
ENHANCING GENE EXPRESSION BY USING
CHROMOSOMAL SEQUENCES THAT FLANK AN
EFFICIENTLY EXPRESSED TRANSGENIC LOCUS

ANONA JANE CRANSTON

Ph.D. Thesis
University of Edinburgh
1996



ABSTRACT

The β -lactoglobulin (BLG) gene encodes a major whey protein in the milk of ruminants. Studies of the ovine BLG gene demonstrated that the gene was expressed specifically and efficiently in the mammary gland of transgenic mice. The promoter of the BLG gene has been used to target the expression of human proteins to the mammary gland in both transgenic mice and transgenic sheep. High levels of expression of some foreign proteins were obtained in some cases but many of the hybrid transgenes were expressed poorly and appeared susceptible to position effects. The aim of this work was to test whether the chromosomal sequences flanking an efficiently expressed transgenic locus could be used to improve transgene expression.

The AATB transgene comprising 4.2 kb of the ovine BLG promoter sequences linked to human α 1-antitrypsin (hAAT) genomic sequences gave high levels of hAAT expression in the mouse mammary. One line in particular (AATB 46.2) expressed hAAT at 6-8 mg/ml. This line was reported to have low transgene copy number, that could have been accommodated in a single cosmid vector. The aim of this work was to investigate the role of the flanking murine sequences in the high expression of the AATB transgene for this line.

This thesis details the restriction characterisation of the AATB 46.2 genomic locus and the cloning of most (if not all) of the locus in several cosmids. A total of 4.6 kb of the 5' flanking murine sequences and 4.7 kb of the 3' flanking murine DNA were successfully isolated in separate cosmid clones. Comparison of the clones with genomic DNA demonstrated the locus had been cloned without rearrangement. Two secondary transgenes TAB and TA were constructed from these clones. TAB comprised the cloned 5' and 3' sequences flanking a single AATB transgene, while TA comprised the AATB transgene with only the 3' murine DNA from the AATB 46.2 locus. Transgenic mice were generated using these constructs.

Expression of hAAT in the TAB transgenic mice was found to be improved compared with the AATB transgenic mice. Also the frequency of expression of the human protein by transgenic mice was increased. Ten of eleven transgenic founders/lines expressed the transgene at greater than 40 μ g/ml compared with eleven

of twenty four for AATB. The levels of hAAT expression in the secondary transgenic mice were generally very high and appeared to correlate with transgene copy number. By contrast expression of hAAT was not observed in any of the TA transgenic lines. The presence of either the 5, murine DNA alone or both the 5' and 3' murine DNA, may be required to bring about the improved expression of the biomedical transgene.

The cloned murine sequences confer improved expression upon the AATB transgene. It now remains to be seen if they can improve the expression of other biomedical transgenes regulated by the BLG promoter, and transgenes regulated by other promoters that direct expression in tissues other than the mammary gland

Acknowledgement

There are many people to thank for the supportive role they have played in the completion of this work. I thank my supervisor Dr John Clark for his patience during the writing of this thesis. I would also like to thank Dr John Mullins, Dr Ian Garner and Dr. Mike Dalrymple for useful discussion at various stages during this work.

I thank in particular many of the staff of the Molecular Biology Division of the Roslin Institute. I thank Alison Cowper for dealing with my constant queries in the early stages of this work. I particularly wish to thank Dr. Pamela Brown for her support throughout and for the excellent technical advice and teaching received. I thank also Ed Hitchin , Katrina Gordon and Ray Ansel for their encouragement and support.

I thank the Staff of PPL Therapeutics Ltd for making my six months in industry such a memorable and enjoyable experience and for their work in the production of the TAB transgenic mice.

Lastly I would like to thank my friends and family for their support and understanding throughout this work.

Dedication

To my friends and to Ian with love

Abbreviations

APS	ammonium persulfate
hAAT	human α_1 -antitrypsin
BLG	β -lactoglobulin
bp	base pairs
CAT	chloramphenicol acetyl transferase
DNA	deoxyribonucleic acid
DNase 1	deoxyribonuclease 1
dpm	disintegrations per minute
EDTA	Ethylenediaminetetra-acetic acid
kb	kilobases (1000 x bp)
MAR	matrix attachment region
MOPS	3-(N-morpholino)-propanesulphonic acid
MW	molecular weight
d NTP	2'-deoxy (N) 5'-triphosphate (N= adenosine, cytidine, guanosine)
OD _n	optical density _{wavelength in nanometers}
PEG	polyethylene glycol
RNA	ribonucleic acid
RNase A	ribonuclease A
SAR	scaffold attachment region
SDS	sodium dodecyl sulphate
Tris	tris (hydroxymethyl)aminoethane
TEMED	N, N, N', N' tetramethylethylenediamine

WAP	wehy acidic protein
w/v	weight/volume
g	gram
mg	milligram
μg	microgram
ml	millilitre
μl	microlitre
M	molar
mM	millimolar
μM	micromolar

CHAPTER 1 : INTRODUCTION	1
1.1 : Transgenic Biology	1
1.1.2 : The Generation Of Transgenic Animals By Retroviral Insertion Of DNA	2
1.1.3.1 : The Generation Of Transgenic Animals By Pronuclear Injection	2
1.1.3.2 : Efficiency Of Generating Transgenic Animals By Pronuclear Injection	3
1.1.4 : Generation Of Transgenic Animals By Sperm Mediated Transfer	5
1.1.5.1 : Modification of the Germline By Homologous Recombination	6
1.1.5.2 : Mouse Embryonic Stem Cells	7
1.1.5.3 : Specific Modifications can be Made to the Murine Genome by Gene Targeting in ES cells	8
1.1.5.4 : The Efficiency Of Producing Transgenic Mice By Homologous Recombination	12
1.1.5.5 : ES Like Cells of Large Animals Could Have Applications in Biotechnology	15
1.1.5.6 Summary	16
1.2 THE MECHANISM OF TRANSGENE INTEGRATION	17
1.2.1 : Characteristics Of Integrated Foreign DNA	17
1.2.2 : Fate Of Experimentally Introduced DNA In Mammalian Cells	17
1.2.3 : Efficiency Of Circularisation Of Foreign DNA	19
1.2.4 : The Mechanism Of Concatamer Formation	21
1.2.5 : Studies Of Integrated Transgene Loci	21
1.2.6 : Concatamer Integration Into The Host Genome	27
1.3 : POSITION EFFECTS IN TRANSEGENIC ANIMALS	28
1.3.1 : Position Effects	28
1.3.2 : Position Effect And Transgene Methylation	29
1.3.3: Transgenes And Position Effect Variegation	30
1.3.4 : Positive Position Effects	32
1.4 : ELEMENTS INVOLVED IN GENE AND TRANSGENE REGULATION	32
1.4.1 : Elements other than the Basic Promoter are Required for Appropriate Transcription of Eukaryotic Genes	32
1.4.2 : The Role Of Introns In Transgene Expression	34
1.4.3 Enhancers	36
1.4.4 : Locus Control Region	40
1.4.5 : Scaffold/Matrix Attachment Regions	45
1.4.6 : Non MAR Boundary Elements	51
1.5 : TRANSGENIC ANIMALS IN THE PRODUCTION OF BIOMEDICAL PROTEINS	53
1.6. : LACTATION	56
1.7.1: The Hormonal Regulation Of Milk Protein Genes	57
1.7.2 : The Casein Genes	58
1.7.3 : The Rodent Whey Acidic Protein (WAP) Gene	59
1.7.4 : The Ovine BLG Gene	60
1.8 : OVINE β-LACTOGLOBULIN AS A TRANSGENE IN MICE	63
1.9 : AATB THE TRANSGENE	64
1.10 : THE AIM OF THIS WORK	68

CHAPTER 2 : MATERIALS AND METHODS	68
2.1 : Mice Used	68
2.2 : Recombinant plasmid clones	68
2.3 : Non-recombinant plasmid vectors and bacterial host strains used	68
2.4 : Enzymes, antibiotics and specialised chemicals	69
2.5 : Media	70
2.6 : Autoradiography and Photography	71
2.7 : Preparation of High Molecular Weight DNA from Mouse Liver	71
2.8 : Isolation of RNA by RNAZOL	72
2.9 : Deproteinisation of nucleic acids using phenol/chloroform extraction	73
2.10 : Ethanol precipitation of nucleic acids	74
2.11 : Spectrophotometric analysis to determine concentration of DNA and RNA samples.	74
2.12 : Restriction Endonuclease digestion of DNA	75
2.13 : Agarose Gel Electrophoresis	75
2.14 : Acrylamide gel Electrophoresis	76
2.15 : Western blotting	77
2.16 : Transfer of nucleic acids onto a solid support	78
2.17 : Gel purification of DNA	79
2.18 : Radioactive labelling of DNA	81
2.19 : Hybridisation	81
2.20 : Construction of a Cosmid Library of AATB 46.2	82
2.21 : Phage Cloning of the Secondary transgene	85
2.22 : Erase a Base deletions of 3' murine DNA	88
2.23 : Sequencing	89
2.24 : Generation of Transgenic Mice	90
2.25 : Tail screening by PCR	90
2.26 : Analysis of Milk Samples	91
2.27 : Hybridisation of Dried Agarose Gels (Unblots)	92

CHAPTER 3 : RESTRICTION ENZYME ANALYSIS OF HIGH EXPRESSING AATB LOCI	95
3.1 INTRODUCTION	95
3.1.1 The Apparent Position Effect Of AATB 35	95
3.1.2 The Proposed Approach	97
3.2 ESTIMATING THE COPY NUMBER OF AATB 35	104
3.2.1 Copy Number Estimation Of AATB 35 By Comparison With Plasmid Controls	104
3.2.2 Copy Number Estimation Of AATB 35 By Junction Fragment Comparisons	111
3.3 ESTIMATION OF THE COPY NUMBER OF AATB 46.2	120
3.3.1 The Apparent Position Effect In AATB 46.2	120
3.3.2 Copy Number Comparison of AATB 35 with AATB 46.2	120
3.3.3 Copy Number Estimation Of AATB 46.2 By Dilution Blot Analysis Of AATB 46.2	121
3.4 THE ALTERNATIVE APPROACH	125
3.4.1 Constructing A Secondary Transgene From Cosmid Clones Harboured The AATB 46.2 Flanking DNA	125
3.5 THE RESTRICTION MAP OF THE LOCUS OF AATB 46.2	129

CHAPTER 4 : CLONING AND CHARACTERISATION OF THE AATB 46.2 LOCUS	138
4.1 INTRODUCTION	138
4.1.1 Construction of the Cosmid Library	138
4.1.3 Restriction Mapping Of The Cosmid Clones	140
4.2 RESTRICTION ANALYSIS OF CLONES ISOLATED FROM THE LIBRARY	148
4.2.1 Restriction Analysis Of Clone 5	148
4.2.2 Restriction Analysis Of Clone 10	160
4.2.3 Restriction Analysis Of Clone 9	173
4.2.4 Restriction Analysis Of Clone 8	189
4.3 Comparison Of Clones With Genomic DNA	208

CHAPTER 5 : THE CONSTRUCTION OF THE SECONDARY TRANSGENE214

5.1 Introduction	214
5.2 : The Pre-integration site of AATB 46.2 has not been Characterised	214
5.3 : Only Seven Bases are Missing At The 5' Chromosomal/Transgene Junction	215
5.3.1 : The Immediate 5' Flanking DNA is AT Rich	216
5.4 : Fifty Base Pairs of AATB have been Deleted At the 3' junction of AATB 46.2	224
5.4.1 : The 3' Flanking Murine DNA of AATB 46.2 Contains Murine Repetitive Elements	225
5.4.2 Rational For Use of a Phage Vector to Construct the Secondary Transgene	243
5.4.3 Construction of the Secondary Transgene	250
5.4.4 : Phage Cloning of the Secondary Transgene	251

CHAPTER 6 : EXPRESSION OF THE SECONDARY TRANSGENES IN THE MAMMARY GLAND OF TRANSGENIC MICE	253
6.1 INTRODUCTION	253
6.2 EXPRESSION OF AATB IN SECONDARY TRANSGENIC MICE	254
6.2.1 Generation of Transgenic Mice	254
6.2.2 Estimation of Copy Numbers	256
6.2.3 Expression of hAAT in TAB Transgenic Mice	266
6.2.4 Lactation Difficulties in TAB Transgenic Mice	277
6.2.5 The Tissue Specificity Of The TAB Transgene	278
6.2.6 hAAT Expression Per Copy Of The TAB Transgene.	284
6.3 Analysis of the TA lines	288
6.4 Discussion	288

CHAPTER 7 : DISCUSSION	291
7.1 : A Summary of This Work	291
7.2 : Further transgenic experiments	293
7.3 : The Tissue Specificity of TAB is the Same as AATB	295
7.4 : The Murine Sequences May Contain Boundary Elements	296
7.5 : Can these Murine Sequences Confer their effect in Dairy Animals ?	297
References	300

Figures

Chapter 1

FIGURE 1.1 A : INSERTION VECTORS B : REPLACEMENT VECTORS	11
FIGURE 1.2 : POSITIVE NEGATIVE SELECTION FOR HOMOLOGOUS RECOMBINATION IN GENE TARGETING EXPERIMENTS - A : HOMOLOGOUS RECOMBINATION - B : RANDOM INTEGRATION.....	14
FIGURE 1.3 : STRUCTURE OF TWO INTEGRATED TRANSGENE LOCI	25

Chapter 3

FIGURE 3.1 : PROPOSED APPROACH FOR STUDY OF POSITION EFFECT IN LINE AATB 35.....	99
FIGURE 3.2 : ORIGINAL COPY NUMBER ESTIMATE OF LINE AATB 35	101
FIGURE 3.3: PROBES USED IN GENOMIC MAPPING OF TRANSGENIC MICE.....	103
FIGURE 3.4 : DETERMINATION OF AATB 35 COPY NUMBER BY COMPARISON WITH PLASMID CONTROL AUTORADIOGRAPH OF COPY NUMBER CONTROL EXPERIMENT.....	106
FIGURE 3.5 : COMPARISON OF THIRD GENERATION MOUSE OF AATB 35 WITH GO MICE OF OTHER AATB TRANSGENIC LINES	110
FIGURE 3.6 : AUTORADIOGRAPH OF ECO RV DIGESTED GENOMIC DNA OF LINE AATB 35	113
FIGURE 3.7 : DENSITOMETRY OF THE AUTORADIOGRAPH SHOWN IN FIGURE 3.6.....	115
FIGURE 3.8 : COMPARISON OF AATB 46.2 WITH AATB 35.....	119
FIGURE 3.9 : ESTIMATION OF COPY NUMBER OF AATB 46.2 BY DILUTION BLOT ANALYSIS.....	124
FIGURE 3.10 : THE ALTERNATIVE APPROACH.....	128
FIGURE 3.11 : DETECTION OF JUNCTION FRAGMENTS	132
FIGURE 3.12 : IDENTIFICATION OF 5' HYBRIDISING AND 3' HYBRIDISING JUNCTION FRAGMENTS.....	134
FIGURE 3.13 : ESTIMATED RESTRICTION MAP OF LOCUS OF AATB 46.2.	136

Chapter 4

FIGURE 4.1 : SUPER-COS-ONE VECTOR.....	143
FIGURE 4.2 : AATB PROBES USED FOR SCREENING THE LIBRARY AND CHARACTERISATION OF THE CLONES	144
FIGURE 4.3 : RESTRICTION SITES IN THE AATB TRANSGENE.....	147
FIGURE 4.4 : ETHIDIUM STAINED AGAROSE GEL OF CLONE 5 RESTRICTION DIGESTS.....	155
FIGURE 4.5 : AUTORADIOGRAPH OF SOUTHERN BLOT OF GEL IN FIGURE 4.4	157
FIGURE 4.6 : RESTRICTION MAP OF CLONE 5.....	159
FIGURE 4.7 : ETHIDIUM STAINED AGAROSE GEL OF CLONE 10 RESTRICTION DIGESTS	168
FIGURE 4.8: AUTORADIOGRAPH OF SOUTHERN BLOT OF GEL IN FIGURE 4.6.....	170
FIGURE 4.9 : RESTRICTION MAP OF CLONE 10.....	172
FIGURE 4.10 : ETHIDIUM STAINED AGAROSE GEL OF CLONE 9 RESTRICTION DIGESTS.....	180
FIGURE 4.11 : AUTORADIOGRAPHS OF SOUTHERN BLOT OF GEL IN FIGURE 4.10.....	182
FIGURE 4.12 : ETHIDIUM STAINED AGAROSE GEL OF DOUBLE DIGESTS OF CLONES 8 AND 9	184
FIGURE 4.13 : AURORADIOGRAPHS OF SOUTHERN BLOT OF AGAROSE GEL IN FIGURE 4.12.....	186
FIGURE 4.14 : RESTRICTION MAP OF CLONE 9	188
FIGURE 4.15 : RESTRICTION DIGESTS OF CLONE 8 (FIRST DNA PREPARATION).....	197
FIGURE 4.16 : RESTRICTION DIGESTS OF CLONE 8 (SECOND DNA PREPARATION).....	199
FIGURE 4.17 : ETHIDIUM STAINED AGAROSE GEL OF CLONE 8 RESTRICTION DIGESTS	201
FIGURE 4.18 : AUTORADIOGRAPHS OF SOUTHERN BLOT OF GEL IN FIGURE 4.17.....	203
FIGUR 4.19 : ANALYSIS OF THE SMALL BGL II FRAGMENTS	205
FIGURE 4.20 : RESTRICTION MAP OF CLONE 8	207
FIGURE 4.21 : DIRECT COMPARISON OF CLONES 8 AND 9 WITH GENOMIC DNA OF AATB 46.2	210
FIGURE 4.22 : COMPARISON OF GENOMIC MAP WITH MAPS OF CLONES 5,8,9 AND 10	212

Chapter 5

FIGURE 5.1 : SEQUENCING OF THE TRANSGENE/CHROMOSOMAL JUNCTION OF CLONE 9	219
FIGURE 5.2 A : AT RICH SEQUENCES OF THE IMMEDIATE 5' FLANKING DNA	221
FIGURE 5.2 B : SEQUENCE OF 5' END OF ECO RV JUNCTION FRAGMENT.....	223

<u>FIGURE 5.3 : SEQUENCING OF THE 3' TRANSGENE/CHROMOSOMAL JUNCTION OF CLONE 8</u>	229
<u>FIGURE 5.4: REGIONS OF CLONE 8 MURINE DNA SEQUENCED</u>	231
<u>FIGURE 5.5 (A) : SEQUENCE OF SECTION (A)</u>	233
<u>FIGURE 5.5 (B) : SEQUENCE OF SECTION (B)</u>	237
<u>FIGURE 5.5(C) : SEQUENCE OF SECTION (C)</u>	240
<u>FIGURE 5.6 : PROBING GENOMIC DNA WITH MURINE SEQUENCES OF CLONE 8</u>	242
<u>FIGURE 5.7(A) ; THE LARGE NOT I FRAGMENTS FROM CLONE 8 AND 9 ARE SUITABLE FOR CONSTRUCTING THE SECONDARY TRANSGENE</u>	245
<u>FIGURE 5.7 (B) : JOINING THE TWO NOT I SRF I FRAGMENTS GIVES AN AATB TRANSGENE FLANKED BY THE CHROMOSOMAL DNA OF AATB 46.2</u>	247
<u>FIGURE 5.7 (C) : THE LAMBDA DASH II PHAGE CLONING VECTOR</u>	249

Chapter 6

<u>FIGURE 6.1 : PROBE USED TO ESTIMATE THE COPY NUMBER OF TAB TRANSGENIC MICE</u>	260
<u>FIGURE 6.2 : SOUTHERN BLOT FOR COPY NUMBER ESTIMATION OF ROSLIN INSTITUTE GENERATED TAB TRANSGENIC MICE</u>	262
<u>FIGURE 6.3 ; SOUTHERN BLOT FOR COPY NUMBER ESTIMATION OF PPL GENERATED TAB TRANSGENIC MICE</u>	265
<u>FIGURE 6.4 : SDS PAGE OF ROSLIN INSTITUTE TAB TRANSGENIC MOUSE MILKS</u>	269
<u>FIGURE 6.5: SDS PAGE OF PPL TAB TRANSGENIC MOUSE MILKS</u>	271
<u>FIGURE 6.6 : WESTERN BLOT OF TAB MILK SAMPLES</u>	274
<u>FIGURE 6.7: COMPARING HAAT EXPRESSION OF TAB TRANSGENIC MICE WITH AATB TRANSGENIC MICE</u>	276
<u>FIGURE 6.8 : NORTHERN BLOT ANALYSIS OF TOTAL RNA FROM TRANSGENIC MICE AATB 46.2 AND TAB 36.2</u>	281
<u>FIGURE 6.9 : NORTHERN BLOT ANALYSIS OF TOTAL RNA FROM TRANSGENIC MOUSE TAB 36.2</u>	283
<u>FIGURE 6.10 : PLOT OF COPY NUMBER AGAINST HAAT EXPRESSION LEVEL IN TAB TRANSGENIC MICE</u>	287

CHAPTER 1 : INTRODUCTION

1 1 1: Transgenic Biology

Transgenic animals are animals that have integrated foreign DNA into their germ line as a consequence of experimental introduction of DNA (Palmiter and Brinster, 1985). Transgenic animals not only exhibit unique phenotypes , but they also pass those traits to their offspring (Palmiter and Brinster, 1985,1986; Jaenisch 1988). The development of methods for engineering and introducing functional foreign genes into whole organisms has contributed to understanding the factors which control the expression of specific genes and in the identification of genes and processes which regulate cell proliferation, differentiation and interactions. Transgenic livestock have been produced that express economically important proteins in their blood and milk.

Following the isolation of mouse embryonic stem cells it has been possible to make defined mutations in the murine genome by homologous recombination. This technology has enabled the generation of transgenic mice carrying mutations in specific genes (Hasty et al., 1991) and more recently to replace genes with a similar gene from another species (Stacey et al., 1994).

The following sections review the techniques used to generate transgenic animals, factors affecting transgene expression and the current state of production of economically important proteins in transgenic livestock.

1.1.2 : The Generation Of Transgenic Animals By Retroviral Insertion Of DNA

The production of transgenic animals has been ongoing since 1976. The first transgenic mouse strain generated had the Moloney Leukaemia retrovirus DNA stably integrated into the germ line (Jaenisch 1976). The method used was to infect mouse embryos with retroviruses early in development. This method of generating transgenic animals had several disadvantages. The integration of the foreign DNA into the genome occurred at various stages of embryonic development, and thus the foreign DNA was not always integrated in all somatic tissues or the germ line. In addition the transgenic animals generated using this approach often contained more than one site of insertion because more than one cell in the early embryo was infected with the virus. This required the strains of mice to be outbred to segregate the loci and generate animals harbouring only one site of insertion. Furthermore the amount of foreign DNA that can be accommodated within a retroviral vector is limited to approximately 8kb, and this is inadequate for many transgenic experiments.

1.1.3.1 : The Generation Of Transgenic Animals By Pronuclear Injection

In 1980 a breakthrough in the generation of transgenic mice hailed the advent of a new and exciting means of using whole animals to analyse gene expression. The technique was to inject naked DNA into the pronucleus of a newly fertilised mouse egg (Gordon et al., 1980). It was found, using this technique, that the injected DNA integrated into the genome, and could be inherited by the offspring of the founder transgenic animals. By injecting into the pronucleus the introduced DNA could

integrate into the genome before or at the first round of DNA replication and cleavage, and could thus contribute to the genome of all somatic tissues and of the germ line. Integration of the foreign DNA can also occur after the first zygotic division, this is shown by the generation of mosaic founder animals that, on average, harbour less than one copy of the transgene per cell and the generation of founder animals that transmit the transgene at a frequency significantly less than 50% (Wilkie et al., 1986).

1.1.3.2 : Efficiency Of Generating Transgenic Animals By Pronuclear Injection

The efficiency of generating transgenic animals by pronuclear injection is dependent on a number of parameters. As with any new technique the parameters were at first poorly defined. In early experiments Brinster and co-workers carried out a fairly detailed study of factors that affect the efficiency of generating transgenic foetuses by pronuclear injection (Brinster et al., 1985). The study investigated the effect of (i) DNA concentration, (ii) buffer, (iii) nuclear versus cytoplasmic injection, (iv) plasmid form, and (v) genetic background.

The experiment was performed using a single 4.8 kb plasmid. The plasmid was injected in four different forms, supercoiled, restricted and blunt ended, restricted with a single enzyme generating compatible ends, and restricted with two enzymes giving dissimilar ends. Plasmid form was found to have an effect on the efficiency of generating transgenic foetuses. The finding was that the injection of linear DNA was five times more efficient than supercoiled DNA. No significant difference was found between the linear molecules that had similar or dissimilar ends. Although the data

was not conclusive there was a suggestion that blunt ended DNA fragments may be less efficient than those with staggered ends.

Generally the foreign DNA was found to integrate at a single site in the genome. The site of integration of the transgene was thought to be random. The number of copies of the foreign DNA molecule was found to vary between 1 and several hundred copies and these copies were usually found to be orientated in a head to tail manner. The predominance of this arrangement of transgenes within transgenic animals relative to tail to tail or head to head arrangements led to the proposal that the arrays of transgenes were formed by homologous recombination between the injected fragments of DNA (Brinster et al., 1981; Palmiter et al., 1982; Palmiter and Brinster, 1985).

In the early transgenic experiments expression of most transgenes were obtained but the levels of expression and the tissue specificity of expression were often unpredictable and variable. This was generally a consequence of transgene design and transgenic animals soon provided much information about gene expression and the regulatory sequences required for appropriate gene expression. The purpose of many transgenic studies is to elucidate the involvement of a gene in normal developmental processes. The analysis of transgene expression can demonstrate the presence or lack of the required *cis*-acting sequences for its appropriate and abundant expression.

Although the pattern of expression of some of the early transgenes showed correct tissue specificity in the lines that expressed the gene, levels of expression were often disappointing. Frequently many of the lines generated did not detectably express the transgene. Such difficulties with transgene expression were believed to be a result of the repressive effects of the chromatin structure of the genomic DNA into which

they had integrated. This was termed the position effect and is discussed widely in the literature (e.g. see Al Shawi and Bishop 1990) although the nature of these effects are still understood incompletely. Some transgenes, however, were less susceptible to position effects than others and it soon became evident that position effects could be minimised by the inclusion of sufficient regulatory sequences within the transgene (Section 1.4 and references within).

Pronuclear injection is, nevertheless, an inefficient process even in mice. Many fertilised ova are required, which necessitates the use of large numbers of donor animals. Of the embryos that survive the injection procedure only ten percent give rise to transgenic offspring even in the hands of a practised microinjectionist.

When the technique is used to generate transgenic large animals it is even more inefficient and the number of fertilised ova surviving the injection procedure and giving rise to transgenic animals is less than in mice (Clark et al., 1990). Unlike mice, many large domestic animals such as cows, sheep and goats have only one or two offspring (Pigs are an exception in this respect). Thus the number of donor and recipient animals required is considerably greater than for a transgenic mouse experiments and the production of large transgenic animals is, overall, very inefficient and expensive.

1.1.4 : Generation Of Transgenic Animals By Sperm Mediated Transfer

Sperm mediated transfer involves incubating the transgene DNA with spermatozoa which are then used to fertilise eggs *in vitro*. Only one publication to date claims to have produced transgenic mice using this method (Lavitrano et al.,

1989). The result has never been successfully repeated even by those experienced in the generation of transgenic animals (Brinster et al., 1989). Six years on other groups are still trying to use this technique to produce transgenic animals, particularly in livestock species, despite this there remains no published evidence of its effectiveness for the generation of transgenic animals.

1.1.5.1 : Modification of the Germline By Homologous Recombination

Yeast is used extensively for genetic studies. Although it is unicellular, it is a lower eukaryote, and its genetics are thus thought to be more representative of the higher eukaryotes than prokaryotes. Indeed, many of the findings in yeast have been extrapolated and found to bear relevance to higher eukaryotic systems.

Notwithstanding the similarities a major difference between yeast and mammals is the frequency with which homologous recombination occurs between experimentally introduced DNA and the genome. In yeast homologous recombination events are at least ten times more frequent than random integration events (Hinnen et al 1978). In mammals the ratio is less than 1:100 (Roth and Wilson 1988). A number of factors are likely to be responsible for the 1000 fold difference in the frequency of homologous recombination. One obvious factor that may be significant is a substantial difference in genome size between *Saccharomyces cerevisiae* and higher eukaryotes (14×10^7 and 3×10^9 bp respectively), thus the target genomic sequence is more dilute. Thus unlike yeast, mammalian cells integrate foreign DNA into their genome principally by non homologous, or illegitimate recombination.

A number of experiments studying homologous recombination have been done in established mammalian cell lines. In these experiments defined genetic changes were introduced and mutation of both selectable genes and cellular genes (using selectable markers) have been performed successfully (Foiger et al., 1982). Such experiments demonstrated the feasibility of creating defined mutations in mammalian cells and proved useful in the elucidation of the mechanisms of homologous recombination in mammalian cells, although their usefulness in the study of gene function in the whole animal was limited.

1.1.5.2 : Mouse Embryonic Stem Cells

The isolation of proliferating mouse embryonic stem (ES) cells opened new possibilities for the study of gene function in transgenic animals. ES cells were derived from the inner cell mass of the mouse blastocyst and maintained in culture in the undifferentiated state. When re-introduced into a recipient blastocyst, they contributed efficiently to the formation of all tissues in a chimeric mouse, including the germline (Evans and Kaufman, 1981; Martin, 1981).

ES cells could be manipulated *in vitro* without losing their ability to contribute to the germ line of chimeras. Robertson et al., (1986) demonstrated that ES cells could generate germline chimeras even after transfection with the neomycin resistance gene and selection for G418 resistant cells. Similarly ES cells which had been selected for spontaneous mutation of the HPRT gene by growth in media containing 6 thioguanine were shown to be able to generate germline chimeras (Hooper et al., 1987).

1.1.5.3 : Specific Modifications can be Made to the Murine Genome by Gene Targeting in ES cells

The first targeted disruption of a murine gene was performed by Thomas and Capecchi (1987). These workers targeted the endogenous HPRT gene in ES cells. This gene is on the X chromosome, therefore, only one mutant copy was required to generate the HPRT⁻ phenotype in male ES cells. ES cells harbouring the disrupted HPRT gene could be selected by growth in the presence of 6-thioguanine (6T-G, the cytotoxic base analogue).

The experiments were executed using two types of gene targeting vector: the replacement vector and the insertion vector. Replacement vectors recombine into the genome by homologous double cross-over or gene conversion resulting in the replacement of endogenous DNA by exogenous sequences (see figure 1 A). By contrast insertion vectors recombine into the genome by a crossover resulting in the insertion of the entire vector sequences into the endogenous locus. This type of vector results in direct repeats at either side of the selectable marker (see figure 1 B).

In these experiments the selectable marker was the neomycin resistance gene and cells surviving in the presence of G418 and 6-TG were found to have lost HPRT activity due to targeted disruption of the HPRT gene. Unfortunately unlike the HPRT gene most cellular genes are not selectable. Targeting of non-selectable endogenous genes requires screening of large numbers of cells for novel endogenous/vector junctions by either Southern blotting or PCR. As previously discussed the frequency of homologous recombination in mammalian cells is low and screening for homologous recombination in events is a significant undertaking.

Using ES cell technology it has been possible to introduce null mutations in endogenous mouse genes by integration of a selectable marker into the gene of interest. Many hundreds of so called 'knockout' mice have been generated to test the function of endogenous genes. A recent comprehensive list of knockout targeting experiments has been produced by Brandon et al., (1995).

A double replacement technique has been used to replace the mouse α -lactalbumin (α -lac) gene with its human counterpart (Stacey et al., 1994). In this work the entire mouse gene was replaced with its human counterpart leaving no vector sequences at the target locus. The work was performed in HPRT-deficient ES cell line (HM-1) which remain totipotent in culture in the absence of feeder cells (Magin et al., 1992). The use of these cells permitted the exchange of the mouse α -lac sequences for the HPRT gene followed by a second replacement of the targeted HPRT gene with the human α -lac sequences. This approach exploits the ability to select first for and then against HPRT expression in the HM-1 cell line.

This type of approach could be used to study transcriptional regulation of endogenous mouse genes by the introduction of subtle mutations in the regulatory regions or protein structure and function by the introduction of amino acid changes through mutation of the protein coding regions.

Figure 1.1

The figure shows the differences between insertion and replacement targeting vectors. sm = a selectable marker (e.g. the neomycin gene resistance (neo^r) gene) that can be selected for in cultured cells. Insertion of the selectable marker into an essential part of the target gene inactivates it.

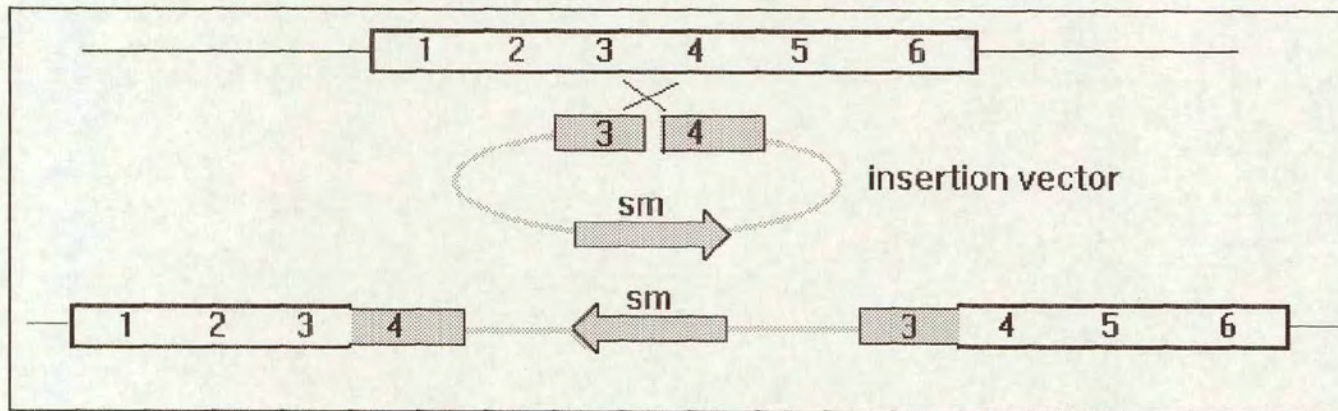
A : Insertion Vectors

The figure gives a diagrammatic representation of insertion of an insertion vector into the genome via homologous recombination. Insertion vectors are designed such that the regions of homology at the ends of the linearised vector are adjacent to each other in the genomic target. Homologous recombination occurs via a single crossover event resulting in duplication of the regions of homology flanking the site of insertion

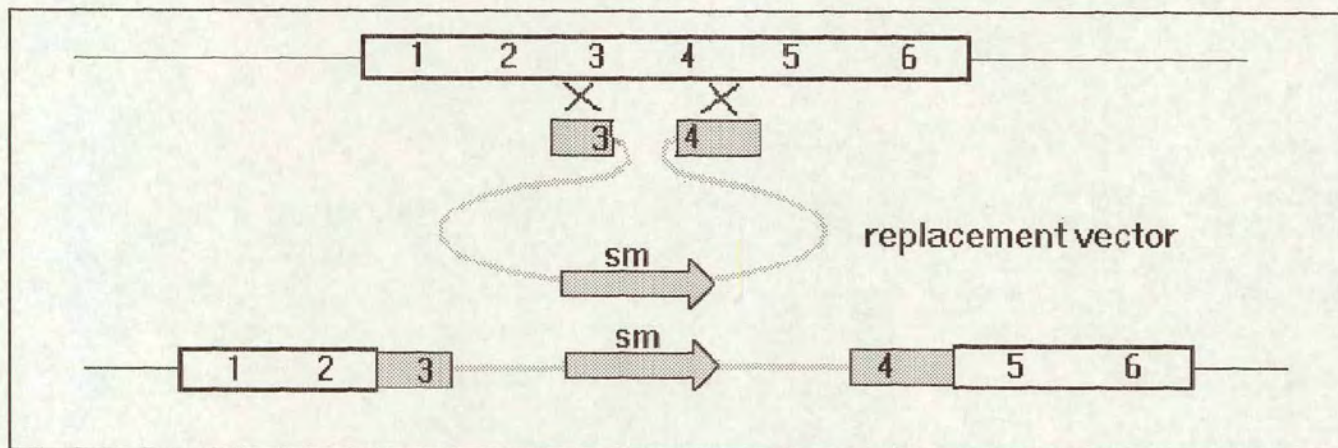
B : Replacement Vectors

The figure shows a diagrammatic representation of insertion of a replacement vector into the genome via homologous recombination. The regions of homology at either end of the linearised vector are not adjacent to each other in the genomic target and insertion occurs via a double cross over or gene replacement event resulting in insertion of the entire vector at the target site. Unlike insertion vectors the targeting event occurs without causing duplication of the genomic sequences flanking the insertion site.

A



B



1.1.5.4 : The Efficiency Of Producing Transgenic Mice By Homologous Recombination

The use of a single selectable marker does not differentiate between random insertion of the exogenous DNA and insertion via homologous recombination. Because of the low frequency of homologous recombination that occurs in ES cells strategies have been devised to enrich for homologous recombination events.

Positive negative selection (PNS) involves the use of two selectable markers (see Capecchi, 1989 for a review). One of which can be selected for (positively selected) and the other which can be selected against (negatively selected). The vector must be of the replacement type (see previous section and figure 1 A). The positive selection marker is placed between the arms of homology to the target gene (figure 1.2 A). Thus upon homologous recombination into the genome it is retained., whereas the negative selectable marker is placed outside the arms of homology and is lost in homologous recombination events (figure 1.2 A). By contrast on random insertion (by non homologous recombination) into the genome the negative selectable marker is retained (figure 1.2. B), resulting in cell death when this marker is selected against.

A 2000 fold enrichment for ES cells containing targeted disruptions of the HPRT and int-2 genes over the total number of neomycin resistant cells was reported by Mansour et al (1988) using PNS. However other workers have reported considerably lower levels of enrichment using this type of approach (Li et al., 1992, Stacey et al., 1994). Nevertheless, adoption of a positive negative selection system significantly enhances the ability to recover targeted events.

Figure 1.2 : Positive Negative Selection For Homologous Recombination in Gene Targeting Experiments

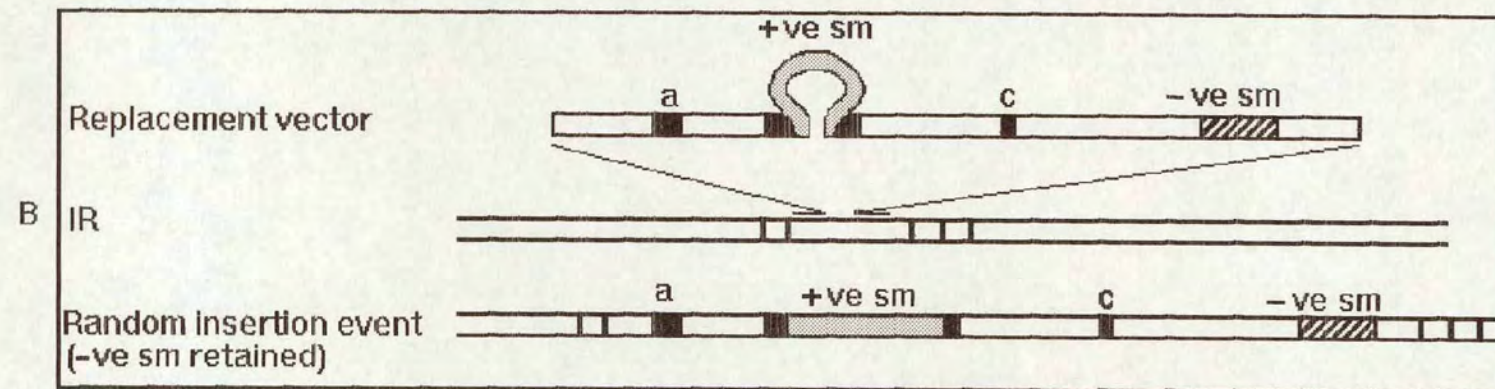
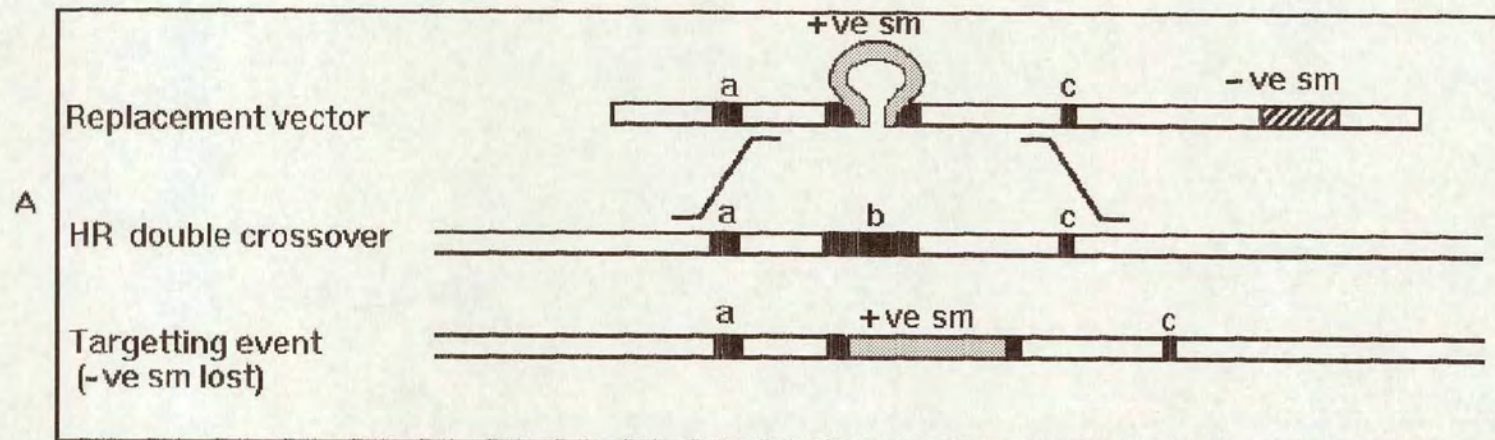
The frequency of homologous recombination events in mammalian cells is low compared with random insertion events. Positive negative selection has been employed by some workers to enrich for cells that have integrated the targeting vector via homologous recombination. The approach uses two selectable markers, one that can be selected for and another that can be selected against. +ve sm = a selectable marker that can be selected for (e.g. neo^r), - ve sm = a selectable marker that can be selected against (e.g. HSV-tk). The vector is always of the replacement type. The positive selectable marker is placed within the arms of homology to the target genomic locus and the -ve selectable marker is placed at one end of the vector outwith the regions of homology.

A : Homologous Recombination

The figure shows integration of a replacement vector designed for PNS into the genome by homologous recombination (HR). The targeting event occurs via a double cross over resulting in the homologous regions of the vector and all sequences between them being inserted into the genome. Since the negative selectable marker is not flanked on both sides by sequences homologous to the genomic target it is lost. Cells that have undergone a targeting event thus survive negative selection.

B : Random Integration

The figure shows integration of a replacement vector designed for PNS into the genome by illegitimate recombination (IR). Illegitimate recombination occurs through the ends of the linearised DNA and results in the insertion of the entire vector into the genome. Both selectable markers are present and cells that have undergone integration of the vector by this mechanism should not survive selection against the negative selectable marker.



1.1.5.5 : ES Like Cells of Large Animals Could Have Applications in Biotechnology

The ability to insert transgenes at a known locus in large domestic animals has many advantages. The availability of large animal embryos is limiting, and the technique of pronuclear injection is considerably less efficient in farm animals than in mice. Thus the ability to maintain targeted ES cells in culture affords the thorough analysis of positive ES cell colonies prior to blastocoel injection and implantation. By targeting transgenes into a specific permissive site position effects that repress transgene expression could be minimised. However to date the only true embryonic stem cell that has been isolated is that of the mouse. Despite intensive research a true embryonic stem cell has not been isolated from any other mammal (Wilmot and Campbell, 1994, Clark, 1996).

Cell lines that resemble embryonic stem cells have been isolated from sheep (Clark et al., 1992). This cell line (designated TNT 4) was obtained by microdissection and explantation of the embryonic discs of early sheep embryos. Unlike mouse ES cells these cells have a flattened morphology and express markers associated with differentiation such as cytokeratin, and nuclear lamin A/C. Recently, however, live lambs have been obtained by nuclear transfer from the TNT 4 cell line into enucleated sheep oocytes (Campbell et al., 1996).

It has not yet been demonstrated that these cells can be genetically manipulated in culture as mouse ES cells can be. However it is possible that domestic animals harbouring targeted transgene insertions will be available sometime in the foreseeable future using this and/or similar cell lines and nuclear transfer technology.

1.1.5.6 Summary

Transgenic animals that carry foreign DNA in their genome as a result of experimental manipulation have now been generated by a number of techniques. Insertion of genes using retroviral vectors allows one single copy of the transgene to integrate randomly into the genome. However this method often results in multiple integration sites and the amount of DNA that can be inserted is limited. Retroviruses also exhibit species specificity which also limits the use of this approach. For these reasons among others this method is not widely used for the generation of transgenic animals.

Pronuclear injection is currently the most commonly used technique for the generation of transgenic animals. Large transgenes can be introduced to the germline using this technique. Unfortunately the efficiency of generating transgenic animals using this technique is low, particularly in livestock. The injected DNA generally integrates at a single site and the number of integrated transgenes varies from one to hundreds of copies. Like retroviral insertion the site of integration is thought to be random. Many transgenes exhibit "position effects" which are thought to result from the influence of local chromatin structure on the integrated transgenes.

With the development of mouse embryonic stem cells it has become possible to target the insertion of transgenes to a particular genomic site via homologous recombination. This technology allows precise mutations to be made in the mouse genome and offers great potential for the study of gene regulation and protein function. However to date targeted mutation of the genome is only possible in the mouse.

1.2 THE MECHANISM OF TRANSGENE INTEGRATION

1.2.1 : Characteristics Of Integrated Foreign DNA

Outlined previously were some of the characteristics of transgenic animals produced by pronuclear injection. Analyses of the transgene loci indicate the integration of a single transgene is a rare event and transgenes are generally integrated as an array. The number of transgenes within the array varies from two to several hundred and they are normally orientated in a head to tail manner (e.g. Brinster et al., 1981; Palmiter et al 1982). These characteristics have also been demonstrated for the loci of foreign DNA inserted into the chromosomes of cultured cells that have been stably transfected by a variety of techniques (calcium phosphate co-precipitation, DEAE-dextran treatment, and direct microinjection). Thus it has been proposed that the mechanisms underlying the integration of foreign DNA in both systems are the same (Bishop and Smith 1989). The mechanisms by which the arrays of foreign DNA are built up and the mechanisms involved in chromosomal insertion have been the subject of numerous studies. Based on these studies mechanisms for both concatamer formation of the foreign DNA and integration of the DNA into the host genome have been proposed (Bishop and Smith, 1989, Wall and Seidel 1992).

1.2.2 : Fate Of Experimentally Introduced DNA In Mammalian Cells

Since the predominant arrangement of transgenes in integrated arrays is head to tail it has been proposed that the arrays must be established by homologous recombination between the injected molecules (Palmiter et al., 1982). Homologous

recombination occurring between similar linear fragments would only regenerate linear fragments; yet linear input DNA has been found to integrate more efficiently than circular DNA and has been demonstrated to integrate as concatamers (Brinster et al., 1981). These concatamers cannot be formed from linear fragments unless they undergo changes within the cell, following injection (Brinster et al., 1981; Palmiter and Brinster, 1985; Bishop and Smith, 1989).

Many studies have been performed that have investigated the fate of foreign DNA after entry into the cell. The studies have been performed mainly in culture cells and generally using viral genomes to characterise the changes that occur in the DNA upon transfection, electroporation and microinjection. These studies have been extensively reviewed by Bishop and Smith (1989).

That circularisation of linear input molecules occurs upon entry into cultured cells has of course been established using the SV40 viral genome. The life cycle of the SV40 virus requires that the viral genome be in the form of a circular molecule. If the viral genome is linearised prior to injection then the SV40 virus is still capable of eliciting infection demonstrating that the circularisation must have occurred following transfection (Wake et al. 1985)

Southern and Berg, (1982), performed a study in which they introduced a plasmid pSV2neo (comprising pBR322 and SV40 sequences with an inserted neomycin resistance gene: into COS7 cells. Transfections were performed with linearised plasmid and supercoiled plasmid. The results demonstrated that the cells transfected with linear plasmid now harboured populations of supercoiled DNA linear DNA and open circle DNA. It was evident that circularisation of the linear input DNA had occurred within the cells. Restriction of the nuclear DNA with two enzymes,

demonstrated that some of the input DNA was still in the original linear form. The importance of these studies is that they demonstrate that circularisation of input DNA does occur upon entry to the cell.

1.2.3 : Efficiency Of Circularisation Of Foreign DNA

The efficiency of circularisation of introduced DNA following transfection or microinjection was performed by Kopchick and Stacey (1984). In this study the avian sarcoma virus was used. The avian sarcoma virus has viral genes *gag*, *pol*, *env*, and *src* flanked by LTR's. The virus molecule was cloned into the pBR322 plasmid. The plasmid and virus was introduced by transfection or microinjection in three states (i) circular (ii) linearised within the plasmid sequences and (iii) linearised within the *pol* sequences of the viral DNA. The molecules were introduced into cultured avian cells. The circular molecules and those linearised within the plasmid sequences were capable of yielding full length viral mRNA initiated and terminated at the LTRs. Molecules linearised within the *pol* sequences of the viral genome would only give rise to the full length viral genome if ligation occurs. Mutation of the *pol* region inactivates the gene, which prevents virus infection.

Cytoplasmic microinjection of these three types of molecule gave similar virus titres indicating that the *pol* gene had been efficiently and accurately ligated within the cells. By comparison, transfection of the three differently treated molecules demonstrated that the virus titres of the circular molecule and those linearised within the plasmid sequences were similar but those linearised within the *pol* viral sequences

gave a 95% reduction in activity. This was found not to be the result of inefficient ligation but rather the introduction of mutations over several base pairs upon ligation.

Thus mutation occurred upon ligation of transfected molecules whereas microinjected molecules linearised within the *pol* sequences did not undergo mutation. Analysis of a molecule linearised within the *pol* gene using restriction enzymes with non complementary ends prevented the production of functional *env* mRNA which indicates that single stranded protruding ends of microinjected molecules need to be complementary in order to undergo efficient ligation within microinjected cells. In contrast transfected molecules treated in the same way were ligated regardless of their structure

Numerous other studies utilising the SV40 viral genome have demonstrated clearly that ligation of linear input DNA occurs within cells (Wilson et al., 1982; Wake et al., 1984; Roth et al., 1985). The study by Kopchick and co-workers (1984) in particular, indicates that this ligation event is particularly accurate and efficient when the input DNA is microinjected and has compatible ends. In the experiment the DNA was microinjected into the cytoplasm of mammalian cells in culture. By contrast production of transgenic animals involves the injection of DNA directly into the male pro-nucleus. Palmiter et al., (1982) reported the circularisation of linear input DNA following nuclear injection in mouse eggs. It may be that DNA injected directly into the nucleus of embryos undergoes the same processes as DNA injected into the cytoplasm of cultured mammalian cells.

1.2.4 : The Mechanism Of Concatamer Formation

Analysis of transgene arrays in transgenic mice have clearly shown that the arrays are predominantly arranged in a head to tail orientation. The relative paucity of head to head and tail to tail orientations lead to the proposal that the mechanism of concatamer formation must be via homologous recombination (Brinster et al., 1981; Palmiter et al., 1982)

Bishop and Smith (1989), proposed that the concatamers are formed by homologous recombination between either circularised input molecules and linear molecules, or between circularly permuted molecules and linear molecules. Circularly permuted molecules are molecules formed by random breakage of circularised input DNA (see figure 1.3). If this is so then the transgene-transgene junctions would be the self ligated ends of the input molecule and the integrity of the transgene-transgene junction will depend on damage sustained to the ends of the input molecule before self ligation occurs. Studies outlined in the two previous sections (Kopchick and Stacey, 1984; Palmiter et al., 1982) have demonstrated that circularisation of the input DNA occurs upon introduction of linear DNA to the cell or embryo, Finally, the nature of the ends of the input molecules may have important consequences for the efficiency of this process and thus the integrity of the transgene-transgene junctions.

1.2.5 : Studies Of Integrated Transgene Loci

Clues to the mechanism of concatamer formation and integration can be gleaned from detailed analysis of integrated transgene arrays. A number of detailed studies of integrated transgene loci have been performed. Three studies have shown

the regeneration of a restriction site between transgenes in arrays present in transgenic animals (Palmiter et al., 1982; Yamamura et al., 1985). The transgene ends in each cases were compatible. This is consistent with the efficient self ligation of microinjected DNA.

A study of seven transgene-transgene junctions obtained by PCR amplification of the junctions from two concatameric integrants of the RSVCat gene demonstrated the transgene ends at junctions were missing 1 to 62 bases (Rohan et al. 1990). It has been demonstrated that compatible ends are required for the efficient ligation of microinjected input DNA in cultured cells (see section 2.3) The ends of the injected transgene in the Rohan study were non compatible and minimal exonuclease activity appears to have occurred at all transgene-transgene junctions suggesting the efficient self ligation of microinjected DNA is homology dependent.

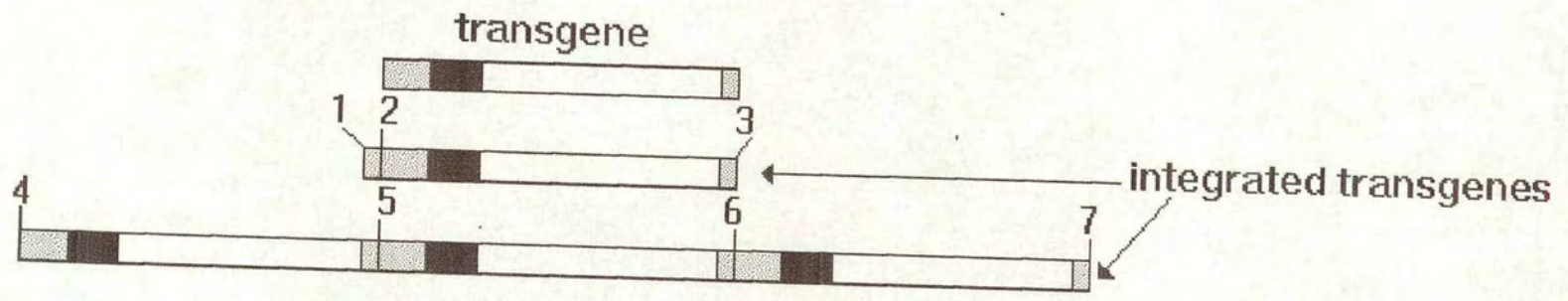
Two other transgene loci have been cloned and the fine detail of the junctions both transgene-transgene and transgene-chromosomal have been determined (Hamada et al., 1993). In total, three transgene-transgene junctions were analysed in this study, all of which were virtually intact. Two nucleotide deletions occurred in two of the three junctions. The transgene in this study was restricted using Pvu1 and was blunt ended prior to injection. As in the Rohan study, the transgene ends lacked homology. It cannot be ruled out that the lack of homology at the transgene ends reduced the efficiency of self ligation of the microinjected transgene and was causative in the two nucleotide deletion that occurred at two of the transgene-transgene junctions.

In the Hamada study four transgene-chromosomal junctions were analysed (see figure 1.3 for maps of the integrants). Two integrants were analysed, each had two transgene-chromosome junctions. Only one of the four transgene-chromosome

junctions occurred well within the transgene (Junction 1 Figure 1.3). The other three junctions occurred within four nucleotides of the ends of the transgene. The maps of the integrants are given in Figure 1.3.

Figure 1.3 : Structure of two Integrated Transgene Loci

The above diagram is a representation of the two transgene integrants analysed by Hamada et al., (1993). The upper diagram represents the transgene used to generate the transgenic mice. Below this is a representation of the first transgene integrant which comprises one nearly intact copy of the transgene and at the 5' end of this is 233-235 bp of the 3' end of another transgene. It has two transgene/genome junctions (1 and 3), and one transgene/transgene junction (2). The second integrant contains three transgenes. It has two transgene/transgene junctions (5 and 6) and two transgene/genome junctions (4 and 7). Only minor deletions of up to 3 nucleotides had occurred at all three transgene/transgene junctions. Three of the four transgene/genome junctions occur within 1-4 nucleotides of the 5' or 3' end of the injected transgene ends. (Adapted from Hamada et al. 1993)



The authors propose the pre-integration concatamer is linear since the chromosome/transgene junctions occur mainly at the ends of transgenes. A circularly permuted concatamer would generate junctions at random positions within the transgene. They propose that the linear concatamer is built up by the homologous recombination between linear and circular molecules. They also propose that the small nucleotide deletions occurring at the transgene-transgene junctions occurred by limited exonuclease activity at the ends of the transgenes prior to self ligation.

1.2.6 : Concatamer Integration Into The Host Genome

It has been proposed the integration of transgenes occurs randomly at chromosomal break points (Brinster et al.,1985). Limited information is available about the pre-integration site of transgene loci, and much of the available information comes from studies where the phenotype of the transgenic mouse has been abnormal. Many of the flanking genomic sequences that have been studied have undergone rearrangement of some form, including insertion, duplication and deletion. In fact to date no transgene locus has been demonstrated to be a straightforward insertion of a transgene array into the host genome. No extensive homology has been demonstrated to occur between the host genomic sequences of transgene loci and the transgene sequences. On studying two transgene loci Hamada et al.(1993) propose that the integration of the transgene array into the host genome occurs by end ligation of the transgene concatamer and chromosome breakage which may be stabilised initially by limited 2 to 3 nucleotide homologies that occur between the transgene and genome ends

1.3 : POSITION EFFECTS IN TRANSGENIC ANIMALS

1.3.1 : Position Effects

A transgenic experiment involves the generation of several lines harbouring the same transgene. Each transgenic founder animal is the result of a unique set of molecular events. The copy number of the transgene and the position of insertion into the genome will differ for each founder animal. Expression of the transgene is usually similar in all individuals of a given line but expression between lines varies. The range of tissues in which a given transgene is expressed is generally constant throughout the lines generated but the level of transgene expression more often than not bears no relationship to the transgene copy number.

The variability in the expression levels of the same transgene in different lines is thought to be the result of the effects of the neighbouring chromosomal DNA at the integration site. For many transgenes the expression level of the transgene is poor; and it is thought that the chromatin structure of the host genome at the region of insertion represses the transgene promoter (Palmiter and Brinster, 1986; Wilson et al., 1990). Such effects, where the chromosomal location of a transgene is thought to modulate its expression are termed "position effects". Position effect has been used to explain poor levels of expression, aberrant expression and variability of expression within a single line. In most cases although the assumption is reasonable our understanding of these effects is lacking.

Only one study to date has formally demonstrated that the aberrant expression in a line of transgenic mice was the result of a position effect. Transgenic mice

harbouring a transgene with the MUP promoter, and HSV TK coding region show high levels TK expression in the liver, the testes, and the preputial gland (Al-Shawi et al., 1990). Of five lines generated only one did not express the transgene in the preputial gland. The locus of this line was cloned and used to generate secondary transgenic mice. Very little surrounding genomic DNA was included in the secondary transgene. All of the secondary transgenic lines generated expressed the transgene at high levels in the liver, testes and preputial gland demonstrating that the lack of preputial expression in primary transgenic line was a consequence of its genomic environment and not the result of a mutation in the transgene.

1.3.2 : Position Effect And Transgene Methylation

It has been proposed that the methylation of DNA in animals is involved in general repression of transcription (for a review see Ceder, 1988). Methylation of genomic DNA occurs on cytosine residues at CpG sequences. Regions of DNA within the genome have been shown to be particularly rich in such CpG methylatable sequences; these regions have been termed CpG islands (Bird et al., 1985). The distribution of methylated cytosine residues in DNA undergoes characteristic changes during differentiation of a cell lineage and the loss of methylated cytosines from promoter regions correlates with activation of many tissue specific genes during differentiation in vivo (Ceder, 1988, Jaenisch and Jahner 1984). Methylation of promoter regions of endogenous genes on the inactive X-chromosome has been proposed to have a role in repressing gene expression, the corresponding allele of the active X chromosome is found to be unmethylated (Swain et al., 1987). However the role of methylation in transcriptional repression of genes is not clear. The mechanisms

of *de novo* methylation are poorly understood and there remains some debate as to whether methylation is a cause of gene repression or a consequence of the same.

However transgene repression has been demonstrated to correlate with methylation in a number of cases (Hadchouel, et al., 1987; Allen, et al., 1990). The transcriptional state of the transgene in these studies has been shown to be dependent on the parental origin of the transgene. The activity of the transgenes have been shown to correlate directly with the methylation status, in that methylation of the transgene inactivated the transgene. McGowan and co-workers (1989) have studied a line of transgenic mice harbouring a lac Z reporter transgene. This transgene was shown to exhibit cellular mosaic expression and the level of mosaicism was shown to correlate inversely with the transgenes methylation.

1.3.3: Transgenes And Position Effect Variegation

The phenomenon of position effect variegation (PEV) has been widely studied in *Drosophila* (Henikoff, (1994) and references therein). This effect is typically associated with chromosomal rearrangements that place a gene, normally positioned within euchromatic chromatin, in close proximity to heterochromatin. Heterochromatin is normally associated with and limited to specific regions of the genome, predominantly the centromeres and telomeres. Centromeric and telomeric DNA is rich in satellite sequences, and the non histone protein composition of these regions differs from the rest of the chromatin. Much of the DNA within heterochromatin is thought to be transcriptionally inactive. It is not known what mechanisms delimit the regions of constitutive heterochromatin but it is clear from

studies in *Drosophila* that a germline translocation event may allow the spread of heterochromatin into regions which are normally euchromatic causing their repression. This repression is not exhibited in all cells and the transcriptional state of the gene varies in individual cells, giving rise to a variegated phenotype.

The variability in transgene expression between lines of transgenic mice harbouring the same construct has not been usually studied at the level of the individual cell. Two possibilities exist for the variation in transgene expression in a given tissue. Firstly the rate of transcription of the transgene from the promoter varies between the lines. Secondly that the level of expression of the transgene is the same for each copy but the number of cells within the tissue expressing the transgene differs between the lines. Such heterocellular expression of a transgene has been demonstrated in transgenic mice harbouring a transgene encoding β -galactosidase driven by the promoter and upstream regulatory regions of the α -globin gene (Robertson et al.1995). This study demonstrates clearly the heterocellular pattern of expression normally associated with PEV but does not investigate the underlying cause of the variegated expression.

The BLG transgene is normally expressed in a predictable position independent manner and at stable levels within a given line. However two lines of mice harbouring this transgene demonstrated variability of transgene expression within individuals of the line. In situ hybridisation of metaphase chromosome spreads has demonstrated that the transgene array in this instance has integrated close to the centromere (Dobie. et al., 1996). In situ hybridisation of mammary tissue from individuals of these lines demonstrate variegation of transgene expression within the tissue. In the light of their results and that of others Dobie and colleagues have suggested that both the integration

site and large copy number of some transgene loci may pre-dispose transgenes to PEV resulting in variable expression within individual transgenic lines. The frequency of PEV occurring at transgene loci is not known but it may have important implications for the application of transgenic technology.

1.3.4 : Positive Position Effects

Transgenes generally comprise a cell type specific promoter and either homologous or heterologous protein coding sequences. Generally transgenes are expressed in the target tissue but often they are also expressed in tissues in which the promoter is not normally active. The ovine BLG gene (Farini and Whitelaw, 1995) and human β -globin gene exhibits such ectopic expression (Blom van Assendelft et al., 1989). This type of expression is referred to as positive position effects and is thought to be a position dependent phenomenon, because only a percentage of transgenic lines harbouring the same transgene exhibit ectopic expression. Positive position effects can arise as a result of integration close to an enhancer type element or from a lack of repression. However, this type of low level aberrant expression of transgenes has not been studied in detail.

1.4 : ELEMENTS INVOLVED IN GENE AND TRANSGENE REGULATION

1.4.1 : Elements other than the Basic Promoter are Required for Appropriate Transcription of Eukaryotic Genes

Before the advent of transgenic animals many of the basic features of genes involved in transcription had been elucidated. Promoter regions had been detected and

sequences within them required for polymerase binding had been defined. The discrepancy in size between precursor RNA's and the mRNA's had led to the discovery of the intron/exon architecture of eukaryotic genes and thus to the machinery involved in RNA splicing. The transcription termination and poly A site had also been delineated.

The analysis of transgene expression has contributed to the elucidation of additional elements involved in gene regulation. Many early transgenes were expressed poorly. At least two reasons for this have been identified. Firstly, the lack of sufficient regulatory sequences within the injected DNA and, secondly, the repression of transgene expression by the chromatin structure of the neighbouring genomic DNA. The latter reflects position effects and is discussed earlier in the introduction to this thesis.

A number of transgenic animal experiments have demonstrated the expected tissue-specific (e.g. The α -globin gene cluster, Sharpe et al., 1993) and temporal profile of expression (e.g. the BLG gene, Simons et al 1987) and freedom from position effects (e.g. The β -globin gene with LCR elements, Grosveld et al 1987). Analysis of such transgenes has elucidated the function a number of elements involved in gene regulation. These elements are found both within and outwith the transcription unit, some of them are many kilobases away from the transcription initiation site of the gene (see sections 1.4.2 - 1.4.6 for discussion).

