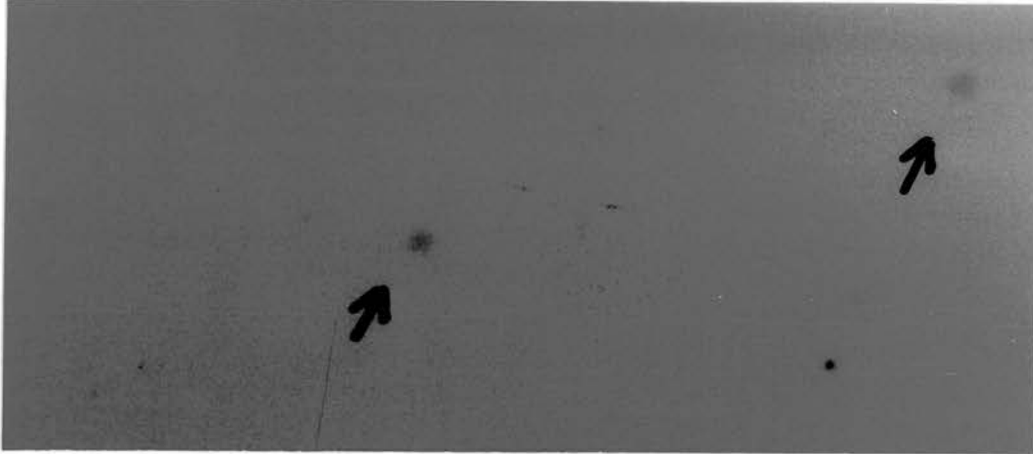
The third probe which was used to screen the P1 library filters was ACIXSR which, like ACY7, was used in the pulsed-field mapping in Chapter 5. Using ACIXSR, five potential positives clones were identified, these are shown in diagram 7.2. The two filters 100.5.48 and 100.5.4 are duplicate filters of the same set of clones. The two positive colonies which were identified in each of these two filters showed the same orientation to each other although they were not in the same position on each filter. As discussed above, this suggested that the same two clones had been identified. The third potentially positive colony on filter 110.5.36 was not matched by a similar positive on the duplicate filter. However the appearance of this colony and comparison to the level of background hybridisation suggested that it was a positive clone.

7.3.1 Identification of the positive clones

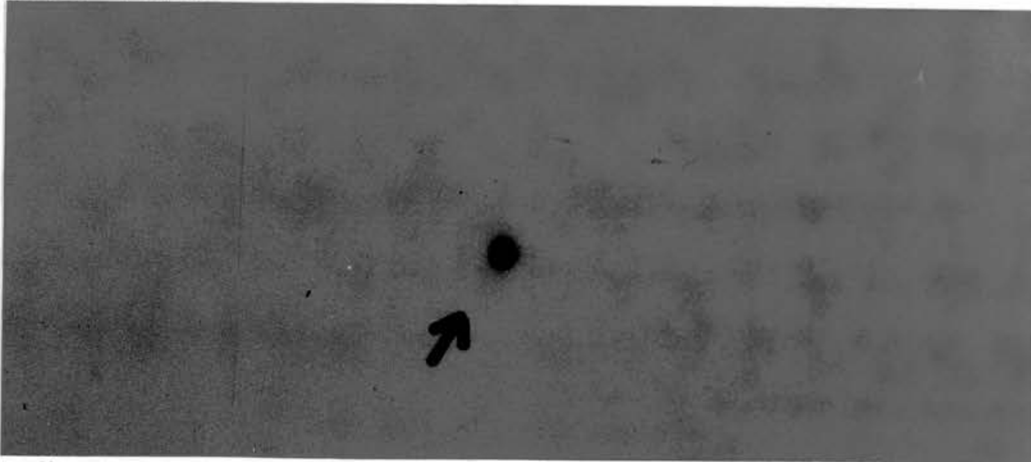
The potential positive clones were identified by their filter and grid numbers and were requested from ICRF, they are listed in table 7.3. As expected, the positive clones identified in filters 100-5-4 and 100-5-48 were found to be duplicates of the same P1 clones, thus three P1 clones were returned for analysis.



Filter 100.5.48



Filter 100.5.4



Filter 110.5.36

Diagram 7.2: Potential positive clones identified by hybridisation with ACIXSR

Five potential positive clones on three different filters were identified by screening the ICRF P1 library using the probe ACXISR. The positive clones on the duplicate filters 100.5.48 and 100.5.4 look like a matching pair of clones.

More surprisingly, the clones J1174 and M0356, which had been identified in duplicate and seemed to be promising as potential positive clones, had also been identified and requested by two other groups. Neither group were using probes specifically from chromosome 7p. It was surprising that these clones had been identified and requested by other groups who were not working on the same chromosomal region since the probe which had been used, ACIXSR, has been shown to be single copy in the pulsed field experiments in Chapter 5.. The subsequent analysis of these P1 clones made this observation less surprising.

7.4 Analysis of the P1 clones

The clones were supplied by ICRF as agar stabs, on receipt they were streaked out onto LB/kanamycin agar plates. From these plates 16 to 18 single colonies were selected and transferred to Hybond N^+ filters and to fresh agar plates as described in section 2.7. The colonies were screened by filter hybridisation using the probe ACIXSR and the results of this experiment are shown in diagram 7.4. ACIXSR hybridised to several of the clones from M0356 and J1174, although the signals were surprisingly weak, requiring an overnight exposure to the X-ray film. Less surprisingly, no positive clones were identified from clone L101, the clone which had been identified only on one filter.

7.4.1 Fluorescent *in situ* hybridisation.

DNA was extracted from a large scale culture (induced with IPTG) of a positive clone from both J1174 and M0356. The DNA was LiCl/PEG purified and then biotin labelled for use in fluorescent *in situ* hybridisation (see section 2.14). The biotin labelled probes were hybridised to metaphase chromosome spreads of a normal male which were prepared by T. Johnson.

Filter Number	Clone position		Clone identified
100-5-4	X = 136	Y = 12	ICRF P700M0356
	X = 112	Y = 20	ICRF P700J1174
100-5-48	X = 64	Y = 60	ICRF P700M0356
	X = 40	Y = 68	ICRF P700J1174
110-5-36	X = 43	Y = 63	ICRF P700L101

Table 7.3: The positive clones requested from ICRF.

The five positive clones shown in diagram 7.2 were identified by their grid numbers and requested from ICRF. As expected, the two pairs from filters 100.5.48 and 100.5.4 identified the same two clones, thus three clones were returned for analysis.

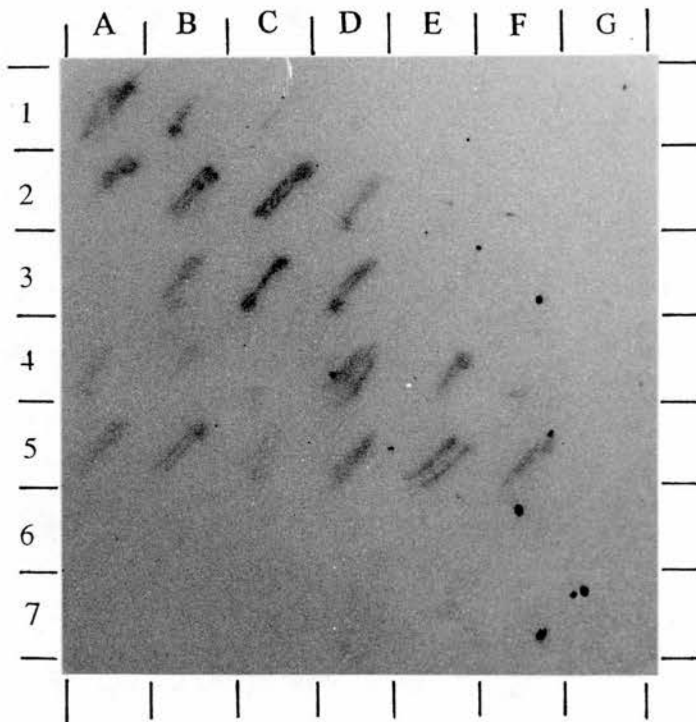


Diagram 7.4: Screening single colonies from the P1 clones.

Single colonies from the agar stabs returned by ICRF were screened by filter hybridisation using ACIXSR. Clones A1-D3 are from clone J1174, E3 - F5 are from clone M0356 and clones G5 - G7 are from L101. Several positives are seen for clones J1174 and M0356,, there are no positive colonies from L101.

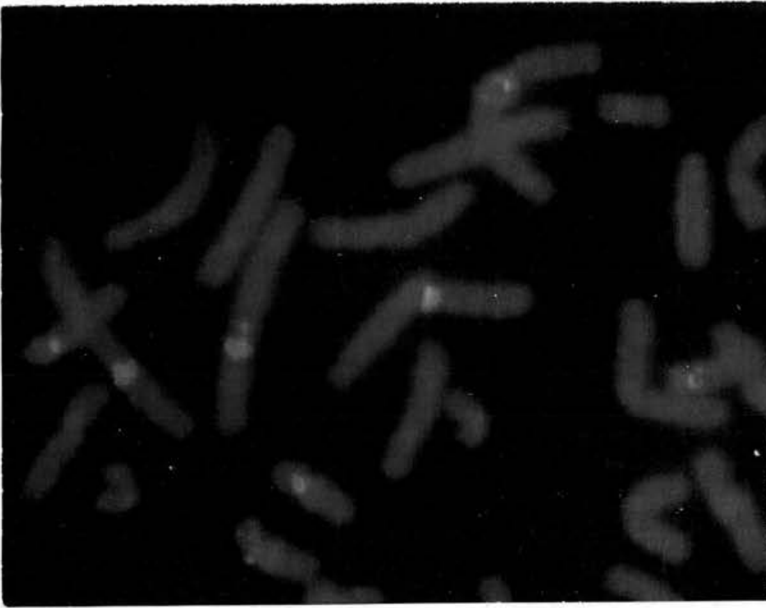
The FISH hybridisation results showed that both probes hybridised strongly to a region near the centromeres of several pairs of chromosomes (see diagram 7.5). Analysis of the metaphase spreads was conducted by J. Fantes (MRC Human Genetics Unit, Edinburgh). Clone J1174 appears to hybridise to chromosomes 1, 2, 15 and 16. The signal on the centromere of chromosome 2 is especially strong. The second clone, M0356, hybridises to the centromeres of chromosomes 1, 5 and 19. Thus the pattern of hybridisation is clearly different for the two probes, suggesting that they contain different sequences. No hybridisation to chromosome 7p could be detected for either clone despite considerable effort.

7.4.2 Restriction digests of P1 clones.

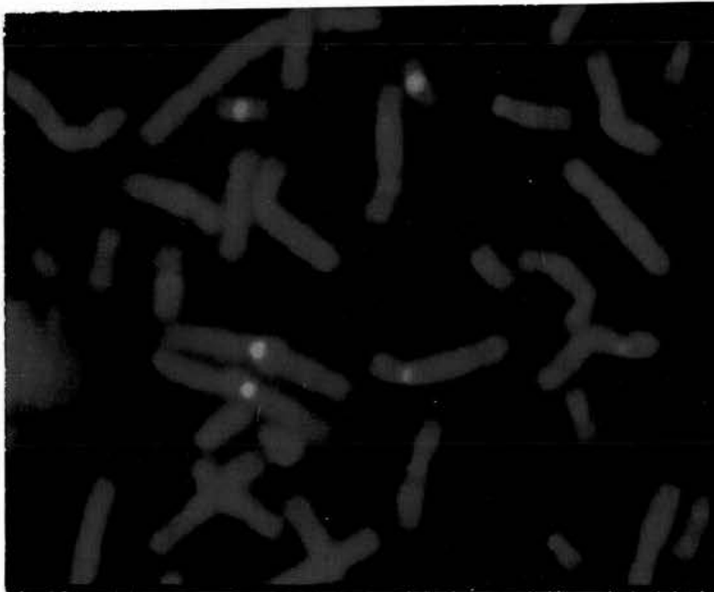
Although no hybridisation to chromosome 7p could be detected for either of the two P1 clones it remained theoretically possible that the clones contained single copy sequence from the PDGFA region. The strong hybridisation to the centromeric regions of the other chromosomes could swamp the signal from hybridisation to 7p. With this in mind the P1 clones were analysed further by restriction digests.

Restriction digests were performed using the enzymes NotI, SalI, SfiI, EcoRI, SpeI, BglII, XbaI and HindIII. These digests suggested that the total size of J1174 is around 30 kb, while M0356 seems to be about 20 kb. The P1 vector is about 16 kb in size, which means that the clones carry inserts of 14 kb and 4 kb respectively.

Hybridisation of the ACIXSR probe to Southern blots of the EcoRI digests of the P1 clones was particularly interesting. The probe hybridised to four bands in the digest of J1174 and to three bands from clone M0356. Since ACIXSR is derived from an EcoRI/SalI digest of λAC10 (see diagram 5.2) no more than one band should be hybridised by ACIXSR.



(a)



(b)

Diagram 7.5: Fluorescent *in situ* hybridisation analysis of the P1 clones

(a) P1 clone J1174 hybridised to a normal, male metaphase chromosome spread. The probe hybridises to the centromeric region of chromosome 1, 2, 15 and 16.

(b) M0356 hybridises to the centromeres of chromosomes 1, 5 and 19.

In the SpeI and BglII digests the probe ACIXSR hybridised to bands which corresponded to the expected sizes of bands derived from the vector arms. A very low level of hybridisation of the probe to the vector sequence may have been expected since the probe was found to faintly hybridise to the background grid from the filters during the library screening. In addition ACIXSR shows a low level of hybridisation to the DNA of the ladders of size markers. However, hybridisation to the P1 vector arms by ACIXSR does not explain the failure of the P1 clones to hybridise to 7p22 in the FISH experiments. These clones were identified in duplicate on the gridded library filters because they showed a much higher level of hybridisation than the background grid. The apparent size of the clones and the unusual hybridisation pattern from the restriction digest strongly suggested that the P1 clones do not come from the region upstream of PDGFA.

7.4.3 PCR Screening of the P1 clones.

The results described above suggest that the P1 clones M0356 and J1174 do not correspond to the sequence of the upstream region of PDGFA, to confirm this a PCR assay was developed. For this, ACIXSR was partially sequenced using the vector primer Sp6. From this sequence a second primer was synthesised (PACIXSR2) which was used to 'walk' further into the clone. The sequence data is shown in diagram 7.6. This sequence was sufficient to allow two PCR primers to be made: ACIXSR91 and ACIXSR260 to amplify a 189 bp fragment. The PCR reaction conditions are shown in table 2.2.

3′ - GATCCACTGCAGCGCACTCCTGCTGTGCACCCCTGCAGTG

CACCCCAGAACCTAAGCAGGGCCTGGCCGGGCCGCCCACA

ACIXSR91 PACIXSR2

ACC**ACGGAGCTGGGGTGGGTGTG**TTCCCTGCACCCAAGGC

CGGCAGCCAGGCGTGGAGGAGGGGGCGTGTGAGTGAGGGC

ACAGCCTGGTGGGAGCATGTTCCTGCGTTTCCTGTGCCAG

CCGCTGTTCCCAGCCCTCTCCATGGATTAACTCACTCTCT

ACIXSR260

CTATTCCTGATTTCCCTGGAGTGTAGGCA**AATCAGCCCCA**

CTTCACAGCTGAGGACGCTGAGGGCAGTCGTGGTAAGGCC

AGGTCCTGGCAGGCTGGCCCGGAGTGGATCAGCTCCTCTC

CACCTGG - 5′

Diagram 7.6: Sequence data from the ACIXSR probe

Sequence data generated from the 3' end of ACIXSR using the sequencing primer Sp6 (from the vector sequence, not shown) and primer PACIXSR2 (broken underline). The PCR primers ACIXSR91(upstream) and ACIXSR20 (downstream), shown in bold and underlined, were synthesised from this sequence and used to develop a PCR assay.

This PCR assay was used to analyse DNA extracted from 'minipreps' of all of the P1 clone colonies which were hybridised by ACIXSR (see diagram 7.4). As shown in diagram 7.7, PCR products of the expected size (189 bp) were seen only with human gDNA and control plasmids pACCATIX and pACIXHN. These two plasmid are derived from the same bacteriophage clone as ACIXSR. No PCR products can be seen with the two P1 clones DNA samples (diagram 7.8). The failure of the PCR on the P1 clones implies that the sequence identified by the primers is not present in the clone.

7.5 Conclusion and discussion.

The P1 cloning system has been shown in previous mapping projects to be more successful than cosmids in spanning some regions such as rDNA repeat sequences (Francis 1994). In the attempts to map the Di George syndrome region on chromosome 22q, P1 clones were identified from a region which had been found to be unstable in YAC clones (Halford 1993). Unfortunately the P1 clones have not been found to be useful for mapping PDGFA.

The P1 clones which were identified from the library have been fully analysed. The results of the FISH experiments, the restriction digests and the PCR assay all indicate that these clones do not correspond to the 5' region of the PDGFA gene. For the purpose of developing a physical map of the PDGFA locus these P1 clones were not useful and thus were not analysed any further.

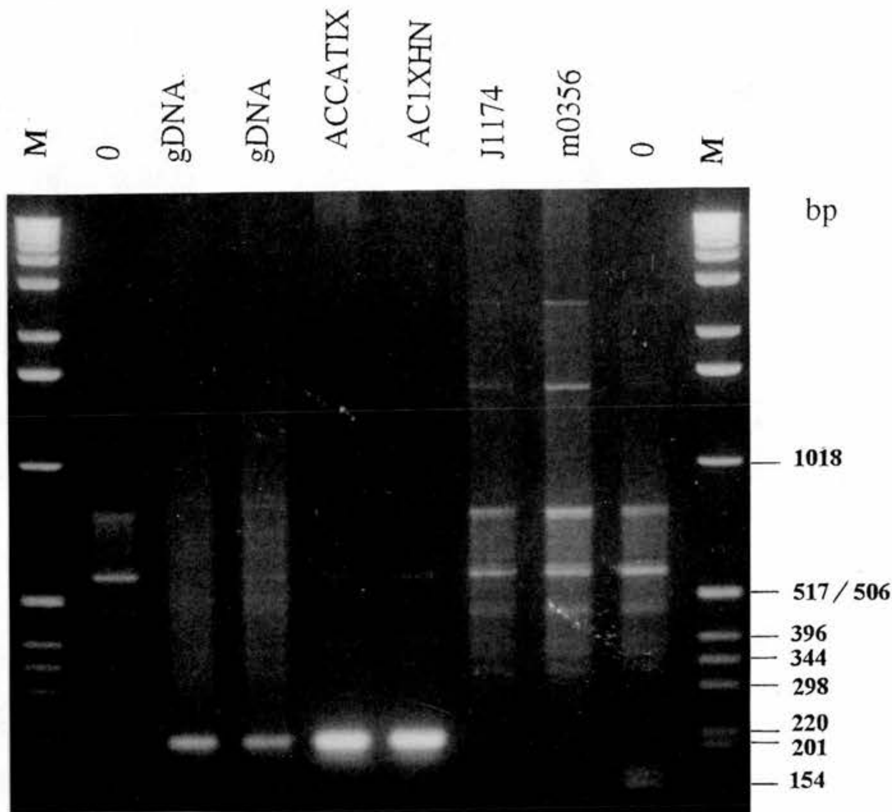


Diagram 7.7: PCR analysis of the P1 clones.

The primer pair ACIXSR91/ACIXSR260 were used to analyse the positive P1 clones identified by the filter hybridisation shown in diagram 7.4. The expected 189 bp PCR product is seen only in the positive controls of human gDNA and the plasmid clones pACCATIX and pACIXHN. No PCR products are seen for either of the P1 clones which were analysed (J1174 and MO356).

This result was surprising since the two clones J1174 and M0356 were identified in duplicate using the two sets of filters. Hybridisation of ACIXSR to the clones was confirmed on receipt of the agar stabs. This suggests that the probe ACIXSR which was used to identify these clone does cross-hybridise specifically with some sequence in the clones although the molecular basis of this cross hybridisation is not known.

The pattern of hybridisation to the chromosome centromeres in the FISH experiments was unexpected since the clones were identified using the single copy probe ACIXSR (see Chapter 5). The chromosome centromeres are known to contain highly repetitive DNA, suggesting that the P1 clones, J1174 and M0356, contain sequences similar to the centromeric repeats. However neither of the two different patterns of hybridisation shown by these clones corresponds to a known distribution of centromeric repeat sequences. It is tempting to speculate as to whether the upstream region of the PDGFA locus may contain similarly repetitive DNA sequences. The presence of such sequences upstream from PDGFA would help to explain the failure to identify YAC clones spanning the gene. However in the absence of any data for the region surrounding the PDGFA gene such suggestions can only be highly speculative.

The presence of a repeat motif in the P1 clones might also explain why these clones had been identified and requested by two other research groups, neither of whom were working on chromosome 7p. It is also possible that these clones had been identified by other groups by coincidence or that more that one clone was present in the DNA at these grid co-ordinates. It may be that the clones contain several repeated sequences or they may contain a centromere repeat sequence which has not been described. It would be interesting to explain the FISH pattern which has been observed, however this avenue of inquiry was not relevant to the aims of this project and thus was not pursued.

8. Chapter 8: Telomere YACs

8.1 Chromosome telomere YAC clones

The linkage analysis described in Chapter 4 has shown PDGFA to have a distal location on chromosome 7p. At present no marker distal to PDGFA has been described and there is no telomere probe available for chromosome 7p. Rough estimates of the distance between PDGFA and the chromosome telomere suggested that the gene may be less than 1 Mb from the chromosome end (discussed in section 4.10). The subtelomeric regions of chromosomes have been found to be underrepresented in YAC clone libraries (Cohen 1993) and this may explain the failure to find positive YAC clones during the screening of the YAC libraries described in Chapter 6.

