Long range analysis of the mammalian casein locus

Anthony M. Tomlinson

A thesis presented for the degree of Ph.D.
Institute of Cell and Molecular Biology
University of Edinburgh
September 1999

Declaration

The work presented in this thesis is my own, unless otherwise stated. The experiments were designed by myself in collaboration with my supervisors Dr. Michael Dalrymple, Dr. Ian Garner, Dr. Angelika Schnieke and Prof. David Melton.

Abstract

Current notions of genome organisation hold that chromatin is structurally and functionally partitioned into domains delimited by specialised chromatin elements. In transgenic studies a variety of chromatin domains and functional elements isolated from them have been shown to direct physiological levels of gene expression independent of the integration position of the transgene within the genome. This would make them useful components of transgene constructs which are intended to give reliably high levels of tissue-specific gene expression. Since most chromatin domains are larger than can be accommodated by conventional cloning vectors, the advent of systems for handling very large DNA molecules *in vitro* has greatly facilitated this work. In particular, yeast artificial chromosomes (YACs) have proved useful in the discovery and study of chromatin domain elements such as insulators, locus control regions and nuclear scaffold attachment regions.

The principal aim of this project was to assay for position independent, copy number dependent and tissue- and developmental stage-specific gene expression from the human and murine casein gene loci in transgenic mice. Two overlapping YACs covering the murine casein gene locus were restriction mapped, generating the first reported physical map of this locus. The order of the five casein genes is α–β–γ–ε–κ; as in other mammals the κ casein gene, though evolutionarily unrelated to the other members of the locus, is closely linked, which may imply that casein gene expression is under locus control. The YACs are collinear within the region of overlap, suggesting that neither is rearranged. One of these YACs was manipulated with the intention of inserting a reporter gene under the control of an ovine β-lactoglobulin promoter downstream of the murine β casein polyadenylation site, so that gene expression from the YAC could be detected against a murine background in cell culture or transgenic mice. In parallel, a YAC bearing the human casein locus was used to establish transgenic mice. Unexpectedly, none expressed detectable levels of human casein RNA when analysed by northern blot, although four produced human κ casein RNA that was detectable by reverse transcriptase polymerase chain reaction, indicating basal levels of transcription of the gene. The reasons for this are unknown, though initial characterisation of genomic DNA from the transgenic animals suggests that the YAC is not intact in any of the eight lines. The human casein YAC was also transfected into HC11 murine mammary epithelial cells with a view to assessing the level of human casein gene

expression from the YAC *in vitro,* but unfortunately no stably transfected cell lines were established. These results and their implications for the use of large DNA constructs in transgenic mammals are discussed.

Acknowledgements

Firstly I must thank my supervisors Mike Dalrymple, Angelika Schnieke, Ian Garner and David Melton for their support and patience throughout this project. Ian Garner in particular offered constructive criticism, support, encouragement and the occasional tab at all the right moments. I am also very grateful to the staff and management of PPL Therapeutics Ltd. for the opportunity to pursue this PhD, and to my labmates past and present for putting up with me, especially when I was writing up. You can have your computers back now. There is a list of very important people that all deserve a mention here, though more because they're my friends than for any trivial reasons such as material contributions to this work. In no order at all they are Geoff, Al, Ste, Paul, John, Miriam, Peet, Gordon('s), Celia, Simon, Jose, Sarah, Tim, Jenny, Sebastienne, Tim, Brian, Byro and Julian. Cheers. Special mention must go to my ex-flatmates Miriam, Julian and Celia for that fantastic year when I sorted myself out and learned advanced dinner party etiquette at Miriam's knee. We were cool.

Serious bits: thanks to Sarah for helping me with the thesis and lots of other things, and for being so diverting. Thanks too to the Lads for crisis care and toxicity management when I needed it (not to mention outrageously salubrious dosshouse facilities), and for continuity of care despite my fickle dalliances with wives, careers and the like. Not so serious bits: no navel-gazing episode would be complete without mention of the monster, the black bomb, the tractor, the binned-it and Cliffhanger – all had their day, and all made a lasting impression on me (some more literally than others).

Among the many great folk at PPL there are some that stand out. Andy Carver and Mike Dalrymple have sadly left along with Jayne Wu (née Keenan) and Anona Cranston (née Turnbull), but in times gone by they showed me the nobility of labour in quite unexpected ways. Still hanging on are other stalwarts like Ann Scott, Liz Emslie and Simon Temperley, who can still recall the great days when the mood is on them. Special thanks to Sebastienne, Chris and Jenny for valued friendship. Thanks also to those without whom this thesis would have been much worse. Katrina Gordon tried valiantly to teach me cell culture, Yvonne Gibson kept track of all of my mice far better than I did, and Chris Coles kept me well supplied with reading material, snide humour and Square Crisps. Ian Garner gave staunch unwavering support while I

was writing up, without which I would have been sunk, and the staff and students of the Wellcome Trust 19th Summer School convinced me that there was still some merit in molecular biology after all.

I also want to thank my parents. Mum taught me that I was clever and Dad taught me that I wasn't stupid (and to weld, of course). Thank you both. You were right though - I'd probably have been happier in the Saracen's.

Finally, thank you Karla. Rah rah PhD, yahda yahda, an' all that jazz.....

Abbreviations

°C	degrees Celcius
µg	microgramme
µl	microlitre
BLG	β-lactoglobulin
bp	base pairs
cDNA	complementary DNA
d	days
DNA	deoxyribonucleic acid
DNaseI	deoxyribonuclease I
dNTP	2'-deoxy (N) 5'-triphosphate; N can be adenosine (A), cytidine (C), guanosine (G) or thymidine (T)
EDTA	ethylenediaminetetra-acetic acid
fg	femtogramme
g	gramme
h	hours
kb	kilobase pairs
l	litre
LCR	locus control region
M	molar
MAR	matrix attachment region
mg	milligramme
min	minutes
ml	millilitre
MOPS	3-(N-morpholino)-propanesulphonic acid
MW	molecular weight

N	normal (molar equivalent)
NaOAc	sodium acetate
ng	nanogramme
NH_4OAc	ammonium acetate
nl	nanolitre
nt	nucleotides
OD_n	optical density$_{\text{wavelength in nanometres}}$
pg	picogramme
pl	picolitre
RNA	ribonucleic acid
RNase A	ribonuclease A
rpm	revolutions per minute
rRNA	ribosomal RNA
SAR	scaffold attachment region
SDS	sodium dodecyl sulphate
sec	seconds
TEMED	N, N, N', N' tetramethylethylenediamine
tris	tris (hydroxymethyl) aminoethane
tRNA	transfer RNA
UHP	ultra high purity (water)
w/v	weight per unit volume
WAP	whey acidic protein
xg, g	times the force of gravity

Figures

Tables

Contents

to whom I owe the leaping delight

1. Chapter 1 - Introduction

1.1 Transgenic animals as tools for research and protein production

The use of transgenic animals in research and biotechnology is now well established (reviewed by Clark *et al.* 1992). The core technology consists of the introduction of a cloned DNA construct into the germ line of the animal, either by gene targeting in embryonic stem cells, transfer of transfected cells into enucleated oocytes (nuclear transfer) or, more commonly, by pronuclear injection of a fertilised egg (reviews: Palmiter and Brinster, 1986; Moreadith and Radford, 1997; Wolf *et al.* 1998). Transgenic animals of a number of species have now been described including mice, rats, rabbits, sheep, goats, cattle and pigs. Transgenic mice have generally been the research tool of choice while transgenic livestock have been used for agricultural strain improvement and the large-scale production of therapeutic proteins, notably in the milk (reviewed by Clark, 1992). The potential of this system is widely recognised, and therapeutic proteins have already been expressed at economically-viable levels in the milk of transgenic sheep (reviewed by Clark *et al.* 1992a). Compared to *in vitro* cell-based sources transgenic animals produce protein which is relatively pure, appropriately modified and efficiently secreted. Therapeutic proteins are also purified from natural human sources but in many cases the protein is present at low concentration in the source material and purification is therefore prohibitively expensive or impossible. In addition the risk of disease transmission, especially from human blood products, is significant. Transgenic animals might circumvent all of these problems by producing large quantities of easily purified human protein in a disease-free environment at a very low cost.

The use of transgenic animals has greatly advanced our understanding of gene expression, and in particular the role of chromatin structure in controlling transcription. Only in whole animals can the full complexity of gene control by developmental stage, tissue type and hormonal action be studied. Only whole animals undergo embryogenesis, which has important implications for the expression of a number of genes. The remodelling of chromatin during development of a differentiated cell type is not replicated in tissue culture, where introduced DNA encounters a fully differentiated chromatin structure immediately. This may be important in such phenomena as position effect variegation (reviewed by Karpen, 1994), the role of introns in transgene expression (Brinster *et al.* 1988) and the opening of chromatin by locus control regions (LCRs;

reviewed by Kiousiss and Festenstein, 1997). In addition, cell culture does not necessarily replicate the full range of gene expression seen in the tissue, especially when cell-cell interactions are important controls, as is the case in the mammary gland, for example.

1.1.1 Generation of transgenic mice by pronuclear microinjection

Since the pioneering work of Palmiter and Brinster (reviewed by Palmiter and Brinster, 1986) this technology has become widespread. It is by far the most common means of generating transgenic mice and other animals, and technical manuals are available (*e.g.* Hogan *et al.* 1986). DNA is usually injected into the pronuclei of fertilised mouse eggs as pure linear molecules in a low salt buffer. The eggs are then transferred into the oviduct of day 1 pseudopregnant foster mice and carried to term. The pups are screened for the presence of the transgene, usually by PCR and/or Southern blot of tail biopsy DNA. Typically, 10 % of live births are transgenic (reviews: Palmiter and Brinster, 1986; Pinkert, 1994), although much higher and lower frequencies may be observed.

Among the poorly understood phenomena associated with transgenesis the most striking is the mechanism of integration of the transgene into the host genome as tandem arrays of between 1 and over 200 copies. Integration rarely occurs at more than a single site which suggests that the introduction of a break in the genome may be the rate limiting step. The process by which the injected DNA forms tandem arrays is not understood. A model has been proposed by Bishop and Smith (1989) in which injected DNA fragments are circularised in the egg and homologous recombination with other linear molecules extends the circle to a large circular head-tail concatemer. The circular array may then be broken randomly and incorporated opportunistically into the genome at a chromosomal break. This model accounts for tandem arrays and for the rarity of integration, because chromosomal breaks are rare. However; it does not explain why linear molecules are integrated preferentially over circular molecules (reviewed by Palmiter and Brinster, 1986). In addition the prediction of the model that the end copies of a transgene array will be incomplete is not borne out by direct cloning of transgene-transgene and transgene-genome junctions (A. Turnbull, personal communication; McFarlane and Wilson 1996; A. Tomlinson and M. Dalrymple, unpublished results).

1.1.2 Transgenic livestock

There are several technical and logistical problems associated with extending the pronuclear microinjection technology from mice to livestock. Typically, low numbers of ova are obtained from superovulated livestock and so high numbers of animals must be superovulated and subjected to rapid surgery to obtain sufficient numbers of eggs. Bovine ova have been obtained from slaughterhouses and matured and fertilised *in vitro*, but the viability of embryos produced in this way is low (Krimpenfort *et al.* 1991; summarised by Clark *et al.* 1992a). Large animals rarely carry more than twins, although pigs are a notable exception, and pregnancy rates are generally poor, so large numbers of recipients are usually needed. Lastly, the generation time of large animals is very long in research terms, often involving a wait of many months between microinjection and phenotypic analysis. These factors and the extremely high cost of farm animal husbandry on this scale compared to mouse husbandry mean that transgenic farm animals represent a major undertaking in time and resources (reviews: Clark *et al.* 1992; Wall and Seidel, 1992; Pursel and Rexroad, 1993). However, the potential benefits to medicine and agriculture are sufficient to warrant continued research in this area. To date, transgenic goats, sheep, pigs and cows have been generated (*e.g.* Clark *et al.* 1989; Wall *et al.* 1991; Ebert *et al.* 1991; Krimpenfort *et al.* 1991), and three therapeutic proteins produced in transgenic animals are currently undergoing clinical trials. Human α–1 antitrypsin (AAT) has been expressed in transgenic sheep (Wright *et al.* 1991) and is currently in Phase IIb clinical trials as a treatment for lung damage in victims of cystic fibrosis. Human antithrombin III has been produced in transgenic goats (Edmunds *et al.* 1998) and is currently in clinical trials as an adjunct to anticoagulant treatment during coronary artery bypass grafting (Scrip 2192, 1996). Human α–glucosidase, produced in transgenic rabbits (Bijvoet *et al.* 1998), is currently in clinical trials as a treatment for Pompe's disease (Scrip 2328, 1998). All of these proteins are secreted into the milk of the transgenic animal.