1.4.2 : The Role Of Introns In Transgene Expression

Prokaryotic genes comprise uninterrupted stretches of coding DNA sequence. By contrast most eukaryotic genes have protein coding regions (exons) that are interrupted by non coding DNA sequences (introns) (Breathnach et al., 1977 and Jeffreys and Flavell 1977). Eukaryotic genes are also transcribed from the transcription initiation site to the termination site giving rise to a precursor RNA. The final messenger RNA (mRNA), however, is only a fraction of the size of the primary transcript as it contains only the protein coding sequences (exons) plus additional 5' and 3' untranslated regions (UTRs). All other sequences (the introns) are removed during nuclear RNA processing (splicing). The basic machinery involved in RNA splicing has been elucidated (for a review see Watson et al., 1987). Sequences at exon intron junctions are usually well conserved and consensus sequences have been determined for splice donor and splice acceptor sites (Breathnach and Chambon, 1981). Splicing progresses by formation of an intermediate termed a lariat (Keller, 1984) and the formation of this intermediate involves sequences 20- 55 nucleotides upstream of the splice acceptor site (reviewed by Padgett et al., 1986). Many components are required to form an active splicing complex (spliceosome) including U1 and U2snRNPs (Black et al., 1985). The processes of nuclear RNA processing remain the focus of intensive research efforts. Nevertheless, there are still many aspects of intron splicing that are not understood.

Reed and Maniatis (1986) have shown that sequences adjacent to the splice site can affect the accuracy and efficiency of splice site selection. It has also been demonstrated that the selection of splice sites is dependent on their location within the

pre-mRNA (Nelson and Green, 1988). Pre-mRNA's are likely to have defined structures and the context of the splice site may influence the nuclear splicing process.

Yull et al., (1995 (a)) have observed aberrant splicing of RNA transcribed from a transgene expressing factor IX from a cDNA construct driven by the BLG promoter. The RNA was properly processed in the liver (the normal site of factor IX production) but was spliced using a cryptic splice site in the mammary gland. These authors suggest that there may be tissue specific differences in the trans-acting factors involved in nuclear RNA processing.

It has been demonstrated that for some genes introns carry important regulatory elements. An enhancer region has been found in intron 2 of the apolipoprotein B gene (Brooks et al 1994) and intragenic regulatory regions have been reported in numerous other genes; these include human growth hormone gene (Moor et al., 1985), rat growth hormone gene (Sap et al., 1990), PDGF-B (Franklin et al 1991).

The ovine BLG gene appears to require introns for efficient expression. The entire gene was expressed efficiently in the mammary gland of mice and appeared to be expressed in a position independent and copy number dependent manner (Whitelaw et al., 1992). However intronless BLG transgenes were much less efficiently expressed (Whitelaw et al., 1991). Less than 50% of the transgenic lines expressed BLG and levels were twenty fold reduced on that of the entire gene. This suggests the introns of BLG harbour important regulatory elements or serve some structural function for the efficient transcription of the gene.

Many proteins are transcribed from very large genes (e.g. factor VIII, Davie 1986) or genes which have not yet been cloned. The cDNA for such genes can more

readily be obtained and thus used experimentally. Transgenic animal experiments utilising cDNA sequences attached to a heterologous promoter have not usually been very successful (Simons et al., 1988; Shani et al., 1992; Velandar et al., 1992). A number of experiments have been performed where the inclusion of introns within a transgene have elevated expression or proven to be essential for expression of the transgene (Brinster et al., 1988; Buchman and Berg 1988; Archibald et al., 1990; Whitelaw et al., 1991). Heterologous introns have been inserted into transgenes and in some cases have been found to have a small positive effect on transgene expression (Palmiter et al., 1991, Choi et al., 1991) but this approach is not universally effective suggesting that the role of introns in transgene expression is more than a simple requirement to enter the RNA splicing pathway. The mechanisms underlying this effect are not fully understood.

1.4.3 Enhancers

Our initial understanding of gene regulation was obtained from studies of bacterial genes. The initial characterisation of eukaryotic promoters revealed similar arrangements of sequence elements to that discovered in prokaryotic genes. Many eukaryotic promoters possess the conserved TATAAT motif placed 20 to 30 bp upstream of the transcriptional start site. This motif is essential not for transcription per se, but for the accurate placement of RNA polymerase II at the transcriptional start site. Other promoter sequences found immediately upstream of the start site in prokaryotes have also been conserved in some eukaryotic genes. Such sequences immediately surrounding the start site were found to be sufficient to direct RNA polymerase to the start site and drive efficient expression of genes in prokaryotes. The

TATA box and upstream conserved sequences collectively specified the promoter (Hawley and McLure, 1983). Not surprisingly studies of eukaryotic gene regulation in cell culture demonstrated that the basic promoter, as defined in prokaryotes, was not sufficient to regulate the often complex expression of eukaryotic genes.

The transcription of viral genes in infected culture cells was the subject of the early studies of transcriptional regulation in eukaryotic cells. Viruses are unable to replicate DNA or synthesise protein outwith the host cell and thus rely almost entirely on host factors for the transcription of their genes and translation of their RNA. SV40 genes are transcribed efficiently in many culture cells and thus provided a model for eukaryotic gene transcription. Experiments using the SV40 virus demonstrated that efficient transcription of viral genes did not occur if 150 bp upstream of the basal promoter was deleted (Benoist and Chambon 1981).

Within this region a 72 bp element occurred twice. Deletion of one of the repeated elements did not adversely affect transcription of the viral genes, where as deletion of both 72 bp elements drastically reduced gene expression. These regions of DNA which were not part of the basic promoter but were nonetheless required for efficient transcription, were termed enhancers. Functional assays of viral gene expression in cell culture defined the characteristics of the viral enhancers (reviewed by Atchison, 1988).

It was found that viral enhancers could only increase transcription if they were present in *cis*. Nevertheless such sequences could exert their effect over distances of several kilobases. The orientation of the sequences with respect to the basal promoter elements did not appear to be important for them to exert their effect and enhancers were able to drive the efficient expression of genes other than those with which they

are normally associated. The orientation of the enhancer also appeared to be unimportant (Banerji et al., 1981).

Subsequently sequences which exert an enhancing effect on gene expression were also found in cellular genes. An enhancer element was found to be present in the immunoglobulin heavy chain gene (Banerji et al., 1983, Gillies et al., 1983). Efficient expression of the Ig heavy chain requires the locus to undergo rearrangement. This rearrangement event is B-cell specific and precedes the expression of the Ig heavy chain. Analysis of this phenomenon led to the discovery of an enhancer positioned between the (D) and (C) regions. Rearrangement of the gene locus brings a promoter of the (V) region to within 2.5kb of the enhancer resulting in elevated expression of V_{κ} .

In order to investigate the effect the enhancer had on a heterologous promoter the Ig heavy chain enhancer was linked to β -globin gene (Banerji et al., 1983). This gene is normally expressed in erythroid tissue not in lymphoid tissue. The expression of the gene was increased dramatically by the presence of the Ig heavy chain enhancer in cell lines of lymphoid derivation. Moreover the positioning of the enhancer with respect to the β -globin promoter was of no importance. Thus the Ig heavy chain enhancer can (i) enhance expression of heterologous promoters, (ii) enhance expression of a gene regardless of its position with respect to the promoter and (iii) exert its effect over long distances. These characteristics are consistent with those of viral enhancers. Importantly it only enhances expression in cells of lymphoid derivation.

Numerous other enhancers have now been identified in eukaryotic genes by their ability to increase transcription of cellular genes in transfection assays. Detailed

analysis of enhancers, both viral and mammalian, demonstrate they are composed of multiple genetic elements (modules), each of which binds one or more transcriptional activator proteins (Dyran, 1989 and references therein). The factors bound at the enhancer site are generally a combination of ubiquitous and cell specific DNA binding proteins.

Evidence is accruing that the binding of unrelated transcriptional activator proteins to enhancer modules is co-operative (Adams and Workman 1995). It is currently thought that the binding of tissue specific factors to the enhancer region initiates the binding of other DNA binding factors to sites in the immediate vicinity, in a co-operative fashion. The binding of the *trans* activating factors is proposed to displace nucleosomes and via this mechanism may alter the chromatin structure of the region making it more accessible to other *trans* acting factors. Experimentally this is demonstrated by a general increase in DNase I sensitivity and accessibility of DNA to restriction endonucleases upon factor binding to the enhancer. Currently it is thought that the proteins bound at different enhancer modules interact via protein-protein contacts to form a three dimensional structure which then interacts with RNA polymerase II.

The ability of enhancers to confer a cell type and temporally specific profile of expression upon a gene is likely to be the key to understanding the differential expression of genes throughout development. Thus experiments have been devised to identify enhancer elements and analyse the expression profile they bestow on reporter gene constructs harbouring a minimal promoter. Enhancer trapping experiments were first devised in *Drosophila* (O'Kane & Gehring, 1987) and have subsequently been adapted for the mouse (Gossler et al., 1989). This approach allows the detection of

enhancers which display an interesting expression profile throughout development, and the subsequent analysis of the genes with which they are normally associated.

It is likely that enhancer elements play a central role in the opening of a chromatin domain and activation of genes (Ellis et al., 1996). Nevertheless even in the presence of enhancer elements transgenes are often still poorly expressed in the whole animal. Thus it is clear that enhancers are necessary elements for the temporal and cell type specific expression of eukaryotic genes but may not be sufficient for copy number dependent expression of transgenes.

1.4.4 : Locus Control Region

Using various approaches a number of different types of element have been demonstrated to be essential for accurate and abundant gene expression. Nevertheless expression of transgenes even in the presence of enhancers was still unpredictable.

The first predictable expression of a transgene in transgenic mice was demonstrated by Grosveld and his colleagues (Grosveld et al., 1987). The group were studying the regulatory sequences of the human β -globin like gene cluster. This is a cluster of 5 genes 5' ϵ - γ^G - γ^A - δ - β 3' which span 70 kb in total. The genes are expressed throughout development in the order in which they occur in the genome.

Early transgenic experiments showed disappointing expression of the human β -globin transgene in mice (Magram et al., 1985, Townes et al., 1985, Kollias et al., 1986, 1987). The expression was low and highly variable. A group of genetic diseases termed the thalassemias are associated with mutations throughout the gene cluster (see Weatherall 1993 and 1995 for a review). The study of the Dutch $\gamma\beta$

thalassemia demonstrated that the disease was the result of a 100 kb deletion extending into the β -globin like gene locus (Taramelli et al 1986). This deletion did not include the β -globin gene or any of its immediate 5' regulatory regions. Despite the gene being intact no β -globin gene expression was obtained from the affected allele. DNase I sensitivity studies demonstrated that the chromatin of the affected allele was maintained in an inactive (closed) conformation and thus was not expressed (Kioussis et al., 1983). This suggested that sequences other than those immediately 5' and 3' of the β -globin gene controlled its activation.

Clusters of tissue specific DNase I superhypersensitive (HS) sites had been detected at each end of the locus 5' of the ϵ -globin gene and 3' of the β -globin gene (Tuan et al., 1985; Forrester et al., 1986). These sites were considered to be candidate sites for regulation of the β -globin like gene locus. The effect of their inclusion in a β -globin transgene was analysed in transgenic mice (Grosveld et al., 1987). The transgene comprised the HS cluster from the 5' end of the locus, the human β -globin gene and the 3' HS cluster from the locus. The profile of expression of this globin gene minilocus was very predictable. The transgene was consistently expressed in transgenic foetuses and related to the transgene copy number. The results of this experiment gave rise to the concept of the Locus Control Region (LCR).

The LCR is defined as a DNA sequence which confers transgene expression that is independent of its site of integration in the host genome, but dependent on its copy number. As with the definition of other gene regulatory elements this definition reflects the nature of the functional assay used. The mechanisms by which these sequences bring about this effect remain the subject of intensive research.

All four 5' HS sites have been cloned and sequenced. Footprinting analysis of the HS sites have demonstrated binding sites for numerous DNA-binding proteins, both ubiquitous and erythroid specific, in each of the HS regions (Talbot et al., 1991; Lowrey et al., 1992; Philipsen et al., 1993; Stamatoyanopoulos et al., 1995; Kotow and Orkin 1995). Functional analysis of constructs harbouring specific wild type 5'HS's have demonstrated that although two of the HS sites 5'HS2 (Ellis et al., 1993, Caterina et al 1994) and 5'HS3 (Fraser et al., 1990; Philipsen et al., 1993) show partial LCR activity, no single HS gives the full LCR effect. Both 5'HS2 and 5'HS3 can drive position independent expression of β -globin transgenes when more than one copy of the transgene is present, but only at a fraction of the level expressed by the full LCR (Grosveld et al., 1987, Strouboulis et al., 1992).

Since no single LCR element has been identified that drives high level, position independent and copy number independent expression of the β -globin transgene, it is thought that all four HS elements are required for the full expression of all genes in the locus. It has been proposed that the individual elements within the LCR combine to form a single large nucleoprotein complex that interacts with and thus activates each of the genes in the locus (Fraser et al., 1993). In this model, the promoters of the different genes in the locus would be in competition for the LCR (Behringer et al., 1990). By default, promoters closest to the LCR would be the most frequently expressed. Thus ϵ -globin would be the most frequently expressed gene. Down regulation of the more proximal genes is thought to occur by the sequential expression of repressors of the more proximal genes throughout development. Studies of some of the thalassemias suggested the presence of such repressors and more recently the DNA

binding proteins GATA-1 and YY1 have both been found to be repressors of the ϵ -globin gene (Raich et al., 1995).

The β -globin LCR has been shown to enhance the expression of the Thy 1 gene and tk-neo in MEL cells but not in non erythroid cells so the globin LCR confers its effect in an erythroid specific manner (Blom van Assendelft et al., 1989). This is not entirely surprising since some of the *trans*-acting factors bound to the HS sites have been demonstrated to be erythroid specific and essential for LCR activity (Kotow and Orkin, 1995). The LCR has no effect on the expression of transgenes in other non erythroid tissues.

The profile of expression conferred on transgenes by the globin LCR could also conceivably be produced by the presence of an insulator element. This possibility has been excluded by studying single copy insertions of the β -globin minigene in transgenic mice. The high level tissue specific expression of the single integrants was observed but positive position effects resulted in low levels of expression in other tissues demonstrating that the single copy transgenes are not insulated (Ellis et al., 1993). Moreover transgenes containing partial LCR fragments are capable of position independent expression (Fraser et al., 1990) and position effects are seen in β -globin constructs that do and do not contain the LCR. Also the generalised DNase sensitivity exhibited by the β -globin locus (Forrester et al., 1990) extends well beyond the regions used in the transgenic mice experiments. Thus the ability of the LCR to override repressive position effects, but not prevent positive position effects, suggests that the globin LCR functions by a dominant positive activation mechanism as opposed to providing a boundary function (Dillon and Grosveld., 1993).

The β -globin LCR regulates the expression of a multigene family. It determines (i) the erythroid specific expression (ii) developmental expression of the genes (gene switching) and (iii) the level at which the genes are expressed. Although the β -globin LCR is now the paradigm of LCR function it regulates the expression of a highly complex multigene locus and is thus unlikely to be typical of other regulatory regions that confer a similar LCR effect on less complex gene loci.

Other LCR like elements have been identified by their ability to produce high level, position independent, copy number dependent expression on genes expressed in transgenic mice. The human CDC2 gene exhibits copy number related expression in transgenic experiments (Greaves et al., 1989), as do the human ADA gene (Arnow, et al., 1995), β -myosin heavy chain gene (Knotts et al., 1995), mouse tyrosinase gene (Schedl et al., 1993), PEPCK gene (Cheyette et al 1992) and chicken lysozyme gene (Bonifer et al., 1990). The regulatory regions of all of these genes correlate with DNase I HS sites. The chicken lysozyme transgene harboured the putative domain boundaries which are thought to insulate the transgene from position effects. Subsequent experiments (Bonifer et al., 1994) demonstrated that the putative domain boundaries were not required for the position independent expression of the chicken lysozyme transgene but did appear to reduce the level of ectopic expression of the transgene in other tissues. No boundary elements have as yet been identified for any of the other aforementioned genes. However regulatory regions outwith the immediate HS region are required to confer the LCR effect on the ADA gene in transgenic mice (Aranow et al., 1995). It has been proposed by these workers that sequences other than the enhancer are present in the regulatory region which facilitate the enhancer

function. These regions have been termed “facilitators” but their mechanism of action is unknown.

1.4.5 : Scaffold/Matrix Attachment Regions

Microscopic examination of the lampbrush chromosome from the oocyte using phase contrast light microscopy demonstrated the chromosome consisted of a dense longitudinal axis from which loops of DNA emanated (Gall 1956, reviewed in Watson et al 1988). Visualisation of these structures led Gall and others to propose that a further level of DNA compaction (beyond that achieved by the sequestering of DNA on histone octamers) is brought about by the folding of the chromatin fibre into loops attached at their bases to a central support.

Further electron microscopy studies on histone depleted metaphase chromosomes of HeLa cells showed loops of DNA attached to a central protein scaffold. Similar findings were obtained from interphase nuclei. (Paulson & Laemmli 1977). These findings provided further support for the putative loop arrangement of DNA in chromosomes.

Studies were performed which began to shed light on the nature of the supporting structure of chromosomes. The nuclear matrix is the insoluble structure which remains after the treatment of isolated nuclei with DNase I and the extraction of proteins (the histone proteins in particular) using high salt solutions. It is the same size and shape as the nucleus, contains the nuclear lamina, remnants of the nucleoli and an internal proteinaceous network (Berezny and Coffey, 1974). The nuclear scaffold is obtained by treatment of isolated nuclei with a mild detergent lithium diiodosalicylate (Mirkovitch et al., 1984). This removes the histones and some non

histone proteins. Although similar to the matrix it differs in its overall protein composition (Belgrader et al., 1991).

In 1985 Berrios et al., established that topoisomerase II is a component of the nuclear matrix of *Drosophila* nuclei. Previously a protein of 170 kd had been identified as the major component of the metaphase scaffold in HeLa cells. (Lewis & Laemmli, 1982). This protein was subsequently demonstrated to be topoisomerase II. Immunolocalisation of topoisomerase II has shown that it is located within an element forming the longitudinal axis of the chromatid in both 'native like' and histone depleted chromosomes (Berrios et al., 1985).

Genomic DNA sequences that bind the nuclear scaffold (SAR's) were identified using the nuclear 'halo' mapping procedure (Mirkovitch et al., 1984). Matrix binding sequences in cloned DNA (MAR's) were identified by Cockerill and Garrard (1986). Both MAR's and SAR's were found to be AT rich stretches of DNA ranging from 100 bp to several hundred bp long.

These findings were consistent with the findings of Moreau et al (1982). These workers demonstrated the systematic punctuation of DNA with early melting zones which they termed 'A+T rich linkers'. These regions were AT rich, 800bp long (on average) and were clustered or spaced at intervals of 10 to 40 kb. To correlate this with function they analysed cloned genomic DNA from chick and *Drosophila*, and demonstrated that A+T linkers flanked the 15 kb chick α -globin gene cluster. The 15 kb of DNA within the linkers correlated well with extent of DNase I sensitivity previously established for this gene. The chick β -globin gene cluster was also framed 5' by a single A+T rich linker and a cluster of four such linkers at the 3' end. Similarly the ecdysone inducible P1 protein gene from *Drosophila* was shown to be

delimited by A+T rich linkers. It was proposed by these workers that the A+T rich linkers defined domains of the eukaryotic genome that may correlate with putative functional units. S/MAR's are AT rich sequences that have been shown to coincide with the boundaries of actively transcribed genes (Gasser and Laemmli, 1986) it may be that the AT linkers described by Moreau and colleagues are S/MAR's since they are AT rich and are located at the boundaries of genes.

It has been shown that genes are preferentially sensitive to DNase I digestion in tissues in which they are transcribed.(Weintraub and Gourdine, 1976; Phi-Van and Strätling, 1988; Levy-Wilson and Fortier, 1989). Chromatin which is sensitive to DNase I digestion is referred to as open chromatin, that which relatively refractory to DNase I digestion is referred to as closed chromatin. The region of DNase I sensitivity is not restricted to the coding region of active genes but extends to the flanking regions. The extent of general DNase I sensitivity of a number of genes has been mapped. The DNase I sensitivity of the chicken lysozyme gene has been rigorously studied in parallel with matrix binding studies. There is no sudden demarcation of the open and closed chromatin of the chicken lysozyme gene, rather the level of sensitivity gradually declines both in the 5' and 3' direction for some distance from the transcription unit of the gene, which is centrally placed within the DNase I sensitive region. The regions of the chicken lysozyme gene found to bind nuclear matrix co-localised with the 5'boundary of the DNase I sensitive domain (Jantzen et al., 1986 ; Phi-Van and Strätling, 1988). Similar findings were obtained with parallel DNase I and MAR mapping of the human apolipoprotein B gene (Levy-Wilson and Fortier, 1989). The finding that S/MAR regions are positioned on the

boundaries of open and closed chromatin of actively transcribed genes implied that S/MAR's may have a functional role in gene transcription.

The S/MAR regions of the chicken lysozyme gene have been called "A elements". The assessment of their possible function was carried out using a minidomain with the enhancer region and promoter of the chicken lysozyme gene driving the expression of a CAT (chloramphenicol acetyltransferase) reporter gene. Two minidomains were tested for expression in transfection assays. The mini-domain described above and another identical but with the 5' A element of the chicken lysozyme gene placed at each end of the construct. In transient transfection assays of chicken promacrophage cells no difference in expression was detected between the two constructs suggesting that the A element was not a classical enhancer. However in stable transfection assays the A elements significantly enhanced expression of the reporter gene over that of the minidomain lacking the A elements (Steif et al., 1989). The A elements conferred position independent expression of the reporter gene from the chicken lysozyme enhancer and promoter. The 5' A element was also capable of augmenting transcription of heterologous enhancer-promoter elements (Phi-Van et al., 1990; Sippel et al., 1992). Further examination of the effect of A elements on expression indicated that the A element could only augment expression of the gene if a functioning regulatory unit is present between the A elements (Sippel et al., 1992). The results of the transfection studies suggested that the A elements have an important role in the position independent expression of the associated gene.

McKnight et al (1992) tested the effect of the 5' A element of the expression of WAP transgenes in transgenic mice. The WAP transgene had been tested previously in transgenic mice and expression was observed in only 50 % of the lines at low levels

and its regulation did not mimic that of the endogenous gene (Burdon et al., 1991). The transgenic mice harbouring the A elements and WAP transgene were produced by co-injection of the A elements and the transgene since stable plasmid constructions could not be obtained. The A elements did not increase the level WAP expression in the lines but did allow low level WAP expression in all lines. The inclusion of the A elements also appeared to permit regulation of the WAP transgene throughout lactation in a manner that paralleled the endogenous WAP gene. The authors claimed that these results demonstrated the MAR could establish independent genetic domains in transgenic mice, but the expression of the WAP gene in these experiments was not copy number dependent indicating that the A elements did not fully insulate the WAP transgenes.

The role of the A elements in chicken lysozyme gene expression have also been tested in transgenic mice (Bonifer et al., 1990). The transgenic mice harbouring the entire chicken lysozyme gene including the 5' and 3' A elements expressed the transgene in a tissue specific manner. The expression levels were directly related to the transgene copy number in all of the founder animals. The study demonstrated that the chicken lysozyme gene locus behaves as a regulatory unit independent of its chromosomal position. However, further experiments (Bonifer et al., 1994) demonstrated that the removal of the A element did not significantly affect the position independent copy number dependent expression of the gene in transgenic mice and resulted only in an increase in ectopic expression. Therefore it has not been conclusively demonstrated that the 5' A element of the chicken lysozyme gene acts as an insulating element *in vivo*. SAR sequences from the human β -interferon putative boundary domains (Bode and Maass, 1988) have also been tested in transgenic mice

for their ability to act as boundary elements for a construct containing the murine HSP70.1 promoter linked to a firefly luciferase reporter gene (Thompson et al 1994). Neither elevated expression, copy number dependence nor elimination of position effects was observed in differentiated tissues suggesting that these sequences were unable to act as boundary elements. Currently there is no strong evidence for MAR's functioning as boundary elements *in vivo*.

A number of proteins have now been identified that bind to S/MAR's. These include; histone deacetylase (Hedzel et al., 1991), ARBP (von Kries et al., 1991), SATB1 (Dickson et al., 1992), SAF A (Romig et al., 1992), ATF transcription factor (Dworetzky et al., 1992) CHD1 (Stokes and Perry, 1995), nucleolin (Dickson and Kohwi-Shigematsu, 1995) and Bright (Herrscher et al., 1995). The functions of these proteins *in vivo* are as yet undetermined. Some of them have been demonstrated to be an integral part of the nuclear matrix (Histone deacetylase, ARBP, SAF-A and SATB1). Others have been found to be partitioned between the insoluble matrix fraction and soluble fraction of nuclear proteins (nucleolin and Bright) and CHD 1 has been shown to be associated with bulk chromatin rather than the nuclear matrix. ARBP and SAF-A are abundant nuclear proteins and for this reason have been proposed to have a structural role in the looping of DNA. SAF-A self polymerises and has been demonstrated, using electron microscopy, to form looped structures in DNA. SATB1 and Bright are tissue specific MAR proteins expressed in the human thymus and B cells respectively. The presence of tissue specific MAR binding proteins suggests that MARs have a role in gene transcription. Bright has a domain that shares identity with SWI1, a component of the SWI/SNF complex of *trans*-activating factors.

CHD1 contains both a chromatin organisation modifier (chromo) domain and a helicase/ATPase domain. The binding of such proteins which are capable of affecting local chromatin structure is consistent with MARs having a role in the control of local chromatin structure and consequently a role in transcriptional control.

1.4.6 : Non MAR Boundary Elements

Putative boundary elements were found at the 5' and 3' ends of the *Drosophila* 87A7 heat shock locus (Udvardy et al., 1985). These unusual nucleoprotein structures were named specialised chromatin structures or *scs* elements. The *scs* elements are not matrix bound. Each of the *scs* elements (*scs'* and *scs*) have two DNase I hypersensitive sites flanking a DNase I resistant core of 250-230 bp (Kellum and Schedl, 1991). The ability of the *scs* elements to insulate a transgene from position effects was demonstrated using the *Drosophila* white gene as a reporter (Kellum and Schedl, 1991). The white gene functions autonomously in each cell of the *Drosophila* eye. The level of white gene expression is reflected in eye colour which is easily measured and the white gene has been shown to be susceptible to position effects. Comparison of the white transgene flanked with the *scs* elements with the reporter gene that was not flanked with the *scs* elements clearly demonstrated that the *Drosophila scs* elements were capable of insulating a reporter transgene from position effects. These same workers then demonstrate further the ability of the *scs* elements to function as domain boundaries by an Enhancer blocking assay (Kellum and Schedl, 1992). A MAR sequence was unable to prevent the action of an enhancer when placed between the enhancer and the promoter elements suggesting that not all sequences that bind to the matrix can function as insulators. Studies are now being performed to

isolate proteins which bind to the *scs* elements. Zhao et al., (1995) have isolated a protein BEAF 32 which binds *scs*'. Ongoing work to identify further proteins which are associated with the *scs* elements may explain the molecular mechanism by which the elements exert their effect.

The 5' HS4 site of the chicken β -globin gene has also been proposed to act as a boundary element (Chung et al., 1993). This region of the chicken β -globin gene locus is characterised by a constitutive HS site in all tissues. The presence of a constitutive HS site at the 5' end of the β -globin locus appears conserved in human mice and chicken and it has been proposed to act as an insulator preventing the influence of the potent LCR spreading in the 5' direction. The element has not been demonstrated to bind to the nuclear matrix and it may be quite different from the MARs that have been proposed to function as boundary elements. Chung et al performed an enhancer blocking assay in erythroleukemia cells to demonstrate the HS site functioned as a boundary and found it insulated a reporter gene from the influence of a nearby LCR, when positioned between them. They also assayed the ability of the HS site to insulate the white gene in *Drosophila* and found that the element insulated the gene from position effects. However a single HS site was not able to provide complete insulation in either of these experiments. Moreover the element has not been tested in transgenic mice which would appear to be the most rigorous system for testing the function of putative boundary elements.



1.5 : TRANSGENIC ANIMALS IN THE PRODUCTION OF BIOMEDICAL PROTEINS

Many human diseases arise from a lack of functional proteins which fulfil important physiological roles. For such diseases simple replacement therapy can alleviate the symptoms of the disease and in some instances slow its progression. Widespread replacement therapy of such physiologically active proteins requires adequate volumes of the protein be synthesised and purified at low cost. For proteins required in high dosage (eg. human α_1 -antitrypsin) the volumes required can be considerable.

Some proteins for replacement therapy have traditionally been isolated from animal sources. Examples of this are insulin and calcitonin (Bowman and Rand 1984). Much of the insulin used clinically, until quite recently, was extracted from either porcine or bovine pancreas. Molecular species differences are too small to influence its metabolic activity as human insulin differs from porcine and bovine insulin by only one and three amino acids, respectively. Nevertheless they are foreign proteins and allergic reactions do occur as a result of prolonged administration of the foreign insulin or other co-purified protein impurities.

For blood disorders such as the haemophilia's the required biomedical proteins are purified from human blood. There is of course a significant risk of disease agents being transmitted from the blood donor to the recipient of the purified blood product. For these reasons among others alternative methods of producing important biomedical proteins have been sought.

Human proteins can be made by recombinant bacteria (Pestka 1983) (Johnson,1983) or yeast (Sleep et al., 1987, Casolaro et al., 1987) Unfortunately such

systems are only suitable for the production of relatively simple proteins that do not require post translational modification. Human insulin, which is non-antigenic, is now available as a result of recombinant technology (Johnson 1983), and is now routinely used. The production of more complex mammalian proteins using bacteria or yeast generally results in a product of low biological activity and/or stability, as prokaryotes and lower eukaryotes are incapable of carrying out the post translational modifications normally present on many mammalian proteins.

Culture systems of transformed mammalian cells, hybridomas and insect cell lines have also been employed for the production of biomedical proteins. Such cell culture systems are capable of performing post translational modification of proteins and have produced some biomedical proteins with high efficacy (Garver et al., 1987; Pavirani et al., 1987; Ogonah et al., 1996). One of the difficulties with mammalian cell systems is striking a balance between the high protein production per cell and high cell density (Hodgson, 1991,1992). Such systems are currently used to produce biomedical proteins but the cost of this method of production is generally high.

Transgenic animals have been used to produce biomedical proteins directly. It was proposed soon after the production of the first transgenic mice that transgenic farm animals could be used as bioreactors for the production of important biomedical proteins (Palmiter et al., 1982, Lovel-Badge 1985). With the use of specific promoters the protein could be secreted into body fluids which could then be harvested for purification of the biomedical protein. Candidate fluids were blood saliva, urine and milk. The protein could be purified from the fluid and the transgenic animals would provide an infinitely replenishable source of the protein.

The expression of biomedical proteins in the blood is problematical in that the biologically active protein has access to all body tissues and may have serious harmful effects. Nevertheless transgenes encoding human haemoglobin have been successfully expressed in the blood of transgenic swine (Swanson et al., 1992) with no apparent deleterious effects. Expression of transgenes in the mammary gland and the targeting of the secreted biomedical protein to milk has attracted much attention (reviewed by Maga and Murray, 1995). The volume of milk produced during lactation is considerable, particularly in ruminants the amount of protein production attainable during one lactation in dairy cows is approximately 240 kg/animal/lactation and in dairy sheep approximately 12 kg/animal/lactation (Clark et al., 1990). Given these characteristics it seemed reasonable that dairy animals could be used as bioreactors for the production of biomedical proteins. By targeting the expression of a given gene to the mammary gland using a milk protein gene promoter the protein of interest could pass through a mammalian cell capable of carrying out post translational modifications that may be required and secreting the protein in a medium from which it should be relatively easily harvested.

The mammary gland is an organ which is relatively autonomous. Thus the expression of the transgene in this organ should have little opportunity to interfere with the wider physiology of the transgenic animal. Throughout lactation milk protein genes are abundantly expressed, and are thus candidates genes to provide the promoter to drive efficient expression of transgenes encoding biomedical proteins, in the mammary gland.

A number of transgenes have now been expressed in the milk of large domestic animals under the direction of milk protein gene promoters (Wright et al., 1991,

Greenberg et al., 1991, Velandar et al., 1992). Large numbers of transgenic animals can be produced from a single transgenic founder which expresses the protein at commercially viable levels. All offspring of a given founder as a rule produce the protein at similar levels. Consistent and high levels of expression of the protein of interest in milk facilitate large scale purification.

1.6. : LACTATION

The mammalian mammary gland undergoes extensive changes during lactation and pregnancy. During pregnancy the mammary gland grows a large lobule-alveolar epithelial tissue and ductal network, during late gestation the gland differentiates and alveolar structures are formed at the ends of the ductal tree. The differentiation accompanies the acquisition of the ability of the epithelial cells to synthesise milk proteins and secrete milk (Topper and Freeman 1980). After parturition the volume of milk secreted increases to a maximum and then gradually declines with time . Lactation is influenced by a number of endocrine hormones, cell-cell interactions and external stimuli such as milk removal and temperature.

The composition of milk is species dependent and the regulation of milk production has been demonstrated to vary from species to species (Jenness, 1982). Because of the large cost involved in the production of transgenic large animals transgenes are more often than not tested in the mouse.. The effects of species variation in the hormonal control of lactation need to be considered for the meaningful analysis of pilot transgenic experiments in mice.

1.7 1: The Hormonal Regulation Of Milk Protein Genes

The tissue and developmental stage-specific regulation of milk protein gene expression is influenced by a variety of factors including peptide and steroid hormones, cell-cell, and cell substratum interactions. Lactogenesis in the mammary gland requires a large quantitative increase in the synthesis of enzymes, metabolites and messenger RNA's associated with the secretion of the unique components of milk. There is no single hormone that initiates lactation. Rather, a cascade of events occurs in the endocrine system during the third trimester of gestation that prepares the mammary gland for the secretion of milk. The secretion of milk is closely integrated with parturition (Tucker, 1988). In cows the levels of glucocorticoids and growth hormone peak at parturition. The level of prolactin peaks just prior to this at about three or two days before parturition as do the levels of estradiol-17 β and the level of prostaglandin F also peaks about one day before parturition (Tucker, 1988). The combination of changes in the levels of these factors is thought to bring about the second stage of lactation. Importantly the levels of progesterone in the serum fall dramatically at about five days prior to parturition and coincide with the rise in the levels of these other lactogenic factors. Progesterone has been demonstrated to be an inhibitor of lactation (reviewed by Tucker, 1988). The following paragraphs review some of the most recent transgenic work on the regulation of milk protein genes. It seems that there are important similarities in milk protein gene promoters and similar DNA motifs are present within the promoter and enhancer regions but there appear to be important differences in the detailed mechanisms of regulation of different milk protein genes.

1.7.2 : The Casein Genes

The casein genes are a family of genes that are present in a cluster. This gene cluster is present on chromosome 6 in the cow and on chromosome 5 in the mouse. The bovine casein gene cluster has been reported to be within a 185-kb locus (Threadgill and Womack, 1990) and the genes are arranged α_{s1} , β , α_{s2} and κ casein. It is believed that the casein genes are co-ordinately regulated and that this regulation may involve elements similar to the LCR of the globin gene cluster (Maga and Murray 1995). This proposal is supported by transgenic experiments using the rat β -casein gene as a transgene in mice. In these studies the casein gene was expressed in a tissue specific manner that was developmentally regulated but the maximal level of expression of the transgene was 100 fold less than that of the endogenous mouse β -casein gene (Lee et al., 1988). The CAT reporter constructs driven by the β -casein gene promoter were susceptible to position effects and demonstrated a 500-fold range of expression in the CAT activity of the lines generated (Lee et al., 1989). These studies showed that the minimal promoter fragment encompassing 0.5-kb of 5' flanking sequence, non coding exon 1 and 0.5 kb intron A was sufficient for the mammary specific expression during lactation but it was clear that other *cis*-acting sequences required for the normal high level of expression of this gene during lactation were lacking in these transgenes. In pseudopregnant rabbit β -casein gene induction is seen on administration of the pituitary hormone prolactin. Rabbit mammary tissue appears to be more sensitive to β -casein gene induction by prolactin than mammary tissue from other animals. The reason for this is thought to be that prolactin increases the secretion of progesterone from the corpus luteum in rats and mice and progesterone then acts as an inhibitor of lactogenesis. When progesterone is absent

from the system prolactin is shown to be lactogenic. Thus it would appear that the induction of the casein genes in many species is indeed dependent on the presence of prolactin (Reviewed by Tucker, 1988). Studies on the rabbit α_{s1} -casein have shown that the transcription of this gene is induced by prolactin (and the workers demonstrated that the region responsible for this responsiveness was contained within a 655-bp fragment located 5' of the coding sequences). In subsequent studies on this fragment an MGF like binding sequence was found within the fragment (see section 1.7.4)

1.7.3 : The Rodent Whey Acidic Protein (WAP) Gene

Studies of the hormonal requirements for WAP gene regulation performed on various explant and cell culture systems have indicated that WAP induction is dependent upon the synergistic action of, glucocorticoids, insulin, and prolactin. The molecular mechanisms by which these factors interact to regulate WAP gene expression are only now beginning to be understood. In the previous section it was highlighted that prolactin appears to be one of the major determinants in the regulation of the casein genes. However the induction of WAP gene expression in the absence of prolactin is considerably more sensitive to glucocorticoids than is the case for the casein genes. A 68 fold increase in the level of WAP mRNA was obtained when insulin and hydrocortisone were added to mammary explant cultures this compared with only minimal increase in casein gene mRNA (Hobbs et al., 1982) Thus it appears that although prolactin is required for optimal WAP gene expression the induction of this gene is less sensitive to prolactin but is considerably more sensitive to glucocorticoids than the casein genes. A recent study identified glucocorticoid receptor

binding sites flanking NF1 sites that have been demonstrated to be within a previously identified hypersensitive site in the 5' flanking region of the rat WAP gene (Li and Rosen 1994). These sites were identified using *in vitro* DNase I footprinting. It was demonstrated that these sites were able to confer dexamethasone inducibility to a construct with a minimal thymidine kinase promoter driving the expression of a CAT (chloramphenicol acetyltransferase) reporter gene. Transgenic mice harbouring the rat WAP transgene were adrenalectomised and administration of dexamethasone was demonstrated to be essential for the maintenance of the expression of the transgene, a correlation of the glucocorticoid induction of the WAP gene and the appearance of a DNase I hypersensitive site was demonstrated. The workers conclude that the hormone induced chromatin changes in the 5' flanking regions of the WAP gene are a direct consequence of the binding of the glucocorticoid receptor to the glucocorticoid half sites identified by the footprinting assay. The glucocorticoid receptor binding sites did not conform to the accepted consensus binding sequence for glucocorticoid receptor but the identified footprint site did contain GR binding half sites. It is thought that the binding of the GR to these sites is cooperative and may involve interaction with other nuclear factors involved in WAP gene transcription.

1.7.4 : The Ovine BLG Gene

As with many other milk protein genes the molecular mechanisms regulating the tissue specific and developmental control of the ovine BLG gene are only just beginning to be understood. Studies on this milk protein promoter in combination

with studies done using the β -casein gene have helped to elucidate a pathway by which the prolactin receptor mediates effects on milk protein gene promoters.

The prolactin receptor is a member of the cytokine receptor superfamily. A variety of cytokines, lymphokines and growth factors function by interacting with receptors that are members of this receptor family. These receptors share extracellular motifs and have limited similarity in their cytoplasmic domains. Although unlike the protein tyrosine kinase receptors, the receptors of the cytokine receptor family do not possess a catalytic domain, binding of these receptors to their ligand induces tyrosine phosphorylation via a family of cytoplasmic protein tyrosine kinases named the Janus kinase family or JAKs. JAKs are phosphorylated on tyrosine residues upon binding of a cytokine to its receptor causing an upregulation of the JAK enzymatic activity. The activation of the JAK in turn leads to the phosphorylation of the receptor and activation of a transcription factor belonging to the STAT family. These STAT transcription factors bind to GAS sites in some gene promoters bringing about an effect on gene transcription. The prolactin receptor has been demonstrated to be associated with the JAK2 tyrosine kinase (Rui et al., 1994).

A factor that was originally thought to be mammary specific was found to bind to the promoter region of the ovine BLG gene (Watson. et al., 1991) This factor named MPBF was thought to be related to the factor identified by (Schmitt-Ney. et al., 1991) that bound to the promoter of the β -casein gene. The factor in that study was called MGF. The binding sites for these two factors were found to be very similar. The MPBF/MGF sites were also found to be present within an enhancer implicated in substratum dependent prolactin induction of the β -casein promoter

driving the expression of reporter genes in the CID9 mammary cell line (Schmidhauser et al., 1992).

Three MPBF sites occur within the 410-bp promoter of the ovine BLG promoter that has been demonstrated to be sufficient to drive the tissue specific and position independent expression of the BLG transgene in mice (Whitelaw et al., 1992). The MPBF sites in the BLG transgene have been further studied to elucidate their role in ovine BLG transgene expression in the mouse. Mobility shift assays were performed to identify mutations in the three MPBF binding sites that abolished binding of the factor (Burdon et al., 1994). The three sites had different affinities for the MPBF transcription factor. One site designated STM had a high affinity for the factor, another site designated A3 had an intermediate affinity for the factor and the third site A1 only weakly bound the factor. Three transgenes harbouring mutations were used to generate transgenic mice. One transgene harboured a mutation that abolished binding in the high affinity site only, another transgene harboured mutations in the two weakest binding sites A1 and A3 and the third transgene harboured mutations that abolished binding in all three binding sites. Analysis of the BLG RNA levels in eight transgenic lines for each transgene demonstrated that the mutations did not abolish expression of the transgene but the level of expression of the A1/A3 mutation was reduced by half and with all three binding sites mutated the level of expression was reduced four fold. Thus the MPBF binding sites were demonstrated to be necessary for the maximal expression of the BLG gene driven by the -410 -bp promoter. The transgene was still expressed in a position independent manner. Ectopic expression of BLG was detected in the salivary gland but the level of expression in this tissue was very low. It was demonstrated that the MPBF factor is

also present at low levels in the salivary gland and in the liver at 2% and 4% respectively of the levels of MPBF in the lactating mammary gland.

The MGF factor has since been shown to be STAT 5 (Wakao et al., 1994). Suggesting that the prolactin receptor signals via the JAK/STAT signalling pathway. The receptor is associated with JAK2 and JAK2 phosphorylates STAT5 activated STAT5 then binds the MPBF/MGF sites present in milk protein gene promoters. These MGF binding sites closely resemble the GAS binding site consensus sequence. It is likely that maximal expression of BLG is brought about by the synergistic effects of a number of different binding factors influenced by a number of different signalling pathways.

1.8 : OVINE β -LACTOGLOBULIN AS A TRANSGENE IN MICE

The milk protein gene promoter of interest in this study is that of the ovine β -lactoglobulin (BLG) gene (Ali and Clark 1988). The BLG gene encodes one of the major whey proteins in ruminants. This gene was first studied as a potential gene to provide a promoter to drive the expression of heterologous biomedical proteins in the milk of sheep. Test studies were performed in mice using the entire ovine BLG gene and a considerable amount of its 5' regulatory sequences as a transgene. Despite the fact that the rodent possesses no endogenous BLG gene the ovine BLG protein was expressed efficiently in the mammary gland of transgenic mice. Nearly all lines of mice generated which harboured this transgene expressed the protein in their milk (Simons. et al., 1987). Thus the transgene was expressed in a position independent manner. Moreover the expression level of the protein in the milk correlated with the

copy number of the transgene in any given line (Whitelaw et al., 1992). Thus the expression of the transgene was also copy number dependent. The profile of expression of the protein throughout pregnancy and lactation was similar to that in sheep. Thus despite the fact that rodents do not normally express this protein during lactogenesis the regulatory sequences of this ovine gene are correctly interpreted in the murine mammary gland.

The expression of the BLG gene on the whole was at first thought to be tissue specific but recent data shows that the protein is expressed at very low levels in other tissues (Farini and Whitelaw, 1995). The low level ectopic expression of this transgene in mice may well be the consequence of the lack of a repressor that is normally present in sheep as ectopic expression of BLG was not detected in sheep (Ali, 1989). Thus in this respect the mouse may not be the ideal model for testing this promoter when linked to biomedical proteins that may have either a pharmacological or deleterious effect because of ectopic expression.

The high levels of mammary specific expression of the ovine BLG transgene in transgenic mice demonstrated that this was a good promoter for driving high level expression of biomedical proteins in the milk of domestic farm animals. In order to test the ability of the BLG promoter to drive the expression of heterologous proteins a number of experiments were conducted using the promoter and 5' flanking sequences of the ovine BLG gene fused a heterologous protein coding region.

1.9 : AATB THE TRANSGENE

The AATB transgene encodes human α_1 -antitrypsin. This protein is a major protease inhibitor of a number of human serine proteases including elastase, trypsin, chymotrypsin, thrombin, and bacterial proteases. The most important inhibitor action is that against neutrophil elastase, a protease which degrades elastin of the alveolar walls as well as other structural proteins of a variety of tissues. Insufficient circulating levels of α_1 -antitrypsin result in lung tissue damage, leading eventually to the chronic and debilitating disease emphysema. The serum concentration of α_1 -antitrypsin in humans is normally about 1.3 mg/ml.

α_1 -antitrypsin shows a considerable amount of genetic variability. The M variant which can be further classified into subtypes, is the most common allele in all populations. The S allele reaches polymorphic frequencies in many populations (this allele encodes protein which has reduced stability) as does the Z allele, (which encodes protein which is poorly secreted). Both the S and Z alleles are associated with reduced circulating levels of the protease inhibitor and with emphysema. Simple replacement therapy of the protein considerably slows the advance of the disease.

α_1 -antitrypsin is a single chain glycoprotein. It has three N-linked carbohydrate side chains. The side chains are not essential for its biological activity, but synthetic AAT produced in bacteria and yeast lack the carbohydrate moieties and display a markedly reduced half life (Casolaro et al., 1987). Mammalian culture cells produce correctly glycosylated hAAT but the yields are low, and overall, the technology expensive. This is problematic when the protein is required in high dosage for replacement therapy. It was thought that the expression of the gene in the mammary gland of transgenic sheep could lead to the secretion of correctly glycosylated protein into the milk from where it could be easily harvested.

The AATB transgene contains 4.2 kb of 5'untranslated sequences from the ovine BLG gene. The coding sequences are genomic hAAT M1 sequences minus intron one (Archibald et al., 1990). The intron was removed by using an hAAT cDNA to provide the first 80 bp of the hAAT "minigene". Introns were included in the transgene because they have been shown to influence the efficiency of expression of some transgenes and previous transgenic mouse experiments, using the BLG promoter and an hAAT cDNA, were not successful (Archibald unpublished data).

Two studies were performed using this transgene. In the original study (Archibald et al., 1990) thirteen transgenic lines were analysed for the production of the human protein in their milk. Of thirteen lines five expressed the transgene in their milk. Three of these expressed the transgene at levels between 0.5 and 1.0 mg/ml. Another line AATB 35 gave very good levels of expression (7mg/ml).

In the second study (Carver et al., 1993) the level of hAAT was assayed from the milk of eleven female founder animals. In this study all transgenic animals were found to express the protein in their milk. Nevertheless there was a 30 000 fold range in the expression of the human protein between the animals. The levels detected most of the lines were very low indeed and would not have been detected using the protein assays in the original study. Two lines were found to contain high levels of the human protein in their milk 8 mg/ml and 21 mg/ml. Overall the results of the two independent studies using the AATB transgene correlate very well.

It is quite clear in both studies that the level of expression of the transgene did not correlate in any way with the transgene copy number. Interestingly the two stable lines expressing relatively high levels of hAAT, AATB 35 (Archibald et al.,1990) and

AATB 46.2 (Carver et al., 1993), appeared to be generated from transgene loci of very low copy number (one and three respectively).

These studies demonstrated that it is possible for the BLG promoter to drive high levels of expression of heterologous transgenes in the mammary gland of transgenic mice. AATB has also been demonstrated to drive efficient expression of hAAT in the milk of transgenic sheep (Carver et al., 1993). To date a flock of transgenic sheep harbouring the construct have been produced which stably express the human protein in their milk at 13-17 g/l. Subject to successful clinical trials hAAT from transgenic sheep will be available for therapeutic use in the foreseeable future.

To date AATB has been one of the most successful transgenes for the production of a biomedical protein in the mammary gland of transgenic animals. It appears that sheep express the AATB transgene more predictably than transgenic mice, which is not entirely surprising since the transgene utilises an ovine promoter to drive its transcription. Although the ovine BLG promoter gave a position independent copy number dependent expression of the genomic BLG gene, the promoter did not give the same predictable expression profile for heterologous genes.

1.10 : THE AIM OF THIS WORK

The production of transgenic animals as bioreactors for the production of biomedical proteins was first spoken of in the scientific literature soon after the generation of the first transgenic mouse by pro-nuclear injection (Brinster et al 1985). Since then numerous studies have been undertaken both in the murine and other animals to achieve this aim.

The vast majority of transgenes encoding biomedical proteins are subject to position effects and the expression levels are erratic and unreliable. Given the high cost of generating transgenic live stock a transgene promoter that gives reliable good levels of expression in a targeted organ is the ideal promoter.

The ovine BLG promoter has been demonstrated to be capable of driving the high level of expression of human alpha-1-antitrypsin in the mammary gland of transgenic mice and sheep. Two lines of transgenic mice were generated that gave unusually high levels of the human protein in their milk. In both cases this was thought to be due to the influence of the murine chromosomal DNA surrounding the transgene locus. The aim of this study was to investigate the role the flanking murine sequences played in expression of the transgene in one of these lines. And hopefully isolate through this investigation the sequences responsible for high level expression, with a view to using the sequences or the information obtained about their role in transgene expression, to improve the predictability of the BLG promoter when used to drive the expression of heterologous protein in the mammary gland.

CHAPTER 2 : MATERIALS AND METHODS

2.1 : Mice Used

Deoxyribonucleic acid (DNA) for Southern blot analysis was prepared from liver tissue of transgenic mice of lines AATB 35 (Archibald et al., 1990), AATB 46.2 (Carver et al., 1993) and non transgenic mice (F1) using method 2.7. DNA for construction of the cosmid library was obtained from liver DNA of a transgenic mouse of line AATB 46.2 (AATB 46.2.36).

2.2 : Recombinant plasmid clones

AATB ; P-poly111-I with BLG-h α_1 AT insert supplied by Dr. A. Archibald

p8 α 1ppg : human AAT cDNA clone supplied by Dr. G. Kelsey

2.3 : Non-recombinant plasmid vectors and bacterial host strains used

Vectors

Bluescript SK+ ; 2961 bp phagemid carrying ampicillin resistance gene (Stratagene)

SuperCos 1 ; 7937 bp cosmid carrying ampicillin resistance gene, dual *cos* sites, SV40 promoter and neomycin resistance gene (Stratagene)

Lambda DASH II ; was the lambda phage vector used to clone the secondary transgene (Stratagene)

Bacterial host strains and genotypes

DH5 α (F⁻, ϕ 80dlacZ Δ M15, Δ (lacZYA-argF)U169, deoR, rec A1, endA1, hsdR17(r_k⁻, m_k⁺), supE44, λ ⁻thi-1, gyrA96, relA1)

NM554 (recA13, araD139, Δ (ara-leu)7696, Δ (lac)l7A, galU, galK, hsdR, rpsL(str^r), mcrA, mcrB)

XL1-Blue MRA (Δ (mcrA)183, Δ (mcrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, gyrA96, relA1, lac)

XL1-Blue MRA (P2) XL1-Blue MRA (P2 lysogen)

2.4 : Enzymes, antibiotics and specialised chemicals

Amersham : [³⁵S]dATP- α -S 3000 ci/mmol, α -³²P-CTP 3000 ci/mmol, Hybond N, random prime labelling kit, USB sequencing kit

Boehringer Mannheim : proteinase K, X-GAL, IPTG, calf intestinal phosphatase (CIP), caesium chloride (CsCl), all restriction endonucleases used (unless otherwise stated)

Costar : Spin-X DNA purification columns

Difco : bactotryptone, yeast extract

Gibco BRL : DH5 α competent cells

ICN Pharmaceuticals : NZ amine (casein hydrosylate)

New England Biolabs : β -agarase

Pharmacia : universal primer M13, Klenow fragment DNA polymerase 1

Promega : Erase-a-Base system

PRU : goat anti α_1 AT serum

SAPU : horseradish peroxidase conjugated sheep/goat IgG antiserum.

Sigma : diethylpyrocarbonate (DEPC), TEMED, sodium dodecyl sulphate (SDS) ampicillin, human α_1 -antitrypsin, 4-Chloro-1-naphthol, hydrogen peroxide

Stratagene : SuperCos 1 cosmid vector kit, Lambda DASH 11 undigested vector kit, Gigapack 11 gold *in vitro* phage packaging extract, Srf 1 restriction endonuclease

2.5 : Media

Broths

All *E.coli* strains were grown in either LBM (10g tryptone, 5g yeast, 5g NaCl, 2g MgCl₂.6H₂O per litre) or in NZY (10g NZ Amine, 5g NaCl, 2g MgSO₄.7H₂O, 5g yeast extract per litre, adjusted to pH 7.5 with NaOH) liquid medium.

Plates

LBM plates for plasmid subcloning were poured using 10g bactotryptone, 5g yeast extract, 5g NaCl 15g agar per litre and ampicillin to a final concentration of 50µg/ml added to cooled medium immediately before pouring. LBM plates for phage cloning were prepared as described but with the omission of ampicillin and addition of 10 ml per litre 1M MgSO₄.7H₂O just before pouring. Top agar overlays were poured from 10g bactotryptone, 5g yeast extract, 8g of agar, 5g NaCl per litre and 10mls 1M MgSO₄.7H₂O added just before pouring. NZY plates for the cosmid library were poured using, 10g NZ Amine, 5g NaCl, 2g MgSO₄.7H₂O, 5g yeast extract adjusted to pH 7.5 with NaOH with 15g agar per litre and ampicillin at a final concentration of 50µg/ml.

2.6 : Autoradiography and Photography

Agarose gels stained with 0.5 μ g/ml of ethidium bromide were photographed using Polaroid type 667 film in a land camera, illuminating from below using a transilluminator of wavelength 302nm.

Autoradiographs were produced using Agfa-Curix film RP1. Films were exposed for the times stated at either -70°C or room temperature with Dupont intensifying screens unless otherwise stated.

2.7 : Preparation of High Molecular Weight DNA from Mouse Liver

The mouse was dissected and placed immediately into ice-cold SSCT (1 x SSC + 10mM Tris HCl pH.7.4). The liver tissue was minced and washed several times in ice-cold SSCT to remove red blood cells and then resuspended in 6 ml of ice-cold RSBE/NP40 (10 mM Tris, 10mM NaCl, 2mM EDTA pH. 7.4, 5% NP40 and 1mM PMSF (added just before use)). The minced liver tissue was homogenised using a 10 ml Dounce homogeniser. Transferred to a 30 ml Corex tube and spun at 4.5 rpm in an SS34 Sorval rotor for 5 minutes at 4°C. The supernatant was discarded and the pellet gently resuspended in another 6 ml of RSBE/NP40 (this wash was repeated twice more). Finally the nuclei were resuspended in 2 ml of RSBE (10 mM Tris, 10mM NaCl, 2mM EDTA pH. 7.4) followed by the addition of 2 ml of SNET (1% SDS, 0.6 M Na Cl, 10 mM EDTA, 20 mM Tris pH. 7.4) whilst vortexing the suspension. Proteinase K was added to a final concentration of 200 μ g/ml and the

viscous solution was incubated at 37°C for two hours. The DNA was then deproteinised by phenol and phenol/chloroform extraction.

2.8 : Isolation of RNA by RNAZOL

(adapted from Chomczynski and Sacchi, 1987)

All solutions , glassware and plastics were soaked in 0.1% solution of DEPC, then autoclaved.

Solution

23.64 g Guanidinium Thiocyanate (4M)

1.25 mls of 1M Na Citrate pH. 7.0 (25 μ M)

0.5 g Sarcosyl (1%)

Made up to 50 mls with distilled deionised water

360 μ l of β -mercaptoethanol (0.1 M)

For use in RNA isolation one volume of water saturated phenol was added to one volume of the above solution along with 0.1 volumes of 2M Na Acetate pH.4.0

Method.

Tissues were dissected from mice of lines AATB 46.2 and TAB 36.2 and snap frozen in liquid nitrogen. The samples (approximately 100mg) were slightly thawed for ten minutes on ice before addition of 2mls of RNAZOL stock solution and homogenised at high speed for 1 minute. The homogenate was transferred to Eppendorf tubes in 1 ml aliquots and 100 μ l of chloroform added to each. The samples were shaken vigorously and left on ice for 15 minutes. They were then centrifuged for 15 minutes at 4°C. The upper aqueous phase was removed and an

equal volume of isopropanol added before placing the samples at -20°C for 45 minutes. The samples were then spun at high speed for 15 minutes in an Eppendorf microfuge. The resultant pellet was washed twice with 75% ethanol and air dried. The RNA pellet was then dissolved in 300 μl of DEPC treated distilled deionised water.

2.9 : Deproteinisation of nucleic acids using phenol/chloroform extraction

(a) General protocol

DNA samples were deproteinised using phenol/chloroform extraction. The Generally one phenol extraction one phenol/chloroform and one chloroform extraction was performed although the number of extractions did vary depending on the quality of the sample. Phenol extraction of DNA entailed adding one volume of water saturated phenol to the lysed nuclei or bacteria, mixing thoroughly, and separating the organic and aqueous phases by centrifugation. The aqueous phase was recovered and the extraction procedure repeated with one volume of phenol/chloroform (1:1). The final extraction was performed with chloroform/isoamyl alcohol (24:1) to remove residual traces of phenol from the aqueous phase. The DNA was then recovered by ethanol precipitation.

(b) Of DNA for library construction

The solution of DNA from the lysed nuclei was gently extracted once with 0.5 volumes of phenol and 0.5 volumes of chloroform and again with 1 volume of chloroform. Following the gentle but thorough mixing of the organic and aqueous

phases they were separated by centrifugation at room temperature. To prevent breakage of the DNA the organic phases were removed from the bottom of the tube using a pasteur pipette and the DNA transferred into a clean tube by gentle pouring. The DNA was then placed inside dialysis tubing and dialysed against 1 litre of TE buffer (10mM tris pH 7.4, 1mM EDTA) overnight. The next day the DNA was further dialysed against 3 litres of TE buffer overnight. The DNA was of high molecular weight but very dilute to concentrate the DNA it was dialysed four times, for two hours each dialysis, against 100ml volumes of butan-2-ol. All dialysis was performed at 4°C with gentle stirring.

2.10 : Ethanol precipitation of nucleic acids

DNA and RNA was recovered by precipitation with 2.5 volumes of 100% ethanol containing 0.3 volumes of NaAcetate pH 5. The ethanol mix and phenol extracted DNA/RNA sample were mixed, genomic DNA was spooled out using a glass rod and partially air dried before being dissolved in TE buffer (10mM tris pH 7.4, 1mM EDTA pH 8). Plasmid DNA and RNA samples were incubated at -70 for 30 minutes. This was then centrifuged and the resulting pellet of nucleic acids washed twice with 70% ethanol air dried and redissolved in TE buffer. Genomic DNA was stored at 4°C, plasmid DNA at -20°C and RNA at -70°C.

2.11 : Spectrophotometric analysis to determine concentration of DNA and RNA samples.

Aliquots of 5-10 µl of the sample nucleic acid were added to TE buffer to give a final volume of 1 ml. The optical density of the diluted DNA/RNA was then

determined using a spectrophotometer with a deuterium light source at 260 and 280 nm. An optical density reading of 1 equals 50 μ g DNA/ml; 40 μ g RNA/ml.

2.12 : Restriction Endonuclease digestion of DNA

(a) Genomic DNA

10 μ g of DNA was digested in a volume of 200 μ l with 2 units of enzyme/ μ g of DNA according to the manufacturers directions. Digestion was allowed to proceed for at least 5 hours then the digested DNA was phenol/chloroform extracted and ethanol precipitated. The pellet was resuspended in 40 μ l of TE buffer. 0.2 volumes of loading dye (30% ficoll 400, 0.1% SDS, 40mM EDTA pH8, 1.2 mg/ml bromophenol blue) were added then the sample was heated to 65°C for 10 minutes before being loaded on an agarose gel.

(b) Plasmid /cosmid DNA

Restriction digests of plasmid or phage DNA were carried out at a concentration of enzyme of less than 1 μ g/10 μ l with the concentration of enzyme not exceeding 10%. The incubation time varied from 1 to 2 hours. Small aliquots of these digests were checked on agarose minigels for completeness of digestion.

2.13 : Agarose Gel Electrophoresis

(a) Of DNA

Digested genomic DNA was electrophoresed in 0.7 % agarose gels. Large (20cm x 20cm) gels were run at 30-40 volts overnight and small (10cm x 12cm) were

run at 20 volts overnight. Digested plasmid DNA of small recombinant clones (AATB, TAB subclones) were electrophoresed in 0.8-1% agarose gels at 50-70 volts for 30 minutes to several hours depending on the resolution required. Digests of the cosmid clones were electrophoresed in 0.3% agarose gels and run at a maximum of 30 volts at 4°C (to prevent the gel melting) until the fragments were sufficiently well resolved (usually 8 hours or overnight). All agarose gels used for DNA electrophoresis were made from and run in buffer containing 40mM tris acetate, 2.5mM EDTA/Na pH7.7 and 0.5µg/ml of ethidium bromide.

(b) Of RNA

10µg equivalents of RNA samples were electrophoresed in denaturing agarose gels (1.5% agarose, 2.2M formaldehyde, 10mM MOPS (3-(N-morpholino)-propanesulphonic acid) pH 7). 20µl of sample buffer (50% formamide, 2.2M formaldehyde, 0.5mM EDTA/Na, 10mM Sodium phosphate pH 8) was added to each RNA sample, on ice, immediately after heating the sample at 65°C for 10 minutes. The gel was run in 1x MOPS at 25 volts overnight

2.14 : Acrylamide gel Electrophoresis

(a) Denaturing sequencing gel

Sequencing reaction were electrophoresed in 0.8% denaturing acrylamide gels (31.5g Urea, 7.5 ml of 10xTBE pH 8.8, 11.25 ml of 40% acrylamide/Bis solution (supplied by Bio-Rad), 25 ml of distilled deionised water, 400µl of 10% APS (ammonium persulfate), 40µl TEMED). The gels were cast using 4mm spacers and run on a Bio-Rad vertical sequencing gel apparatus. The running buffer was 1 x TBE and the gel run at 50 watts for 4 to 6 hours. Once the run was finished the sequencing

gels were fixed in 10% acetic acid for 20 minutes, then transferred onto Whatmann 3MM paper and dried at 80°C under vacuum on a gel drying apparatus.

(b) SDS PAGE

SDS polyacrylamide gel electrophoresis was used to analyse the protein composition of the milk samples of the transgenic mice. The resolving gel was a 10% acrylamide gel (16.2 ml dH₂O, 10 ml 1.5M Tris-HCl pH8.8, 400µl 10% SDS, 13.3 ml 30% Acrylamide/Bis, 100µl 10% APS, 20µl TEMED). The gel was poured, overlaid with distilled water and left to polymerise overnight. The next morning the water was removed and a 4% stacking gel (6.1 ml dH₂O, 2.5 ml 0.5M Tris-HCl pH 6.8, 1.3 ml 30% acrylamide/Bis, 50µl 10% APS, 5µl TEMED) poured on top of the polymerised resolving gel. A comb was then inserted into the gel and left in-situ until the stacking gel was sufficiently well polymerised and wells were formed (approximately 1 hour). Once the gel was polymerised the wells were washed and then filled with running buffer (15g Tris base, 72g glycine, 5g SDS made up to 5 litres with dH₂O). Samples were run in the stacking gel for 45 minutes at 26mA and in the resolving gel for a further 3.5 hours at 36 mA in tris-glycine running buffer. Once the run was complete The gels were stained and fixed (20ml 100% trichloroacetic acid, 74ml acetic acid, 500ml Methanol, 0.4g coomassie blue, made up to 1 litre with dH₂O) overnight. The next morning the gels were washed several times with destain (70ml acetic acid, 230ml ethanol made up to 1 litre with dH₂O) to remove the excess coomassie stain.

2.15 : Western blotting

Following electrophoresis of the milk samples in a 10% denaturing acrylamide gel the proteins were transferred to a nitro-cellulose membrane by electroblotting. The

current was 0.8 mA/cm^2 the anode buffer was 0.3M Tris, 20% methanol in distilled water pH 10.4 and the cathode buffer was 25mM Tris, 40mM hexanoic acid and 20% methanol in distilled water pH 9.4). Following electroblotting the nitro-cellulose filter was immersed in a 5%(w/v) solution of skimmed milk in PBS. Human α_1 AT was identified on the immunoblot filters by incubating the filter with goat anti- α_1 AT serum (1;200dilution of serum in 10mls skimmed milk at 37 °C for 2 hours), followed by incubation of the filter with anti-sheep/goat IgG serum conjugated to horseradish peroxidase. The filter was given four 30 minute washes in 20 ml of skimmed milk solution after each incubation. Finally the filter was developed in 4-Chloro-1-naphthol and H_2O_2 .(4ml 4-Chloro-1-naphthol at 3mg/ml, 1ml 1MTris-HCl pH 7.5, 19ml dH_2O , 50 μl 3% H_2O_2)

2.16 : Transfer of nucleic acids onto a solid support

(a) Southern blotting

Before transfer of DNA to Hybond N nylon membrane (Amersham) the DNA was depurinated by immersion in 0.25M HCl for 2x 15 minutes. The DNA was then denatured by immersion in 1.5M NaCl, 0.5M NaOH for 40 minutes changing the solution after 20 minutes. The gel was then neutralised using a 0.5M solution of Tris-HCl pH 7.4 containing 1.5M NaCl. The gel was gently shaken in all solutions to aid diffusion of the solutions throughout the gel. The gel was then rinsed in 2x SSC (20 x SSC is 175.3g NaCl, 88.2g tri-sodium citrate to a final volume of 1 litre NaOH or HCl used as appropriate to bring the solution to pH 7). The DNA was then transferred by Southern blotting (Southern, 1975, Manniatis, 1982). The bottom reservoir contained

10 x SSC. The gel was inverted and laid on top of the 3MM wick from the bottom reservoir. The filter was then placed on top of the gel followed by 3MM paper soaked in 2 x SSC. Finally a layer of paper towels were placed on top of the gel to draw the solution through the gel and transfer the DNA onto the Nylon membrane. Transfer was allowed to proceed overnight. After removal of the filter from the apparatus it was rinsed in 6 x SSC and then partially dried. The DNA was then fixed to the nylon membranes using UV-irradiation (wavelength 302 nm) using a pre-calibrated stratalinker.