The candidate chromosome 7 telomere YAC clones were screened for the presence of the PDGFA gene: if PDGFA mapped very close to the telomere the gene itself may have been contained within the YAC clones. The clones were also analysed in an attempt to identify a polymorphic marker which could be used to complete the chromosome 7p linkage map and a single copy DNA probe which could be used to extend the pulsed field restriction map and to screen the YAC library

8.2 Telomere YAC vectors

Cloning chromosome telomeres in YACs uses a modified YAC vector with a single yeast telomere. The telomere YAC vector, pTYAC1, which has been used to develop a telomere YAC library is described in section 1.13.4. This approach is possible because the essential elements required for telomere function are conserved between yeast and humans, allowing the yeast cell to recognise the human DNA as a functional telomere. Whilst the cloning of human telomere-associated DNA using the telomere YAC vector cloning system is conceptually simple, it is difficult to isolate clones which contain single copy DNA from specific individual telomeres.

The DNA adjacent to the telomere consists of long tracts of repetitive sequence which can be present at several different telomere loci (Brown 1990a). The extent of this repetitive telomere associated DNA varies between telomeres. In addition specific individual telomeres show polymorphic variation between individuals (Brown 1990a, Wilkie 1991). However, the development of telomere YAC libraries (Riethman 1989) means that the human chromosome telomeres are now rapidly being cloned and mapped.

8.3 The human telomere YAC Library

A library of potential human telomere YAC clones in the size range 50 -250 kb has been developed. Transformants were screened with the human telomere repeat (TTAGGG)n to identify potential telomere clones (Riethman 1989). Four YAC recombinant clones from this library which have been identified by FISH and PCR as candidates for the telomere of chromosome 7p were sent as a kind gift by H. Riethman. The details of these YACs are listed in table 8.1. Note that y2185 contains two YACs, y2185e is a subclone of the larger YAC in y2185. The largest of these candidate chromosome 7p telomere YACs is 260 kb making these clones relatively small in comparison to conventional YACs. As table 8.1 shows, none of these YACs have, at present, been mapped uniquely to chromosome 7.

YAC clone	Size of YAC Bands	NIGMS chromosome localisation by PCR	FISH localisation
y2178	170 kb	2,7,(5,8-10,13-22)	7p22, 5q53, 3q29, 1p15, 8p23, 1p36, 1q32-41, Yq12, 16q24, 19p13.3, 6q27, 4q28
y2185	255, 145 kb	7 (2)	7p22, 7cen, 7q32-34, 11p15, 10cen, 16p13.3, 16q24, 5q35, 19p13.3, 20p12, 20q13.3, 1p36, 1q41, 1q44, 9p24, 9q34, 4q28, Yq12
y2185e	260 kb	7 (2)	as for clone y2185
y2200	60 kb	3,7,9	7p22, 3q29 (weak signals)

Table 8.1: Candidate telomere YACs for 7p.

The candidate telomere YACs for the chromosome 7p telomere, y2178, y2185, y2185e and y2200 were a kind gift from H. Riethman. YAC clone y2185e is a subclone of clone y2185 containing only the larger of the two YACs.

8.4 Analysis of the telomere YAC clones for PDGFA

8.4.1 PCR analysis of the telomere YACs

A crude DNA extract of the YAC along with the genomic DNA of the yeast host was prepared from each clone (see section 2.4.3). This crude preparation was purified by extraction with phenol and chloroform and reprecipitated in ethanol.

The tYAC clones were analysed with three different PCR assays using the primer pairs ApaL/844 from exon 3, 430/731 from intron 6 and 2260/2790 from intron 1 of PDGFA. The PCR assays using primers ApaL/844 and 430/731 are described in the YAC library screening in Chapter 6 (see diagram 6.1). Primers 2260/2790 span the GC repeat in intron 1 of PDGFA which is described in Chapter 3 (section 3.6.1). The reaction and cycling conditions for all the PCR assays are listed in table 2.2.

None of the three PCR assays which were used to screen the candidate telomere YACs gave a PCR product for any of the four YAC clones which were analysed. As an example, the PCR reactions from the 430/731 primer pair PCR are shown in diagram 8.2. This PCR assay included three different gDNA positive controls (marked '+') all of which gave the expected 331 bp product. No PCR product can be seen in any of the tYAC DNA samples. These three PCR assays span the length of PDGFA and these results strongly suggest that these YAC clones do not contain the gene.

8.4.2 Analysis of the telomere YACs by filter hybridisation

It is possible that the absence of PCR products from the crude yeast gDNA extracts described above was due to PCR failure, thus the clones were also screened by filter hybridisation. Agarose blocks containing each of the four telomere YAC clones

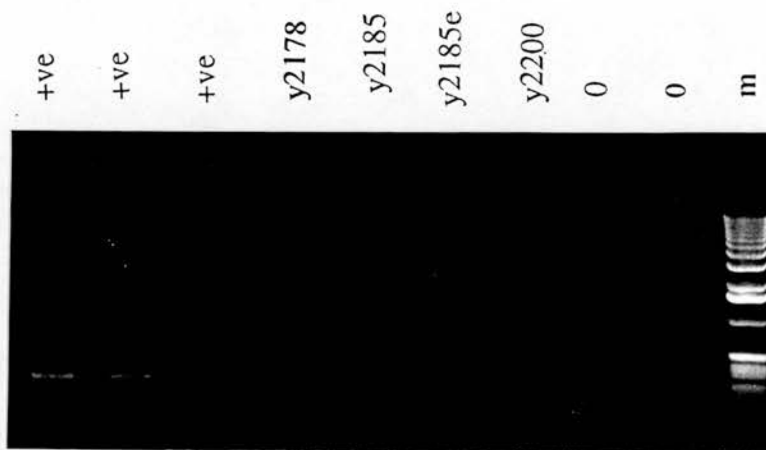


Diagram 8.2: PCR analysis of the telomere YAC clones for PDGFA

PCR analysis using the primer pair 430 and 731 with yeast gDNA from the telomere YAC clones y2178, y2185, y2185e and y2200. The expected 330 bp PCR product is seen in the three human gDNA control samples but not with any of the tYACs.

were prepared as described in section 2.4.7. The yeast chromosomes were separated by pulsed field gel electrophoresis in a 1% agarose gel at 15°C for 22 hours at 200 volts using 30 second pulses. The pulsed-field gel was blotted onto Hybond N^+ membrane and then hybridised with radiolabelled probes.

The filter was hybridised with the probe ACY7 which spans intron 1 to exon 4 of PDGFA. This probe was also used in the pulsed field restriction mapping (see diagram 5.2). Following this, the blot was stripped and reprobed using MS31. The MS31 probe identifies the nearest proximal locus to PDGFA in the linkage analysis described in Chapter 4. The probe was kindly supplied by A. Jeffreys and is described in section 4.9. The telomere YACs were screened for MS31 since it was theoretically possible that MS31 may be present in the absence of PDGFA. Such a situation could arise if the data from the linkage analysis is incorrect or if the telomere YAC is rearranged and does not represent the DNA sequence at the 7p telomere despite being derived from that region. Finally, the blot was stripped and probed with radiolabelled Cot1 DNA. Cot1 DNA will hybridise to human repeat DNA sequences and confirm that the YAC clones contain human DNA.

Neither of the probes ACY7 or MS31 hybridised to any of the four telomere YACs indicating that neither locus was present in the clones. Diagram 8.3 shows a pulsed-field gel and the results of probing the blot with ACY7 and Cot1. The Cot1 DNA hybridised to all four tYACs confirming the presence of human repeat sequences in all of the clones. Hybridisation of the Cot1 DNA to the filter was conducted last, after the same filter had been probed with both ACY7 and MS31. Since the DNA could have failed to transfer onto the membrane during the Southern blotting, or could have been removed when the blots were stripped, this also confirmed the presence of the tYAC DNA on the Hybond N^+ membrane.

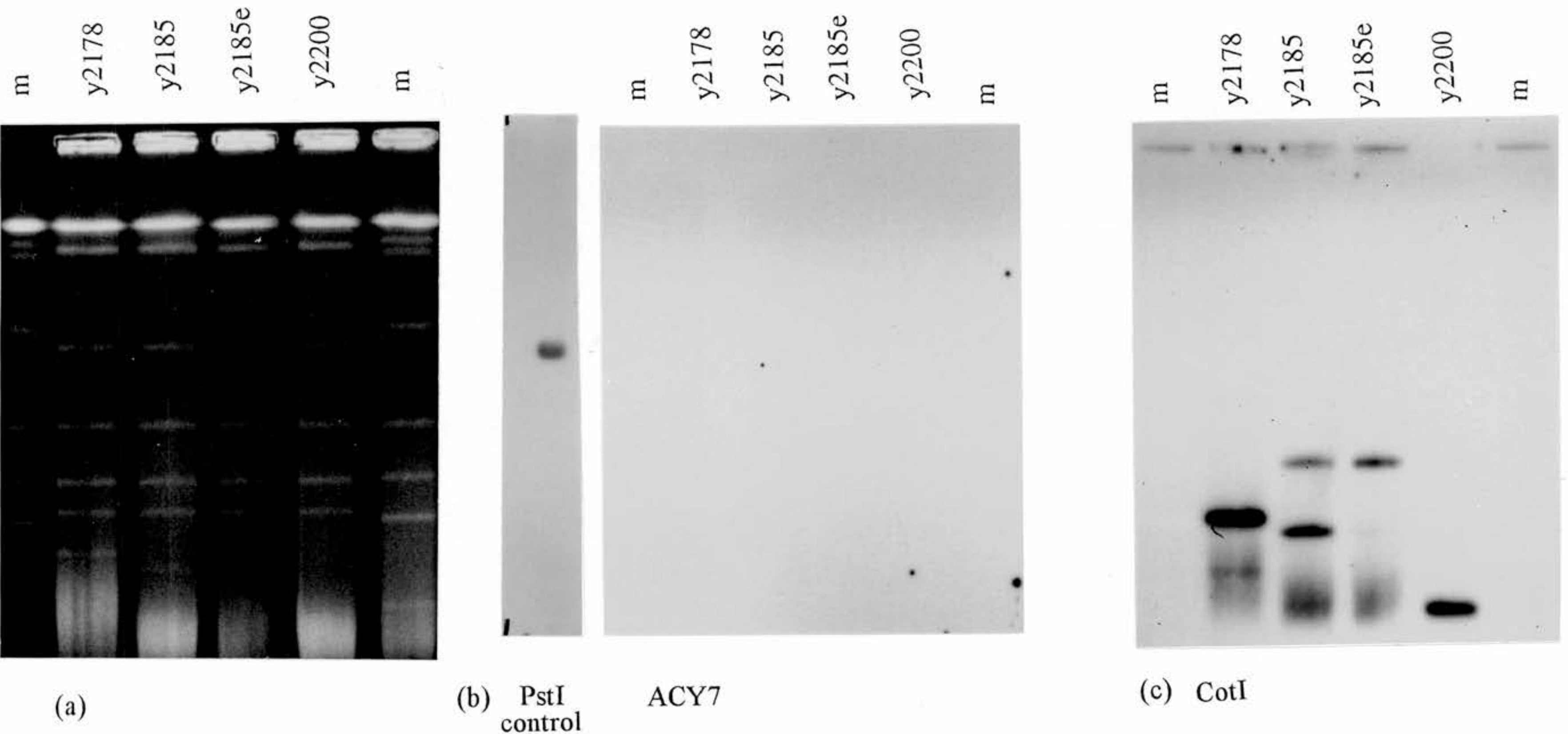


Diagram 8.3: Filter hybridisation analysis of the telomere YAC libraries

(a) Pulsed-field gel electrophoresis of the YAC clones. (b) Probing a Southern blot of the pulsed-field gel with ACY7. A PstI digest of gDNA was included in the hybridisation as a positve control. (c) Subsequent hybridisation of CotI DNA to the same blot. CotI DNA hybridises to human DNA repeat sequences confirming that the YAC clones do contain a human DNA insert.

8.5 Screening the whole telomere YAC library

While the whole telomere YAC library is not made widely available, it has been screened by PCR for the presence of PDGFA. The three PCR assays 631/631, ApaL/844 and 430/731 were used by E Green (NIH, Bethesda, USA) to screen this library at the same time as he screened the other YAC libraries described in Chapter 6. No positive clones were identified although this is not surprising given that the most likely clones, the candidates for the 7p telomere, have been analysed more fully as described above.

8.6 Further analysis of the telomere YACs

The results described above clearly show that PDGFA is not present in the telomere YAC clones which are candidates for the 7p telomere. This is not entirely surprising since the largest of these clones is 260 kb, the gene would have to be very close to the telomere to be cloned in these YACs. Further analysis of the telomere YACs was conducted since a single copy probe or a polymorphic marker derived from the telomere would be very useful in extending the genetic linkage map and to continue to attempt to build a physical map of the region.

A single copy probe would be useful to extend the pulsed-field restriction maps of the region and, if PDGFA and the telomere are in close proximity, the two loci may map to the same restriction fragment. Such a probe would also be useful for screening the YAC and P1 libraries. While such clones would not be expected to contain PDGFA since the libraries have already been screened for the gene, YAC clones from the subtelomeric region would be extremely useful for developing a physical map of the region.

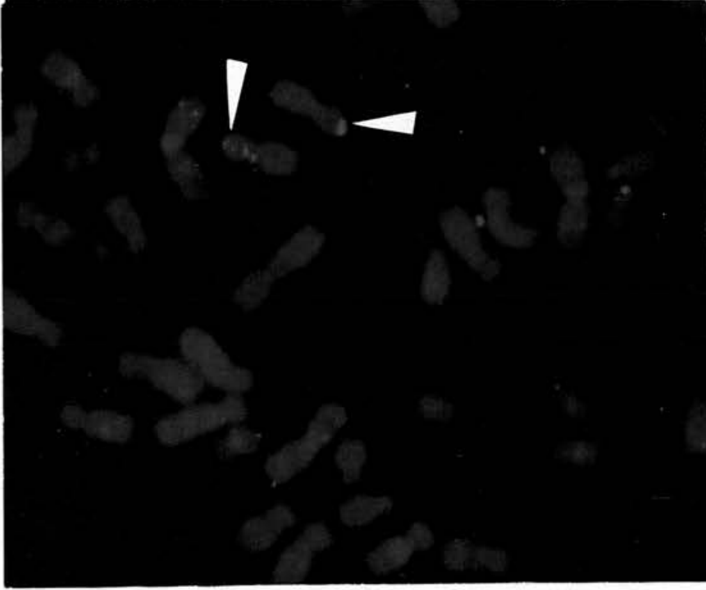
A polymorphic locus for the region would allow the linkage map of chromosome 7p to be completed by providing a telomere marker. Such a marker would allow an estimation of the distance between PDGFA and the telomere although the genetic distance may not correspond to the physical distance since the telomeric regions of chromosomes are known to have high recombination rates.

8.6.1 Fluorescent *in situ* hybridisation

To confirm that these telomere YAC clones do hybridise to the chromosome 7p telomere, fluorescent *in situ* hybridisation experiments were conducted. For these a substantial amount of the YAC DNA (about 1 μg) was required. To isolate the YAC bands from the genomic DNA of the yeast host cell, the whole chromosomes were separated by pulsed field electrophoresis. The bands corresponding to the YACs were cut from the agarose gel and the DNA was extracted using the 'Geneclean' method. Fragments of random DNA sequence were amplified from the purified YAC DNA by PCR. This PCR reaction was conducted using a method and oligonucleotide primers from J. Warner, it is described in section 2.11.1. The PCR products obtained were of random size producing a smear of DNA sized 220 bp to 1 kb when separated in an agarose gel. About 1 μg of this randomly sized DNA was biotin labelled and used for fluorescent *in situ* hybridisation to metaphase chromosome spreads from a normal male.

The biotin labelled DNA fragments from each of the telomere YACs was found to hybridise to large number of both interstitial and telomeric loci. Amongst this, hybridisation to chromosome 7p telomere was confirmed for all four of the candidate telomere YACs. As an example, diagram 8.4 shows the PCR products from y2185 (a), and y2185e (b). The probe hybridises to the chromosome 7p telomere (marked with an arrow) in addition to many other loci.

(a)



(b)

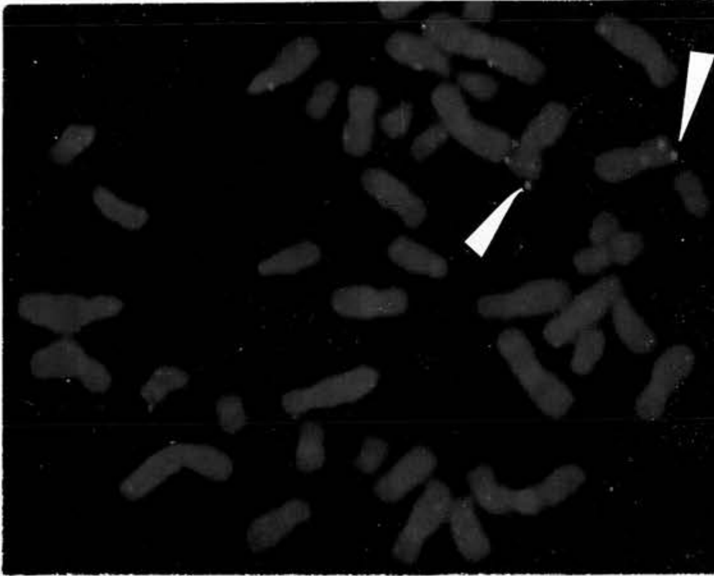


Diagram 8.4: Fluorescent *in situ* hybridisation using y2185 and y2185e
FISH analysis using biotin labeled, randomly amplified DNA from (a) tYAC y2185 and (b) tYAC y2185e. Hybridisation to chromosome 7p (marked with the arrows) can be seen, as well as hybridisation to a large number of other sites including other telomere loci. The chromosome identification and analysis was done by J. Fantes (MRC Human Genetics Unit, Edinurgh).

This pattern of hybridisation of the YAC DNA to many different loci across the genome was expected given the repetitive nature of telomeric DNA sequences. It is consistent with the possibility that these YAC clones are derived from the 7p telomere.

8.6.2 Screening for CA repeats in the YAC clones

CA dinucleotide repeats are frequently the source of highly polymorphic loci. The randomly amplified DNA from the YAC clones was screened for the presence of such repeats using a $(GT)_{15}$ oligonucleotide primer. The random PCR products which were generated for the FISH experiments (described above in section 8.6.1) were separated in an agarose gel to check their average size. This agarose gel was blotted using Hybond N^+ and then probed using a $(GT)_{15}$ oligonucleotide primer.

Hybridisation to the YAC DNA was very weak which suggests that there are few CA repeat sequences in the YAC clones. Thus efforts were directed towards isolating end sequence from the YACs rather than searching for a polymorphic CA repeat.

8.7 Isolation of endclones from YACs

The isolation and analysis of YAC clones is restricted by the large size and low copy number of the vector. A number of methods based on the use of PCR have been developed to isolate and characterise the clones. Several of these methods, such as Alu-Alu PCR and Alu-vector PCR depend on the presence of Alu sequences within the DNA insert at a suitable distance from each other or from the vector/insert boundary to allow amplification by PCR. Other approaches such as plasmid rescue

and vectorette PCR depend on the presence of specific restriction sites within a suitable distance from the vector/insert boundary.

For the analysis of the candidate telomere YACs none of these methods are ideal since they are all based on the assumption that the sequence of the cloned DNA and the distribution of features such as Alu repeats and restriction sites is random. This is not generally the situation in the telomeric regions of chromosomes. However, since the insert end of the telomere YAC clones will be furthest away from the true telomere repeats and closer to landmarks of interest on chromosome 7p such as PDGFA, it was decided to attempt to clone the ends of the telomere YACs. Unlike conventional YACs, each telomere YAC contains only one arm and only one insert end can be cloned for each.

Two approaches, those of plasmid rescue and vectorette PCR were attempted. Of these approaches the plasmid rescue attempts were not successful and are not described, the end-cloning using vectorette PCR is described below.

8.7.1 Vectorette PCR end cloning

Vectorette end-cloning uses PCR to amplify the end fragment of the human DNA insert in the YAC clones (Riley 1990). The YAC clone is digested with a restriction enzyme and a common linker is ligated to the DNA fragments. Terminal sequences are then amplified using a vector specific primer and a linker specific primer. The linker, or vectorette cassette which is ligated to the digest DNA contains an area of mismatch. The linker specific primer matches the lower sequence of the cassette. It cannot therefore be used to seed the PCR reaction until its complementary strand has been synthesised from the vector primer. This ensures that PCR products can only be amplified from DNA molecules which have both the vector and cassette sequences.

Total yeast DNA, including the telomere YAC, was digested with the enzymes Sau3AI leaving the DNA fragment with a GATC overhang or with RsaI leaving a blunt ended fragment. The digested DNA was purified and then ligated with the vectorette cassette, with a GATC overhang or blunt ended as appropriate. The primer sequences and PCR conditions are described in section 2.12. The resultant PCR products were checked by digestion with EcoRI since only genuine YAC end-clones should contain an EcoRI site at the junction between the vector and the cloned insert. PCR products which were shown to cut with EcoRI were cloned into the vector PCR-Script using the Stratagene PCR cloning kit. The resultant bacterial colonies were screened directly using the primer pair T3 and T7 as described in section 2.11.