1.1.3 Gene targeting in mice

The generation of mouse clones from cultured embryonic stem (ES) cells is an established technology (Thompson *et al.* 1989; reviewed by Moreadith and Radford, 1997). ES cells lines are pluripotent, may be stably propagated in culture and when microinjected into an early embryo can contribute to all tissues of the resultant chimera. DNA may be introduced and

randomly integrated into ES cells as an alternative to the microinjection of DNA into zygotes. This has been used to introduce YACs into the mouse germline (Jakobovits *et al.* 1993). However, the greatest potential of ES cells lies in their amenability to modification by homologous recombination (HR; Capecchi, 1989), which has facilitated the generation of genetically modified mice by gene targeting (reviewed by Brandon *et al.* 1995). The targeting construct is transfected into ES cells and transfected cells are screened for the HR event. Positive cells are microinjected into recipient blastocysts which are then transferred into foster animals and carried to term. The resultant chimeras are bred to establish germline transmission of the desired genotype. This technology has mostly been used to generate loss of function or "knockout" mutations, but has also been used to introduce genes into specific loci, make defined changes in genes or control sequences and to replace one gene with another. This latter procedure involves two targeting events, inserting the hypoxanthine phosphoribosyltransferase (HPRT) counterselectable marker as an intermediate which is removed in the second targeting event, leaving only the intended alteration to the mouse genome (Reid *et al.* 1990).

Thus, gene targeting in murine ES cells is a powerful technique in that it allows predetermined changes to be made to the mouse genome which may not readily be achieved by any other means. However, gene targeting is time-consuming compared with pronuclear microinjection. In addition, expression of a transgene may be improved as a result of the multimerisation that often accompanies random transgene integration. Nevertheless, for investigation of the subtleties of gene expression at natural loci and for definitive loss of function studies, gene targeting in mice remains the paradigm. There are efforts to develop ES cells for other species which have been partially successful; rat (Iannaccone, 1994) and pig (Wheeler, 1994) chimeras have been reported. However, germline transmission has not been demonstrated in either case. Thus, to date this technology is confined to mice.

1.1.4 Cloning by nuclear transfer

Another method for the introduction of exogenous DNA into the germline of mammals has recently emerged with the generation of viable sheep (Campbell *et al.* 1996; Wells *et al.* 1997; Schnieke *et al.* 1997; Wilmut *et al.* 1997), cattle (Cibelli *et al.* 1998), and mice (Wakayama *et al.* 1998) by nuclear transfer from cultured primary cells. Cloning was first achieved in mammals

by embryo splitting (Willadsen, 1979) but separation of an embryo into more than two parts gives drastically reduced viability (reviewed by Wolf *et al.* 1998). Nuclear transfer advanced this technology because single embryonic cells could be transferred into enucleated oocytes to give viable offspring, thus generating numbers of cloned animals from a single early embryo. Eventually these cells were successfully cultured for increasing periods of time prior to nuclear transfer, and ultimately animals were derived from cells taken from proliferating culture, thus giving access to potentially unlimited numbers of cloned animals as well as the possibility of cell-mediated transgenesis in non-murine species (reviewed by Wolf *et al.* 1998).

The advantages of this technique over earlier cloning experiments using *ex vivo* embryonic cells are that it allows much larger clones of animals to be generated and that the cells may be modified in culture, generating cloned transgenic animals which would greatly reduce the wastage inherent in traditional microinjection-based transgenesis.

There are a number of further advantages that nuclear transfer might have over microinjection for the generation of transgenic animals. All of the animals derived from the same cell population would be of the same sex which would be useful for production of proteins in the mammary gland, for example. Mosaicism would be avoided and thus founder (G_0) animals could be used for production, which would save a great deal of time when dealing with large animals. Generation of large numbers of animals for production purposes would be simpler and faster than by normal means, and the genetic stability of the transgene would not be a factor if the production animals were all founders. Screening of transfected cell lines might reduce the number of animal lines which have to be produced to obtain high expression and maximise the potential of a given construct. Ultimately, gene targeting followed by nuclear transfer could allow work in other species to parallel the ES cell-based work in mice. The first transgenic cloned mammals have already been produced (Schnieke *et al.* 1997; Cibelli *et al.* 1998), proving that at least in principle this technology is viable.

However, there are many technical problems associated with nuclear transfer at present. Transfer of cells from adult somatic tissue is possible (Wilmut *et al.* 1997) but in practice there should be no need to use adult tissue to establish suitable cell lines. Nuclear transfer from foetal cells is more straightforward but still relatively inefficient compared to embryo transfer

following microinjection. The value of screening for transgene expression in cultured cells would far outweigh this problem but to date this is not feasible for genes which are not normally expressed in those cells. Work is underway to establish cell lines which may mimic terminally differentiated tissue such as the mammary epithelium, whilst remaining amenable to nuclear transfer (Gordon *et al.* in press). The combination of HR and nuclear transfer which would allow gene targeting experiments to be carried out in non-murine mammals is a distant goal at present, but would be of considerable use in both research and medicine. For example, deletion of porcine cell-surface markers might overcome hyperacute rejection of porcine-human xenografts.

1.2 Transgene expression

1.2.1 Position effects

Expression levels of most transgenes are low with respect to the endogenous gene, vary over several orders of magnitude between transgenic lines, are unrelated to copy number and may be spatiotemporally inappropriate (reviewed by Brinster *et al.* 1986). These results have been ascribed to influences of the chromatin surrounding the site of integration, which had been named the position effect in earlier studies of natural translocations and integrations, particularly in *Drosophila* (reviewed by Udvardy, 1999). The subsequent discovery of sequences which could overcome these effects has greatly advanced the study of genome structure, particularly the involvement of chromatin in the control of gene expression.

The lack of correlation between transgene expression and copy number is most often cited as the definitive evidence for a position effect. A position effect may be more accurately defined as a change from the expected transgene expression level due to its position in the genome. This could include an increase in transcription as well as a decrease, and also a change in the pattern of expression. In most cases in fact, the exact expression level and pattern of a transgene cannot be precisely predicted, because position effects can never be ruled out. However, by generating many lines of transgenic animals, the general expression pattern of a transgene can be estimated.

Position effects are classically due to translocations of endogenous genes and here it is straightforward to describe the change in expression. In this context, a position effect is

diagnosed when a chromosomal rearrangement causes a change in the expression pattern of a gene without perturbation of the transcription unit or proximal promoter, although this may be difficult to demonstrate definitively if the gene has not been fully characterised. For example, Nishio *et al* (1994) discovered an additional exon of the dystrophin gene some 500 kb 5' to the previously described promoter.

Position effects have been studied extensively in *Drosophila.* The P transposon system allows single copy transgenics to be made and for the transgene to transpose inducibly. This facilitates the study of single copies of the transgene in multiple integration sites. Multiple copy genes are frequently silenced in *Drosophila,* and indeed a similar system has been proposed to explain the generally lower levels of gene expression per copy that are seen in multiple copy transgenic vertebrates (Dorer, 1997). As yet there is little concrete evidence to support this hypothesis.

Stable position effects are thought to occur when by translocation or transgenesis a gene is separated from some of its cognate control elements, or brought under the influence of newly-juxtaposed elements, and thus fails to express efficiently. The loss of elements which protect the gene from neighbouring chromatin effects might result in silencing or inappropriate activation. Alternatively the effect of a gene's cognate enhancer might be diluted by competition from a newly juxtaposed promoter for the enhancer activity. Finally, if a gene is normally activated by an LCR (see below), removal of the LCR might result in a combination of inappropriate activation and general silencing due to the loss of a chromatin-opening function (reviewed by Higgs, 1998). Stable position effects lead to reproducible expression between animals of the same transgenic line.

Position effect variegation (PEV) has been extensively studied in yeast and *Drosophila.* It is characterised by a variation in expression of the transgene, or translocated gene, between animals of the same line, and also between cells within the same tissue. PEV is thought to occur when a gene is introduced near to heterochromatin such that the heterochromatin may engulf and silence the gene in some cells but not others. At some point in development the heterochromatinisation of a cell lineage appears to be fixed, because in variegated tissue clonal inheritance of expression characteristics is observed. Interestingly, two *Drosophila* genes have been moved from heterochromatin to euchromatin, and display PEV in their new locations

(reviewed by Kleinjan and van Heyningen, 1998). This suggests that heterochromatin may not simply be a stochastically silent transcriptional region but may in fact be a fundamentally different transcriptional area of the genome, possibly mediated by the physical separation of heterochromatin and euchromatin within the interphase nucleus (reviewed by Karpen, 1994).

PEV has not been definitively demonstrated in vertebrates,. There are several reports which describe mosaicism in the expression of transgenes and invoke PEV as a probable cause (Pretcheva *et al.* 1994; Dobie *et al.* 1996; Robertson *et al.* 1995) but not all of these transgenes are inserted into obvious heterochromatin. In *Drosophila* tandem gene arrays tend to become heterochromatinised (Dorer and Henikoff, 1994) and this has been suggested to occur in vertebrates (Dobie *et al.* 1996, Dorer, 1998). However, there are few supporting data at present and the occasionally described reduction in expression per copy could in most cases be due to titration of *trans*-acting factors. Davis and McDonald (1988) used double transgenics and nuclear run-on analyses to show that the average rate of transcription per gene decreased with increased copy number of an elastase I (EI) transgene, while 250 copies of this gene had no deleterious effect on the expression of a growth hormone transgene driven by the same EI promoter at a different integration site. However, although this study demonstrated a lack of competition for the enhancer; there may be other essential factors that bind elsewhere in the EI transgene which are titrated in multiple copy lines.

Dobie *et al* (1996) described PEV in transgenic mice carrying an ovine β–lactoglobulin (BLG) transgene. Here there was a clear correlation between apparent PEV and pericentromeric integration, as shown by fluorescent *in situ* hybridisation (FISH). In this paper it was reasoned that the transgene repeats themselves might attract heterochromatinisation similarly to repeated sequences in *Drosophila* (Dorer and Henikoff, 1994). Elliot *et al* (1995) described an apparent PEV when transgenic mice were generated carrying a human CD2 transgene including the LCR (Greaves *et al.* 1989) linked to the immunoglobulin heavy chain enhancer Eμ. Normally the CD2 transgene would be expressed position independently and copy number dependently, but in the presence of Eμ some lines displayed PEV. In the silenced loci the chromatin was found to be resistant to nuclease digestion, specifically lacking the hypersensitive sites characteristic of the CD2 LCR. In addition to demonstrating that the effects of the LCR may be overcome by the presence of other elements *in cis* and therefore that the LCR may not give truly position-

independent expression, these results have also been interpreted to support the hypothesis that heterochromatinisation may be potentiated by transgene repeats in vertebrates (Dorer, 1998). However, two of the five PEV lines are described as low copy number and one is said to be two copy. At least one of the non-PEV lines is also two copy. Thus, transgene copy number alone has no proven correlation with PEV in this experiment. It is possible that additional factors such as the position of integration have an effect on the probability of gene silencing in the variegating mice.

Whatever its aetiology, and whether or not it is analogous to PEV in *Drosophila,* variegated transgene expression in vertebrates has been unequivocally demonstrated and could pose major practical difficulties in the application of transgenesis for medicine. For example, in a mouse model of Duchenne muscular dystrophy (DMD), gene therapy by transgenesis was impaired by mosaic expression of the transgene (reviewed by Dorer, 1998). In the remodelling of porcine cell surface markers to render pig organs suitable for transplantation into humans, mosaicism could mean that a proportion of the cells of the organ remain antigenic and result in rejection. In addition, the recent reports connecting vertebrate PEV with LCR function and chromatin dynamics (reviewed by Kioussis and Festenstein, 1997) suggest that this phenomenon might soon be incorporated into the overall picture of gene expression in vertebrates, and prove to be more than a curiosity of transgene expression.

1.2.2 Introns and transgene expression

Owing to the relative ease with which they can be cloned and the occasionally prohibitive size of genomic constructs, cDNAs have often been introduced into transgenic animals. Although cDNAs are expressed efficiently in cell culture, in transgenic animals efficient cDNA expression is rare (reviewed by Palmiter and Brinster, 1986), though with some exceptions (Ebert et al. 1991; Velander et al. 1992). This has resulted in the interesting finding that introns, particularly intron 1 of several genes (Brinster *et al.* 1988; Whitelaw *et al.* 1991; Aronow *et al.* 1992; Hurwitz *et al.* 1994), can enhance the transcription of many cDNA and minigene constructs. Heterologous introns seem to function as well as homologous introns (Palmiter *et al.* 1991). The mechanism of this phenomenon is unknown, but may provide insights into chromatin remodelling at transcription as well as a practical means of enhancing transgene expression from

cDNAs. A construct based on a published ovine β–lactoglobulin (BLG) expression vector but with the first intron and non-coding exon from the bovine β–casein gene has resulted in expression of transgenic cDNAs which is comparable to that of the equivalent genomic construct (I. Garner, unpublished results).