(b) Northern blotting

RNA gels do not need to be treated before blotting. The gels were rinsed in 1 x SSC and blotted onto Hybond N membrane as described above.

2.17 : Gel purification of DNA

A number of methods were used for the recovery of DNA from agarose gels.

(a) Spin-X columns

Small fragments to be used as DNA probes were recovered by cutting the band out of the gel (generally 0.8% agarose) and placing in the upper chamber of a spin-X column. The column was then placed at -70°C for 30 minutes to freeze the gel and thus fractionate the agarose matrix. The tube was then removed from the freezer the gel allowed to thaw, and then spun at high speed in an Eppendorf microfuge. The agarose is retained in the upper chamber, which was discarded and the buffer and much of the DNA was at the bottom of the tube. To precipitate the DNA 0.3 volumes of NaAcetate pH 5 were added to the solution followed by two volumes of 100%

ethanol. The tube and contents were then incubated at -70°C for 30 minutes. Following this the tube was then spun at high speed in a microfuge for 15 minutes at 4°C . The supernatant was removed and the DNA pellet washed twice in 70% ethanol before air drying. Finally the pellet was dissolved in an appropriate volume of TE buffer (generally $50\mu\text{l}$) depending on the size of the pellet.

Electrophoresis into 15% PEG

Some larger fragments for subcloning were isolated using this method (e.g. Eco RV fragment of Clone 9). The fragment run in an ethidium bromide stained gel was visualised by illumination with low intensity UV light. A piece of the gel (of similar shape and size to the migrating band) immediately in front of the band was removed using a sterile scalpel. This trough was then filled with a solution of 15% polyethyleneglycol (PEG) in TE. The gel was then replaced into the electrophoresis tank which now contains only enough buffer to be level with the top of the gel but not submerge it. electrophoresis was allowed to proceed again for ten minutes to run the DNA into the trough. The PEG was removed and the gel illuminated with UV light to check the band has been removed. DNA removed in this way was used directly in ligation reactions.

β -Agarase

Fragments used in the construction of the secondary transgene were isolated using this method. The fragments were run in a 1% LMP agarose gel containing $0.5\mu\text{g/ml}$ ethidium bromide. Once resolved the piece of gel containing the fragment was removed. Two volumes of 1x agarase buffer were added to the gel slice and this

was left to diffuse for 2 x 30 minutes. The buffer was then removed by aspiration and the gel slice melted at 65°C once melted the gel slice was incubated at 37°C and allowed to cool to this temperature. β -agarase (1 unit per 100 μ l of gel) was then added to the melted gel slice and digestion allowed to proceed for one hour at 37°C. Once digestion was complete the tube was cooled on ice for 15 minutes and then spun at high speed in an eppendorf microfuge. Any undigested agarose formed a pellet in the bottom of the eppendorf tube. The supernatant containing the DNA was removed, phenol/chloroform extracted and finally ethanol precipitated.

2.18 : Radioactive labelling of DNA

Probes were radiolabeled according to the random priming labelling method devised by Feinberg and Vogelstein, (1983). Either Amersham's or Boehringer Mannheim's random prime labelling kits were used according to the manufacturers directions. Generally 25ng of DNA were labelled. Incorporation of the radioactive label was checked by counting a 1 μ l aliquot dropped onto a glass fibre filter which was then washed with 5% trichloroacetic acid and recounted. Incorporation as determined by Cherenkov counting was routinely in excess of 1×10^9 dpm/ μ g DNA.

2.19 : Hybridisation

All hybridisations were performed in phosphate buffer (0.5M sodium phosphate pH 7.2, 7% SDS, 1mM EDTA/Na) according to Church and Gilbert, (1984). All filters were pre-hybridised in buffer for 1 hour before addition of the radiolabelled probe. The probe was denatured by the addition of 0.25 volumes of

freshly made NaOH 5 minutes before addition to the hybridisation chamber. All hybridisations were performed at 65°C in a perspex hybridisation chamber or in hybridisation bottles. Filters were washed at 65°C in 40mM sodium phosphate, 1% SDS or 0.1-0.2 x SSC, 0.1% SDS. All washes were high stringency.

2.20 : Construction of a Cosmid Library of AATB 46.2

(a) Preparation of vector DNA

20µg of SuperCos1 vector (Evans et al., 1989) was digested with Xba restriction endonuclease in a total volume of 200µl in standard buffer conditions. The digested vector was then subject to phenol/chloroform extraction, ethanol precipitation and the pellet resuspended in 20µl distilled, deionised water giving a final concentration of 1µg/µl. The digested vector was then treated with calf intestinal alkaline phosphatase (CIP). Following this treatment the protein was extracted using phenol/chloroform and ethanol precipitated. The pellet was resuspended in 20µl of TE buffer. Finally the vector was digested with Bam HI restriction endonuclease followed by phenol/chloroform and chloroform extraction and ethanol precipitation. the pellet was resuspended in 20µl of TE buffer.

(b) Nde II partial digestion of the AATB 46.2 chromosomal DNA

Nde II is an isoschizomer of Sau 3A and generates ends compatible with ligation into the Bam HI cloning site of the cosmid vector. A series of test partial digests were performed on liver and DNA isolated by the methods described in sections 2.7 and 2.9(b) of this chapter. Digestion using 0.5 units of Nde II was allowed to proceed for 5 to 45 minutes. The test digestion reactions were stopped by

removal of an aliquot from a 100 μ l reaction and adding 10 μ l of loading dye. The time point samples were run on a 0.3% agarose gel with DNA size markers. The bulk of the digested DNA migrated with an apparent size of 50 kb after a 30 minute digestion. Two large scale digestion of 250 μ g of DNA was then performed using conditions determined from the test digest (total reaction volume 550 μ l). The digest was stopped by the addition of 9 μ l of 0.5M EDTA, pH8.0.

(c) Size selection of digested genomic DNA

The entire DNA digest was loaded on a 10%-40% sucrose gradient. The gradient was spun at 25000rpm in an SW28 Rotor in a Beckman ultracentrifuge for 16 hours at 21°C. 1 ml fractions were obtained by piercing the bottom of the polypropylene tube with a needle and allowing the solution to drip into Eppendorf tubes placed underneath it. Aliquots of the fractions were run on a 0.4 % agarose gel. All fractions containing DNA of an apparent size of 30-50 kb were pooled and ethanol precipitated. A total of 1.5 μ g of genomic DNA was obtained from this procedure. The isolated DNA was then CIP treated for 60 minutes at 37°C followed by phenol extraction, chloroform extraction and ethanol precipitation. The pellet was resuspended in 10 μ l of TE buffer giving a final concentration of 0.05 μ g/ μ l.

(d) Ligation of DNA and packaging of the cosmid library

The ligation reaction contains 0.5 μ g of Nde II digested DNA, 0.2 ug of Xba I/Bam HI digested cosmid vector, 2 μ l of 10mM ATP and 6.8 μ l of distilled-deionised water. The ligation reaction was incubated at 14°C overnight. Packaging of the DNA was performed using Stratagene's Gigapack Gold II in vitro phage packaging extracts

according to the manufacturers directions. Only 4 μ l of ligation reaction were added to the extracts and incubation of the extracts and ligated DNA was allowed to proceed for 2 hours at 22 °C.

(e) Plating of the Cosmid Library

Screening of the cosmid library involves the fixing of DNA from the colonies to a hybridisation membrane. The colonies of the library were grown on hybond N gridded filters [20 000 colonies per filter (132mm diameter)]. The filters were placed on top of LB plates containing 50 μ g/ml of ampicillin. The colonies were spread on the filter by dropping the recombinant bacteria onto the centre of the filter and shaking at high speed (Luckman R100 rotatest shaker, speed 10) for several minutes until the solution is close to the edge of the filter. To prevent excessive growth of the colonies the filters were incubated overnight at 30°C.

(f) Replica Plating of the Cosmid Library

The filters were replica plated on identical hybond N filters. Two replicas were made of each filter. The library filters were placed on sterile 3MM paper (autoclaved between sheets of glass in a dry autoclave for several hours). All manipulation of the filters were performed with sterile forceps which were ethanol flamed after each use. A hybond filter was then placed on top of the library filter, covered with another sheet of sterile 3MM and a heavy stack of glass plates placed on top for 30 seconds. The filters were marked using a needle and ink so they could be orientated following hybridisation. The replica filters were placed on 140mm LB amp plates and grown for 5 hours before treatment of the filter to fix the cosmid DNA to the Hybond

membrane. The original filters were placed on fresh LB amp plates and grown for 5 hours at 37°C to allow the colonies to recover. Following incubation the master plates were stored at 4°C and regularly placed on fresh LB amp plates. The replica filters were prepared for hybridisation by placing it colony side up for 30 seconds in the surface of Whatmann 3MM paper soaked with 0.5M NaOH, then another sheet of Whatmann paper soaked with 1M Tris-HCl, pH 7.6 for 30 seconds. Following this the filters were immersed in 1M Tris-HCl, pH 7.6, 1.5M NaCl and the bacterial debris removed by gently rubbing the filter with a gloved hand. The filters were air dried till they were almost dry then the DNA was fixed to the filter by exposure to UV light (wavelength 302nm) in a pre-calibrated stratalinker. Finally the filters were immersed quickly in chloroform to remove any remaining bacterial debris washed in 1M Tris-HCl, 1.5M NaCl dried and wrapped in saran ready for hybridisation.

2.21 : Phage Cloning of the Secondary transgene

(a) Preparation of the lambda DASH II vector

5µg of Lambda DASH II undigested vector (Stratagene) was digested with Not I restriction endonuclease in a total volume of 50µl. The digest was allowed to proceed for two hours. The phage DNA was then extracted once with phenol/chloroform and once with chloroform. An equal volume of 4M NH₄OAc (ammonium acetate) was added to the aqueous phase followed by 2.5 volumes of ethanol. The DNA was then pelleted by spinning at high speed in an Eppendorf microfuge. The pellet was washed twice in 70% ethanol air dried and dissolved in 50µl of TE buffer.

(b) Preparation of the insert

The insert was prepared from the 15.2 kb Not I fragments of clones 8 and 9. The Not I fragments were gel purified and then digested with Srf I. The fragments containing the murine DNA and part AATB transgene were gel purified and ligated together in the presence of Not I restriction digest to suicide ligations that occurred between the Not I ends of the two fragments. Analysis of the ligation reaction on agarose an agarose gel demonstrated the presence of several products of ligation. Only the large 19.9 kb fragment was gel purified for ligation into the lambda DASH II vector.

(c) Ligation of phage vector and secondary transgene

The ligation reaction contained 0.5µl of ligase buffer (Boehringer Mannheim), 0.5µg of Digested lambda DASH, 0.75µg of insert DNA and 2.5 units of T4 DNA ligase. The reaction was incubated at 4°C overnight. For efficient packaging it is recommended that the ligation reaction be carried out at DNA concentrations in excess of 0.2 µg/µl. In this instance the DNA concentration per microlitre is 0.25µg.

(d) Packaging of ligation reaction

The recombinant phage was packaged using Gigapack II Gold *in-vitro* lambda packaging extracts (Stratagene). The packaging procedure was carried out following the manufacturers directions. 4µl of the ligation reaction was added to the extracts. the incubation was allowed to proceed for 2 hours. Finally 500µl of SM buffer was added to the packaging extract.

(e) Plating the phage clones

The host cell strain was XL1-Blue MRA (P2). A culture of host cells were grown in LB medium containing 0.2% maltose and 10mM Magnesium sulphate at 37°C for 5

hours with shaking. The culture was then spun for 10 minutes at 2000 rpm. The cells were then gently resuspended in 10mM MgSO₄. The cells were then diluted to an OD₆₀₀ of 0.5 with sterile 10mM MgSO₄. 600µl of cells were infected with a range of volumes of the packaged ligation reaction. (100µl, 50µl, 10µl and 1µl). The phage were incubated with the cells for 15 minutes at 37°C with gentle shaking to allow the phage to attach to the cells. The phage were grown on NZY bacterial plates (see media, section 2.. The infected cells were added to 3ml of top agar (NZY broth and 0.7% (w/v) agarose) at 48 °C mixed gently then poured onto the plates. The plates were placed at 37 °C overnight. A total of 35 plaques were obtained each was screened for the presence of 5' and 3' AATB sequences. All but one of the clones hybridised with both probes. Two phage clones were picked for large scale preparation of the phage.

(f) Large scale preparation of Lambda phage

To obtain enough phage DNA in order to excise the cloned DNA insert a large bulk culture was processed (Blattner et al.,1977). 2×10^5 phage were absorbed to 32 ml of XL1-Blue MRA cells which had been concentrated into 0,2 volumes of 10mM MgSO₄. The absorption was left at 37 °C for 20 minutes before being added to 500mls of prewarmed NZY. This was shaken at 37 °C until lysis occurred, after which 5 ml of chloroform was added, shaking continued for 30 minutes before 3mg of RNase and 3mg of DNase were added per flask. The flask was left to sit at room temperature for 40 minutes before being cleared of debris by centrifuging at 10 000 rpm for 10 minutes in a Sorval GSA rotor. 10% PEG₆₀₀₀ and 1M NaCl were added to the phage containing supernatant, dissolved and left at 4°C overnight. The precipitated phage were pelleted by centrifugation at 10 000rpm for 15 minutes at 4°C in a Sorval

GSA rotor. The phage pellet was dissolved in 8 ml of phage buffer, residual debris was extracted by vigorous shaking with an equal volume of chloroform and clearing by brief centrifugation at 3 000rpm for 2 minutes. The top phase was removed and overlaid on a CsCl step gradient (consisting of 1.5 ml 56% CsCl at the bottom of the tube, 1.5 ml of 45% CsCl on top of this and 2.5 ml of 31% CsCl on top of this) The step gradients were centrifuged in an Beckman SW 28 rotor at 23 000rpm and 4°C for 2 hours 15 minutes with zero deceleration. The phage heads band between the 45% and 56% CsCl layers and are visible as a slight bluish coloration. The band was removed through a wide bore needle and dialysed against two changes of 50mM tris-HCl pH7.9, 10mM MgCl₂, 10mM NaCl. To isolate the phage DNA the dialysed phage heads were extracted twice with phenol/chloroform and twice with chloroform and ethanol precipitated

2.22 : Erase a Base deletions of 3' murine DNA

The Erase-a-Base systems designed for the rapid construction of plasmid clones containing progressive unidirectional deletions in any inserted DNA. The system is based on the procedure developed by Henikoff (1984). Exonuclease III is used to specifically digest insert DNA from a 5' protruding or blunt end restriction site. The adjacent sequencing primer binding site is protected from digestion by a 4 base 3' overhang restriction site. Exonuclease III removes a single strand of the DNA starting from the 5' overhang. The reaction proceeds at a steady rate. The reaction is stopped by placing aliquots in S1 nuclease buffer which has a low pH and contains zinc cations, both of which inhibit further exonuclease activity. The S1 nuclease removes the single stranded tail produced by exonuclease III and results in blunt ends at either

end of the clone. These are then ligated and transformations performed so the vector containing the reduced insert can be obtained in sufficient quantities for sequencing reactions to be performed.

Using this method nested deletions were made in the 3' murine flanking DNA that was subcloned in bluescript phagemid vector using the Not I and Hind III sites in the multiple cloning site. In all 2.3 kb of the murine DNA was sequenced using this method (see Chapter 5). The subclone was digested with Cla I and Kpn I. Cla I (which cuts nearest the insert) generates a 5' overhang and digestion with this enzymes makes the DNA 5' of the restriction site susceptible to Exonuclease III digestion. Whereas digestion with Kpn I (which cuts nearest the priming site) generates a resistant 3' overhang. The Kpn I site is adjacent to the T7 and M13 priming sites of the vector and the Cla I site is immediately 3' of the insert. The remainder of the procedure was carried out as per the manufacturers directions

2.23 : Sequencing

Sequencing of double stranded recombinant clones was performed using a USB sequencing kit.. The DNA template was denatured by the addition of 2 μ l of denaturing mix (300 μ l of 5M NaOH, 2 μ l of 0.5M EDTA, 200 μ l dH₂O) to a total volume of 20 μ l incubated on ice for 15 minutes. The solution was then neutralised by the addition 110 μ l of neutralising mix (8 μ l 1M Tris-HCl pH 4.5, 3 μ l 3M NaAc pH 5.2, 100 μ l 100% ethanol). Following this addition the tube is spun at high speed in a microfuge and the pellet of denatured DNA resuspended in 6 μ l of dH₂O and 2 μ l Sequenase buffer and 1 μ l of primer (concentration 0,8 μ M). This is heated to 65°C for 2 minutes and then the reaction is allowed to cool slowly so the template and

primer anneal. From this stage onwards the sequencing reactions are essentially the same as that devised by Sanger et al., (1977). [³⁵S]dATP- α -S was used as the radioactive label. The kit was used according to the manufacturers directions

2.24 : Generation of Transgenic Mice

Transgenic mice were generated by the injection of gel purified insert DNA of TAB and TA plasmids (at a concentration of 1.5 ng/ μ l) into pronuclear mouse eggs, obtained from superovulated C57BL/6 x CBA F1 females after mating with F1 males. Injected eggs were cultured overnight, cleaved embryos were transferred into the oviducts pseudopregnant MF1 recipients (Simons et al., 1987).

2.25 : Tail screening by PCR

Screening G₀ offspring generated from pronuclear injected eggs, for transgenic mice was done using tail DNA. Approximately 1-1.5cm of mouse tail was removed and placed in 1.5ml of digestion buffer (0.3M NaAcetate, 10mM Tris-HCl pH 7.9, 1mM EDTA, 1% SDS with 200 μ g/ml proteinase K) at 37°C overnight, with shaking. The next day the digested sample is vortexed to disaggregate debris. Proteins and other unwanted matter were removed by phenol/chloroform extraction. of 750 μ l of the residue. 25 μ l of 3M NaAcetate were added to the extracted aqueous phase followed by 1 volume of isopropanol. The mixture was mixed well and left at room temperature for 10 minutes for the DNA precipitate to form. The DNA was pelleted

by spinning briefly at high speed in an Eppendorf microfuge. The pellet was washed twice in 70% ethanol air dried and resuspended in 250 μ l of TE buffer.

PCR reactions were then performed with 1 μ l of the DNA. One μ l of DNA from each animal was placed in a 500 μ l Eppendorf tube under 50 μ l of mineral oil. The DNA was denatured by heating for ten minutes at 95°C for 10 minutes on a PCR block.. 50 μ l of PCR sample mix was then added to each tube (5 μ l 10 x PCR buffer, 8 μ l of NTP mix, 5 μ l of DMSO, 0.5 μ l of transgene primer mix, 0.5 μ l control primer mix (HPRT) dH₂O to a final volume of 50 μ l mix plus 0.25 μ l of Taq polymerase for each tube). The samples were then subject to 30 cycles of 91.5°C for 30 seconds then 60°C for 5 minutes. Once the program was complete aliquots of the reactions were analysed on 2% agarose gels.

Control primers amplify a 332 bp segment of the HPRT gene. The BLG primers amplify a 246 bp segment of the 5' end of the BLG gene. The generation of the smaller PCR product is diagnostic of the presence of the AATB transgene.

2.26 : Analysis of Milk Samples

(a) Radial immunodiffusion

Milk samples were obtained from all animals on the 11th day of lactation following oxytocin administration. The milk samples were defatted by dilution the milk 1/5 with sterile distilled water, vortexing the samples briefly, and re-suspending the casein pellet gently in the lower aqueous phase. The aqueous phase was then gently removed from the bottom of the Eppendorf tube and the milk fat left behind. for loading on the partigen plates the samples were further diluted 1/4, 1/20 and 1/100

giving final dilutions of 1/20, 1/100 and 1/500. Samples that did not give a measurable precipitate at these dilutions were loaded at the 1/5 strength. 20 μ l of the diluted milk samples were loaded in the wells of human α_1 -antitrypsin LC-Partigen plates (Behring diagnostics). These plates were left at room temperature for 5 days to develop. The precipitin rings formed by each sample were measured. The concentration of the hAAT in the milk samples were determined from a calibration curve obtained from standards loaded on the same plate. Southern blotting was performed on mice that were PCR positive to verify the result.

(b) for SDS PAGE

The samples for SDS PAGE were also diluted 1/5 and defatted. Aliquots of defatted milk were further diluted 1/5 in sample buffer (6.4 ml distilled H₂O, 2 ml 0.5M Tris HCl pH 6.8, 2g glycerol, 3.2 ml 10% SDS, 0.8 ml 2-mercaptoethanol, 2 ml 0.05% bromophenol blue) giving a final dilution of 1/25. The diluted samples were boiled for four minutes and then returned to the pre-boiling volume by the addition of sterile distilled water. 5 μ l of the treated sample was loaded per lane in a further 30 μ l of sample buffer.

2.27 : Hybridisation of Dried Agarose Gels (Unblots)

The approach is to hybridise radiolabelled probes directly to nucleic acids trapped within dried gels or "unblots" (Shinnick et al., 1975; Purello and Balazs 1983; Tsao et al 1983; Wallace and Miyada 1987). The method works well with oligonucleotide probes giving an at least five fold higher signal intensity than transfers (Wallace and Miyada 1987). Since probes generated by random priming of purified DNA fragments are longer their accessibility to the DNA molecules trapped in the gel

presents a problem; hybridisation of unblots with such probes give rise to unacceptably high background hybridisation (Schinnick et al., 1975). This problem can be overcome by chemically fragmenting the random primed probes (using HCl) before hybridising them with the unblot (Stoye J.P. personal communication). The advantage of the unblot over conventional Southern blotting is that the result can be obtained in less time and the sensitivity is greater (single copy junction fragments visible after a few hours exposure to X-ray film). However stripping the unblot is more difficult than stripping the signal from Southern blots and thus the number of hybridisations that can be performed on each unblot is limited relative to a nylon filter obtained by Southern blotting. The protocols used for probe preparation drying and hybridisation of the gels are given below.

(a) Chemical fragmentation of the Probe

Probes were labelled as detailed in section 2.18. The probe was purified by increasing the probe volume to 100 μ l loading the entire volume to a 1 ml G50 sephadex column equilibrated with TE buffer and then spinning the column at 1.6 g for 4 minutes in a Sorval HB4 swing out rotor. One μ g of sheared salmon sperm DNA was added to the resultant purified probe and adjusted to 0.1M HCl by the addition of one tenth volume of 1M HCl. This was incubated at 37°C for 40 minutes followed then by the addition of a half volume of 1M Tris HCl pH7.5. The probe was denatured by boiling for ten minutes and was then incubated on ice for five minutes before being added to the hybridisation chamber.

(b) Drying of Agarose Gel

Following FIGE (field inversion gel electrophoresis) the 0.7% agarose gel (Sigma) was immersed in 0.5M NaOH, 1.5M NaCl for 30 minutes (solution changed

after 15 minutes). It was then rinsed thoroughly in distilled water and neutralised by immersion in 0.5M Tris-HCl, 1.5M NaCl with gentle shaking for 30 minutes at room temperature (solution changed after 15 minutes). The gel was then placed on two sheets of Whatmann 3MM and the other side covered with plastic wrap. It was then placed paper side down on a gel drier and dried at room temperature under vacuum for 30 minutes until nearly flat. The heater of the gel drier was then turned to 60°C and the gel dried for a further 20-30 minutes until it is the thickness of X-ray film; care should be taken not to overdry the gel to the thickness of plastic wrap. At this point the gel was marked by punching holes in two corners of the gel using a Pasteur pipette. The gel will never fully rehydrate and can be removed from the paper backing by wetting the unblot and paper and placing it on plastic to which it adheres to preferably and then removing the Whatmann

(c) Hybridisation of the unblot

The gel was placed in a perspex hybridisation box with 100mls of 5x SSPE (1x SSPE contains 150mM NaCl, 10mM NaH₂PO₄, 1mM EDTA with pH adjusted to 7.4 with NaOH) and 0.1% SDS at 55°C. The denatured probe was then added to the box and hybridisation carried out overnight at 55°C. Following hybridisation the unblot was washed four times (5 minutes each) in hybridisation buffer at room temperature, then a further twice (30 minutes each) in 0.1xSSC, 0.1%SDS at 55°C. The gel was then exposed to AGFA Curix X-ray film at -70°C with Du-Pont intensifying screens.

CHAPTER 3 : RESTRICTION ENZYME ANALYSIS OF HIGH EXPRESSING AATB LOCI

3 1 INTRODUCTION

3.1.1 The Apparent Position Effect Of AATB 35

As discussed in the Introduction to this thesis, the promoter region of the ovine BLG gene has been shown to drive efficient expression of BLG protein in the mammary gland of transgenic mice (Simons et al., 1987). In these experiments BLG was shown to have the expected profile of expression throughout pregnancy and lactation. A subsequent study of BLG expression in transgenic mice demonstrated the expression of BLG in transgenic mice was position independent and copy number dependent (Whitelaw et al., 1992).

The work in this thesis further examines the expression of a hybrid transgene, AATB, that encodes an important biomedical protein, α_1 -antitrypsin. This transgene comprises the ovine BLG promoter linked to a minigene encoding human α_1 antitrypsin (hAAT) (Archibald et al., 1990). The protein coding regions of the AATB transgene are that of the M1 variant of the hAAT gene (Ciliberto et al., 1985). The hAAT sequences included within the transgene are 6.3 kb in size. Intron 1 has been removed but all other introns are present. Also included in the AATB transgene were 1.3 kb of 3' untranslated sequences of the human AAT gene.

Two studies had been performed previously in mice with the AATB transgene. Unlike the ovine BLG transgene, the hybrid transgene was not expressed in a position independent and copy number dependent fashion. In the first study (Archibald et al., 1990) six of the thirteen lines generated did not detectably express the transgene. The lower limit of detection in this study was 0.5 $\mu\text{g/ml}$. Three lines expressed the

transgene in the range of 0.5 mg/ml to 1 mg/ml. But only one line, AATB 35, gave a particularly good expression level of 7 mg/ml. The expression of hAAT in the milk of AATB transgenic mice did not correlate with the transgene copy number

Similar results were obtained in the second study using the same AATB transgene (Carver et al., 1993). In the second study all of the eleven female founder mice expressed the protein in their milk. Again, the levels of expression, were on the whole, very low. Only three of the eleven founder mice expressed at levels greater than 0.2 mg/ml, and only two of the three gave particularly good levels of expression. AATB 46.2 was reported to give 8mg/ml and AATB 45.5, 12.5 mg/ml of hAAT in their milk. The levels of expression obtained from transgenic mice in this study also did not correlate with transgene copy number.

AATB 35 appeared to harbour a low transgene copy number. The unusually high expression of hAAT in line AATB 35, relative to the other lines harbouring the construct, could have been the result of either a mutation occurring within the transgene sequences, or a direct consequence of the nature of the murine genomic sequences into which the transgene had inserted (Al-Shawi et al., 1990)). Position effects have been widely mentioned in the literature (see introduction to this thesis). The term is generally used to explain unexpected patterns of expression (in terms of tissue specificity or levels) obtained in transgenic animals. A positive position effect was considered to be a likely explanation for the unusually high levels of expression of hAAT in line AATB 35 and AATB 46.2. The aim of this work was to study this position effect.

3.1.2 The Proposed Approach

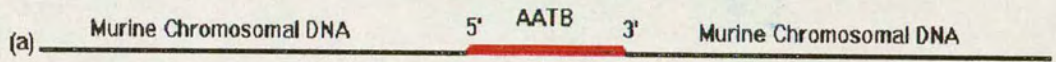
AATB 35 was reported to carry a single copy of the AATB transgene (Archibald et al., 1990). The AATB transgene was 10.6 kb in size. Thus, the single transgene locus of line AATB 35 could have been accommodated easily in a cosmid vector (insert size 32-42 kb). A number of clones harbouring the transgene locus ought to have been detected by screening a cosmid library of the line. The clones would have varied only in the amount of both 5' and 3' flanking DNA present in the insert. Each of the clones would then have been used to generate secondary transgenic animals. Comparison of the hAAT expression in the secondary transgenic lines with the expression in the original AATB transgenic mice (Archibald et al., 1990 & Carver et al., 1993) would have demonstrated if the presence of the murine flanking DNA from AATB 35 gave improved expression of hAAT in transgenic mice. Thus comparison of hAAT expression between the secondary transgenic lines, harbouring different clones, could have been an approach to define the position of sequences that mediate the positive position effect. A diagrammatic representation of the approach is given in figure 3.1

Secondary transgenic animals could also have been generated by re-introduction of the cloned AATB transgene with minimal flanking DNA. This would have demonstrated if the high expression of AATB 35 was due to mutation of the transgene sequences (Al Shawi et al 1990).

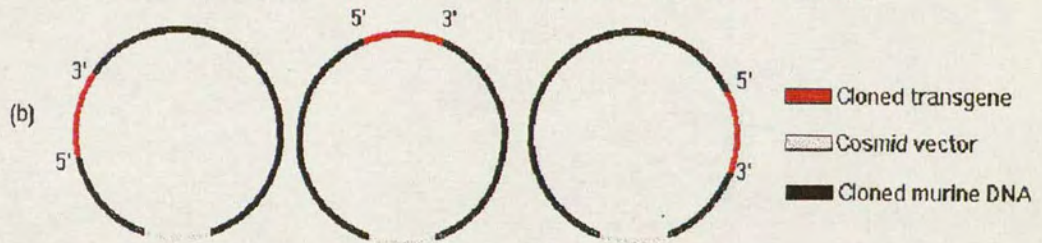
Figure 3.1 : Proposed Approach for study of position effect in line AATB 35

The diagram represents the proposed approach for the study of the position effect in line AATB 35. Section (a) is a representation of the reported single copy transgene locus of AATB 35 (Archibald et al., 1990). It shows the single AATB transgene and the flanking murine, chromosomal DNA. Section (b) represents a number of clones isolated from a cosmid library of line AATB 35. The clones differ in the amount of 5' and 3' flanking chromosomal DNA present in the cloned insert. Section (c) represents the loci of the secondary transgenic animals, (n) represents the number of transgenes present in the locus. This would vary for each line. Comparison of hAAT expression in the secondary transgenic mice with that of the primary AATB mice generated by Archibald et. al. and Carver et al. would have demonstrated whether the murine sequences had any positive effect on both the level and frequency of expression of AATB.

Locus of line AATB 35



Clones Isolated from cosmid library of line AATB 35 (Minimum insert 32 kb)



Generation of Secondary transgenic mice by pronuclear injection of inserts

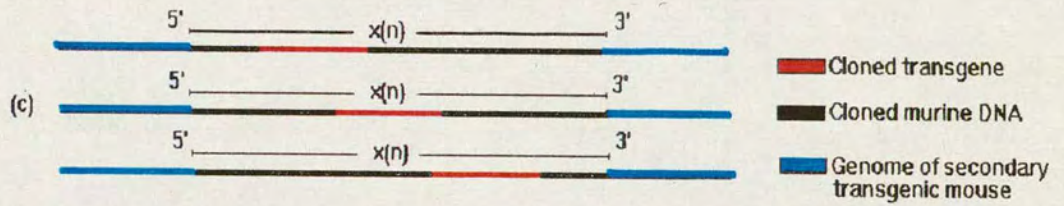


Figure 3.2 : Original copy number estimate of line AATB 35

The figure shows the results of the original screen of the AATB lines by Southern blot analysis (courtesy of Alan Archibald). Each of the lanes contain 10 µg of mouse tail DNA. The probe is the entire AATB construct including the p-Polly III-I vector (digested with Bam HI). The copy number of all transgenic founder mice was estimated by comparing the intensity of the bands on the autoradiograph with that of one copy and ten copy equivalent plasmid controls. Bam HI generated two internal fragments from AATB of 6.3 kb and 1.8 kb. Fragments of this size are generated by Bam HI digestion of AATB 35. The intensity of these fragments is low and was comparable with that of the single copy plasmid control, suggesting AATB 35 had a low copy number locus. However the presence of a 2.5 kb Bam HI band from AATB 35 intimates that AATB 35 may not actually be single copy. A 2.5 kb Bam HI fragment is not an internal fragment of AATB, but would be generated from an AATB head to tail array.

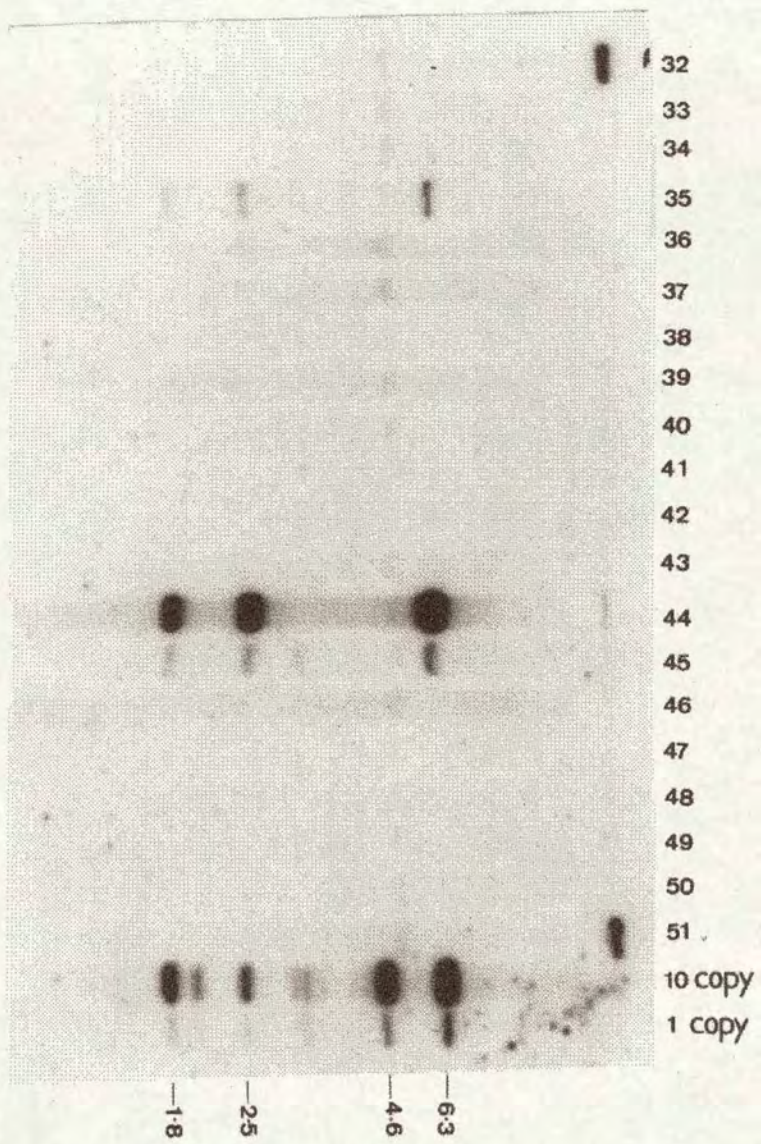
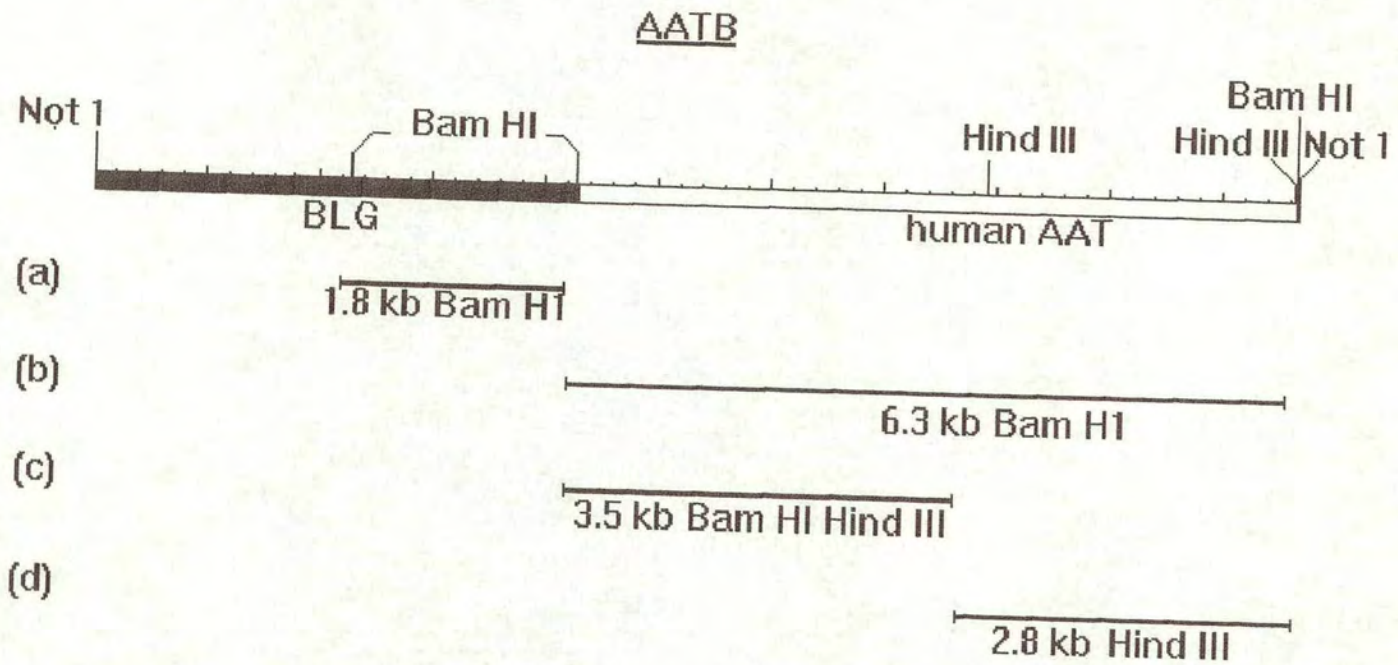


Figure 3.3: Probes Used In Genomic mapping Of Transgenic Mice

The AATB transgene has three Bam HI sites. Digestion of AATB gives two internal fragments one 1.8 kb in size (probe (a)) and a larger fragment of 6.3 kb (probe (b)). These fragments were gel purified and used as probes in the estimation of copy number and the genomic mapping of the transgenic mice. AATB also had two Hind III restriction sites generating an internal 2.8 kb fragment (probe c) that was used to identify which of the junction fragments contained 3' AATB sequences (figure 3.12). A fourth probe (probe d) generated by Bam HI / Hind III double digestion was used to probe the unblot (figure 3.11).



3.2 ESTIMATING THE COPY NUMBER OF AATB 35

3.2.1 Copy Number Estimation Of AATB 35 By Comparison With Plasmid Controls

A more detailed analysis of the AATB 35 locus was carried out before starting the experiments. It had previously been reported that AATB 35 harboured a single copy of the AATB transgene (Archibald et.al., 1990). The result was obtained by comparing the intensity of internal fragments of the transgene, on Southern Blots, with 1 copy and 10 copy plasmid controls (Figure 3.2). This method of estimating transgene copy number is standard practice, but does need to be interpreted with a certain amount of caution, as significant errors can be made using this technique.

Since a single copy or very low copy locus was crucial for the planned approach of this work it was important to establish definitively the copy number of the locus. The copy number estimation was repeated employing the same method used previously. The results are shown in Figure 3.4.

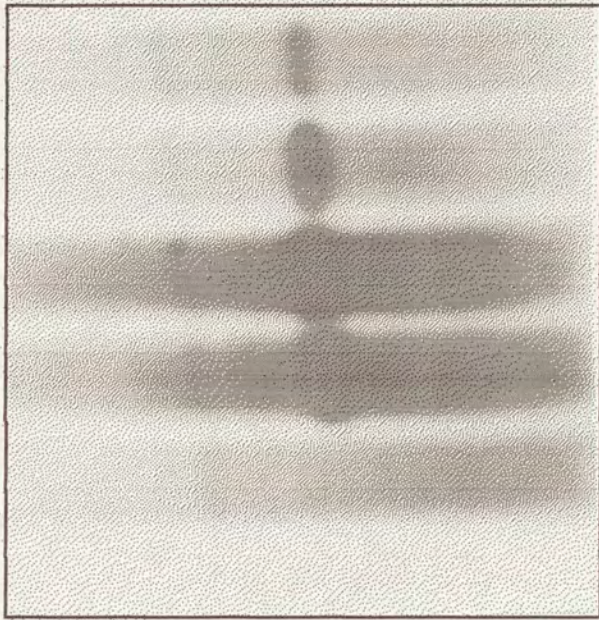
It is quite clear from this Southern blot that the locus of AATB 35 has multiple copies of the transgene. Comparison of the internal fragment with the 5 copy plasmid control would indicate that the transgene copy number of the locus is considerably in excess of five copies and certainly not one as reported (Archibald et al., 1990). As previously emphasised this method of copy number estimation is susceptible to a number of errors. The accuracy of this method relies on three factors; firstly; accurate determination of DNA concentration, secondly; accurate serial dilution of the plasmid controls and thirdly; accurate pipetting of generally viscous genomic DNA. Thus this method of determining transgene copy number is not entirely reliable. Nevertheless

Figure 3.4 : Determination Of AATB 35 Copy Number By Comparison With

Plasmid Control Autoradiograph of copy number control experiment

The figure shows a repeat of the copy number estimate of line AATB 35 using plasmid copy number controls. All the lanes contain 10 ug of mouse genomic DNA prepared from liver. The plasmid copy number controls were loaded in 10 µg of non transgenic mouse DNA. The lane labelled AATB(1) contained 10 µg of genomic DNA of a female third generation mouse AATB 35.11.60. The lane labelled AATB(2) contained 10 µg of genomic DNA of male third generation mouse AATB 35.8.11. The control lane contains 10 µg of liver DNA from a non transgenic mouse. DNA in all lanes was restricted with Bam HI endonuclease. Bam HI generates two internal fragments from AATB a 6.3 kb fragment from the human AAT sequences and a 1.8 kb fragment from the ovine BLG sequences. The probe used in this instance was probe (b) Figure 3.3. The Southern blot was laid down with X-ray film for four days at -70°C with intensifying screens. The copy number of both AATB 35 mice was similar and appeared to be well in excess of 5 copies.

5.3



1 copy

5 copy

AATB(1)

AATB(2)

control

the disparity of the results is larger than expected and, was not easily explained by simple technical error.

The animals analysed in the Southern blot in figure 3.4 are third generation, whereas the original copy number determination was performed on the founder AATB 35. It is conceivable that this animal may have been a somatic mosaic. If this were so then the copy number of the animal would be an underestimate since only a percentage of the cells would harbour the transgene array. However the frequency of transmission of the transgene to the offspring was close to fifty percent (data not shown). This implied that the animal was not germline mosaic but does not eliminate the possibility of somatic mosaicism.

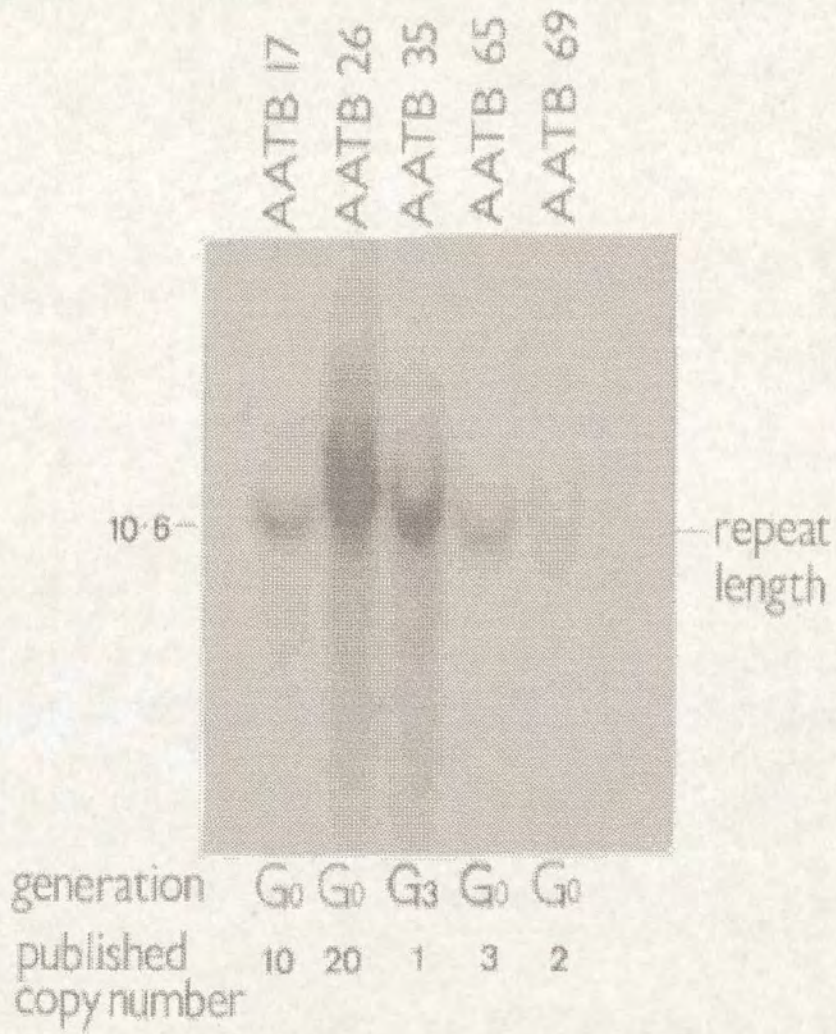
Another possibility is that the locus of AATB 35 had undergone expansion over the generations. This is unlikely as both third generation animals analysed in figure 3.4 were obtained from different offspring of the AATB 35 founder. It is unlikely that expansion of the locus would occur similarly in two independent transgenic mice. Moreover, if this were so then a concomitant change in the level of expression of hAAT would be expected. The level of human AAT in the milk of a third generation animal of line AATB 35 was measured on the 11th day of lactation by radial immunodiffusion, and found to be equivalent to the level determined in the G₁ animals assayed previously (Archibald et al., 1990). The founder AATB 35 was male and thus no milk sample was analysed at G₀. It is likely that AATB 35 was somatic mosaic but in the absence of DNA from AATB 35 this could not be proved.

The original strategy for this project had been to clone the entire AATB locus from this line into a cosmid vector (see figure 3.1). As the AATB transgene is 10.6 kb in size this would only have been possible if the locus harboured three or less copies of

the AATB transgene. This would, nevertheless, considerably diminish the amount of flanking chromosomal DNA within the inserts of the cosmid clones. Since the murine sequences mediating the effect might have been some distance from the transgene locus this had an important bearing on the suitability of the planned approach for the study of this line. To be absolutely sure of these results another method was employed to accurately establish the transgene copy number of line AATB 35.

Figure 3.5 : Comparison Of Third Generation Mouse of AATB 35 With Go Mice Of Other AATB Transgenic Lines

The figure shows the results of a Southern blot of DNA of various AATB transgenic mice. Each lane contains 10 ug of genomic DNA digested with Eco RV restriction endonuclease. The DNA from AATB 17, 26, 65 and 69 were the tail DNA from the original screening of the AATB lines generated by Archibald et al. this lab. The DNA of AATB 35 is from a third generation transgenic mouse and is liver DNA. All of the lines on the Southern blot generated a 10.6 kb band. This is evidence that these AATB lines harbour a head to tail array of the AATB transgene. The relative intensities of these bands gave an indication of the relative copy numbers of the lines. The published copy number for each line is given below each lane. Qualitatively the results of this blot support the published copy number for the all of the lines except AATB 35.



3.2.2 Copy Number Estimation Of AATB 35 By Junction Fragment Comparisons

It was apparent from the previous result that the locus of AATB 35 contained an array of transgenes. In order to gain further information about the array, liver DNA from AATB 35 was restricted with Eco RV, an enzyme which cuts at a unique site within the AATB transgene. This analysis confirmed both the orientation of the transgenes within the array and allowed a more accurate determination of their number. Head to head and tail to tail orientations would have given fragments of a defined length for each enzyme: such fragments (13.2 kb for head to head, and 8 kb for tail to tail, upon digestion with Eco RV) were not detected. Complex loci which have other non transgene DNA present, or transgenes harbouring deletions, would produce bands of sizes which were not predicted from the known restriction map of AATB, in addition to two junction fragments. The generation of only three bands upon Eco RV digestion of the genomic DNA (figure 3.6) demonstrated that the transgene array of AATB 35 had a simple head to tail arrangement. The only internal fragment from the array was a 10.6 kb AATB repeat length fragment. This indicates that all the transgenes within the array were orientated in head to tail fashion. The other two fragments of 9 kb and 6.5 kb were assumed to be the transgene/chromosomal junction fragments. The intensity of the three fragments were compared by densitometry of the autoradiographs. Figure 3.7 shows the densitometry of the 9kb junction fragment and the 10.6 kb junction fragment.

Junction fragments are usually single copy. Using a probe which spans the restriction site and comparing the intensity of the internal array fragments to the junction fragment in each lane gave a clearer indication of the copy number of the locus than using plasmid copy number controls.

Figure 3.6 : Autoradiograph Of Eco RV Digested Genomic DNA of Line AATB

35

All the lanes contain 10 ug of liver DNA from AATB 35.8.11, digested with Eco RV. The numbers above the lanes are the time (in hours) digestion was allowed to proceed, o/n= overnight digestion with extra enzyme added. The time course of digestion was performed to ensure that the fragments being compared on the Southern blot were indeed the junction fragments and the internal AATB fragment from the array: as opposed to fragments generated by either partial digestion of the DNA, or star activity of the enzyme. The digested DNA was electrophoresed in a 0.7% agarose gel. The gel was Southern blotted onto hybond N membrane and hybridised with probe (b), Figure 3.3. Only three fragments were generated: One intense AATB repeat length fragment and two less intense junction fragments (sizes given in kb on the left). The intensity of the two junction fragments was similar. Each junction fragment is present in only one copy per genome and thus comparison of the signal from the junction fragments with that of the repeat length fragment gave an indication of the number of transgenes present within the array (see figure 3.7).

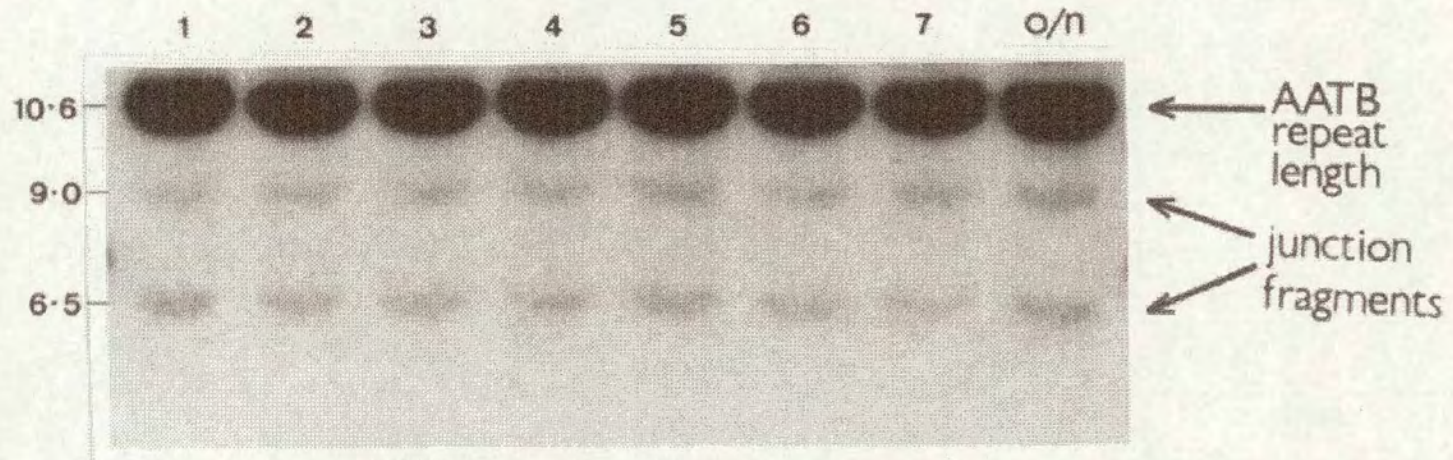


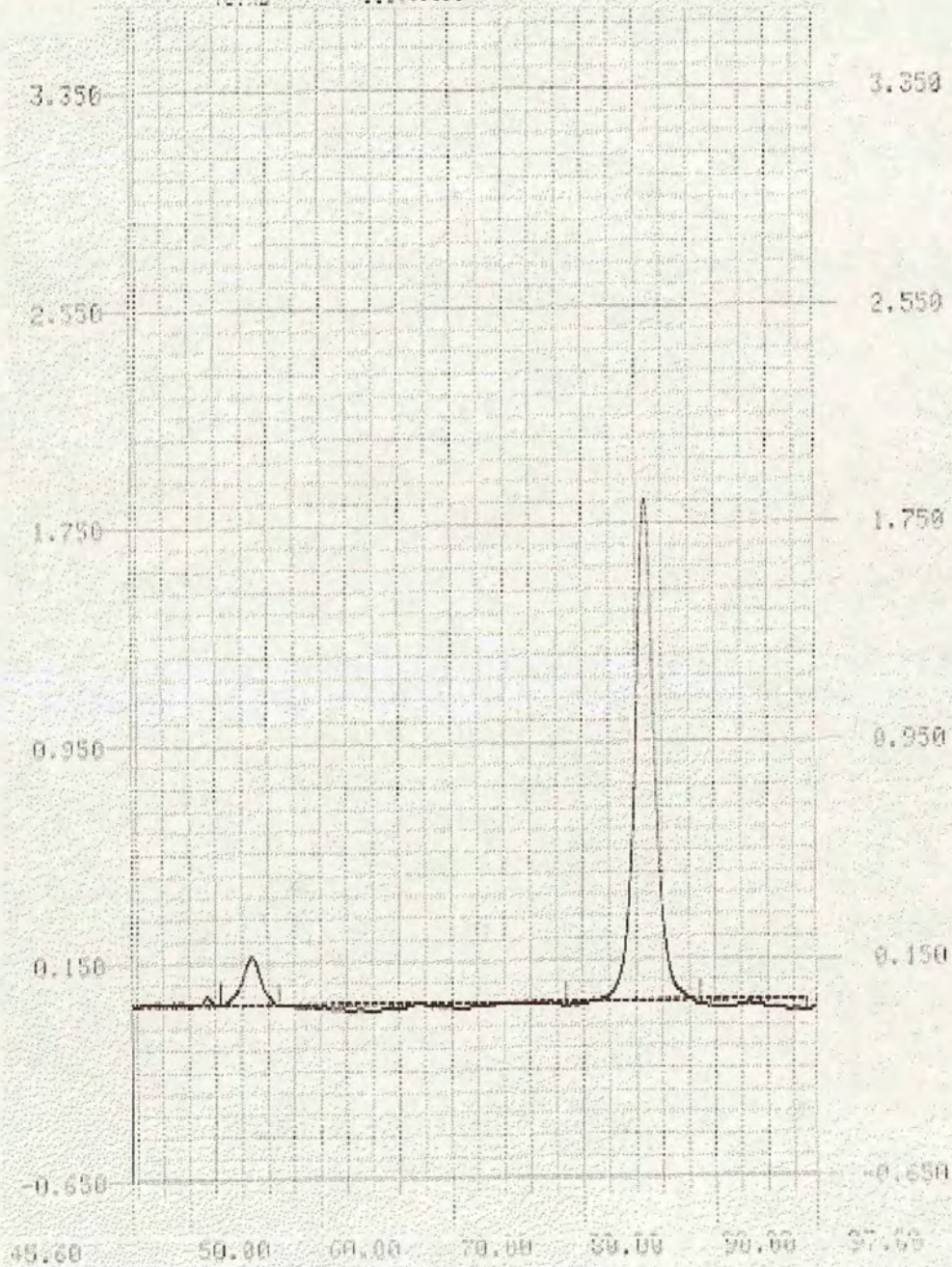
Figure 3.7 : Densitometry Of The Autoradiograph Shown In Figure 3.6

The intensity of the three Eco RV fragments (figure 3.6) was compared by densitometry. All lanes were scanned using a Shimadzu densitometer. The intensity of both junction fragments were found to be very similar but the background was always greater for the larger junction fragment which was closest to the repeat length band. The background of the repeat length band and the larger junction fragment was similar: this allowed an accurate comparison of the two fragments. The figure shows the densitometry result of the larger junction fragment and the repeat length fragment of lane 4, figure 3.6. The ratio of the intensities is one to ten. This result cannot be considered absolutely quantitative because of the non linear response of X-ray film. Nevertheless it does show that the locus of AATB 35 will not fit within a single cosmid vector.

LANE NO. 1

NO.	Y POS.	AREA	CONC.	MARK	%
1	54.8	7217.525		U	8.3
2	84.7	79298.659		U	91.6

TOTAL 86516.184



The densitometry (figure 3.7) confirms the result from the plasmid copy control experiment. The ratio between the junction fragment and the internal fragment is one to ten. This method of estimating transgene copy number is not exact and a number of points need to be taken into consideration. Firstly nothing is known about the nature of the junction fragments and thus of the proportion of AATB sequences within those fragments capable of hybridising to the probe. The assumption was made that the probe sequences 5' of the Eco RV site were also present in the 5' junction fragment and similarly the probe sequences 3' of the Eco RV site were also present in the 3' junction fragment. The Eco RV restriction site is not quite in the centre of the probe sequences. The significance of this would appear to be minimal as analysis of the two junction fragment bands by densitometry shows the intensity of the bands is not significantly different (data not shown). However the internal fragment hybridises with the entire probe whereas the junction fragments only hybridise with half of the probe sequences. Finally the response of X-ray film is only linear over a very limited range of signals, for this reason the weak signal from the junction fragment may be under represented relative to the intense signal of the internal fragment and vice versa. The Southern blot shown in Figure 3.5 suggests that the copy number of AATB 35 may be considerably more than eleven (suggested by the densitometry result). Figure 3.5 shows the autoradiograph of a Southern blot of a number of AATB lines digested with Eco RV. Although the junction fragments are not clear for any of the lines on this blot. It is clear that all the AATB lines analysed on this blot gave rise to an AATB repeat length fragment on digestion with Eco RV. This demonstrated the presence of a head to tail transgene array in all of these lines. A copy number standard was not loaded on the Southern in figure 3.5. Nevertheless the figure provides qualitative

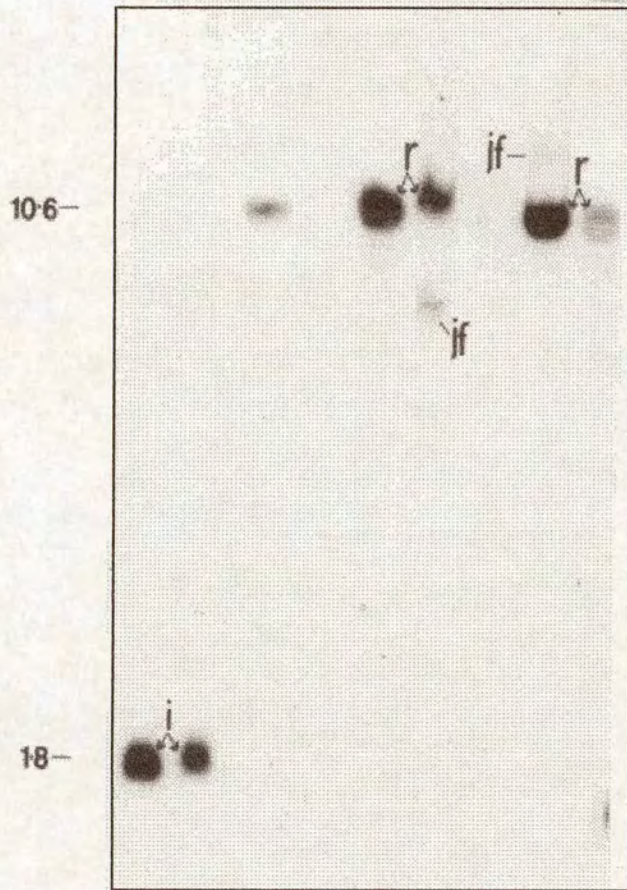
support for the published copy numbers of the AATB lines analysed (other than AATB 35). This supports the assertion that the underestimation of the AATB 35 copy number was not the result of an error in the plasmid copy number controls but was the result of somatic mosaicism of the founder.

Taking all of these factors into account the fact remains that the copy number of AATB 35 was in excess of what had been expected and posed a real problem for the planned approach for studying position effects with this line. A locus of this size could not conceivably be accommodated within either a cosmid or even a P1 vector. In addition it had been thought that the 7mg/ml expression of hAAT in line AATB 35 was driven from one single copy of the AATB transgene. The analysis of the copy number outlined in this chapter shows that the transgene copy number is probably in excess of eleven copies. This means that the expression per copy of the transgene is at best 0.64 mg/ml.

Figure 3.8 : Comparison of AATB 46.2 With AATB 35

The figure shows an autoradiograph of a Southern blot of AATB 35 and AATB 46.2 liver DNA probed with probe (a) figure 3.3. C= control (non transgenic liver DNA), P= one copy equivalent plasmid control (p-Poly-III-I-AATB digested with Not I), 35= AATB 35 liver DNA (10 µg), 46.2= AATB 46.2 liver DNA (10 µg), i= internal AATB fragment, jf= chromosomal/transgene junction fragment, r= AATB repeat length. Comparison of the 1.8kb internal Bam HI fragment of AATB 35 and AATB 46.2 demonstrated that the copy number of AATB 46.2 was less than that of AATB 35. The generation of a repeat length fragment with Eco RV and Nhe I demonstrated that both AATB 35 and AATB 46.2 harbour a head to tail array of transgenes. Comparison of the intensity of the repeat length band generated by the two lines supports the data of the internal fragment: indicating the copy number of AATB 46.2 was lower than that of AATB 35. However comparison of the Eco RV junction fragment of AATB 46.2 with the repeat length fragment suggested that more than one repeat length fragment was generated by Eco RV digestion of AATB 46.2. The information from this blot indicated that the copy number of AATB 46.2 was at least three but probably more.

BamHI EcoRV NheI
C 35 462 P C 35 462 C 35 462



3.3 ESTIMATION OF THE COPY NUMBER OF AATB 46.2

3.3.1 The Apparent Position Effect In AATB 46.2

Only two stable lines of AATB transgenic animals were generated which expressed the human protein at relatively high levels. The other stable high expressing line was generated in the second study (Carver et al., 1993). The line AATB 46.2 generated at Pharmaceutical Proteins Limited (now PPL Therapeutics) had very similar characteristics to AATB 35. It expressed human α_1 -AT at a comparable level (6-8 mg/ml) and the locus was also reported to be low copy. The high expression level of the human protein in the milk of lactating females had been maintained through several generations. If the copy number of this line was significantly lower than that of AATB 35 this line would be more amenable for the study of position effects using the previously outlined approach. Thus the copy number of AATB 46.2 was also investigated further.

3.3.2 Copy Number Comparison of AATB 35 with AATB 46.2

The first criterion was to compare the copy number of the AATB 46.2 with that of AATB 35. Liver DNA of one mouse from each line was compared using Southern blot analysis. The result of this analysis is given in figure 3.8. Bam HI generates two internal fragments of the AATB transgene. Comparison of the intensity of the Bam HI internal fragments from the two lines demonstrates the copy number of AATB 46.2 is lower than that of AATB 35. Therefore it was worth proceeding with a more in depth investigation of the copy number of this line as it was likely to be more

suitable for the study. However the enzymes Eco RV and Nhe I generated a 10.6 kb fragment that indicated the presence of a head to tail array of transgenes within the locus of both lines. The intensity of the repeat length fragment of AATB 46.2 was less than that of AATB 35 consistent with the copy number of AATB 46.2 being lower than that of AATB 35. However junction fragments liberated from the loci by these enzymes were visible in all lanes. The intensity of the junction fragments (assumed to be the intensity of a single copy fragment) was less than that of the internal 10.6 kb band generated from both lines. Densitometric analysis of this Southern blot (data not shown) demonstrates that the transgene copy number of AATB 46.2 is at least three.

A copy number of greater than three prohibits the cloning of the entire AATB transgene locus in a single cosmid vector. It was, therefore, important to more accurately determine the copy number of AATB 46.2. The approach taken in order to circumvent the non-linear response of X-ray film was to perform a dilution blot analysis.

3.3.3 Copy Number Estimation Of AATB 46.2 By Dilution Blot Analysis Of AATB 46.2

To obtain a quantitative result from the type of analysis done in figures 3.6 and 3.7 it was essential that both signals strong and weak were still within the linear range of X-ray film. The rationale of the dilution blot is to dilute the 10.6 kb internal fragment produced from the array till the signal from this, equates with that of the junction fragments, when hybridised with the same probe. The comparison in this case is like with like (weak signals with weak signals) so the effect of the non linear response of the film is circumvented.