This approach generated end-clones for all four of the telomere YACs. The PCR products for y2178 (1C) were amplified from the Sau3A1 digests, while the RsaI digests led to PCR amplified products from y2185 (6B), y2185e (7A) and y2200 (8C). As shown in diagram 8.5, the PCR product from y2185e (7A) does not cut with EcoRI. It cannot bridge the vector/insert junction and thus was not analysed further. Further attempts to generate a subclone from y2185e were not successful.

PCR products were also obtained from Sau3A1 digests of y2185 and y2200. The fragments produced from these samples appeared to be smaller than those shown in diagram 8.5, thus they were not investigated. Unfortunately, the PCR products produced for all three of the successfully cloned YACs are very small, containing only about 70 bp of the human insert DNA. This is not surprising since the method has only been successful using the frequently cutting restriction enzymes. Other, less frequently cutting, restriction enzymes such as BamHI, BglII and BclI were also used to digest the yeast/YAC DNA in attempts to endclone the YACs. None of these digests produced PCR fragments.

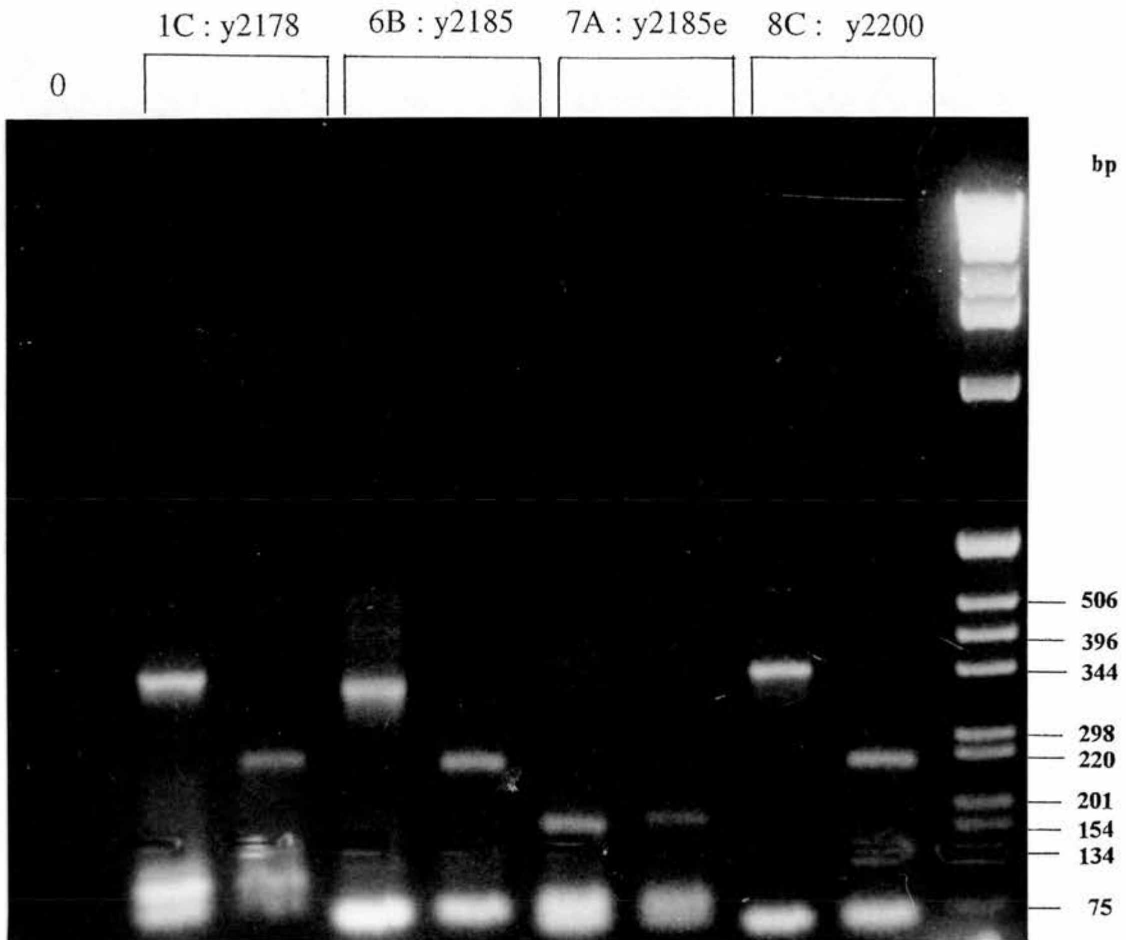


Diagram 8.5: The vectorette PCR products digested with EcoRI
The vectorette PCR products amplified from YACs y2178, 2185, 2185e and 2200. The samples are undigested in the first lane of each pair and digested with EcoRI in the second lane of each pair. Samples 1C (y21780), 6B (y2185) and 8C (y2185e) all cut with EcoRI confirming that they are derived from the vector/insert boundary. Sample 7A (y2185e) does not cut with EcoRI and was not analysed any further.

The PCR products 1C, 6B and 8C were sequenced using the primers T3 and T7. The end clone sequences are shown in diagram 8.6. Sequencing these cloned PCR products confirmed that they are derived from the vector/insert boundary of the YAC clones. However, as expected from the size of the PCR products, the length of the inserts which have been subcloned are very short. This made the development of PCR assays difficult.

8.7.2 Development of PCR assays for the YAC endclones

The sequence derived from the vectorette end-cloning of the telomere YACs was used to scan the GenEmbl databases: no sequences with significant homology were identified. However, searching the GDB database found primer sequences for PCR 'STSs' for the end sequence from two of the YAC clones: y2178 and y2185. These had been submitted by H. Riethman at the same time as the end-clones presented above were being sequenced for this project. One of these primers, sv-1-2178F matches the sequence in end clone 1.2 (see diagram 8.6) which was also derived from y2178.

The primers for these two PCR assays were made and tested on the YAC clones. Primer pair sv-1-2178F and sv-1-2178R are derived from y2178 while sv-1-2185F and sv-1-2185R are from y2185. The sequence for the primers and the conditions used for the PCR reactions are described in table 2.2.

End clone 1.2 (y2178)

PBR-4177

TTGAAGCATTTATCAGGGTTATTGTCTCATGAGCCGGATACATATTTGAATGTA
TTTAGAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATC

EcoRI

ACGAGGCCCTTTCGTCTTCAA**GAATTCAAGGAAAAGAAGTTAAATAATTTAAAA**

D-580

GAGTACGTTTCGGGTACAGCTATTTAGCACTAGACAGTAAGATTACAGCGATTC
TCGTACGAACGGTTACGATTCG

End clone 6.25 (y2185)

PBR-4177

TTGAAGCATTTATCAGGGTTATTGTCTCATGAGCCGGATACATATTTGAATGTA
TTTAGAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATC

EcoRI

ACGAGGCCCTTTCGTCTTAA**GAATTCATGTCTGTCTGAGCCTAAGCCTCTGAGC**

D-580

ATTCTCTGGAGCGCCATGCTGTGTGGACAGGCTGTGACCAAGGAGAGGACAGCG
ATTCTCGTACGAACGGTTACGATTCG

End clone 8.15 (y2200)

PBR-4177

TTGAAGCATTTATCAGGGTTATTGTCTCATGAGCCGGATACATATTTGAATGTA
TTTAGAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATC

EcoRI

ACGAGGCCCTTTCGTCTTAA**GAATTCACCCTTCCTCTGCCTTTTTGTTCTATCT**
AGGCCCTCAGCCTTTGGGGAGTGGATTTTCCCACTCACTCCCCTGATCCAAGGA

D-580

GAGGACAGCGATTCTCGTACGAACGGTTACGATTCG

Diagram 8.6: Sequence from the Vectorette endclones of the tYACs
Endclones from the tYACs y2178, y2185 and y2200. The size of each of these clones is very small containing only about 65 to 90 bp of the YAC insert DNA(in bold). The YAC vector primer sequences PBR-4177 and the vectorette primer sequence D-580 are marked.

PCR analysis showed that both PCR assays, using primer pairs sv-2178F/R and sv-1-2185F/R, amplified products from all four tYAC clones. These PCR assays were used to analyse a panel of chromosome hybrid cells containing fragments of human chromosomes 7 and 2. The PCR products generated using primer pair sv-1-2178F/R are shown in diagram 8.7. The amplified PCR product is seen with all four tYACs and with gDNA. Hamster and mouse gDNA controls do not amplify confirming that the products seen from the cell hybrid lines are derived from human DNA. All three samples from chromosome 2 and the majority of the panel of chromosome 7 hybrids amplify with this PCR assay. One sample, corresponding to 7 cent-qter, has not amplified. This is thought to be a false negative, there was no sample left in the eppendorf tube which was rinsed with 5 μl of dH_2O to attempt this PCR reaction. The regions of chromosomes 2 and 7 contained within each hybrid sample are indicated on the diagram. Samples containing only chromosome 2 or 7q clearly amplify confirming that this PCR assay is not unique for the 7p telomere.

Thus these PCR assays do not identify unique sites. This means that the endclones of these telomere YACs do not contain unique or single copy DNA. Further efforts to isolate single copy sequence which could be used as a 7p telomere STS or probe will require cloning methods which search the whole cloned insert rather than just its end.

8.8 Discussion and conclusion

Screening both the individual candidate telomere YACs for chromosome 7p and the whole telomere YAC library (Riethman 1989) has found that this library does not contain clones spanning PDGFA. The absence of PDGFA in the candidate chromosome 7p telomere YACs is not surprising since the largest of these was 260 kb. The RARE cleavage experiments presented in Chapter 9 measure the distance between PDGFA and the telomere of chromosome 7p.

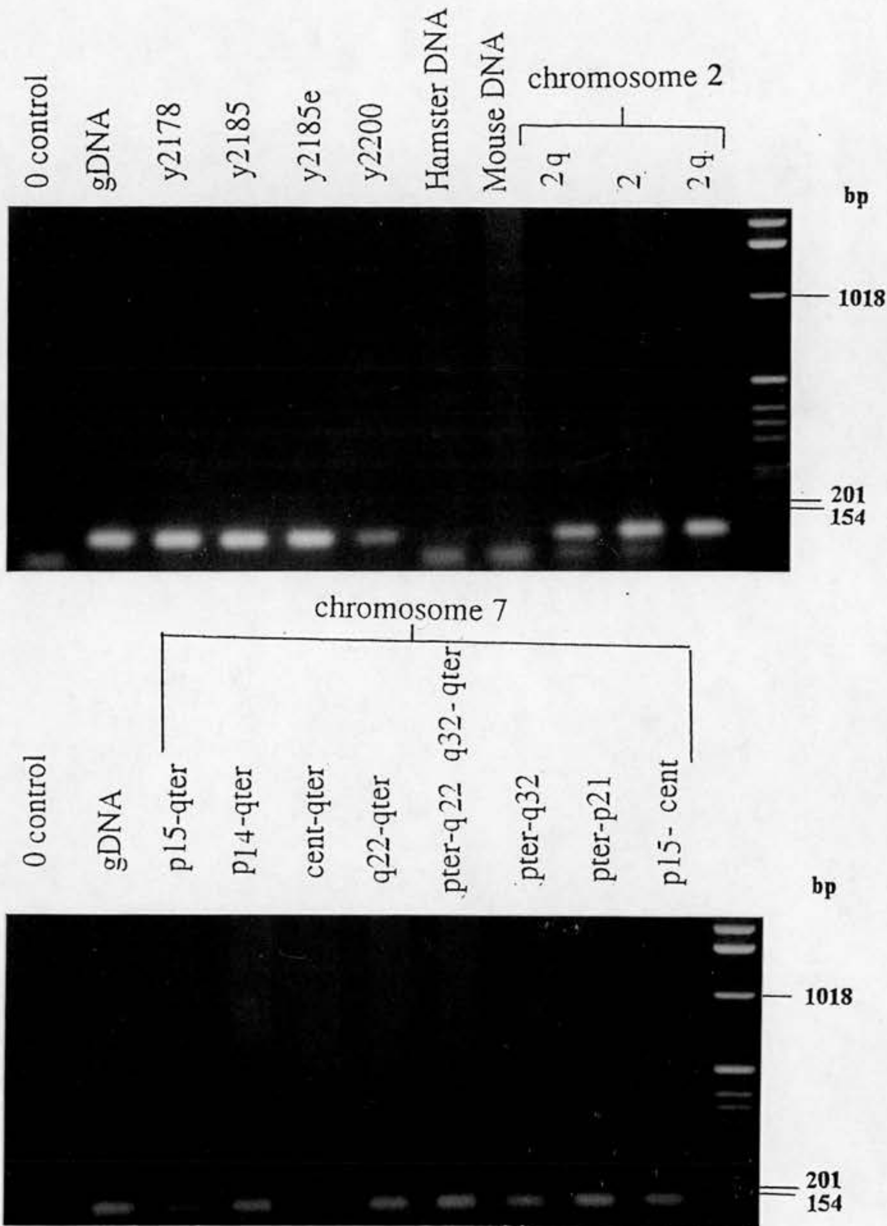


Diagram 8.7: PCR analysis of somatic cell hybrids using primers sv-1-2178F/R PCR analysis of gDNA, the tYACs and somatic hybrid cell lines containing fragments from human chromosomes 2 and 7. Mouse and hamster gDNA did not amplify confirming that the PCR product is specific to human DNA. The 98 bp PCR product is seen in cell lines which contain DNA from chromosomes 2 and 7q only, thus this PCR assay does not identify a single copy sequence from chromosome 7p and is not a true 'STS'. The absence of a PCR product in sample 7cen -qter is thought to be a false negative arising from a failed PCR reaction (see section 8.7.2).

FISH analysis of random DNA amplified from these candidate 7p telomere YAC clones has confirmed that the clones do hybridise to the telomere of chromosome 7p. However, as shown in diagram 8.4, these clones also hybridise to a number of other chromosomes telomeres as well as to other interstitial loci. This is not surprising since it is known that the repetitive DNA in the subtelomeric region can be present at several different telomeric loci.

End-cloning the DNA inserts by Vectorette PCR yielded very small fragments. H. Riethman has had more success using plasmid rescue to clone the ends of these YACs and has developed PCR assays for two of the clones. Unfortunately these PCRs do not identify unique sites on chromosome 7p and are therefore not useful STSs. While the end of these telomere YAC clones are not unique sequences it is possible that the clones do contain 7p unique sequence within their length.

Further analysis of these clones has not been undertaken since they are also being analysed by H. Riethman. The cloning of other chromosome telomeres using this telomere YAC library have been reported recently (Riethman 1994) and it seems likely that a similar report for chromosome 7p is simply a matter of time. In addition a second large group (Helms 1993) have also indicated that they have candidate clones for 7p. However neither of these two groups has reported any substantial data for the chromosome 7p telomere. The availability of unique DNA sequence in the form of an STS or a probe would be very valuable to continue the attempts to map PDGFA. Of most value would be a polymorphic marker for the 7p telomere which could be used to complete the linkage map.

The main aim of this project was to map PDGFA rather than the chromosome 7p telomere. Given the time constraints that are inevitable during any project, along with the knowledge that at least two large groups are attempting to clone the 7p telomere it was decided not to pursue this avenue of inquiry. Other approaches towards mapping PDGFA were sought. Given that the direct mapping

approaches of genetic mapping by linkage and physical mapping through building clone contigs had been exhausted a less direct approach was required. For this the technique of RecA restriction endonuclease (RARE) cleavage mapping was used to measure the distance between PDGFA and the chromosome 7p telomere. This approach had been used to measure distances near to the chromosome 4p telomere during the hunt for the Huntington's chorea gene. The use of the RARE cleavage mapping technique is described in the next chapter.

9. Chapter 9: RecA-assisted restriction endonuclease (RARE) cleavage.

9.1 Introduction

The preceding chapters in this thesis describe the screening of YAC, P1 and telomere YAC libraries in an attempt to find clones spanning PDGFA. Without clones from the region it is not possible to develop clone contigs to build a physical map of the region. Given the absence of such clones, an alternative approach is required.

The most obvious approach is to determine the physical distance between the PDGFA gene and other landmarks on chromosome 7p. Attempts have already been made, in the pulsed field mapping project in Chapter 5, to measure the physical distance between PDGFA and the linked marker MS31 (*D7S21*). None of the other loci used in the linkage mapping are suitable candidates to be physically mapped to PDGFA. However, the evidence from the linkage analysis described in Chapter 4 suggests that PDGFA is in a sub-telomeric location on chromosome 7p. No markers have been mapped distal to PDGFA, thus the telomere itself is the most obvious landmark to which PDGFA could be mapped.

As discussed in section 4.10, the distance between PDGFA and the telomere was not easy to determine from the linkage data. However estimates had suggested that the distance was likely to be between 1 Mb to 5 Mb, with the possibility that it was less than 1 Mb. Given that the distance from the gene to the telomere could be relatively small, the technique of RecA-assisted restriction endonuclease (RARE) cleavage was used. The technique and its application to the mapping of PDGFA is described below.

9.2 RecA assisted restriction endonuclease (RARE) cleavage

RecA assisted restriction endonuclease (RARE) cleavage is used to perform sequence-specific cleavage of genomic DNA. It was described by Ferrin (1991, 1993) who used RecA protein to bind a large (60 base) primer to a known EcoRI restriction site. RecA binds the primer to the homologous sequence in duplex DNA creating a triple helix, this protects the restriction site from methylation which can then be used to block all other EcoRI sites on the chromosome. Subsequent removal of the primer and restriction with EcoRI should generate a fragment of DNA from the telomere to the targeted restriction site which can be identified by PFGE and Southern blotting. This procedure is illustrated in diagram 9.1. Using the original protocol, attempts to cleave fragments greater than a few hundred kilobases were unsuccessful (Ferrin 1991). However, a modification to this protocol which raises the reaction pH from 7.5 to 7.85 was found to reduce non-specific cleavage and this modified protocol has been used to generate fragments as large as 1.7 Mb (Ferrin 1993, 1994).

9.3 Application of the RARE cleavage technique

To apply the RARE technique to mapping the distance between PDGFA and the chromosome 7p telomere there were a number of requirements. Firstly a suitable EcoRI site was required along with the sequence of the flanking DNA to allow a large 60 base primer to be designed. An EcoRI site in intron 4 was selected, the sequencing of this region is described below in section 9.4.

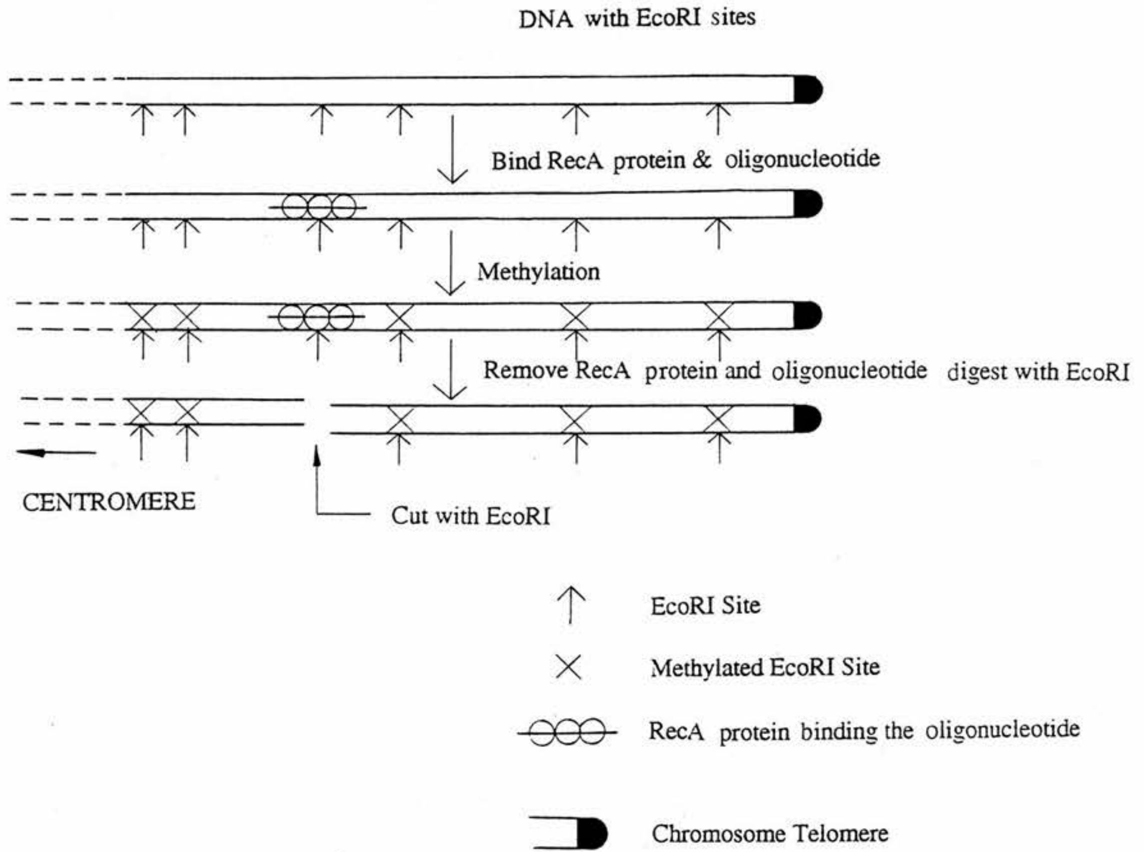


Diagram 9.1: RecA-assisted restriction endonuclease (RARE) cleavage
Schematic illustration of RecA-assisted restriction endonuclease (RARE) cleavage, adapted from Ferrin (1993) to show the application of the technique to identifying telomere fragments.