1.3 The role of chromatin structure in gene expression

The current view of the genome is that it consists of essentially discrete transcriptional domains which are somehow prevented from interaction with neighbouring domains by specialised chromatin elements. Constructs which show position independent expression in transgenic animals are thought to be able to create discrete chromatin domains autonomously and therefore to contain all of the necessary components of such domains. Prevailing ideas on how these domains may be maintained and some of the elements which may be involved are discussed below (reviews: Sippel *et al.* 1992; Kollias and Grosveld, 1992; Sippel *et al.* 1993; Geyer, 1997; Kioussis and Festenstein, 1997; Hart and Laemmli, 1998; Udvardy, 1999). Broadly, these elements may be delineated into LCRs, insulators and structural chromatin elements.

Actively transcribed genes are invariably found to reside within a region of chromatin with increased sensitivity to DNase I digestion, which may extend far to either side of the gene. "Open" or "decondensed" chromatin is also referred to as euchromatin, contrasting with "closed", "condensed", heterochromatin. In general, euchromatin is replicated early in the cell cycle, is physically decondensed, concentrates in the interior of the nucleus, is CpG-hypomethylated and contains hyperacetylated histones, while heterochromatin generally has the opposite properties (reviewed by Higgs, 1998). In addition to the mild hypersensitivity of open chromatin to DNaseI digestion, there are smaller discrete DNase I hypersensitive sites (HSs), which are some 100 fold more sensitive to DNaseI than bulk chromatin. These are sites of protein-DNA interaction as determined directly and by the presence of consensus binding sites for DNA-binding proteins (reviews: Sippel *et al.* 1993; Higgs, 1998).

Thus, active transcription of a gene is associated with a decondensed, open chromatin configuration and the presence of DNase I hypersensitive protein binding sites. However, the causal relationship between the chromatin opening, binding of protein factors and active

transcription is open to question. In particular, whether open chromatin is a requirement for transcriptional activation or a direct topological result of active transcription has not been unequivocally resolved.

The description of the position effect led to a functional assay for complete transcription loci, because only a construct containing all of the regulatory elements associated with a gene will be expressed in a correct, position-independent, copy number dependent pattern. A number of loci have now been used to successfully reproduce this pattern of expression in transgenic animals (Grosveld *et al.* 1987; Chamberlain *et al.* 1988; Lang *et al.* 1988; Greaves *et al.* 1989; Bonifer *et al.* 1990; Aronow *et al.* 1992; Carson and Wiles, 1993; Neznanov *et al.* 1993; Schedl *et al.* 1993a; Palmiter *et al.* 1993; Diaz *et al.* 1994; Talbot *et al.* 1994; Jones *et al.* 1995; May and Enver, 1995; reviews: Higgs, 1998; Udvardy, 1999) of which the most intensively studied is the human β-globin locus (Grosveld *et al.* 1987; Blom van Assendelft *et al.* 1989). A new regulatory element, the locus control region (LCR), was described in these studies. The LCR was defined functionally as able to confer high (*i.e.* physiological), tissue specific, position independent and copy number dependent levels of expression upon a β globin transgene. A number of discrete functions have since been ascribed to LCRs (reviewed by Higgs, 1998) but broadly these may be divided into chromatin remodelling and transcriptional control. The globin LCR, for example, seems to mediate the formation of an erythroid-specific open chromatin domain and to strongly enhance the transcription of the globin genes at appropriate developmental stages. Since the β globin locus contains the archetypal LCR a summary of the key findings in this area is presented below, followed by a brief consideration of some of the chromatin remodelling functions associated with other gene loci reported in recent years.

1.4 The β globin LCR

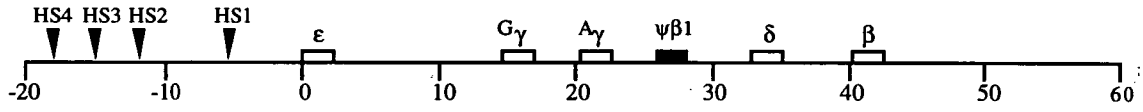
The human β–globin locus comprises 5 closely related genes, ε which is expressed in the yolk sac, γ^G and γ^A which are expressed in the foetal liver, and δ and β which are expressed in the adult bone marrow and peripheral reticulocytes; there is also a nonexpressed pseudogene (Figure 1.1; reviews: Maniatis *et al.* 1981; Kollias and Grosveld, 1992; Martin *et al.* 1996). The genes are located at the locus in order of their expression during development (Figure 1.1), and are all transcribed from the same strand. Many natural mutations in the β globin locus have been

described which perturb the normal patterns of expression and lead to clinical disorders including sickle-cell anaemia and various βthalassaemias. Analysis of one such lesion (Hispanic thalassaemia) which deleted 35 kb of DNA approximately 15 kb upstream of the locus led to the description of the LCR. The LCR consists of 4 sites, HS1 to HS4, which are extremely hypersensitive to DNase I and are located approximately 12 to 20 kb upstream of ε globin (Figure 1.1). These sites are developmentally stable but are only present in erythroid cell lineages. In addition, there is a constitutive hypersensitive site, 5' HS5, located further upstream, which has been reported to have an insulator function (Li and Stamatoyannopoulis, 1994; Chung *et al.* 1993; reviewed by Geyer, 1997). Recently the existence of a sixth DNaseI hypersensitive site (HS6) has been reported in mice (Bender *et al.* 1998) but this does not appear to be conserved in humans. In erythroid cells the globin locus resides in a region of open chromatin which extends 3' from HS5 for approximately 200 kb and appears to be dependent upon the LCR, as suggested by transgenic studies and naturally occurring mutations (Kioussis *et al.* 1983; Grosveld *et al.* 1987; Forrester *et al.* 1990).

Following the report by Grosveld *et al* (1987) that a construct based on the human β globin locus including the LCR was expressed in a correct, position independent pattern in transgenic mice, several groups began dissecting the structure and function of HS1 to HS4 (reviews: Kollias and Grosveld, 1992; Grosveld *et al.* 1993; Engel, 1993; Wood, 1996; Martin *et al.* 1996; Kioussis and Festenstein, 1997; Higgs, 1998; Udvardy, 1999). In general, studies have involved constructs containing subsets of the LCR in attempts to dissect the chromatin opening function from the transcriptional activator function, and to define minimal sequences required for each. Broadly, these attempts were unsuccessful until very recently, although some useful insights were gained into the LCR function. There are numerous conflicts in the data, which mainly seem to depend on the exact constructs and criteria employed. For example, Ellis *et al* (1993) suggest that deletion of the GATA-1 or USF transcription factor binding sites in a 215 bp synthetic HS2 construct has no significant effect on the expression of a linked β globin gene, but using a 1.9 kb HS2 fragment Caterina *et al* (1994) demonstrate that deletion of these sites leads to a 75 % diminution in β globin mRNA levels. It seems that in isolation a given LCR element may exhibit different effects from those observed in a more natural context. In addition, the spatial organisation of the control elements may be of functional significance. Finally, although the correlation of DNaseI hypersensitivity with LCR function cannot be denied, there is no

reason to suppose that all of the functionality of the LCR is confined to the hypersensitive sites, since no report has demonstrated full LCR activity from a miniconstruct containing only HSs 1-4 without any intervening DNA. In this light, it would seem prudent to view with caution all results obtained by deletion and mutation of small portions of the β globin LCR, since the interactions between 'core' sequences and globin promoters, especially at multiple copy (Ellis *et al.* 1993, 1996), cannot be assumed to accurately represent the function of the locus as a whole. In support of this, several groups have more recently begun studying the globin LCR in a more natural chromosomal context, using low copy number transgenics, YACs and somatic cell hybrids carrying human chromosome 11 (Ellis *et al.* 1996; Gribnau *et al.* 1998; Reik *et al.* 1998; Epner *et al.* 1998) and these studies seem to be yielding less ambiguous data than the earlier work with small constructs.

Figure 1.1: The human β–globin locus



Erythroid-specific DNAase I hypersensitive sites are shown by arrows.
ψβ1 is a non-expressed pseudogene.

HS2 of the β globin LCR has classical enhancer activity in that it functions as an enhancer of heterologous genes in transient assays, while the other three HSs do not. HS1 appears to have little independent function, and natural mutations deleting HS1 have no apparent phenotype. Similarly, HS4 has little or no function in isolation. HS3 appears to direct position-independent expression to an extent, and also to enhance expression when stably integrated. However, these properties are merely generalisations and the complexities of the reported data over the last decade could allow almost any interpretation to be made if viewed injudiciously. Recently a clearer view has emerged of the regulation of the β globin locus but there are still some outstanding controversies. These are discussed below.

1.4.1 Chromatin opening by the β globin LCR

The human β globin locus is characterised by a region of general erythroid-specific,

developmentally stable DNaseI I hypersensitivity relative to the surrounding chromatin. This open chromatin conformation is dependent upon the presence of the LCR. There is evidence that the opening of chromatin is independent of gene expression in as much as it precedes globin transcription (Forrester *et al.* 1986; Groudine *et al.* 1983; Forrester *et al.* 1987; reviewed by Grosveld *et al.* 1993). Also, in experiments where low levels of overall expression are seen, expression is still position independent (Fraser *et al.* 1993). It may be that the LCR autonomously opens the chromatin in erythroid cells, and that this open conformation facilitates the binding of transcription factors to both the LCR and the proximal promoters of the globins (Grosveld *et al.* 1987; Shesely *et al.* 1991; Kim *et al.* 1992). Reitman *et al* (1993) report that when linked to a truncated chicken β globin gene missing its promoter, the 3' enhancer (corresponding to LCR HS2 in humans) is not DNaseI I hypersensitive. However, only six such lines were generated and the HS was created in three of these. In addition, it should be noted that while the chicken and human β globin loci are often considered together, the rearrangement within the chicken relative to the human locus complicates the picture, and the functions of the HSs may not be strictly identical in both species. Finally, HS2 is generally accepted to have no chromatin-opening function in isolation (Ellis *et al.* 1993) and therefore a construct containing only HS2 would be susceptible to chromatin condensation which would presumably preclude formation of the HS.

Ellis *et al* (1996) recently conducted a careful analysis of the functions of the four HSs in single copy transgenic mice. HS3 was found to be necessary and sufficient to direct position independent expression and therefore is probably responsible for chromatin opening by the LCR. In the same paper they confirm the inability of HS2 to direct position-independent expression at single copy.

1.4.2 Transcriptional activation by the β globin LCR

HS2 is the only HS of the LCR capable of enhancing expression in transient assays (reviewed by Grosveld *et al.* 1993) and for this reason has received a great deal of attention in loss of function studies (Ryan *et al.* 1989; Caterina *et al.* 1991; Ellis *et al.* 1993; Caterina *et al.* 1994). However HS2 is only capable of expressing β globin at 40 % of wild type levels, and although it does seem to be able to direct position independent expression in transgenic animals (Fraser *et al.*

1993), this is only at high copy number (Ellis *et al.* 1993, 1996). The core of HS2 consists of binding sites for the haematopoietic transcription factor NF-E2, flanked by binding sites for the transcription factors GATA-1, USF and SP1 (Ellis *et al.* 1993; Caterina *et al.* 1994). Deletion of any of these sites reduces expression levels of a linked β globin gene, but no significant effect on position independence is observed unless the entire HS2 is deleted (Caterina *et al.* 1994). Ellis *et al* (1993) showed that the dimer of NF-E2 binding sites is sufficient to direct modest levels of position independent βexpression in multiple copy arrays. This result may indicate that multiple protein-DNA interactions are responsible for LCR function *in vivo*.

Binding sites for the haematopoietic factors NF-E2 and GATA-1 have also been shown to be necessary for the formation of HS4 (Stamatoyannopoulis *et al.* 1995) and are present in a similar conformation in HS3 (Pruzina *et al.* 1994; Stamatoyannopoulis *et al.* 1995). Thus these factors seem to be important in the formation of the LCR, although their functional significance is as yet unclear. As discussed below, in isolation individual HSs have differing activating properties with different genes of the locus, although no variation of this activation function with development has been demonstrated.