The dilution factor required to bring the internal fragment to the same intensity as the junction fragment will give an estimated copy number of the array and hence of the line (which is the dilution factor plus 1). The dilution blot analysis is shown in figure 3.9. For the intensity of the repeat length fragment to match that of the junction fragment requires between a one in four, and, one in eight, dilution. Thus the repeat length fragment appeared to contain the equivalent of six copies of the transgene. Hence the total number of copies of the transgene in AATB 46.2 was approximately seven. Although the copy number of AATB 46.2 was less than that of AATB 35, the locus was still too large to be accommodated in a single cosmid clone.

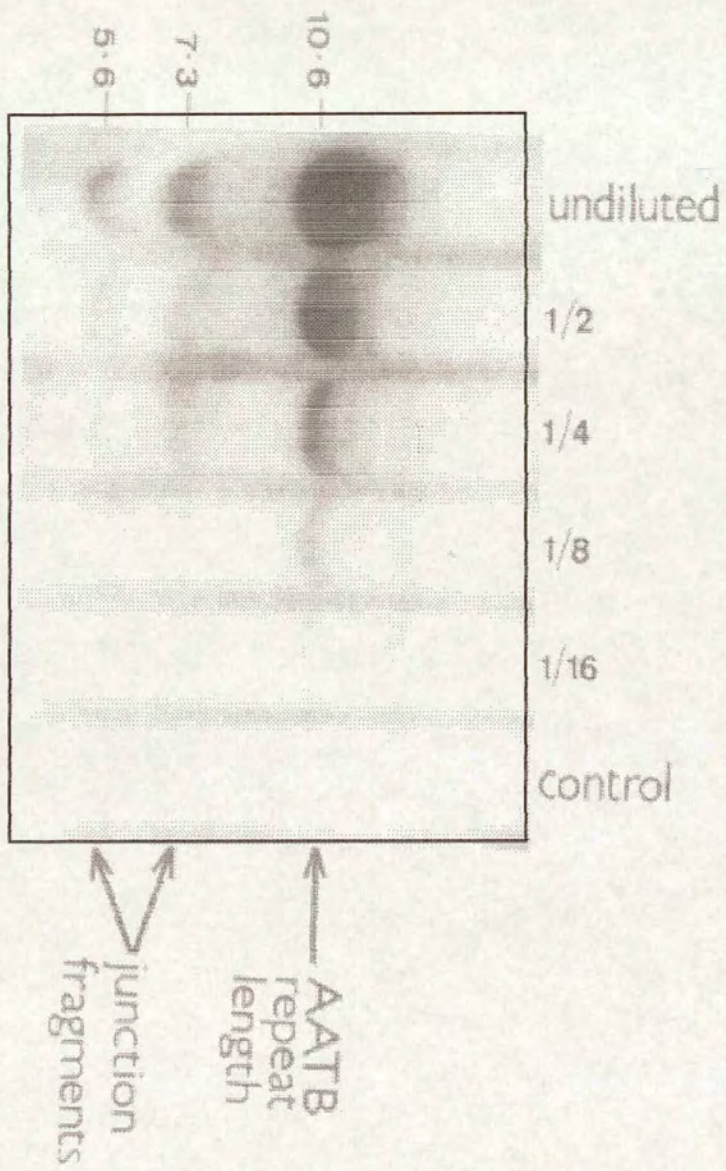
There were only two AATB lines from the previous studies that were stable and gave high levels of hAAT expression. From the more careful blotting analysis in this study neither of these lines were shown to harbour the very low transgene copy number loci anticipated, on the basis of previous experimental results.

The lowest copy locus AATB 46.2 could conceivably be cloned in a P1 vector. This would allow the entire transgene array plus some of its flanking regions to be cloned. Unfortunately inserts in P1 vectors are frequently unstable. The insert in this instance would contain regular repeating units of AATB and it was considered technically unwise to use this approach. The locus of both lines could easily be accommodated within a YAC but such vectors harbour very large inserts and this would be a very non informative way of regenerating the position effect.

Nevertheless the murine genomic DNA flanking the transgene arrays in both lines were apparently permissive for high levels of expression of human α_1 -AT. The nature of these sequences and their role in promoting the high levels of transgene expression was, therefore, still of considerable interest.

Figure 3.9 : Estimation Of Copy Number Of AATB 46.2 By Dilution Blot Analysis

The figure shows a Southern blot of an AATB 46.2 dilution gel. Each of the lanes contain 10 µg of DNA, digested with Eco RV. The probe was probe b figure 3.3 The numbers above each lane gives the dilution factor of the transgenic DNA. The lanes vary only in the ratio of AATB 46.2 DNA and non transgenic mouse DNA present. The first lane contained 10 µg of liver DNA from a mouse of line AATB 46.2. The next lane contained a 1:1 ratio of non transgenic liver DNA and AATB 46.2 liver DNA (5µg of each). The third lane contains a 1:3 ratio of transgenic to non transgenic DNA. The fourth lane contains a 1:7 ratio of transgenic to non transgenic DNA and the fifth lane contains a 1:15 ratio. The final lane contains only non transgenic DNA. The dilution of the transgenic DNA reduces the intensity of the repeat length band in accordance with the dilution factor. Comparing across the lanes, the intensity of the single copy junction band (in the lane containing the undiluted transgenic DNA) with the diluted repeat length band allows the number of transgenes liberated from the array by Eco RV to be estimated. The intensity of the junction fragments lies between that of the repeat length band diluted 1/4 and 1/8. This suggests that approximately six AATB repeat length fragments were liberated from AATB 46.2 by Eco RV. The copy number of AATB 46.2 is approximately seven copies.



3.4 THE ALTERNATIVE APPROACH

3.4.1 Constructing A Secondary Transgene From Cosmid Clones Harboured The AATB 46.2 Flanking DNA

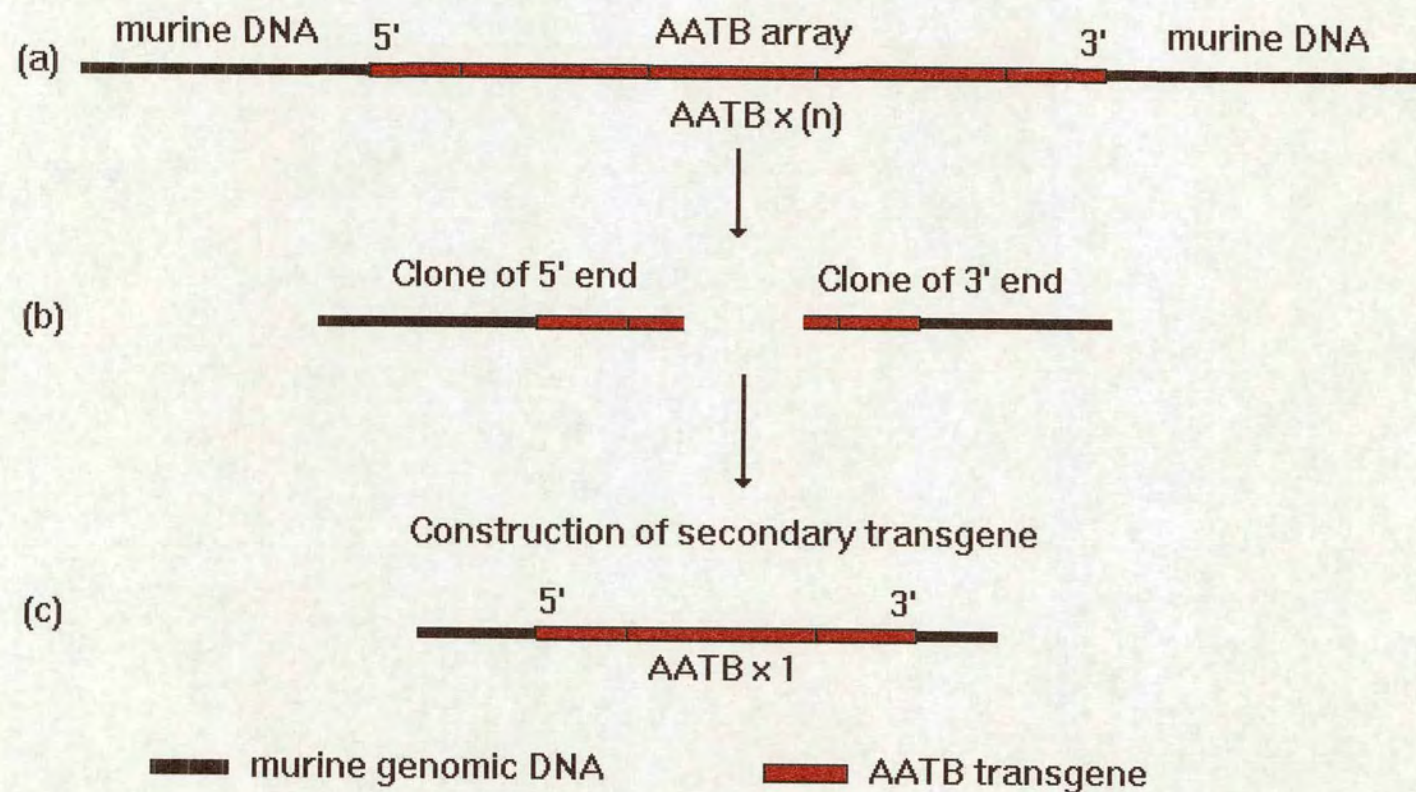
The locus of all the lines of AATB analysed appeared to harbour a head to tail array (Figure 3.5). It was thus concluded that the concatameric arrangement of transgenes within the array of AATB 35 or AATB 46.2 was not responsible for the high expression of the hybrid transgene in these lines. The approach decided upon was to proceed with the cosmid vector cloning of the locus (see figure 3.10). The proposal was to construct a cosmid library from liver DNA of an AATB 46.2 transgenic mouse, screen the library for clones that contained the transgene and isolate from these clones those that contained flanking genomic DNA. These flanking genomic sequences would be used to construct a secondary transgene comprising just one single copy of the AATB transgene, plus the chromosomal DNA from both the 5' and 3' ends of the locus.

The approach was not as neat as the initial proposal in which the locus itself remains unchanged and a number of clones containing varying amounts of genomic flanking DNA both 5' and 3' were likely to be isolated. It would be impossible to regenerate the position effect per se using this approach since the locus would be dramatically altered. Nevertheless, a study of the chromosomal flanking sequences and their role in the high level expression of either line was still possible. AATB 46.2 was the line of choice for the study since it had the lowest transgene copy number thus minimising as much as possible the possibility of recombination occurring between

direct repeats harboured within the cosmid inserts and maximising the chances of cloning both the 5' and the 3' flanking sequences.

Figure 3.10 : The Alternative Approach

The figure is a diagrammatic representation of the planned approach for the study of the position effect in AATB 46.2. (a) Represents the locus of AATB 46.2, (n) is the number of transgenes within the array. (b) represents clones of the 5' and 3' end of the loci which harbour significant amounts of flanking murine DNA. (c) represents the secondary transgene construct made from the isolated clones and one intact copy of the AATB transgene.



3.5 THE RESTRICTION MAP OF THE LOCUS OF AATB 46.2

A map of the locus of AATB 46.2 was required before the locus could be cloned from a cosmid library of the line. A number of enzymes were identified which cut either uniquely or infrequently within the AATB transgene (see figure.4.3). Restriction of AATB 46.2 genomic DNA with such enzymes followed by Southern blotting and hybridisation with 5' and 3' probes allowed the generation of a map of the locus.

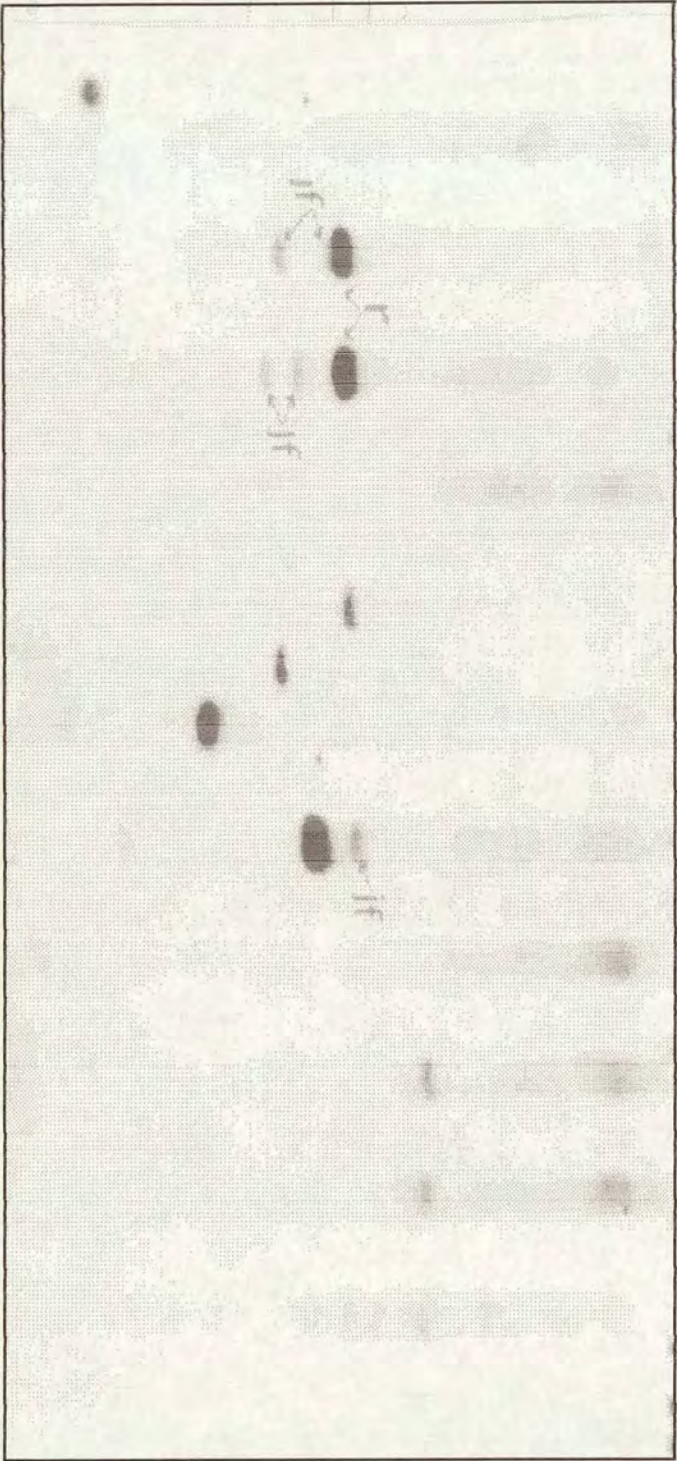
It has previously been established that the locus harbours a simple array of transgenes oriented in head to tail fashion. This array generates intense internal fragments on Southern blot analysis. All weaker fragments detected were assumed to be the junction fragments of the locus. Southern blot analysis of Bam HI digested AATB 46.2 DNA detected only one junction fragment because of the probes used (figure 3.12). likewise only one Hind III junction fragment was detected (figure 3.11). Two junction fragments were detected upon digestion with Eco RV and Nhe I (figures 3.11 and 3.12). Sufficient information was available from these blots to confirm the structure of clones harbouring genomic DNA from the AATB 46.2 locus. The genomic map of the AATB locus is given in figure 3.13.

The figure gives a restriction enzyme map of the AATB 46.2 locus. The locus has six copies of the transgene within the array. The two transgenes abutting the Chromosomal DNA are represented as entire copies. But the extent of transgene damage (if any) at the 5' end of the array is not clear from this analysis. A 10.6 kb junction fragment can be detected from the locus with the 6.5 kb Bam HI probe. This demonstrates that the Bam HI restriction site that is normally present at the very 3' end

of the AATB transgene is missing in one of the transgenes. The single copy fragment is 10.6 kb, this is marginally larger than expected (8.9 kb) if the missing restriction site is from a transgene within the array. Thus the damage has occurred to the transgene abutting the host chromosomal sequences at the 3' end of the locus. This suggests that this transgene has sustained some damage either before, or during integration into the genome. No 3' Hind III junction fragment is detected with the same probe which suggests the Hind III site situated 150 bp from the 3' end of the AATB transgene is still present on the damaged transgene.

Figure 3.11 : Detection Of Junction Fragments

The figure shows an unblot of variously digested AATB 46.2 liver DNA probed with probe b, figure 3.3. The DNA was digested with the enzymes indicated. The gel was a 1% LMP agarose gel. The digested DNA was loaded on the gel and subjected to field inversion gel electrophoresis (FIGE). Each lane contained 10 µg of digested DNA. The gel treated with HCl followed by NaOH and finally tris HCl pH 7.4. It was then dried and hybridised directly with the probe. The AATB fragment generated by enzymes Sal I, Not I and Pvu I would appear to be too large to have entered the gel under the conditions used (see materials and methods). Enzymes Bgl II and Sca I generate one single large fragment. The actual size of the fragment could not be determined as fragments greater than 30 kb were not resolved. Enzymes Nhe I and Eco RV generate the repeat length fragment (r) and two other faint fragments which are believed to be the chromosomal/transgene junction fragments (jf). Hind III generates the 7.8 kb fragments expected from an AATB head to tail array and a less intense fragment which is believed to be the junction fragment. The lane labelled 1 copy contains a one copy equivalent plasmid control which is P-Polly-III-I-AATB digested with Not I (i.e. a one copy equivalent of the injected transgene. The lane labelled Bam HI (human) contains 10 µg of human mammary tumour DNA digested with Bam HI.



Sal I
Nhe I
EcoRV
Not I
1 copy
Bam HI (human)
Sac I
Hind III
Pvu I
Bgl II
Sca I
Sma I

22
13.3
9.7
5.1

Figure 3.12 : Identification of 5' Hybridising And 3' Hybridising Junction Fragments

The figure shows the results of a Southern blot analysis to detect which of the junction fragments hybridised with a 5' AATB probe (a, figure 3.3) and which with the 3' AATB probe (d, figure 3.3). r= AATB repeat length fragment (10.6 kb), i= internal AATB fragment and jf= chromosomal/transgene junction fragment. Each of the lanes contained 10 µg of liver DNA digested with either Bam HI , Eco RV or Nhe I. For each enzyme the first lane is the result of hybridisation of the filter with probe a. The blot was exposed to X-ray film for 48 hours, at -70 ° C with intensifying screens. The second lane is the result of hybridising the same filter with probe c. In this instance the blot filter was exposed to X-ray film for 72 hours, at -70 ° C with intensifying screens. The size of the junction fragments are given in kb.

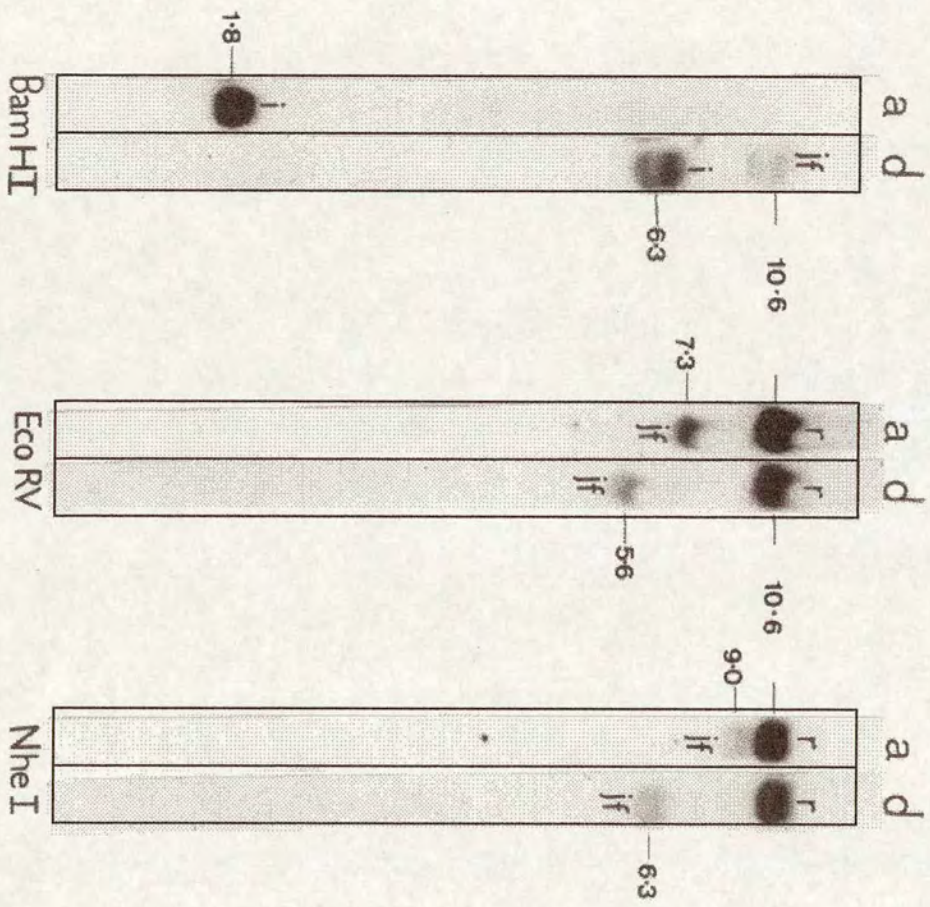
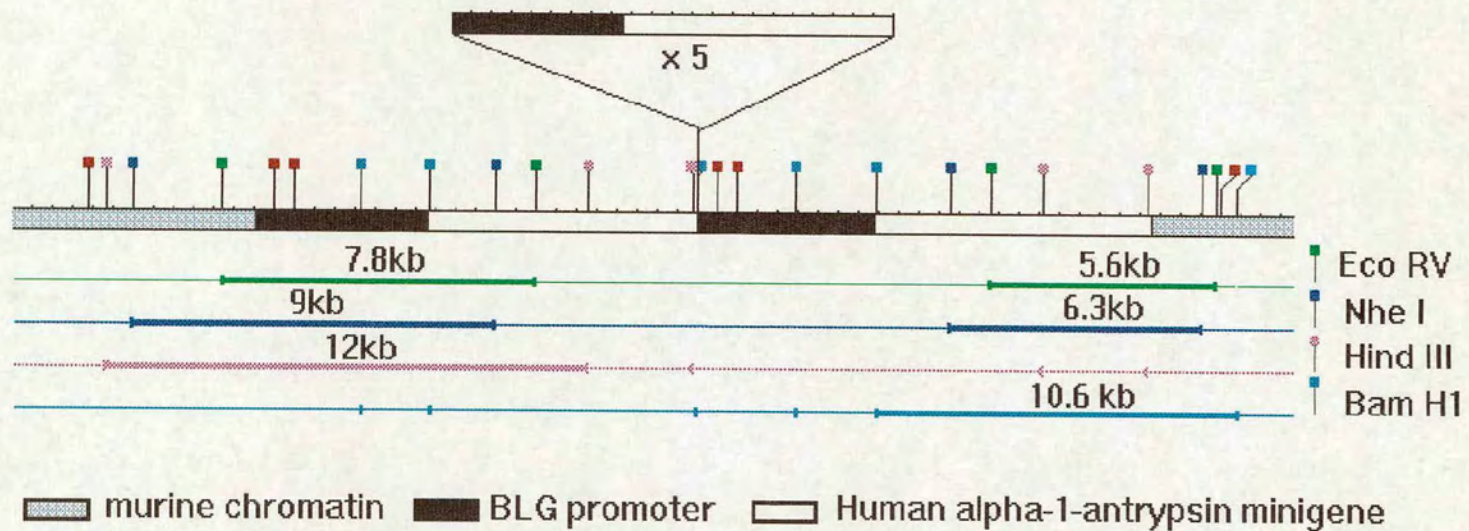


Figure 3.13 : Estimated restriction Map Of Locus Of AATB 46.2

The figure shows a restriction map of the locus of AATB 46.2 as determined by southern blot analysis (detailed earlier in this chapter). The number of copies of the transgene is approximately seven (see figure 3.9). Fragments, that appeared to be the junction fragments, were identified using a four restriction endonucleases and 5' and 3' AATB probes (see figures 3.11 and 3.12). These junction fragments are represented in the figure by thick lines, the size (in kb) of the junction fragment is given above each. The 3' junction fragments were detected with a 2.8 kb probe from the 3' end of AATB (probe d figure 3.3). No 3' Hind III junction fragment was detected (data not shown) that suggested that the 3' most Hind III site of AATB was present and thus the majority of the AATB transgene at the 3' end of the transgene array. However the detection of a Bam HI junction fragment with this probe suggested that the Bam HI site situated 20 bp from the 3' end of AATB was missing from the transgene at the 3' end of the array. The probe used to detect the 5' junction fragments was an internal fragment that did not span the 5' most AATB sequences (probe a figure 3.3) and thus was not as informative about the amount of transgene present at the 5' end of the transgene array.



SUMMARY

Two lines of mice harbouring the AATB transgene were found to express the transgene at unusually high levels in the milk of transgenic mice. The BLG promoter has been used to drive the expression of biomedical proteins in the mammary gland of transgenic animals. Unfortunately the expression levels of heterologous transgenes driven by the BLG promoter are unpredictable and more often than not disappointing. The aim of this work is to investigate the role of the surrounding murine chromosomal sequences in the high level expression of the transgene in one of the lines.

The proposal was to clone the entire locus in a cosmid vector and generate from the clones a number of transgenic lines with a new secondary construct which is the AATB transgene from the high expressing line along with significant amounts of the flanking murine DNA. Crucial to this approach was the presence of a single or very low copy transgene locus within one of the lines.

It has been demonstrated in this Chapter that neither of the high expressing lines possesses a locus of sufficiently low copy number to make the planned approach feasible. An in depth investigation of the transgene copy number in both lines has been performed.

The higher than expected transgene copy number ruled out the original proposal as a method of investigating the position effect of either line AATB 35 or AATB 46.2. An alternative approach to study the position effect in AATB 46.2 was designed. This involves constructing and screening a cosmid library from line AATB 46.2. Clones isolated which harboured the 5' and 3' murine flanking DNA can then be used to construct a secondary transgene to investigate the role of the murine chromosomal DNA, surrounding the locus.

CHAPTER 4 : CLONING AND CHARACTERISATION OF THE AATB 46.2 LOCUS

4.1 INTRODUCTION

4.1.1 Construction of the Cosmid Library

A cosmid library was constructed in Stratagene's Super-cos vector (Evans et al. 1989) (see Figure 4.1). This vector is capable of accommodating 32 to 42 kb of insert DNA. The vector carries a neomycin (*neo*) resistance marker and an SV40 promoter for neomycin gene expression in eukaryotic cells. It has T3 and T7 promoters to make RNA probes from the insert that can, for example, be used for cosmid walking. The vector also has two *cos* sites to allow efficient cloning of non size selected DNA. However, it was found that the DNA had to be size selected to establish the library in this instance.

The genomic DNA was prepared from high molecular weight liver DNA of a transgenic mouse of line AATB 46.2 (Carver et al., 1993), (see Materials and Methods for details of library construction). The DNA was partially digested with Nde II (an isoschizomer of Sau 3A), size selected on a sucrose gradient and dephosphorylated before ligation into the Bam HI cloning site of the Super-Cos-One vector. The bacterial strain infected with the packaged cosmid vector was *E. coli* NM554. This strain harbours a RecA mutation rendering it recombination deficient. It also harbours mutations affecting the *E. coli* restriction system (*mcrA* and *mcrB*) which prevent the cutting of methylated genomic DNA cloned within the cosmid vector.

4.1.2 Screening The Cosmid Library

A library of approximately 8×10^5 clones was obtained. It had been calculated that 2.2×10^5 would need to be screened for a reasonable probability of cloning the locus. This figure was calculated using the equation bellow. (Glover, DNA Cloning, Vol. I, 1985,p13)

$$N = \frac{\ln(1-P)}{\ln(1-x/y)}$$

N = number of colonies required to be screened

P = an arbitrary probability of finding a particular sequence of DNA

x = the insert size

y = genome size

The arbitrary probability used was 95 % (see Glover, DNA Cloning, Vol. 1, 1985, p13). The murine genome is approximately 3×10^6 kb and the average insert size was assumed for the purposes of the calculation to be 40 kb.

The library was titered and the colonies plated at approximately 2×10^4 per filter of 15 cm diameter. A total of 4×10^5 colonies were screened to maximise the chances of cloning the murine flanking DNA of the AATB 46.2 locus.

The library was screened using only the BLG sequences of the AATB transgene (probe (b), Figure 4.2). This was to prevent the detection of cosmid clones harbouring the mouse, endogenous AAT gene which may have cross hybridised with a probe containing hAAT sequences. Multigenic clones were obtained on the primary screen. Seven were obtained on the secondary screen. Analysis of these seven clones by

Southern Blotting demonstrated that only four clones genuinely harboured AATB sequences. Table 4.1 shows the clones obtained at each stage of the screening process.

Table 4.1 : Number of positive clones obtained on each screen

Clones	Primary screen	Secondary screen	Southern blot
1	+	+	
2	+	+	
3	+		
4	+	+	same as 5
5	+	+	+
6	+		
7	+		
8	+	+	+
9	+	+	+
10	+	+	+

4.1.3 Restriction Mapping Of The Cosmid Clones

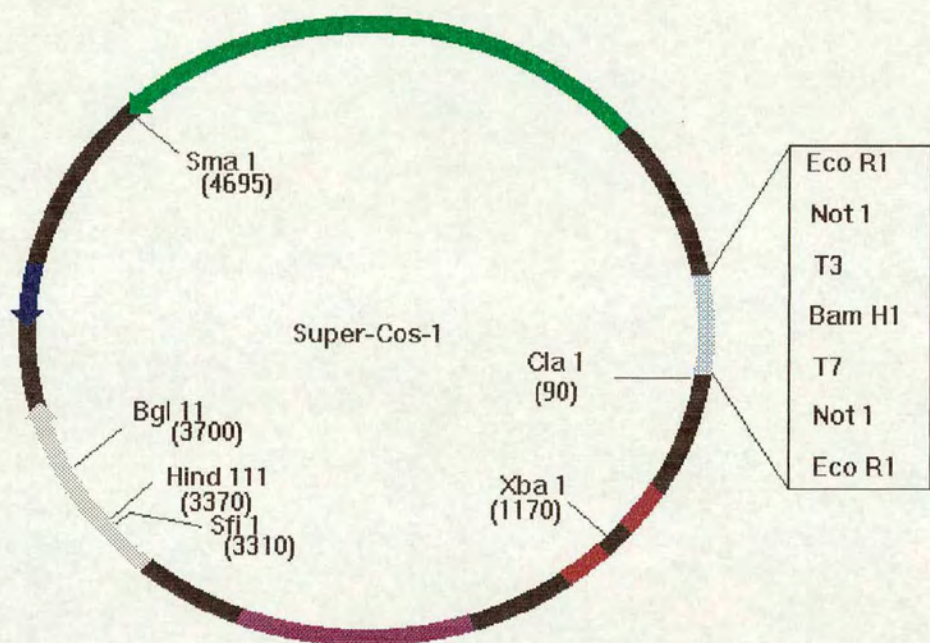
The DNA of all clones obtained on the secondary screening of the library was digested with a number of restriction enzymes. The restriction patterns of these clones were compared with the genomic map of the AATB locus (see chapter 3, Figure 3.13). Three of the clones appeared to be isolated as a result of false hybridisation to the probe. The restriction map of these clones displayed no relationship to the genomic

map of the AATB locus, and no positive signals were obtained from the Southern blots hybridised with AATB probes.

Two clones (5 and 10) gave a simple restriction pattern entirely consistent with being internal to the transgene array (Figures 4.4 - 4.9). All of the insert, in both of these clones, was of AATB origin. These clones had given the strongest signals on the primary screening and this was consistent with several copies of AATB being present in the insert. Two other clones (8 and 9) gave restriction fragments expected of the AATB transgene. These clones also gave restriction fragments consistent with the junction fragments of the genomic maps, and were thus likely to contain the mouse chromosomal DNA flanking the transgene array. The detailed restriction enzyme analysis of these four clones is presented in the following sections.

Figure 4.1 : Super-Cos-One Vector

The Super-Cos-One vector is 7937 bp in size (Evans et al., 1989). In packaging 1080 bp are lost between the cos sites resulting in a final vector size of 6857 bp, after infection of E. Coli NM554. The positions of some of the restriction sites, of importance in the mapping studies, have been given.



Multiple cloning site

Ampicillin resistance gene

Origin of replication

Neomycin resistance gene

SV40 promoter

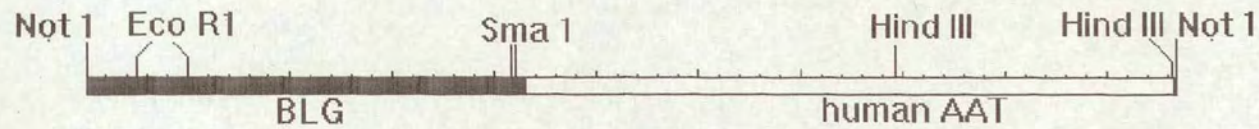
lambda cos sites

Figure 4.2 : AATB Probes Used For Screening The Library And

Characterisation Of The Clones

The figure shows the probes used for the screening of the library and the characterisation of the clones which are detailed in this Chapter. It shows a diagrammatic representation of the AATB transgene and the position of the enzymes used to generate the probes. Probe (a) is the entire AATB transgene which was removed from its vector sequences with restriction endonuclease Not I. Probe (b) was generated by digesting AATB with Not I and SmaI, it contains almost all of the ovine BLG sequences present in the AATB transgene, and was used to screen the cosmid library. Probe (c) is generated by digesting AATB with Eco RI and Not I. This double digest generates two small fragments of 615 bp and 555bp which co-migrate in agarose gel electrophoresis. Thus the probe consists of two fragments which together span more than 1 kb of the 5' end of AATB. Probe (c) was used to characterise the clones isolated from the library. Probe (d) is the 2.8 kb Hind III fragment from the 3' end of the AATB transgene. Probe (d) was also used to characterise the clones isolated from the library. All of the probes were purified by electrophoresis in 0.7 % agarose gels (see Chapter 2, Materials and methods for details of probe purification)

AATB

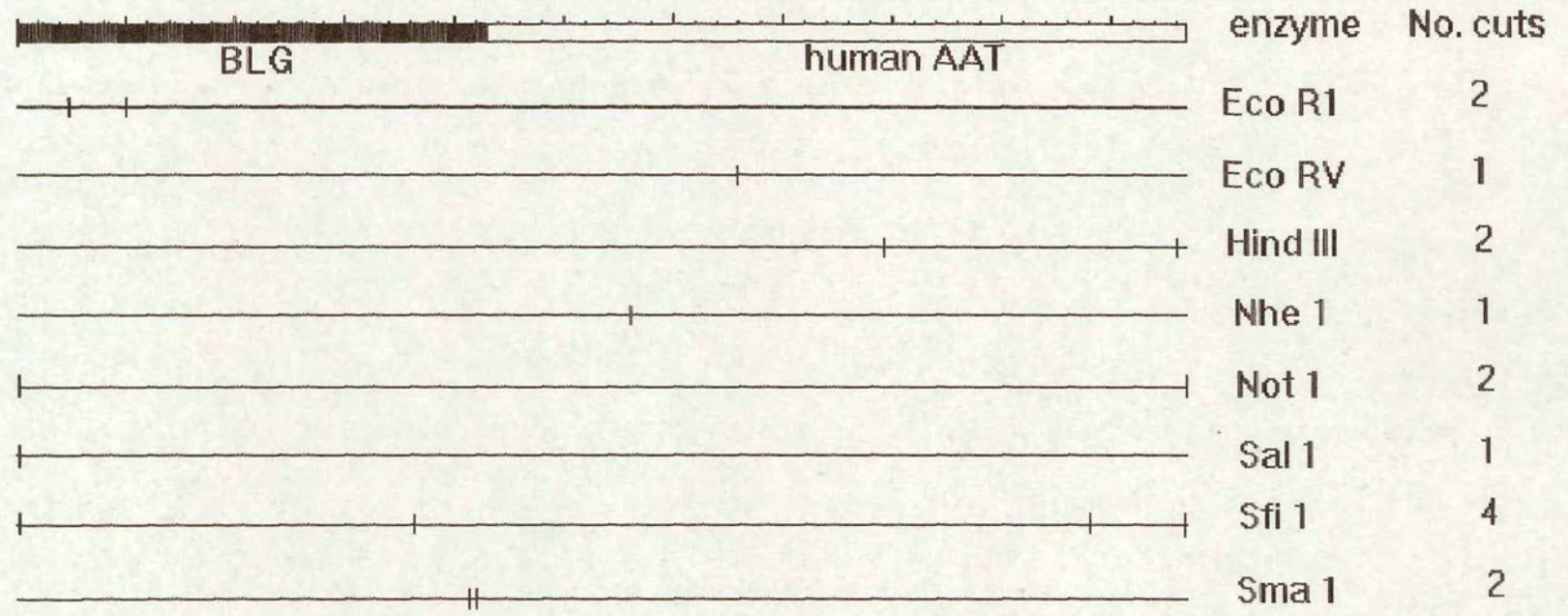


- (a) entire transgene (10.6 kb)
- (b) library probe (4.2 kb)
- (c) 5'probe (1 kb)
- (d) 3' probe (2.8 kb)

Figure 4.3 : Restriction Sites In The AATB Transgene

A number of restriction enzymes were used in the mapping of the AATB 46.2 genomic locus and in the characterisation of the positive clones obtained from the library screen. The figure shows the position, within the AATB transgene, of the restriction sites, of enzymes used in the mapping studies,. The number of recognition sites for each enzyme is given. Sma I cuts twice in AATB but the sites are only 58 base pairs apart and can thus, for the purpose of rough mapping of the cosmid clones be considered to cut at a unique site in AATB.

AATB



4.2 RESTRICTION ANALYSIS OF CLONES ISOLATED FROM THE LIBRARY

4.2 1 Restriction Analysis Of Clone 5

The restriction analysis of Clone 5 is detailed in Figures 4.4 and 4.5. Figure 4.4 shows the restriction fragments generated by digestion of clone 5 with various restriction enzymes. The gel shown in the figure is a 0.3 % agarose gel. The use of low percentage gels was essential to allow the resolution of the large fragments generated from clone 5 by a number of the enzymes used. Figure 4.5 shows an autoradiograph of the Southern blot of the gel, following hybridisation of the filter with the entire transgene probe (probe (a), figure 4.2). figure 4.6 is a diagrammatic representation of the restriction map of clone 5 deduced from the information in the agarose gel and Southern blot, in conjunction with the known restriction maps of the AATB transgene and the Super-Cos-One vector.

In all eight restriction enzymes were used for the restriction analysis of clone 5. Cla I, Bgl II, Not I, Eco RI, Sal I, Nhe I, Eco RV and Sma I. Figure 4.1 shows the position of the restriction sites of Cla I, Bgl II and Sma I in the cosmid vector. Figure 4.3 gives the position of the Not I, Eco RI, Sal I, Nhe I Eco RV and Sma I sites in the AATB transgene. In this section the information obtained from each digest is considered in turn. The Southern blot of the gel in figure 4.4 was only probed with the entire transgene probe. Hybridisation with this probe alone was sufficient to establish that the insert of Clone 5 harboured only AATB sequences.

Cla I was chosen for the analysis of the cosmid clones because it cuts at a unique site at position 90 in the cosmid vector (see figure 4.1). There was no Cla I

restriction site present in the AATB transgene (full sequence of AATB courtesy of Mike Dalrymple, PPL Therapeutics Ltd.). The enzyme generated only one large fragment on digestion of clone 5 (figure 4.4) demonstrating the presence of a large insert in the cosmid vector. The fragment hybridised with the AATB probe showing that the insert harboured AATB sequences.

Bgl II was chosen because it also cut at a unique site within the cosmid vector (see figure 4.1) and did not cut within AATB. This enzyme was a key enzyme in the mapping of the clones. It had a six base pair recognition site and would thus be expected to cut several times within a cosmid insert of greater than thirty-two kb. Digestion of clone 5 with Bgl II generated a single large fragment. The enzyme merely linearised the clone indicating that the large insert present in this clone had no Bgl II restriction site. As with the fragment generated by Cla I digestion this fragment hybridises with the AATB transgene probe (figure 4.5, Bgl II).

Digestion of clone 5 with Not I was more informative. The cosmid vector had two Not I sites situated 5' and 3' of the cloning site (see figure 4.1). Thus Not I was used to liberate the insert from the cosmid vector. This enzyme has an eight base pair recognition site that consists only of G and C bases. Cleavage by this enzyme is inhibited by methylation at cytosine bases. DNA isolated from eukaryotes is partially methylated and thus the enzyme cuts infrequently in eukaryotic, genomic DNA (average fragment size greater than 400 kb, (Brown. and Bird., 1986). Not I was the enzyme used to isolate the AATB transgene from the p Poly III-I vector prior to pro-nuclear injection of the transgene. Thus the injected AATB transgene had complementary, GC rich, four base pair, 5' overhang at either end. There is no internal Not I site in the AATB transgene. Digestion of clone 5 with this enzyme

generated four fragments (figure 4.4, Not I). One of the fragments was 6.8 kb, the expected size of the cosmid vector. This was the only one of the four fragments that did not hybridise with the AATB probe (figure 4.5 Not I) (it was assumed that this fragment was the cosmid vector but this has not been proven by hybridising the filter with a cosmid probe). The entire insert hybridises with the AATB probe indicating that the insert is of AATB origin. Not I also generated a 10.6 kb fragment. This was the size of the injected AATB transgene. The fragment hybridised with the AATB transgene probe and was assumed to be individual transgenes generated from the transgene array. The fragment is more intense than the other fragments in the lane. This suggests that there is more than one intact transgene within the insert of clone 5. The generation of this fragment demonstrated that the Not I sites have been regenerated in the transgene array. The other two fragments generated by Not I were 9.4 kb and 4.8 kb in size. These fragments are situated at the 5' and 3' end of the insert respectively (See figure 4.6, Not I and later section on Sma I digest) and were the result of Nde II digestion of AATB transgenes present in the array.

Digestion of Clone 5 with Eco RI was consistent with the Not I digestion of the clone. Like Not I, Eco RI liberated the insert from the cosmid vector (see figure 4.1). A 6.8 kb fragment was generated from clone 5 by digestion with this enzyme (figure 4.4, Eco RI). This fragment was the same size and intensity as the 6.8 kb fragment liberated by Not I digestion of the clone (compare Not I digest with Eco RI digest figure 4.4, fragments labelled (C)). This fragment did not hybridise with the AATB transgene probe (Figure 4.5, Eco RI) and was assumed to be the cosmid vector. Eco RI cuts twice within AATB. At position 611 and 1169 (see figure 4.3). Digestion of AATB with this enzyme liberated a 558 bp internal fragment. A head to tail array of

transgenes would have been expected to generate a 10055 bp fragment and the small internal fragment previously mentioned. Fragments of this size were visible on the gel shown in figure 4.4. Both of these fragments also hybridised with the AATB probe.(figure 4.5, Eco RI). One other fragment is generated from Clone 5 by this enzyme. The fragment is 3.6 kb and hybridises with the AATB probe. This size of this fragment was consistent with Eco RI digestion of the 4.8 kb truncated transgene present at the 3' end of the insert (see figure 4.6, Not I and Eco RI).

Enzymes Sal I, Nhe I and Eco RV all cut once within the AATB transgene (see figure 4.3) and not at all in the cosmid vector. Each of these enzymes generated only two fragments on digestion of clone 5. Each generated an AATB repeat length fragment of 10.6 kb that hybridised with the AATB transgene probe (figure 4.4 Sal I, Nhe I and Eco RV and figure 4.5 Sal I, Nhe I and Eco RV). Each of the enzymes would have been expected to generate such a fragment from a head to tail AATB array. The second fragment is 21.5 kb. This fragment (generated by each of these enzymes) incorporates the cosmid vector and the remainder of the insert following removal of the AATB repeat length fragments. The repeat length fragments and the 21.5 kb fragments generated by Sal I and Nhe I are of equal intensity (Eco RV has demonstrated star activity and was not suitable for this analysis). This suggested that the two fragments contained equal masses of DNA. Since the larger fragment was two times bigger than the repeat length fragment this indicated that two repeat length fragments must have been generated from clone 5 by digestion with Sal I and Nhe I. In conjunction with the information generated by the Not I digest this suggests that the insert of clone 5 harbours two intact copies of the AATB transgene and two truncated copies of 9.4 kb and 4.8 kb. This gives a total insert size of 35.5 kb. This was within

the range expected for a cosmid vector insert (minimum insert 32 kb, maximum insert 42.kb).. The vector could not in fact have accommodated three intact transgenes and the two truncated transgenes as this would have been outwith the maximum insert size. Similarly one intact transgene and the truncated transgenes would have been below the expected insert size for the cosmid vector

Sma I was the only enzyme used in the analysis of clone 5 that cut within the vector (see figure 4.1) and within AATB (see figure 4.3). Sma I cuts twice in AATB (at position 4115 and 4173) but the restriction sites are only 58 bp apart. For the purpose of the rough mapping of the cosmid clones this enzyme was considered to cut once within AATB. Digestion of clone 5 with other enzymes detailed in this section had demonstrated that the insert of clone 5 was an integral part of the AATB transgene array of line AATB 46.2 (see figure 3.13, Chapter 3). As Sma I cut within the insert and the vector digestion of the clone with this enzyme allowed the orientation of the cosmid and the insert to be ascertained. The enzyme generated 3 fragments. The AATB repeat length fragment and two fragments of 6.6 kb and 3.9 kb. The 3.9 kb fragment did not hybridise with the AATB transgene probe suggesting it was of cosmid origin. The other two fragments hybridised with the transgene probe indicating that they contained AATB sequences. The 6.6 kb fragment gave a weak signal, because only 2.9 kb of the fragment is of AATB origin and the remaining sequences are derived from cosmid (see figure 4.6).

The restriction map of clone 5 is shown in figure 4.6. Digestion with Not I demonstrated that the insert contained an integral part of the AATB 46.2 transgene array. The results of digestion with Eco RV, Nhe I and Sal I demonstrated the transgenes in the insert were oriented head to tail (as expected from the genomic

mapping, Chapter 3). These enzymes also confirmed the presence of two intact copies of AATB within the insert in addition to the two truncated copies identified by Not I digestion.

The insert size is 35.5 kb and incorporates 3 transgene /transgene junctions. The injected transgene possessed Not I complementary ends. The Not I site has been faithfully regenerated at all three transgene/transgene junctions present in the clone.

The aim of the cloning was to isolate the murine sequences flanking the AATB 46.2 transgene array. As it was clear from the preliminary mapping studies that clone 5 had only AATB sequences within the insert, no further analysis was performed on this clone.

Figure 4.4 : Ethidium Stained agarose gel of Clone 5 Restriction Digests

The figure shows a 0.3% agarose gel stained with ethidium bromide. M=raoul molecular weight markers. The clone was digested with ClaI, BglII, Not I, Eco RI, Sal I, Nhe I, Eco RV, and Sma I. Not I liberates the insert from the cosmid vector as does Eco RI. A 6.8 kb fragment (the expected size of the packaged cosmid vector) is generated by both of these enzymes. Not I, Sal I, Nhe I, Eco RV, and Sma I all generate a 10.6 kb fragment. This is the repeat length of the AATB transgene. The generation of this fragment with these enzymes indicates the presence of a head to tail array of transgenes within the insert.(see text for details). The generation of this fragment with Not I indicates that the Not I restriction site has been regenerated between all of the transgenes present in the insert of this clone. Cla I and Bgl II cut once in the cosmid vector and not at all in the AATB transgene. These enzymes linearise the cosmid, generating one single large fragment which co-migrates with the 48 kb molecular weight marker. Restriction of Clone 5 with Sma generates only three fragments, one is 10.6 kb (AATB repeat length). Sma I cuts within the vector (see Figure 4.1) and within the AATB transgene (see figure 4.3) thus restriction of the clone with this enzyme allows the orientation of the vector with respect to the insert to be ascertained. Fragments marked (C) are fragments which do not hybridise with the transgene probe (see figure 4.5) and are thus believed to be cosmid.

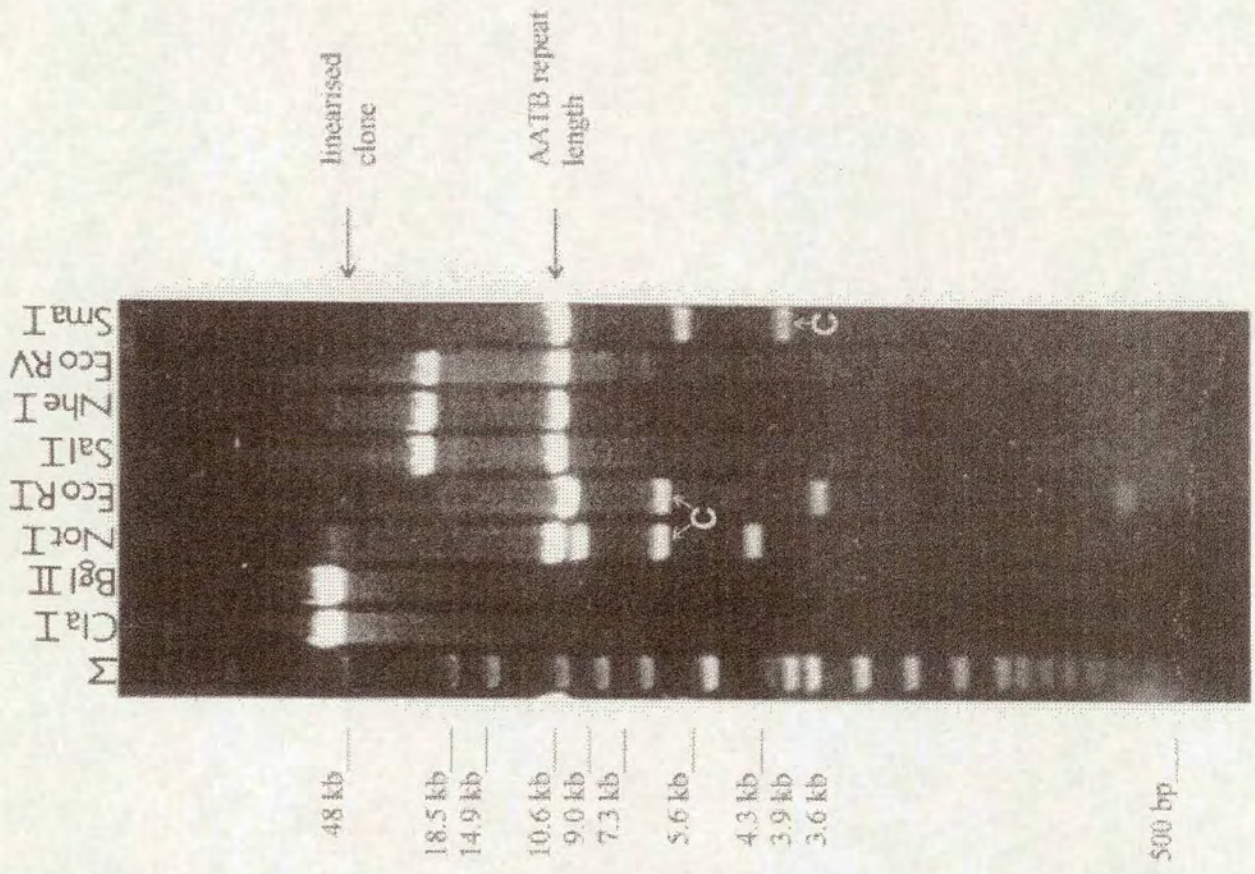


Figure 4.5 : Autoradiograph of Southern Blot Of Gel In Figure 4.4

The figure shows the autoradiograph of a Southern blot of the gel shown in figure 4.4. The blot was probed with the entire AATB transgene (probe (a), figure 4.2) and then exposed to X-ray film for 18 hours, at -70° C, with intensifying screens. The majority of fragments seen on the stained gel in figure 4.4 hybridise with the transgene probe. Only the 6.8 kb fragments generated by Not I and Eco RI and the 3.9 kb Sma I fragment do not hybridise (see fragments marked C in figure 4.4). In each case the size of the fragment is that predicted for the cosmid vector. All fragments generated from the insert of clone 5 hybridise with the AATB transgene, indicating that the entire insert is of AATB origin.

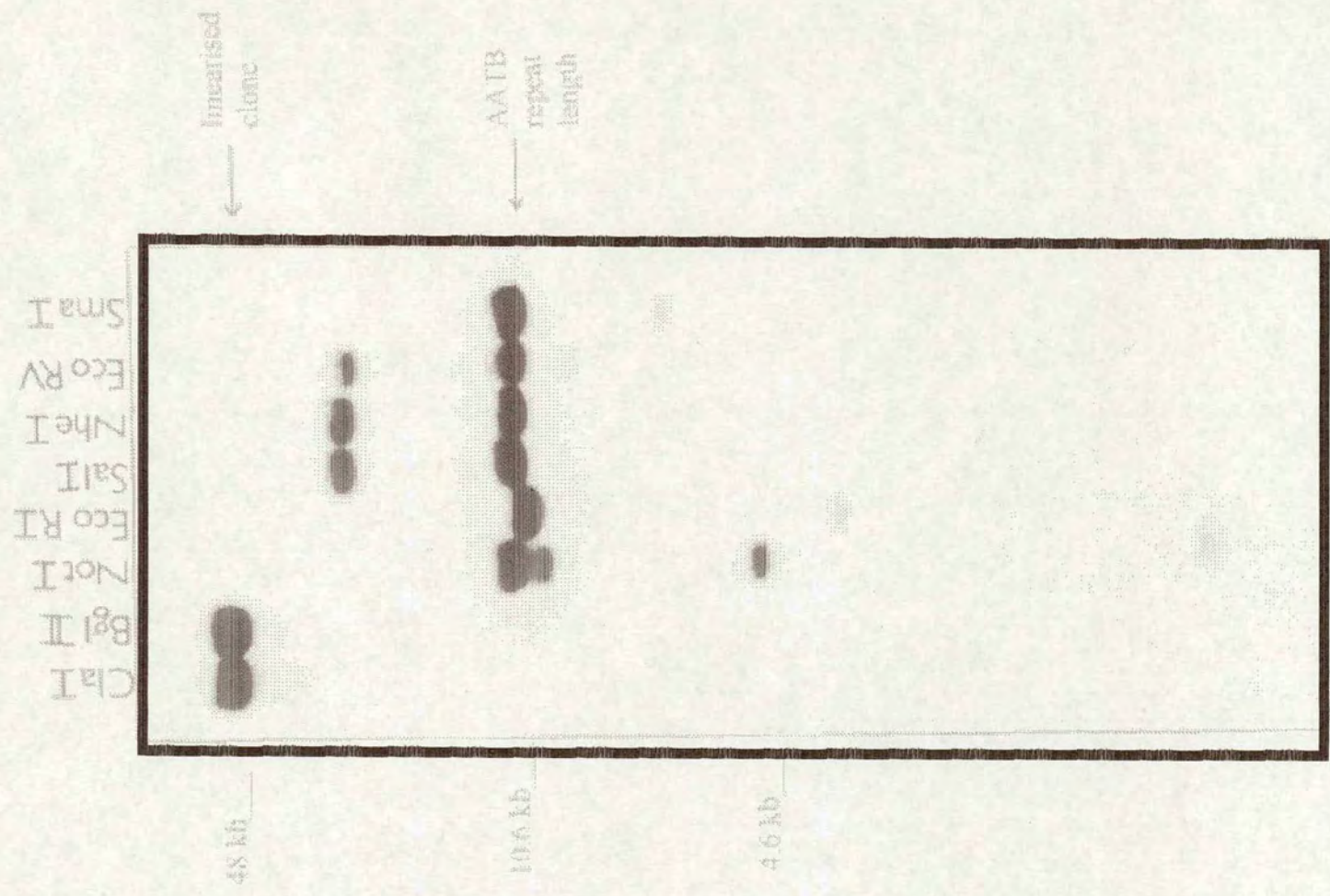
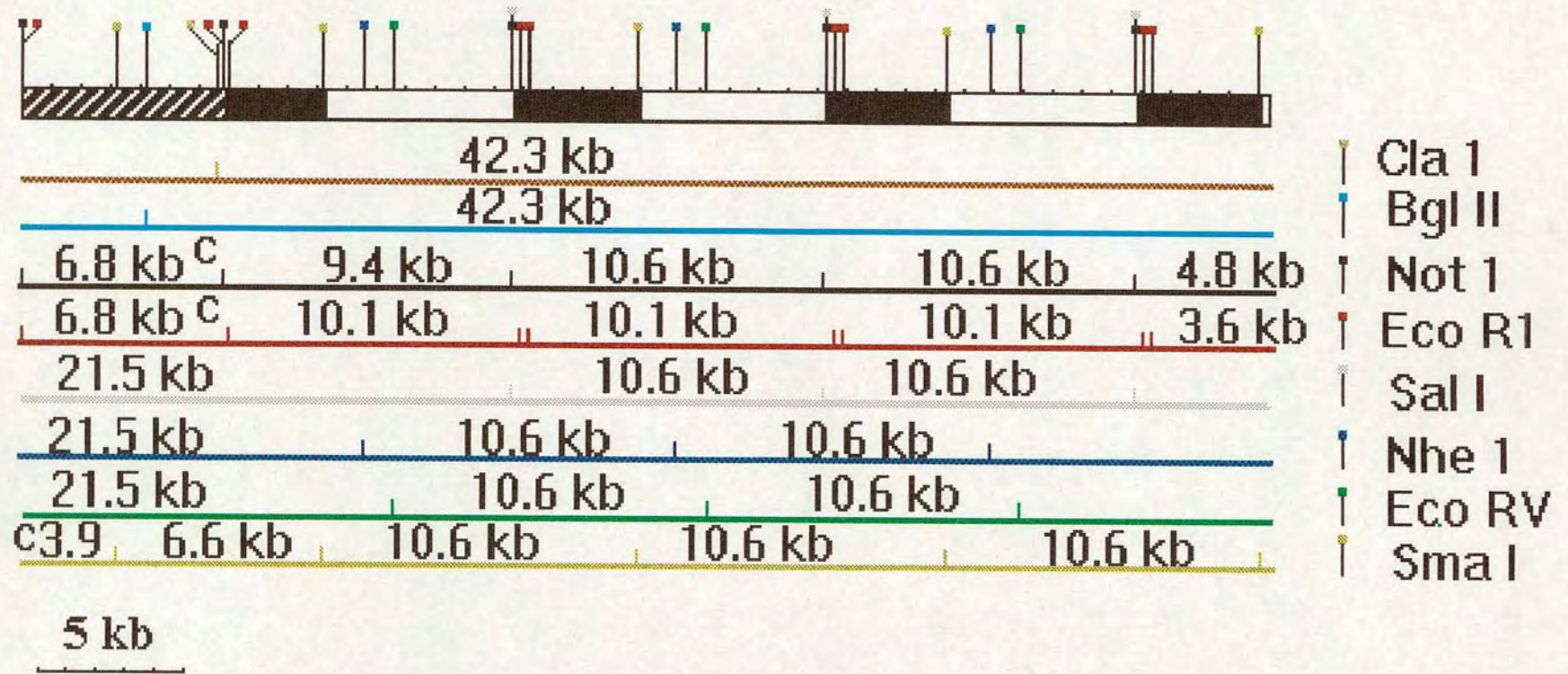


Figure 4.6 : Restriction Map Of Clone 5

The insert of clone 5 harbours part of the transgene array of AATB 46.2. The insert has two intact transgenes and two truncated transgenes of 9.4 and 4.8 kb (see Not I digest, figure 4.4 and 4.5). The Not I site has been regenerated at all three transgene/transgene junctions present in the insert. The 4.8 kb transgene contains the 5' end of AATB. This is ascertained by the generation of a 3.6 kb fragment on digestion of clone 5 with Eco RI (figure 4.4). All the transgenes in the insert are oriented head to tail (demonstrated by the generation of the AATB repeat length by enzymes Nhe I, Sal I and Eco RV, that cut only once in AATB). This orientation was expected from the results of the genomic mapping studies detailed in Chapter 3. Digestion with Sma I demonstrates that the 9.4 kb part transgene is at the 5' end of the insert (with respect to the 5' 3' orientation of AATB) and the 4.8 kb part transgene is present at the 3' end of the insert. The three fragments which do not hybridise with the AATB transgene probe (probe (a) figure 4.2) are labelled (c). These fragments are assumed to be cosmid.



Cosmid

 BLG sequences

 hAAT sequences

- ↑ Cla I
- ↑ Bgl II
- ↑ Not I
- ↑ Eco R1
- ↑ Sal I
- ↑ Nhe I
- ↑ Eco RV
- ↑ Sma I

4.2.2 Restriction Analysis Of Clone 10

Clone 10 was mapped using ten different restriction enzymes : Cla I, Bgl II, Not I, Sal I, Eco RV, Sfi I, Nhe I Eco RI, Hind III and Sma I. The positions at which these enzymes cut the cosmid vector and AATB are shown in figures 4.1 and 4.3 respectively. The fragments resulting from digestion of clone 10 with each of the enzymes listed were separated on an 0.3% agarose gel run in TAE buffer with 0.5 µg/ml ethidium bromide (shown in figure 4.7). The gel was Southern blotted and the filter hybridised sequentially with two AATB probes ; one from the 5' end of the AATB transgene (probe (c), figure 4.2) and one from the 3' end of the AATB transgene (probe (d), figure 4.2). The use of the two probes enabled more information to be extracted from the blots than is possible with the single AATB probe used to characterise clone 5 (see section 4.2.1).

Cla I cuts at position 90 in the cosmid vector (figure 4.1) and has no recognition site in AATB. Restriction of clone 10 with this enzyme gave a single large fragment that migrated with the 48 kb molecular weight marker (see figure 4.7, Cla I). The fragment hybridised with both 3' and 5' AATB probes.(figure 4.8 A, Cla I and figure 4.8 B, Cla I). This indicated the presence of a large insert that harboured AATB sequences.

Bgl II similarly linearised the clone generating a single large fragment (figure 4.7, Bgl II). This fragment also hybridised with both AATB probes. As with clone 5 the insert of clone 10 appeared to have no Bgl II restriction site. This enzyme would normally have been expected to cut several times in a cloned insert of this size. The

paucity of Bgl II restriction sites in the insert was entirely consistent with the insert harbouring only AATB sequences, as the AATB transgene contained no Bgl II site.

The Not I restriction digest generated four fragments (figure 4.7, Not I). The cosmid vector had two Not I sites; one either side of the cloning site (see figure 4.1). Thus this enzyme liberated the insert from the vector. As expected a 6.8 kb band was present in the Not I lane of figure 4.7 (fragment labelled (c)). This fragment did not hybridise with any of the AATB probes, and it was therefore presumed that this fragment was the vector. The Not I digest produced three fragments from the insert. Not I restriction sites would not be expected to occur at such a high frequency in genomic DNA. Restriction of clone 10 with this enzyme generated a 10.6 kb fragment which hybridised with both 5' and 3' AATB probes. 10.6 kb was the size of the injected AATB transgene. The generation of this fragment and the presence of so many Not I sites in the insert was suggestive of the insert harbouring an AATB array. The generation of single 10.6 kb transgenes from the cloned array with Not I demonstrated the faithful regeneration of the Not I site between individual AATB transgenes present in the array confirming the analysis of Clone 5 (see previous section).

The intensity of the 10.6 kb fragment was greater than that of the 6.8 kb cosmid fragment. The differential in the intensity of the two fragments was not commensurate with their difference in size. This suggested that more than one entire transgene was liberated from the cosmid insert. The other two fragments generated by Not I are relatively small and are only faintly visible on the gel (figure 4.7, Not I). These fragments are 2.8 kb and 3.2 kb in size. The 3.2 kb fragment hybridises with the 3' AATB probe (probe (d), figure 4.2) but does not appear to hybridise with the 5'

AATB probe (probe (c), figure 4.2). The 2.8 kb fragment appears not to hybridise with either probe. Information from other restriction digests of this clone however suggests that this fragment is the 5' end of a truncated AATB transgene (see next section on Sal I digest).

Sal I was chosen for the analysis because it cuts at a unique site in the AATB transgene (see figure 4.3). The restriction site was 25 bp from the 5' Not I site of the AATB transgene. There was no Sal I site in the cosmid vector. Digestion of clone 10 with this enzyme generated only two fragments; an AATB repeat length fragment and a 12.8 kb fragment. The repeat length fragment hybridises with both AATB probes which shows it was the AATB transgene. The fragment was as intense as the 10.6 kb fragment generated by Not I and more intense than the 12.8 kb fragment in the same lane (figure 4.7, Sal I). This suggested that the enzyme liberated more than one AATB transgene from the array. The 12.8 kb fragment was the remainder of clone 10 following the removal of the intact AATB transgenes. The Not I digest generated two small fragments of 2.8 kb and 3.2 kb. The size of the larger Sal I fragment was consistent with being the cosmid vector (6.8 kb) plus these two fragments. The 12.8 kb fragment hybridised with both the 5' and 3' AATB probes. The 3.2 kb Not I fragment hybridised with the 3' AATB probe so the hybridisation of the larger Sal I fragment with the 5' AATB probe suggested that the 2.8 kb Not I fragment was in fact the 5' end of an AATB transgene. The results of the Sal I digest in combination with those of the Not I digest revealed the entire insert was of AATB origin. Thus like clone 5, clone 10 harboured an insert that was an integral part of the AATB transgene array.