Since the orientation of the PDGFA gene on chromosome 7p was not known, it was important to have two probes flanking the selected EcoRI site. It was not known which probe would identify the telomere fragment. This requirement for two single copy probes on either side of the selected EcoRI site restricted the choice of restriction sites. The EcoRI site in intron 4 is flanked by the probes ACY7 and AC41HR which are shown in diagram 5.1. Since both of these probes had been used before they were known to be suitable single copy DNA sequences.

Obviously this technique is based on the use of PFGE techniques which are described in Chapter 2, pulsed field mapping had already been used to develop a long range restriction map around PDGFA, this is described in Chapter 5.

9.4 A suitable EcoRI site in intron 4 of PDGFA

As mentioned above, the EcoRI restriction site in intron 4 of PDGFA was selected (see diagram 5.1). It lies 1.1 kb downstream from exon 4, beyond the minisatellite which flanks exon and intron 4 and is shown in diagram 5.1. Sequence data was available for the first 750 bp of intron 4, this sequence was used to synthesise the first of the two primers (P844rev and P950) which were used to extend the sequence downstream to beyond the EcoRI site (shown in diagram 9.2).

A 60 base primer (RARE1) which contained the EcoRI site at its centre was synthesised, its sequence is marked on diagram 9.2. The primer was purified to remove debris and incompletely synthesised primers by separation in a polyacrylamide gel (see section 2.13).

SalI Exon 4

GTCGACCCCACGTCCGCCAACTTCCTGATCTGGCCCCCGTGCGTGGAGGTGAAACGCTGCACCGGCTGCT 70

GCAACACGAGCAGTGTCAAGTGCCAGC*CCTCCCGCGTCCACCACCGCAGCGTCAAGGTGAGCCCTCCCCT* 140

CCCCTCCCCTCCCCTCCCCTCCCCTCCTCTCCTCTCCCCTCCCCTCCTCTCCTCTCCCCTCCCCTCCTCT 210

CCCCTCCCCTCCCCTCCCCACCCCAAGCCCGGGGCCGCCCCTACCAAGCAGCCAGGGTCCCTCCTGCATC 280

CACCACCACAGCGTCAAGGTGAGCCCTGCCCTCCCCTCCTCTCCCCCTCCCTCCCCTCCCCTCCCCACCC 350

GAGCCCAAGGGCCGCCGCGCCGCCCCTACCTAGCAGCCAGGGTCCCCTCCCGGGTCCACCACCGCAGCGT 420

CAAGGTGAGCCCTGCCCCTCCCCTCCCCTCCCCACCCGAGCCCGGGGCCGCCCCTACCAAGCAGCCAGGG 490

TCCCTCACGCATCCAACACTGCAGCATCCAGGTGAGCCCTGCCCCACGCAGGCCTGAGGGTCGCCCCTAC 560

ApaL

*TGGGTTGCCAGGGTCCCTCC*TGGGAGGCCACAAGGTAGCCCCTGGTGTGGCTGGTGGG**GTGCAC**GGCTG 630

CTGCTCCCATCTCCCTGGCAGGTACACTGAGGCCCAGGGAGGAGAAGGAGCACCTGGGGGGGCGGGAAAG 700

P844rev

GCAGGGCCGGGCGGGGCTTCTGCTCTGCTGCTCTGGCGACGAGCCTGGCCTCCTTTCCACGCCCTAGACC 770

ACAGTGAGGGGCCTTGCGAGGTGTCCAGCTGCTTAGCCGGAAGCTGTCTGCGTCCCCGGGCCTGGAGGGT 840

CCTGTGCGGGCTCTTCAGTGGGGGCCGAGGTGCATGGCAGGAGGAGTAGGGCTGGATAGGCAGAGCCTGC 910

ATTGCCTTTTGAAGTGCCTGACCTTTGGAGGGCCCCCCGGCTTCCCCCCTAAAGCAGCCCAGGAGAAGG 980

P950

AGAAGGAGTGCTTGGGGTGGGCTCGAGGCCAGACAGCAGACTTCCTTCCGTAGAATCCCAGGGCGAGCCC 1050

RARE1 EcoRI

CTCTTTCTTTAAGTTGGGGGTGGTGCCTGCCTTCTTCAGGCTGCTTTAGCA**GAATTC**CATCCACTAGGAG 1120

CTGATTAGCAACAAATGTATTTCTCAAAGTTCTGGAGGCTGGAAGTCCAAGATCAGGGCGCCGGCAGACT 1190

GTGTCTCCGCTGCAGCCGTGTTTTCTGGGTTCATAGATGACACCTTCTCGCTCTGCCATCAGTGTGAGG 1258

Diagram 9.2: Sequence data from intron 4

This 1.26 kb of sequence spans from the SalI site in exon 4 (boxed) to beyond the EcoRI site in position 1100 (marked in bold). It was extended from the available sequence using primers P844rev and P950. The minisatellite in intron 4 is shown underlined in italics, between positions 98 and 580. The ApaL site at position 620 (in bold) marks the intron 4 dimorphism which was used in the linkage analysis in Chapter 4. The 60 base oligonucleotide (RARE1) which spans the EcoRI site and was used in the RARE cleavage experiments is indicated (double underlined).

9.5 The RARE experiments

In preparation for the RARE experiments, a control experiment was conducted to measure the amount of methylase required to methylate the DNA samples. The methylation step is critical in this procedure and it is often difficult to achieve adequate methylation of the entire DNA sample. The genomic DNA samples were methylated using 0 to 800 units of EcoRI methylase by incubation at 37°C for 1 hour. The samples were then equilibrated in the appropriate buffer and digested with 100 units of EcoRI using the standard conditions described in section 2.5.1. The samples were separated in a pulsed field gel using 10 second pulses and are shown in diagram 9.3. As this gel shows, maximal methylation is seen with 600 units of methylase, thus in the subsequent experiments 600 units of EcoRI methylase were used.

The methods used for the RARE cleavage are described in Chapter 2, section 2.13. In the early experiments three samples with 360, 680 and 1200 ng of oligonucleotide primer were used since the primer concentration had been found to be critical (Ferrin 1993), however the modified pH of the buffer (Ferrin 1994) seems to have removed this problem. In the experiments conducted for this project, the primer concentration was not found to be a critical factor and only one concentration of primer (680 ng) was used in the later experiments.

In each experiment a number of control reactions were run. These included a 0 control from which EcoRI methylase and EcoRI were omitted, a methylation control from which RecA protein and the oligonucleotide primer had been omitted and an EcoRI digest control from which the RecA protein, oligonucleotide primer and the EcoRI methylase had been omitted. These are shown on the experiments illustrated in diagrams 9.4 to 9.6. The challenging nature of this technique meant that these controls were found to be essential. The appearance of the control samples in the resolved pulsed field gel were a good indicator as to whether the reactions had been satisfactory before the gel was blotted and probed. Of critical importance were the

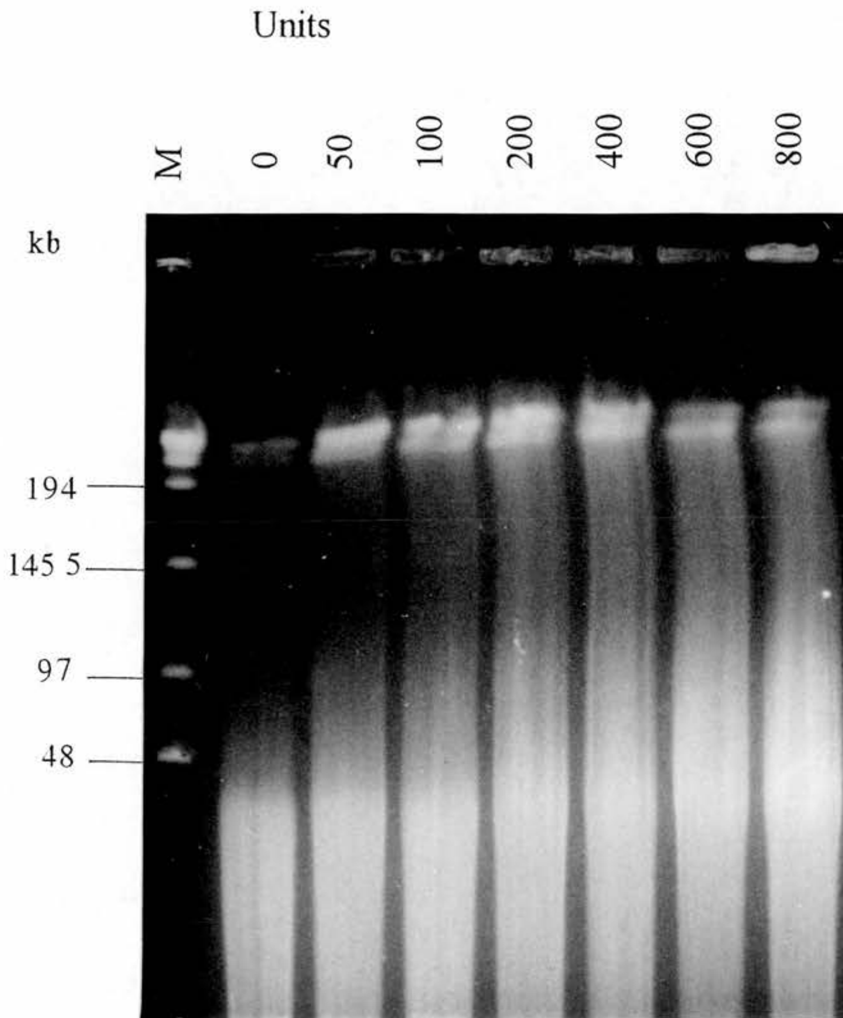


Diagram 9.3: EcoRI methylation of genomic DNA

EcoRI digests of gDNA samples which had been methylated with 0 to 800 units of EcoRI methylase. Maximal methylation is seen with 600 units of methylase, thus 600 units were used in the subsequent RARE cleavage experiments.

methylation and the EcoRI digestion. Washing and equilibration of the agarose samples in the appropriate buffers had to be rigorous to ensure that the EcoRI digestion of the DNA sample would be complete.

9.6 Results of the RARE cleavage experiments

Selecting appropriate conditions for the pulsed field gel separation of the RARE treated samples was not straight forward. The expected size of the PDGFA/ telomere fragment was not known. In addition, the orientation of the gene on chromosome 7 was not known, thus it was not possible to predict which of the two probes, ACY7 and AC41HR, would hybridise to that fragment. To overcome this problem the gels were run with two sets of samples, after blotting the membrane was cut in two and the two halves were probed separately with the two selected flanking probes ACY7 and AC41HR.

Diagram 9.4 shows a pulsed-field gel which was produced using ramped pulses of 15 to 60 seconds, it has been divided and hybridised with both probes. This diagram shows two bands sized about 630 to 679 kb in the left half of the gel which has been probed with AC41HR. The right hand of the gel which has been probed with ACY7, does not show any bands. No bands are hybridised in the control sample lanes. The two fragments hybridised in this gel are close to the limiting mobility band, this region of pulsed field gels is not well resolved, so the bands cannot be sized accurately.

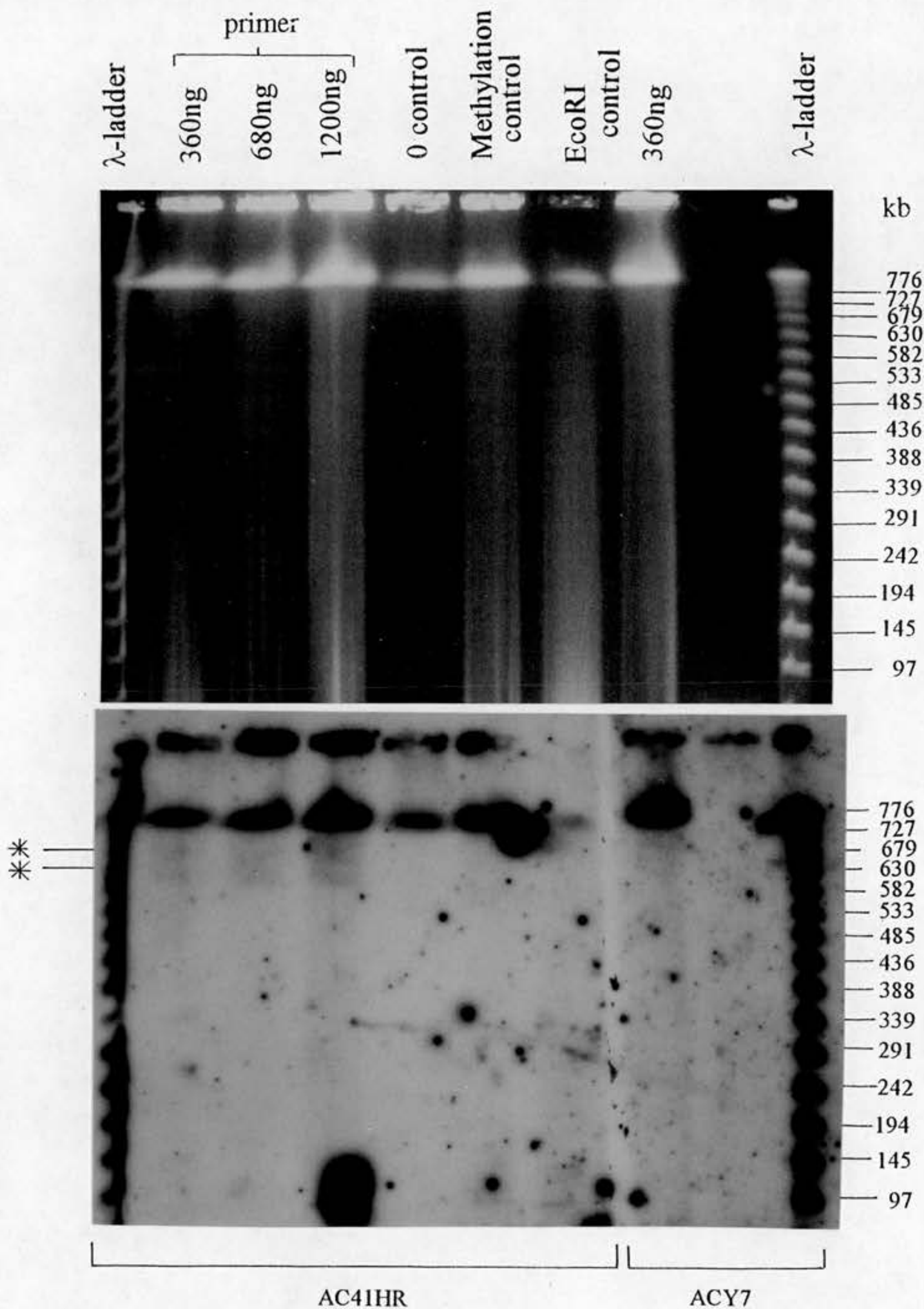


Diagram 9.4: RARE cleavage experiment 1

(a) The RARE samples were separated using pulse times of 15 to 60 seconds. (b) Probing the left half of the gel with AC41HR identified two fragments of 630 and 679 kb (marked '*'). In the right half of the gel, probed with ACY7, no bands are identified. The autorad was exposed to the hybridised blot for 21 days.

The experiment was repeated using ramped pulse times of 45 to 120 seconds to optimise the separation of fragments in the 500 kb to 1 Mb size range. This second experiment, probed with both ACY7 and AC41HR is shown in diagram 9.5. As this diagram shows, AC41HR hybridises two bands sized about 630 kb and 679 kb in both hybridisations. Following this hybridisation, the blots were stripped and reprobed with the alternative probe to the previous hybridisation. On hybridisation to the right hand side of the blot AC41HR identified the same two bands. These fragments are not hybridised in the control sample lanes or with the probe ACY7.

Thus these experiments indicate that the PDGFA gene is about 650 kb from the chromosome 7p telomere and that the gene is positioned on the chromosome with its 5' end centromeric and its 3' end telomeric.

9.7 Polymorphic length variation in the sub-telomeric region

As shown in diagrams 9.4 and 9.5 the probe AC41HR hybridises to two bands. All of these gDNA samples are derived from the same individual (a normal male). This pattern of hybridisation suggests that the sub-telomeric region of chromosome 7p shows variation in its length. A polymorphic length variation has been described for several other chromosomes telomeres (discussed below in section 9.8) and it would not be surprising to find a similar length variation in the chromosome 7p telomere.

To determine whether the chromosome 7p telomere does show a polymorphic variation in length, fresh blood samples were collected from a further four unrelated individuals and used to make gDNA agarose blocks (as described in Chapter 2).

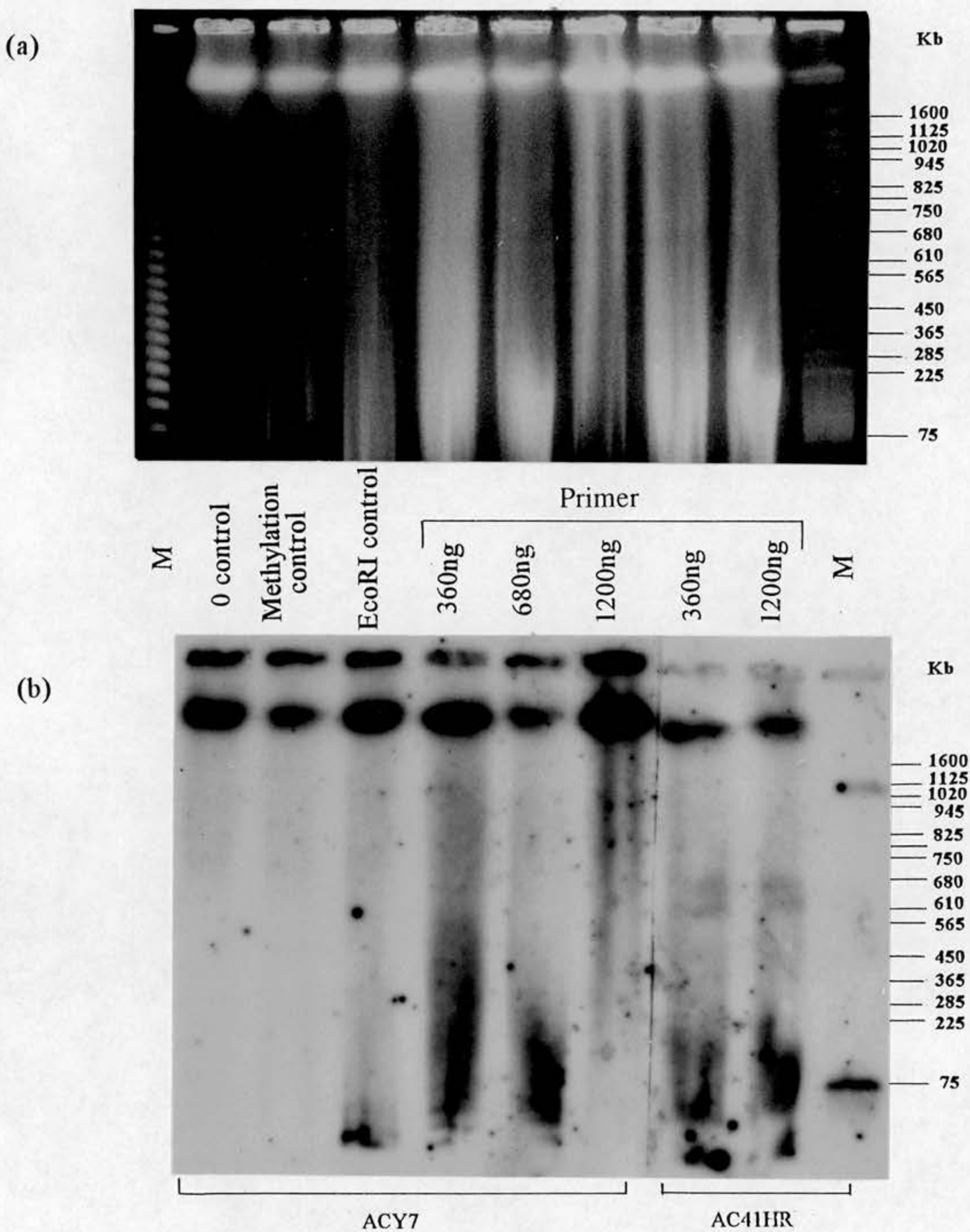


Diagram 9.5: RARE cleavage experiment 2

(a) PFGE separation of RARE samples using ramped pulse times of 45 to 120 seconds, (b) hybridised with ACY7 (left side) and AC41HR (right side). ACIXSR hybridises bands of 630 kb and 679 kb, autorad exposed to blot for 21 days.

The RARE experiments were repeated using these additional samples. For these experiments, only one primer concentration of 360 ng of RARE1 was used. The filters from the gels were probed with AC41HR only. The hybridised blot was exposed to a phosphoimager cassette for 28 days and then scanned to produce the image shown in diagram 9.6.

The results of this RARE cleavage experiment suggest than individuals SI, LL and JW all carry two different sized telomere fragments. The lower band is also present in individual LS, unfortunately the quality of the image makes it difficult to determine whether LS has a second, large band. No bands can be seen for individual DC. This absence of bands for individual DC seems likely to be due to imaging process since the sample in the agarose gel looks adequately methylated.

Thus the telomere region of chromosome 7p may show length variation similar to that described for other chromosome telomeres. With the data available it is not possible to confirm this or to estimate the number and size of different lengths or the level of heterozygosity. The difficulty in resolving small size differences between such large fragments (650 kb) makes it difficult to determine whether the bands hybridised in this experiment are the same size. Clearly further experimental data would be required to resolve these questions, this is discussed further below in section 9.8.