The LCR as a whole must be viewed as a strong activator of transcription within the locus, which is associated with several general and erythroid-specific transcription factor binding sites. Despite considerable effort by a number of research groups the mechanisms of this activation and its variation between genes and developmental stages remain unclear. Martin *et al* (1996) suggest that enhancers create favourable chromatin conditions for transcription rather than increasing the rate of transcription initiation *per se,* which would make enhancers and LCRs functionally similar. The best evidence to support this position comes from single cell expression analysis in variegating cell populations (Walters *et al.* 1995, 1996) where it has been demonstrated that enhancers increase the probability of gene expression without significantly altering the expression per cell, a role that has been a major feature in descriptions of LCRs. However, in this context the successful dissection of the position-independence function from the classical enhancer function within the globin LCR (Ellis *et al.* 1996) makes no sense - HS2 and HS3 clearly play different roles, at least in isolation. It may be that the classical enhancer activity of HS2 - the ability to enhance transcription of a linked transgene in a transient assay - is irrelevant to its function when stably integrated, but this seems unlikely given that strong

potentiation of expression coupled to sensitivity to position effects are the undisputed characteristics of HS2. Similarly, it now seems very clear that HS3 can suppress position effects, but has only a modest ability to increase transcription of linked globin genes. This, according to Martin *et al* (1996), would identify HS3 as an enhancer, yet in transient assays HS3 has no enhancer activity. Perhaps in time the original definition of an enhancer will be abandoned, but until then it must be concluded that the HS2 enhancer does not suppress position effects, while the position effect-suppressing HS3 element does not have enhancer activity. However, it may be that HS2 increases the probability that transcription factors will interact with the promoter, while HS3 increases and ensures the accessibility of HS2 to those factors which mediate its activity. Taken together, these two functions would allow HS2 to act as a enhancer as postulated by Martin *et al* (1996) while still depending on HS3 for reliable function in different chromosomal positions.

1.4.3 Developmental regulation of the β globin locus

Fraser *et al* (1993) analysed the developmental pattern of expression of the human globins in transgenic mice when linked to each HS of the LCR. The constructs were based upon a 36 kb restriction fragment encompassing the γ, δ and β globin genes with a single HS cloned in 3.4 kb 5' to the γ^G globin gene. γ and β globin steady-state mRNA levels were analysed in whole embryo, 16 day foetal liver and adult blood.

HS1 gave very low overall levels of β and γ globin mRNA, 4 % and 2 % respectively of the murine βH1 (embryonic) and βmaj globin (foetal and adult) mRNA levels. However, these levels were 50 fold higher than those obtained with a control construct with no HS at all. Expression was copy number dependent, although only two lines were examined. γ globin was expressed in the embryo but not in the foetal liver, which is a target of normal γ globin expression. β globin, which is normally expressed in foetal liver and adult blood, was expressed at a low level in all tissues under HS1. HS2 gave similar results to HS1, but the overall level of expression was higher. Unlike HS1, HS2 did not stimulate β globin transcription over γ globin transcription. Again, although only two lines were examined, expression was related to copy number.

HS3 was the only HS which drove γ globin transcription in foetal liver. γ globin mRNA levels in the embryo and foetal liver were comparable to those obtained with the entire locus in transgenic mice (Strouboulis *et al.* 1992). In contrast to the γ globin pattern, β globin steady-state mRNA levels were 15 % of endogenous levels in the foetal liver, and at 20 % in the adult blood. The entire locus gives 90 - 100 % β globin expression (Strouboulis *et al.* 1992).

HS4 gave low levels of γ globin mRNA (2 % of endogenous levels) in the embryo and no detectable mRNA in the foetal liver. β globin mRNA was 8 % of endogenous levels in the foetal liver but 25 % in adult blood, the highest levels of β globin expression observed in this study.

These results raise a number of interesting points. As has been seen by other groups (Magram *et al.* 1985; Kollias *et al.* 1986; Shih *et al.* 1990; Starck *et al.* 1994), in the absence of any LCR elements the globins still exhibit developmental regulation of expression. This, and the developmental stability of the LCR, has led to the belief that it is developmentally inactive. Fraser *et al* (1993) hold that their results suggest that the LCR is not developmentally neutral. However, results obtained with individual HSs must be viewed with caution, as many results point to a concerted action of the LCR as a whole and any HS in isolation may therefore not function as if it were part of the LCR. Also, the data obtained without an LCR cannot be ignored. γ globin expression in the foetal liver requires HS3, but silencing of the γ globin gene in adult tissues is autonomous (no requirement for the LCR). Therefore it may be that γglobin expression is regulated developmentally at the promoter, but that HS3 is needed for this regulation to result in detectable mRNA levels in the foetal liver. In support of this, hereditary persistence of foetal haemoglobin (HPFH), where silencing of γglobin in adult tissues is inoperative, can be caused by deletions and point mutations in the region from -117 to -202 bp from the γ globin cap sites (Berry *et al.* 1992; reviewed by Engel, 1993). Functional deletion studies of the γ globin promoter in transgenic mice confirm the presence of proximal sequences required for developmental regulation, especially silencing (Stamatoyannopoulis *et al.* 1993). Finally, no stage-specific proteins have been detected bound to HS3 or the LCR in general (Reitman *et al.* 1993), and the pattern of DNaseI I hypersensitivity within the LCR is developmentally invariant. Thus the role of the LCR in developmental regulation of γ globin transcription is not obvious, while the major role of the promoter seems incontrovertible.

The situation for ε globin is very similar to that for γ globin, in that it is autonomously silenced in foetal and adult tissues and requires at least part of the LCR for high levels of transcription in the yolk sac (Raich *et al.* 1990; Shih *et al.* 1990).

The results for βglobin are more complex. In the absence of HS3 and HS4, β globinis transcribed prematurely in the embryo, though at low levels (Fraser *et al.* 1993). The pattern of expression is correct in the presence of either HS3 or HS4. Thus silencing of the β globin gene is at least partly dependent upon the LCR. However, several reports suggest that competition between β globin and the other globin genes for the LCR is responsible for the developmental regulation of β globin (Choi and Engel, 1988; Shesely *et al.* 1991; Hanscombe *et al.* 1991; Berry *et al.* 1992; Kim *et al.* 1992) . In HPFH, the persistence of γ globin transcription is accompanied by a decrease in β transcription (reviewed by Engel, 1993). In the absence of the other globins, β globin expression is developmentally inappropriate and elevated in the presence of HS2, HS3 or HS4 (Fraser *et al.* 1993), suggesting that the LCR alone is not sufficient for developmental regulation of β globin. In murine erythroleukaemia (MEL) cells carrying a single human chromosome 11, which behave as adult erythroid cells, human βglobin transcription is abolished in the presence of an inducible selectable marker gene within the LCR (Kim *et al.* 1992). Hanscombe *et al* (1991) generated transgenic mice with constructs in which the LCR was linked to the γand β globin genes in different orders, and also with a construct containing the LCR, the β globin gene and the α globin gene, which are normally unlinked. When the β globin gene was 3' to the γ globin gene normal expression patterns were seen in embryo and foetus (adult animals were non-viable). However, when the β globin gene was 5' to the γ globin gene (*i.e.* closest to the LCR) β was expressed in the yolk sac; γ globin expression was unaltered in terms of developmental regulation, but was lower overall. Similarly, the presence of α globin between the β globin gene and the LCR gives rise to the correct pattern of β globin expression, even though the α globin gene does not normally interact with this LCR. When the β globin gene is closer than the α globin gene to the LCR the embryonic silencing of β globinis abolished. Thus the presence of another gene between the βglobin gene and the LCR is necessary for correct developmental regulation of β globin transcription.

In summary, the picture of developmental regulation of the β globin locus is broadly thus: the

ε globin gene and the two γ globin genes are autonomously regulated by their proximal promoters, being expressed in the correct developmental pattern in the absence of the β globin gene and also in the absence of the LCR, although low overall levels of expression make interpretation difficult. In contrast, the β globin gene is silenced in embryonic tissue only in the presence of the LCR (specifically HS3 and/or HS4) and the γ globin gene, or indeed another gene, provided it is proximal to the LCR relative to the β globin gene. Correspondingly, correct expression of β globin in adult tissues requires the silencing of upstream embryonic/foetal genes.

1.4.4 Models of LCR function

In light of the above results, the most convincing model is a development of that first suggested by Choi and Engel (1988) for the chicken β globin locus, in which competition for activation by the LCR between the globins is responsible for the gene switching in the presence of stage-specific gene silencers. The notion that competition for the LCR between promoters regulates expression within the locus has largely been accepted (reviewed by Dillon and Grosveld, 1993). The early proposal that competition was reciprocal (Behringer *et al.* 1990; Enver *et al.* 1990) has been disproved by work in transgenic systems (Hanscombe *et al.* 1991; Kim *et al.* 1992) and also by looking at natural genetic lesions (Berry *et al.* 1992). Instead, a polar competition model has emerged, where the order of the genes and their relative distances from the LCR are crucial in determining which is expressed at a given stage. However, the issue is complicated by the finding that εand γ globins are autonomously developmentally regulated, and do not require either the β globins or the LCR for the correct pattern of expression (although in the absence of the LCR expression levels are low and subject to position effects).

The model which best fits the data would hold that in embryonic tissues the LCR opens chromatin in the erythroid cells and strongly enhances ε and γ globin gene expression. In foetal liver cells, stage-specific factors binding to the εglobin promoter silence ε globin expression, but γ globin is still expressed at high levels and silences the more distal βglobin gene by competition for the LCR, which is closer to the γglobin gene than the βglobin gene and so interacts with the γglobin promoter preferentially. In adult erythropoietic cells the γ globin genes are silenced by factors binding to their promoters and interaction with the LCR is blocked. Now the nearest active (non-silenced) gene to the LCR is the βglobin (and δ) gene, and so the

adult globins are transcribed at high levels. In fact, the autonomous silencing of the ε and γ globin genes is consistent with this model, which would predict that those genes closest to the LCR would always be activated by it unless silenced, while distal genes would naturally be silenced by competition with the proximal genes and so would not be expected to be silenced autonomously (Hanscombe *et al.* 1991).

A refinement of the polar competition model asks whether the LCR 'seeks' promoters by diffusion through solution, or by tracking along the DNA until a promoter is reached. The main results in favour of the diffusion model come from Kim *et al* (1992) and Hanscombe *et al* (1991), where relative distance from the LCR is seen to be important, which is predicted by the diffusion model but not by the tracking model (reviewed by Dillon *et al.* 1997).

The above model is general in that it does not attempt to describe the conformation of the LCR or the nature of the interaction between it and the globin promoters. In general, most results point towards a view of the LCR as a single functional unit, but there are other models in the literature. Engel (1993) has proposed that the LCR is an 'amalgam' rather than a 'complex', suggesting that the functional linkage of the HSs is not reflected by a physical interaction. The stage-specific silencing of an individual globin gene is suggested to occur by a stable interaction between the promoter, a HS of the LCR, and a negative protein factor. In this context an LCR HS is proposed to act as part of a stage-specific silencing 'complex'. However, the data do not support this, and there are several results which suggest that the LCR constitutes a single regulatory unit in erythroid cells. Firstly, expression of all globins is higher in the presence of the entire LCR than any single HS or combination of HSs, including high copy number arrays (Fraser *et al.* 1993; Reitman *et al.* 1993). Secondly, DNaseI hypersensitivity pattern of the LCR is developmentally stable, even though different HSs seem to be most important at different developmental stages (Forrester *et al.* 1987; Tuan *et al.* 1989; Fraser *et al.* 1993). Finally, competition for the LCR occurs between the β and γ globin genes even though different HSs maximally stimulate their expression (Choi and Engel, 1988; Dillon and Grosveld, 1993; Fraser *et al.* 1993). If γ and β interact with different HSs, competition can only occur if these elements are stably complexed and acting as a single regulatory unit.

Also, in the absence of the LCR developmental silencing of the ε and γ globin gene still occurs,

so there is no reason to suppose that the LCR would be necessary to silence them. In addition, HS3 is invoked as the element responsible for silencing the γglobin genes (Engel, 1993), but there are two γ globin genes and only one HS3. It might be expected that a γ globin gene not bound to HS3 would be transcribed in adult cells under this model unless it is autonomously silenced, which would obviate the model.

In the absence of any data suggesting that the LCR performs a silencing function, and in the light of data which suggest that ε and γare autonomously silenced while βis silenced in early stages by competition for the LCR, a model based on negative regulation by the LCR seems less credible than a model based on positive interaction with the LCR which may be blocked by stage specific silencer factors.