Like Sal I, restriction enzymes Eco RV and Nhe I cut at a unique site in AATB and not at all in the packaged cosmid vector. These enzymes also generated only two fragments; an AATB repeat length fragment and a larger fragment of greater than 20 kb. Both of the fragments hybridised with both AATB probes (figure 4.2 (c) and (d)). The importance of these enzymes in the mapping of the clones is that they demonstrated, by the generation of an AATB repeat length fragment, that the transgenes in the insert were orientated head-to-tail. Enzyme Eco RV appeared to exhibit some star activity (see faint bands in Eco RV lane, figure 4.7 and figure 4.8 B) but the two fragments generated by Nhe I appeared to contain equal masses of DNA. Since the larger fragment was approximately 23 kb (just over twice the size of the repeat length fragment) Nhe I must have liberated two AATB transgene fragments from the insert. Digestion of clone 10 with Not I showed the part transgenes present in the insert to be 3.2 and 2.8 kb. In the light of this Nhe I could only have generated two AATB fragments from the insert, if the insert contained three whole copies of AATB (see figure 4.9). Three intact transgenes and two truncated transgenes of 2.8 kb and 3.2 kb gave a total insert size of 37.8 kb. This was within the expected range for an insert in the Super-Cos One vector (32 kb - 42 kb).

Sfi I and Hind III cut more than once in the AATB transgene (see figure 4.3) giving internal AATB fragments of defined length. Both enzymes also cut once in the cosmid vector (Figure 4.1). Sfi I cuts AATB four times. One site is 7bp from the 5' Not I site, the second at position 3578, the third at position 9872 and the fourth at position 10602 only eight base pairs from the Not I site at the 3' end of the transgene (see figure 4.3). The enzyme thus generates AATB internal fragments of 6.2 kb and 3.6 kb and 800 bp. Four fragments were visible in the Sfi I lane of figure 4.6. Two of

the fragments were more intense than the other two (see fragments labelled (i) in Sfi I, figure 4.6). These intense fragments were the correct size to be the internal AATB fragments. This data is consistent with there being more than one AATB transgene within the clone 10 insert. The 800 bp fragment is not seen because fragments of less than 1.8 kb were run off the bottom of the gel. The other two fragments are 5.1 kb and 6.9 kb. The former fragment hybridised with the 5' AATB probe (probe (c) figure 4.2) whereas the latter hybridised with the 3' AATB probe (probe (d) figure 4.2). The size and hybridisation pattern of these fragments in conjunction with the data from the Not I digest allowed the orientation of the vector and insert to be established (see figure 4.8)

Hind III cuts twice within AATB (see figure 4.3). This enzyme generates one internal AATB fragment of 2.8 kb from a single AATB transgene, and generates an additional 7.8 kb internal AATB fragment from a head to tail AATB array. Fragments of 2.8 kb and 7.8 kb were generated by digestion of clone 10 with Hind III (see fragments labelled (i), Hind III, figure 4.7). The 7.8 kb fragment was more intense than other fragments in the lane; suggesting that the insert of clone 10 contained a head-to-tail array of several AATB transgenes. The 2.8 kb fragment was the smallest of the fragments liberated from clone 10 by Hind III. If the stoichiometry of the fragments in this lane had been one to one, this fragment should have been the least intense; however, this fragment was more intense than either the 5.2 kb fragment or the 4.5 kb fragment consistent with the fragment being an internal AATB fragment from a multicopy AATB array.

The 5.2 kb Hind III fragment hybridised with the 5' AATB probe. It was not confirmed that this fragment contained cosmid sequences but the size was consistent

with this being the 3' most truncated transgene (detected by Not I digestion) and 2.3 kb of cosmid vector incorporating the T7 priming site (see figure 4.9). The 4.5 kb fragment was assumed to be the remaining cosmid sequences (including the T3 priming site) from the Hind III site in the cosmid vector to the first Hind III site at the 5' end of the insert (see figure 4.9). This fragment does not hybridise with any of the AATB probes (figure 4.7 A and figure 4.7 B). This is consistent with the fragment comprising mainly cosmid sequences and only 0.4 kb of AATB sequences outwith those used as AATB probes used in this analysis (see figure 4.9, Hind III).

Sufficient information was obtained from the gel in figure 4.6 and the Southern blots shown in figure 4.7 to establish that the insert of clone 10 like that of clone 5 was an integral part of the transgene array of AATB 46.2. The insert of clone 10 was shown to contain no Bgl II site. This would be expected of clones containing part of the transgene array, since AATB does not contain a Bgl II site. Not I digest of the clone demonstrated that the insert contained more than one intact transgenes and that the Not I site had been regenerated at all four transgene/transgene junctions contained within this insert (see also analysis of clone 5, section 4.2.1). The Not I restriction digest also demonstrated that two part transgenes were present in the array. The results of the Sal I, Eco RV, Nhe I, Hind III and Sma I digests demonstrate that the transgene copies are arranged in a head-to-tail fashion; a result consistent with the genomic mapping studies of AATB 46.2 detailed in Chapter 3. The Nhe I digest in particular demonstrated that the insert of clone 10 must contain three whole copies of the transgene in addition to the two truncated transgenes. The fragments generated from digestion with Sfi I and Hind III (which cut within AATB and the cosmid) demonstrated that the truncated transgenes were present at either end of the insert

abutting the cosmid vector. The insert obviously contains no murine DNA and was not further investigated in this work.

Figure 4.7 : Ethidium Stained Agarose Gel Of Clone 10 Restriction digests

Clone 10 was digested with a number of restriction enzymes. The resultant fragments were separated by electrophoresis in a 0.3% agarose gel run in TAE buffer containing ethidium bromide at 0.5 µg/ml. M= raoul molecular weight markers (Appligene). Clone 10 was digested with ClaI, Bgl II, Not I, Sal I, Eco RV, Sfi I, Nhe I, Eco RI Hind III and Sma I. Cla I and Bgl II enzymes each cut at a single site within the cosmid vector, digestion with either of these enzymes generates one single linear fragment. This indicates that the large insert contains no recognition site for these enzymes. Bgl II would be expected to cut several times in an insert of this size. The lack of Bgl II sites in the insert is consistent with the insert being entirely of AATB origin. Not I, Sal I, Eco RV, Nhe I and Sma I all generate a 10.6 kb fragment. This is the AATB repeat length and the generation of this fragment with these enzymes is indicative of an head to tail AATB array being present within the insert. Not I and Eco RI both liberate the insert from the cosmid vector c= the 6.8 kb fragment liberated by these enzymes that is assumed to be the cosmid vector. Enzymes Sfi I and Hind III cut more than once within the AATB transgene (see figure 4.3), i= the internal AATB fragments generated by digestion of clone 10 with these enzymes. The intensity of the internal fragments is greater than all other fragments generated by these enzymes indicating the presence of several copies of AATB within the insert.

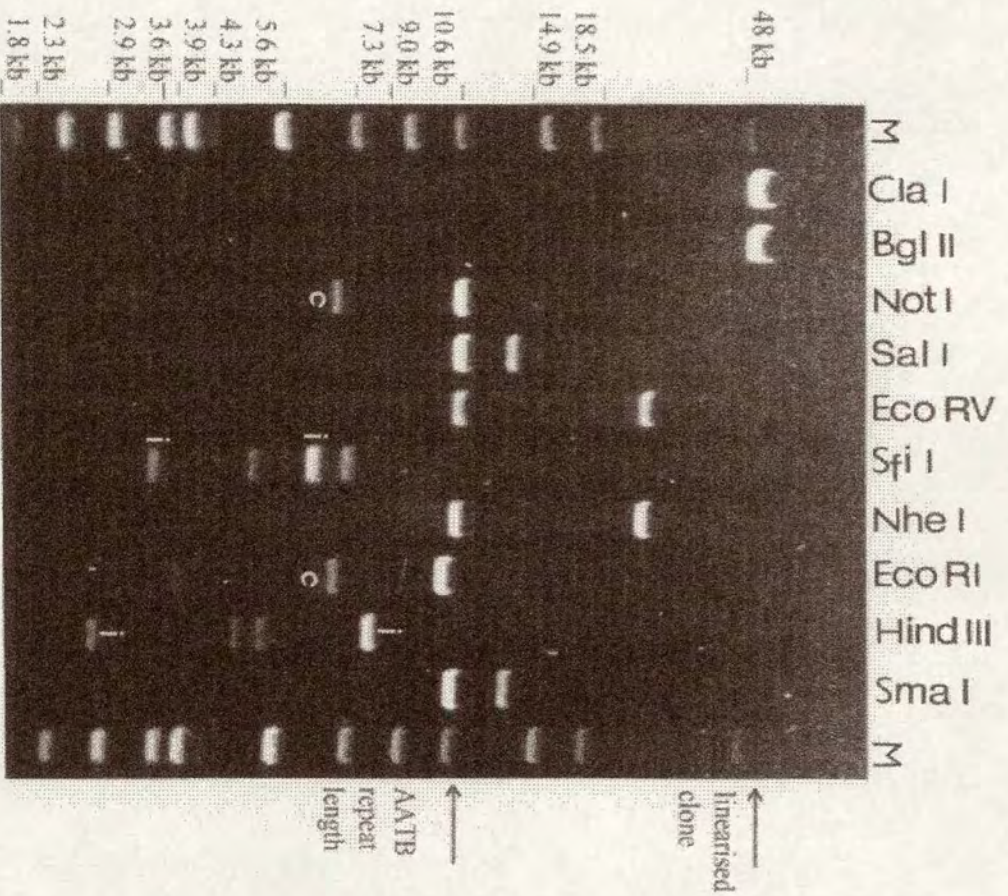


Figure 4.8: Autoradiograph Of Southern Blot Of Gel In Figure 4.7

A.

The gel in figure 4.7 was Southern blotted on hybond N membrane (Amersham) and hybridised with a probe from the 5' end of AATB (probe (c), figure 4.2). The filter was exposed to X-ray film, for two days, at room temperature, with intensifying screens. The linearised clones hybridise with this probe as do the AATB repeat length fragments. Only two of the Hind III fragments hybridised with this probe: the 7.8 kb internal fragment ((i) figure 4.7) and the 5.25 kb fragment. Only two of the Sfi I fragments also hybridised with this probe: the 6.3 kb internal fragment ((i) figure 4.7) and the 5.1 kb fragment. The filter was then stripped and laid down with X-ray film to ensure removal of the signal, before being reprobbed

B.

The gel in figure 4.6 was Southern blotted on hybond N membrane (Amersham) and hybridised with a probe from the 3' end of AATB (probe (d), figure 4.2). The filter was then exposed to X-ray film, for 1 hour, at -70°C , with intensifying screens. The linearised clones also hybridise with this probe as do the AATB repeat length fragments. Only one Hind III fragment hybridises with this probe, this was expected since the Hind III internal fragment is the probe used. Two Sfi I fragments hybridise with this probe: the 6.9 kb fragment and the 6.3 kb internal fragment ((i) figure 4.7).

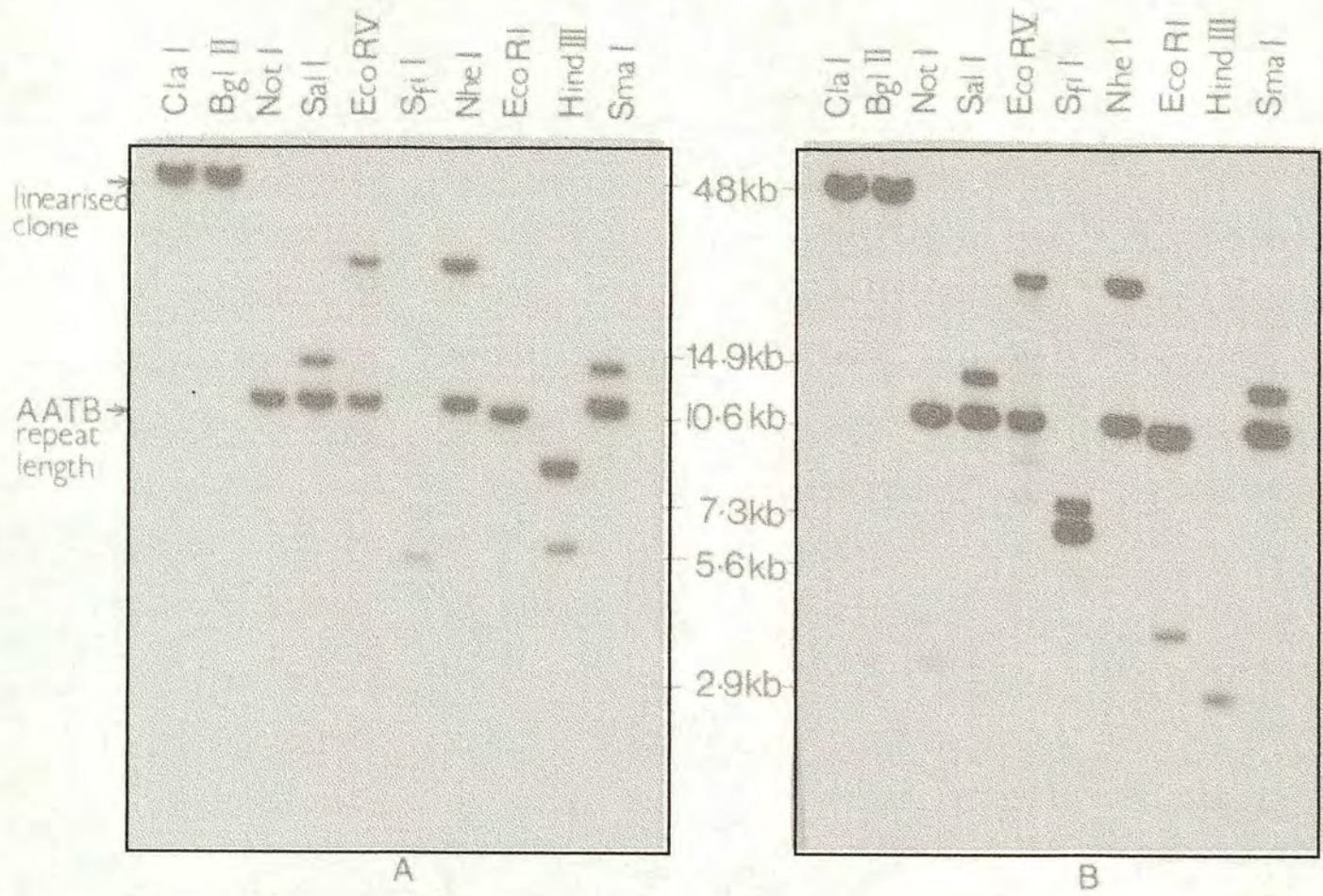
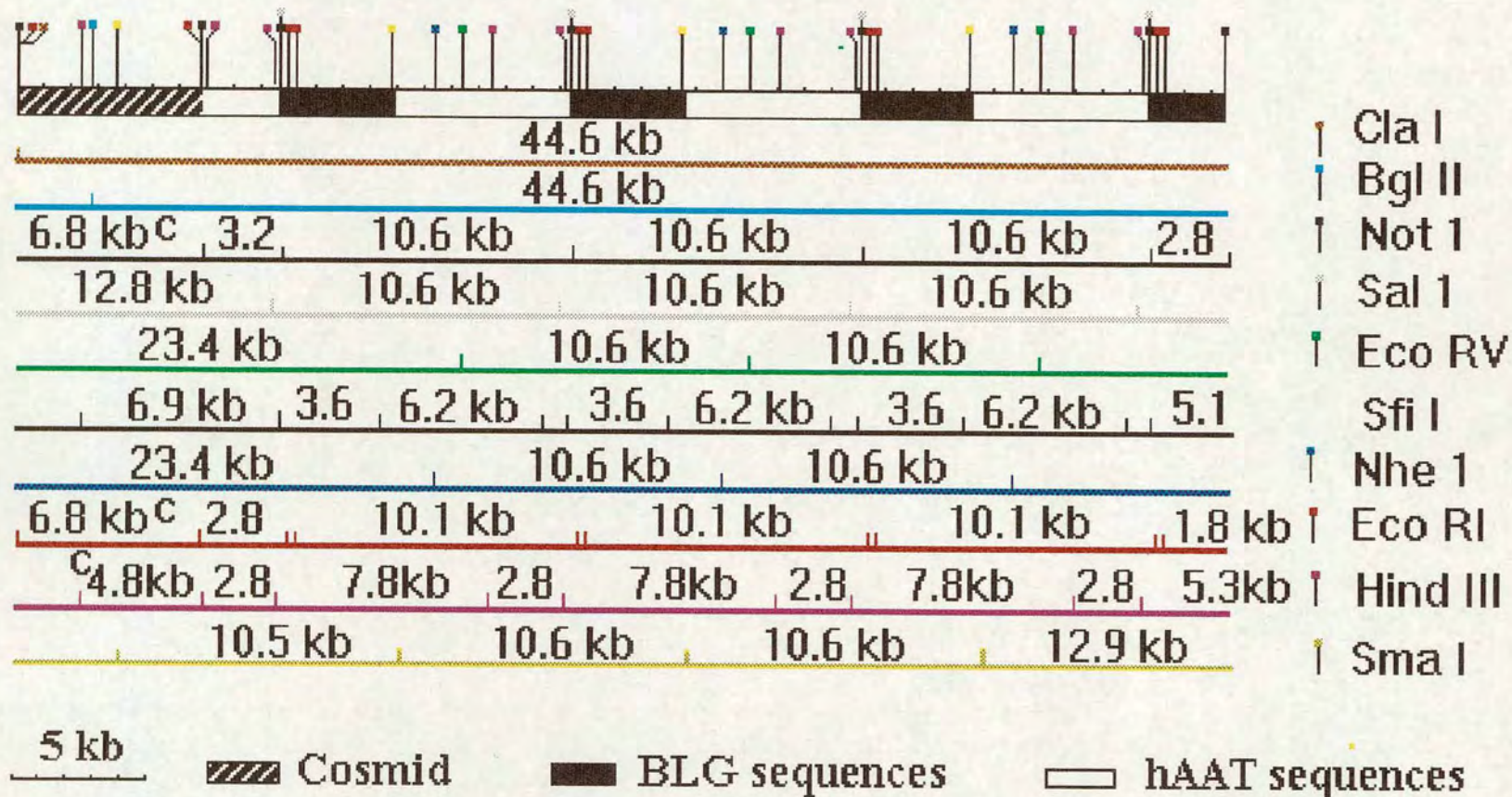


Figure 4.9 : Restriction Map Of Clone 10

The insert of clone 10 harbours part of the transgene array of AATB 46.2. The insert has three intact transgenes and two part transgenes of 3.2 kb and 2.8 kb (see Not I digest, figure 4.7). The Not I site has been regenerated at all four transgene/transgene junctions present in the insert. Digestion of the clone with Eco RV, Nhe I and Sal I liberated a 10.6 kb AATB repeat length fragment demonstrating the head to tail orientation of the transgenes in the insert. The 3.2 kb truncated transgene is at the 5' end of the insert (with respect to the 5' 3' orientation of AATB) and is the 3' end of an AATB transgene. The 2.8 kb fragment is located at the 3' end of the insert and is the 5' end of an AATB transgene. Thus at least 5 of the transgenes present in the array of AATB 46.2 are orientated in a head-to-tail manner (the copy number of AATB 46.2 is 7, see Chapter 3). Both Not I and Eco RI generate a 6.8 kb fragment that did not hybridise with AATB probes. This fragment is believed to be the cosmid (see fragments labelled c).



4.2.3 Restriction Analysis Of Clone 9

The restriction analysis of clone 9 was carried out in a manner similar to that described for clone 10 (section 4.2.2). The same ten restriction enzymes were used: Cla I, Bgl II, Not I, Sal I Eco RV, Sfi I, Nhe I, Eco RI, Hind III and Sma I. The analysis of clone 9 was performed more rigorously than that of clone 10, or clone 5, because it harboured DNA that was not of AATB origin. Four probes were used in the restriction characterisation of clone 9: The entire transgene (probe (a), figure 4.2), a probe from the 5' end of the AATB transgene (probe (c), figure 4.2), a probe from the 3' end of AATB (probe (d), figure 4.2), and the entire cosmid (figure 4.1). The analysis demonstrated that the insert of clone 9 spanned the chromosomal/transgene junction at the 5' end of the AATB 46.2 array. The insert of clone 9 contained 4.6 kb of murine DNA. The remaining insert was three AATB transgenes arranged head to tail (see figure 4.14).

Restriction of clone 9 with Cla I generated a single fragment. The fragment co-migrates with the 48 kb molecular weight marker and, hybridised with all four probes (a,c and d, figure 4.2, and cosmid). That result demonstrated that the clone contained a large insert which contained AATB sequences from both ends of the AATB transgene.

Restriction of Clone 9 with Bgl II generated two fragments (figure 4.10, Bgl II). Since one Bgl II site was known to be present in the cosmid vector, the second Bgl II site, required to generate the two fragments from the circular clone, must have been within the insert. AATB contains no Bgl II site therefore the insert contained DNA that was not of AATB origin. The larger of the two fragments (figure 4.10, Bgl II) hybridised with all AATB probes used in the analysis (figure 4.11 A and B: and

figure 4.13 A) and with the cosmid probe (figure 4.13 B). The 7.3 kb Bgl II fragment did not hybridise with any AATB probes including the entire transgene (Figures 4.11 A and B, and figure 4.13 A). It did, however, hybridise with the cosmid probe (figure 4.13 B). This fragment contains some of the cosmid vector sequences and some of the non hybridising DNA from the insert.

Digestion of clone 9 with Not I generated only three fragments (figure 4.10, Not I and figure 4.12). Not I sites were present at either side of the cloning site in the Super-Cos -One vector and enabled the separation of the insert from the vector (figure 4.1). One of the three fragments generated was 6.8 kb in size (the expected size of the cosmid vector) and hybridised with the cosmid probe (Figure 4.13 B). The other two fragments were generated from the insert. One was the same size as the injected AATB transgene (10.6 kb). The other fragment was larger, migrating a little slower than the 14.9 kb molecular weight marker. Both fragments hybridised with 5' and 3' AATB probes (figure 4.11 A and B) but not with cosmid (figure 4.13 B). The generation of the 10.6 kb fragment which was more intense relative to other fragments in the lane (figure 4.10, Not I) suggested the presence of more than one copy of the AATB transgene within the insert. As in the clones described earlier in this chapter the Not I site is regenerated at all transgene/transgene junctions cloned in the insert.

The 15.2 kb fragment contains the 5' most AATB transgene of the AATB 46.2 array and the murine DNA which harbours the Bgl II site. Double digestion of clone 9 with Not I and Bgl II reduces the 15.2 kb Not I fragment to 11.7 kb (compare Not I digest of clone 9 (N 9) with Not I/Bgl II digest of clone 9 (NB 9), figure 4.12): giving a fragment of 3.5 kb which did not hybridise with AATB or cosmid probes (see fragment marked with arrow in Not I/Bgl II digest of clone 9 (NB 9) in figure 4.12).

The double digest of clone 9 with Not I/Bgl II also liberates two cosmid hybridising fragments of 4.2 and 2.6 kb (figure 4.12, NB 9). The generation of these fragments was predicted from the known restriction sites of these enzymes in the packaged cosmid vector.

Restriction enzymes Eco RV and Nhe I cut only once in AATB (figure 4.3) and not at all in the packaged cosmid vector. Both of these enzymes generate the AATB repeat length fragment from clone 9. In each case the repeat length fragment is more intense than other bands, of similar size, in the same lane. This result demonstrates the transgenes present in the insert are arranged in a head to tail manner. The clones described earlier in this chapter (clone 5, section 4.2.1 and clone 10, section 4.2.2) contained only AATB sequences within their inserts, and generated only two fragments when digested with either Eco RV or Nhe I. In contrast clone 9 generated three fragments on digestion with either Eco RV or Nhe I (see figure 4.10). The smaller fragments produced by each enzyme (see fragments labelled (a) and (b) figure 4.10) did not hybridise with the 3' AATB probe but did hybridise with the 5' AATB probe. The size of these fragments (7.3 kb for Eco RV and 9 kb for Nhe I) was entirely compatible with that of the 5' hybridising junction fragments identified for these enzymes in the genomic mapping studies (see Chapter 3). This suggests the non hybridising DNA of the clone 9 insert was murine DNA flanking the AATB 46.2 transgene locus.

Hind III also produced a fragment of a size that correlated with the 5' Hind III junction fragment: as determined from the AATB 46.2 genomic mapping studies (see fragment labelled (c), figure 4.10). The 11.4 kb fragment hybridised with the 5' AATB probe but not with the 3' AATB probe (figure 4.11 A and B respectively).

Hind III also generated the predicted 7.8 kb and 2.8 kb internal fragments from the AATB array present in the insert. These fragments were more intense than the other fragments. This was consistent with the data from other digests that indicated the presence of more than one transgene in the insert. Neither of the remaining two fragments, generated by digestion of the clone with Hind III, hybridised with the 5' or 3' AATB probes, and were thus assumed to be cosmid. The size of these fragments was consistent with the orientation of the cosmid and vector deduced from the the Sfi I digest.

Digestion of Clone 9 with Sfi I also generated fragments whose size (6.2 kb, 3.6 kb and 800 bp), probe specificity, and hybridisation intensity was that expected for the internal AATB fragments generated by this enzyme. The other two fragments generated by Sfi I digestion (3 kb and 9 kb) did not hybridise with the AATB probes. The 9 kb fragment contained both cosmid and murine sequences from the Sfi I site in the cosmid vector to beyond the T3 primer site, and the entire 4.6 kb of murine DNA present in the insert. AATB has an Sfi I site at both ends of the transgene (figure 4.3). The generation of this fragment from this clone indicated that the transgene abutting the junction at the 5' end of the AATB 46.2 locus was either intact or very nearly intact, since the 5'Sfi I site of AATB includes the last four bases of the Not I site. The nature of the transgene end was demonstrated later by sequencing across the transgene chromosomal junction of the cloned insert (see Chapter 5). The generation of the 3kb fragment indicated that the 3' most Sfi I site of AATB (see figure 4.3) was not present in the 3' most transgene of the insert of clone 9.

Digestion of clone 9 with Sal I generated only two fragments. As with the other enzymes which cut once in AATB, Sal I generated an AATB repeat length

fragment. The second fragment was greater than 20 kb. Both fragments hybridise with the 5' and 3' AATB probes. The intensity of the two Sal I fragments on the agarose gel appears to be the same (figure 4.10). This indicates the fragments are present in equal mass. Since the larger fragment is twice the size of the repeat length fragment, Sal I must liberate two transgenes from the insert. The intensity of the larger band on the autoradiographs of the Southern blot, probed with either 5' or 3' AATB probes, was approximately half that of the repeat length band (see figure 14.11 A and B). This agrees with the larger band having the same amount of DNA but only half the amount of hybridising sequences as the repeat length band. The intensity of the repeat length fragment both on the agarose gel and on the autoradiographs was the same as the repeat length fragments generated by Not I, Eco RV, and Nhe I, suggesting all of these enzymes liberate two transgene copies from the insert. Therefore the insert harboured three transgenes and the murine flanking DNA. The total insert size was 36.4 kb which is within the expected range for a Super-Cos -One insert (32-42 kb).

In summary Clone 9 appears to contain the 5' chromosomal/transgene junction from the AATB 46.2 locus. Bgl II (which does not cut AATB) cuts within the insert and the cosmid generating a 7.3 kb fragment which does not hybridise with AATB probes. A double digest of the clone with Bgl II and Not I liberates a 3.5 kb fragment that does not hybridise with either AATB or cosmid probes. Thus the insert of the clone harbours DNA of non AATB origin.

The generation of fragments with Eco RV, Nhe I, and Hind III which are similar in size to the 5' hybridising junction fragments, determined for these enzymes

in the genomic mapping studies (Chapter 3), suggests that the non hybridising DNA is the murine DNA flanking the AATB 46.2 locus.

The intensity of the AATB repeat length fragment generated by Sal I is similar to that of the larger fragment generated by this enzyme. The differential in size of fragments suggests that the repeat length fragment must contain the equivalent of two copies of the transgene to be present in equal mass. In conjunction with similar information determined from other restriction digests of the clone this shows the presence of three transgenes in the insert, in addition to the 4.6 kb of murine DNA. Not I restriction of the insert demonstrated that the Not I restriction site has been regenerated at both transgene/transgene junctions present in the insert of clone 9. The results of the Sfi I and Sal I restriction digests also suggest that the 5' most transgene at the chromosomal transgene junction has sustained little or no damage prior to integration into the mouse genome.

Figure 4.10 : Ethidium Stained Agarose Gel Of Clone 9 Restriction Digests

The gel is a 0.3 % agarose gel stained with ethidium bromide at 0.5 $\mu\text{g/ml}$. The enzymes used for each digestion are given above each lane. M= raoul molecular weight marker (Appligene) The size of the markers is given in kb on the left. L= the linearised clone. R= the AATB repeat length fragment generated by enzymes Not I, Sal I, Eco RV, Nhe I and Sma I. a= the 5' Eco RV junction fragment, b= the 5' Nhe I junction fragment and c= the 5' Hind III junction fragment. The intensity of the repeat length fragment is greater than that of other fragments (Sal I lane excepted: see text for details) indicating the presence of more than I AATB transgene in the insert. The generation of fragments a, b and c by Eco RV Nhe I and Hind III respectively: that are the expected size of the 5' junction fragments (as determined in the genomic mapping studies), suggested this clone harboured the 5 chromosomal/transgene junction of AATB 46.2.

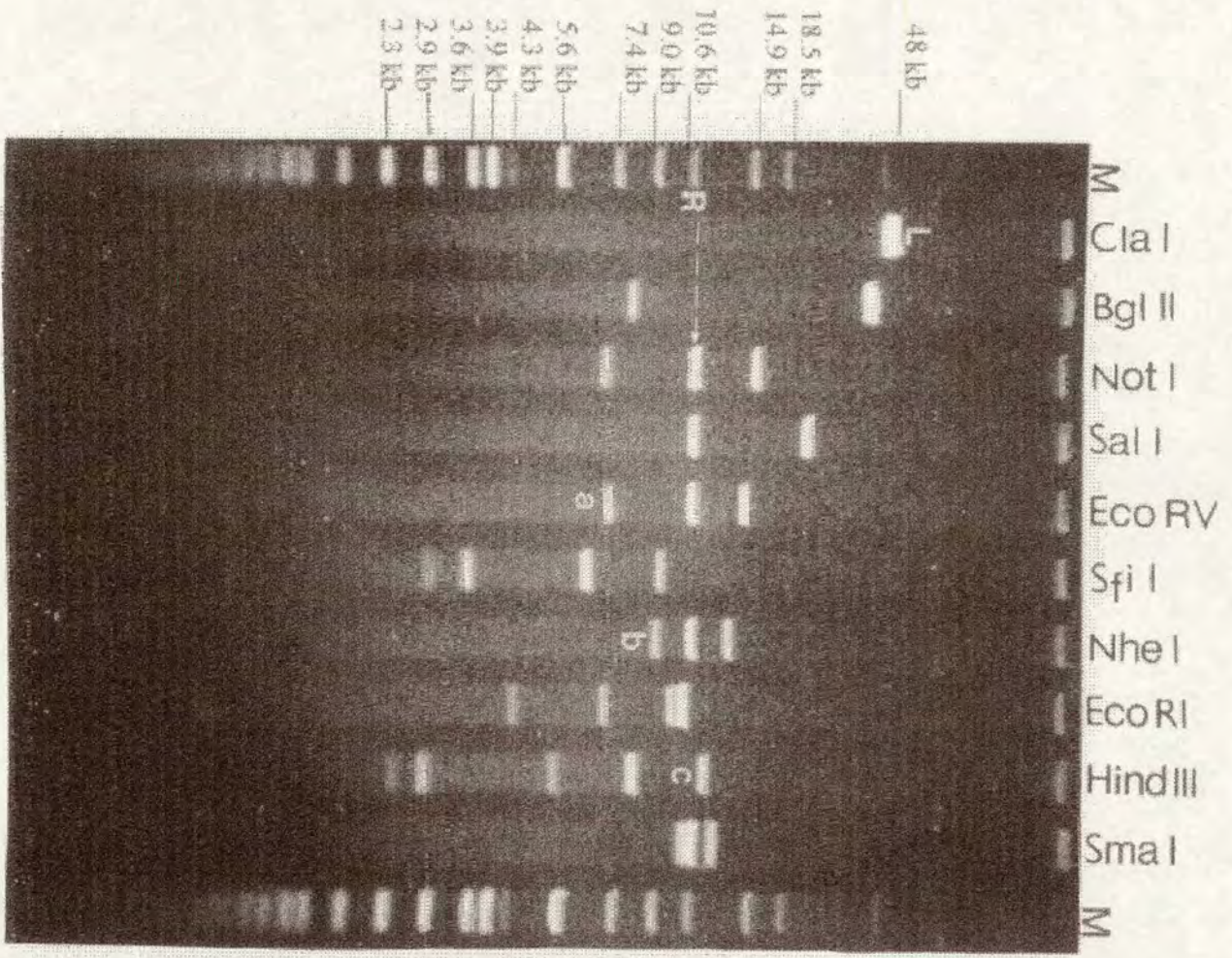


Figure 4.11 : Autoradiographs of Southern Blot of Gel in Figure 4.10

A.

The agarose gel in figure 4.10 was Southern blotted onto Hybond N nylon membrane (Amersham). The filter was then hybridised with a 5' AATB probe (probe (c), figure 4.2). The autoradiograph was produced by exposing the filter to X-ray film, for two hours, at -70°C , with intensifying screens. The 7 kb Bgl II fragment does not hybridise with this probe. The linearised clone (Cla I) does as did the repeat length fragments generated by Not I, Sal I, Eco RV, Nhe I and Sma I. The suspected junction fragments a, b and c all hybridise with this probe: consistent with them being the 5' junction fragments of AATB 46.2. The filter was then stripped and exposed to X-ray film to ensure removal of the signal, before being reprobated.

B.

This is an autoradiograph of the filter in figure 4.11 A. The filter was in this instance hybridised with the 3' AATB probe (probe (d), figure 4.2). The filter was exposed to X-ray film, for 1.5 hours, at room temperature, with intensifying screens. The linearised clone (Cla I) also hybridised with this probe. This suggested the presence of entire AATB transgenes in the insert. The repeat length fragments generated by Not I, Sal I, Eco RV, Nhe I and Sma I also hybridised with this probe in addition to the 5' probe (figure 4.11 A). This demonstrated that the 10.6 kb fragment present in many of the lanes was the AATB transgene. The suspected 5' junction fragments a, b and c (figure 4.10) did not hybridise with this probe.

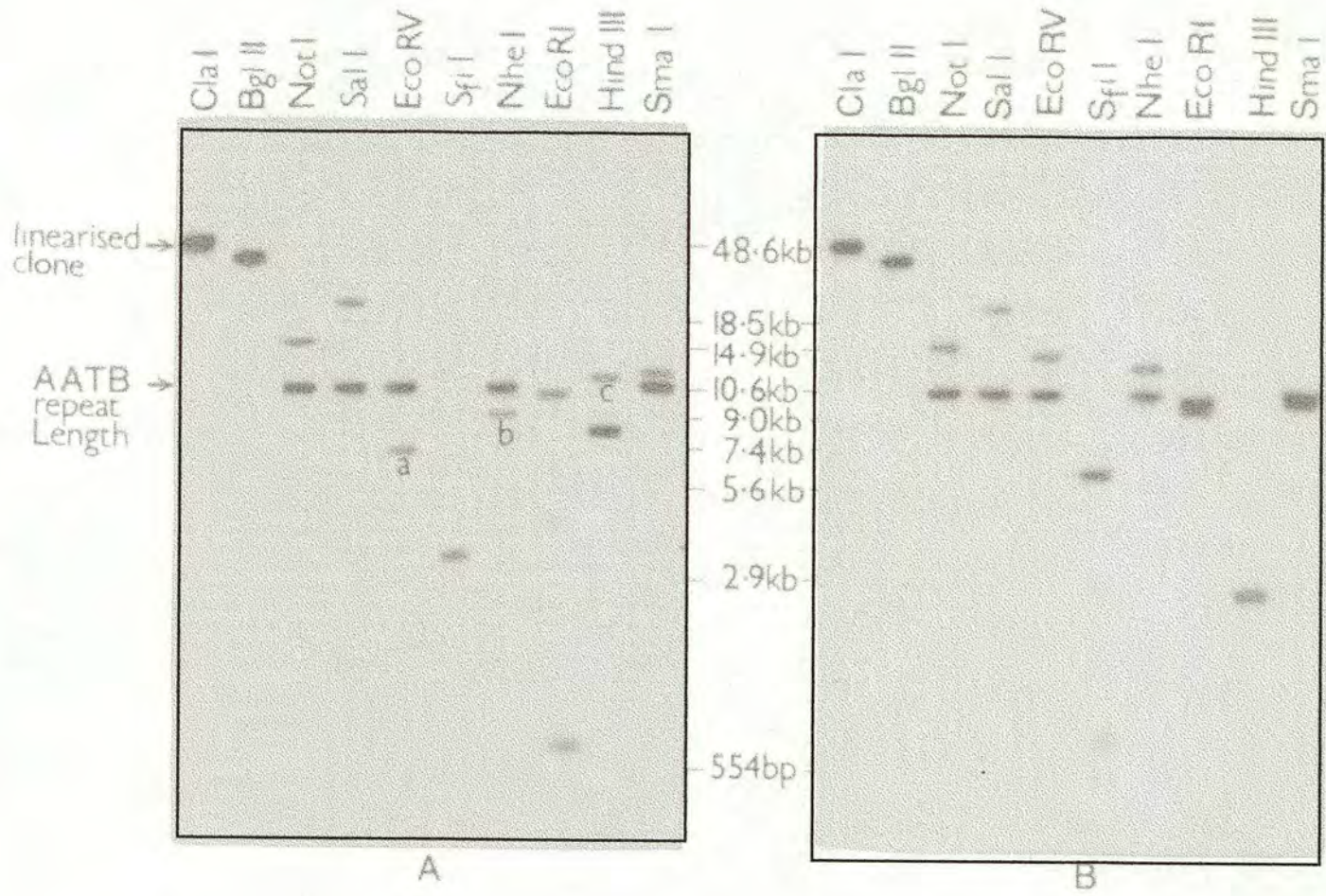


Figure 4.12 : Ethidium Stained Agarose Gel of Double Digests of Clones 8 and 9

The figure shows a 0.3% agarose gel stained with ethidium bromide 0.5 µg/ml. M= raoul molecular weight markers, (sizes given in kb on the right). The gel shows similar digests of clone 8 (8) and clone 9 (9) run in adjacent lanes. The clones were digested with Not I, (N); Bgl II, (B); and Eco RV, (E) and doubly digested with Not I/Bgl II, (NB); Not I/Eco RV, (NE); and Bgl II/Eco RV. This gel was Southern blotted and hybridised with transgene and cosmid probes. The results of the hybridisations are shown in figure 4.13. Fragments which did not hybridise with either of the probes are indicated with an arrow. These fragments were believed to be murine in origin and be the flanking DNA from the AATB 46.2 locus.

N B E NB NE BE
89 89 89 89 89 89 M

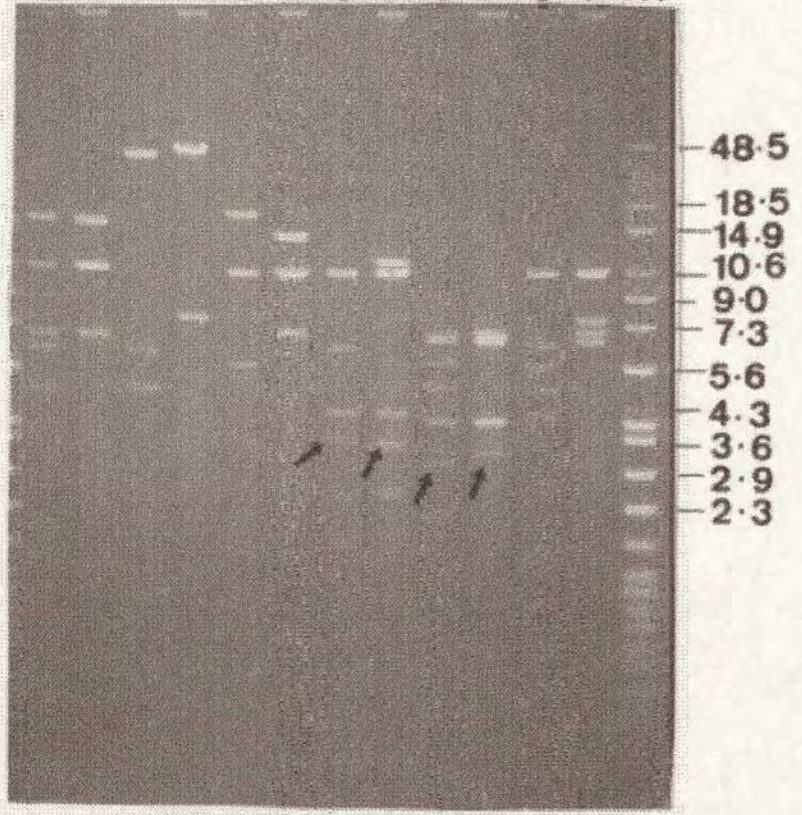


Figure 4.13 : Autoradiographs of Southern Blot of Agarose Gel in Figure 4.12

A.

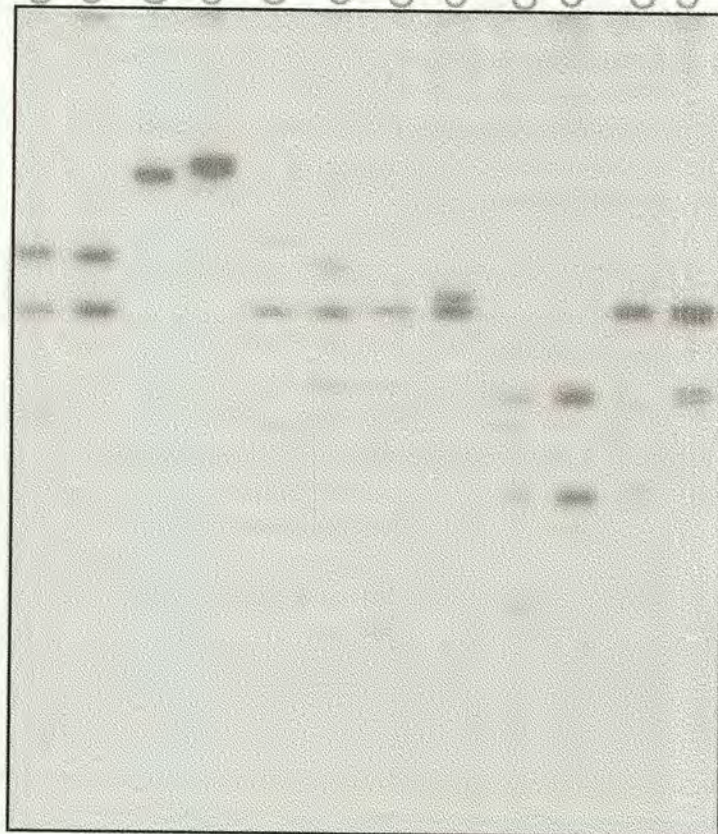
Is the autoradiograph of a Southern blot of the gel shown in figure 4.12: probed with the entire AATB transgene (probe (a), figure 4.2). The numbers between the two autoradiographs are the size of the raoul molecular weight markers in kb (see figure 4.12). Clone 8 (8) and clone 9 (9) were similarly digested and run in adjacent lanes. N = Not I, B = Bgl II, E = Eco RV, NB = Not I/Bgl II double digest, NE = Not I/Eco RV double digest and BE = Bgl II/Eco RV double digest. The filter was stripped and exposed to X-ray film to ensure removal of the signal, before being reprobated

B.

Is the autoradiograph of a Southern blot of the gel shown in figure 4.12: probed with all of the cosmid sequences. The sizes of the raoul molecular weight markers are shown on the left (see figure 4.12). Clone 8 (8) and clone 9 (9) were similarly digested and run in adjacent lanes. N = Not I, B = Bgl II, E = Eco RV, NB = Not I/Bgl II double digest, NE = Not I/Eco RV double digest, BE = Bgl II/Eco RV double digest.

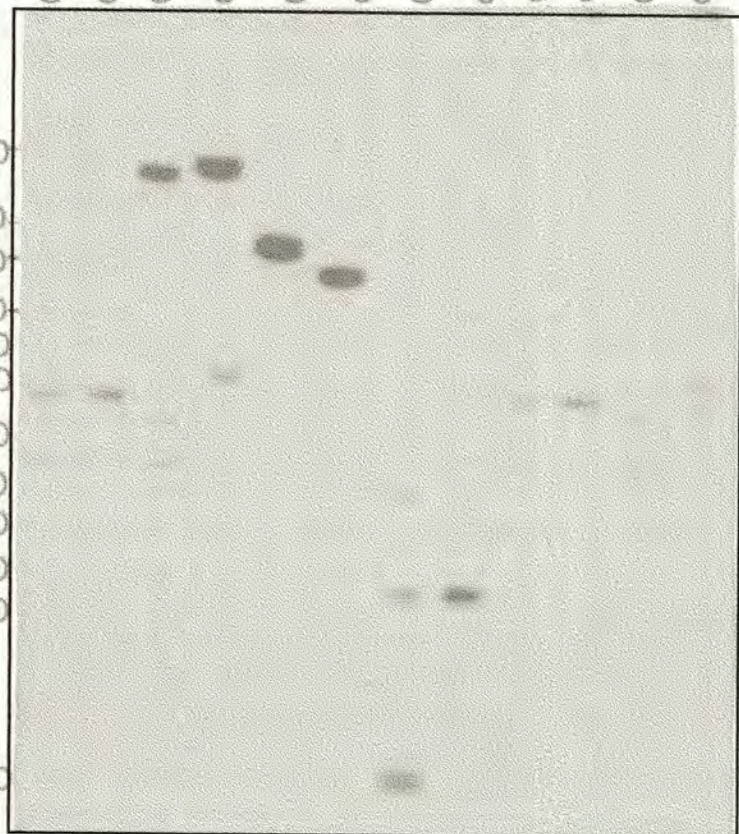
Comparison of Clone 8 and Clone 9 showed important differences between the clones. Digestion of clone 9 with Not I generated only one single 6.8 kb fragment that hybridised with cosmid sequences, whereas, clone 8 generated two fragments that hybridised with the cosmid clone: one of 6.8 kb and another of 4.8 kb. Not I would be expected to generate only one 6.8 kb fragment that hybridises with cosmid as this enzyme liberates the insert from the vector. All of the fragments generated by Bgl II hybridise cosmid only the large fragment hybridises with AATB. Not I/Bgl II double digest of clone 8 generates 5 fragments visible on the gel (figure 4.12). Two hybridise with cosmid and another two hybridise with AATB. The remaining fragment does not hybridise with either probe indicating that it is murine in origin. An additional fragment of approximately 750 bp hybridised with the cosmid probe but not with AATB. Not I/Bgl II digestion of Clone 9 generated five fragments. two hybridised with AATB and two others hybridised with cosmid. The remaining fragment did not hybridise with either probe and was believed to be murine in origin.

N B E NB NE BE
 8 9 8 9 8 9 8 9 8 9 8 9



A

N B E NB NE BE
 8 9 8 9 8 9 8 9 8 9 8 9

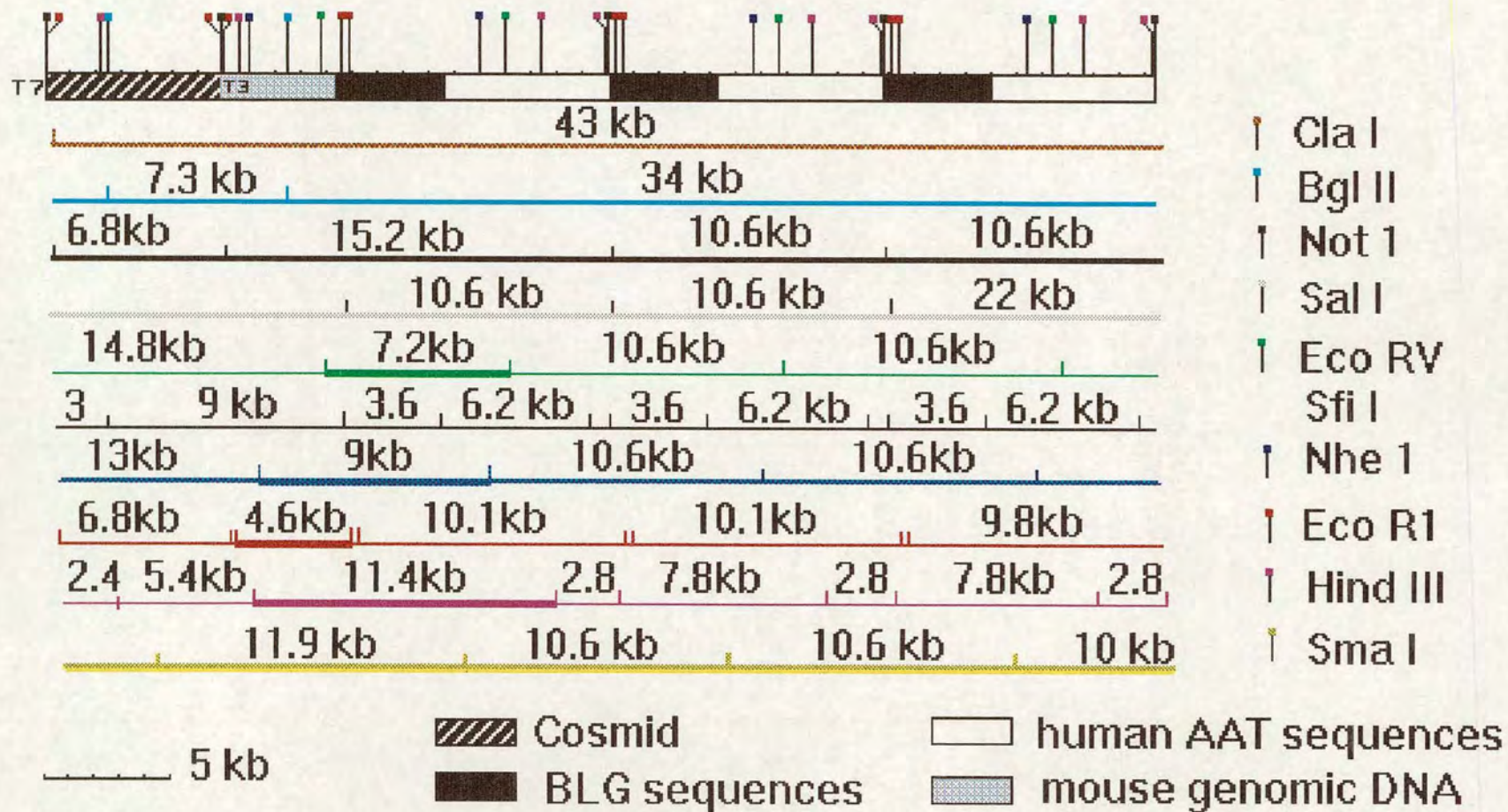


B

48.5kb
 18.5kb
 14.9kb
 10.6kb
 9.0kb
 7.3kb
 5.6kb
 4.3kb
 3.6kb
 2.9kb
 2.3kb
 754bp

Figure 4.14 : Restriction Map of Clone 9

The clone contains part of the AATB transgene array of AATB 46.2 and 4.6 kb of murine flanking DNA. Double digests with Not I/ Bgl II and Not I/ Eco RV generated fragments that did not hybridise with either AATB of cosmid suggesting that they were murine in origin. Fragments generated by Eco RV, Nhe I and Hind III are the correct size to be the 5' hybridising junction fragments detected in the genomic mapping studies (see Chapter 3 and figure 4.22). These fragments hybridise only with the 5' AATB probe and are represented in the figure by thick lines. The insert spans three transgene/transgene junctions and the Not I restriction site is regenerated at all three. Digestion of the clone with Eco RV, Nhe I and Sal I generated a 10.6 kb AATB repeat length fragment, demonstrating that the transgenes in the array are arranged head-to-tail.



4.2.4 Restriction Analysis Of Clone 8

The restriction analysis of clone 8 was carried out as for clone 9. Ten enzymes were used in the restriction mapping: Cla I, Bgl II, Not I, Sal I, Eco RV, Sfi I, Nhe I, Eco RI, Hind III and Sma I. A total of four probes were used in the analysis: three AATB probes (a, c and d, figure 4.2) and the cosmid probe (the entire cosmid sequences, figure 4.1). The insert of clone 8 was found to contain sequences that did not hybridise with the cosmid or AATB probes. The generation of these fragments suggested the presence of murine genomic DNA in the insert. Moreover, Eco RV and Nhe I generated fragments that were the expected size of the 3' AATB 46.2 junction fragments (identified in the genomic mapping studies) that hybridised only with 3' AATB sequences. Thus clone 8 appeared to harbour the 3' AATB 46.2 chromosomal/transgene junction.

The initial plasmid preparation of clone 8 gave low yields of DNA relative to all other clones from which DNA was prepared (see table 4.1): suggesting there was some problem with this clone. This preparation of the clone 8 showed very faint bands of 4.8 kb and 2.4 kb in samples of uncut cosmid run on an 0.3% agarose gel (data not shown) and in Cla I digested clone 8 (figure 4.15). However later plasmid preparations contained large amounts of the small contaminating entity (figure 4.16, all lanes). The contaminant would appear to be supercoiled as two bands are visible in uncut cosmid samples and upon digestion with Cla I, Sal I, Eco RV and Nhe I (Sal I, Eco RV and Nhe I do not cut within the cosmid vector). However upon digestion with Bgl II, Not I, Eco RI and Sma I a single 4.8 kb band is generated. All of the aforementioned enzymes that apparently linearise this contaminant cut within the cosmid vector.

Analysis of clone 8 demonstrated the presence of an additional part-cosmid of 4.8 kb that was an integral part of the clone. The contaminant present in the DNA preparation of clone 8 was also 4.8 kb. This was suggestive of recombination occurring within the clone. The 'recombining out' of the 4.8 kb fragment from clone 8 would have been expected to give two products: (a) a recombinant that contains the AATB sequences, murine DNA and only a 6.85 kb cosmid vector and; (b) a small part cosmid vector of 4.8 kb. A fragment that was likely to be the larger of the two products of recombination (clone 8 minus the 4.8 kb cosmid) was visible on agarose gels following electrophoresis of *Cla* I digested DNA (see figure 4.16). The intensity of this fragments in the ethidium stained gel was very weak suggesting only very low levels of this recombinant were present.

It would appear that the plasmid preparation contained three separate cosmids, two of which had arisen as a result of recombination of the original clone . The original clone contains the AATB sequences, murine sequences and the two cosmids (one complete 6.8 kb and one truncated 4.8 kb). The low concentration of the large recombinant would suggest that recombination occurs at a low frequency. The large excess of the smaller recombinant is probably due to its replicative advantage over the other two clones since it is approximately one tenth of their size.

The remainder of this section deals with the restriction mapping of clone 8. All mapping studies outlined in this section were performed on the first DNA preparation of clone 8 that contained only minimal amounts of the recombined molecules.

Digestion of clone 8 with *Cla* I linearises it (figure 4.17, *Cla* I). The linearised clone was large and hybridises with AATB probes (Figure 4.18 A and B). This indicated that the clone had a large insert harbouring AATB sequences.

Digestion of clone 8 with Bgl II generated three fragments: one large fragment of greater than 30 kb and two smaller fragments of 6.3 kb and 4.8 kb. Super-Cos One has one Bgl II site (see figure 4.1). The generation of three fragments with this enzyme indicated the presence of two further Bgl II sites within the clone. The AATB transgene has no Bgl II sites so the clone 8 insert contained DNA that was not of AATB origin. All three fragments hybridised with the cosmid probe (figure 4.13 b). Only the largest of the three fragments hybridised with the AATB probe (figure 4.13 a). It was at first surprising that all three fragments generated hybridised with the cosmid. The intensity of the 6.8 kb and the 4.8 kb fragment was similar suggesting a molar ratio of one to one. This indicates that the 4.8 kb fragment was generated from digestion of clone 8 rather than being present as a separate entity.

Not I digestion of clone 8 generated five fragments: the fragments were 15.2 kb, 10.6 kb, 6.8 kb 6.3 kb and 4.8 kb. Three of the fragments generated hybridised with the 3' AATB probe (probe (d), figure 4.2). These were the 15.2 kb, 10.6 kb and 6.2 kb fragments. Two of these three fragments also hybridised with the 5' AATB probe (probe (c), figure 4.2). The fact that the 6.3 kb fragment did not hybridise with probe d suggested that the fragment was the 3' end of a truncated AATB transgene (see figure 4.20). The 10.6 kb fragment hybridised with both AATB probes and was the size of the original injected AATB transgene. Its generation from the clone 8 insert by Not I suggested that this transgene was part of an array and, that the Not I site had been regenerated between it and adjacent transgenes (see also analysis of clones 5, 10 and 9). The intensity of this fragment relative to the 15.2 kb fragment demonstrated that only one single 10.6 kb transgene was liberated from the insert. The 15.2 kb fragment contained both 3' and 5' AATB sequences and thus contained

most if not all of the AATB transgene. None of the AATB hybridising fragments hybridised with the cosmid probe. The results of Not I/Bgl II double digests suggest that the 15.2 kb fragment also contained murine DNA (see later this section).

Not I liberates the insert from the cosmid vector. The predicted size of the cosmid vector is 6.8 kb. As expected the 6.8 kb Not I fragment hybridised with the cosmid probe. In addition the 4.8 kb fragment also hybridised with the cosmid probe. This was most unexpected and suggested that this clone was not as straightforward as the others (clones 5, 10 and 9) that generated only the expected single 6.8 kb cosmid hybridising band upon Not I digestion.

A Not I/Bgl II double digest was performed on clone 8 (figure 4.12). Three fragments were found to hybridise with the cosmid probe (figure 4.13 b). Two of the fragments (4.2 kb and 2.6 kb) were expected too, as these were the predicted cosmid fragments, based on the known restriction map of Super-Cos-One. The third fragment was not expected. It was only 750 base pairs and, because of its small size, had not been visible on the ethidium bromide stained gel. No such small cosmid hybridising fragment was detected on single Bgl II or Not I digests. Close examination of the small fragments generated by Not I and Bgl II on 1% agarose gels found no small fragments were generated by Not I digestion, but three small fragments were generated upon digestion of clone 8 with Bgl II (see figure 4.19 A). These fragments did not hybridise with either AATB or cosmid probes indicating they were probably murine in origin. These fragments were later sequenced (see chapter 5). A fourth small fragment was generated on double digests (see figure 4.19 B) that hybridised with the cosmid probe only. This fragment was cloned and the ends were sequenced. It spanned the T3 site from one cosmid vector leading into the sequences of another

cosmid at position 3130. Thus the truncated vector was adjacent to the intact vector in clone 8 and contained all the sequences from 180 bp 5' of the Sfi I site (see figure 4.1) to the cloning site beyond the T7 promoter sequence. This is 4.8 kb: the size of the extra cosmid hybridising fragment in the Not I digests and, moreover, the size of the contaminating recombinant fragment. The truncated cosmid had lost the T7 promoter site the lambda cos site and the SV40 promoter but retained the cosmid origin of replication and the ampicillin resistance gene. Thus clone 8 contained two bacterial origins of replication.

In addition to the three small Bgl II fragments (see figure 4.19 A) Not I/Bgl II double digestion of Clone 8 generated another fragment which did not hybridise with any of the probes used in the analysis. This fragment was 3.6 kb in size and was generated by Bgl II digestion of the 15.3 kb Not I fragment (see figure 4.12). The small Bgl II fragments total approximately 1 kb of DNA. The total amount of mouse genomic DNA in clone 8 is at least 4.6 kb.

Eco RV was used to map the AATB 46.2 genomic locus (see Chapter 3). This enzyme generated two junction fragments from AATB 46.2: one 5' hybridising fragment of 7.3 kb (isolated in clone 9, see previous section), and one 3' hybridising fragment of 5.6 kb. Eco RV digestion of clone 8 generated a fragment, that hybridised with the 3' AATB probe but not the 5' probe, that was 5.6 kb (see band labelled (a), figure 4.18 B). The presence of murine DNA in clone 8 was shown by Not I/Bgl II digests of clone 8. The generation of the 5.6 kb fragment upon digestion with Eco RV was good evidence that the murine DNA was the flanking murine DNA of the AATB 46.2 locus.

Further corroborative evidence that clone 8 spanned the 3' chromosomal/transgene junction was obtained from the Nhe I digest. This enzyme generated a 6.3 kb junction fragment from the AATB 46.2 genomic locus (see chapter 3). Digestion of clone 8 with Nhe I also generated a 6.3 kb fragment that hybridised only with the 3' AATB probe (see fragment labelled (b), figure 4.18 B). Nhe I also generated a 2 kb fragment that did not hybridise with either AATB or cosmid probes. The murine DNA in clone 8 appears to be the AATB 46.2, 3' chromosomal/transgene junction. The murine DNA incorporated in clone 8 contained two Nhe I sites.

Both Eco RV and Nhe I generate an AATB repeat length fragment. The intensity of the fragment generated by these two enzymes is greater than the intensity of the 10.6 kb fragment generated by Not I and Sal I (see figure 4.17). More than one AATB repeat length was liberated from clone 8 by Nhe I and Eco RV. The 6.2 kb truncated transgene shown to be present by Not I digest contains both the Eco RV and Nhe I restriction sites in AATB (see figure 4.3). From the Not I data two other transgenes were shown to be present. The generation of two 10.6 kb fragments from the insert of clone 8 demonstrates that the three transgenes are arranged head to tail.

Both Nhe I and Eco RV also generate another fragment in addition to the aforementioned fragments. In each case this was larger than the other two fragments and comprised the remainder of the clone following the removal of the two AATB repeat length fragments and the junction fragments. These fragments are 17.6 kb for Eco RV and 14.8 kb for Nhe I (figure 4.15, and 4.17). Fragments of this size could only have been generated, from clone 8, by these enzymes, if the 4.8 kb Not I fragment was an integral part of the clone. Neither of these large fragments hybridised with the 5' and 3' AATB probes (probes (c) and (d) respectively, figure 4.2).

However the 17.6 kb Eco RV fragment was shown to hybridise with the entire AATB transgene (figure 4.13 a) and the cosmid probe (figure 4.13 b). The 17.6 kb Eco RV fragment therefor did contain AATB transgene sequences, but not the 5' most or 3' most sequences.

In summary clone 8 appears to span the 3' chromosomal/transgene junction from the AATB 46.2 locus. It contains at least 4.6 kb of murine genomic DNA. Not I digest of clone 8 generates a 15.3 kb fragment that contains the 3' most transgene of the array and the murine DNA. Another intact transgene was present that was liberated by Not I digestion. This indicated that the Not I site was regenerated between the transgenes (figure 4.20). The generation of two AATB repeat length fragments from the insert by Eco RV and Nhe I (both of which cut only once within AATB) showed that the transgenes present in the array were all orientated head to tail (figure 4.20).

Clone 8 also contained an extra part cosmid which had not been anticipated. Yields of DNA from clone 8 were always inferior to preparation of other clones suggesting that the presence of either the murine DNA or more likely the additional cosmid sequences were causing some problem. The extra cosmid sequences 'recombined out' of the clone readily resulting in a small plasmid being the most abundant species in DNA preparations from the clone (see figure 4.1 B). The recombination of the clone is not entirely surprising since the presence of the additional cosmid sequences constitute a direct repeat. Homologous recombination between these direct repeats would be expected to result in the excision of the intervening DNA (Lewin, Genes III, (1987), p598).

Figure 4.15 : Restriction Digests of Clone 8 (First DNA Preparation)

The figure shows a 0.3% agarose gel stained with ethidium bromide following the electrophoresis of various restriction digests of clone 8 (first preparation of cosmid DNA from this clone). M= raoul molecular weight markers, the size of the markers, in kb, are given on the right. The clone is linearised with Cla I, the linearised clone is indicated with an arrow. Many of the digests give rise to a 10.6 kb fragment, which is the AATB repeat length: these fragments are also indicated with an arrow. The small white arrows indicate the small contaminating fragments seen in this first preparation of DNA from this clone. The concentration of DNA in these fragments is considerably less than any other fragment on the gel even the 2.5 kb Eco RI fragment and the 2 kb Nhe I fragment. These fragments are barely visible in samples of uncut clone 8 on agarose gels. This in combination with their low concentration suggests they are contamination, rather than products of digestion of clone 8 with the named enzymes.