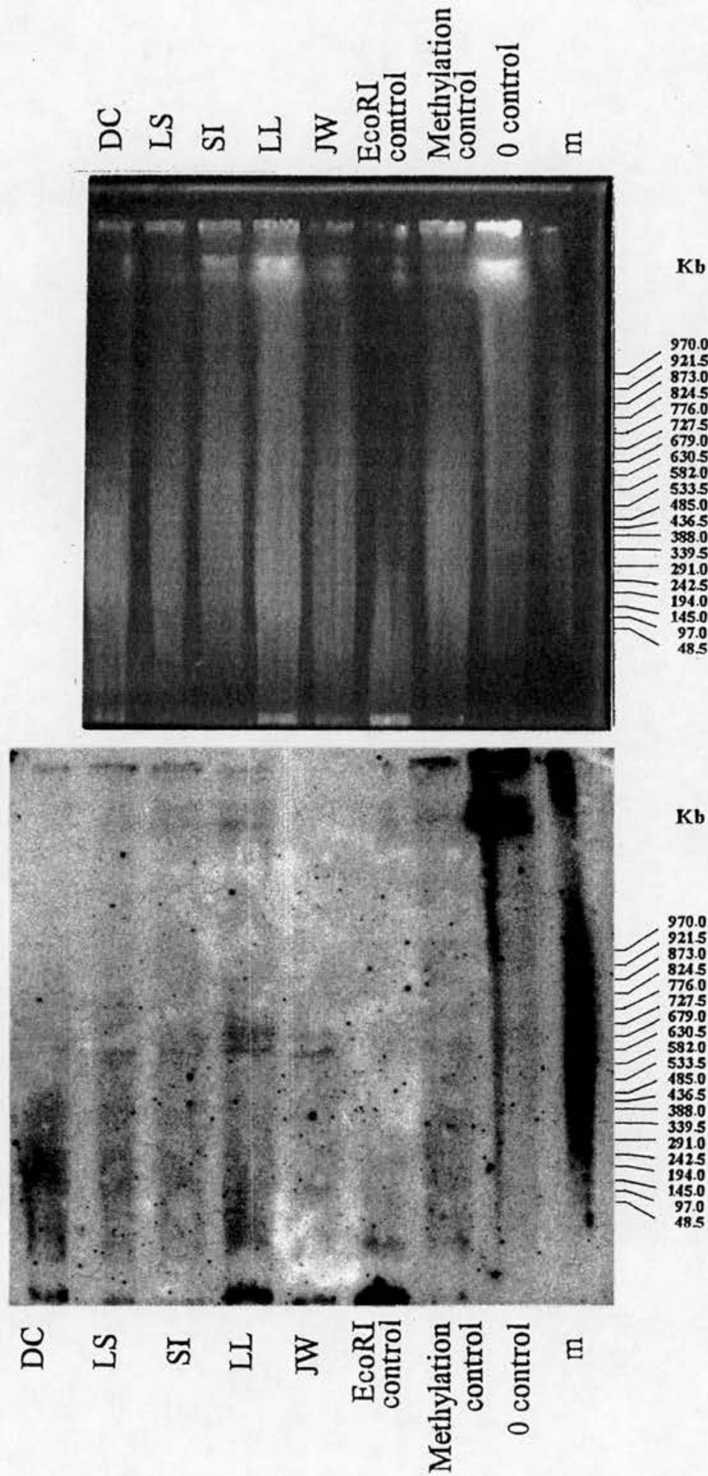


Diagram 9.6: RARE experiment 3

(a) PFGE separation using ramped pulses of 45 to 120 seconds of gDNA samples from 5 unrelated individuals, (b) hybridised with AC41HR. The results are discussed in section 9.7. The phosphoimager cassette was exposed to the blot for 28 days.

9.8 Discussion

These RARE experiments identify discrete DNA fragments in the genomic DNA samples when probed with AC41HR, placing PDGFA about 650 kb from the chromosome 7p telomere. Thus PDGFA is located less than 1 Mb from the telomere, confirming the estimations of this distance which were discussed in Chapter 4 (section 4.10) following the results of the linkage analysis. The implications of a location so close to the chromosome telomere for a growth factor gene such as PDGFA are discussed in the conclusion to this thesis in Chapter 10.

The PDGFA/telomere fragment is hybridised by the downstream probe AC41HR, thus the gene is positioned with its 5' end centromeric and its 3' end telomeric. This means that the coding region of the gene separates its control sequences from the telomere of the chromosome. Thus telomeric deletions of chromosome 7 would not be expected to delete the promoter region of the gene without also removing the coding region of the gene. This is in contrast to the orientation of the α-globin gene which has been shown to be very close to the telomere on chromosome 16p (Wilkie 1990a). However it is not known what effect telomeric deletions of chromosome 7p have on the expression of PDGFA.

This orientation of PDGFA means the 5' end of the gene lies closest to the proximally linked locus MS31 (*D7S21*). Thus, only the upstream probe ACIXSR could have been useful in the attempts to measure the physical distance between PDGFA and MS31 in the pulsed field restriction mapping experiments described in Chapter 5. Given the failure to establish physical linkage between PDGFA and MS31 in the pulsed-field gel experiments in Chapter 5, it would be interesting to use the RARE cleavage technique to measure the distance between these two loci.

The final experiments (shown in diagram 9.6) hybridised different sized PDGFA/telomere fragments in at least three different individuals. This size variation may be due to the subtelomeric repeat sequences. These lie proximal to the simple tandem repeats and are highly variable in length and degree of repetitiveness. These sub-telomeric repeat sequences may be present on several different chromosomes but entirely absent from others (Brown 1990a, deLange 1990, IJdo 1992). These regions show variation in length and in the type of repeats which are present at the chromosome telomere between different individuals, making the subtelomeric region highly polymorphic.

Polymorphic length variation has been described for several of the chromosome telomeres which have been cloned and characterised including 7q (Hing 1993, Dietz-Band 1990), 2q and 8p (Macina 1994). The most striking example is that of chromosome 16p where the α-globin gene has been has been shown to be 170 , 350 or 430 kb from the chromosome 16p telomere (Wilkie 1991). The two most common alleles, the 170 kb and 350 kb sizes have been found to contain different repeats with no sequence homology between the two. Thus the two chromosome 16p telomeres carried by one individual may be completely different and may be more similar to the telomeres of other chromosomes. The results of the experiments presented above show that a similar length variation is present on chromosome 7p, it would be interesting to investigate this further.

It important to acknowledge that these experiments do not provide direct evidence that the hybridised DNA fragments do correspond to a true telomere fragment. However the results presented above have been shown to be consistent both within the same individual and between unrelated individuals. No bands were seen with the probe ACY7 or in any of the control sample lanes, thus the DNA fragments which are hybridised are specific to the probe AC41HR, require the presence of the RecA bound oligonucleotide primer and methylation of the DNA. All of these factors strongly suggest that the hybridised band does correspond to a PDGFA/telomere DNA fragment.

Alternative explanations for the observed pattern of hybridisation can be theorised. One possible alternative explanation would be a duplication of the sequence around the protected EcoRI site. If the sequence duplication was located about 650 kb downstream from PDGFA and positioned in the opposite orientation to the sequence within the gene then this pattern of hybridisation would be seen. However, such an explanation seems unlikely given the additional evidence for a telomeric location for PDGFA provided by other experiments such as the linkage analysis described in Chapter 4.

In the absence of a unique telomere probe for chromosome 7p, there is no easy way to confirm whether this hybridised fragment does correspond to the telomere. The telomere probe TelBam11 does hybridise to the chromosome 7p telomere, however it is not unique and also hybridises to several other chromosome telomeres (Brown 1990a). Such a probe would be expected to hybridise to the 650 kb band seen in the experiments presented above, however it would be expected to hybridise to many other telomere fragments as well. This would be expected to produce a high level of background.

The autorads which are shown above were exposed to the hybridised filters for at least 21 nights at - 70 °C. The signals are relatively weak and an increase in the level of background above that which is seen in these exposures was expected to obliterate the specific signals which can be seen. Thus these filters were not hybridised with non-specific telomere probes. Since at least two large research groups are attempting to clone the chromosome 7p telomere it seems likely that single copy probes from the region will soon be available (see section 8.8). When a unique probe for the chromosome 7p telomere does become available it would be interesting to reprobe these RARE experiments to confirm that the hybridised fragments are true telomere fragments.

10. Chapter 10: Conclusion and discussion.

10.1 A telomeric location for PDGFA

The introduction to this thesis proposed two possible locations for *PDGFA*, confining the gene to two relatively small cytogenetic regions: (i) between the breakpoint in 0044 Rag 1-15 and the proximal breakpoint in DSE or (ii) between the distal breakpoint in DSE and the chromosome 7p telomere. These two possible locations for the gene are illustrated in diagram 1.3. The mapping of PDGFA to about 650 kb from the chromosome 7p telomere means that the deletion in patient DSE cannot be a true terminal deletion and that the second of the two possible locations for the gene has been found to be correct. A telomeric location for a gene, especially a growth factor such as PDGFA is interesting. Thus, the major features of the telomeric regions of chromosomes, the consequences for physical mapping and the implications of such a location on the expression and function of PDGFA are discussed below.

10.2 The telomere sequence and structure

Human chromosomes end with tandem repeats of the sequence (TTAGGG)n. The sequence of these tandem repeats are well conserved through evolution with the G rich strand corresponding to the 3' end of the chromosome. These repeats at the telomere end overcome the problems with replication of the far end of the DNA strand. DNA polymerases require an RNA primer and synthesise DNA in the 5' to 3' direction. Thus a short length of DNA equivalent to the size of the RNA primer is lost from the chromosome end during replication. The use of the specialised DNA polymerase, telomerase, which carries a (TTAGGG)n RNA template allows this sequence to be replaced. These (TTAGGG)n repeats alone appear to be sufficient for telomere function. Individuals with terminal deletions of chromosome 16p have been described in whom the broken chromosome had been 'healed' by the direct

addition of the (TTAGGG)n repeats to the breakpoint site (Wilkie 1990, Flint 1994). These 'healed' chromosomes function normally.

The telomere ends appear to give stability to the chromosome: double strand breaks in the chromosome or loss of all telomere repeats cause the chromosomes to become unstable. As a consequence of this, a large number of chromosome breakage-reunion events are seen which lead to the development of chromosome rearrangements and ultimately cell death. The progressive loss of telomere repeats in somatic cell lines is thought to be related to the cellular ageing process, this is discussed further in section 10.5.

Proximal to the simple tandem repeats described above is a more complex region of repetitive DNA sequences. These subtelomeric repeats which are highly variable in length and degree of repetitiveness are described in section 9.8. The RARE experiments which are described in Chapter 9 show that the telomeric region of chromosome 7 shows length variation similar to that described for other chromosome telomeres. The role of this sub-telomeric repeat region is not clear, it may simply act as a buffer between the telomere and essential DNA sequences protecting them against telomere loss. These sequences may be involved in processes such as chromosome pairing in meiosis. However, it is also possible that these sequences have no direct role and have arisen as a consequence of recombination or translocations between non-homologous chromosome telomeres (reviewed in Kipling 1995).

10.2.1 Minisatellite sequences in subtelomeric regions

The sub-telomeric regions of chromosomes have been shown to contain a large number of minisatellite repeat elements (Royle 1988, Hing 1993) the reason for this clustering of sequences is not known. It is possible that this clustering of

minisatellite sequences may explain the increase in recombination which has been described in the telomeric regions of the chromosomes. However it is also possible that these minisatellite sequences have arisen as a consequence of that increase in recombination.

It is interesting to have found two minisatellite sequences within the PDGFA gene. As described in section 3.10, these two repeat elements are separated by only 1 kb, however their sequence and structure indicate that they are not related. Interestingly the two other pairs of closely physically linked minisatellite sequences which are described in Chapter 3 were also found in subtelomeric regions. It is not clear whether or not these closely linked minisatellites have arisen by coincidence.

10.3 Development of maps of telomeric regions

10.3.1 Linkage analysis

The linkage data which is shown in Chapter 4 gave unexpectedly large values for the genetic distances between PDGFA and the nearest proximal loci. The finding that the PDGFA gene is very close to the chromosome telomere made this result less surprising. An increase in recombination has been observed in the telomeric regions of other chromosomes (for example Burmeister 1991, Helms 1992). This increase in recombination towards the telomere, which is greater in male meioses then female, is thought to be due to 'hot-spots' of high recombination frequency rather than a generalised increase. For example, comparing the genetic and physical maps for the long arm of chromosome 21 found that one region (21q22.3) which accounted for 40% of the length of the genetic map corresponded to only 10 % of the cytogenetic length of the chromosome, indicating a small region of dramatically increased recombination (reviewed in Kipling 1995).

10.3.2 Long range restriction maps

Generation of a single copy telomere probe for chromosome 7p would allow the extension of the pulsed field restriction map described in Chapter 5. Extending the restriction map to the telomere should be possible given that the region is relatively small. Such information may help to identify CpG islands and indicate whether there are any other genes distal to PDGFA. However similar restriction mapping in the pseudoautosomal region of the X and Y chromosomes found a large number of rare cutter restriction sites in the terminal region of the chromosome. These sites were particularly common in the last 200 kb of the chromosome (Petit 1988). If a large number of sites are also present in the telomere region of chromosome 7p further probes may be required. This cluster of restriction sites described in the telomere of the pseudoautosomal region were spread randomly and did not appear to identify CpG islands. It is possible that this pattern of restriction sites is related to the function of the telomere.

10.3.3 YAC clone contig maps

The comparison of the linkage, YAC and NotI restriction maps for chromosome 21 identified a number of discrepancies most importantly the order of some loci were different (Chumakov 1992, Ichikawa 1993). Comparison of the YAC clone contigs to the NotI restriction map identified the majority of the discrepancies;while the linkage and NotI restriction maps were in much closer agreement. In particular, discrepancies were identified around the telomeric regions of chromosome 21 which seems to be relatively unstable in clones (Tanzi 1988). Cloning the region around PDGFA will depend on the identification of telomeric clones. The evidence from other chromosomes suggests that this region will be problematic and that alternative approaches to YAC cloning would be advised, this is discussed further below.

10.4 Representation of loci in YAC clone libraries

The subtelomeric regions of chromosomes are known to be underrepresented in YAC libraries. The majority of the YAC clone libraries are made from partial EcoRI digests of the target DNA which is cloned into an EcoRI site in the YAC vector. Since only some EcoRI sites will be cut, the chance of cloning a individual fragment will reduce as it gets nearer to the telomere and there are fewer distal EcoRI sites. Obviously the requirement for flanking restriction sites means that standard YAC cloning vector systems cannot clone the true telomere end of the chromosome.

Alternatively, the region surrounding PDGFA may be unstable or unclonable in YACs. As discussed in section 6.5, regions which contain repetitive DNA sequences are thought to be particularly unstable in YAC clones. Since the sub-telomeric regions of chromosomes are known to contain long tract of repeats this may explain the failure to identify YAC clones spanning the gene.

With the data available it is not possible to determine whether the PDGFA region was not identified in the YAC clone libraries through chance, a systematic selection against the region as a result of the way in which the libraries are constructed or because the region is unclonable in YACs. If PDGFA was not in the libraries which were screened due to a combination of decreased likelihood and chance, screening of further YAC libraries may eventually identify clones spanning PDGFA. However, the screening of the YAC libraries has been exhaustive and it seems reasonable to conclude that future attempts to clone the region around PDGFA should use alternative approaches.

10.4.1 Development of chromosome telomere clones

The telomeres of chromosomes are cloned using a modified or 'half-YAC' vector and a telomere YAC library is described in Chapter 8. The development of telomere YAC s quickly allowed the cloning of the 7q telomere (Riethman 1993) and the pseudoautosomal X/Y telomere (Brown 1990b). The cloning of these chromosome telomeres proved to be particularly easy because they carry largely unique DNA sequences. The cloning of other chromosome telomeres is complicated by the presence of the sub-telomeric repeats. These make it very difficult to identify unique sequence and mean that the telomere YAC clones need to be relatively large to overlap with conventionally cloned DNA. Despite these problems, to date a large number of human chromosome telomeres have been cloned and characterised. These include 1q (Negorev 1994), 2q (Macina 1994), 4p (Youngman 1992, Bates 1990), 4q (Weber 1990), 7q (Riethman 1993), 9q (Guerrini 1990) 16p (Wilkie 1991), 21q (Reston 1995) and the psuedoautosomal telomere (Brown 1990b).

Following the results of the linkage analysis and the screening of the standard YAC clone libraries, candidate chromosome 7p telomere YAC clones were screened for PDGFA (see Chapter 8). The largest of these YACs was 260 kb while the distance between PDGFA and the telomere of chromosome 7p subsequently shown to be about 650 kb, thus it is not surprising that these telomere clones did not include PDGFA. Analysis of the telomere YACs was then concentrated on identifying single copy DNA which could be used as a telomere probe. This analysis was frustrated by the difficulty in isolating single copy DNA sequence, a consequence of the sequence structure of the subtelomeric regions. The problems associated with identifying unique sequence from the telomere YAC clones are explained well in Negorev (1994). However, since several groups are working on this problem, it seems likely that the 7p telomere will eventually be cloned.

10.5 The implications of a telomere location for PDGFA

10.5.1 Chromosome telomeres in malignancy

Expression of PDGF is frequently seen in mesenchymal tumours, and this may result in important autocrine stimulation of tumour growth. However, the molecular mechanisms underlying aberrant growth factor synthesis by tumours are largely unknown. One phenomenon which accompanies tumour cell proliferation is reduction in length of the terminal telomere repeat. Loss of all telomere repeats results in instability of the chromosome ends and chromosome breakage-reunion events (Hastie 1989). This may explain the increased number of chromosome rearrangements seen in malignant cells. Such subtelomeric rearrangements can have dramatic effects on expression of nearby genes; for example truncation of the 16p telomere can silence expression from the nearby α-globin cluster (Wilkie 1990b). Characterisation of these 16p deletions has shown that expression of the α-globin gene was silenced even when the gene itself was intact. It has been speculated that the deletion affected control sequences related to the gene.

Because of these phenomena, a sub-telomeric location for PDGFA is of great interest. The orientation of the PDGFA with its 5' or promoter region centromeric would mean that telomere deletions would not be expected to delete just the control sequences of the gene.

10.5.2 Telomere length and ageing

The length of the telomere repeat sequences has been shown to be greater in sperm than in somatic tissues such as fibroblasts or lymphocytes suggesting that somatic tissue cells lose their telomere DNA sequences. This loss of telomere sequences is progressive and has been shown to be related to donor age (Hastie 1990). In culture, the telomere length of fibroblasts decreases with each cell passage by up to 50 bp per generation and up to 2 kb are lost before cell death. In the last few

doublings of senescent fibroblasts in culture dramatic increase in the number of chromosome-chromosome fusions is seen (reviewed in Greider 1990, Harley 1991, 1991). It has been suggested that this progressive shortening of the chromosome telomeres is related to the phenomenon of cellular ageing or senescence. However, it is important to note that while telomere loss is implicated in the process of cellular senescence of the cell this does not explain its significance to the ageing of the organism. The progressive loss of telomere repeats may have interesting consequences for the expression of sub-telomeric gene such as PDGFA which is involved in the control of cell growth.

10.5.3 Ring Chromosome 7 and PDGFA

A subtelomeric location for PDGFA makes the abnormalities seen in ring chromosome 7 interesting. As discussed in section 1.9, the two constant findings amongst patients carrying a ring chromosome 7 were growth retardation and skin lesions which may predispose to malignancy.

It has been suggested that the clinical manifestation of ring chromosome syndromes may not be related to loss of DNA sequence. In one study of three patients with ring chromosome syndrome, telomere repeats were identified at the fusion site by *in situ* hybridisation, no deletions of DNA could be detected (Pezzelo 1993). High resolution cytogenetic analysis of a fourth patient with ring chromosome syndrome also failed to identify any deletion of DNA (Sawyer 1993). Thus it has been proposed that the clinical problems may be caused by instability of the ring chromosome during somatic cell growth leading to chromosome rearrangements and aneuploidy in some cells in the somatic tissues. However the clinical symptoms of the described patients with ring chromosome 7 include specific abnormalities in addition to the general symptoms associated with ring chromosome such as growth retardation. These specific abnormalities, which are described in section 1.10.4,

included skin lesions bearing some similarity to the *patch* mutant mouse (see section 1.7.1). In one patient, a boy who died at 20 months, gross abnormalities in development of the brain were described. PDGFA is known to be expressed by neurones and is thought to be involved in the development of the brain. It seems likely that these patients with ring chromosome 7 may have deletions of PDGFA. It would be interesting to conduct molecular studies to establish whether the gene has been deleted.

10.6 Future work

This mapping project has used the standard approaches of linkage analysis, long range restriction mapping and screening YAC libraries in addition to the technique of RARE cleavage. To develop further information about the region surrounding PDGFA, a major goal must be to clone the chromosome telomere itself, as described in Chapter 8, several large research groups are working towards this. The development of a single copy probe and a polymorphic marker would allow the completion of the linkage and pulsed field restriction maps.