It has also been suggested that the LCR plays no direct role at all in enhancing the expression of the globin genes, but merely maintains an open chromatin domain within which the genes are activated or silenced according to the suite of transcription factors bound to their proximal promoters (Groudine and Weintraub, 1982; Martin *et al.* 1996). This may seem at odds with the current body of data, but Martin *et al* (1996) point out that there is no definitive evidence for globin switching as a dynamic process; the suggestion is rather that it may be a stably established chromatin formation in different erythroid cells. Wiejgerde *et al* (1995) used FISH in single foetal erythroid cells to study the transcription of globin genes, and found that 85% of cells contained one species of globin transcript only (either γor β) during a switching period, while some 15% contained both transcripts. This was interpreted as evidence that only one gene was transcribed at a time, the double signals arising from persistence of transcripts in cells where a switch had recently taken place to the other gene, which supports the "looping out" model where the LCR interacts directly with globin promoters as a single holocomplex. However, Martin *et al* (1996) interpret the double signal as evidence of simultaneous transcription of globin genes, which is inconsistent with the "looping out" model, and instead propose that the data presented by Wiejgerde *et al* (1995) are consistent with their hypothesis that the LCR is not involved in switching, and that globin gene transcription is in fact a stochastic process.

To address this ambiguity, Gribnau *et al* (1998) performed single cell primary transcript FISH using a more stringent technique which removes all nascent transcripts (by allowing enough time

for processing and degradation) during a reversible inactivation of transcription elongation, before the block is removed allowing transcription to resume. In this manner they removed all previously synthesised and nascent primary transcripts from the cells so that any globin primary transcripts detected should be nascent at the time of assay.

The results of this experiment unambiguously demonstrate that only one β globin gene is transcribed at a time in any one erythroid cell and thus support the hypothesis that a single regulatory unit, the LCR, is dynamically alternating between globin promoters in globin gene switching. Double transcripts were detected with a lag time (from release of the block to intron 2 detection) of approximately double that of single transcripts, indicating switching. Although there is still no evidence that the LCR physically interacts with globin promoters, the hypothesis that globin transcription is a stochastic process operating independently at each promoter can now be rejected.

1.5 Other long range control elements

Over the last decade, a wealth of elements have been described which are reported to have LCR function (reviews: Higgs, 1998; Udvardy, 1999). Some of the more significant examples are briefly described below.

The structure of the β globin locus is conserved in mammals and chickens, although in chickens a duplication event has complicated the issue. Due to its similarities to the human β globin locus, the chicken locus will not be discussed here.

The chicken lysozyme locus has been extensively studied, its role as a marker of macrophage lineages paralleling that of β globin as a marker for erythroid lineages (reviewed by Sippel *et al.* 1992). Chicken lysozyme constitutes a major component of egg albumen, and is also expressed in the myeloid lineage of the haematopoietic cells. The gene lies within a 20 kb region of general DNaseI I hypersensitivity (Strätling and Dölle, 1986; Jantzen *et al.* 1986) at the borders of which are found the so-called A elements (Phi-Van and Strätling, 1988; Sippel *et al.* 1992, below) which have matrix attachment activity *in vitro* as well as a constitutive transcriptional activating activity in transgenic studies with homologous and heterologous genes, though not in

transient assays. Within this domain reside up to 9 DNase I hypersensitive sites comprising a complex system of tissue specific enhancers, promoter elements and a silencer element (reviewed by Sippel *et al.* 1992). The entire domain functions in a developmentally correct, position independent pattern in transgenic mice (Bonifer *et al.* 1990), and thus may be described as possessing a locus control function (Bonifer *et al.* 1990; Sippel *et al.* 1992; Sippel *et al.* 1993). The most elegant dissection of the regulatory elements of this locus has been reported by Bonifer *et al* (1994), in which the entire domain and deletions thereof were analysed in transgenic mice. Specifically, the work focused on the -6.1 kb early enhancer, which is DNaseI I hypersensitive in all appropriate cell types (oviduct, macrophage precursor cell lines and primary macrophages), the -2.7 kb late enhancer, which is only hypersensitive in the late stages of macrophage differentiation, (reviewed by Sippel *et al.* 1992), and the two flanking A elements. In contrast to expected results, deletion of the A elements did not abolish copy number dependent expression provided all of the *cis*-regulatory elements were intact, but it did result in an increased frequency of ectopic expression of the transgene, which was aggravated by the additional deletion of either enhancer. Deletion of either the -6.1 kb or the -2.7 kb enhancers resulted in a loss of position independence.

These results suggest that all of the *cis*-regulatory elements within the lysozyme domain are necessary for locus control function and that they act cooperatively rather than hierarchically, and also that the A elements do not protect from negative position effects in the context of the entire locus, but seem to insulate the gene from possible stimulation by elements outside the domain (see later discussion).

The human αglobin locus is evolutionarily related to the β locus (although divergent in structure), and coordinate expression of both loci results in the various stage-specific heterodimeric haemoglobins, each formed by a gene product from each locus (summarised by Higgs *et al.* 1990). Following the description of a natural deletion of 65 kb upstream of the α locus, resulting in repression of both α globins (Hatton *et al.* 1990), Higgs *et al* (1990) analysed the deleted region and fragments thereof linked to an α gene in stably transfected MEL cells and transgenic mice. Two erythroid specific HSs some 30 to 40 kb upstream of ζ globin were discovered, and a construct containing these directed high levels of developmentally correct αglobin transcription in transgenic mice, although copy number, which was analysed

over a much broader range than for the β LCR, was not correlated with expression level in high copy individuals.

The human major histocompatibility complex (MHC) class I antigen HLA-B7 gene is expressed on the cell surface of a wide range of cell types at different levels, is induced by viral infection and lymphokines, and may be repressed by some oncogenes (summarised by Chamberlain *et al.* 1991). Remarkably, a construct containing the HLA-B7 gene and just 0.66 kb of 5' and 2.0 kb of 3' flanking DNA in transgenic mice is expressed in a position independent, tissue specific pattern at levels approaching endogenous, and is inducible by the lymphokine IFNγ (Chamberlain *et al.* 1988). Deletion studies suggest that sequences from -0.66 kb to -0.11 kb are necessary and sufficient to direct high, position independent, IFNγ inducible expression of HLA-B7 (Chamberlain *et al.* 1991).

Human CD2 is expressed on the surface of differentiating T cells, and is present on most thymocytes and essentially all peripheral T cells (summarised by Greaves *et al.* 1989). A construct with the 15 kb CD2 gene, 4.5 kb of 5' and 9 kb of 3' flanking sequences was expressed in a copy number dependent pattern on T cells and megakaryocytes of transgenic mice at similar levels to the endogenous gene (Lang *et al.* 1988). Greaves *et al* (1989) demonstrated that T cell specific HSs in the 3' flank were necessary for high level tissue-specific expression, and sufficient to direct position independent expression of a thymus specific (Thy-1) and a heterologous (β globin) gene in the T cells of transgenic mice. Thus this element comprises a functional LCR. Interestingly, β globin was not prevented from expressing in the foetal liver, nor was Thy-1 silenced in the brain, indicating that the CD2 LCR does not possess any insulating or tissue-specific silencing function.

Deletion studies of the 3' flank of human CD2 revealed a 1.5 kb minimal region that conferred high level, copy number dependent expression in transgenic mice (Lang *et al.* 1991), and also assigned the enhancer and LCR functions to CD2 HS1 and CD2 HS3 respectively. HS1 had previously been shown to function as an enhancer in transient assays (Lake *et al.* 1990).

More recently, the expression of various CD2 constructs in the thymus of transgenic mice has been studied at the level of single cells (Festenstein *et al.* 1996), a study which provided the first

direct link between LCR function and PEV in vertebrates. Constructs in which HS3 was deleted expressed in a copy number dependent manner in most cases, but in some animals this was not the case. The transgene in these animals was found to have integrated pericentromerically. When thymus cells from these animals were flow-sorted according to CD2 expression, they separated into CD2+ and CD2- populations, which displayed open or closed chromatin around the transgene respectively. Thus in the absence of HS3 a CD2 transgene is expressed copy-number dependently in most animals, but when integrated pericentromerically is stochastically silenced with accompanying changes in chromatin structure. This study was soon followed by similar work with the human β–globin LCR (Milot *et al.* 1996a).

There are two reports of Alu sequences which appear to be implicated in LCR-like functions. Neznanov *et al* (1993) conducted deletion analyses of the 5' and 3' flanks of the human keratin 18 (K18). They found that with both flanks expression of a K18 transgene was indistinguishable from the endogenous mouse gene, but in the absence of the 3.5 kb 3' flank position independent expression is lost in some tissues but not in others. In addition, a TK reporter construct flanked by the 5' and 3' regions in transgenic mice was expressed in a position independent but not tissue specific pattern. The flanking regions have no obvious HSs or MARs, but do have *Alu* sequences. In the absence of the 3' flank, Thorey *et al* (1993) demonstrated that transcription of an *Alu* element in the 5' flank was necessary for copy number dependent expression in the kidneys of transgenic mice. Regulation of this locus is extremely complicated, but some locus control function is clear within the constructs used, although more work is necessary to obtain a clear picture. Aronow *et al* (1992) report a major regulatory complex in intron I of the human adenosine deaminase (ADA) gene, consisting of six DNaseI I hypersensitive sites, HSI to HSVI, and also *Alu* elements. HSIII has classic tissue specific enhancer activity in transient assays with a CAT reporter gene, but is subject to position effects in transgenic mice. Addition of non-HS sites flanking HSIII, including an *Alu* element, removes this position effect. Since CAT is activated tissue specifically and position independently, this represents an LCR function, the first described within a gene. The roles of the *Alu* elements in these loci are unclear, but active transcription of the element seems to be important for function.

The metallothionein genes MTI and MTII are expressed in essentially all cell types, and are regulated by environmental factors (metal ions) and cell-cell signalling. The MTI promoter

directs transgene expression in mice, but at low levels, subject to position effects and not related to copy number. The two MT genes are closely linked in mice, flanked by DNaseI I hypersensitive sites. When these sites were used in a mini-locus containing MTI, they were able to direct position independent copy number dependent expression in transgenic mice (Palmiter *et al.* 1993), which was stimulated correctly by environmental and developmental signals (metallothionein transcription is stimulated in the foetal liver). Thus these elements comprise an LCR function. However, only two of four heterologous genes showed position independent expression when driven by these regions, and MTI is expressed in over 50 % of transgenic lines even in the absence of the flanking elements, so the LCR function may be comparatively modest.

The *albino* phenotype of mice is caused by a point mutation in the tyrosinase gene, which spans 80 kb of chromatin. Introduction of a tyrosinase minigene including 5.5 kb of promoter sequences rescues the *albino* phenotype, but is subject to position effects (summarised by Ganss *et al.* 1994). Schedl *et al* (1993a) generated transgenic *albino* mice with a 250 kb yeast artificial chromosome (YAC) harbouring the tyrosinase gene and 155 kb of upstream sequences. All transgenic mice were fully pigmented, indicating correct expression of the tyrosinase gene. A HS 12 kb upstream of the gene was demonstrated to be melanocyte-specific (Ganss *et al.* 1994) and was used to drive expression of the tyrosinase minigene in transgenic mice. Coat colour and eye colour (which tissues derive from the two lineages in which tyrosinase is expressed) were found to roughly correlate with copy number, but some ectopic expression was seen which had not been detected with the YAC construct. Thus the YAC construct was correctly expressed, but the minigene construct has only imperfect LCR function, either because the upstream element is incomplete or because the minigene lacked elements necessary for tissue specificity. Work in YACs with the entire gene should address this question.

1.5.1 *Structural chromatin elements and insulators*

Along with the idea of a transcription domain within which influence from chromatin outside the domain is negligible, comes the idea that the domain has borders. Matrix attachment regions or scaffold attachment regions (MARs and SARs) have been proposed as the boundaries of these domains. The topological loops of up to 100 kb that chromatin is found in have been proposed as physical transcription domains (reviews: Cook, 1989; Eissenberg and Elgin, 1991). However,

some MARs reside within genes, which would suggest that a MAR is no barrier to the transcription complex, and therefore no barrier to other protein-DNA interactions. It is worth noting here that current *in vitro* assays for MAR function may introduce artefactual DNA-protein interactions, so that not all MAR elements may be functional *in vivo*.

Stief *et al* (1989) demonstrated that the chicken lysozyme 5' A element is able to protect a -6.1 kb enhancer/lysozyme promoter/CAT construct from position effects in stably transfected promacrophage cell lines, and report an enhancer function of the A element which differs from classical enhancers in that it is not apparent in transient transfection assays. Interestingly, the expression levels are not strictly copy number dependent, but rather seem to increase by more than the increase in copy number. McKnight *et al* (1992) analysed the effect of the A element in transgenic mice, using a construct linking the 5' MAR to the gene for whey acidic protein (WAP) under its own promoter. Expression of WAP was developmentally correct (*i.e.* lactation and mammary gland specific), which contrasts with the results from WAP transgenes in the absence of the MAR. However, no copy number dependence or elevation of expression levels were seen. Clark *et al* (1994) linked the A element to a chloramphenicol acetyltransferase (CAT) transgene driven by the ovine β–lactoglobulin (BLG) promoter. No significant increase in the frequency of CAT expression or in the correlation of CAT protein with transgene copy number was observed. Thus the A element is able to buffer a gene from some position effects in different tissues, but the protection is imperfect. It must be concluded from the limited data available that the A element is not simply an efficient 'insulator'. It has intrinsic enhancer-like properties, while its ability to create a topological chromatin loop, or confer copy number dependent expression, has not been proven (see also Dillon, 1993).