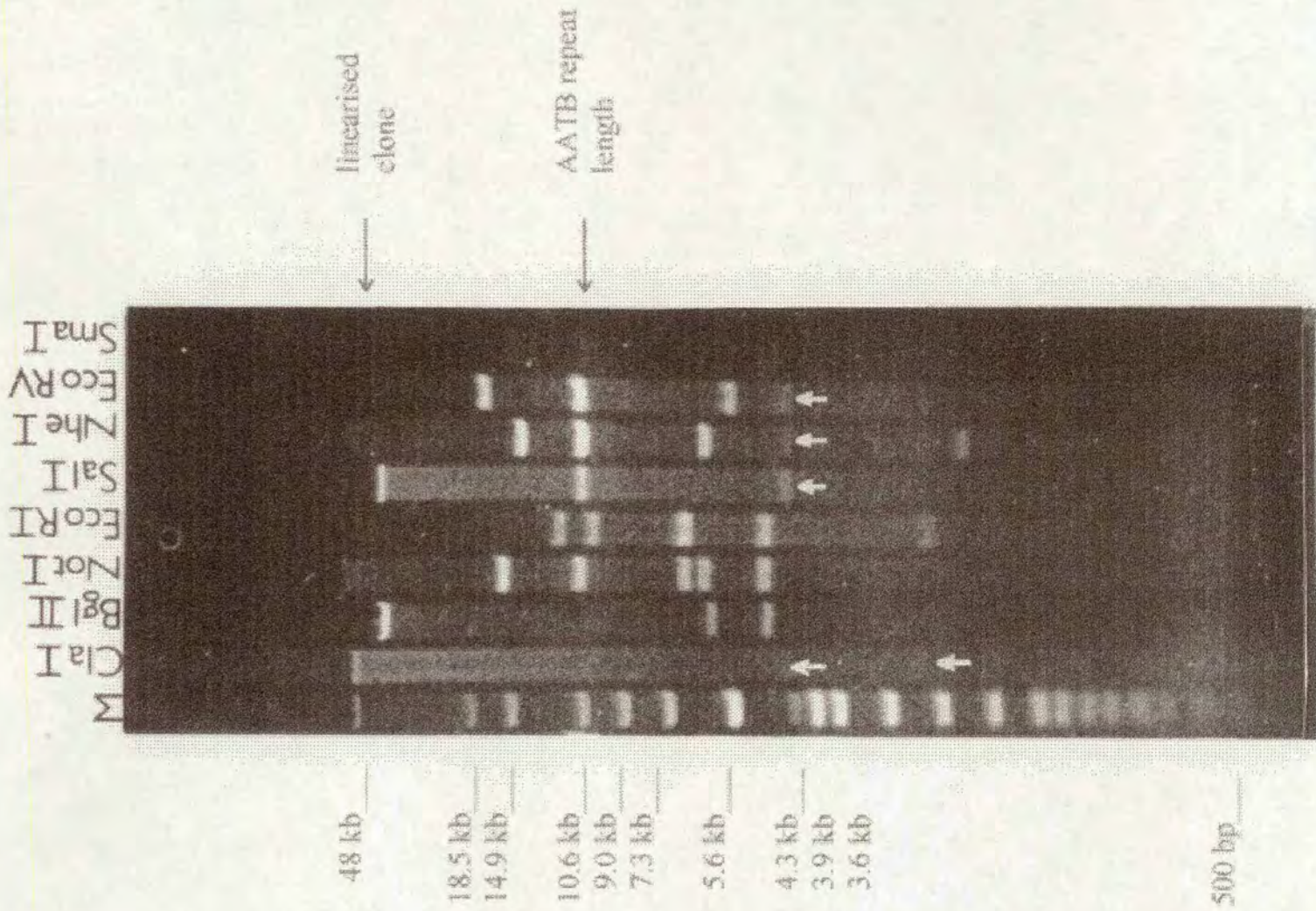


Figure 4.16 : Restriction Digests of Clone 8 (Second DNA preparation)

The figure shows a 0.3% agarose gel stained with ethidium bromide following the electrophoresis of various restriction digests of clone 8 (second preparation of cosmid DNA from this clone). M= raoul molecular weight markers, the size of the markers, in kb, are given on the right. The clone is linearised with Cla I, the linearised clone is indicated with an arrow. Many of the digests give rise to a 10.6 kb fragment, which is the AATB repeat length: these fragments are also indicated with an arrow. The small white arrows indicate the small contaminating fragments seen in this second preparation of DNA from this clone. The concentration of DNA in these fragments is considerably more than any other fragment on the gel. These fragments are the same size as the faint contaminating bands seen in the first preparation of DNA from this clone (see figure 4.15). Enzymes Cla I, Sal I, Eco RV and Nhe I appear not to cut within the contaminating cosmid, as two bands are visible in these lanes (s= supercoiled, o= open circle). Enzymes Not I, Bgl II, Eco RI and Sma I linearise the supercoiled contaminant giving a single band of 4.8 kb (L=linearised).

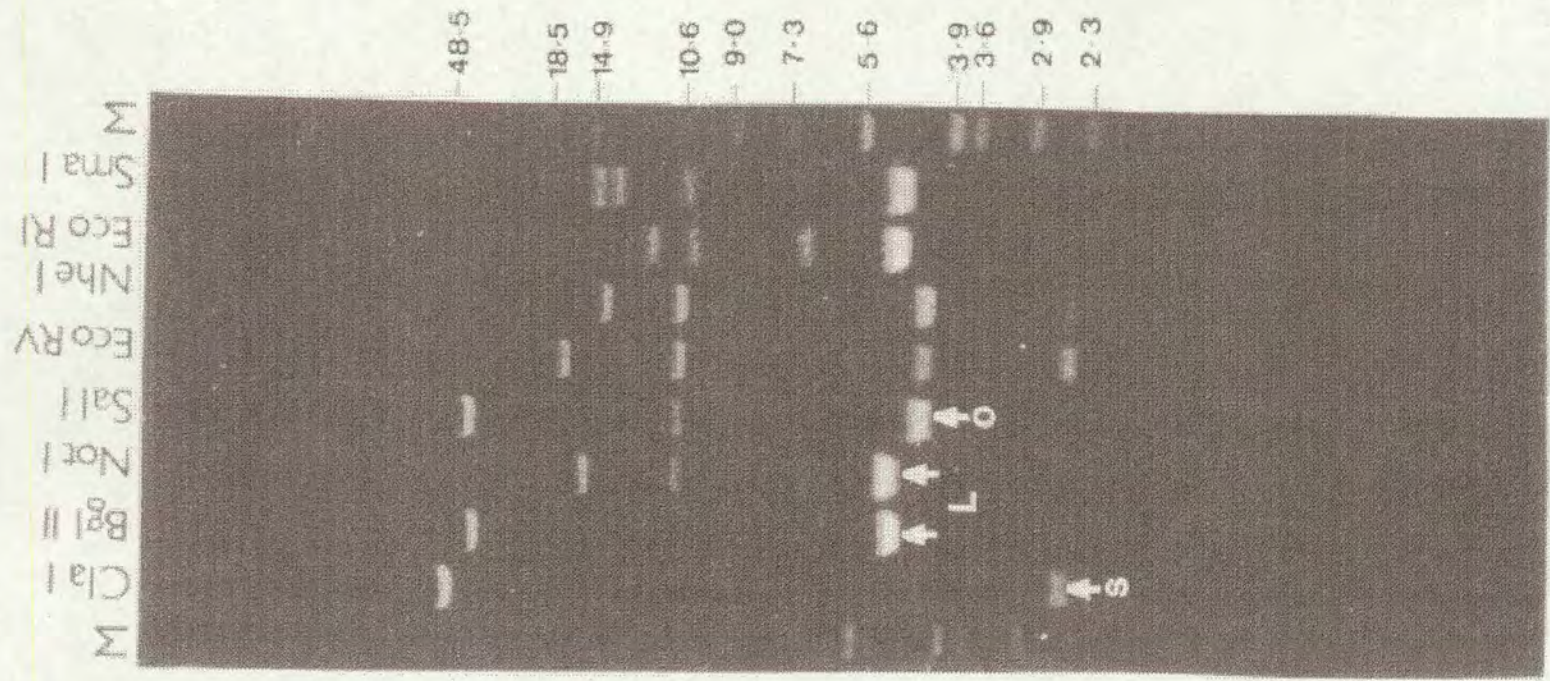


Figure 4.17 : Ethidium Stained Agarose Gel of Clone 8 Restriction Digests

The gel is a 0.3% agarose gel stained with ethidium bromide at 0.5 $\mu\text{g/ml}$. The enzymes used for each digest are given above each lane. M= raoul molecular weight marker (Appligene). The size of the markers is given in kb on the left. L= the linearised clone. R= the AATB repeat length fragment generated by enzymes Not I, Sal I, Eco RV, Nhe I and Sma I. a= the 3' Eco RV junction fragment and b= the 3' Nhe I junction fragment. These fragments are the same size as the 3' hybridising Eco RV and Nhe I junction fragments determined in the genomic mapping studies. This suggests that the insert of clone 8 spans the 3'transgene/chromosomal junction of AATB 46.2.

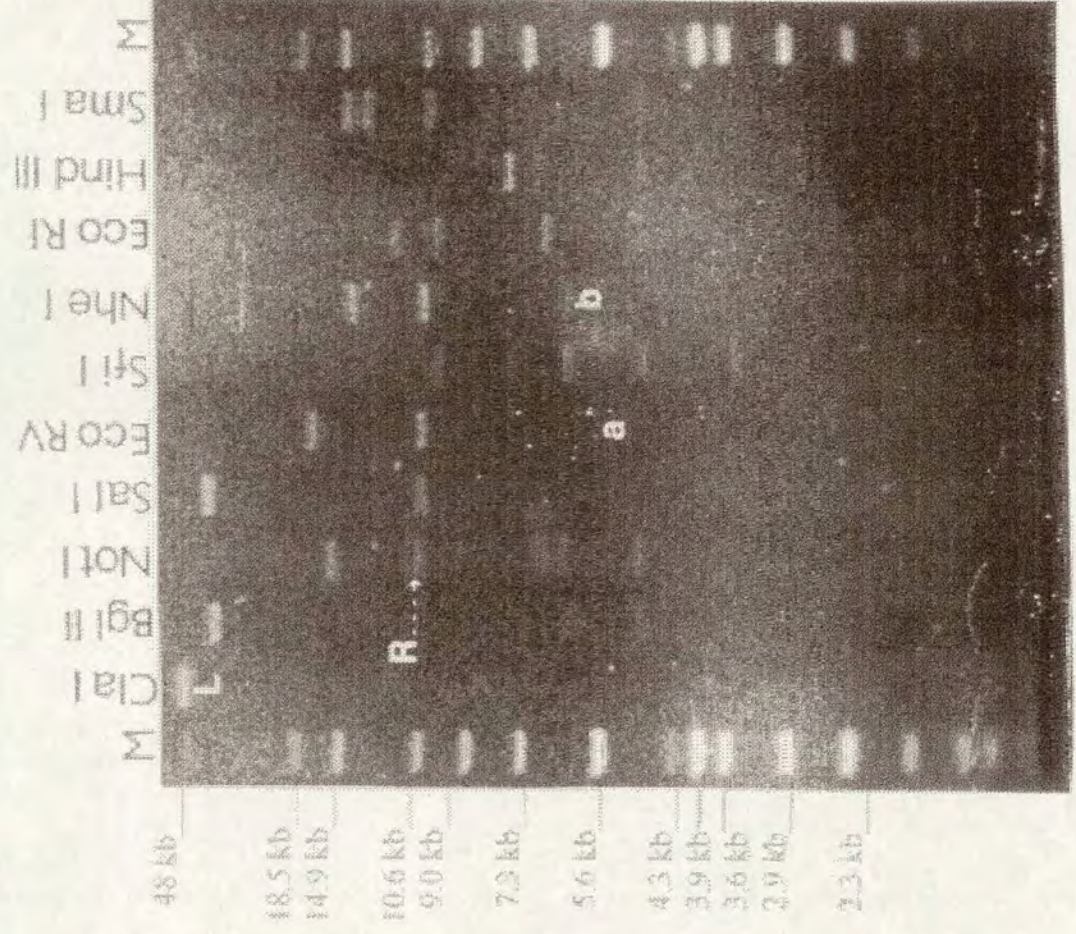


Figure 4.18 : Autoradiographs of Southern Blot of Gel in Figure 4.17

A.

The agarose gel in figure 4.17 was Southern blotted onto hybrid N nylon membrane (Amersham). The filter was hybridised with a 5' AATB probe (probe (c), figure 4.2). The autoradiograph was produced by exposing the filter to X-ray film, for 16 hours, at room temperature with intensifying screens. The filter was stripped and exposed to X-ray film to ensure removal of the signal, before being reprobated.

B.

This is an autoradiograph of the filter in figure 4.18 A. The filter in this instance was hybridised with the 3' AATB probe (probe (d), figure 4.2). The filter was exposed to X-ray film for 48 hours, at room temperature with intensifying screens.

Digestion of clone 8 with enzymes *Eco* RV and *Nhe* I generated fragments which were of the size expected of the 3' hybridising junction fragments (determined in the genomic mapping studies). Both of these fragments hybridised with the 3' probe but not with the 5' probe. This data further supports the contention that this clone harbours the 3 chromosomal/transgene junction of AATB 46.2. *Eco* RI generates a 2 kb fragment and *Nhe* I generates a 1.6 kb fragment (see figure 4 15) that did not hybridise with either AATB probe used in this analysis. This suggested the clone contained murine DNA.

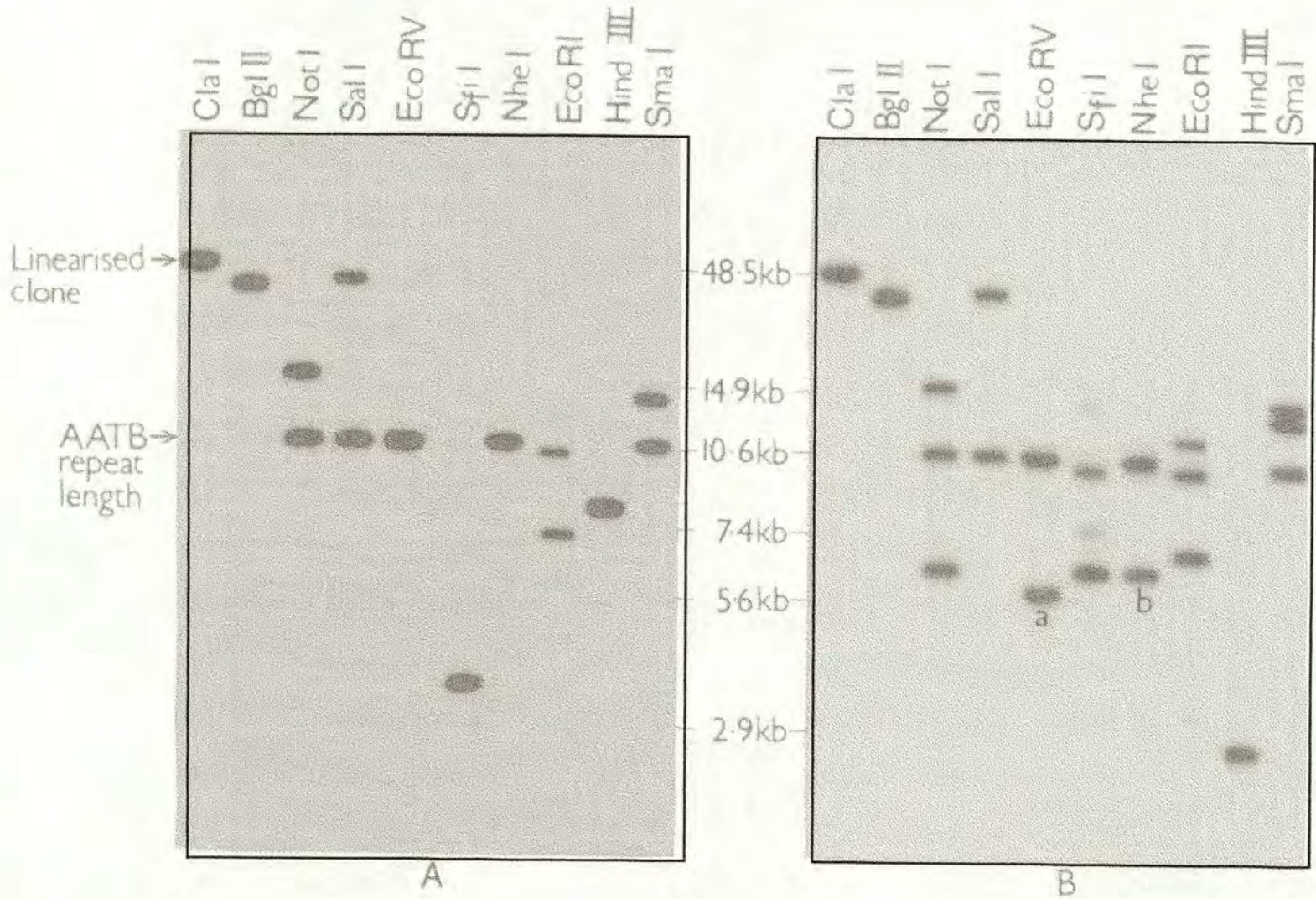


Figure 4.19 : Analysis of the Small Bgl II fragments

A. The figure shows a 1% agarose gel stained with ethidium bromide at 0.5 µg/ml. The sample is clone 8 digested with Bgl II. The gel demonstrates that the non AATB DNA present in the clone contains a cluster of Bgl II sites. digestion of the clone with this enzyme generates three small fragments (a, b and c) that do not hybridise with either AATB or cosmid probes (see figures 4.13 A and B)

B. The figure shows a 1% agarose gel stained with ethidium bromide at 0.5 µg/ml. The sample is clone 8 digested with both Not I and Bgl II. Double digestion generates the same three fragments (a, b and c) visualised on Bgl II digestion and one additional small fragment (d). This fragment hybridises with cosmid probe (see figure 4.13 B, NB digest of clone 8). This fragment was sub-cloned and sequenced and shown to be part of the SuperCos I vector (see figure 4.20, insert). Sequencing of the fragment demonstrated that a second truncated cosmid vector is also present in clone 8 and abuts the intact cosmid vector.

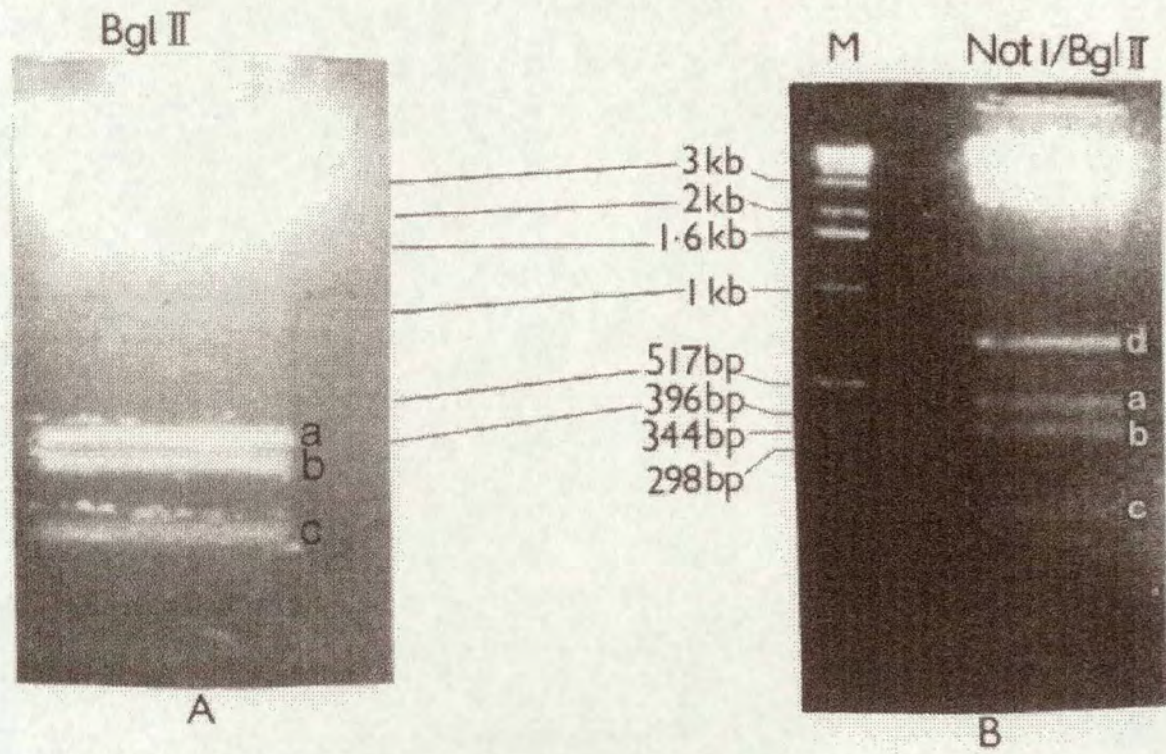
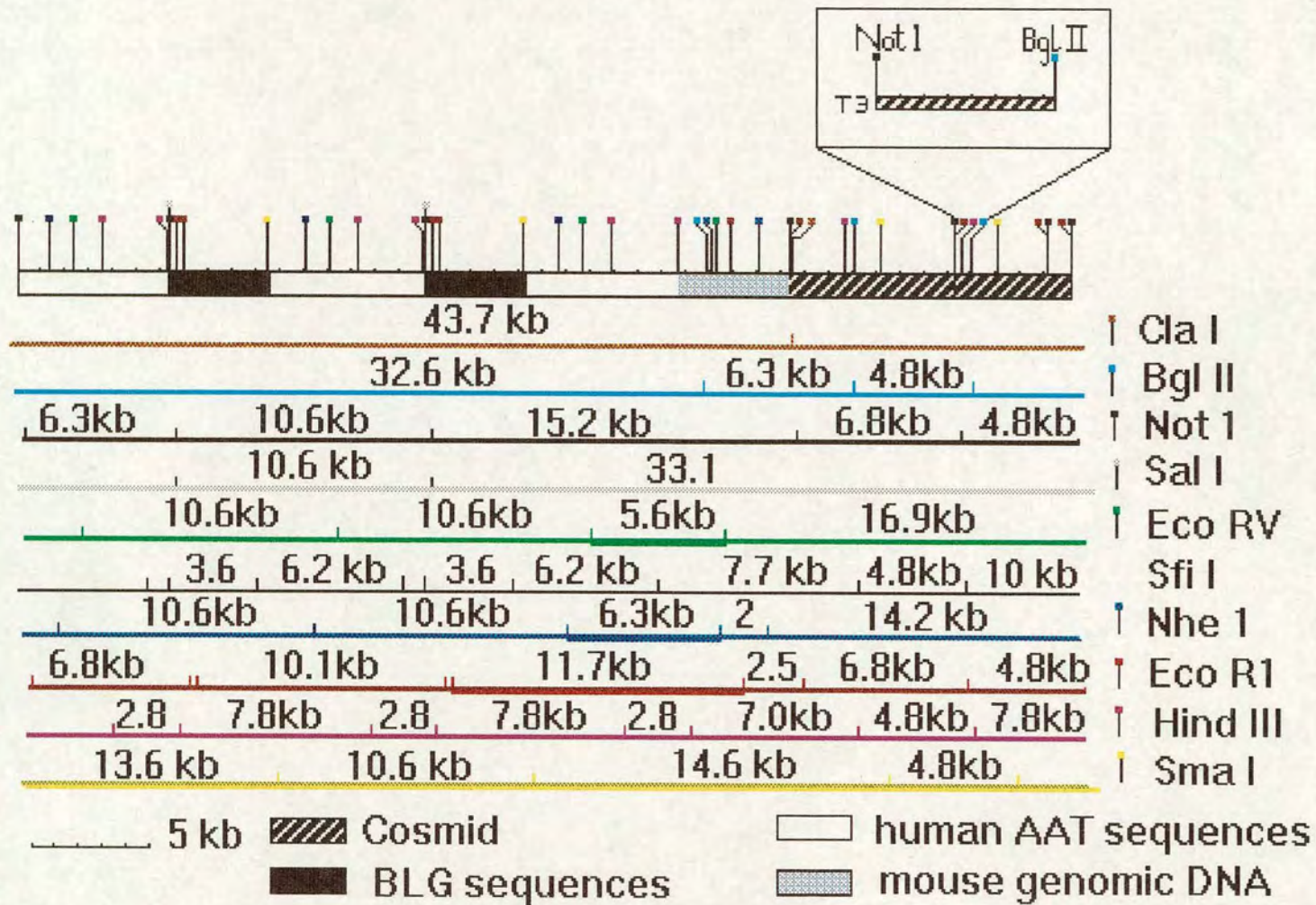


Figure 4.20 : Restriction Map of Clone 8

The figure shows the restriction map of clone 8 determined from the data presented in this section. The clone harbours one truncated copy of AATB that is 6.3 kb in size and is located at the 5' end of the insert abutting the vector, an intact copy of AATB that is liberated from the insert by Not I, and the transgene/chromosome junction at the 3' end. The junction fragment liberated by Not I is 15.2 kb in size and hybridises with both 5' and 3' AATB probes (probes (c) and (d), figure 4.2) which indicated that the fragment contained most of the AATB sequences. Hind III digestion of the clone indicates that the Hind III site 120 bp from the 3' end of AATB is present at the junction demonstrating the transgene at the junction is almost complete. The junction fragments generated by Eco RV, Nhe I digests and Eco RI digests are indicated using thick lines. Clone 8 also contains a duplication of 4.8 kb of the cosmid sequences. The intact 6.8 kb cosmid and the 4.8 kb cosmid abut each other and the Not I/Bgl II fragment that was sequenced is shown in the insert.



4.3 Comparison Of Clones With Genomic DNA

The transgene array/chromosomal junctions for both the 5' and 3' ends of the AATB 46.2 locus were successfully isolated from the library screen. The amount of mouse chromosomal DNA was approximately 4.7 kb in clone 8 and 4.6 kb in clone 9. The larger Not I fragments from both clones 8 and 9 spanned the chromosomal/transgene junction. These fragments hybridised to both 5' and 3' AATB probes. This indicated the transgenes at both the 5' and 3' ends of the array contained much of the 3'most and 5'most AATB sequences.

To demonstrate that the junctions were indeed cloned without rearrangement. Genomic DNA of AATB 46.2 and the clone inserts have been compared by Southern blotting. Control mouse DNA laced with single copy equivalents of Clone 8 or Clone 9 was digested with restriction enzymes Eco RI and Eco RV and subject to electrophoresis adjacent to similarly digested AATB 46.2 genomic DNA. The size of the 5' and 3' hybridising fragments of the clones and genomic DNA are thus compared directly. Autoradiographs of the Southern blot (figure 4.21) probed with 5' and 3' AATB probes confirms the clones do indeed span the junction fragments of both ends of the AATB 46.2 locus. No major rearrangements of the locus have occurred during the cloning procedure.

Figure 4.21 : Direct Comparison of Clones 8 and 9 With Genomic DNA of AATB 46.2

The figure shows an autoradiograph of genomic DNA run along side control DNA laced with one copy equivalents of the cosmid clones 8 and 9. Two enzymes were used for this analysis Eco RV and Eco RI.

A

Is probed with a probe from the 5' end of AATB (probe c, figure 4.2). All three lanes show the expected intense signal of the internal AATB fragments from the array (10.6 kb for Eco RV and 10.1 kb for Eco RI). Clone 8 shows no junction fragment signal on hybridisation with this probe. The signal from the junction fragments is indicated by **jf**. The signal from Clone 9 and genomic DNA are identical demonstrating that clone 9 harbours the 5' transgene/chromosomal junction of AATB 46.2. This result shows that no rearrangement of the locus has occurred during cloning of the locus.

B

Is the same blot as in figure A probed with 3' AATB sequences (probe (d), figure 4.2). All three lanes show the expected signal of the internal fragments of the AATB array. Clone 9 shows no other signal with this probe but clone 8 and the genomic DNA give identical signals from the junction fragments (**jf**) demonstrating that clone 8 harbours the 3' end of the AATB 46.2 locus (the junction fragments appear to be of slightly different size this is because the DNA in lane 3 appears to have run anomalously and it is my belief these fragments are indeed the same size). The result demonstrates that the 3' transgene/chromosomal junction was cloned without rearrangement.

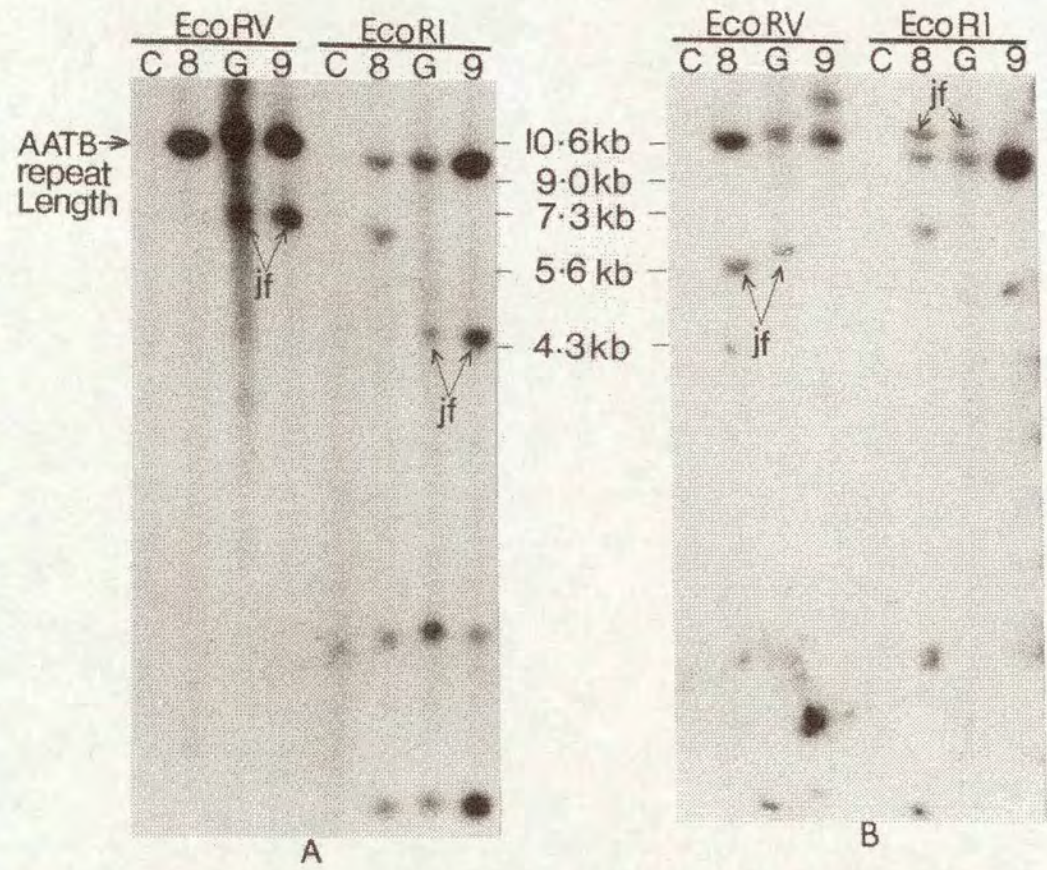
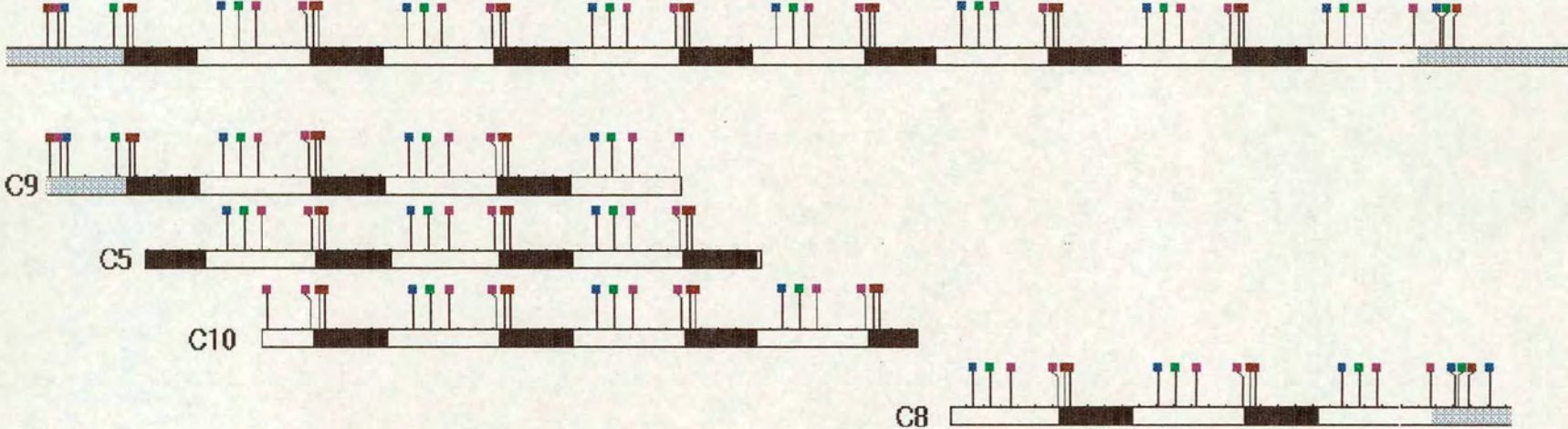


Figure 4.22 : Comparison of Genomic Map With Maps of Clones 5,8,9 and 10

The figure shows the restriction map of the genomic locus compared with that of the clones isolated from the cosmid library. The genomic map was obtained from the mapping studies detailed in Chapter 3. The genomic mapping studies indicate that AATB 46.2 harbours seven copies of AATB and thus contains six transgene/transgene junctions. The mapping of the clones is detailed in Chapter 4. Alignment of the clones with the genomic map indicate that clones 9,10 and 8 together span the AATB locus except for a few hundred base pairs. Since it is not possible to align clones 5 and 10 exactly it is not certain that the entire locus is contained within the four cosmid clones but it is nevertheless very likely.

genomic locus



mouse genomic DNA BLG sequences human AAT sequences

Discussion

This chapter describes the successful isolation of cosmid clones harbouring the murine DNA flanking the AATB 46.2 transgene locus. Only one clone including the 5' flanking DNA, and one containing the 3' flanking DNA were obtained. Both of the clones contain approximately 4.6 kb of murine DNA plus two intact copies and one part copy of the AATB transgene.

Other clones were isolated which contained only DNA of AATB origin. These clones were internal to the array each contained at least three transgene/transgene junctions. It is likely that together the clones span the entire transgene locus of AATB 46.2. Each clone harbouring the junction fragments of the locus spanned at least two transgene /transgene junctions. In every case the Not I site was present between the transgenes. The genomic mapping studies detailed in chapter 3 show clearly that the transgenes in the array are arranged in a head to tail manner. This has been confirmed by the cloning of the locus, detailed in this chapter. It has been widely demonstrated that transgenes present in an array are generally oriented head to tail. The paucity of head to head and tail to tail orientation of transgenes within transgene loci has lead to the proposal that transgene arrays do not form by simple end on ligation of the linear injected fragments, but rather by mechanisms involving homologous recombination. The loci of both AATB 35 and AATB 46.2 are no exception to the rule. All transgenes within the array of each line are oriented head to tail. Thus it seems unlikely that the Not I site is regenerated between the transgenes by simple end on ligation despite the fact that end on ligation would be expected to be very efficient in this instance because the enzyme generates a 5', GC rich, 4 base pair overhang.

Nevertheless the Not I restriction site is present at all transgene/transgene junctions within the array. This is suggestive of efficient ligation occurring between the Not I "sticky ends". Circularisation of linear input DNA has been shown to be efficient and accurate upon cytoplasmic injection into mammalian culture cells (Kopchick and Stacey, 1985). The Not I site could be re-established in the injected AATB transgene by circularisation of the injected fragment if this occurred before nibbling occurs at the transgene ends. Some studies have reported the regeneration of a restriction site between transgenes within an array (Palmiter et al., 1982, Ninomiya et al., 1988,) but this has never been demonstrated in a cloned transgene locus. A study performed by Hamada et al., (1993) investigated the nature of four transgene/transgene junctions "nibbling" of the transgenes had occurred in all instances. The transgene in the Hamada study was blunt ended prior to injection. Ligation of blunt DNA fragments is believed to be less efficient than that of fragments with cohesive termini and, thus blunt ended transgenes may be more susceptible to end damage than those having complementary ends.

Kopchick and Stacey (1985) found that the nature of the ends of the linear fragments had a large bearing on the fidelity of circularisation of cytoplasmically injected fragments . The difference between the Hamada study and this study demonstrate this is probably also true of pronuclear injection. It has been proposed that the head to tail transgene arrays are built up by homologous recombination between circular molecules or circular and linear molecules or circularly permuted molecules. The presence of an intact Not I site between each of the transgenes in the AATB 46.2 locus would suggest that all of the transgenes of the array have undergone efficient high fidelity circularisation prior to their inclusion in the array.

CHAPTER 5 : THE CONSTRUCTION OF THE SECONDARY TRANSGENE

5.1 Introduction

The previous chapters outline the restriction mapping and cloning of the AATB 46.2 locus. The aim of this work was to investigate the role of the murine sequences flanking the transgene array, of this locus, in driving high level hAAT expression in this line. Unfortunately the transgene array of the AATB 46.2 locus contained several more copies of the transgene than had been anticipated at the outset of this work (see chapter 3). It was therefore, not possible to clone the entire high expressing locus for generation of secondary transgenic mice.

It was therefore proposed to construct a secondary transgene which harboured a single intact copy of AATB flanked by the cloned 5' and 3' flanking chromosomal DNA of AATB 46.2. Two cosmid clones (clone 8 and clone 9), each containing almost 5kb of the flanking murine DNA were obtained from the library screen (see Chapter 4). These clones were used to generate a secondary transgene comprising one copy of AATB flanked by these sequences (see figure 3.10). The construction of this transgene is detailed in this Chapter.

5.2 : The Pre-integration site of AATB 46.2 has not been Characterised

An attempt was made to establish the nature of the pre-integration site of the AATB 46.2 locus. It was hoped that Southern blotting would establish if the flanking 5' and 3' flanking DNA of AATB 46.2 was contiguous or whether rearrangement of the genomic locus had occurred.

Similarly digested AATB 46.2 liver DNA and non transgenic control DNA were subject to agarose gel electrophoresis, Southern blotting and probing with 3' murine DNA from the AATB 46.2 locus. The result was a smear in all lanes (figure 5.6). This type of result is diagnostic of the presence of repetitive sequences in the probe. Because of this, no information regarding the pre-integration site was obtained from this Southern. The experiment was not repeated with probes obtained from the 5' murine DNA, and it is not currently known if the locus is a result of straightforward insertion of a transgene array into the genome or if the chromosomal site of insertion has been rearranged. No integrated transgene locus that has been subject to detailed analysis has been demonstrated to be a straightforward insertion of a transgene array into the genome. Integration has always been associated with either simple deletions (Hamada et al., 1993) or other rearrangements (Palmitter et al., 1987). Since this locus appears to more straightforward than many analysed to date, mapping of the pre-integration site of this locus may provide further information about the nature of transgene integration.

5.3 : Only Seven Bases are Missing At The 5' Chromosomal/Transgene Junction

Restriction mapping of clone 8 and clone 9 indicated that the transgenes at both the 5' and 3' end of the transgene concatamer of line AATB 46.2 were almost complete. Sequencing was performed to determine more precisely the nature of the chromosomal/transgene junction of clone 9. A 7.8 kb Eco RV junction fragment, from the clone, was sub-cloned into bluescript phagemid (Stratagene). Using a primer to the 5' end of the AATB transgene (kindly provided by Anthony Thomlinson, PPL

Therapeutics Ltd.) the chromosomal/transgene junction was sequenced (see figure 5.1). Sequencing demonstrated that the transgene abutting the chromosomal DNA was missing only seven base pairs (including the four base single stranded end generated by Not I digestion of P-Polly III-I-AATB) (see figure 5.1). This small deletion was the only perceptible damage to the most 5' copy of AATB in the array.

5.3.1 : The Immediate 5' Flanking DNA is AT Rich

The transgene concatamer had integrated next to a thirty base pair microsatellite, comprising ten TAA trinucleotide repeats (see figure 5.2 A). A further 145 bp of the murine DNA immediately 5' to the transgene array was sequenced and the entire region was found to be AT rich (80 % AT: see figure 5.2 B). Mammalian genomes are typically 40-60% AT (Lewin 1987) rich and an AT content of 80 % is unusually high.

AT rich regions of chromosomes have been shown to bind the Nuclear Matrix. Though this region is AT rich it does not however contain motifs that have been identified in matrix attachment regions. Two regions have been highlighted in the immediate 5' flanking DNA (figure 5.2 A). Region (a) ATATAT is the same as a previously identifies nucleation site for base un-pairing that was identified to be important in the binding of a MAR binding proteins SAT B1 (Dickenson et al., 1992) and nuclonin (Dickenson et al., 1995). There are of course thousands of ATATAT motifs in the genome which but the presence of this nucleation site in such an AT rich environment may be significant. The other region is the microsatellite (AAT)₁₀ this region may also have high base un-pairing potential. This is a property of some but not all AT rich sequences and has been proposed to be important in the regulation of

superhelical strain of transcribed loci and therefore important in the regulation of gene transcription per se (Bode et al., 1992).

A further 342 bp of murine DNA from clone 9 was also sequenced (figure 5.2 B). This region was situated approximately 1 kb upstream of the 5' AATB transgene. Unlike the immediate 5' flanking DNA the sequence of this upstream region was not more than averagely AT rich. It does however contain an AT dinucleotide repeat of twenty base pairs long which is very likely to have high base un-pairing potential. Other than this no significant motifs were found in these sequences. A search of the EMBL database with the sequence of both of these regions, using the FASTA program, detected no significant homologies with other sequences in the database.

Figure 5.1 : Sequencing Of The Transgene/Chromosomal Junction Of Clone 9

The 7.8 kb Eco RV junction fragment of clone 9 was sub-cloned into bluescript phagemid (Stratagene) and the junction (region A) was sequenced using a primer of the 5' AATB sequences (BLG amp 10: kindly donated by A. Thomlinson, PPL Therapeutics Ltd.). The 5' most transgene of the AATB 46.2 array is almost intact. Only seven bases are missing from the transgene end as it would have been on injection. These seven base pairs include the four base pair single stranded end of the transgene generated by Not I digestion. The murine sequences immediately 5' of the transgene end are that of an AAT microsatellite that is 30 bp long. The regions underlined with heavy black lines are those sequenced. The sequence of these regions is given in figures 5.2 A and B

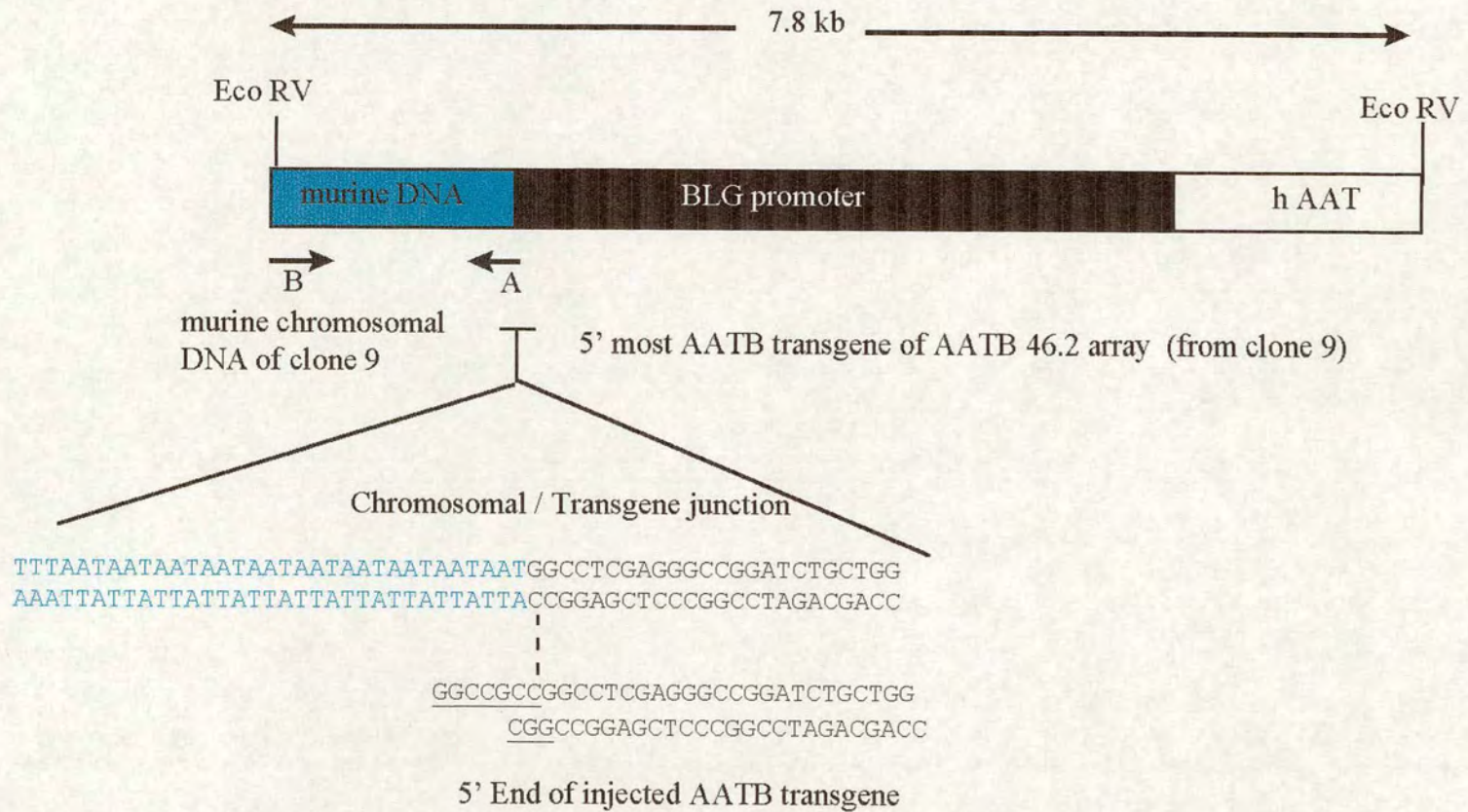
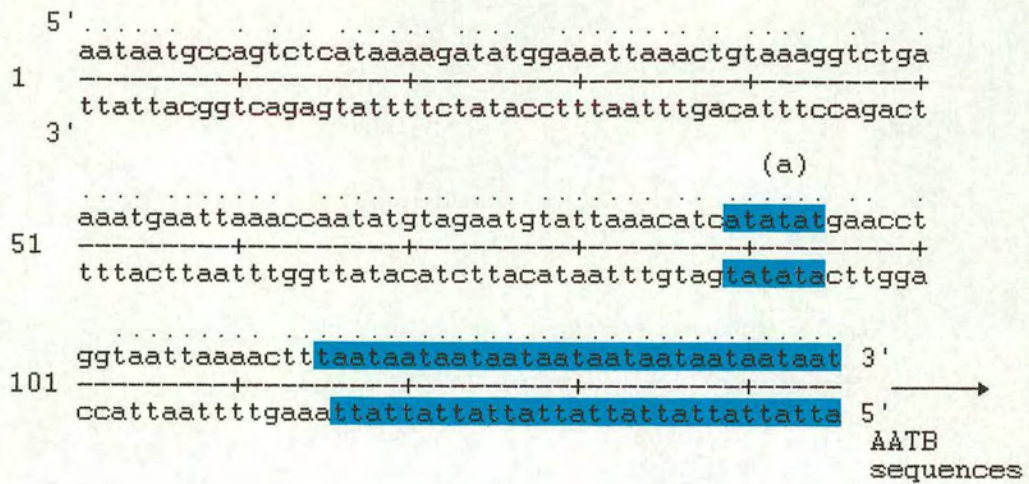


Figure 5.2 A : AT Rich Sequences Of The Immediate 5' Flanking DNA

A total of 145 bp of murine flanking DNA were sequenced using the BLG amp 10 primer. These sequences proved to be very AT rich. Highlighted in blue are regions (a) = an ATATAT element proposed to be a base un-pairing nucleation site (Dickenson et al., 1992) and (b) = 30 bp microsatellite composed of an AAT trinucleotide repeat. Dots are placed above A and T residues present in the sequence to demonstrate visually the AT richness of the sequenced region.



Including Trinucleotid AAT Repeat

Total Number Of Bases = 145 bp
 Total Number of A & T = 116 bp
 Percentage of A & T = 80 %

Excluding Trinucleotide AAT Repeat

Total Number of Bases = 115
 Total Number of A & T = 86
 Percentage of A & T = 75%

Figure 5.2 B : Sequence of 5' End of Eco RV Junction Fragment

The Eco RV Junction Fragment of clone 9 was sub-cloned into bluescript phagemid and the 5' end sequenced using the T3 and SK priming sites of bluescript. The sequence of this region (region (b), figure 5.1) is shown in the figure. The region highlighted in blue is a dinucleotide repeat (AT)₁₀ such sequences are thought to have high base un-pairing potential and may have a role in gene transcription.

from bluescript prining site

→

5' tttgacat tttgagtgagaacagttctcagcctttgogccagttcaggag
1 +-----+-----+-----+-----+
3' aaactgtaaaactcactcttgtcaagagtcggaaacgoggtcaagtcctc

51 gaggcacacocggctaataatgctgcagagtaagtctttcttaagtttagtatt
+-----+-----+-----+-----+
ctccgtgtggccgattacgacgtctcattcagaaagaattcaaatcataa

101 tctcatgggtcttctgtgtaatgtgctgtaccctgtgtccattttgtcc
+-----+-----+-----+-----+
aagagtaccagaagacacattacaogacatgggacacaggtaaaacagg

151 acaatccagttcttgacagacacattttcttactacctgtttgattttaga
+-----+-----+-----+-----+
tgtaggtcagaactgtctgtgtaaaagaatgatggacaaactaaaatct

201 gcttcg tttctcatccatttgggatgatattaataatgccagttccta
+-----+-----+-----+-----+
cgaagcaaaggagtaggtaaaccctactataattattacggtcagagtat

251 aacgatatggaattaacagtaaatgaataacaatatgtagaatgtattaa
+-----+-----+-----+-----+
ttgctataccttaattgtcatttacttattgttatacatcttacataatt

301 actcatatagactgtataactt **atataatataatataatata** 3'
+-----+-----+-----+-----+ →
tgagtatactgacatatgaa **atataatataatataatata** 5' towards
AATB
sequences

5.4 : Fifty Base Pairs of AATB have been Deleted At the 3' junction of AATB 46.2

Previous mapping studies on clone 8 had determined that the 3' most Hind III site of AATB was present in the transgene at the 3' end of the array. However they also demonstrated that the 3' most Bam HI site of AATB was missing from the last transgene in the array (see Chapter 3, Section 3.5). Clone 8, which harboured the 3' AATB 46.2 chromosomal/transgene junction was unstable however it was possible to sub-clone the 3' junction from the TAB clone.

A Hind III/Not I fragment that spanned the junction was sub-cloned into bluescript phagemid (Stratagene). In order to find out more about the 3' murine sequences a major part of the subclone was sequenced using the Erase-a-Base system (Promega). This system allows nested deletions of the plasmid insert to be made so the insert can then be sequenced from one primer. This clone was sequenced using the M13-20 priming site of bluescript. In total 2.6 kb of the 3' flanking DNA was sequenced. Only one strand was sequenced. The sequencing demonstrated that only 50 bp had been deleted from the 3' most transgene of the array (figure 5.3).

In order to generate a functional secondary transgene, that could be compared with AATB, using the approach outlined in the following sections of this chapter, it was essential the 5'most and 3'most transgenes of AATB 46.2 were nearly complete. The TAB secondary transgene harbours a single AATB transgene that has seven bases missing at the 5' end and 50 bp missing at the 3' end. The lack of the terminal 50 bp of AATB in the secondary transgene was not considered to be important: as the missing 50 bp were in the 3' untranscribed region of hAAT approximately 2kb downstream of the polyadenylation signal of the hAAT minigene.

5.4.1 : The 3' Flanking Murine DNA of AATB 46.2 Contains Murine Repetitive Elements

The regions sequenced are shown in figure 5.4. The sequences are given in figures 5.5 A, B and C. A search of the EMBL data base with these sequences detected three regions that displayed extensive homology with either B1 or B2 murine repetitive elements (see figures 5.5 A and B). The presence of these sequences in the 3' murine DNA was consistent with the findings of the pre-integration Southern (figure 5.6). Other than this no homologies with sequences in the database were found.

The murine B1 repetitive element is an *Alu* type element originally identified by its homology with the human *Alu* consensus sequence. The human *Alu* element is approximately 300 bp and there are an estimated 500 000 such elements dispersed throughout the haploid human genome, which have been calculated to be spaced about 5 000 bp to 8 000 bp apart (for a review see Jelinek and Schmid, 1982). The human *Alu* repetitive elements are a family of elements which are non identical however they are sufficiently similar for a consensus sequence to have been derived. The nucleotide sequence of individual cloned *Alu* family members differ by an average of only 10% from the consensus sequence, and most of these differences appear to be randomly distributed throughout the 300 nucleotide sequence. It has been demonstrated that *Alu* family members are transcribed in vitro by RNA polymerase III yielding discrete low molecular weight RNA molecules, yet despite their high copy number seldom give rise to high levels of polymerase III transcribed RNAs *in vivo*.

Unlike the human *Alu* element the rodent B1 repeat sequence is only 130 bp long but shows significant sequence similarity with the monomer unit of the human *Alu* element suggesting that the human and rodent repetitive elements have a common ancestry. At first it was thought that the B1 murine repetitive elements were not transcribed, however, low level transcription of such elements has been detected and the level is elevated in certain undifferentiated embryonal carcinoma cells and SV40 transformed cell lines (Carey and Singh, 1988). The level of transcription of these elements is low and, although no studies of the chromatin structure of these elements have been published, it is thought that these elements are generally sequestered in transcriptionally inactive chromatin. The B1 element in region B of the 3' flanking murine DNA (figures 5.4 and 5.5 B) has an A rich sequence at one end and is flanked by 13 bp direct repeats. This is consistent with the proposed insertion of these elements into the genome by retroposition.

Two regions of the 3' murine DNA showed homology with murine B2 repetitive element family. The transgene array appears to have integrated into the 5' end of a B2 repetitive element. The second B2 element is 440 bp downstream of the first. Like the B1 element this B2 element has an A rich region at one end and is flanked by 8 bp direct repeats (see figure 5.5 A). B2 elements have also been demonstrated to be transcribed by Polymerase III *in vitro* but these elements are almost entirely sequestered by chromatin proteins in quiescent cells. The transcriptional repression of these elements *in vivo* is thought to be mediated by histone H1. However it has been shown that they become partially unmasked in growing or SV40 transformed cells (Carey et al., 1986). As with the human *Alu*

family no definitive function has been determined for the B1 and B2 repetitive elements.

The 3' flanking DNA did contain numerous so called TG boxes. The TG box has been proposed to be a motif of a subset of matrix attachment regions (see Boulikas, 1993a and 1993b for a review). In addition the 3' flanking DNA contains GT dinucleotide repeats which have been proposed to have a propensity to form Z DNA under certain physiological conditions and have been proposed to have evolutionarily conserved chromosomal positions in *Drosophila* (Pardue et al., 1987). There was also a liberal amount of other microsatellite repetitive DNA scattered throughout the sequenced regions. Only 2 kb of the 3' flanking DNA of clone 8 has not been sequenced.

Figure 5.3 : Sequencing Of The 3' Transgene/Chromosomal Junction Of Clone 8

The 4.8 kb Hind III/Not I fragment of the secondary transgene was sub-cloned into bluescript. Sequencing of the 5' end of the fragment using the KS primer of the bluescript cloning vector allowed the sequencing of the 3' chromosomal/transgene junction of AATB 46.2. The figure shows the sequence at the junction and compares this with the 3' end of the injected AATB transgene. 50 base pairs of AATB were missing from the AATB transgene at the 3' end of the AATB 46.2 transgene array. other than the 7 base pairs missing at the 5' end this was the only perceptible damage sustained by the AATB transgenes of the AATB 46.2 locus. This transgene was used to construct the secondary transgene, thus the AATB transgene in the secondary transgene (TAB) differs from the AATB transgene of the Archibald and Carver studies in this respect.

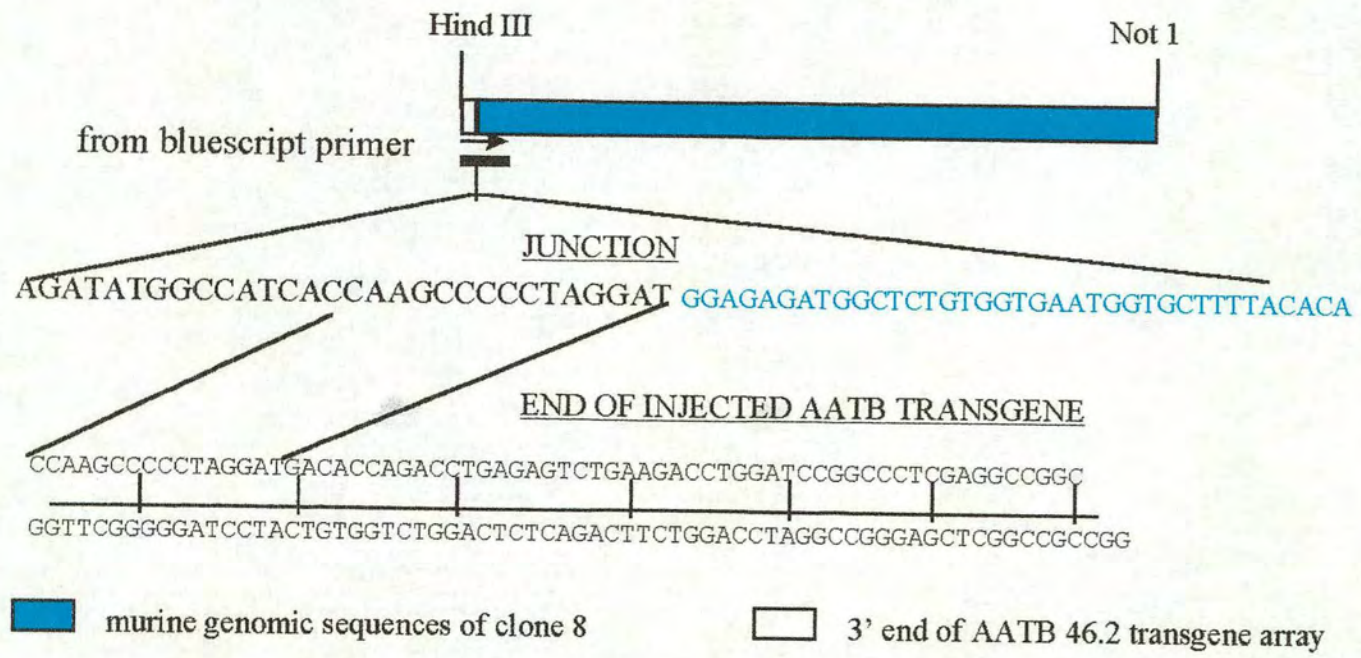
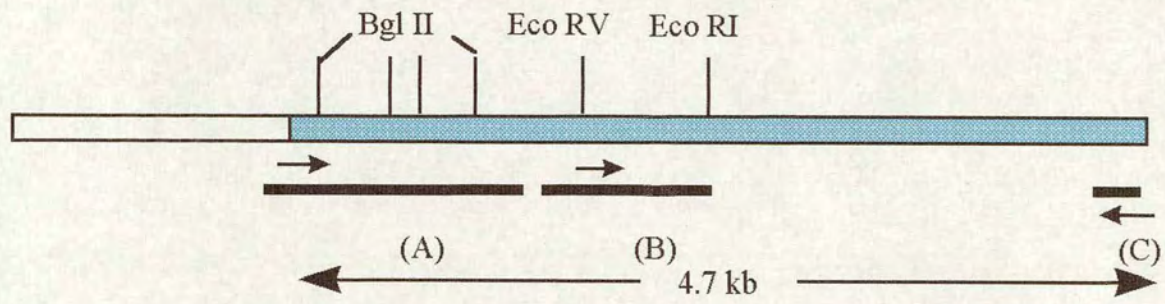


Figure 5.4: Regions Of Clone 8 Murine DNA Sequenced

The figure shows the regions of the murine DNA of clone 8 that were sequenced using the Erase-a-base system (Promega). Section (A) is 1462 bases long, section (B) is 994 bases long and section (C) is 201 bases long. The sequence of these sections and the regions of interest are shown in figures 5.5(A), (B) and (C). The position of restriction sites previously identified in the restriction mapping of the clones are shown.



- 3' end of AATB transgene
- Murine genomic DNA from clone 8
- Regions of DNA sequenced

Figure 5.5 (A) : Sequence of Section (A)

*The figure shows the sequences of the murine chromosomal DNA of clone 8. Section (A) was 1462 bases long. Using the FASTA program these sequences were used to search for homology with sequences in the EMBL data base. The sequences underlined show considerable homology with mouse B2 repetitive elements. The second B2 element has an AT rich region at one end (shown with tilde overlining) and is flanked by short direct repeats (indicated with by *). The regions overlined with dashed lines are GT dinucleotide repeats which have been proposed to be able to adopt Z DNA structure in cells under physiological conditions and proposed to have evolutionarily conserved chromosomal positions in Drosophila (Pardue et al., 1987). The regions overlined with heavy lines are GT boxes (which have been proposed to be a motif of a subset of matrix attachment regions. The positions of the Bgl II restriction sites are shown*

A

junction
↓

AATB Sequences 5' ggagagatggctctgtggtgaatggtgcttttacacag
 3' cctctctaccgagacaccacttaccacgaaaaatgtgtc
 B2 →

39 gtcctgagttcaattctcagcaaccacatggtggctgatgctcactctg
caggactcaagttaagagtogttgggtgtaccaccgactacgagtgaagac

89 gcttgtaggtatatacagaaaaagcattcatatataaaatcaggaat
cgaacatccatataatgtctttttcgtaagtatatattttatgtcctta

139 aaagattaaaaacaaaacatttttncacgggatgaggtttgtgcatggc
tttctaatttttgttttgtaaaaanngtgcctactccaaacacgtaccg

Bgl II
/

189 caagtactcggaaggcagagatcagatctcgagttcaaggccagccttat
gttcatgagccttccgtctctagctctagagctcaagttccgggtcggaata

239 ctacagagtgagtttcaggacattcagggccacagtgaaatcctgtctc
gatgtctcactcaaagtcctgtaagttccagggtgtcactttaggacagag

289 aaaaagacaaaaacaaaaacaaaacaagattctattgttatgtttattat
ttttctgtttttgttttgttttgttcttaagataacaatacaaaataaata

339 tttattttattttttgtgtggctatttgcctgcatggatgcatgtgcacc
aaataaaataaaaaacacaccgataaacggacgtacctacgtacacgtgg

389 acctgtgtgcctagtggtccagaaaaatgccactggacatggagttatacat
tggacacacggatcacaggctcttttacggtgacctgtacctcaatatgta

439 ggtggtgggaattaacaatgggccccctctctacaagagtaagcatttta
ccaccacccttaatttgtaccogggggagagatgttctcatctgtaaaat

 ~~~~~~

489 tattatttttttagttttattttattttatgatgtaagtacactgat  
ataataaaaaaatcaaaaataaaataaaataactacattcatgtgacta  
 ~~~~~~

Bgl II

539 ctgctcttcagacacaccacaagagggcatcagatctcattaaggatggt
 -----+-----+-----+-----+-----+
gacgagaagtctgtgtggtgttctcccgtagtctagagtaattcctacca

589 tgtgagccaccatgtggttgctgggaattgaactcagaaccttcggaaga
 -----+-----+-----+-----+-----+
acactcgggtggtacaccaacgacccttaacttgagtcttggaaaccttct

639 gcagtcagtactcttaaccgctgagccatctctccagcccaagagtattc
 -----+-----+-----+-----+-----+
cgtcagtcattgagaattggcgactcggtagagaggtcgggttctcataag
 B2 *****
 ←----- Bgl II

689 tctctagcccagagagaaaatttgagaaatagtttgaggtaacacaagatc
 -----+-----+-----+-----+-----+
agagatcgggtctctctttaaactctttatcaaactccatgtgttctag

739 ttgtctccacaaaaacaaatgaaacatgaaaaatttctgcacatttagaaa
 -----+-----+-----+-----+-----+
aacagaggtgttttgtttactttgtactttttaagacgtgtaaatcttt

789 ataatttagnnnctgggtggtggtggtgcacatctttaatcccagcactg
 -----+-----+-----+-----+-----+
tattaaatcnnngacccaccaccaccacgtgtagaaattagggtcgtgac

839 ggaggcagaggcaggcagatttctgagttgacggccagcctggtacaga
 -----+-----+-----+-----+-----+
cctcogtctcogtccgtctaaagactcaaactgocggtcggaccatgtct

889 gatgagtcaggacagccagggtacacagagaaagctctgtctcaaaca
 -----+-----+-----+-----+-----+
ctactcaggtcctgtcgggtcccgatgtgtctcttctgagacagagtttgt

939 aaacaaaacaaaacaaaacaaaacaaaacaaaacaaaacaagaaagaaag
 -----+-----+-----+-----+-----+
tttgttttgttttgttttgttttgttttgttttgttttcttcttct

989 gaaggaaggaaggaaggaaggaaggaaggaaggaaggaagaaatagctgcag
 -----+-----+-----+-----+-----+
cttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct

1039 cgtatctcagggtggttgagttatgaaagcctgggtcatgtaattgttaat
+-----+-----+-----+-----+
gcatagaggtccaccaactcaatactttcggacccagttacattaacaatta

1089 ccagcactcaagaagatagagagagggcaggagggtcaaattcaagtcatt
+-----+-----+-----+-----+
ggtcgtgagttcttctatctctctcogtctcccagtttaagttcagtaa

Bgl II

1139 tctgctgccctgggtgagtttgaagcatccttgggattcatgagatcttat
+-----+-----+-----+-----+
agacgacgggaccactcaaacttcgtaggaaccctaagttactctagaata

1189 ttactcgtgtgtgtgtttgtgtgtgtctgtatgtatgtgtgtacatggat
+-----+-----+-----+-----+
aatgagcacacacacaaacacacacagacatacacacacatgtaccta

1239 gtctgtgtgtgtatctgactgtgtatctatgtgtttgtgtatacatatat
+-----+-----+-----+-----+
cagacacacacatagatagacacatagatacacaaacacatatgtatata

1289 gtctacatgtgtctgtgtgtctgtgtgtttatctgtgtgcccaggact
+-----+-----+-----+-----+
cagatgtacacagacacacagacacacacaaaatagacacacgggttctga

1339 taacatgcaaccctgactgacctggagctcactatgaggaccaggccggc
+-----+-----+-----+-----+
attgtacgttgggactgactggacctcgagtgatactcctgggtccggccg

1389 cactgtgaagatcagcctgccattgtgctgagattcggcgtgcaggacca
+-----+-----+-----+-----+
gtgacacttctagtcggacggtaacacgactctaagccgcacgtcctggt

1439 cccctggccataaaacatttatt
+-----+-----+
ggggaccggtat tttgtaaataa

Figure 5.5 (B) : Sequence of Section (B)

*Further sequencing of the murine sequences present in clone 8. The position of the sequenced DNA is shown in figure 5.5. The region is almost 1 kb in length. A FASTA search of the EMBL data base with these sequences found only one region that displayed significant homology to any other sequences. That region has high homology with a mouse B1 repetitive element. This repetitive element contains an A rich region (indicated by ~) and is flanked by direct repeats (indicated by *). Overlined with dashed lines is a 38 bp CA perfect dinucleotide repeat (proposed to be capable of forming Z DNA under physiological conditions). Regions overlined with heavy dark lines are GT boxes.*

(B)

→ from bluescript primer

```

5' tagaactgcaagggcgaccacatgacaaactctcaaccttatcagcatt
1  -----+-----+-----+-----+-----+
3' atcttgacggttccccgctggtgtactggttgagagttggaatagtcgtaa

                                *****
51  cagtgatatcaattacaagaatggagaactggagccgagcatggtggcgc
    -----+-----+-----+-----+-----+
    gtcactatagttaatgttcttacctcttgacctggctogtaccaccgcg
                                *****

atgctttaatcccagcactcgggatgcagagggcagggtggatttctgagct
101 -----+-----+-----+-----+-----+
tacgaaattagggctcgtgagccctacgtctccgtccacctaagactcga
    B1 →

tgaggccagcctggtgctacaaagtgagttccaggacagccagggctata
151 -----+-----+-----+-----+-----+
actccggtcggaccacgatgtttcactcaaggctcctgtcgggtcccgata

                                ~~~~~*****
201  cagagaaaccctgtcttgaaaacccccccccccccccccccccccagaatgga
    -----+-----+-----+-----+-----+
    gtctctttgggacagaaacttttggtttttttttttttttttttcttacct
                                ~~~~~*****
*****
251  gaactgccattgtctgctattggtctctcacaatatcttaatccaatac
    -----+-----+-----+-----+-----+
    cttgacggtaacagacgataacaagagagtttatatagaattaggttatg
    *****

301  tgggaagnatcataaaacaggccccttttaatggattactagaggcagat
    -----+-----+-----+-----+-----+
    acccttcntagtattttgctcggggaaaattacctaatagatctccgtcta

351  ataatatgagtggttgaaacaactgccaatgtaataagagaagatccg
    -----+-----+-----+-----+-----+
    tattatactcacaactttgttgacggttacaattattctctcttaggc

401  ttgtctacagtacttttaatgataaggttctgttctaggacaaagtcaat
    -----+-----+-----+-----+-----+
    aacagatgtcatgaaaattactattccaagacaagatcctgtttcagtta

451  tcctcattgtgatattatgtatccacccaaatggaagcaatggaagtt
    -----+-----+-----+-----+-----+
    aggagtaaacactataataacataggtgggtttaacctcgttaccttcaa

501  tccatcaataaagaaattcactgtaatctgacgctttctgctgtgatgta
    -----+-----+-----+-----+-----+
    aggtagtatttctttaagtacattagactgcgaaagacgacactacat

551  ttgcagcaatgagaatgtaattaacaaaactggagcaacatggaaaaca
    -----+-----+-----+-----+-----+
    aacgctcgttactcttacaataaattggttttgacctcgttgtacctttgt

```


Figure 5.5(C) : Sequence of Section (C)

This region is the 3' most murine DNA of clone 8. A FASTA search of the EMBL data base with these sequences found no significant homologies with other sequences in the data base.