In the absence of clones for the chromosome 7p telomere, an interesting alternative approach to extending the pulsed field mapping of the telomere of a chromosome would be to use a telomere probe which is not unique. Such a probe could be used to probe restriction digests of DNA from a human-hamster hybrid cell line which contains chromosome 7p as its only human contribution. This approach has been used by Burmeister (1991) who used a non-unique subtelomeric fragment from the chromosome 21q telomere to complete their PFGE restriction map. The probe TelBam11 which has been mapped to the 7p telomere in addition to several other chromosomes (Brown 1990a) or an end-clone from the candidate telomere YACs described in Chapter 8 could be suitable candidates for probes. Such an approach would depend on a suitable somatic hybrid cell line being available for culture. As

discussed in Chapter 5, the methylation pattern of DNA derived from cultured cells would be altered. Thus the restriction map analysis of PDGFA would have to be repeated to using the DNA from the cultured cell lines. While the distance of 650 kb between the chromosome telomere and *PDGFA* is relatively small, if a large number of sites are present this approach may fail to link the two loci.

Ultimately the completion of a mapping study of this region depends on the development of a clone contig which spans from the chromosome telomere, across *PDGFA* to proximal loci such as MS31 (*D7S21*). The PDGFA gene itself has already been found to be easily clonable in the bacteriophage vector (Bonthron 1988) and it may be that the region could be mapped using bacteriophage and cosmid vectors. Using cosmid 'walking' to map the region would be laborious and slow given the small size of the cloned inserts in these vectors (<50 kb). However, the inevitable future development of YAC clone contigs for the rest of chromosome 7p may mean that such an approach would only be required to bridge a small gap. Unfortunately the majority of chromosome 7p has not yet been mapped in YAC clone contigs (see diagram 1.10) and this approach could not be undertaken at present.

The large size of inserts which can be carried in YAC clones along with the ease of cloning some regions which have been found to be difficult to clone in *E.Coli* based systems has meant that YACs have become very widely used as the system of choice for long range mapping projects. However regions such as the PDGFA gene which seem to be unclonable in the YAC vector will require alternative approaches. A number of alternative cloning systems to YACs are currently being developed including BACs (bacterial artificial chromosome) and MACs (mammalian artificial chromosomes). It may be that these systems, especially MACs which would be maintained in mammalian cell lines, will be found to allow the cloning of DNA sequences which cannot be maintained in yeast cells. Unfortunately alternative vector cloning systems such as these are still under development and no libraries of clones using these vectors are available.

References

Allitto,BA; MacDonald,ME; Bucan,M; Richards,J; Romano,D; Whaley,WL; Falcone,B; Ianazzi,J; Wexler,NS; Wasmuth,JJ; Collins,FS; Lehrach,H; Haines,JL; Gusella,JF (1991): Increased recombination adjacent to the Huntington disease-linked D4S10 marker. Genomics. 9, 104-112.

Anand,R (Ed.) (1992): Techniques for the analysis of complex genomes. 1st ed. Academic Press, London. 238 pages.

Anand,R; Riley,JH; Butler,JC; Markham,AF (1990): A 3.5 genome equivalent multi acess YAC library: construction, characterisation, screening and storage. Nuc. Acid. Res. 18(8), 1951-1956.

Antequera,F; Boyes,J; Bird,AP (1990): High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. Cell. 62, 503-514.

Antonaides,HN; Galanopoulos,T; Neville-Golden,J; Kirtsy,CP; Lynch,SE (1991): Injury induces in vivo expression of platelet-derived growth factor (PDGF) and PDGF receptor mRNAs in skin epithelial cells and PDGF mRNA in connective tissue fibroblasts. Proc. Natl. Acad. Sci, USA. 88, 565-569.

Antoniades,HN; Scher,CD; Stiles,CD (1979): Purification of human platelet-derived growth factor. Proc. Natl. Acad. Sci, USA. 76(4), 1809-1813.

Badbanchi,F; Otto,M; Kohlhauer,U; Grzeschik,N; Beck,S; Smith,SJL; Swallow,DM; Kooy,RF; Vergnaud,G; Aldred,M; Grzeschik,K-H (1994): The EUROGEM map of human chromosome 7. Eur J Hum Genet. 2(3), 216-217.

Barrett,TB; Benditt,EP (1988): Platelet-derived growth factor gene expression in human atherosclerotic plaques and normal artery wall. Proc. Natl. Acad. Sci, USA. 85, 2810-2814.

Bates,GP; MacDonald,ME; Baxendale,S; Sedlacek,Z; Youngman,S; Romano,D; Whaley,WL; Allito,BA; Poustka,A; Gusella,JF; Lehrach,H (1990): A yeast artificial chromosome spanning the Huntington disease gene. Am. J. Hum. Genet. 46, 762-775.

Berk,BC; Alexander,RW; Brock,TA; Gimbrone,MA; Webb,RC (1986): Vasoconstriction: a new activity for platelet-derived growth factor. Science. 232, 87-90.

Betsholtz,C; Johnsson,A; Heldin,C-H; Westermark,B; Lind,P; Urdea,MS; Eddy,R; Shows,TB; Philpot,K; Mellor,AL; Knott,TJ; Scott,J (1986): cDNA sequence and chromosomal localization of human platelet-derived growth factor A-chain and its expression in tumour cell lines. Nature. 320, 695-699.

Betsholtz,C; Rorsamn,F; Westermark,B; Ostman,A; Heldin,C-H (1990): Analogous alternative splicing. Nature. 344, 299.

Bird,AP (1987): CpG islands as gene markers in the vertebrate nucleus. TIGS 3, 342-347.

Bird,AP; Taggart,M; Frommer,M; Miller,OJ; Macleod,D (1985): A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell. 40, 91-99.

Birren,B; Lai,E (Eds.) (1993): Pulsed field gel electrophoresis a practical guide. 1st ed. Academic Press Inc, London, 253 pages.

Bonthron,DT; Collins,T; Gzeschik,K-H; Van Roy,N; Speleman,F (1992): Platelet-derived growth factor A chain: conformation of localization of PDGF A to chromosome 7p22 and description of an unusual minisatellite. Genomics. 13, 257-263.

Bonthron,DT; Morton,CC; Orkin,SH; Collins,T (1988): Platelet-derived growth factor A chain: Gene structure, chromosomal location and basis for alternative mRNA splicing. Proc. Natl. Acad. Sci, USA. 85, 1492-1496.

Bowden-Pope,DF; Koppen,AV; Schatteman,G (1991): Is PDGF really important? Testing the hypotheses. TIGS 7, 413-418.

Bowen-Pope,D; Vogel,A; Ross,R (1984): Production of platelet-derived growth factor-like molecules and reduced expression of PDGF receptors accompany transformation by a wide spectrum of agents. Proc. Natl. Acad. Sci, USA. 81, 2396-2400.

Bowen-Pope,DF; Malpass,TW; Foster,DM; Ross,R (1984): Platelet-derived growth factor in vivo: levels, activity and rate of clearance. Blood 64(2), 458-469.

Breuton,L; Huson,SM; Winter,RM; Williamson,R (1988): Chromosomal localisation of a developmental gene in man: direct DNA analysis demonstrates that Greig cephalopolysyndactyly maps to 7p13. Am. J. Med. Genet. 31, 799-804.

Brown,WRA; MacKinnon,PJ; Villasante,A; Spurr,N; Buckle,VJ; Dobson,MJ (1990a): Structure and polymorphism of human telomere asociated DNA. Cell. 63, 119-132.

Brown,WA; Dobson,MJ; MacKinnon,P (1990b): Telomere cloning and mammalian chromosome analysis. J. Cell. Sci. 95, 521-526.

Brownstein,BH; Siverman,GA; Little,RD; Burke,DT; Korsmeyer,SJ; Schlessinger,D; Olson,MV (1989): Isolation of single-copy human genes from a library of yeast artificial chromosome clones. Science. 244, 1348-1351.

Brueton,LA; van Herwerden,L; Chotai,KA; Winter,RM (1992): The mapping of a gene for craniosynostosis: evidence for linkage of the Saethre-Chotzen syndrome to distal chromosome 7p. J. Med. Genet. 29, 681-685.

Brunak,S; Engelbrecht,J; Knudsen,S (1991): Prediction of human mRNA donor and acceptor sites from the DNA sequence. J. Mol. Biol. 220, 49-65.

Budowle,B; Chakraborty,R; Giusti,AM; Eisenberg,AJ; Allen,RC (1991): Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am. J. Hum. Genet. 48, 137-144.

Buetow,KH; Weber,JL; Ludwigsen,S; Scherpbier-Haddema,T; Duyk,GM; Sheffield,VC; Wang,Z; Murray,JC (1994): Integrated human genome-wide maps constructed using the CEPH reference panel. Nature Genet. 6, 391-393.

Burke,DT; Carle,GF; Olsen,MV (1987): Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science. 236, 806-812.

Burmeister,M; Kim,S; Price,ER; Lange de,T; Tantravahi,U; Myers,RM; Cox,DR (1991): A map of the distal region of the long arm of human chromosome 21 constructed by radiation hybrid mapping and pulsed-field gel electrophoresis. Genomics. 9, 19-30.

Chotai,KA; Brueton.LA; Herwerden,L van; Garrett,C; Hinkel,GK; Schinzel,A; Mueller,RF; Speleman,F; Winter,R (1994): Six cases of 7p deletion: clinical, cytogenetic and molecular studies. Am. J. Med. Genet. 51, 270-276.

Chu,G; Vollrath,G; Davis,RW (1986): Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science. 234, 1582-1585.

Chumakov,I; Rigault,P; Guillou,S; Ougen,P; Billaut,A; et al. (1992): Continuum of overlapping clones spanning the entire human chromosome 21q. Nature. 359, 380-387.

Cohen,D; Chumakov,I; Weissenbach,J (1993): A first-generation physical map of the human genome. Nature. 366, 698-701.

Collins,T; Ginsburg,D; Boss,JM; Orkin,SH; Pober,JS (1985): Cultured human endothelial cells express platelet-derived growth factor B-chain: cDNA cloning and structural analysis. Nature. 316, 748-750.

Collins,T; Bonthron,DT; Orkin,SH (1987): Alternative RNA splicing affects function of encoded platelet-derived growth factor A chain. Nature. 328, 621-624.

Collins,T; Pober,JS; Gimbrone,MA; Hammacher,A; Betsholtz,C; Westermark,B; Heldin,C-H (1987): Cultured human endothelial cells express platelet-derived growth factor A-chain. Am. J. Path. 126, 7-12.

Collins,T; Young,R; Mendoza,A; Fries,JWU; Williams,AJ; Sultan,P; Bonthron,DT (1991): Regulation of PDGF expression in vascular cells. Atherosclerosis,xx 139-149.

Dausset,J; Cann,H; Cohen,D; Lathrop,M; Lalouel,J-M; White,R (1990): Centre d'Etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. Genomics. 6, 575-577.

DeLozier-Blanchet,CD; Masouye,I; Vollenweider,S (1992): Ring chromosome 7, hyperpigmented skin lesions and malignant melanoma. Am. J. Med. Genet. 43, 1039-1040.

Deuel,TF; Senior,RM; Huang,JS; Griffin,GL (1982): Chemotaxis of moncytes and neutrophils to platelet-derived growth factor. J. Clin. Invest. 69, 1046-1049.

Dietz-Band,J; Riethman,H; Hildebrand,CE; Moyzis,R (1990): Characterisation of polymorphic loci on a telomeric fragment of DNA from the long arm of human chromosome 7. Genomics. 8, 168-170.

Dobson,MJ; Brown,WRA (1992): Cloning human telomeres in yeast artificial chromosomes. Chap. 5. In: Techniques for the analysis of complex genomes. 1st ed. (Ed: Anand,R) Academic press, London, 81-92.

Doolittle,RF; Hunkapiller,MW; Hood,LE; Devare,SG; Robbins,KC; Aaronson,SA; Antoniades,HN (1983): Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding platelet-derived growth factor. Science. 221, 275-277.

Dower,WJ; Miller,JF; Ragsdale,CW (1988): High efficiency transformation of E.Coli by high voltage electroporation. Nuc. Acid. Res. 16(13), 6127-6145.

Drabkin,H; Sage,M; Helms,C; Green,P; Gemmil,R; Smith,D; Erickson,P; Hart,I; Ferguson-Smith,A; Ruddle,F; Tommerup,N (1989): Regional and physical mapping studies characterising the Greig polysyndactly 3;7 chromosome translocation, t(3;7)(p21.1;p13). Genomics. 4, 518-529.

Fantes,JA; Bickmore,WA; Fletcher,JM; Ballesta,F; Hanson,IM; van Heynigen,V (1992): Submicroscopic deletions at the WAGR locus, revealed by nonradioactive in situ hybridisation. Am. J. Hum. Genet. 51, 1286-1294.

Feinberg,AP; Vogelstein,B (1983): A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6-13.

Feinberg,AP; Vogelstein,B (1984): A Technique for radiolabeling DNA restriction endonuclease fragments to high specific activity: Addendum. Anal. Biochem. 137, 226.

Ferns,GAA; Ross,R (1990): Styl RFLP of the human platelet derived growth factor (PDGF) A-chain gene. Nuc. Acid. Res. 18(1), 208.

Ferrin,LJ; Camerini-Otero,RD (1991): Selective cleavage of human DNA: RecA-assisted restriction endonuclease (RARE) cleavage. Science. 254, 1494-1497.

Ferrin,LJ; Camerini-Otero,RD (1993): Synaptic complexes and RecA-Assisted restriction endonuclease cleavage of DNA. Chap. 4. In: Methods in molecular genetics. Vol. 2B. (Ed: Adolph,KW) Academic press, Orlando, 57-66.

Ferrin,LJ; Camerini-Otero,RD (1994): Long-range mapping of gaps and telomeres with RecA-assisted restriction endonuclease (RARE) cleavage. Nature Genet. 6, 379-383.

Flint,J; Craddock,CF; Villegas,A; Bentley,DP; Williams,HJ; Galanello,R; Cao,A; Wood,WG; Ayyub,H; Higgs,DR (1994): Healing of broken human chromosomes by the addition of telomeric repeats. Am. J. Hum. Genet. 55, 505-512.

Francis,F; Zehetner,G; Hoglund,M; Lehrach,H (1994): Construction and preliminary analysis of the ICRF human P1 library. GATA 11(5-6), 148-157.

Gashler,AL; Bonthron,DT; Madden,SL; Rauscher,FJ; Collins,T; Sukhatme,VP (1992): Human paltelet-derived growth factor A chain is transcriptionally repressed by the Wilms tumor suppressor WT1. Proc. Natl. Acad. Sci, USA. 89, 10984-10988.

Graham,JB; Kunkel,GR; Fowlkes,DM; Lord,ST (1990): The utility of a HindIII polymorphism of factor VIII examined by rapid DNA analysis. Brit. J. Haem. 76, 75-79.

Green,ED; Idol,JR; Mohr-Tidwell,RM; Braden,VV; Peluso,DC; Fulton,RS; Massa,HF; Magness,CL; Wilson,AM; Kimura,J; Weissenbach,J; Trask,BJ (1994): Integration of physical, genetic and cytogenetic maps of human chromosome 7: isolation and analysis of yeast artificial chromosome clones for 117 mapped genetic markers. Hum. Mol. Genet. 3(3), 489-501.

Green,ED; Olson,MV (1990): Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc. Natl. Acad. Sci, USA. 87, 1213-1217.

Greenhalgh,DG; Sprigel,KH; Murray,MJ; Ross,R (1990): Platelet-derived growth factor and fibroblast growth factor stimulate wound healing in the genetically diabetic mouse. Am. J. Path. 136(6), 1235-1246.

Greider,C (1990): Telomeres, telomerase and senescence. BioEssays 12(8), 363-368.

Grzeschik,K-H; Tsui,L-C; Green,ED (1994): Report of the first international workshop on human chromosome 7 mapping 1993. Cytogenet. Cell. Genet. 65, 52-73.

Grigoriou,M; Kastrinaki,M-C; Modi,WS; Theodorakis,K; Mankoo,B; Pachnis,V; Karagogeos,D (1995): Isolation of the human MOX5 homeobox gene and localisation to chromosome 7p221.1-p21.3. Genomics. 26, 550-555.

Gronwald,RGK; Adler,DA; Kelly,JD; Disteche,CM; Bowen-Pope,DF (1990): The human platelet-derived growth factor receptor alpha-subunit gene maps to chromosome 4 in close proximity to c-kit. Hum. Genet. 85, 383-385.

Grotendorst,GR; Chang,T; Seppa,HEJ; Kleinman,HK; Martin,GR (1982): Platelet-derived growth factor is a chemoattractant for vascular smooth muscle cells. J. Cell. Physiol. 113, 261-266.

Grotendorst,GR; Martin,GR; Pencev,D; Sodek,J; Harvey,AK (1985): Stimulation of granulation tissue formation by platelet-derived growth factor in normal and diabetic rats. J. Clin. Invest. 76, 2323-2329.

Guerrrini,AM; Ascenzioni,F; Pisani,G; Rappazzo,G; Valle,GD; Donini,P (1990): Cloning a fragment from the telomere of the long arm of human chromosome 9 in a YAC vector. Chromosoma 99, 138-142.

Gyapay,G; Morissette,J; Vignal,A; Dib,C; Fizames,C; Millasseau,P; Marc,S; Bernardi,G; Lathrop,M; Weissenbach,J (1994): The 1993-94 Genethon human genetic linkage map. Nature Genet. 7, 246-249.

Halford,S; Wadey,R; Roberts,C; et al. (1993): Isolation of a putative transcriptional regulator from the region of 22q11 deleted in DiGeorge syndrome, Shprintzen syndrome and familial congenital heart disease. Hum. Mol. Genet. 2, 2099-2107.

Hammacher,A; Hellman,U; Johnsson,A; Ostman,A; Gunnarsson,K; Westermark,B; Wasteson,A; Heldin,C-H (1988): A major part of platelet-derived growth factor purified from human platelets is a heterodimer of one A and one B chain. J. Biol. Chem. 263(31), 16493-16498.

Hammacher,A; Mellstrom,K; Heldin,C-H; Westermark,B (1989): Isoform-specific induction of actin reorganization by platelet-derived growth factor suggests that the functionally active receptor is a dimer. EMBO 8(9), 2489-2495.

Harley,CB (1991): Telomere loss: mitotic clock or genetic time bomb. Mutat Res. 256, 271-282.

Harley,CB; Futcher,AB; Greider,CW (1990): Telomeres shorten during ageing of human fibroblasts. Nature. 345, 458-460.

Hart CE, Bailey,M, Curtis,DA, Osborn,S, Raines,E, Ross,R, Forstrom,JW (1990): Purification of PDGF-AB and PDGFA-BB from human platelet extracts and identification of all three dimes in human platelets. Biochem 29, 166-172.

Hastie,ND; Allshire,RC (1989): Human telomeres: fusion and interstitial sites. TIGS 5, 326-330.

Hastie,ND; Dempster,M; Dunlop,MG; Thompson,AM; Green,DK; Allshire,RC (1990): Telomere reduction in human colorectal carcinoma and with aging. Nature. 346, 866-868.

Heinkoff,S (1984): Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene. 28, 351-359.

Heldin,CH; Westermark,B; Wasteson,A (1979): Platelet-derived growth factor: purification and partial characterisation. Proc. Natl. Acad. Sci, USA. 76(8), 3722-3726.

Heldin,C-H (1992): Structural and functional studies on platelet-derived growth factor. EMBO 11(12), 4251-4259.

Heldin,C-H; Johnsson,A; Wennergren,S; Wernstedt,C; Betsholtz,C; Westermark,B (1986): A human osteosarcoma cell line secretes a growth factor structurally related to a homodimer of PDGF A chains. Nature. 319, 511-514.

Heldin,C-H; Westermark,B (1989): Platelet-derived growth factor: three isoforms and two receptor types. TIGS 5, 108-111.

Helms,C; Mishra,SK; Riethman,H; Burgess,AK; Ramachandra,S; Tierney,C; Dorsey,D; Donis-Keller,H (1992): Closure of a genetic linkage map of human chromosome 7q with centromere and telomere polymorphisms. Genomics. 14, 1041-1054.

Henke,A; Fischer,C; Rappold,GA (1993): Genetic map of the human pseudoautosomal region reveals a high rate of recombination in female meiosis at the Xp telomere. Genomics. 18, 478-485.

Herwerden,L van; Rose,CSP; Reardon,W; Breuton,LA; Weissenbach,J; Malcolm,S; Winter,RM (1994): Evidence for locus heterogeneity in acrocephalosyndactyly: a refined localization for the Saethre-Chotzen syndrome locus on distal chromosome 7p - and exclusion of Jackson- Weiss syndrome from craniosynostosis loci on 7p and 5q. Am. J. Hum. Genet. 54, 669-674.

Hing,A; Helms,C; Donis-Keller,H (1993): VNTR and microsatellite polymorphisms within the subtelomeric region of 7q. Am. J. Hum. Genet. 53, 509-517.

Hoheisel,J; Pohl,FM (1986): Simplified preparation of unidirectional deletion clones. Nuc. Acid. Res. 14, 3605.

Hoheisel,JD; Maier,E; Mott,R; McCarthy,L; Grigoriev,AV; Schalkwyk,LC; Nizetic,D; Francis,F; Lehrach,H (1993): High resolution cosmid and P1 maps spanning the 14Mb genome of the fission yeast *S.pombe*. Cell. 73, 109-120.

Holmgren,L; Glaser,A; Pfeifer-Ohlsson,S; Ohlsson,R (1991): Angiogenesis during human extraembryonic development involves spatiotemporal control of PDGF ligand and receptor gene expression. Development. 113, 749-754.

Hoppe,J; Schumacher,L; Eichner,W; Weich,HA (1987): The long 3'-untranslated regions of the PDGFA and B mRNAs are only distantly related. Febs. Lett. 223, 243-246.