The *scs* elements of *Drosophila* insulate the *white* gene from both positive and negative position effects in *Drosophila* (Kellum and Schedl, 1991) and therefore seem to function as true insulators (reviewed by Sippel *et al.* 1992). 5' HS4 from the chicken globin locus, which is constitutive, has been reported to block the function of a linked LCR element and relieve position effects in *Drosophila* (Chung *et al.* 1993). Similarly, 5' HS5, the constitutive HS from the human β globin locus, has been used to insulate a β globin gene from stimulation by a linked HS3 element. The general region of DNaseI I sensitivity in the globin locus ends abruptly at this element *in vivo* (reviewed by Kollias and Grosveld, 1992), further suggesting that HS5 has an

insulating function.

1.6 Yeast artificial chromosomes

Yeast artificial chromosomes comprise an insert of apparently limitless size plus the three main specialised elements which direct efficient replication, maintenance and segregation of eukaryotic chromosomes, namely an origin of replication (ARS), telomeres (TEL) and a centromere (CEN) (reviewed by Hieter *et al.* 1990). There are also usually auxotrophic marker genes which allow selection in yeast for the arms of the YAC and bacterial sequences allowing propagation in *E. coli.*

ARS (autonomously replicating) sequences, first identified by the ability to confer high frequency transformation and extrachromosomal maintenance upon plasmids, consist of a core of at least one copy of the domain A consensus sequence (A/T)TTTAT(A/G)TTT(A/T), two or three 5' flanking nucleotides and a varying number of 3' flanking nucleotides (reviewed by Newlon, 1988). ARS1, used in pYAC vectors (Burke *et al.* 1987) conforms to this model.

Plasmids and YACs bearing ARS elements are subject to the same replication as normal chromosomes in that they are replicated once in S phase of the cell cycle. It may be that large eukaryotic DNA fragments (such as YAC inserts) contain additional ARS regions, and a number of eukaryotic DNA fragments have been shown to contain functional ARSs in yeast (reviewed by Newlon, 1988).

Plasmids bearing ARSs are characterised by episomal propagation, but segregation of the plasmids at mitosis is poorly regulated, leading to increased copy number in some cells and loss of the plasmid in others. Yeast centromeres were first cloned by their ability to mitotically stabilise such plasmids, and also to direct segregation in the first division of meiosis (reviewed by Newlon, 1988). The yeast centromere consists of three conserved regions. CDEI is an 8 bp conserved palindrome, separated from a 25 bp conserved sequence, CDEIII, which is an imperfect palindrome, by CDEII, which is an A-T rich region some 78-86 bp in length. Deletion studies suggest that the central nucleotide of CDEIII is essential, since point mutations of it decrease chromosome stability by 1000 fold. CDEI is important but not-essential, while the

length and A-T content of CDEII are more important than its sequence (reviewed by Hieter *et al.* 1990).

The structure of yeast centromeres is less complex than that of higher eukaryotic centromeres. For example, no specialised sequence motifs surround yeast centromeres, and yeast centromeres do not form the classic constricted kinetochore of higher eukaryotes, but bind a single microtubule in mitosis in an otherwise undifferentiated structure. Interestingly, yeast centromeres can be inactivated by transcription across them, a feature which was originally used in the study of centromere function and has recently given rise to a system for inducing an increase in YAC copy number (Smith *et al.* 1990, 1992).

Telomeres serve two functions in chromosome maintenance. Firstly, they protect a linear molecule from degradation or repair by the cell, and secondly, they allow the end of the chromosome to be replicated without net loss. Telomere function is exceedingly conserved and heterologous telomeres are functional in yeast, even if they have different sequences to yeast telomeres (reviewed by Newlon, 1988; Blackburn, 1990; Hieter *et al.* 1990). In general, telomeres have a G-rich strand orientated 5' to 3' towards the end of the chromosome, and a C-rich strand. In yeast the sequence is several hundred nucleotides of the repeat $(C_{1\text{-}3}A)_n$. When heterologous telomeres (*e.g. Tetrahymena,* (C_4A_2)) are introduced into yeast on the ends of linear DNA molecules, yeast telomeric sequences are added on, presumably by a telomerase. Most endogenous yeast telomeres are flanked by highly conserved repetitive 6.7 kb Y' elements, but these are of unknown significance and are not essential for telomere function. Another seemingly redundant element, the X element (3.0 - 3.75 kb) is also found at most yeast telomeres, in which it is separated from any Y' sequences present by $C_{1\text{-}3}A$ repeats. Both X and Y' elements contain ARSs.

The *de novo* synthesis of yeast telomere sequences on heterologous telomeres suggests that the process is functionally separate from normal template-dependent polymerisation. Telomerases from several organisms have been studied in detail and found to be riboproteins, the RNA component of which has a sequence complementary to the G-rich strand sequence of the telomere of that organism, and mutated RNA components lead to correspondingly altered telomere repeat sequences. Thus the telomerase is a reverse transcriptase which carries its own

RNA template. The mechanism by which telomerases recognise divergent telomeres is thought to involve specialised secondary structures formed by non-Watson-Crick pairing between G residues in the 3' overhang (reviewed by Blackburn, 1990).

Following the description by Burke *et al* (1987) of a series of YAC cloning vectors (the pYAC vectors) the use of YACs in genomic library construction has become widespread, due to the large insert size and therefore relatively few numbers of clones needed for a representative library. Construction of a YAC library is accomplished by ligation of both YAC arms to fragmented genomic DNA, followed by transformation into yeast and selection for cells carrying both arms of the YAC (reviewed by Hieter *et al.* 1990).

Since the insert size in YACs has no practical upper limit, the introduction of multiple inserts into single YACs during library construction is common. In addition, co-transformation of homologous YACs into single yeast cells can result in chimaerism by recombination between them (Larionov *et al.* 1994). Establishing that a YAC clone contains a single YAC representing a single contiguous intact piece of genomic DNA is an essential step following isolation of a library clone.

1.6.1 Transgenic animals and YACs

The production of transgenic animals from YACs is essentially the same as for more conventional constructs. A major obstacle is the generation of sufficient quantities of suitably pure and intact YAC DNA, since it is by definition present only at single copy in yeast cells. The pCGS966 - type vector presents a useful means of amplifying new YAC clones, and a derivative of pCGS966 can be used to retrofit existing pYAC-based YACs to introduce the inducible promoter and TK marker necessary for amplification.

There are now several reports from different groups in the literature describing the generation of transgenic mice from YACs via ES cells (Jakobovits *et al.* 1993; Strauss *et al.* 1993; Choi *et al.* 1993) and pronuclear injection (Peterson *et al.* 1993; Schedl *et al.* 1992; Schedl *et al.* 1993a). The YAC transgenes are maintained without significant rearrangement although initial rearrangement before integration seems high, and there are no deleterious effects due to the yeast

CEN and TEL sequences (reviewed by Forget, 1993). Therefore YACs represent a workable means of introducing very large segments of genomic DNA into transgenic mice.

1.7 The milk protein genes

The major milk proteins are the caseins and the whey proteins. Together these make up >90 % of the total milk protein by mass (reviewed by Mercier and Vilotte, 1993). In general, the regulation of these genes is under tissue-specific and hormonal control, with synergy of hormonal induction a key feature. Extracellular matrices and cell-cell interactions are also important, and there is evidence that induction of several milk genes involves the removal of specific repressor proteins that silence the genes in appropriate tissues and stages.

1.7.1 β lactoglobulin

β lactoglobulin (BLG) is the major whey protein in ruminants, but is absent from rodents. The 162 residue protein is encoded by a 7 exon gene spanning some 5 kb (Ali and Clark, 1988). In sheep, BLG RNA is already transcribed in mid-gestation. Steady-state BLG mRNA levels increase slowly before and then rapidly after parturition, reaching 5 % of total mammary mRNA by day 20 of lactation. The function of BLG is not clear, but it is thought to play a role in transport of small fatty acids within the mammary gland.

BLG transcription is regulated developmentally and tissue specifically. Like other milk protein genes, it is induced synergistically by prolactin in the presence of glucocorticoid, though it is less responsive to these lactogenic hormones than other genes such as β casein (reviewed by Mercier and Vilotte, 1993). A 16.2 kb construct comprising the BLG gene with 4.2 kb of 5' and varying 3' flanks was expressed at high levels in transgenic mice, and the expression pattern approximately followed that of the endogenous BLG gene in sheep (Simons *et al.* 1987; Harris *et al.* 1991). 406 bp 5' and 1.9 kb 3' to the BLG promoter are the minimum flanking sequences necessary for high level mammary gland specific expression in mice (Whitelaw *et al.* 1992). Dissection of the promoter reveals binding sites for a variety of transcription factors, including the mammary gland factor (MGF or MBPF), which has subsequently been included as a member of the Stat family of signal transducing transcription factors (Wakao *et al.* 1994; Darnell *et al.* 1994) and has been detected in non-mammary tissues. MGF/Stat5 is necessary for transduction

of prolactin activation to the promoters of milk protein genes. (Wakao *et al.* 1994; Gouilleux *et al.* 1994; Demmer *et al.* 1995).

The BLG promoter has been used to direct high levels of mammary specific expression of several transgenes in the mammary gland of mice and sheep (Simons et al. 1988; Clark *et al.* 1989; Archibald *et al.* 1990; Wright *et al.* 1991; Shani *et al.* 1992; Dobrovolsky *et al.* 1993; Hurwitz *et al.* 1994; Clark *et al.* 1997a,b; Korhonen *et al.* 1997; Butler *et al.* 1997; McKee *et al.* 1998; Sola *et al.* 1998; Hyttinen *et al.* 1998). When driven by a BLG promoter, levels of the serine-protease inhibitor α-1 antitrypsin (AAT) have reached 33 mg/ml in transgenic sheep milk, which represents 50 % of the total milk protein (Wright *et al.* 1991; Carver et al. 1992; Carver et al. 1993). However, expression levels are variable with integration position and do not usually correlate with copy number. In particular, cDNA-based BLG constructs are not expressed well. In addition to the inclusion of introns (see Section 1.2.2), a strategy has been developed where the poor cDNA expression is "rescued" by the presence of an entire BLG transgene *in cis* (Clark *et al.* 1992). Initial experiments with the human AAT cDNA demonstrated that if the BLG-cDNA construct was microinjected in the presence of a BLG transgene, the expression level of the cDNA was higher and more consistent than if it was injected without the BLG cointegrant (Clark *et al.* 1992). Three possible models were proposed to explain the rescue effect, suggesting that either the BLG transcription unit had some direct enhancing or insulating effect on the linked transgene, or else that the rescue was an indirect consequence of active transcription of the BLG gene. More recently it has been reported that the BLG transgene must be transcribed but need not be translated to effect the rescue (Yull *et al.* 1997). This suggests that functional elements within the transcription unit are not directly responsible for the rescue, but rather that active transcription of the rescuing transgene somehow confers activation of the rescued transgene. However, in order to abolish transcription the construct used by Yull *et al* (1997) was significantly altered relative to that which was transcribed, and other effects of this perturbation such as altered nucleosome phasing (Lui *et al.* 1995) cannot be ruled out.

Initial rescue experiments used the BLG gene to rescue BLG-driven constructs, and introduced the two constructs by coinjection. Although this undeniably rescued the expression of the coinjected transgene, the expression levels of the rescued and rescuing transgenes were poorly correlated and unpredictable, which has been attributed to varying arrangements at the

transgenic locus (Clark *et al.* 1992). Clark *et al* (1997a) therefore injected double constructs consisting of a BLG gene linked covalently to a BLG-FIX (Factor IX) or BLG-CAT gene. This resulted in silencing of the linked BLG gene itself, which was unexpected. It was postulated that the regular arrangement of rescued and rescuing transgenes might result in some form of repeat-induced silencing. However, there are many examples of transgenes which are not obviously silenced in this manner (*e.g.* Grosveld *et al.* 1987; Greaves *et al.* 1989; Wright *et al.* 1991; Whitelaw *et al.* 1992). The most striking of these are those which are expressed position-independently. It may be argued that in this case whatever function relieved the transgene from position effects would also overcome repeat-induced silencing, and yet BLG has been reported to express position-independently (Whitelaw *et al.* 1992). In addition, although a FIX construct has been successfully rescued by a coinjected BLG transgene in two previous studies (Clark *et al.* 1992; Yull *et al.* 1995) both used a larger BLG construct incorporating some 4.2 kb of the BLG promoter, and also more 5' and 3' BLG DNA flanking the FIX transgene. Thus, there is no direct evidence that the minimal BLG transgene used in the later study would be expected to rescue a coinjected transgene. In the earlier study with AATD (Clark *et al.* 1992) which also used the larger BLG transgene there were two out of the nine BLG-AATD (Clark *et al.* 1992) cointegrated lines which did not express AATD or BLG. This may be taken as evidence that the silencing of BLG by a linked transgene may extend across the larger BLG flanks (Clark *et al.* 1997a) but it seems more likely that these lines were simply silenced by some other means, such as integration of the injected DNA into centromeric heterochromatin. Nevertheless, this is an intriguing result which may reveal new aspects of gene control if repeated.