(C)

5' cagttatgagttotgccatgtgggtggatggatttgatttaatccaggta
1 -----+-----+-----+-----+-----+
3' gtcaatactcagacggtacacccacctacctaaactaaattaggtccagt

tctggaagtcggttgggtgcttttaactgctgagccatctctccagtcctcc
51 -----+-----+-----+-----+-----+
agaccttcagccaaccacgaaaattgacgactcggtagagaggtcagagg

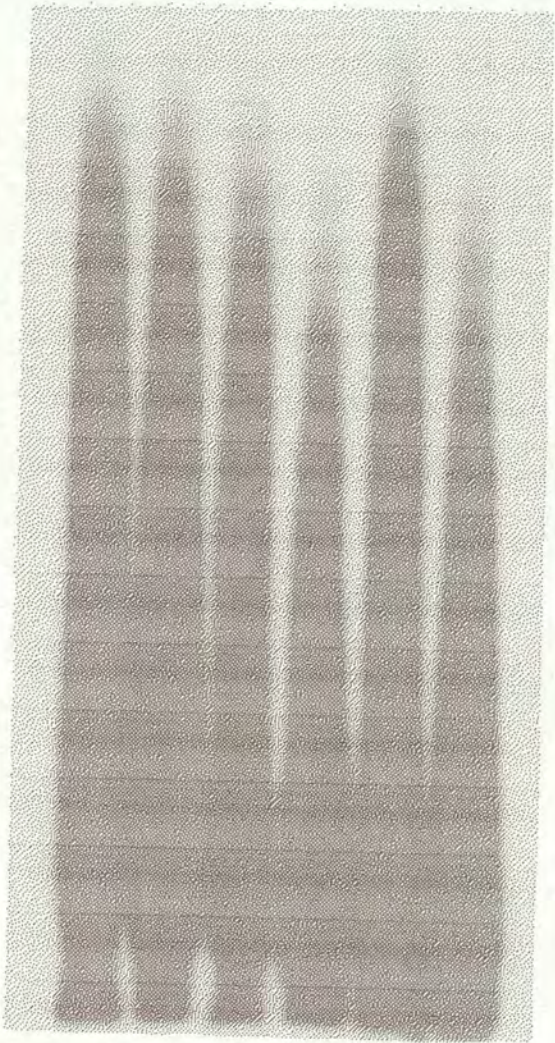
actgtcacctttttattttattttacattttggtatatacatatacataataca
101 -----+-----+-----+-----+-----+
tgacagtggaataaataaataatgtaaaacaatatatgtatgtattatgt

cacatacacatatatacatataactatcacctttaatgttattgcat
151 -----+-----+-----+-----+-----+
gtgtatgtgtatatatgtatgtatattgatagtggaattacaataacgta

← from bluescript primer

Figure 5.6 : Probing Genomic DNA With Murine Sequences Of Clone 8

The Southern blot shows genomic DNA of mouse of line AATB 46.2 (A) and a non transgenic sibling (C) digested with three different endonucleases. The Southern was probed with the 1.6 kb Hind III/Eco RV fragment of the murine sequences of Clone 8 in order to obtain information about the pre-integration site of line AATB 46.2. The resultant smear obtained in all lanes with this probe demonstrated the presence of repetitive DNA in the probe sequences. Subsequent sequencing of this fragment demonstrated the presence of mouse repetitive elements within the probe sequences (see figure 5.5 (A)). No details of the pre-integration site were obtained from this blot.



EcoRV/A
EcoRV/C
NheI/A
NheI/C
HindIII/A
HindIII/C

5.4.2 Rationale For Use of a Phage Vector to Construct the Secondary Transgene

Constructing the secondary transgene in plasmids proved to be impossible because of the inherent instability of cosmid clone 8. Restriction mapping and sequencing of part of this clone demonstrated the clone harboured two origins of replication and two ampicillin resistance genes both originating from the cosmid vector. The clone recombined to produce a small cosmid fragment which was considerably more abundant in DNA preparations than the original clone itself (see Chapter 4, Section 4.2.4). Numerous attempts to sub-clone fragments from clone 8 using plasmids failed.

This was the only clone isolated from the cosmid library that harboured the 3' flanking murine DNA of AATB 46.2. If the secondary transgene was to contain the 3' murine sequences they had to be recovered from this clone. To circumvent the difficulties experienced using plasmids, a phage vector was employed to construct the secondary transgene.

The secondary transgene encompassing the 4.6 kb 5' murine DNA, AATB (10.6 kb) and the 4.7 kb 3' murine flanking DNA, is 19.9 kb in size. This is within the size limit for the insert of many phage vectors, whereas, the size of the recombinant that caused the difficulty in the plasmid sub-cloning experiments was 4.6 kb, and thus too small to be packaged (once ligated into a phage vector). Therefore using a phage vector allowed positive selection for the secondary transgene on the basis of size. Details of the construction and cloning of the secondary transgene are given in the next section.

Figure 5.7(A) ; The Large Not I Fragments From Clone 8 and 9 Are Suitable For Constructing The Secondary Transgene.

Both clone 8 and clone 9 liberate fragments of approximately 15.2 kb on Not I digestion. The figure shows clones 8 and 9 and in particular the position of the Srf I and Not I restriction sites. (F1) = The Not I fragment spanning the 5' AATB 46.2 chromosomal/transgene junction. It includes all the 5' cloned murine DNA of the locus (4.6 kb) and an intact AATB transgene. (F2) = The Not I fragment spanning the 3' AATB 46.2 chromosomal/transgene junction. It contains all of the 3' cloned murine DNA of the AATB 46.2 locus (4.7 kb) and most of the AATB sequences (see chapter 4, analysis of clone 8). Each of the 15.2 kb Not I fragments harbours only one Srf I restriction site situated in the 3' end of the ovine BLG sequences of AATB. Fragments F1 and F2 were isolated by gel electrophoresis in 0.3% agarose gels and used in the construction of the secondary TAB transgene.

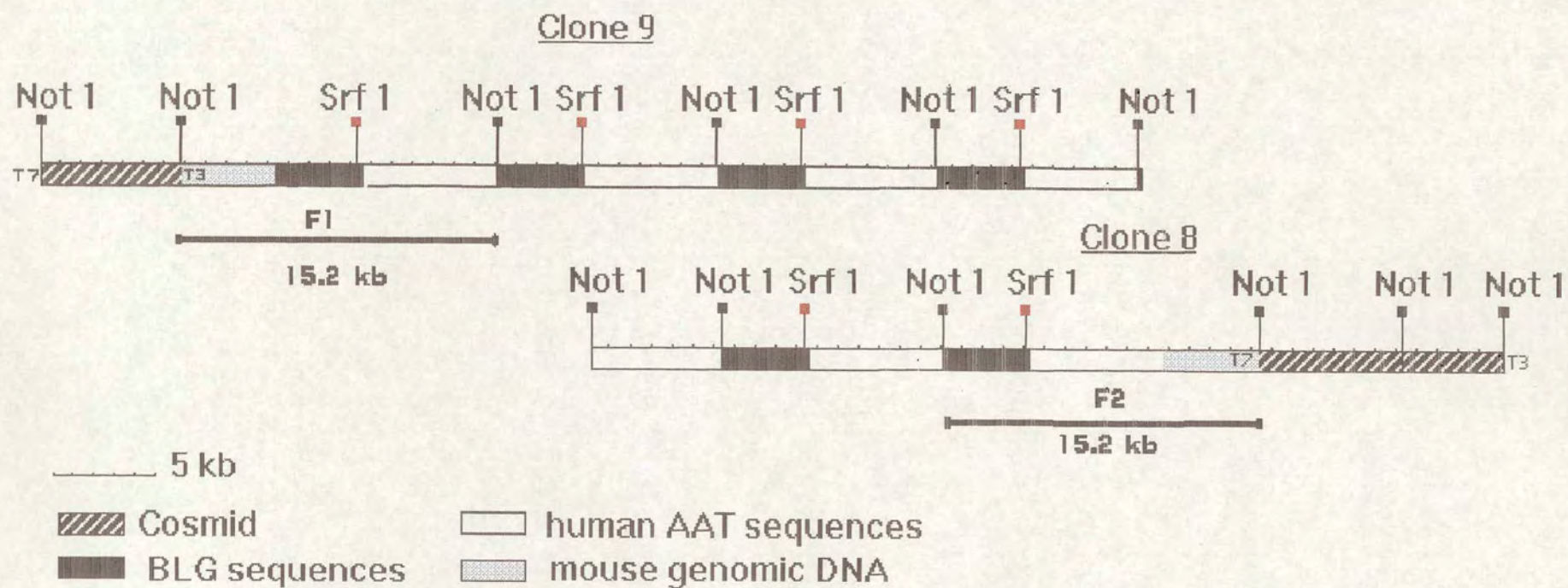


Figure 5.7 (B) : Joining The Two Not I Srf I Fragments Gives An AATB Transgene Flanked By The Chromosomal DNA Of AATB 46.2

The Not I fragments shown in Figure 5.7 (A) were digested with Srf I and the 8.8 kb fragment from clone 9 and the 11 kb fragment from clone 8 were gel purified. The yield from each round of digestion and gel purification was low so this procedure was performed several times and the purified fragments pooled. The two fragments were ligated in the presence of Not I restriction enzyme: to reverse the favourable ligation of the Not I complementary ends. The ligation product (19.9 kb) was separated from the input fragments (8.8 kb and 11 kb) and less abundant products of the ligation reaction (17.6 kb and 22 kb) by gel electrophoresis in 0.3% agarose gels, purified and then ligated into the Not I sites of the lambda DASH II phage vector (Stratagene : figure 5.7 C).

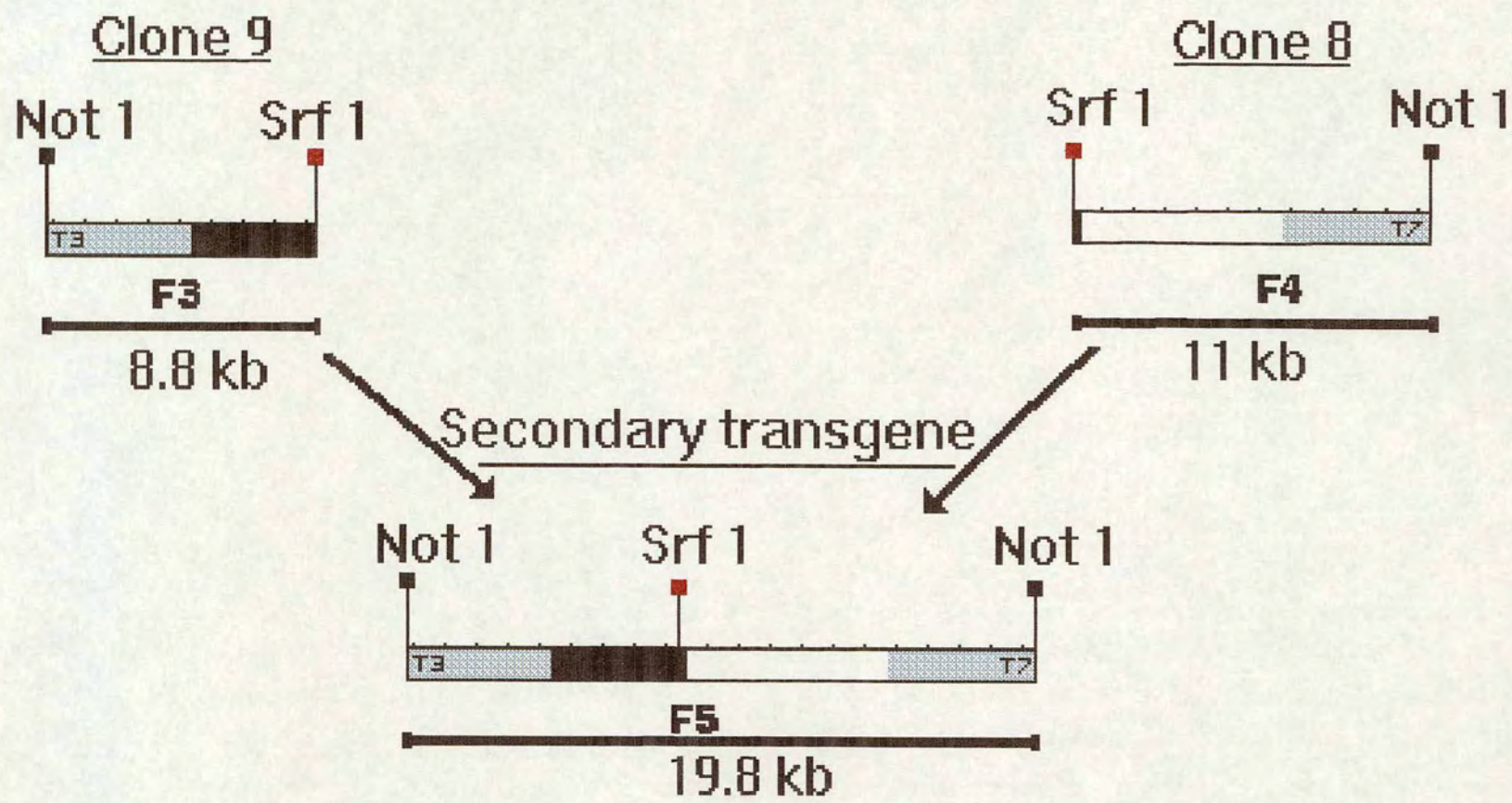
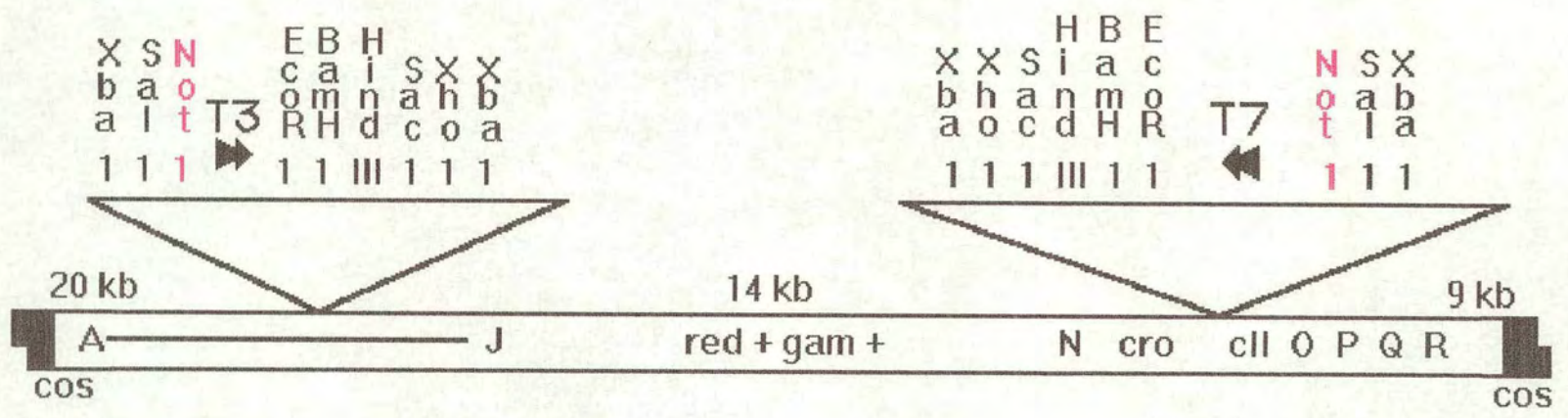


Figure 5.7 (C) : The Lambda DASH II Phage Cloning Vector

The figure is a diagrammatic representation of the Lambda DASH[®] II phage cloning vector (Stratagene). The vector accommodates an insert of 9kb to 23 kb. The Not I sites of the vector (shown in red) were used to ligate the insert into the vector. This removes the T3 and T7 promoters of the vector. The T3 and T7 promoters from the Super-cos-one vector are part of the secondary transgene (see figure 5.7 (B)).



5.4 3 Construction of the Secondary Transgene

Chromosomal DNA flanking the AATB 46.2 locus was cloned in two separate cosmid vectors. Clone 8 harboured the 3' end of the locus and clone 9 the 5' end of the locus. The starting point for construction of the secondary transgene was to isolate a fragment containing the murine flanking DNA and an AATB transgene, from each clone. These fragments were obtained by digesting clones 8 and 9 with Not I. The 15.2 kb fragment from clone 8 harboured the 3' most AATB transgene of AATB 46.2 and all of the cloned 3' flanking murine DNA (F2, figure 5.7 A). Similarly the 15.2 kb Not I fragment of Clone 9 harboured the 5' most AATB transgene of the array and all of the cloned 5' flanking murine DNA (F1 figure 5.7 A).

Each Not I fragment, F1 and F2, contained a unique Srf I restriction site within the BLG sequences of AATB (see figure 5.7 A). Thus F1 and F2 each generated two fragments when digested with Srf I. F1 gave fragments of 8.8 kb (F3, figure 5.7 B) and 6.4 kb, and F2 gave fragments of 11.1 kb (F4 figure 5.7 B) and 4.1 kb. Ligation of F3 and F4 at the Srf I site generates a single AATB transgene flanked 5' and 3' with the appropriate flanking murine DNA of the AATB 46.2 locus.

Fragments F3 and F4 were gel purified before being ligated to produce the secondary transgene (see figure 5.7 B). The presence of the Not I complementary ends 5' of F3 and 3' of F4 required the ligation reaction to be carried out in the presence of Not I restriction enzyme. This was to reverse Not I complementary end ligations and thus drive the ligation reaction in the direction of the less favourable blunt end ligation at the Srf I site.

As expected, ligation of the two fragments generated a number of products (verified by visualisation of the reaction mix on an 0.3% agarose gel stained with ethidium bromide, data not shown). The predominant product was approximately 20 kb in size. This product was assumed to be the desired secondary transgene (F5, figure 5.7 B). Other less abundant products both smaller and larger were present. These products were likely to be the products of ligation occurring between like fragments (i.e. F3 + F3 and F4 + F4). The 19.9 kb product of the ligation was gel purified and ligated into the lambda DASH II cloning vector.

5.4.4 : Phage Cloning of the Secondary Transgene

Lambda DASH II was the cloning vector used to clone the secondary transgene (figure 5.7 C). The vector was prepared for ligation of F5 (see figure 5.7 B) by digestion with Not I (see figure 5.7 C). The Not I sites of the vector were outwith the multiple cloning site and digestion of this vector with Not I resulted in the loss of the T3 and T7 promoters. This was essential as the secondary transgene TAB (F5, figure 5.7 B) is flanked 5' by a T3 promoter and 3' by a T7 promoter. These promoters at either end of the secondary transgene originate from the Super-Cos-One cosmid vector in which the locus of AATB 46.2 was cloned.

A total of 33 plaques were obtained from the phage cloning. Each of the phage clones were screened with probes from the 5' and 3' end of the AATB transgene. All but one of the phage clones hybridised with both probes indicating that sequences from both F3 and F4 were present in the clones.

Two secondary transgenes were isolated from the cloning procedure. The intended secondary transgene (TAB): encompassing 5' murine sequences, AATB, and

3' murine sequences, and another transgene (TA) that contained only AATB and 3' murine sequences. This transgene is the same as the 15.2 kb Not I fragment of clone 8 (F2, figure 5.7 A). It is probable that cloning of this transgene arose by carry over of the non digested F2 through all stages of construction of the secondary transgene.

The size of the TAB transgene is similar to the left arm of the phage vector when digested with Not I. To facilitate the separation of the TAB secondary transgene from the vector sequences the clone was digested with both Not I and Nru I (which cuts at position 4592 of lambda dash II). Both secondary transgenes (TA and TAB) were used to generate transgenic animals. The analysis of these transgenic mice is detailed in chapter 6 of this thesis.

CHAPTER 6 : EXPRESSION OF THE SECONDARY TRANSGENES IN THE MAMMARY GLAND OF TRANSGENIC MICE

6.1 INTRODUCTION

The previous chapters outline the genomic mapping of the high expressing AATB 46.2 locus, the cloning of the locus and, the construction of the TAB and TA transgenes. This chapter describes the generation of transgenic mice carrying the secondary transgenes and the analysis of their expression.

As with the AATB studies, the secondary transgenic animals were generated by pro-nuclear injection. of mouse eggs collected from F1 (C57BL/6 x CBA) females, after mating with F1 stud males (Simons et al., 1987). The number of TAB transgenic mice generated was comparable with the previous AATB studies enabling a direct comparison to be made between the AATB and TAB transgenes.

The analysis of the transgenic mice involved the determination of the transgene copy number for each animal and measurement of hAAT expression levels in G₀ and G₁ animals. The copy number was estimated by comparison with plasmid copy number controls. The protein composition of the transgenic mouse milk was visualised on Coomassie stained SDS PAGE gels. These gels showed the presence of an additional protein in transgenic mouse milk that was not present in the non transgenic control. The identity of the additional protein, as hAAT, was verified by immunoblotting. The amount of hAAT present in the milk was quantified by radial immunodiffusion (RID).

Comparison of the TAB transgenic experiments with the previous AATB experiments (Archibald et al., 1990 and Carver et al., 1993) showed that both the frequency of expression of hAAT in the mammary gland of transgenic mice and the levels of expression were much higher in the TAB transgenic mice. Transgenic mice harbouring the TA transgene did not detectably express hAAT. The conclusion of the secondary transgenic experiments is that the cloned flanking murine sequences from the AATB 46.2 locus, that are present in the TAB transgene markedly improved the expression of the AATB transgene and are probably causative in the high expression of hAAT from the transgene array of AATB 46.2.

6.2 EXPRESSION OF AATB IN SECONDARY TRANSGENIC MICE

6.2.1 Generation of Transgenic Mice

In all twelve transgenic animals/lines were generated harbouring the TAB transgene (Table 6.1). Six transgenic mice/lines were generated at the Roslin Institute. All six founders were female and hAAT expression data was obtained from them. Another set of six TAB transgenic mice were generated at PPL Therapeutics Ltd. In this instance two of the founder mice were female and the others were male. The levels of hAAT in the milk of the two female founder mice were measured, and the males were used to establish lines.

Of the six mice produced at the Roslin Institute five transmitted the transgene to their offspring. One founder failed to transmit, even after extensive breeding. This mouse TAB 80 was presumed not to have the transgene in its germline. Two of the other Go mice, TAB 14 and TAB 52, transmitted the transgene to their offspring at a

frequency considerably less than 50% suggesting that these founders were also germline mosaics. TAB 29, TAB36, and TAB 54 all transmitted the transgene at a frequency very close to 50 %.

Of the six mice generated at PPL Therapeutics Ltd. only one failed to transmit the transgene. This founder TAB 108.22 was male and so no expression data was obtained. Four of the other five founders transmitted the transgene at a frequency close to 50 % suggesting the founder mice were not mosaic. However, subsequent copy number determination of Go mice and their offspring suggested that at least one of the six founders was mosaic. The transmission frequency of TAB 111.4 is not known since this mouse was not bred to produce a line.

Expression data was obtained from the eight female Go animals and from the offspring of five of these animals. The remaining expression data was obtained from the female transgenic offspring of the three male founders that transmitted the transgene. Thus, overall, hAAT expression was analysed in eight independently generated lines plus three Go animals. In two of the lines the transgene copy number rose in the first generation. The copy number of TAB 113-12-2 is double that of the founder male TAB 113-12 (see table 6.3). In TAB 14 a much more dramatic eight fold rise in copy number was seen between the Go and G₁ animals (see table 6.2), however the transgene copy number was stable between G₁ and G₂ animals (data not shown). A small rise in copy number is also apparent between TAB 110.28 and its offspring. These three founder animals were presumed to be mosaic, as were the two Go animals that did not transmit the transgene. The percentage of mosaic Go animals within this study is 42%. This is slightly greater the figure of 33% obtained by Wilkie et al (1986). It is possible that the analysis of Go transgenic animals may not give an

entirely accurate picture of transgene expression because an undetermined percentage of the cells in a given tissue may not carry the transgene. Expression data of female transgenic offspring is free from such ambiguity, since the transgene is present in all somatic cells.

TABLE 6 1 : TAB Transgenic Mice

Go Animal (Sex)	Transgene Transmitted	Offspring analysed
14 (F)	1/36	√
29 (F)	5/9	unable to lactate
36 (F)	11/20	√
52 (F)	2/22	√
54 (F)	2/4	√
80 (F)	0/32	no offspring
108-16 (F)	√	√
108-22 (M)	0/6	no offspring
110-14 (M)	√	√
110-28 (M)	√	√
111-4 (F)	nb	no offspring
113-12 (M)	√	√

Legend to Table 6.1

(M)=male, (F)=female, nb= not bred on

6.2.2 Estimation of Copy Numbers

The method used for the estimation of transgene copy numbers was the same as in the previous AATB studies. The intensity of internal fragments from the

transgene is estimated relative to plasmid copy number controls. In this study the determination of copy number is simplified by the ability to quantify the signals from Southern blots using a Molecular Dynamics Phosphor Imager. Loading across the lanes was controlled for, by using a β -casein probe (kindly donated by S. George).

The two sets of TAB transgenic mice were produced at different sites and several months apart, for this reason the copy numbers in the two sets of animals have been analysed separately. In each case the same plasmid controls and non transgenic DNA have been used so that a comparison can be made between the two sets of copy number results. Figures 6.2 and 6.3 show the Southern blots from which the transgene copy numbers were estimated. Tables 6.2 and 6.3 show the data obtained from the PhosphorImager and the estimated copy number of the TAB transgenic mice. The transgene copy numbers in this study ranged from 5 copies (TAB 54) to greater than 120 (TAB 14.20).

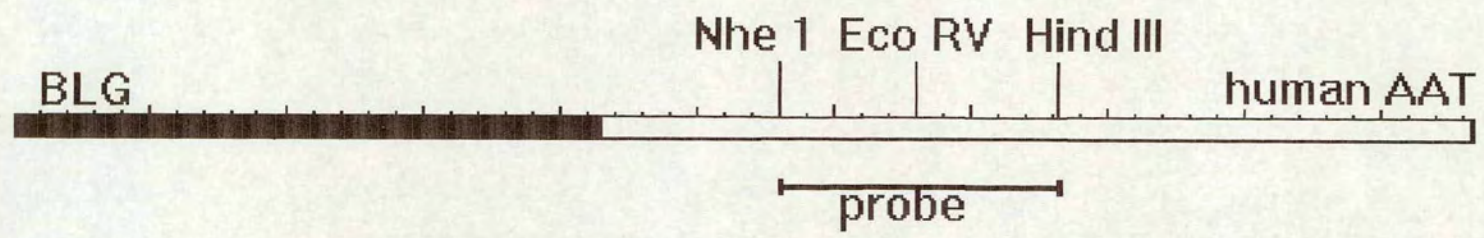
The copy number estimates were performed on tail DNA. This was not ideal as it tends to be more degraded than DNA prepared from soft tissues such as liver. However liver DNA could not be obtained from these animals as they were required for breeding to establish transgenic lines for the milk analysis. The DNA was digested with Eco RV generating the two expected internal fragments of 7.2 kb and 5.6 kb. The signal from both fragments was measured using the PhosphorImager and compared with the 5 copy plasmid control that was also digested with Eco RV. In figure 6.2 the plasmid control was loaded in 10 μ g of non transgenic liver DNA. In figure 6.3 the plasmid control was loaded in 28 μ g of the same non transgenic DNA. Using a β -casein probe (kindly donated by S. George) the loading of each of the lanes

was normalised using the signal from the non transgenic DNA as the standard (see tables 6.2 and 6.3).

Figure 6.1 : Probe Used to Estimate the Copy Number of TAB Transgenic Mice

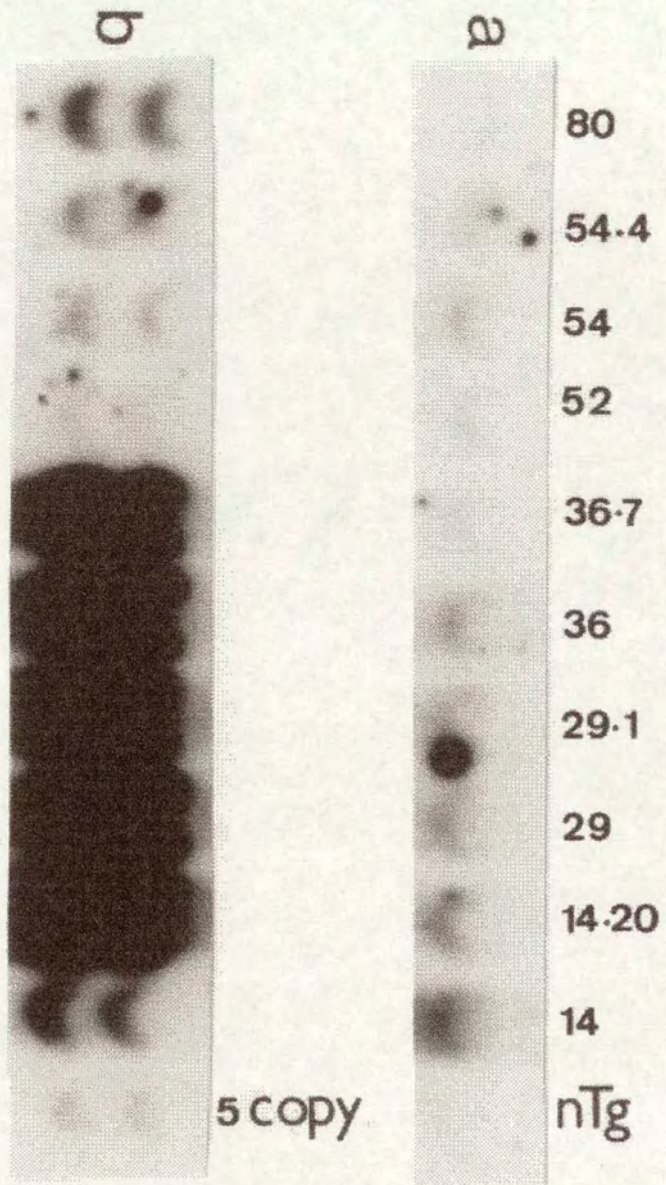
The probe is the Nhe I/Hind III fragment from AATB used for the estimation of copy number of the TAB transgenic mice all of the DNA for copy number estimation was digested with Eco RV. This probe spanned the Eco RV site and detected the two internal Eco RV fragments generated from the TAB transgene.

AATB



**Figure 6.2 : Southern Blot for Copy Number Estimation of Roslin Institute
Generated TAB Transgenic Mice**

The figure shows a Southern blot used for the estimation of the copy number of the TAB transgenic mice generated at the Roslin Institute. The plasmid copy control is a five copy equivalent of the TAB transgene loaded in 10 µg of liver DNA from a non transgenic mouse, digested with Eco RV. Each of the other lanes contains 10 µg of DNA, prepared from tail cuts, digested with Eco RV. (a) Shows the signal obtained with a β-casein probe (donated by S. George). The filter was laid down for 18 hours, at -70 °C, with intensifying screens. The signal was quantified by 16 hour exposure to the PhosphorImager (Molecular Dynamics). The data from the PhosphorImager is shown in Table 2. (b) shows the signal using the AATB probe in figure 6.1. As expected the probe hybridises with two fragments of 7.2 and 5.6 kb. The filter was laid down with X-ray film for 18 hours, at -70 °C, with intensifying screens. The signal from this probe was also quantified by exposure to the PhosphorImager for 3 hours (see table 2)



	β -casein probe (a)	Control/ β -casein probe (b)	AATB probe (c)	Control/ AATB probe (d)	Copy number (d) x 5 (e)	corrected copy number (e) x (b)
control 5 copy	16860	1	44572	1		
TAB 14 (f)	18875	0.89	154594	3.47	17.35	15
TAB 14.20 (f)	36878	0.46	2379203	53.37	266.85	123
TAB 29 (f)	43784	0.38	1698799	38.11	190.55	72
TAB 29.1 (f)	44845	0.37	1428676	32.05	160.25	59
TAB 36 (f)	65210	0.25	1062340	23.83	119.15	30
TAB 36.7 (f)	34755	0.48	872099	19.56	97.8	47
TAB 52 (f)	21427	0.78	58057	1.30	6.5	5
TAB 54 (f)	19879	0.84	59437	1.33	6.65	6
TAB 54.4 (f)	28119	0.59	82895	1.86	9.3	5
TAB 80 (f)	27341	0.61	98037	2.19	10.95	7

Table 6.2

Copy number data from the Phosphorimager for Roslin generated TAB mice

Figure 6.3 ; Southern Blot for Copy Number Estimation of PPL Generated TAB Transgenic Mice

The figure shows a Southern blot used for the estimation of the copy number of the TAB transgenic mice generated at PPL Therapeutics Ltd.. The plasmid copy control is a five copy equivalent of the TAB transgene loaded in 28µg of liver DNA from a non transgenic mouse, digested with Eco RV. Each of the other lanes contains 10 µg of DNA, prepared from tail cuts, digested with Eco RV. (a) Shows the signal obtained with a β-casein probe (donated by Sisilamma George). The filter was laid down for 18 hours, at -70 °C, with intensifying screens. The signal was quantified by 16 hour exposure to the PhosphorImager (Molecular Dynamics). The data from the PhosphorImager is shown in Table 2. (b) shows the signal using the AATB probe in figure 6.1. As expected the probe hybridises with two fragments of 7.2 and 5.6 kb. The filter was laid down with X-ray film for 18 hours, at -70 °C, with intensifying screens. The signal from this probe was also quantified by exposure to the PhosphorImager for 3 hours (see table 2)

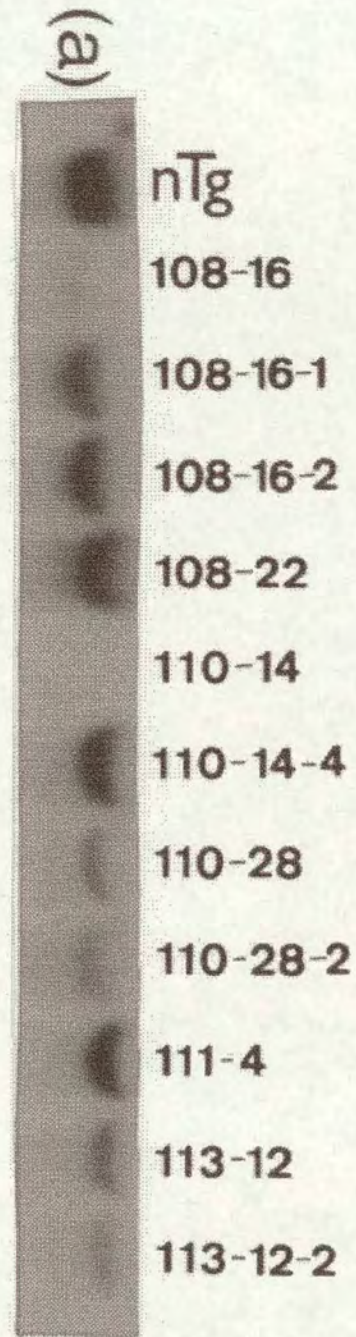


Table 6,3Copy number data from the Phosphorimager for PPL generated mice

	β -casein probe (a)	control/ β -casein probe (b)	AATB probe (c)	control/AATB probe (d)	copy number (d) x 5 (e)	corrected copy No. (e) x (b)
control 5 copy	16515 *	1	5832	1		
108.16 (f)	10835	1.52	7675	1.32	6.58	10
108.16.1 (f)	29072	0.57	25312	4.34	21.7	12
108.16.2 (f)	32982	0.50	31363	5.38	26.89	13
108.22 (m)	37520	0.44	63160	10.83	54.15	24
110.14 (m)	5963	2.78	6462	1.11	5.54	15
110.14.4 (f)	32064	0.51	21120	3.62	18.11	9
110.28 (m)	17962	0.92	46137	7.91	39.55	36
110.28.2 (f)	18139	0.91	63348	10.86	54.3	50
111-4 (f)	28432	0.58	33863	5.81	29.03	17
113.12 (m)	17884	0.92	9211	1.57	7.9	7
113.12.2 (f)	11906	1.38	10876	1.86	9.32	13

* signal normalised to 10 μ g of control DNA (19 μ l of control DNA loaded, The signal was divided by 19 and then multiplied by 6.8 since 6.8 μ l of control DNA is believed to 10 μ g of DNA and is the volume previously loaded to determine the copy numbers of the Roslin mice)

6.2.3 Expression of hAAT in TAB Transgenic Mice

The milk samples of TAB animals were analysed initially by SDS PAGE. This technique allows the separation of the protein components of milk according to their size. Subsequent staining of the gel with Coomassie blue allows visualisation of the protein composition of the milk samples. The technique allows direct visualisation of the hAAT protein in the transgenic mouse milk provided the levels of expression are sufficiently high. The protocol used was adapted from Archibald et al., (1990). The SDS PAGE results are shown in Figures 6.4 and 6.5.

From this analysis of the TAB milk samples the majority of the transgenic mice were shown to express the human protein at high levels. The level of expression between the Go animals and their offspring was generally stable. The clear exception was TAB 14 and TAB 14.20.6. It was plain that the expression level in the G2 animal far exceeded that of the Go. Interestingly this large elevation in hAAT expression correlated with a concomitant rise in the transgene copy number between the founder and her transgenic offspring (see table 6.2).

The stained gels allow direct visualisation of a protein which is present in transgenic mouse milk but absent from non transgenic controls. The protein had the same apparent size as hAAT (54 kD). To verify that it was hAAT a Western blot was performed and probed with goat anti-human AAT serum and anti sheep/goat IgG serum conjugated to horseradish peroxidase. Only one band on the gel reacted with the antibody and this was the same size as the hAAT standard (figure 6.6). No antibody reaction was observed in the non transgenic milk samples. This contrasts with the results of Archibald et al., where cross reaction of the antibody with control mouse milk was observed (assumed to be the mouse endogenous AAT). The Western

blot performed in this study was less sensitive than that performed by Archibald et al., and was probably the reason for the observed difference.

The hAAT in the transgenic mouse milk did not migrate as a tight band. There appeared to be several bands corresponding to hAAT which migrate closely in 10% SDS PAGE gels. AAT produced in human liver has three carbohydrate side chains. The various bands visualised on the Coomassie stained gels probably correspond to differences in the post translational modifications of hAAT produced in the mammary gland. This electrophoretic heterogeneity of the hAAT in transgenic mouse milk was also observed by Archibald et al.. The Western blot is shown in Figure 6.5

A quantitative analysis of the hAAT level in the mouse milk was performed by radial immunodiffusion. The minimum level of hAAT detectable using this method was 40µg/ml. This method in conjunction with radio immunoassay was used to determine hAAT levels in the first AATB study (Archibald et al., 1990) whereas, a much more sensitive ELISA was used in the second AATB study (Carver et al., 1993). The levels of hAAT in the TAB transgenic mouse milks are tabulated in Table 6.4.

Figure 6.4 : SDS PAGE of Roslin Institute TAB Transgenic Mouse Milks

The figure shows a 10 % SDS Poly Acrylamide gel following the electrophoresis of milk samples from a non transgenic mouse and the TAB transgenic mice generated at the Roslin Institute. Lane 2 contains 2.5 μ g of human AAT standard (Sigma). Lane 4 contains the milk sample of AATB 46.2 (the mouse from which the murine sequences in the TAB transgene were cloned). AATB 46.2 is known to express hAAT at 6.25 mg/ml. All of the milk samples were defatted following a 1 in 5 dilution with sterile distilled H₂O. They were then further diluted 1 in five with reducing buffer to give a total dilution of 1 in 25. 5 μ l of each sample was loaded in the gel in 30 μ l of loading buffer. An additional protein (not present in non transgenic milk) is visible in almost all of the transgenic milk samples. This protein had the same apparent size as the hAAT standard (54 kD). The level of expression of the additional protein in TAB 14.20.6 is considerably higher than the level in the founder mouse TAB 14. The levels of expression of the extra protein in TAB 14.20.6 also exceed those of AATB 46.2. The level of hAAT in TAB 29 and TAB 36 also appear to be more than that of AATB 46.2.

Figure 6.5: SDS PAGE of PPL TAB Transgenic Mouse Milks

The figure shows a 10 % SDS Poly Acrylamide gel following the electrophoresis of milk samples from a non transgenic mouse and the TAB transgenic mice generated at PPL Therapeutics. Lane 2 contains 2.5 µg of human AAT standard (Sigma). Lane 4 contains the milk sample of AATB 46.2 (the mouse from which the murine sequences in the TAB transgene were cloned). AATB 46.2 is known to express hAAT at 6.25 mg/ml. All of the milk samples were defatted following a 1 in 5 dilution with sterile distilled H₂O. They were then further diluted 1 in five with reducing buffer to give a total dilution of 1 in 25. 5 µl of each sample was loaded in the gel in 30 µl of loading buffer. An additional protein (not present in non transgenic milk) is visible in almost all of the transgenic milk samples. This protein has the same apparent size as the hAAT standard (54 kD). All of the mice except line TAB 110.14 appear to express hAAT at a level greater than that of AATB 46.2. The level of expression of hAAT in line TAB 110.28 appears to be very high.

(b) Coomassie Stained SDS Gels Of TAB Milk Samples (PPL Mice)

LMW marker	human AAT	non Tg	AATB 46.2 G5	TAB 108.16 G0	TAB 108.16 G1	TAB 110.14 G1	TAB 108.16 G1	TAB 110.14 G1	TAB 110.28 G0	TAB 111.4 G0	TAB 113.12 G1
------------	-----------	--------	--------------	---------------	---------------	---------------	---------------	---------------	---------------	--------------	---------------

hAAT →

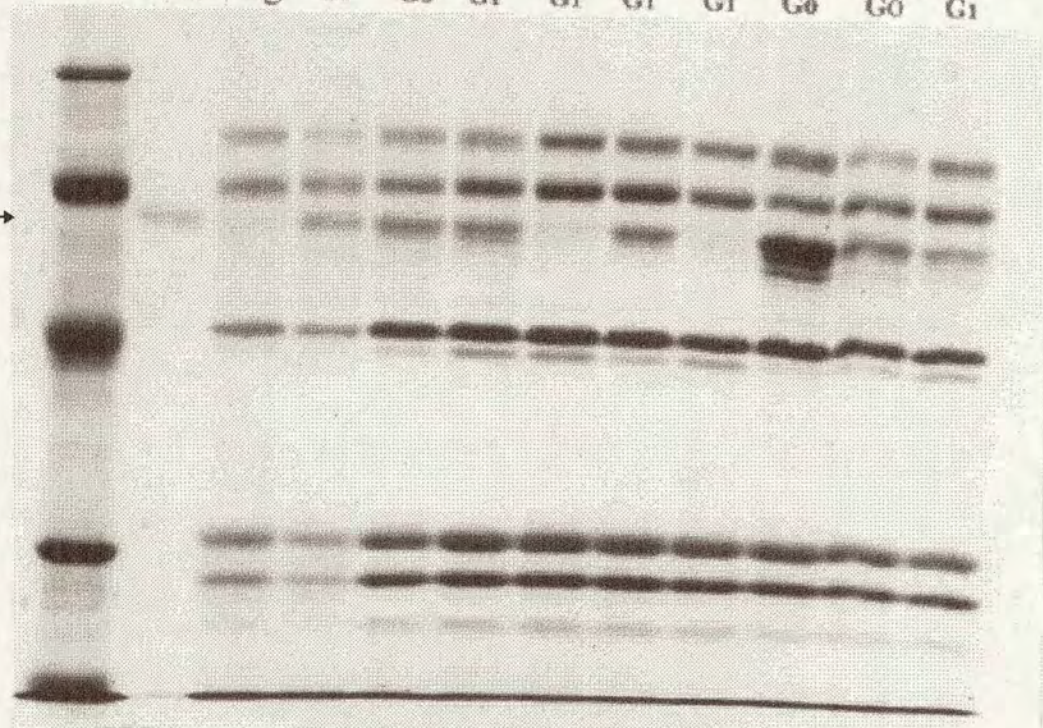


TABLE 6 4**Expression Of hAAT In TAB Transgenic Mice**

Animal or Line	Generation	Expression (mg/ml) (σ)	Standard deviation (mg/ml)
TAB 14	Go	3.48	± 0.65
TAB 14.20.6	G ₂	24.6	± 3.37
TAB 29	Go	14.16	± 0.48
TAB 36	Go	5.25	± 0.67
TAB 36.2	G ₁	7.57	± 0.73
TAB 52	Go	nd	—
TAB 52.1	G ₁	nd	—
TAB 54	Go	0.596	± 0.05
TAB 54.2	G ₁	0.553	± 0.02
TAB 80	Go	0.46	± 0.06
108-16	Go	7.95	± 0.4
108-16-1	G ₁	7.51	± 0.4
110-14-4	G ₁	0.21	± 0.01
110-28-2	G ₁	34.86	± 1.5
111-14	Go	13.5	± 1
113-12-2	G ₁	6.7	± 0.5

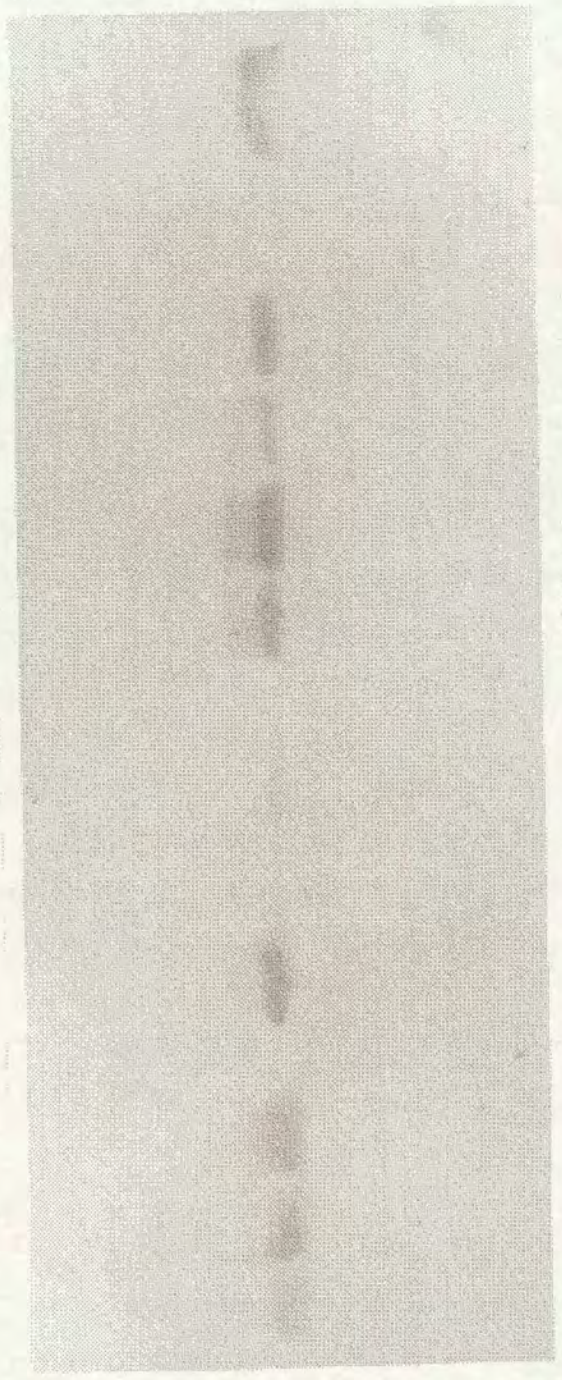
(σ) = average of three independent readings

nd = none detected (limit of detection 40 $\mu\text{g/ml}$)

Figure 6.6 : Western Blot of TAB Milk Samples

The figure shows a Western blot of a 10 % SDS polyacrylamide gel after electrophoresis of the TAB transgenic milk samples. Lane one contains 2.5 ug of hAAT standard (Sigma). Lane 2 contains a sample of non transgenic milk and Lane 3 contains a sample of milk from AATB 46.2. All other lanes contain TAB milk samples from the lines/mice indicated. The proteins were transferred to a nitrocellulose filter by electroblotting. The filter was probed using goat anti- α_1 AT serum [Protein Reference Unit (PRU), Royal Hallamshire Hospital, Sheffield, S10 2JF] and anti sheep/ goat IgG serum conjugated to horseradish peroxidase.

Only one protein hybridises with the antibody. This is the additional protein visualised on the coomassie stained SDS polyacrylamide gels in figures 6.4 and 6.5. The sensitivity of the blot was poor due to poor transfer of the proteins on electroblotting. Nevertheless this western blot verifies the 54 kD additional protein observed in TAB transgenic mouse milk is hAAT.



α_1 AT
nTg

AATB 46-2

TAB 14

29

36

52

54

80

108-16

110-14-4

110-28-2

111-4

113-12-2

Figure 6.7: Comparing hAAT Expression of TAB Transgenic Mice With AATB Transgenic Mice

The figure shows a plot of the expression levels of hAAT from the two previous studies of AATB mice (Archibald et al., 1990 and Carver et al., 1993) and this study of TAB mice. For each group of data the mice expressing hAAT at less than 40 $\mu\text{g/ml}$ are grouped together as are those that express between 40 $\mu\text{g/ml}$ and 1 mg/ml . The column on the left gives the expression level in mg/ml .

This figure demonstrates that the levels of hAAT expression in the TAB mice are on the whole greater than that observed in AATB mice. The numbers of TAB mice expressing hAAT at greater than 40 $\mu\text{g/ml}$ is much improved over the AATB transgenic mice. Only one of eleven TAB mice expressed hAAT at less 40 $\mu\text{g/ml}$, whereas half of the AATB mice from each of the previous studies were found to express hAAT below this level.

expression (mg/ml)	Archibald (AATB)	Carver (AATB)	Cranston (TAB)
35			*
30			
25			*
20			
15			* *
10	*	*	* * *
5		*	*
1			
40-1000 µg / ml	*****	**	***
< 40 µg / ml	*****	*****	*

Figure 6.6 shows the levels of hAAT expression in the AATB transgenic mice (Archibald et al.,1990 and Carver et al., 1993) and the TAB transgenic mice (this study). The hAAT expression levels in the TAB mice are generally high. Only one of the lines TAB 52 did not express the human protein. With the exception of this line the lowest level of expression was 0.2 mg/ml. By contrast more than half of the mice in both AATB studies expressed the transgene at less than 40µg/ml.

Seven of the eleven TAB transgenic founders/lines expressed the human protein at levels considerably greater than 1mg/ml whereas only one of thirteen of the Archibald AATB mice and two of eleven of the Carver AATB mice expressed hAAT at greater than 1 mg/ml. Thus the levels of hAAT expression from the AATB transgene appeared to have been much improved by the presence of the 5' and 3' murine genomic sequences from the AATB 46.2 locus.

6.2.4 Lactation Difficulties in TAB Transgenic Mice

It proved impossible to obtain milk samples from the offspring of TAB 29. The Go animal expressed the human protein at 14.16 mg/ml. Two female offspring were obtained and bred. One of the daughters TAB 29.4 destroyed two consecutive litters only hours after giving birth. Such behaviour is not particularly unusual but is considered important in this instance because the sibling TAB 29.5 was unable to let down her milk. The mammary gland of TAB 29.5 became engorged and she had to be destroyed because she was clearly in distress. Although a large volume of milk was present in the mammary gland, this animal was unable to feed her pups. The reasons for this are not clear. No milk sample was obtained from either of the G₁ animals, thus it is not possible to tell whether the levels of hAAT expression remained steady or

rose in first generation. The transgene copy number in this line remained constant from G₀ to G₁ thus it would seem unlikely that the level of hAAT expression in the milk has significantly increased. Yet the G₀ animal was able to lactate and feed her pups. No explanation can be given for the lactation difficulties of this line.

Interestingly TAB 14.20.6 also had difficulty feeding her pups. During lactation the G₂ animal had a dishevelled appearance and her pups were failed to thrive despite the litter being small. The milk sample obtained from this animal following oxytocin administration was smaller than usual.

Both TAB 29 and TAB 14 express the human protein in their milk at high levels (14.16 mg/ml and 24.6 mg/ml respectively). It is possible that high levels of expression of this foreign protein in the milk can impair lactation. Nevertheless the high expression of the human protein does not appear to have a major detrimental effect on the expression of other major mouse milk proteins (McClenaghan et al., 1990). The levels of the endogenous milk proteins appear to be consistent between the transgenic and non-transgenic animals. Moreover TAB 110-28-2 was able to express hAAT in her milk at almost 35 mg/ml. The lactation difficulties of TAB mice has not been further studied in this work.

6.2.5 The Tissue Specificity Of The TAB Transgene

The AATB transgene is expressed in the mammary gland. Many of the transgenic mice harbouring this transgene also exhibit minimal ectopic expression in the salivary gland (Archibald et al., 1990). The ectopic expression may in part be a property of the ovine BLG promoter in transgenic mice. In depth studies of ovine BLG transgene expression in a range of tissues in transgenic mice have demonstrated

low level ectopic expression in tissues other than the mammary gland (Farini & Whitelaw 1995). The expression of the transgene in these tissues is considerably lower than the expression of the transgene in the mammary gland. The reason for the ectopic expression is currently not known.

It was important to establish whether the murine genomic sequences included in the TAB transgene had an effect on the tissue specificity of the BLG promoter, which drives the expression of this transgene. In order to investigate this, the levels of hAAT transcript were analysed by Northern blotting. The filters were probed with an 800bp Bam HI/Eco RV fragment from p8 α 1ppg (Ciliberto et al., 1985). This probe encompasses most of the hAAT cDNA sequences.

The levels of hAAT mRNA were examined in a range of tissues of AATB 46.2 and a first generation transgenic mouse of line TAB 36. Both lines expressed hAAT at similar levels: AATB 46.2 giving 6.4 mg/ml and TAB 36 giving 7.5 mg/ml. These high levels of mammary expression were borne out by the Northern blot. On long exposures of the probed northern blots low level ectopic expression of the hAAT transcript was detected in the spleen of AATB 46.2 (figure 6.7) and in the salivary gland and kidney of TAB 36 (figure 6.8) but the levels of hAAT mRNA in these tissues was minimal compared with that in the mammary gland. The size of the transcript in the spleen of AATB 46.2 appears to be the same as that of the major transcript in the mammary gland, however, the transcript in the kidney of TAB 36 appears to be smaller. Both the larger and smaller mRNA species are present in the salivary gland of TAB 36. The production of the smaller transcript in the kidney and salivary gland may reflect differences between trans-acting splicing factors in the mammary gland and the kidney and salivary gland (Yull et al., 1995(b)).

Figure 6.8 : Northern Blot Analysis Of Total RNA From Transgenic Mice AATB 46.2 And TAB 36.2

The figure shows total RNA extracted from a range of tissues from AATB 46.2 and TAB 36. The tissues analysed were mammary (m), liver (Li), salivary (sa), kidney (k), spleen (sp), lung (lu) and muscle (m). Ten micrograms of total RNA were loaded from each tissue. The filter was probed with an 800 bp Bam HI/Eco RV fragment from p8 α 1ppg (Ciliberto et al., 1985). This fragment encompasses most of the cDNA sequences of hAAT. The lack of hybridisation in liver: the primary site of mouse endogenous AAT, suggested that the probe does not cross hybridise with the endogenous mouse AAT transcript. In both AATB 46.2 and TAB 36 the mammary gland is the major site of production of the hAAT transcript. Both mice express hAAT protein at similar levels and this is reflected in the similar abundance of the hAAT transcript in this tissue. Low level ectopic expression is exhibited in the spleen of AATB 46.2 and in the kidney of TAB 36. The levels of transcript in these tissues are minimal compared with the mammary expression.

AATB 46.2
m Li sa k sp Lu

TAB 36.2
m Li sa k sp mu

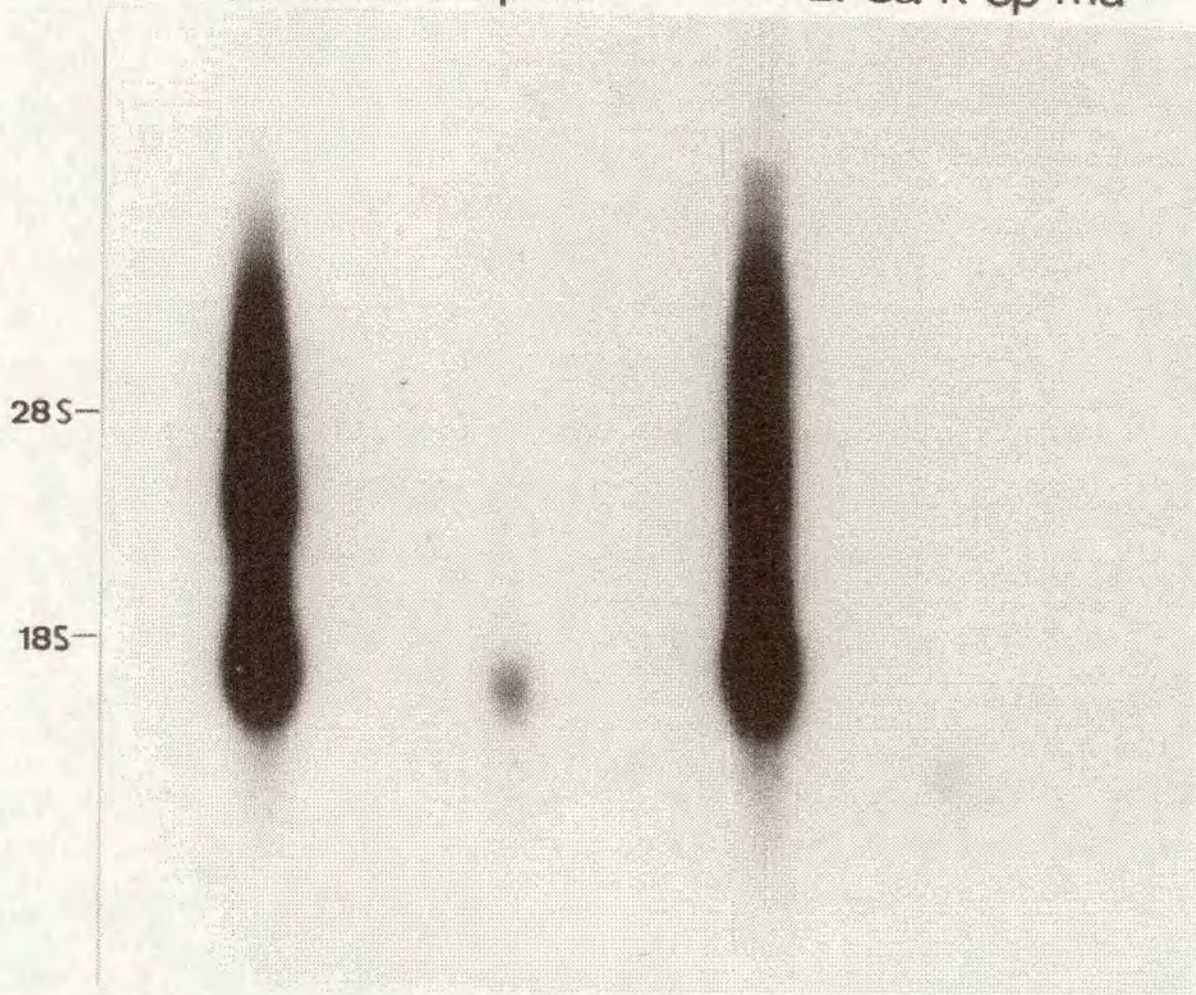
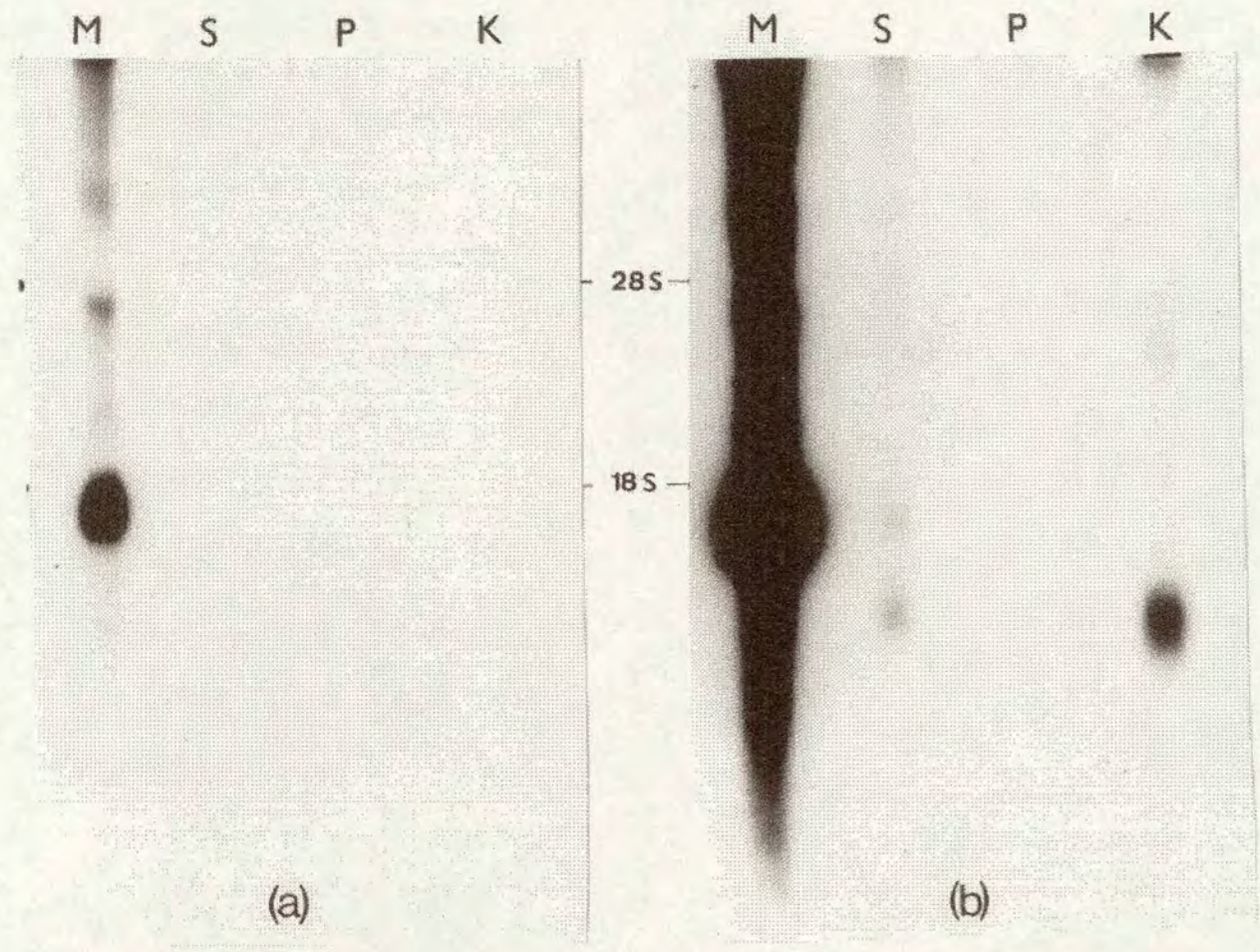


Figure 6.9 : Northern Blot Analysis of Total RNA from Transgenic Mouse TAB 36.2

The figure shows total RNA extracted from four tissues of TAB 36.2. The tissues analysed were mammary (M), salivary gland (S), pancreas (P), and Kidney (K). Ten micrograms of total RNA from each tissue were loaded. (a) is an autoradiograph of the filter following hybridisation of the filter with the radiolabelled Bam HI/Eco RV fragment of p8 α 1ppg and laying down with X-ray film for 1 hour, at -70°C, with intensifying screens. At this exposure hAAT transcript is only detected in the mammary gland. (b) shows a longer exposure of the same filter. It was exposed to X-ray film for 18 hours, at -70°C, with intensifying screens. At this exposure ectopic expression is detected in the kidney and salivary gland. Two species of hAAT transcript were detected in the salivary gland one is the same size as the major mammary transcript the other is slightly smaller. Only the smaller species is detected in the kidney.