Hosang,M; Ronge,M; Wipf,B; Eggimann,B; Kaufmann,F; Hunziker,W (1989): Both homodimeric isoforms of platelet-derived growth factor (AA and BB) have mitogenic and chemotactic activity and stimulate phosphoinositol turnover. J. Cell. Physiol. 140, 558-564.

Ichikawa,H; Hosoda,F; Arai,Y; Shimizu,K; Ohira,M; Ohki,M (1993): A NotI restriction map of the entire long arm of human chromosome 21. Nature Genet. 4, 361-365.

IJdo,JW; Lindsay,EA; Wells,RA; Baldini,A (1992): Multiple variants in subtelomeric regions of normal karyotypes. Genomics. 14, 1019-1025.

Jabs,EW; Li,X; Scott,AF; Meyers,G; Chen,W; Eccles,M; Mao,J-i; Charnas,LR; Jackson,CE; Jaye,M (1994): Jackson-Weiss and Crouzen syndromes are allelic with mutations in fibroblast growth factor receptor 2. Nature Genet. 8, 275-279.

Kikkawa,K; Narahara,K; Tsuji,K; Kubo,T; Yokoyama,Y; Seino,Y (1992): Is loss of band 7p21 really critical for manifestation of craniosynostosis in 7p- ? Am. J. Med. Genet. 45, 108-110.

Kipling,D (Ed.) (1995): The telomere. 1st ed. Oxford University Press, Oxford. 208 pages.

Kohler,N; Lipton,A (1974): Platelets as a source of fibroblast growth promoting activity. Exp. Cell Res. 87, 297-301.

Kruger,G; Gotz,J; Kvist,U; Dunker,H; Erfurth,F; Pelz,L; Zech,L (1989): Greig syndrome in a large kindred due to reciprocal chromsome translocation t(6;7)(q27;p13). Am. J. Med. Genet. 32, 411-416.

Kunz,J; Scherer,SW; Klawitz,I; Soder,S; Du,Y-Z; Speich,N; Kalff-Suske,M; Heng,HHQ; Tsui,L-C; Grzeschik,K-H (1994): Regional localisation of 725 human chromosome 7-specific yeast artificial chromosome clones. Genomics. 22, 439-448. (Short Communication)

Lajeunie,E; Ma,HW; Bonaventure,J; Munnich,A; Merrer,M le; Renier,D (1995): FGFR2 mutations in Pfeiffer syndrome. Nature Genet. 9, 108.

de Lange,T; Shiue,L; Myers,RM; Cox,DR; Naylor,SL; Killery,AM; Varmus,HE (1990): Structure and variability of human chromosome ends. Mol. Cell. Biol. 10(2), 518-527.

Larsen,F; Gundersen,G; Lopez,R; Prydz,H (1992): CpG islands as gene markers in the human genome. Genomics. 13, 1095-1107.

Lathrop,GM; Lalouel,JM; Julier,C; Ott,J (1984): Strategies for mulitlocus linkage analysis in humans. Proc. Natl. Acad. Sci, USA. 81, 3443-3446.

Levinson,B; Kenwrick,S; Lakich,D; Hammonds,G; Gitschier,J (1990): A transcribed gene in an intron of the human factor VIII gene. Genomics. 7, 1-11.

Lewanda,AF; Cohen,MM; Jackson,CE; Taylor,EW; Li,W; Beloff,M; Day,D; Clarren,SK; Ortiz,R; Garcia,C; Hauselman,E; Figueroa,A; Wulfsberg,E; Wilson,M; Warman,ML; Padwa,BL; Whiteman,DAH; Mulliken,JB; Jabs,EW (1994a): Genetic heterogeneity among craniosynostosis syndromes: mapping the Saethre-Chotzen syndrome locus between D7S513 and D7S516 and exclusion of Jackson-Weiss and Crouzon syndrome loci from 7p. Genomics. 19, 115-119.

Lewanda,AF; Green,ED; Weissenbach,J; Jerald,H; Taylor,E; Summar,ML; Phillips,JA; Cohen,M; Feingold,M; Mouradian,W; Clarren,SK; Jabs,EW (1994b): Evidence that the Saethre-Chotzen syndrome locus lies between D7S664 and D7S507, by genetic anlysis and detection of a microdeletion in a patient. Am. J. Hum. Genet. 55, 1195-1201.

Libby,P; Warner,SJC; Ch,MBB; Salomon,RN; Birinyi,LK (1988): Production of platelet-derived growth factor-like mitogen by smooth muscle cells from human atheroma. N. Eng. J. Med. 318, 1493-1498.

Macina,RA; Negorev,DG; Spais,C; Ruthig,LA; Xue-Lan,H; Riethman,HC (1994): Sequence organisation of the human chromosome 2q telomere. Hum. Mol. Genet. 3(10), 1847-1853.

Majesky,MW; Benditt,EP; Schwartz,SM (1988): Expression and developmental control of platelet-derived growth factor A-chain and B-chain/sis genes in rat aortic smooth muscle cells. Proc. Natl. Acad. Sci, USA. 85, 1524-1528.

Majesky,MW; Reidy,MA; Bowen-Pope,DF; Hart,CE; Wilcox,JN; Schwartz,SM (1990): PDGF ligand and receptor gene expression during repair of arterial injury. J. Cell. Biol. 111, 2149-2158.

Matise,TC; Perlin,M; Chakravarti,A (1994): Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linnkage map. Nature Genet. 6, 384-390.

Matoskova,B; Rorsman,F; Svensson,V; Betsholtz,C (1989): Alternative splicing of the platelet-derived growth factor A-chain transcript occurs in normal as well as tumour cells and is conserved among mammalian species. Mol. Cell. Biol. 9(7), 3148-3150.

Mendoza,AE; Young,R; Orkin,SH; Collins,T (1990): Increased platelet-derived growth factor A-chain expression in human uterine smooth muscle cells during the physiologic hypertrophy of pregnancy. Proc. Natl. Acad. Sci, USA. 87, 2177-2181.

Mercola,M; Melton,DA; Stiles,CD (1988): Platelet-derived growth factor A-chain is maternally encoded in Xenopus embryos. Nature. 241, 1223-1225.

Mishra,SK; Helms,C; Dorsey,D; Permutt,MA; Donis-Keller,H (1992): A 2-cM genetic linkage map of human chromosome 7p that includes 47 loci. Genomics. 12, 326-334.

Muenke,M; Schell,U; Hehr,A; Robin,NH; Losken,HW; Schnizel,A; Pulleyn,LJ; Rutland,P; Reardon,W; Malcolm,S; Winter,RM (1994): A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome. Nature Genet. 8, 269-274.

Nakahara,K; Nishimura,H; Kuro-o,M; Takewaki,S; Iwase,M; Ohkubo,A; Yazaki,Y; Nagai,R (1992): Identification of three types of PDGF-A chain gene transcripts in rabbit vascular smooth muscle and their regulated expression during development and by angiotensin II. Biochem. Biophys. Res. Comm. 184, 811-818.

Negorev,DG; Macina,RA; Spais,C; Ruthig,LA; Hu,X-L; Riethman,HC (1994): Physical analysis of the terminal 270 kb of DNA from human chromosome 1q. Genomics. 22, 569-578.

Neil,DL; Vilasante,A; Fisher,RB; Vetrie,D; Cox,B; Tyler-Smith,C (1990): Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. Nuc. Acid. Res. 18(6), 1421-1428.

Ostman,A; Rall,L; Hammacher,A; Wormstead,MA; Coit,D; Valenzuela,P; Betsholtz,C; Westermark,B; Heldin,C-H (1988): Synthesis and assembly of a functionally active recombinant platelet-derived growth factor AB heterodimer. J. Biol. Chem. 263(31), 16202-16208.

Ostmann,A; Andersson,M; Betsholtz,C; Westermark,B; Heldin,C-H (1991): Identification of a cell retention signal in the B-chain of platelet-derived growth factor and in the long splice version of the A-chain. Cell. Reg. 2, 503-512.

Ott,J (Ed.) (1991): Analysis of human genetic linkage. 1st ed., The john Hopkins University press, Baltimore, 30.2 pages.

Paulsson,Y; Hammacher,A; Heldin,C-H; Westermark,B (1987): Possible autocrine feedback in the prereplicative phase of human fibroblasts. Nature. 328, 715-717.

Petit,C; Levilliers,J; Weissenbach,J (1988): Physical mapping of the human pseudo-autosomal region; comparison with genetic linkage map. EMBO 7(8), 2369-2376.

Pezzolo,A; Gimelli,G; Cohen,A; Lavaggetto,A; Romano,C; Fogu,G; Zuffardi,O (1993): Presence of telomeric and subtelomeric sequences at the fusion points of ring chromosomes indicates that the ring syndrome is caused by ring instability. Hum. Genet. 92, 23-27.

Pierce,GF; Mustoe,TA; Altrock,BW; Deuel,TF; Thomason,A (1991): Role of platelet-derived growth factor in wound healing. J. Cell. Biochem. 45, 319-326.

Pierce,GF; Mustoe,TA; Senior,RM; Reed,J; Griffin,GL; Thomason,A; Deuel,TF (1988): In vivo incisional wound healing augmented by platelet-derived growth factor and recombinant c-sis gene homodimeric proteins. J. Exp. Med. 167, 974-987.

Pierce,JC; Sternberg,N (1992): The Bacteriophage P1 cloning system. Chap. 3. In: Techniques for the analysis of complex genomes. 1st ed. (Ed: Anand,R) Academic Press, London, 39-58.

Pledger,WJ; Stiles,CD; Antoniades,HN (1978): An ordered sequences of events is required before balb/c-3T3 cells become committed to DNA synthesis. Proc. Natl. Acad. Sci, USA. 75, 2839-2843.

Raines,EW; Ross,R (1992): Compartmentalization of PDGF on extracellular binding sites dependent on exon-6-encoded sequences. J. Cell. Biol. 116, 533-543.

Rao,CD; Igarashi,H; Chiu,I-M; Robbins,KC; Aaronson,SA (1986): Structure and sequence of the human c-sis/platelet-derived growth factor 2 (sis/PDGF2) transcription unit. Proc. Natl. Acad. Sci, USA. 83, 2392-2396.

Reardon,W; McManus,SP; Summers,D; Winter,RM (1993): Cytogenetic evidence that the Saethre-Chotzen gene maps to 7p21.2. Am. J. Med. Genet. 47, 633-636.

Reardon,W; Winter,RM (1994): Saethre-Chotzen syndrome. J. Med. Genet. 31, 393-396.

Reardon,W; Winter,RM; Rutland,P; Pulleyn.LJ; Jones,BM; Malcolm,S (1994): Mutations in the fibroblast growth factor recepetor 2 gene cause Crouzen syndrome. Nature Genet. 8, 98-103.

Reid,CS; McMorrow,LE; McDonald-McGinn,DM; Grace,KJ; Ramos,FJ; Zackai,EH; Cohen,MM; Jabs,EW (1993): Saethre-Chotzen syndrome with familial translocation at chromosome 7p22. Am. J. Med. Genet. 47, 637-639.

Reston,JT; Hu,X-L; Macina,RA; Spais,C; Riethman,HC (1995): Structure of the terminal 300 kb of DNA from human chromosome 21q. Genomics. 26, 31-38.

Riethman,H; Moyzis,RK; Meyne,J; Burke,DT; Olson,MV (1989): Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector. Proc. Natl. Acad. Sci, USA. 86, 6240-6244.

Riethman,HC; Spais,C; Buckingham,J; Grady,D; Moyzis,RK (1993): Physical analysis of the terminal 240 kb of DNA frome human chromosome 7q. Genomics. 17, 25-32.

Riley,J; Butler,R; Ogilvie,D; Finniear,R; Jenner,D; Powell,S; Anand,R; Smith,JC; Markham,AF (1990): A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosomes (YAC) clones. Nuc. Acid. Res. 18(10), 2887-2890.

Riley,JH; Ogilvie,D; Anand,R (1992): Construction, characterisation and screening of YAC libraries. Chap. 4. In: Techniques for the analysis of complex genomes. 1st ed. (Ed: Anand,R) Academic Press, London, 59-80.

Rizzino,A; Bowen-Pope,DF (1985): Production of PDGF like growth factors by embryonal carcinoma cells and binding of PDGF to their endoderm like differentiated cells. Dev. Biol. 110, 15-22.

Roberts,WM; Look,AT; Roussel,MF; Sherr,CJ (1988): Tandem linkage of human CSF-I receptor (c-fms) and PDGF receptor genes. Cell. 55, 655-661.

Robson,MC; Phillips,LG; Thomason,A; Robson,LE; Pierce,GF (1992): Platelet-derived growth factor BB for the treatment of chronic pressure ulcers. Lancet. 339, 23-25.

Rorsman,F; Bywater,M; Knott,TJ; Scott,J; Betzholtz,C (1988): Structural characterization of the human platelet-derived growth factor A-chain cDNA and gene: Alternative exon use predicts two different precursor proteins. Mol. Cell. Biol. 8, 571-577.

Rorsman,F; Leveen,P; Betzholtz,C (1992): Platelet-derived growth factor (PDGF) A-chain mRNA heterogeneity generated by the use of alternative promoters and alternative polyadenylation sites. Growth Factors 7(3), 241-251.

Ross,R (1986): The pathogenesis of atherosclerosis - an update. N. Eng. J. Med. 314, 488-498.

Ross,R; Glomset,J; Kanya,B; Harker,L (1974): A platelet dependent serum factor that stimulates the proliferation of arterial smooth muscle cells in vitro. Proc. Natl. Acad. Sci, USA. 71(4), 1207-1210.

Ross,R; Matsuda,J; Raines,EW; Gown,AM; Katsuda,S; Sasahara,M; Malden,LT; Masuko,H; Sato,H (1990): Localisation of PDGF-B protein in macrophages in all phases of atherogenesis. Science. 248, 1009-1012.

Royle,NJ; Clarkson,RE; Wong,Z; Jeffreys,AJ (1988): Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics. 3, 352-360.

Rutland,RM; Pulleyn,LJ; Reardon,W; Baraister,M; Hayward,R; Jones,B; Malcolm,S; Winter,S; Oldridge.M; Slaney,SF; Poole,MD; Wilkie,AOM (1995): Identical mutaions in the FGFR2 gene cause both Pfeiffer and Crouzen syndrome phenotypes. Nature Genet. 9, 173-176.

Sambrook,J; Fritsch,EF; Maniatis,T (Eds.) (1989): Molecular biology, a laboratory manual. 2nd ed. 3 vols. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sanchez,A; Chesterman,CN; Sleigh,MJ (1991): Novel human PDGF A gene derived by alternative mRNA splicing. Gene. 98, 295-298.

Sanger,F; Nicklen,S; Coulson,AR (1977): DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci, USA. 74(12), 5463-5467.

Sasahara,M; Fries,JWU; Raines,EW; Gown,AM; Westrum,LE; Frosch,MP; Bonthron,DT; Ross,R; Collins,T (1991): PDGF B-chain in neurons of the central nervous system, posterior pituitary, and in a transgenic model. Cell. 64, 217-227.

Sawyer,JR; Rowe,RA; Hassed,SJ; Cunniff,C (1993): High-resolution cytogenetic characterisation of telomeric associations in ring chromosome 19. Hum. Genet. 91, 42-44.

Scherer,SW; Rommens,JM; Soder,S; Wong,S; Plavsic,N; Tompkins,BJF; Beattie,A; Kim,J; Tsui,L-C (1993): Refined localization and yeast artificial chromosome (YAC) contig-mapping of genes and DNA segments in the 7q21 - q32 region. Hum. Mol. Genet. 2(6), 751. (760)

Scherer,SW; Tompkins,BJF; Tsui,L-C (1992): Short communication: A human chromosome 7-specific genomic DNA library in yeast artificial chromosomes. Mam. Genome 3, 179-181.

Schomig-Spingler,M; Schmid,M; Brosi,W; Grimm,T (1986): Chromosome 7 short arm deletion, 7p21 - pter. Hum. Genet. 74, 323-325.

Schwartz,DC; Cantor,CR (1984): Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell. 37, 67-75.

Seifert,RA; Schwartz,SM; Bowen-Pope,DF (1984): Developmentally regulated production of platelet-derived growth factor-like molecules. Nature. 311, 669.

Senior,RM; Huang,JS; Griffin,GL; Deuel,TF (1985): Dissociation of the chemotactic and mitogenic activities of platelet-derived growth factor by human neutrophil elastase. J. Cell. Biol. 100, 351-356.

Seppa,H; Grotendorst,G; Seppa,S; Schiffmann,E; Martin,GR (1982): Platelet-derived growth factor is chemotactic for fibroblasts. J. Cell. Biol. 92, 584-588.

Shimokado,K; Raines,EW; Madtes,DK; Barrett; Benditt,EP; Ross,R (1985): A significant part of macrophage-derived growth factor consists of at least two forms of PDGF. Cell. 43, 277-286.

Silver,BJ (1992): Platelet-derived growth factor in human malignancy. Biofact. 3(4), 217-227.

Sinclair,K; Warner,JP; Bonthron,DT (1994): The ASP1 gene of Saccharomyces cerevisiae, encoding the intracellular isozyme of L-asparaginase. Gene. 144, 37-43.

Smith,EA; Seldin,MF; Martinez,L; Watson,ML; Choudhury,GG; Lalley,PA; Pierce,J; Aaronson,S; Barker,J; Naylor,SL; Sakaguchi,AY (1991): Mouse platelet-derived growth factor receptor alpha gene is deleted in W19H and patch mutations on chromosome 5. Proc. Natl. Acad. Sci, USA. 88, 4811-4815.

Southern,EM; Anand,R; Brown,WRA; Fletcher,DS (1987): A model for the separation of large DNA molecules by crossed field gel electrophoresis. Nuc. Acid. Res. 15, 5925-5943.

Speleman,F; Craen,M; Leroy,J (1989): De novo terminal deletion 7p22.1-pter in a child without craniosynostosis. J. Med. Genet. 26, 528-532.

Stenman,G; Eriksson,A; Claesson-Welsh,L (1989): Human platelet-derived growth factor A receptor gene maps to the same region on chromosome 4 as the kit oncogene. Gene. Chrom. Canc. 1, 155-158.

Stenman,G; Rorsman,F; Betzholtz,C (1988): Sublocalisation of the human PDGF A-chain gene to chromosome 7, band q11.23, by in situ hybridization. Proc. Natl. Acad. Sci, USA. 85, 180-183.

Stenman,G; Rorsman,F; Huebner,K; Betsholtz,C (1992): The human platelet-derived growth factor A-chain (PDGFA) maps to chromosome 7p22. Cytogenet. Cell. Genet. 60, 206-207.

Stephenson,DA; Mercola,M; Anderson,E; Wang,C; Stiles,CH; Bowden-Pope,DF; Chapman,VM (1991): Platelet-derived growth factor receptor alpha-subunit gene (Pdgfra) is deleted in the mouse patch (Ph) mutation. Proc. Natl. Acad. Sci, USA. 88, 6-10.

Sternberg,N (1990): Bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 kilobase pairs. Proc. Natl. Acad. Sci, USA. 87, 103-107.

Tanzi,RE; Haines,JL; Watkins,PC; Stewart,GD; Wallace,MR; Hallewell,R; Wong,C; Wexler,NS; Conneally,PM; Gusella,JF (1988): Genetic linkage map of human chromosome 21. Genomics. 3, 129-136.

Taylor,RN; Williams,LT (1988): Developmental expression of platelet derived growth factor and its receptor in the human placenta. Mol. Endo. 2, 627-632.

Tommerup,N; Nielson,F (1983): A familial reciprocal translocation t(3;7)(p21.1;p13) associated with the Greig polysyndactyly-craniofacial anomalies syndrome. Am. J. Med. Genet. 16, 313-321.

Tong,BD; Auer,DE; Jaye,M; Kaplow,JM; Ricca,G; McConathy,E; Drohan,W; Deuel,TF (1987): cDNA clones reveal diferences between human glial and endothelial cell platelet-deried growth factor A-chains. Nature. 328, 619-621.

Tsukamoto,H; Sakai,N; Taniike,M; Nakatsukasa,M; Yoshiwara,W; Sakamoto,H; Fujimura,H; Inui,K; Okada,S (1993): Case of ring chromsome 7: the first report of neuropathological findings. Am. J. Hum. Genet. 46, 632-635.

Uberbacher,EC; Mural,RJ (1988): Locating protein-coding regions in human DNA sequences by a multiple-sensor neural network approach. Proc. Natl. Acad. Sci, USA. 88, 11261-11265.

Vilageliu,L; Tyler-Smith,C (1992): Structural instability of YAC clones and the use of recombination deficient yeast host strains. Chap. 6. In: Techniques for the analysis of complex genomes. 1st ed. (Ed: Anand,R) Academic Press, London, 93-112.

Vogelstein,B; Gillespie,D (1979): Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci, USA. 76(2), 615-619.

Vollenweider-Roten,S; Masouye,I; Delozier-Blancher,CD; Saurat,J-H (1993): Cutaneous findings in ring chromosome 7 syndromes. Dermatol. 186, 84-87.

Vortkamp,A; Gessier,M; Grzeschik,K-H (1991): GL13 zinc-finger gene interrupted by translocations in Greig syndrome families. Nature. 352, 539-540.

Vortkamp,A; Thias,U; Gessler,M; Rosenkranz,W; Kroisel,PM; Tommerup,N; Kruger,G; Gotz,J; Pelz,L; Grezschik,K-H (1991): A somatic cell hybrid panel and DNA probes for physical mapping of human chromosome 7p. Genomics. 11, 737-743.