Rescue of BLG-transgenes by an entire BLG gene has been convincingly demonstrated, at least in mice. However, rescue of other promoters with BLG has been less successful, as have efforts to use other rescuing transgenes. Langley *et al* (1998) cointegrated the entire BLG gene with a ribozyme construct under the control of the mouse mammary tumour virus long terminal repeat (MMTV-LTR). In this case one line was generated by coinjection and six more by injection of a double construct consisting of the MMTV-LTR driving the ribozyme (MMTV-RZ5) linked tail-to-tail with the BLG gene including 4.2 kb of 5' and 1.9 kb of 3' flanking DNA. The BLG gene successfully rescued MMTV-RZ5 expression in the mammary gland, but both the ribozyme and BLG mRNA were also detected in a range of other tissues, corresponding to some of those which would normally express MMTV-RZ5 but not significant levels of BLG. In addition, the

temporal control of BLG expression was lost. This may indicate that BLG expression is not after all immune to position effects (Whitelaw *et al.* 1992) as clearly its expression here is influenced by neighbouring chromatin. However, this may be a unique case since the influencing DNA in this instance is not packaged into chromatin when it first encounters the BLG gene and the nuclear environment. For practical purposes, this experiment demonstrates that the BLG transgene rescue is not mammary-specific, which may limit its use for production of bioactive molecules or other proteins which would be deleterious outside the mammary gland. In addition, these data call into question the proposal that BLG transcription is required for rescue (Yull *et al.* 1997). The BLG transgene would not ordinarily be transcribed outside the lactating mammary gland (Simons *et al.* 1987; Whitelaw *et al.* 1992), so it is difficult to see how BLG transcription could lead to rescue in non-mammary tissues. Possibly transcription of the MMTV-RZ5 transgene rescues transcription of BLG in these tissues and the two feed back on each other, but MMTV-RZ5 has been shown to be silent in these tissues in the majority of lines generated in the absence of a BLG cointegrant (L'Hullier *et al.* 1996, Langley *et al.* 1998). Thus, the mechanism of transgene rescue remains obscure.

1.7.2 α lactalbumin

α lactalbumin (αlac) is found in most mammalian milks, and is a calcium metallo-protein, structurally related to lysozyme. αlac functions to convert galactosyl transferase to lactose synthase in the mammary gland (reviewed by Sawyer and Holt, 1993; see also Stinnakre *et al.* 1994). αlac is encoded by a 2 kb gene of 4 exons, and bovine αlac genomic constructs with as little as 750 bp of 5' and 350 bp of 3' flanking sequences are expressed at levels and in a pattern similar to the endogenous gene in transgenic mice (Vilotte *et al.* 1989). In vitro studies have identified binding sites for several mammary nuclear factors including Stat5 (reviewed by Mercier and Vilotte, 1993).

1.7.3 Whey acidic protein

Whey acidic protein (WAP) is the major whey protein in rodents. WAP mRNA constitutes 10 - 15 % of total mRNA in lactating rat, rabbit and mouse mammary epithelial cells. Low levels (1000 fold less) of mRNA are also detectable in virgin and early gestating animals. WAP is poorly expressed in cell culture, which probably reflects a requirement for cell morphology and

cell-cell interaction within the mammary gland which is not replicated in cell culture. There is evidence that post-transcriptional repression may play a part in silencing WAP expression in cell culture (Chen and Bissel, 1989).

A rat WAP genomic fragment with 950 bp of 5' and 1.4 kb of 3' flanking sequence was expressed at 1 - 95 % of endogenous levels in transgenic mice, but expression was initiated earlier in gestation than the endogenous gene. Interestingly, deletion of the 3' flank to 70 bp improved the copy number dependence and frequency of expression, but did not alter the premature initiation.

Two lactating mammary gland specific DNase I hypersensitive sites have been described upstream of WAP. The distal site, at -0.8 kb, is essential for WAP expression in transgenic mice. Glucocorticoid response elements are located within this HS, and the hypersensitivity is dependent upon glucocorticoid induction (Li and Rosen, 1994). Sites binding mammary nuclear proteins in vitro have been mapped to the region from -175 to -88 bp, and recently a negative element has been described which is essential for silencing of WAP in non-mammary cell lines (Kolb *et al.* 1994). The WAP promoter has been used to drive expression of a variety of transgenes (*e.g.* Gordon *et al.* 1987; Yu *et al.* 1989; Velander *et al.* 1992; Devinoy *et al.* 1994). However, expression levels have generally been low and variable, and subject to position dependent ectopic expression compared to the endogenous WAP gene. Nevertheless the WAP promoter has been used to express significant levels of a number of heterologous proteins in the transgenic mammary gland (Pittius *et al.* 1988; Bayna and Rosen, 1990; Reddy *et al.* 1991; Velander *et al.* 1992; Bischoff *et al.* 1992; Hansson *et al.* 1994; Thepot *et al.* 1995; Paleyanda *et al.* 1997; Castilla *et al.* 1998; Butler *et al.* 1997; Lee *et al.* 1998).

As discussed elsewhere, McKnight *et al* (1992, 1995) report the position independent expression of a WAP construct (2.6 kb of 5' and 1.6 kb of 3' flank) when flanked by chicken lysozyme 5' A elements, although copy number dependence was not observed. Thus, existing WAP constructs seem to lack essential flanking elements necessary for insulation of the gene from ectopic stimulation as well as elements necessary for consistently high levels of expression.

1.7.4 The caseins

The caseins are the major mammalian milk proteins. They are thought to perform two main functions, providing a dietary supply of amino acids to the infant and increasing the solubility of calcium (essential for bone growth) as calcium phosphate in the milk. In the presence of calcium, the so-called calcium sensitive caseins of the α and β types form loose, non-crystalline, heavily phosphorylated metallo-protein aggregates termed micelles, which are stabilised by the calcium insensitive κ casein. In the gut the site-specific cleavage of κ casein by rennin causes the milk to clot and remain in the stomach, facilitating digestion (reviewed by Mercier and Vilotte, 1993).

The calcium sensitive caseins are variable in primary and secondary structure, but all possess three basic functional domains. The hydrophobic domain is important in micelle formation, the phosphorylation sites are important in calcium binding, and the highly conserved signal peptide allows efficient secretion of the caseins into the mammary lumen.

In contrast, κ casein is unrelated to the other caseins and does not bind calcium. It is necessary to stabilise the micelle, and the ratio of κ to the other caseins is important in determining micelle size and therefore milk quality (Lin *et al.* 1989; Rollema, 1992). The site-specific cleavage of κ casein in the gut causes the milk to clot and remain in the stomach, facilitating digestion. K casein may be structurally and functionally related to γ fibrinogen, which performs a similar cleavage-induced clotting function in the blood (Jolles *et al.* 1974; Thompson *et al.* 1985; Alexander *et al.* 1988).

All of the bovine casein genes have been cloned and sequenced (Alexander *et al.* 1988; Bonsing *et al.* 1988; Koczan *et al.* 1991; Groenan *et al.* 1993). While the sequences of the casein genes are divergent, certain features of the genes and proteins show the calcium sensitive caseins to be evolutionarily related (Bonsing and McKinlay, 1987). There is clear compartmentalisation of functional domains into discrete exons; no codon in any cloned casein gene is split by an intron. The β casein gene is the smallest of the three bovine genes at 8.5 kb, compared to 17.5 kb for αs1 and 18.5 kb for αs2, but nevertheless has features common to all. Exon I encodes the 5' untranslated region except for 6 nt which are encoded by exon II along with the highly conserved signal peptide. The hydrophobic site is encoded by exon VII, while four minor

phosphorylation sites are encoded by the small exons III, IV, V and VI. The major phosphorylation site is formed by a splice event between exons III and IV, an event which may be selected for since the canonical splice sequence favours the formation of such a site (Jones *et al.* 1985). Exons VIII and IX encode the stop codon and polyadenylation signal. The exon-intron structure of β casein led Rosen and colleagues to propose a model for calcium sensitive casein evolution, where recruitment of functional domains on exons was followed by intragenic and intergenic duplications of an ancestral gene to generate the present genes (Jones *et al.* 1985; Yu-Lee *et al.* 1986; see also Groenan *et al.* 1993). There is no evidence that κ casein is evolutionarily or functionally related to the other caseins.

The caseins of humans, mice, cattle, sheep and rabbits are tightly linked (summarised by Mercier and Vilotte, 1993), and at the start of this project the bovine casein locus had already been mapped by restriction digest (Ferretti *et al.* 1990; Threadgill and Womack, 1990). All of the bovine caseins are present on 200 kb of chromatin. This is not surprising for the calcium sensitive caseins, but the finding that the unrelated κ casein gene is so closely linked to the other members of the family is unexpected, though suggested by the linkage data. It may be that this linkage is indicative of some form of locus control for the caseins, as noted by Alexander *et al* (1988). The murine (this thesis; Tomlinson *et al.* 1996; Rijnkels *et al.* 1997a) and human (Rijnkels *et al.* 1997b) casein loci have since been restriction mapped, and show similar organisation.

Expression studies have focused on β casein, which is the most abundant protein in murine milk. By the start of this project the synergistic action of the lactogenic hormones, demonstrated for the whey proteins, had also been shown for β casein (Schmidhauser *et al.* 1990). Elements of the extracellular matrix are also important (reviewed by Mercier and Vilotte, 1993). Among numerous protein binding sites in the promoter, a Stat5 binding site is essential for prolactin induction, and there is evidence that prolactin induction involves displacement of a repressor protein from the promoter, which may also be the case for WAP (Schmitt-Ney *et al.* 1991; Lee and Oka, 1992). The first intron of β casein is also important for proper expression (Lee *et al.* 1989a, b). More recent advances in the study of β casein expression are discussed in chapter 7.

Casein promoters have been used with some success to direct heterologous gene expression to

the mammary gland (Buhler *et al.* 1990; Meade *et al.* 1990; Greenberg *et al.* 1991; Krimpenfort *et al.* 1991; DiTullio *et al.* 1992; Brem *et al.* 1994; Ebert *et al.* 1994; Maga *et al.* 1994; Uusi-Oukari *et al.* 1997; Cerdan *et al.* 1998; Edmunds *et al.* 1998; Bijvoet *et al.* 1998). Interleukin-2 has been expressed at up to 430 ng/ml in rabbit milk from a 2 kb rabbit β casein promoter (Bühler *et al.* 1990). Using a 21 kb bovine αs1 promoter Meade *et al* (1990) obtained expression of human urokinase at 1-2 mg/ml in the milk of a transgenic mouse. Similarly, Platenburg *et al.* (1994) report the expression of 36μg/ml of human lactoferrin in transgenic mouse milk from a cDNA driven by a bovine α_{s1} promoter, and Brem *et al* (1994) used the same promoter to obtain 1mg/ml of human IGF-1 in transgenic rabbit milk, again from a cDNA. Both subunits of follicle stimulating hormone and the membrane protein cystic fibrosis transmembrane conductance regulator have been expressed under the control of the β casein promoter (Greenberg *et al.* 1991; DiTullo *et al.* 1992). In all these cases expression levels were very variable with the site of integration, which would suggest that there are regulatory sequences outside the region of the casein promoters used in these experiments.

1.8 Summary of project aims

The basic aim of this project was to assay for position-independent, copy number-dependent, physiological levels of gene expression from a YAC carrying the entire mouse or human casein locus in transgenic mice. As well as providing another instance of locus control to add to the growing body of data, this might ultimately lead to improved targeting of transgene expression to the lactating mammary gland. This could have important benefits in biotechnology, where the mammary gland is currently the system of choice for production of biomedical proteins in transgenic animals. Existing casein constructs have been subject to profound position effects, despite using up to 21 kb of 5' flanking DNA in the case of β casein. This suggests that there are other important control elements in the casein locus which have been omitted from the casein constructs used to date. In addition, the fact that the κ casein gene is structurally unrelated to the other casein genes but is tightly linked may be an indication that some coordination of casein gene expression is vested at the locus level. Therefore, a strategy was devised to use a YAC to introduce the entire casein locus into transgenic mice, so that any long range control elements would hopefully be included. The aim was to assay gene expression from the YAC in transgenic mice and tissue culture, with a view to relating expression level to copy number. If copy

number-dependent expression was observed, molecular dissection of the YAC sequences would hopefully identify functionally significant regions.