The AATB and TAB transgenes are expressed mainly in the mammary gland of transgenic mice. Low level ectopic expression of both the BLG transgene and the AATB transgene has been demonstrated in previous studies. The tissue profile of expression of the AATB transgene (which forms the central part of the TAB transgene) does not appear to have been affected by the presence of the murine DNA of the AATB 46.2 locus. The appearance of ectopic expression of the TAB transgene demonstrated these murine sequences were unable to insulate the AATB transgene from positive position effects in tissues other than the mammary gland

6.2.6 hAAT Expression Per Copy Of The TAB Transgene.

The expression levels of the TAB transgene were significantly improved compared to the AATB transgene. To investigate if the expression levels in each of the lines/founders correlated with the transgene copy number the total expression for each line (Table 6.4) was plotted against the estimated transgene copy number (Tables 6.2 and 6.3). This plot is shown in figure 6.9. This analysis demonstrated that the TAB transgene shows a strong trend towards copy number dependent expression. The trend for each of the two sets of mice generated appears to be almost linear. However there appeared to be a difference between the mice generated at the Roslin Institute and those generated at PPL Therapeutics Ltd. It is possible since the copy numbers of these two sets of mice were analysed separately that the difference in expression per copy between the two groups is the result of errors in the estimation of copy number. This was unlikely since the same copy number controls were used in the estimation of copy number. Estimating copy number using plasmid copy number controls is subject to errors (see Chapter 3 of this thesis) even with the aid of a

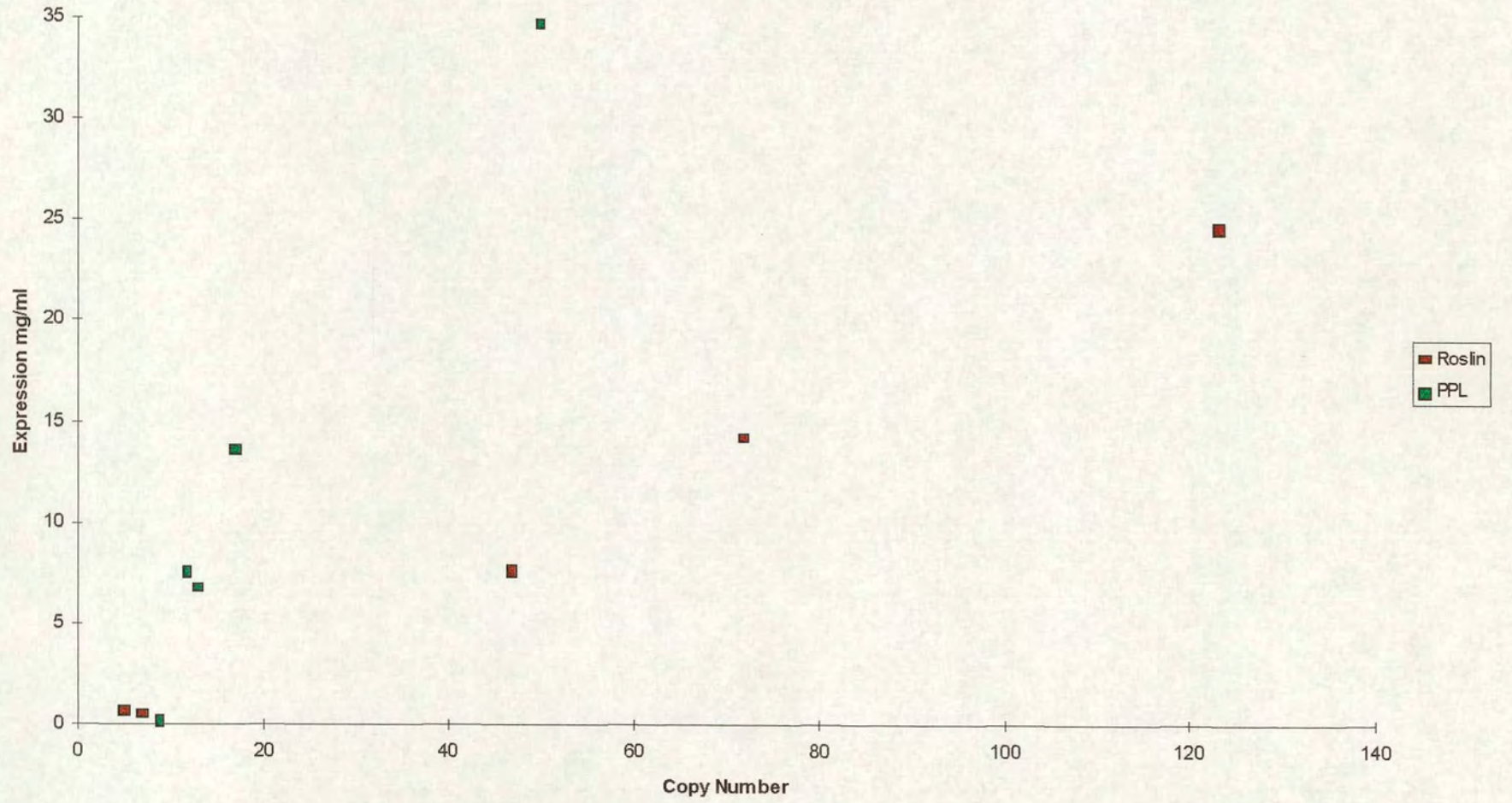
PhosphorImager. The extent of the error in copy number estimation using tail DNA and plasmid controls is difficult to quantify. However the discrepancy in the expression per copy between the two sets of mice is perhaps greater than would be expected from this sort of measurement error.

TAB 36 2 expressed the protein at 7.5 mg/ml and had an estimated copy number of 47, whereas TAB 110.28.2 expresses the protein at nearly 35 mg/ml yet had a similar estimated copy number of 50. The estimated copy numbers of these two mice are very similar yet the expression level of hAAT between the two mice shows an almost 5 fold difference. Similarly TAB 14.20.2 gave expression of 24.6 mg/ml and had an estimated copy number of approximately 120 but TAB 110 28.2 expresses hAAT at 10mg/ml higher but has an estimated copy number of 50 (2.4 times less than TAB 14.20.2). It is unlikely that the errors involved in measuring copy number are so very large and thus the difference in expression per copy number between the Roslin generated mice and the PPL generated mice probably reflects a real difference. No explanation can be given for this observed difference since the genetic background of the mice is the same.

Figure 6.10 : Plot of Copy Number Against hAAT Expression Level in TAB Transgenic Mice

The figure shows a graph of the estimated copy number of the TAB transgenic mice against the measured level of hAAT expressed in the milk of each line/founder. The two sets of mice have been entered on the same graph. The graph shows that the hAAT levels in the transgenic mouse milk rise with rising TAB transgene copy number. The level of expression per copy seems to vary between the two sets of transgenic mice generated for this study. It is not possible to tell if this difference is real or apparent with the data generated in this work as the error involved in calculating the copy numbers has not been quantified. Nevertheless it is quite clear that for both sets of TAB transgenic mice the expression is highest in mice/lines with high transgene copy number and lowest in mice/ lines with low copy number. The figure demonstrates that the murine sequences included in the TAB transgene confer copy number dependent expression on the AATB transgene. Copy number dependent expression was not seen in either of the two previous studies using the AATB transgene.

Expression of hAAT per copy



6.3 Analysis of the TA lines

Two secondary transgenes were obtained from the phage cloning. The analysis of the transgenic mice harbouring the TAB construct was detailed previously in this chapter. The other secondary transgene designated TA contained the same AATB transgene sequences but only the 4.7 kb of 3' flanking murine DNA from Clone 8. Three transgenic founder animals were produced from one round of injections of this construct. The female founder died in mid gestation leaving only two male founders from which to establish transgenic lines. Analysis of the milk samples from second generation females of both TA founder mice found that hAAT was not detectably expressed in the milk of either line. A further twenty TA transgenic females have been analysed. None of the were found to express detectable levels of hAAT in their milk (D. Cheng, unpublished observations). Thus it would appear that the improved profile of expression of the TAB transgene requires the 5' murine sequences. It is not clear whether the 5' sequences alone would confer the improved expression level or if the improved expression is only conferred on the AATB transgene when both the 5' and 3' murine flanking sequences are present.

6.4 Discussion

The profile of expression of the TAB secondary transgene is very predictable. The human protein encoded by the transgene is expressed in practically all of the TAB transgenic mice, and at a level that correlates with the transgene copy number (Figure 6.10). This contrasts with the apparent position dependent expression of the original AATB transgene. No linear relationship is observed between the AATB transgene copy number and hAAT expression.

The findings of the two previous studies of the AATB transgene were similar. The main difference between the studies was the method used for the detection of the hAAT protein in mouse milk. Carver et al. (1993) were able to detect very low levels of the human protein using a sensitive ELISA assay. These workers found expression of hAAT in the milk of all eleven transgenic, founder mice. Nevertheless the levels of expression in many of the transgenic mice were extremely low. Archibald et al. (1990), could detect as little as 5µg/ml of the human protein using Radio Immuno Assay (RIA). Only half of the transgenic mice generated by Archibald et al., (1990) detectably expressed hAAT compared with all of the Go mice in analysed by Carver et al., (1993). However this difference is merely a reflection of the lesser sensitivity of RIA compared with ELISA.

In both studies expression of hAAT per copy of the AATB transgene was generally low, and varied considerably between the lines. By contrast the expression of the TAB transgene was much more consistent and showed a relationship between copy number and level of expression for each line generated.

Of the five transgenic lines/animals generated for this work at PPL Therapeutics Ltd., four of them expressed hAAT at a similar level per transgene copy to AATB 46.2 (the line from which the secondary transgene was cloned). Five of the six TAB transgenic lines/animals generated at the Roslin Institute expressed the transgene at approximately 0.2 mg/ml/copy. This level of expression is similar to the level of expression of AATB 69 (Archibald et al. 1990). The published copy number for AATB 69 is two transgenes. The presence of a low copy number locus in this line is verified by the Southern blot shown in figure 3 6 in Chapter 3 of this work. The

published copy number for AATB 35 (Archibald et al. 1990) is one. It has been demonstrated in this work that the published copy number is an under estimate (See Chapter 3 of this work). The copy number has not been quantitatively determined but has been shown to be in excess of eleven copies. Comparing the copy number of AATB 35 with AATB 46.2 (Figure 4, Chapter 3 of this work) the copy number would appear to be considerably in excess of eleven copies. It is possible that the maximum levels of expression of AATB mice generated by Archibald et al., were in the region of 0.2 to 0.3 mg/ml/copy. Thus the difference in the average hAAT between the TAB mice generated at the Roslin Institute and those generated at PPL Therapeutics Ltd., may reflect the difference in maximum levels of expression displayed previously by the AATB transgenic mice generated at the two different sites. The genetic background of the mice at the two sites is, apparently, similar and there is no obvious explanation for the observed difference in transgene expression.

CHAPTER 7 : DISCUSSION

7.1 : A Summary of This Work

The promoter of AATB is that of the ovine BLG gene. The entire ovine BLG gene was shown to be expressed efficiently in the mammary gland of mice and moreover appeared to be expressed in a position independent and copy number dependent manner (Whitelaw et al 1992). This profile of expression is a remarkable quality displayed by only a small sub-set of transgenes and, is thought to result from the introduction of an entire regulatory unit.

However, expression of constructs comprising the BLG promoter linked to heterologous protein coding regions were generally disappointing. One such transgene is AATB, which comprises the BLG promoter and a human α_1 -antitrypsin minigene (Archibald et al., 1990). Studies performed with AATB demonstrated the transgene was capable of driving high levels of expression in the mammary gland of both sheep and mice. But, high levels of expression were exceptional and the mouse studies clearly demonstrated that AATB was susceptible to position effects (Archibald et al., 1990; Carver et al., 1993). Unusually high levels of expression of hAAT were exhibited by only three murine AATB lines. Two of these lines were stable AATB 35 and AATB 46.2 and appeared to harbour low copy number loci. The high levels of expression in these lines were thought to be the result of integration into a permissive chromosomal site. The aim of this work was to investigate the role of the flanking murine chromosomal DNA in directing high level expression of AATB.

The locus of AATB 46.2 was restriction mapped and cloned. In all 4.6 kb of the 5' and 4.7 kb of the 3' murine flanking DNA was obtained from the library. The flanking murine DNA was incorporated in a secondary transgene (TAB) comprising a single AATB transgene flanked 5' and 3' with the appropriate cloned murine DNA. Analysis of transgenic mice harbouring the TAB secondary transgene demonstrated clearly that the presence of both 5' and 3' murine DNA significantly improved expression of AATB. Only one TAB line did not express hAAT in their milk. The others all expressed more than 0.2 mg/ml. Moreover expression levels of hAAT increased with transgene copy number.

Transgenic animals were also generated carrying another secondary transgene (TA) that comprised AATB and the 3' flanking DNA only. Numerous transgenic mice harbouring this construct have now been generated and surprisingly no hAAT expression was detected in any of the lines generated. The presence of the 3' flanking DNA only; appears to be inhibitory as the TA transgene was less efficient than AATB. This indicates that either the improved AATB expression observed in the TAB transgenic mice require both the 5' and 3' flanking DNA, or the sequences responsible for position independence reside in the 5' flanking DNA.

The aim of this work was to establish whether the DNA sequences flanking an efficiently expressed transgene locus were responsible for the unusually high expression of the transgene locus. The secondary transgenic mice that harboured the transgene exhibited position independent and copy number dependent expression suggesting that the high expression of AATB observed in AATB 46.2 was the result of interaction of the

transgene with sequences at the site of integration. The levels of expression per copy of AATB in the secondary transgenic mice generated at PPL Therapeutics Ltd. appear to be very similar to the expression per copy of AATB in the original transgenic mouse from which the locus was cloned. Thus it would seem that rather than increase the maximal expression per AATB transgene copy, the murine sequences insulate the transgene from repressive position effects. Many questions arise from this work and further studies will be required to establish (a) the precise location of the sequences which confer the effect, (b) their ability to confer position independent copy number dependent expression on a range of other transgenes and (c) mechanism of action

7.2 : Further transgenic experiments

It is however unknown if the murine sequences isolated in this study will be capable of driving efficient expression of other constructs. Experiments are currently underway in cultured cells to establish if these sequences can improve expression of (a) constructs comprising the BLG promoter and heterologous coding sequences (cDNA and genomic) of genes other than human α 1-antitrypsin and (b) constructs driven by promoters other than that of the ovine BLG gene (D. Cheng). The effect of these sequences on promoters which drive gene expression in other tissues (i.e. liver) can be reliably tested in the first instance in well characterised cultured cell lines. Although experiments in cultured cells can give useful information about transgene expression no mammary cultured cell has been shown to retain all of the functions associated with the secretory mammary epithelium. Further transgenic experiments will ultimately be required

to establish the capabilities of the flanking murine sequences of AATB 46.2 to confer position independence on transgenes expressed in the mammary gland and other tissues of transgenic mice.

It is as yet uncertain whether both the 5' and 3' murine flanking DNA or just the 5' flanking DNA alone are required to for position independent and copy number dependent AATB expression. The generation of transgenic mice harbouring the 15.2 kb Not I fragment of clone 9 (AATB plus the 5' flanking DNA only, see figure 5.7 A) will answer this question. Once this question has been answered it should be possible to carry out a deletion analysis in transgenic mice to determine more accurately the location of sequences involved in conferring the observed position independence and copy number dependence.

A range of constructs, that have previously been analysed in transgenic mice are available to test the ability of the murine sequences to confer position independent expression. Yull et al., (1995 a) have recently demonstrated expression of factor IX in the mammary gland of transgenic mice harbouring the FIXD Δ 3' construct. This construct has the same BLG promoter region as AATB and the cDNA sequences of human Factor IX. This construct was co-injected with the entire BLG gene and shown to be expressed at low levels (100ng/ml-60 ug/ml) in the milk of all lines generated. This transgene did not exhibit copy number dependence. This transgene or similar constructs comprising the same BLG promoter as AATB but different protein coding sequences could be used to further test the function of the murine sequences isolated in this work. These sequences can also be tested for their ability to confer position independent copy number dependent expression upon transgenes that are expressed from other tissue specific promoters such as

that of the human α_1 -antitrypsin gene (Yull et al., 1995 b). Such experiments will establish whether the sequences isolated in this work confer position independent expression on transgenes generally or the observed effect of these murine sequences is limited to the AATB transgene.

7.3 : The Tissue Specificity of TAB is the Same as AATB

This work has demonstrated that the murine sequences cloned from AATB 46.2 can confer position independent copy number dependent expression of AATB. The mechanism by which these sequences act is unknown. Position independent copy number dependent expression of transgenes has been associated principally, to date, with the presence of Locus Control Regions. A number of transgenes exhibit this type of expression and have been reported to have LCR activity (see Introduction, section 1.4.4 and references therein). However the flanking murine DNA is not normally associated with either the ovine promoter (mice have no endogenous BLG gene) or the human AAT gene. Association of these sequences with AATB is the result of a random insertion event. LCR's that have been identified to date confer their effect on gene expression in a tissue specific fashion. The β -globin LCR (Grosveld et al., 1987) drives expression of the β -globin genes and heterologous genes in haemopoietic cells, and the CDC2 LCR (Greaves et al., 1989) drives expression of CDC2 and heterologous genes in lymphoid cells. However the mammary specific expression of AATB remains unchanged in the presence of the murine sequences suggesting that the observed enhancement of expression is not the result of integration next to either an LCR or enhancer. It cannot be ruled out

that the transgene array may have integrated next to a mammary specific enhancer but this is considered unlikely especially in the light of the sequences having no significant homology with known sequences in the EMBL database. Furthermore, although enhancers are capable of driving high levels of expression they are often unable to direct position independent and copy number dependent expression.

7.4 : The Murine Sequences May Contain Boundary Elements

Another possible explanation for the position independent expression of the TAB transgene is the presence of boundary elements in the flanking murine DNA. The presence of such elements could allow the BLG promoter to establish and maintain an open chromatin domain that is buffered from possible repressive effects of chromatin lying outwith the boundary elements. Putative boundary elements in both *Drosophila* (Gasser and Laemmli, 1986) and mammals (Steif et al., 1989) have been proposed to be regions of DNA that bind the nuclear scaffold/matrix. S/MAR's have been demonstrated to be located at the ends of genes at the boundary between DNase I sensitive and insensitive DNA. It has been proposed that they allow for the formation of independent regulatory domains, buffering genes within their boundaries from the effects of neighbouring chromatin. The presence of boundary elements would explain how the flanking DNA can confer its effect when flanking an array of approximately 7 transgenes (AATB 46.2) and when flanking individual transgenes within an array (TAB mice). Matrix binding assays have not been performed in this work and the extent of DNase I sensitivity of the TAB transgene in the mammary gland during lactation has not been

investigated. Thus it is unknown if the murine DNA binds the nuclear matrix or even contains regions that are relatively refractory to DNase I. Although it has been proposed that S/MARs can function as boundary elements they have not been conclusively demonstrated to do so in vivo. However it does appear from the literature that S/MARs may have a role in the regulation of transcription by exerting an effect on local chromatin structure (see Introduction, section 1.4.5 and references therein).

The sequence of 2.6 kb of the 3' flanking DNA contains some motifs that have previously been associated with S/MAR's but contains no topoisomerase II sites. The 5' end is 80% AT rich over 145 bp but contains very little in the way of motifs that have previously been identified in MARs. The 5' murine DNA does however contain some AT runs which have been shown to bind proteins that have only recently been isolated on the basis of their ability to bind MARs. Two such proteins Bright (Herrscher et al., 1995) and CHD 1 (Stokes and Perry, 1995) have regions that are homologous with proteins of the SWI SNF family and have been proposed to be involved in transcriptional regulation possibly via chromatin remodelling. It is possible that the enhanced expression observed with the TAB secondary transgene could result from the binding of similar proteins to the AT rich region immediately 5' of AATB.

7.5 : Can these Murine Sequences Confer their effect in Dairy Animals ?

The ability of these murine sequences to confer a position independent and copy number dependent expression on a heterologous construct driven by this promoter is an important result. An increasing number of transgenes have been shown to exhibit position

independent copy number dependent expression. Generally this type of expression is the result of inclusion of sequences outwith the coding region that are normally associated with the gene of interest. Analysis of these transgenes has been informative about aspects of transcriptional control that are currently poorly understood. In this work the sequences that bring about this type of transgene expression are not normally associated with the gene. The BLG promoter is an important promoter that is currently used to drive the expression of biomedical proteins in the milk of transgenic sheep.. Currently all transgenic large animals are generated by pro-nuclear injection. Transgenic experiments with large animals are expensive. The availability of a promoter that can drive expression of biomedical proteins in a position independent manner may significantly reduce the cost of such experiments.

Further studies will establish the capabilities of these sequences to drive the expression of transgenes comprising the BLG promoter and other heterologous protein coding sequences in mice. However to be useful for the commercial production of economically important proteins in transgenic animals these sequences will need to confer position independent copy number dependent expression in dairy animals (the sheep in particular). It is plausible that these sequences will be able to do this since they have been isolated as result of their ability to confer position independent and copy number dependent expression upon a construct driven by an ovine milk protein gene promoter. Currently it is not known how the murine sequences confer this effect. However, the effect is likely to be mediated by specific DNA binding proteins and if the factors involved are conserved between the mouse and sheep then it is possible that expression level of

BLG constructs in transgenic sheep could also be improved by the inclusion of these murine sequences in transgene constructs whose expression is driven by the BLG promoter.

References

- Adams, C.C. and Workman J.L. (1995) Binding of disparate transcriptional Activators to nucleosomal DNA is inherently cooperative *Mol. Cell Biol.* 5 : 1405-1421.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson J.D. (1989). *Molecular biology of the cell*, 2nd edition, New York : Garland Publishing Inc
- Ali, S. (1989) Characterisation of the Ovine Beta-lactoglobulin gene. PhD. Thesis. University of Edinburgh
- Ali, S. and Clark, A.J. (1988). Characterisation of the gene encoding ovine Beta-lactoglobulin. *J. Mol. Biol.* 199 : 415-428
- Allen, N.D., Norris, M.L. and Surani, M.A.H., (1990). Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell* 61 : 853-861.
- Al-Shawi, R., Kinnaird, J., Burke, J., Bishop, J.O., (1990). Expression of a foreign gene in a line of transgenic mice is modulated by a chromosomal position effect. *Mol. Cell Biol.* 10 : 1192-1198.
- Archibald, A.L., McClenaghan, M., Hornsey, V., Simons, J.P. and Clark, A.J. (1990). High-level expression of biologically active human α_1 -antitrypsin in the milk of transgenic mice. *Proc. Natl. Acad. Sci. USA.* 87: 5178-5182
- Aronow, B. J., Ebert, c. H., Valerius, M. t., Potter, S. S., Wiginton, D. A., Witte, D. P. and Hutton, J. J. (1995) Disciting a locus control region : facilitation of enhancer function by extended enhancer-flanking sequences. *Molecular and Cellular Biology* 15 : 1123-1135
- Atchison, M.L. (1988) Enhancers : Mechanisms of Action and cell specificity. *Annual Review of Cell Biology* 4 : 127-53
- Banerji, J., Olson, L. and Schaffner, W. (1983) A lymphocyte-specific cellular enhancer is located downstream of the joining regions in immunoglobulin heavy chain genes. *Cell* 33 : 729-740
- Banerji, J., Rusconi, S. and Schaffner, W. (1981) Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27 : 299-308
- Behringer, R.R., Ryan, T.M., Palmiter, R.D., Brinster, R.L. and Townes, T.M. (1990). Human γ to β -globin gene switching in transgenic mice. *Genes and Development* 4 : 380-389.
- Belgrader, P. Siegel, A.J. and Berezney, R. (1991). A Comprehensive study on the isolation and characterization of the HeLa S3 nuclear marix. *Journal of Cell Science* 98 : 281-91.
- Benoist, C. and Chambon, P. (1981) In vivo sequence requirements of the SV40 early promoter region. *Nature* 290 : 304-310

- Berezney, R. and Coffney, D.S. (1974). Identification of a nuclear protein matrix. *Biochem. Biophys. Res. Commun* 60 : 1410-1417.
- Berrios, M., Osherhoff, N. and Fisher, P.A. (1985). *In situ* localisation of DNA Topoisomerase II, a major polypeptide component of the *Drosophila* nuclear matrix fraction. *Proceedings of the National Academy of Sciences USA* 82 : 4142-4146.
- Bird, A.P. et al. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated CpG rich DNA. *Cell* 40 : 91-99
- Bishop, J.O. & Smith, P. (1989). Mechanism of chromosomal integration of microinjected DNA. *Mol Biol. Med.* 6 : 283-298.
- Black, D.L., Chaubot, B. and Steitz (1985) U2 as well as U1 small nuclear ribonucleoproteins are involved in premenenger RNA splicing. *Cell* 42 : 737-750.
- Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L.A., Grunwald, D.J., Keifer, D.O., Moor, D.D., Sheldon, H.E. and Smithies, O. (1977). Charon phages: safer derivatives of bacteriophage lambda for cloning. *Science* 196 : 161-163
- Blom van Assendelft, G. Hanscombe, O., Grosvald, F. and Gewes, D.R. (1989). The β -globin dominant control region activates homologous and heterologous promoters in a tissue specific manner. *Cell* : 56 : 969-977.
- Bode, J. and Maass, K. (1988). Chromatin domain surrounding the human interferon- β gene as defined by scaffold-attached regions. *Biochemistry* 27 : 4706-4711.
- Bode, J., Kohwi, L., Dickenson, T., Joh, D., Klehr, C., Meilke, T. and Kohwi-Shigematsu, T. (1992). Biological significance of Unwinding Capability of Nuclear Matrix associated DNA's. *Science* 255 : 195-197
- Bonifer, C., Vidal, M. Grosveld, F. and Sippel, A.E. (1990). Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. *EMBO J.* 9 : 2843-2848.
- Bonifer, C., Yanoutsos, N., Krüger, G., Grosveld, F. and Sippel, A.E. (1994). Dissection of the locus control junction located on the chicken lysozyme gene domain in transgenic mice. *Nucleic Acids Research* 22 : 4202-4210.
- Boulikas, T. Homeodomain Protein Binding sites, Inverted repeats, And Nuclear Matrix Attachment Regions Along the Human β -Globin Gene Complex. *J. Cell. Biochem.* 52 : 23-36
- Boulikas, T. Nature of DNA Sequences at the Attachment Regions of Genes to the Nuclear Matrix. (1993). *J. Cell. Biochem.* 52 : 14-22
- Bowman W.C. and Rand, M.J. (1984). *Textbook of Pharmacology*, Blackwell scientific publications pp 19.54
- Brandon, E. P., Idzerda, R.L., McKnight, G.S. (1995) Targeting the mouse genome : a

compendium of knockouts (part I, II, III). *Current Biology* 5 : 625-634, 758-765, 873-881.

Breathnach, R., Mandel, J.L., and Chambon (1977). Ovalbumin gene is split in chicken DNA. *Nature* 270 : 314-319.

Breathnach, R. and Chambon, P. (1981) Organisation and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50 : 349-383.

Breathnach, R., and Chambon, P. (1981). Organisation and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50: 349-383

Brinster, R.L. Chen, H.Y., Trumbauer, M.E., Senear, R., Warren, R. and Palmiter, R.D. (1981). Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. *Cell* 27 : 223-231.

Brinster, R. L., Allen, J. M., Behringer, R. R. Gelinis, R. E. and Palmiter, R. D. (1988) Introns increase transcriptional efficiency in transgenic mice. *Proceedings of the National Academy of Sciences (USA)* 85 : 836-840

Brinster, R. L., Chen, H. Y., Trumbauer, M. E., Yagle, M. K. and Palmiter, R. D. (1985) Factors affecting the introduction of foreign DNA into transgenic mice by microinjecting eggs. *Proceedings of the National Academy of Science (USA)* 82 : 4438-4442

Brinster, R.L., Sandgren, E. P., Behringer, R.R., and Palmiter, R.D. (1989) No simple solution for making transgenic mice. *Cell* 59 : 239-241

Brooks, A.R. Nagg, B.P., Taylor, W.S. Simonet, J. Taylor, J. and Levy-Wilson, B. (1994). Sequences containing the second-intron enhancer are essential for transcription of the human apolipoprotein B gene in the livers of transgenic mice. *Mol. Cell Biol.* 14 : 2243-2256.

Brown, W.R.A. and Bird, A.P. (1986) Long-range restriction site mapping of mammalian genomic DNA. *Nature* 322 : 477

Buchman, A.R. and Berg, P. (1988). Comparison of intron-dependent and intron-independent gene expression. *Mol. Cell Biol.* 8 : 4395 - 4405.

Burdon, T., Sankaran, L., Wall, R.J., Spencer, M. and Henninghausen, L. (1991). Expression of a Whey Acidic Protein Transgene during Mammary Development. *J. Biol. Chem.* 266 : 6909-6914.

Burdon, T.G., Maitland, K.A., Clark A.J., Wallace R. and Watson, C.J. (1994). Regulation of the sheep β -lactoglobulin gene by lactogenic hormones is mediated by a transcription factor that binds an interferon- γ activation site-related element. *Mol. Endo.* 8: 1528-1536

Campbell, K.H.S., McWhir, J., Ritchie, W.A. and Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. *Nature* 380 : 64-66.

Capecchi, M.R., (1989) The new mouse genetics: altering the genome by gene targeting., Trends in Genetics Vol 5: 70-76

Carey, M.F. and Singh, K. (1988). Enhanced B2 transcription in simian virus 40-transformed cells is mediated through the formation of RNA polymerase III transcription complexes on previously inactive genes. Proc. Natl. Acad. Sci. USA. 85 : 7059-7063

Carey, M.F., Singh, K., Botchan, M. and Cozzarelli, N.R. (1986). Induction of specific transcription by RNA polymerase III in transformed cells. Mol. Cell. Biol. 6 : 3068-3076

Carver, A.S., Dalrymple, M.A., Wright, G., Cottom, D., Reeves, D.B., Gibson, Y.H., Keenan, J.L., Barrass, J.D., Scott, A.R., Colman A. and Garner I. (1993). Transgenic livestock as bioreactors: stable expression of human alpha1-antitrypsin by a flock of sheep. Bio/Technology 11: 1263-1270

Casolaro, M.A., Fells, G., Wewers, M., Pierce, J.E., Ogushi, F., Hubbard, R., Sellers, S., Forstrom, Lyons, D., Kawasaki, G. and Crystal, R.G. (1987). Augmentation of lung antineutrophil elastase capacity with recombinant human α_1 -antitrypsin. Journal of Appl. Physiol. 63 : 2015-2023.

Caterina, J.J., Ciavatta, D.J., Donze, D., Behringer, R.R. and Townes, T. (1994). Multiple elements in human β -globin locus control region, 5' HS2 are involved in enhancer activity and position-independent, transgene expression. Nucleic Acids Research 22 : 1006-1011.

Ceder, H. (1988). DNA methylation and Gene Activity. Cell 53 : 3-4.

Cheyette, T.E. Ip, T., Faber, S., Matsui, Y. and Chalkley, R. (1992) Nucl. Acids Res. 13: 3427-3433

Choi, T., Huang, M., Gorman, C. and Jaenisch, R. (1991) A generic intron increases gene expression in transgenic mice. Mol. Cell Biol 11 : 3070-3074.

Chomczynski, P. and Sacchi N (1987). Single step method of RNA isolation by Acid Guanidium Thiocyanate-Phenol-Chloroform extraction. Anal. Biochem. 162 : 156-159

Chung, J.H. Whiteley, M. and Felsenfeld, G. (1993). A 5' element of the chicken β -globin domain serves as an insulator in human erythroid cells and protects against position effects in drosophila. Cell 74 : 505-514.

Church, G.M. and Gilbert, W. (1984) Genomic Sequencing. Proc. Natl. Acad. Sci. USA 81 : 1991-1995

Ciliberto, G., Dente, L. and Cortese, R. (1985). Cell-Specific Expression of a Transfected Human α_1 -Antitrypsin Gene. Cell 41:531-540

Clark A.J., Simons J.P. and Wilmut, I. (1992). Germline manipulation : applications in agriculture and biotechnology. In : Grosveld, F., Kollias, G., (eds) Transgenic animals

pp. 250-251.

Clark, A. J., Archibald, A. L., McClenaghan, M., Simons, J.P., Whitelaw, C. B. A. and Wilmut, I. (1990) The germline manipulation of livestock : progress during the past five years. *Proceedings of the National New Zealand Society of Animal Production* 50 : 167-179

Clark, A.J. (1996) Genetic modification of milk proteins. *Am. J Clin Nutr* 63: 633S-638S

Cockerill, P.N. and Garrard, W.T. (1986). Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topo-isomerase II sites. *Cell* 44 : 273-282.

Davie, E.W. (1986). Introduction of the blood coagulation cascade and the cloning of blood coagulation factors. *J. Protein Chemistry.* 5 : 247-253.

Dickenson, L.A. and Khowi-Shigematsu, T. (1995). Nucleolin is a matrix-attachment region DNA-binding protein that specifically recognises a region with high base-unpairing potential. *Mol. Cell Biol.* 15 : 456-465.

Dickenson, L.A., Joh, T., Kohwi, Y. and Kohwi-Shigematsu, T. (1992). A time specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* 70 : 6331-645.

Dillon, N. and Grosveld F. (1993). Transcriptional regulation of multigene loci: multilevel control. *Trends in Genetics* 9 : 134-137.

Dobie, K.W., Lee, M., Fantos, J.A., Graham, E., Clark, A.J., Springbett, A. Lathe, R. and McClenaghan, M., (1996). Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. *Proceeding of the National Academy of Science USA.* 93 : 6659-6664.

Dworetzky, S.I. Wright, K.L. Fey, E.G., Penman, S. Lian, J.B., Stein, J.L. and Stein, G.S. (1992). Sequence-specific DNA-binding proteins are components of a nuclear matrix-attachment site. *Proceedings of the National Academy of Sciences USA* 89 : 4178-4182.

Dynan, W. S. (1989) Modularity in promoters and enhancers. *Cell* 58 : 1-4

Ellis, J. Tan-Un, K.c., Harper, A., Michalovich, D., Yannoutsos, N. Philipsen, S. and Grosveld, F. (1996), A dominant chromatin opening activity in 5' hypersensitive site 3 of the human β -globin locus control region. *EMBO J.* 15: 562-568

Ellis, J., Talbot, D., Dillon, N. and Grosveld, F. (1993). Synthetic human β -globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. *EMBO J.* 12 : 127-134.

Evans, G.A., Lewis K. and Rothenberg, B.E. (1989). High efficiency vectors for cosmid microcloning and genomic analysis. *Gene* 79:9-20

- Evans, M.J. and Kaufman, M.H. (1981). Establishment in culture of pluripotent cells from mouse embryos. *Nature* 292 : 154-156.
- Farini, E., and Whitelaw, C.B.A. (1995). Ectopic expression of β -lactoglobulin transgenes. *Mol. Gen. Genet.* 246: 734-738
- Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragmentys to high specific activity. *Anal Biochem.* 132 : 6-13
- Fitzgerald, M. and Shenk, T. (1981). The sequence S'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNA's. *Cell* 24 : 251-260.
- Folger, K.R., Wong, E.A., Wahl, G. and Capecchi, M.R. (1982). Patterns of integration of DNA microinjected into cultured mammalian cells : evidence for homologous recombination between injected plasmid DNA molecules. *Mol. Cell Biol.* 2 : 1372-1387.
- Forrester, W.C., Eprer, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T. and Groudine, M. (1990). A deletion of the β -globin locus activation region causes major alteration in chromatin structure and replication across the entire β -globin locus. *Genes and Development* 4 : 1637-1649.
- Forrester, W.C. Thompson, C., Elder, J. and Groudine, M. (1986). A developmentally stable chromatin structure in the human β -globin gene cluster. *Proceedings of the National Academy of Sciences USA.* 83 : 1359-1363.
- Franklin, G.C., Donovan, M., Adam, G.I.R., Holmgren, L., Pfeifer-Ohlsson, S. and Ohlsson, R., (1991) Expression of the human PDGF-B gene is regulated by both positively and negatively acting cell type specific regulatory elements located in the first intron. *EMBO J.* 10 : 1365-1373.
- Fraser, P. Hurst, J., Collis, P. and Grosveld, F. (1990). Dnase 1 hypersensitive sites 1,2 and 3 of the human β -globin dominant control region direct position independent expression. *Nucleic Acids Research* 18 : 3503-3508.
- Fraser, P., Pruzina, S., Antoniou, M. and Grosveld, F. (1993). Each hypersensitive site of the human β -globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes and Development* 7 : 106-113
- Gall, J.G. (1956). On the submicroscopic structure of chromosomes. *Brookhaven Symp. Biol.* 8 : 17-32.
- Garver, R.I, Chytil, A., Karlsson, S., Fells, G.A., Brantly, M.L., Coutney, M., Kantoff, P.W., Nienhuis, A.W., Anderson, W.F. and Crystal, R.G. (1987). Production of glycosylated physiologically "normal" human α_1 -antitrypsin by mouse fibroblasts modified by insertions of a human α_1 -antitrypsin cDNA using a retroviral vector. *Proceedings of the National Academy of Sciences USA* 84 : 1050-1054.
- Gasser, S. and Laemmli, U.K. (1986). Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of D.

melanogaster. Cell 46 : 521-530.

Gillies, S. D., Morrison, S. L., Oi, V. T. and Tonegawa, S. (1983) A time-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33 : 717-728

Gordon, J. W., Scangos, G. A., Plotkin, D., Barbosa, J. A. and Ruddle. (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. Proceedings of the National Academy of Sciences USA 77 : 77380-77384

Gossler, A., Joyner, A. L., Rossant, J. and Skarnes, W. C. (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science 244 : 463-465

Greaves, D.R., Wilson, F.D., Lang, G. and Kioussis, D. (1989). Human CD2 3'-flanking sequences confer high-level T-cell specific, position-independent gene expression in transgenic mice. Cell 56 : 979-986.

Greenberg, N.M., Anderson, J.W., Hsueh, A.J.W. Nishimori, K., Reeves, J.J., DeAvila, D.M., Ward, D.N. and Rosen, J.M. (1991). Expression of biologically active heterodimeric bovine follicle-stimulating hormone in milk of transgenic mice Proceedings of the National Academy of Sciences USA 88 : 8327-8331.

Grosveld, F., Blom van Assendelft, G., Greavas, D. R. and Kollias, G. (1987) Position independent, high level expression of the human β -globin gene in transgenic mice. Cell 51 : 975-985

Hadchouel, M., Farza, H., Simon, D., Tiollais, P. and Pourcel, C. (1987). Maternal inhibition of hepatitis B surface antigen gene expression in transgenic mice correlates with de novo methylation. Nature 329 : 454-456.

Hamada, T., Sasaki, H., Seki, R. and Sakaki, Y. (1993). Mechanisms of chromosomal integration of transgenes in microinjected mouse eggs : sequence analysis of genome-transgene and transgene-transgene junctions at two loci. Gene 128 : 197-202.

Hasty, P., Ramirez-Solis, R., Krumlauf, R. and Bradley, A. (1991). Subtle mutation in the hox 2.6 locus in ES cells. Nature 350: 243-246

Hawley, D. K. and McClure, W. R. (1983) Compilation and Analysis of Escherichia Coli Promoter DNA sequences. Nucleic Acids Research 11 : 2237-2255

Hendzel, M.J., Delcuve, G.P. and Davie J.R. (1991). Histone Deacetylase is a component of the internal nuclear matrix. J. Biol. Chem. 266 : 21936-21942.

Henikoff S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing Gene 28 : 35-359

Henikoff, S. (1994). A reconsideration of the mechanism of position effect. Genetics 138 : 1-5.

Herrscher, R.F. Kaplan, M.H. Lelsz, D.L., Das, C. Scheuermann, R. and Tucker,

- P.W. (1995). The immunoglobulins heavy-chain matrix-associating regions are bound by Bright : a B cell-specific trans-activator that describes a new DNA-binding protein family. *Gen. Dev.* 9 : 3067-3082.
- Hinnen, A., Hicks, J.B. and Fink, G.R. (1978). Transformation of yeast. *Proceeding of the National Academy of Science USA.* 75 : 1929-1933.
- Hobbs, A.A., Richards, D. A. and Rosen J.M. (1982). Complex regulation of casein gene expression. *J. Biol. Chem.* 257: 3598-3605.
- Hodgson, J. (1992). Whole animals for wholesale protein production. *Bio/technology* 10: 863-866
- Hodgson, J. (1991). Cell culture over-sold. *Bio/technology* 9: 810-812
- Hooper, M.L., Hardy, K., Handyside, A., Hunter, S. and Monk, M. (1987). HPRT deficient (Lesch-Nyhan) mouse embryos derived from germ line colonisation by cultured cells. *Nature* 326 : 292-295.
- Jaenisch, R. (1976) Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proceedings of the National Academy of Sciences (USA)* 73 : 1260
- Jaenisch, R. and Jahner, D., (1984). Methylation, expression and chromosomal position of genes in mammals. *Biochem, Biophys Acta* 782 : 1-9.
- Jaenisch, R., (1988). Transgenic animals. *Science* 240: 1468-1474
- Jantzen, K., Fritton, H.P. and Iyo-Kemenes, T. (1986). The DNase 1 sensitive domain of chicken lysozyme gene spans 24kb. *Nucleic Acids Research* 14 : 6085-6099.
- Jeffreys, A.J., and Flavell, R.A. (1977). The rabbit beta-globin gene contains a large insert in the coding sequence. *Cell* 12 : 1097-1108.
- Jenness, R. (1982). Inter-species comparison of milk proteins. In Fox, P.F. (ed) *Developments in dairy chemistry (vol 1)*. Applied Science Publications, New York.
- Johnson, I.S. (1983). Human insulin from recombinant DNA technology. *Science* 219 : 632-637
- Keller, W. (1984). The RNA Lariat. A new ring to the splicing of mRNA precursors. *Cell* 39 : 423-425.
- Kellum, R. and Schedl, P. (1991). A position-effect array for boundaries of higher order chromosomal domains. *Cell* 64 : 941-950.
- Kellum, R. and Schedl, P. (1992). A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol. Cell Biol.* 12 : 2424-24231.
- Kioussis, D., Vanin, E., de Lange, T., Flavel, R. A. and Grosveld, F. (1983) β -globin gene inactivation by DNA-translocation *Nature* 306 : 662-666

- Knotts, S., Rindt, H. and Robbins, J. (1995). Position independent expression and developmental regulation is directed by the β myosin heavy chain gene's 5' upstream region in transgenic mice. *Nucl. Acids Res.* 23: 3301-3309.
- Kollias, G., Hurst, J., de Boer, E. and Grosveld, F. (1987) A tissue and developmental specific enhancer is located downstream of the human β -globin gene. *Nucleic Acids Research* 15 : 5739-5747
- Kollias, G., Wrighton, N., Hurst, J. and grosveld, F. (1986) Regulated expression of human $A\gamma$ -,B- and hybrid $\gamma\beta$ -globin genes in transgenic mice : manipulation of the developmental expression patterns. *Cell* 46 : 89-94
- Kopchick, J.J. and Stacey, D.W. (1984). Difference in intracellular DNA ligation after microinjection and transfection. *Mol. Cell Biol.* 4 : 240-246.
- Kotow, K. and Orkin, S. H. (1995) Dependence of Globin Gene Expression in Mouse Erythroleukemia cells on the NF-E2 heterodimer. *Molecular and Cellular Biology* 15 : 4640-4697
- Lavitrano, M., Camaloni, A., Fazio, V. M., Dolci, S., Farace, M. G. and Spadajora, C. (1989) Sperm cells as vectors for introducing foreign DNA into eggs : genetic transformation of mice. *Cell* 57 : 717-723
- Lee, K-F., Atiee, S.H. and Rosen, J.M. (1989). Differential regulation of rat β -casein-chloramphenicol acetyl-transferase fusion gene expression in transgenic mice. *Mol. Cell. Biol.* 9: 560-565
- Levy-Wilson, B. and Fortier, C. (1989). The limits of DNase 1 sensitive domain of the human apolipoprotein gene coincide with the locations of chromosomal anchorage loops and define the 5' and the 3' boundaries of the gene. *J. Biol. Chem.* 264 : 21196-21204.
- Lewin, B. (1987). *Genes* III. 3 rd edition. Publisher; John Wiley and Sons Inc.
- Lewis, C.D. and Laemmli, U.K. (1982) Higher order metaphase chromosome structure: evidence for metaloprotein interactions *Cell* 29: 171-181
- Li, E. Bestor, T.H. and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69 : 915-926.
- Li, S. and Rosen, J.M. (1994). Glucocorticoid regulation of whey acidic protein gene expression involves hormone induced alterations of chromatin structure in the distal promoter region. *Mol. Endo.* 8: 1328-1335
- Lovell-Badge R.H. (1985). New advances in the field. *Nature* 315 : 628-629.
- Lowrey, C.H., Bodine, D.M. and Nienhuis, A.W. (1992). Mechanism of DNase 1 hypersensitive site formation within the human globin locus control region. *Proceedings of the National Academy of Sciences USA* 89 : 1143-1147.
- Maga, E.A. and Murray J.D. (1995). Mammary gland expression of transgenes and the

potential of altering the properties of milk. *Bio/Technology* 13 : 1452-1456.

Magin, T.M., McWhir, J. and Melton, D.W. (1992). A new mouse embryonic stem cell line with good germ line contribution and gene targeting frequency. *Nucleic Acids Research* 14 : 3795

Magram, J., Chuda, K. and Courtantini, F. (1985) Developmental regulation of a cloned adult β -globin gene in transgenic mice. *Nature* 315 : 338-340

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). *Molecular cloning: A laboratory manual*. Cold Spring Harbour Laboratory Press, New York.

Mansour, S.L., Thomas, K.R. and Capecchi, M.R. (1988). Disruption of the proto-oncogene *int-2* in mouse embryo derived stem cells : a general strategy for targeting mutations to non selectable genes. *Nature* 336 : 348-352.

Martin, G. (1981). *Proceeding of the National Academy of Science USA*. 78 : 7634-7638.

McClenaghan, M., Springbett, A., Wallace R., Wilde C.J. and Clark, A.J.(1995). Secretory proteins compete for production in the mammary gland of transgenic mice. *Biochem. J.* 310 : 637-641

McGowan, R., Campbell, R., Peterson, A. and Sapienza, C. (1989). Cellular mosaicism in the methylation and expression of hemizygous loci in the mouse. *Genes and Development* 3 : 1669-1676.

McKnight, R.A. Shamay, A., Sankaran, L., Wall, R.J. and Henninghausen, L. (1992). Matrix-attachment regions can impart position independent regulation of a tissue specific gene in transgenic mice. *Proceedings of the National Academy of Sciences USA*. 89 : 6943-6947.

Mirkovitch, J., Mirault, M.E. and Laemmli U.K. (1984). Organisation of the higher order chromatin loop : specific DNA attachment sites on nuclear scaffold. *Cell* 39 : 223-232

Moore, D.D., Marks, A.R., Buckley, D.I., Kapler, G., Payvar, F. and Goodman, H.M. (1985). *Proceedings of the National Academy of Sciences USA* 82 : 699-702.

Moreau, J., Marcaud, E., Maschat, F., Kejzlarova-Lepesant, J., Lepesant J-A. and Scherver, K. (1982). A+T-rich linkers define functional domain in eukaryotic DNA. *Nature* 295 : 260-262.

Nelson, K.K. and Green M.R. (1988). Splice-Site selection and Ribonucleoprotein complex assembly during in vitro pre-mRNA splicing. *Genes and Development* 2 : 319-329.

Ninomiya, T., Hoshi, M. and Yuki, A. (1988). Structures of integrated DNA containing human adenovirus EIA genes in transgenic mice. *Agric. Biol. Chem.* 52 : 2537-2546.

O'Kane, C. J. and Gehring, W. J. (1987) Detection in situ of genomic regulatory elements in *Drosophila*. *Proceedings of the National Academy of Sciences (USA)* 84 : 9123-9127

Oganoh, O.W., Freedman R.B., Jemkins, N., Patel, K. and Rooney, B.C. (1996). Isolation and characterisation of an insect cell line able to perform complex N-linked glycosylation on recombinant proteins. *Bio/Technology* 14 : 197-202.

Padgett, R.A. Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986). Splicing of meneger RNA precursors. *Ann. Rev. Biochem.* 55: 1119-1150.

Palmiter, R. D. and Brinster, R. L. (1985) Transgenic Mice. *Cell* 41 : 343-345

Palmiter, R. D., Chen, H. Y. and Brinster, R. L (1982) Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. *Cell* 29 : 701-710

Palmiter, R. D., Sandgren, Brinster, R. L. (1991) Heterologous introns can enhance expression of transgenes in mice. *Molecular and Cellular Biology* 11 : 3070-3074

Palmiter, R.D., Brinster, R.L., Hammer, R.E., Trumbauer, M.E., Rosenfeld, M.G., Birnberg, M.G. and Evans, R.M. (1982). Dramatic growth of mice that develop from eggs microinjected with metallothionein - growth hormone fusion genes. *Nature* 300 : 611-615.

Palmiter, R.D., Chen, H.Y., and Brinster, R.L. (1982). Differential regulation of metallothionein-thymidine kinase genes in transgenic mice and their offspring. *Cell* 29 : 701-710.

Palmiter, R.D. and Brinster, R. L. (1986) Germline transformation of mice. *Annual Revue of Genetics* 20 : 465-499

Pardue, M.L., Lowenhaupt, K., Rich, A. and Nordheim, A. (1987). (dC-dA)_n. (dG-dT)_n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. *EMBO* 6 : 1781-1789

Paulson, J.R. and Laemmli, U.K. (1977). The structure of histone depleted metaphase chromosomes. *Cell* 12 : 817-828.

Pavirani, A., Skern, T., Le Meur, M., Lutz, Y., Lathe, R., Crystal, R.G., Fuchs, J.P., Gerlinger, P. and Courtney M. (1989). Recombinant proteins of therapeutic interest expressed by lymphoid cell lines derived from transgenic mice. *Bio/Technology* 7 : 1049-1053.

Pestka, S. (1983). The purification and manufacture of human interferons. *Sci.Amer.* 249 : 36

Philipsen, S. Pruzina, S. and Grosveld, F. (1993). The minimum requirements for activity in transgenic mice of hypersensitive site 3 for the β -globin locus control region.

EMBO Journal 7 : 655-664.

Philipsen, S., Pruzina, S. and Grosveld, F. (1993) The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the β -globin locus control region. EMBO Journal 12 : 1077-1085

Phi-Van, L. and Strätling, W.H. (1988). The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. EMBO J. 7 : 655-664.

Phi-Van, L. von Kreis, J.P., Ostertag, W. and Strätling, W.H. (1990). The chicken lysozyme 5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. Mol. Cell. Biol. 10 : 2302-2307.

Proudfoot, N.J. and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. Nature 263 : 211-214.

Purrello, M. and Balazs, I. (1983) Direct hybridisation of labelled DNA to DNA in agarose gels. Anal. Biochem. 128 : 393-397

Raich, N., Clegg, C. H., Groffi, J., Roméo, P.H. and Stamatoyannopoulos, G. (1995) GATA1 and YY1 are developmental repressors of the human ϵ globin gene. EMBO Journal 14 : 801-809

Reed, R. and Maniatis, T. (1986). A Role for exon Sequences and Splice-Site Proximity in Splice-Site Selection. Cell 46 : 681-690.

Rhui, H., Kirken, R.A. and Farrar, W.L. (1994). Activation of receptor associated tyrosine kinase JAK 2 by prolactin. J. Biol. Chem. 269: 5364-5368

Robertson, E., Bradley, A. Kuehn, M. and Evans, M., (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature 323 : 445-447.

Robertson, G. , Garrick D., Wu, W., Kearns, M., Martin, D. and Whitelaw, E. (1995). Position-dependent variegation of globin transgene expression in mice. Proceeding of the National Academy of Science USA. 92 : 5371-5375.

Rohan, R.M., King, D and Frels, W.I. (1990). Directing sequencing of PCR - amplified junction fragments from tandemly repeated genes. Nucleic Acids Research. 18 : 6089-6095.

Romig, H. Fackelmayer, F.O., Renz, A., Ramsperger, V. and Richter, A. (1992). Characterisation of SAF-A a novel nuclear DNA binding protein, from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. EMBO J. 11 : 3431-3440.

Roth, D. and Wilson, J. (1988). Illegitimate recombination in mammalian cells in: Kucherlupati, R. and Smith, G.R. (eds) Genetic Recombination pp 621-654, ASM,

Washington, D.C.

Sanger, F., Nicklen, S. and Coulson A.R. (1977) DNA sequencing with chain - terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 : 5463-5467

Sap, J. de Magistris, L., Stunnenberg, H. and Vennstrom, B. (1990). A major thyroid hormone response element in the third intron of the rat growth hormone gene. EMBO J. 9 : 887-896.

Schedl, A., Montoliu, L., Kelsey, G. and Schutz, G. (1993). A yeast artificial chromosome covering the tyrosinase gene confers copy number dependent expression in transgenic mice. Nature 362 : 258-260

Schmidhauser, C., Casperson, G.F., Myers, C.A., Sanzo, K.T., Bolten, S. and Bissel, M.J. (1992). A novel transcriptional enhancer is involved in the prolactin-and extracellular matrix dependent regulation of the β -casein gene expression. Mol. Cell. Biol. 3: 699-709

Schmitt-Ney, M., Doppler, W., Ball, R.R., and Groner, B. (1991). β -casein gene promoter activity is regulated by the hormone mediated relief of transcriptional repression and a mammary gland specific factor. Mol. Cell. Biol 11: 3745-3755

Shani, M., Barash, I. Nathan, M., Ruia, G., Searfoss, G.H., Dekel, I., Fearman, A., Givol, D. and Hurwitz, D.R. (1992). Expression of human serum albumin in the milk of transgenic mice. Transgenic Research 1 : 195-208.

Sharpe, J.A., Wells, D.J., Whitelaw, E., Vyas, P., Higgs, D.R. and Wood, W G. (1993). Analysis of the human α -globin gene cluster in transgenic mice. Proc. Natl Acad Sci. USA 90, 11262-11266

Shinnick, T.M. Lund, E., Smithies, O. and Blattner, F.R. (1975) Hybridisation of labeled RNA to DNA in agarose gels. Nucl. Acids Res. 2 : 1911-1929

Simons, J.P., McClenaghan, M. and Clark, A.J. (1987). Alteration of the quality of milk by expression of sheep β -lactoglobulin in transgenic mice. Nature 328: 530-532

Simons, J.P., Wilmot, I., Clark, A. J., Archibald, A.L., Bishop, J.O., and Lathe, R. (1988). Gene transfer into sheep. Bio/Technology 6 : 179-183

Sippel, A.E, Saueressig, H., Winter, D., Grewal, T., Faust, N., Hecht, A. and Bonifer (1992) The regulatory Domain Organisation of Eukaryotic Genomes: Implications for stable gene transfer. In Transgenic Animals Eds Grosveld F. and Klollias G.

Sleep, D. Bellfield, G.P., Balance, D.J., Steven, J., Jones, S., Evans, L.R., Moir, P.D. and Goody, A.R. (1991). *Saccharomyces cerevisiae* strains that overexpress heterologous proteins. Bio/Technology 9 : 183-187.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98 : 503-517

Southern, P.J., and Berg, P. (1982). Transformation of mammalian cells to antibiotic

resistance with a bacterial gene under control of the SV40 early region promoter. *Journal of Molecular and Applied Genetics* 1 : 327-341.

Srouboulis, J., Dillon, N. and Grosveld, F. (1992). Developmental regulation of a complete 70 kb human β -globin locus in transgenic mice. *Genes and Development* 6: 1857-1864

Stacey, A. Schnieke, A., McWhir, J., Cooper, J., Coleman, A., Melton, D.W. (1994). Use of double-replacement gene targeting to replace the murine α -Lactalbumin gene with its human counterpart in embryonic stem cells and mice. *Mol. Cell Biol.* 14 : 1009-1016.

Stamatoyanopoulos, J. A., Goodwin, A., T. and Lowrey, C. (1995) NF-E2 and GATA binding motifs are required for the formation of DNase 1 hypersensitive site 4 of the human β -globin locus control region. *EMBO J.* 14 : 106-116

Steif, A., Winter, D.M., Strätling, W.H. and Sippel, A.E. (1989). A nuclear DNA attachment element mediates elevated and position independent gene activity. *Nature* 341 : 343-345.

Stokes, D.G. and Perry, R.P. (1995). DNA binding and chromatin localisation properties of CHD1. *Mol. Cell Biol.* 15 : 2745-2753.

Swain, J.L., Stewart, T.A. and Leder, P. (1987). Parental legacy determines methylation and expression of an autosomal transgene: A molecular mechanism for parental imprinting. *Cell* 50: 719-727

Swanson, M.E., Marlin, M.J., O'Donnell, K. Hoover, K., Lago, W., Huntress, V., Pamous, C.T., Pinkert, C.A., Pilder, S. and Logan, J.S.. Production of functional human hemoglobin in transgenic swine. *Bio/Technology* 10 : 557-560.

Talbot, D., and Grosveld, F. (1991) The 5'HS2 of the globin locus control region enhances transcription through the interaction of multimeric complex binding at two functionally distinct binding sites. *EMBO J.* 10 : 1391-1398

Taramelli, R., Kioussis, D., Vanin, E., Bartram, K., Groffen, J., Hurst, J. and Grosveld, F. (1986) $\gamma\delta\beta$ -Thalassaemias 1 and 2 are the result of a 100kb deletion in the human β -globin cluster. *Nucleic Acids Research* 14 : 7017-7029

Thomas, K.R. and Capecchi M.R. (1987). Site directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51 : 503-512.

Thompson, E.M., Christians, E., Stinnakre, M-G. and Renard, J-P. (1994). Scaffold attachment regions stimulate HSP70.1 expression in mouse pre-implantation embryos but not in differential tissues. *Mol. Cell Biol.* 14 : 4694-4703.

Threadgill, D.W. and Womack, J.E. (1990). Genomic analysis of the major bovine milk protein genes. *Nucl. Acids Res.* 18: 6935-6942

Topper, T.Y. and Freeman, C.S. (1980). Multiple hormone interactions in the

developmental biology of the mammary gland. *Physiol. Rev.* 60: 1049-1106

Townes, T. M., Lingrel, J. B., Chen, H. Y., Brinster, R. L. and Palmiter, R. D (1985) Erythroid specific expression of human β -globin genes in transgenic mice *EMBO J.* 4 : 1715-1723

Tsao, S.G.S., Brunk, C.F. and Pearlman, R.E. (1983) Hybridisation of nucleic acids directly in agarose gels. *Anal. Biochem.* 131: 365-372

Tuan, D., Solomon, W, Li, Q. and London, M. (1985). The ' β -like globin' gene domain in human erythroid cells. *Proceedings of the National Academy of Sciences USA.* 82 : 6384-6388.

Tucker, A.H. (1988). Lactation and its Hormonal Control: In "The Physiology of Reproduction, Eds, Knobil, E. and Neill, J. et al. Raven Press, Ltd., New York.

Uvardy, A., Maine, E., and Schedl, P. (1985). The 87A7 chromomere : identification of novel chromatin structures flanking the hear shock locus that may define the boundaries of higher order domains. *Journal Molecular Biology* 185 : 341-358.

Van der Ploeg, L. H. T., Konings, A., Oort, M., Roos, D., Bernini, I. and Flavell, R. A. (1980) *Nature* 283, 637-642

Velander, W.H., Johnson, J.L., Page, R.L., Russel, C.G., Subramanian, A., Wilkins, T., Gwazdauskas, F.C., Pittius, C. and Drohan, W.N. (1992). High-level expression of a heterologous protein in the milk of transgenic swine using the cDNA encoding human protein C. *Proc. Natl. Acad. Sci. USA* 89: 12003-12007

von Kreis, J.P. Buhrmester, H. and Strätling, W.H. (1991). A matrix/scaffold attachment region binding protein : identification, purification and mode of binding. *Cell* 64 : 123-135.

Wakao, H., Gouilleux, F. and Groner, B. (1994). Mammary gland factor (MGF) is a novel member of a cytokine regulated transcription factor gene family and confers prolactin response. *EMBO J.* 13: 2182-2191

Wake , C.T. Vernaleone, F., and Wilson, J.H.(1985). Topological requirements for homologous recombination among DNA molecules transfected into mammalian cells. *Mol. Cel. Biol* 5 : 2080-2089.

Wall, R.J. and Seidel, G.E. (1992). Transgenic farm animals - a critical analysis. *Theriogenology* 38: 337-357

Wallace, R.B. and Miyada C.G. (1987). Oligonucleotide probes for the screening of recombinant DNA libraries. *Methods Enzym.* 152 : 432-442

Watson, C.J., Gordon, K.E., Robertson, M. and Clark, A.J. (1991). Interaction of DNA binding proteins with a milk protein gene promotor *in vitro*: identification of a mammary gland specific factor. *Nucl. Acids Res.* 19: 6603-6610

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. and Weiner, A.M. (Eds)

(1987). *Molecular Biology of the Gene*. Benjamin/Cummings Publishing Company, Inc.

Weatherall, D. J (1993) *Thalassemia*, in *The Molecular Basis of Blood Diseases* (Stamatoyannopoulos, G., Nienhuis, A. W., Majerus, P. W. and Varmus, H. (eds) W.B. Saunders)

Weatherall, D. J (1995) The molecular basis for phenotypic variability of the common thalassaemias. *Molecular Medicine Today* 1 : 15-20

Weintraub, H. and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. *Science* 193 : 848-856.

Whitelaw, C.B.A. Archibald, A., Harris, S., M^cClenaghan, M., Simons, J.P. and Clark, A.J. (1991). Targeting expression of the mammary gland : intronic sequences can enhance the efficiency of gene expression in transgenic mice. *Transgenic Research* 1 : 3-13.

Whitelaw, C.B.A. Harris, S., M^cClenaghan, M., Simons, J.P. and Clark, A.J. (1992). Position independent expression of the ovine β -lactoglobulin gene in transgenic mice. *Biochem. J.* 286 : 31-39.

Wilkie, T. M., Brinster, R. L. and Palmiter, R. D. (1986) Germline and somatic mosaicism in transgenic mice. *Developmental Biology* 118 : 9-18

Wilmot I. and Campbell, K. (). Bovine embryo multiplicatin: progress and challenges

Wilson, C., Bellen, H.G. and Gehring, W.J. (1990). Position effects on eukaryotic gene expression. *Annu. Rev. Cell Biol.* 6 : 679-714

Wilson, J.H., Berget, P.B. and Pipas, J.M. (1982). Somatic cells efficiently join unrelated DNA segments end to end. *Mol. Cell. Biol.* 2 : 1258-1269

Wright, G., Carver, A. Cotton, D., Reeves, D., Scott, A., Simons, P., Wilmot, I., Garner, I. and Coleman, A. (1991). High level expression of active human alpha-1-anti-trypsin in the milk of transgenic sheep. *Bio/Technology* 9 : 830-834

Yamamura, K., Miki, T., Suzuki, N., Ebihara, T., Kawai, K., Kumahara, Y. and Honjo, T. (1985). Introduction of mouse CE genes into C057 cells and fertilized mouse eggs. *Journal of Biochemistry* 97 : 333-339.

Yull, F., Harold, G., Wallace, R., Cowper, A., Percy, J. Cottingham, I. and Clark, A.J. (1995). Fixing human factor IX (fIX) : Correction of a cryptic RNA splice enables the production of biologically active fIX in the mammary gland of transgenic mice. *Proceedings of the National Academy of Sciences USA* 92 : 10899-10903.

Yull, F.E., Wallace, R.M. and Clark A.J. (1995). Feasibility of Adapting the *E.coli lac* Operator/Repressor System for use as and Inducible Switch in the Liver of Transgenic Mice. *Transgenics* 1 : 477-486

Zhao, K. Hart, C.M. and Laemmli, U.K. (1995). Visualisation of chromosomal

domains with boundary element-associated factor BEAF 32. Cell 81 879-889.