Walker,LN; Bowen-Pope,DF; Ross,R; Reidy,MA (1986): Production of platelet-derived growth factor-like molecules by cultured arterial smooth muscle cells accompanies proliferation after aterial injury. Proc. Natl. Acad. Sci, USA. 83, 7311-7315.

Wang,L-M; Killary,AM; Fang,X-E; Parriott,SK; Lalley,PA; Bell,GI; Sakaguchi,AY (1988): Chromosome assignment of mouse insulin, colony stimulating factor 2, and low-density lipoprotein receptors. Genomics. 3, 172-175.

Waterfield,MD; Scrace,GT; Whittle,N; Stroobant,P; Johnsson,A; Wasteson,A; Westermark,B; Heldin,C-H; Huang,JS; Deuel,TF (1983): Platelet-derived growth factor is structurally related to the putative transforming protein p28sis of simian sarcoma virus. Nature. 304, 35-39.

Weber,B; Colins,C; Robbins,C; Magenis,RE; Delaney,AD; Gray,JW; Hayden,MR (1990): Characterisation and organisation of DNA sequences adjacent to the human telomere associated repeat (TTAGGG)n. Nuc. Acid. Res. !8, 3353-3361.

Westermark,B (1990a): Review: The molecular and cellular biology of platelet-derived growth factor. Acta. Endocrinologica. 123, 131-142.

Westermark,B; Siegbahn,A; Heldin,C-H; Claesson-Welsh,L (1990b): B-type receptor for platelet-derived growth factor mediates a chemotactic response by means of ligand-induced activation of the receptor protein-tyrosine kinase. Proc. Natl. Acad. Sci, USA. 87, 128-132.

Wilcox,JN; Smith,KM; Williams,LT; Schwartz,SM; Gordon,D (1988): Platelet-derived growth factor mRNA detection in human atherosclerotic plaques by in situ hybridisation. J. Clin. Invest. 82, 1134-1143.

Wilkie,AOM; Buckle,VJ; Harris,PC; Lamb,J; Barton,NJ; Reeders,ST; Lindenbaum,RH; Nicholls,RD; Barrow,M; Bethlenfalvay,NC; Hutz,MH; Tolmie,JL; Weatherall,DJ; Higgs,DR (1990a): Clinical features and molecular analysis of the alpha thalassemia/mental retardation syndromes. I. Cases due to deletions involving chromosome band 16p13.3. Am. J. Hum. Genet. 46, 1112-1126.

Wilkie,AOM; Lamb,J; Harris,PC; Finney,RD; Higgs,DR (1990b): A truncated human chromosome 16 associated with alpha thalassaemia is stabilised by addition of telomeric repeat (TTAGGG)n. Nature. 346, 868-871.

Wilkie,AOM; Higgs,DR; Rack,KA; Buckle,VJ; Spurr,NK; Fischel-Ghodsian,N; Ceccherini,I; Brown,WRA; Harris,PC (1991): Stable length polymorphism of up to 260kb at the tip of the short arm of human chromosome 16. Cell. 54, 595-606.

Wilkie,AOM; Slaney,SF; Oldridge,M; Poole,MD; Ashworth,GJ; Hockley,AD; Hayward,RD; David,DJ; Pulleyn,LJ; Rutland,P; Malcolm,S; Winter,RM; Reardon,W (1995): Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzen syndrome. Nature Genet. 9, 165-172.

Yarden,Y; Escobedo,JA; Kuang,W-J; Yang-Feng,TL; Daniel,TO; Tremble,PM; Chen,EY; Ando,ME; Harkins,RN; Francke,U; Fried,VA; Ullrich,A; Williams,LT (1986): Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors. Nature. 323, 226-232.

Yeh,H-J; Ruif,KG; Wang,Y-X; Parks,WC; Snider,WD; Deuel,TF (1991): PDGF A-chain gene is expressed by mammalian neurons during development and in maturity. Cell. 64, 209-216.

Young,RM; Mendoza,AE; Collins,T; Orkin,SH (1990): Alternatively spliced platelet-derived growth factor A-chain transcripts are not tumour specific but encode normal cellular proteins. Mol. Cell. Biol. 10(11), 6051-6054.

Youngman,S; Bates,GP; Williams,S; McClatchey,AI; Baxendale,S; Sedlacek,Z; Altherr,M; Wasmuth,JJ; MacDonald,ME; Guesella,JF; Sheer,D; Lehrach,H (1992): The telomeric 60 kb of chromosome arm 4p is homologous to telomeric regions on 13p, 15p, 21p and 22p. Genomics. 14, 350-356.

Zhang,FR; Aurias,A; Delatre,O; Stern,MH; Benitez,J; Rouleau,G; Thomas,G (1990): Mapping of human chromosome 22 by in-situ hybridisation. Genomics. 7, 319-324.

Appendix I Suppliers of chemicals and enzymes

(i) Suppliers of chemicals

Acetic acid	A & J Beveridge, Edinburgh, Lothian, UK.
Acetylated BSA	New England Biolabs, Hitchin, Hertfordshire, UK.
Acrylamide	Merk Ltd, Lutterworth, Leices, UK.
ADP	Boehringer Mannheim UK Ltd, Lewes, East Sussex, UK.
Agarose	GibcoBRL, Life Technologies Ltd, Renfrewshire,UK.
Ammonium acetate	Sigma, Poole, Dorset, UK.
Ammonium persulphate	GibcoBRL, Life Technologies Ltd, Renfrewshire,UK.
Ampicillin	NBL, Cramlington, Northumberland, UK.
ATP-γ-S	Sigma, Poole, Dorset, UK.
Avidin FITC	Vector, Peterborourgh, UK.
Antiavidin	Vector (as above).
Bacto-agar	Difco, East Molesey, Surrey, UK.
Bacto-tryptone	Difco (as above).
Bacto-yeast extract	Difco (as above).
Bio-16-dUTP	Boehringer Mannheim UK Ltd, Lewes, East Sussex, UK.
Bromophenol blue	Sigma, Poole, Dorset, UK.
Calcium chloride	Merk Ltd, Lutterworth, Leices, UK.
Chloramphenicol	NBL, Cramlington, Northumberland, UK.
Chloroform	A & J Beveridge, Edinburgh, Lothian, UK.
CotI DNA	Gibco-BRL, Life Technologies Ltd, Renfrewshire,UK.
DAPI	Sigma, Poole, Dorset, UK.
Dextran sulphate	Pharmacia, Milton Keynes, UK.
Disodium hydrogen phosphate	A & J Beveridge, Edinburgh, Lothian, UK.
Dithiothreitol	Sigma, Poole, Dorset, UK.
DMSO	Sigma (as above).
EDTA	NBL, Cramlington, Northumberland, UK.
EGTA	Boehringer Mannheim UK Ltd, Lewes, East Sussex, UK.
Ethanol	A.A.H. Pharmaceuticals Glasgow, Strathclyde.
Ethidium Bromide	NBL, (as above).
Ficoll	Pharmacia, Milton Keynes, UK.
Formamide	Merk Ltd, Lutterworth, Leices, UK.
Geneclean Kit	Stratech Scientific Ltd, Luton, Bedfordshire, UK.
Glucose	Sigma, Poole, Dorset, UK.
Glycerol	A & J Beveridge, Edinburgh, Lothian, UK.
Glycine	Sigma, Poole, Dorset, UK.
Glycogen	Boehringer Mannheim UK Ltd, Lewes, East Sussex, UK.
Hybond N^+	Amersham International Plc, Chalfont, Buckinghamshire, UK.
Hydrochloric acid	A & J Beveridge, Edinburgh, Lothian, UK.
IPA	Rathburn, Walkerburn, Peebleshire, UK.
IPTG	Gibco BRL, Life Technologies Ltd, Inchinnan, Renfrewshire,UK.
Kanamycin	Gibco-BRL (as above).
LMP agarose	Gibco BRL (as above).
Lymphoprep	Nycomed, Sheldon, Birmingham, UK.
Magnesium acetate	Sigma, Poole, Dorset, UK.
Magnesium chloride	Merk Ltd, Lutterworth, Leices, UK.
Magnesium sulphate	Sigma, Poole, Dorset, UK.
Mineral oil	Sigma (as above).
Nitric acid	A & J Beveridge, Edinburgh, Lothian, UK.
Nucleotides	Pharmacia, Milton Keynes, UK.
Orange G	Merk Ltd, Lutterworth, Leices, UK.
PBS tablets	Sigma, Poole, Dorset, UK.

Phenol	Rathburn, Walkerburn, Peebleshire, UK.
PMSF	Sigma, Poole, Dorset, UK.
Potassium acetate	Sigma (as above).
Potassium chloride	A & J Beveridge, Edinburgh, Lothian, UK.
Proteinase K	Boehringer Mannheim UK Ltd, Lewes, East Sussex, UK.
Radioactive isotopes	Amersham International Plc, Chalfont, Buckinghamshire, UK.
Random hexamer oligonucleotides	Pharmacia, Milton Keyes, UK.
RecA protein	Cambio Ltd, Cambridge, UK.
Salmon sperm DNA	Sigma, Poole, Dorset, UK.
Scintran scintillation fluid	Merk, Lutterworth, Leices, UK.
SDS	Aldridge, Gillingham, Dorset, UK.
Sephadex	Pharmacia, Milton Keynes, UK.
Silver nitrate	Merk, Lutterworth, Leices, UK.
Sodium carbonate	Merk (as above).
Sodium chloride	A & J Beveridge, Edinburgh, Lothian, UK.
Sodium dihydrogen phosphate	A & J Beveridge, Edinburgh, Lothian, UK.
Sodium dodecyl sulphate	Gibco BRL, Life Technologies Ltd, Renfrewshire,UK.
Sodium hydroxide	Merk, Lutterworth, Leices, UK.
Sorbitol	Sigma, Poole, Dorset, UK.
Spermidine	Sigma (as above).
Sucrose	A & J Beveridge, Edinburgh, Lothian, UK.
TEMED	Sigma, Poole, Dorset, UK.
Tetracycline	NBL, Cramlington, Northumberland, UK.
Trichloroacetic acid	Merk, Lutterworth, Leices, UK.
Tris	Gibco BRL, Life Technologies Ltd, Inchinnan, Renfrewshire,UK.
Tween-20	Sigma, Poole, Dorset, UK.
Urea	Merk, Lutterworth, Leices, UK.
X-gal	Gibco-BRL, Life Technologies Ltd, Inchinnan, Renfrewshire,UK.
Xylene cyanol	Sigma, Poole, Dorset, UK.
Zymolyase	ICN Biomedicals,

(ii) Suppliers of enzymes

Enzyme	Supplier
Alw44I	Promega, Milton Keynes, UK.
ApaLI	Gibco-BRL, Life Technologies Ltd, Inchinnan, Renfrewshire,UK.
BamHI	New England Biolabs, Hitchin, Hertfordshire, UK.
BglII	NBL, Cramlington, Northumberland, UK.
BspDI (isoschizomer of ClaI)	New England Biolabs, Hitchin, Hertfordshire, UK.
BstZI (isoschizomer of EagI)	Promega, Milton Keynes, UK.
DNA polymerase I	Gibco-BRL, Life Technologies Ltd, Inchinnan, Renfrewshire,UK.
DNase	Boehringer Mannheim UK Ltd, Lewes, East Sussex, UK.
EcoRI	New England Biolabs, Hitchin, Hertfordshire, UK.
EcoRI methylase	New England Biolabs (as above).
ExoIII nuclease	Gibco-BRL, Life Technologies Ltd, Inchinnan, Renfrewshire,UK.
HaeIII	Gibco-BRL (as above).
HincII	NBL, Cramlington, Northumberland, UK.
HindIII	NBL (as above).
Klenow fragment (DNA polymerase I)	NBL, (as above).
KpnI	NBL (as above).
MluI	NBL (as above).
MspI	NBL (as above).
Mung bean nuclease	Boehringer Mannheim UK Ltd, Lewes, East Sussex, UK.
NotI	New England Biolabs, Hitchin, Hertfordshire, UK.
NruI	New England Biolabs (as above).
PCR-script cloning kit	Stratagene, Cambridge, UK.
Promega magic minipreps	Promega, Milton Keynes, UK.
PstI	NBL, Cramlington, Northumberland, UK.
PvuI	Pharmacia, Milton Keynes, UK.
PvuII	NBL, Cramlington, Northumberland, UK.
RNase	Gibco-BRL, Life Technologies Ltd, Inchinnan, Renfrewshire,UK.
RsaI	Gibco-BRL (as above).
SacII	New England Biolabs, Hitchin, Hertfordshire, UK.
SalI	New England Biolabs (as above).
Sau3AI	New England Biolabs (as above).
SfiI	New England Biolabs (as above).
T4 kinase	NBL, Cramlington, Northumberland, UK.
T4 ligase	Gibco-BRL, Life Technologies Ltd, Inchinnan, Renfrewshire,UK.
TA cloning kit	Invitrogen, Abingdon, UK.
Taq polymerase	Perkin Elmer, Warrington, UK.
XbaI	Gibco-BRL, Life Technologies Ltd, Inchinnan, Renfrewshire,UK.
XhoI	Gibco-BRL (as above).

Publication arising from data presented in this thesis

THE EUROGEM Map of Human Chromosome 7

F. Badbanchi[a]
M. Otto[a]
U. Kohlhauer[a]
N. Grzeschik[a]
S. Beck[a]
S.J.L. Smith[g]
D. M. Swallow[b]
R.F. Kooy[c]
G. Vergnaud[d,f]
M. Aldred[e]
K-H. Grzeschik[a]

[a] Med. Zentrum Für Humangenetik, Marburg, FRG;
[b] MRC Human Biochemical Genetics Unit, University College London, UK;
[c] Department of Medical Genetics, University of Groningen, The Netherlands;
[d] Laboratoire de Génétique des Espèces, Institut de Biologie, Nantes, France;
[e] MRC Human Genetics Unit, Western General Hospital, Edinburgh, UK;
[f] Centre d'Études du Bouchet, Vert le Petit, France;
[g] Human Genetics Unit, Department of Medicine, University of Edinburgh, UK

The first NIH-CEPH consortium map was based mainly on anonymous RFLP markers and functional genes [17]. Provisional index maps presented at the First International Workshop on Chromosome 7 Mapping 1993, by two groups, attempted to integrate a number of the first generation microsatellite markers [2], typed on 8 of the CEPH families, with other PCR-based markers like EGFR, GCK or TCRB, MUC3, ERV3, TCRG, and IL6 [17,18].

The EUROGEM map is exclusively based on 30 markers for which new data were generated by genotyping of the 40 CEPH families. It includes 10 Généthon markers, other microsatellites and RFLP markers, selected by the Resource Centre or the consortium, and 5 new dinucleotide microsatellite markers identified in Marburg: D7S1491, D7S1492, D7S1493, D7S1494, D7S1495. Details on the markers used to build the map can be obtained from the Genome Data Base (GDB).

All genotype information was error-checked visually and entered into the GENBASE program. A second check was made by running the LINK2SUM program to test for inconsistencies in the data (allelic exclusion and non-paternity). All possible errors were verified by retyping and either corrected or the conflicting genotype was deleted from the dataset before inclusion into the study. The analyses for map building were carried out using CRI-MAP, starting from the order predicted by the Généthon markers published by Weissenbach et al. [2], and including the additional markers in a stepwise fashion. For marker systems typed in the EUROGEM project for which there was already some data present in the CEPH version 6 database, in our map building effort this information was overruled by our new data. A first preliminary CRI-MAP 'build' was attempted on all newly genotyped markers and selected markers from the CEPH database. Markers at odds of <1000:1 were excluded. Local support for the preliminary map was assessed using the CRI-MAP options 'flips2' and 'flips4'. Subsequently all non-EUROGEM markers were eliminated.

The resulting EUROGEM map shown in the figure is a baseline map of 30 markers (out of 40 typed by the EUROGEM labs) for which new information was available ordered in unique positions with odds of at least 1000:1. The order of markers which had also been used by other groups agrees well with the one published previously [17,2]. In particular, the order in 7q36 (cen - D7S468 - D7S22 - tel) which seemed to be inverted in the map of Hill et al. [18] (cen - D7S22 - D7S468 - tel) is back to the order indicated by Helms et al. [17] and Gurreri et al. [19], probably due to the availabilty of new data in the present study. The map is 268.3 cM (sex-averaged), 337.4 cM (female), and 209.3 cM (male) in length. Physical locations had been described for 5 of the loci mapped in our study. In addition, we assigned the 5 microsatellite markers developed in our group by FISH and analysis of a somatic cell hybrid panel (data not shown) to regions of chromosome 7. As a result, 10 of the 30 markers on the genetic map have a well-defined physical position allowing the integration of genetic and physical maps: PDGFA - 7p22; D7S1495 - 7p22; D7S1492 - 7p22; D7S513 - 7p21; D7S484 - 7p14 to p15; D7S506 - 7p11.2; MUC3 - 7q22; D7S1493 - 7q22; D7S1494 - 7q36; D7S1491 - 7q36. Several important disease loci have been mapped to chromosome 7 recently: Split hand/split foot malformation (SHFD1) to 7q21-7q22; a deletion breakpoint interval in 7q22 in myeloid disorder; Williams-Beuren-Syndrome (WBS) to 7q11.23; two retinitis pigmentosa genes on 7q and 7p; a "complex bilateral polysyndactyly"-gene and a gene for triphalangeal thumb, both to the 7q36 region; a gene for holoprosencephaly to 7q36, and a gene for Saethre-Chotzen-type of craniosynostosis to distal chromosome 7p.

The integration of the markers on the genetic maps of this chromosome into physical maps and contiguous collections of clones during the next round of EUROGEM will define the position of these genes more precisely and thus speed up their identification.

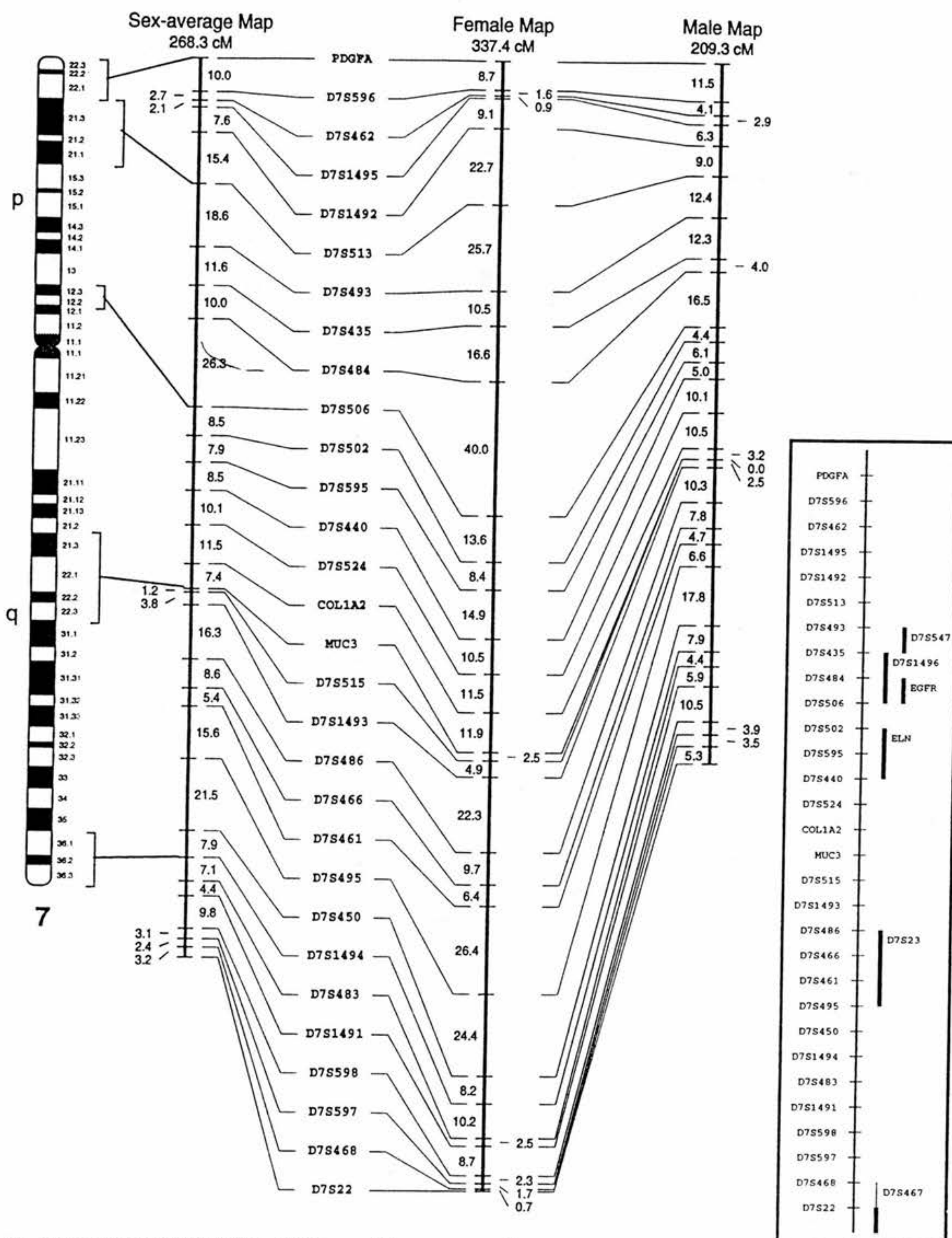


Fig. 7. The EUROGEM Map of Human Chromosome 7