2. Chapter 2 - Materials and methods

2.1 Suppliers of materials

All laboratory reagents were obtained from Fisons Chemicals, Sigma Chemical Co, BDH, Gibco BRL Life Technologies and FMC, except where listed below.

2.1.1 Bacterial and yeast media

Bio 101 Inc.

Sigma Chemical Co.

Gibco BRL Life Technologies

2.1.2 PFGE reagents

2.1.2.1 Restriction and other enzymes

New England Biolabs

2.1.2.2 Agarose

Gibco BRL (Rapid Agarose)

FMC (SeaKem Gold, SeaPlaque, Nusieve GTG)

2.1.3 Enzymes

All enzymes were purchased from Boehringer Mannheim except where otherwise stated.

2.1.4 Southern blotting materials

Schleicher and Schuell

Sartorius

2.1.5 Radiolabelled reagents

[α-^{32}P]dCTP 3000Ci/mmol, NEN Life Science Products

[γ-^{32}P]dATP 6000Ci/mmol, NEN Life Science Products

2.1.6 Mice

C57bl6/CBACA F_1 hybrids, Harlan UK.

2.2 Buffers and solutions - Table 2.1

1M NaPi:	1M NaH_2PO_4, pH adjusted to 7.2 with 1M Na_2HPO_4
50x TAE:	2M Tris, 1M acetic acid, 50mM EDTA, pH8.0
10x TBE:	890mM Tris, 890mM boric acid, 20mM EDTA, pH 8.0
10x TE:	100mM Tris.HCl, 10mM EDTA, pH8.0
1x ET:	10mM Tris, 50mM EDTA. pH8.0
10x TlowE:	100mM Tris, 1mM EDTA, pH8.0
10x HEPES buffer	0.5M HEPES (N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]), 10mM EDTA, pH 8.0
20x SSPE	3M NaCl, 177mMJ NaH_2PO_4, 20mM EDTA, pH7.4 via NaOH
C&G buffer:	0.5M NaPi pH7.2, 7% SDS, 1mM EDTA
Denaturant:	0.5M NaOH, 1.5M NaCl
GTE	50mM glucose, 10mM EDTA, 25mM Tris-HCl pH 8.0
KOAc	60 ml 5M potassium acetate, 11.5 ml glacial acetic acid per 100mls. The resultant solution is 3M potassium and 5M acetate, pH4.8
NaOH/SDS	0.2N NaOH, 1% SDS, made fresh
NDS	50mM Tris.HCl pH9.0, 3% N-lauroylsarcosine, 0.5M EDTA, 1mg/ml proteinase K (added fresh).
Neutralising buffer:	3M NaCl, 0.5M Tris.HCl, pH7.0
PEG solution:	10mM Tris, 1.0mM EDTA, 100mM LiAc, 40% w/v PEG_{3350}
10x TE/LiAC:	100mM Tris, 10mM EDTA, 100mM LiAc
Tail buffer:	0.3M NaOAc, 50mM KCl, 1.5mM $MgCl_2$, 10mM Tris.HCl (pH 8.5), 0.5% NP40, 0.5% Tween-20
SE	1M Sorbitol, 50mM EDTA
SCE	1M Sorbitol, 50mM EDTA, 0.1M sodium citrate, pH5.8 via citric acid.
$CaCl_2$ solution	0.1M Tris.HCl, pH7.4, 0.1M $CaCl_2$
PEG/$CaCl_2$	45% (w/v) $PEG_{3350,}$, 10mM Tris.Cl pH7.4, 10mM $CalCl_2$
Yeast lysis solution (YLS)	0.5M Tris.HCl pH9.0, 3% N-lauroylsarcosine, 0.2M EDTA, 1mg/ml proteinase K (added fresh).

2.3 Media for yeast and bacterial culture

Table 2.2: Bacterial and yeast media

Lauria Broth (LB)	10g bacto-tryptone, 5g bacto-yeast extract, 5g NaCl per litre, pH 7.,2 (NaOH) purchased in gelatin pellets from Bio101 Inc.
Lauria Agar (L)	As above but with 15g agar per litre
2YT	16g bactotryptone, 10g yeast extract, 5g NaCl per litre.
Terrific Broth (TB)	12g bacto-tryptone, 24g bacto-yeast extract, 4ml glycerol per 900 ml, plus 100ml 0.17M KH_2PO_4/0.72M K_2HPO_4 (added after autoclaving)
L-Amp plates	L-agar plus 100 μg/ml ampicillin, added when molten at 60°C after autoclaving
LAXI plates	L-Amp plus 0.004% X-Gal (5-bromo-4-chloro 3-indolyl β-D-galactosidase, 2% stock in dimethylformamide) and 100μM IPTG (isopropyl-β-D-thiogalactopyranoside, 100mM stock in water), added at 60°C after autoclaving.
YPD broth	10g yeast extract, 20g peptone, 20g dextrose per litre. Purchased from Bio 101 Inc.
YPD agar	As above but with 2% select agar
Dropout base (DOB)	1.7g yeast nitrogen base, 20g dextrose, 5g ammonium sulphate per litre. Purchasd from Bio 101 Inc.
Dropout base agar (DOBA)	As above but with 1.7% agar
Synthetic defined medium (SD)	6.7g yeast nitrogen base, 20g dextrose per litre, pH5.8. Supplemented with at least lysine, adenine, histidine and isoleucine.

2.3.1 Yeast "dropout" powders

These are blends of all essential nutrients in the concentrations below except for the "dropout" nutrient, and are used to select for prototrophy for that nutrient. Dropout powders were purchased from Bio 101.

Table 2.3: Yeast medium supplements

Nutrient	Final concentration (µg/ml)
Adenine	20
L-Arginine.HCl	20
L-Histidine.HCl	20
L-Isoleucine	30
L-Leucine	30
L-Lysine.HCl	30
L-Methionine	20
L-Phenylalanine	50
L-Theronine	200
L-Tryptophan	20
Uracil	20
L-Valine	150

2.4 Plasmid DNA stocks

All proprietary plasmids were obtained from the appropriate supplier or subcultured from purchased plasmid stocks. Stocks of all plasmids were stored as *E. coli* aliquots at –70°C, as frozen DNA at –20°C or as precipitated DNA under ethanol at –70°C. All of these methods store plasmids essentially indefinitely, but of the three storage under ethanol is the most robust in that accidental thawing would not harm the sample. Aliquots of plasmids in use were maintained in TE at either 4°C or –20°C.

2.5 Genomic DNA stocks

Stocks of genomic DNA were stored at 4°C in TE.

2.6 Bacterial and yeast host strains

Bacterial host strains were obtained from Stratagene or Boehringer Mannheim. Yeast strain AB1380 was obtained from the American Type Culture Collection. YPH499 and YPH500 were obtained from Stratagene.

Table 2.4: Yeast and bacterial host strains

Strain	Genotype	Reference
DH5αF'	*supE44, ΔlacU169(ϕ80lacZΔM15), hsdR17, recA1, endA1, gyrA96, thi-1*	Hanahan 1983
XL1-Blue	*recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac [F', proAB, lac^{q}ZΔM15, Tn10 (tet^R)]*	Bullock *et al.* 1987
SCS110	*rpsL (Str^r), thr, leu, endA, thi-1, lacY, galK, galT, ara, tonA, tsx, dam, dcm, supE44, Δ(lac-proAB), [F', traD36, proAB, lacI^{q}ZΔM15]*	Yanish-Perron *et al.* 1985
AB1380	*MATα, ura3-5, trp1, ade2-1, can1-100, lys2,1, his5*	Burke *et al.* 1987
YPH499	*MATa, ura3-52, lys2-801^{amber}, ade2-101^{ochre}, trp1Δ63, his3-Δ200, leu2-Δ1*	Sikorski and Hieter 1989
YPH500	*MATα, ura3-52, lys2-801^{amber}, ade2-101^{ochre}, trp1Δ63, his3-Δ200, leu2-Δ1*	Sikorski and Hieter 1989

2.7 Plasmid DNA preparation

Plasmid DNA was prepared by a number of different protocols. Generally, large numbers of clones were prepared by a "miniprep" method, either alkaline lysis and phenol or using the Wizard kit from Promega. Wizard, Wizard 377 and Wizard Plus were all used with no appreciable difference being observed between DNAs prepared by any of the three kits.

2.7.1 Alkaline Lysis miniprep

E. Coli was grown at 37 °C overnight, with shaking, in 3 ml of rich medium (e.g. 2YT + 10 μg/ml ampicillin) in a 14 ml culture tube (Falcon).

1.5 ml of culture were transferred to a 1.5 ml microcentrifuge tube and centrifuged for 2 min at full speed (13,500 rpm) in a benchtop microfuge. The supernatant was removed and the cell pellet resuspended in 100 µl of ice cold GTE (50mM glucose, 25mM Tris pH8.0, 10mM EDTA) supplemented with 4 mg/ml lysozyme.

After incubation for a few minutes at RT, 200 µl of NaOH/SDS (0.2N NaOH, 1% SDS) were added. The tubes were mixed by inversion. The tubes were left on ice for 5 min.

150 µl of KOAc solution (pH4.8) were added, and the tubes mixed by vigorous inversion or by gentle vortexing while inverted. The tubes were left on ice for 5 min.

The tubes were centrifuged for 5 min at 13,500 rpm in a microfuge, after which the supernatant was carefully removed to a fresh tube, taking care not to disturb the pellet of cell debris and genomic DNA.

An equal volume of phenol/chloroform/IAA (25:24:1) was added to each tube, and mixed by vortexing. After centrifugation for 2 min at 13,500 rpm, the upper (aqueous) phase was transferred to a fresh tube. This was repeated using 24:1 chloroform/IAA to remove traces of phenol, and then the plasmid DNA was precipitated by the addition of 2 vol of 100 % ethanol.

After incubation on ice or at -20 °C for 10 min, the DNA was pelleted by centrifugation for at least 10 min at 13,500 rpm, washed once with 70 % ethanol, and air-dried briefly RT (overdrying renders plasmid DNA difficult to redissolve). The DNA was then redissolved in 100 µl of TE or TlowE.

2.7.2 Large scale "midi" plasmid DNA prep

Large quantities of plasmid DNA were either prepared by alkaline lysis followed by CsCl gradient centrifugation, or by Qiagen column using the tip-500 kit.

2.7.2.1 CsCl protocol I

An overnight culture in 50 ml was inoculated with a single colony from a fresh plate. Medium was typically 2YT supplemented with 10 μg/ml ampicillin. The culture was incubated at 37 °C overnight, with shaking, before pelleting the cells for 15' at 3000 rpm in a benchtop centrifuge such as the Beckman GPR. The cell pellet was resuspended in 5 ml of GTE, 10 μg/ml lysozyme and incubated on ice for 5 min.

10 ml of freshly made 0.2 N NaOH, 1% SDS was added and mixed by gentle inversion. Following 10 minutes on ice, 7.5 ml of solution III were added and the tubes left on ice for a further 10 min, before centrifugation at 3,000 rpm for 20 min to pellet the cell debris and genomic DNA.

To the supernatant was added 0.5 vol of 7.5 M NH4OAc, and the tubes were mixed by inversion and left at RT for 20 min. The precipitated protein was then pelleted by centrifugation at 3,000 rpm for 20 min.

2/3 vol of isopropanol was added to the supernatant and left at RT for 30 min to precipitate nucleic acids. After a 20 min centrifugation at 3,000 rpm, the nucleic acid pellet was washed with 70 % ethanol and redissolved in TE supplemented with 10 μg/ml RNase A. The solution was incubated for 1 h at 37 °C.

1.5 g of CsCl was added to each tube, and the resultant solution was transferred to a Beckman 2 ml polyallomer ultracentrifuge tube. 120 μl of 10 mg/ml EtBr were added to each tube, and the tubes balanced and heat-sealed then mixed by inversion.

The plasmid DNA was banded by centrifugation at 80,000 g for 16 h in a Beckman TL100 ultracentrifuge using the TL-A 100 rotor.

The plasmid band was visualised with UV light (at 320 nm) if necessary, and removed by piercing the tube with a syringe. The DNA was recovered by 2-fold dilution with water, repeated washing with an equal volume of isoamyl alcohol to remove the EtBr, and precipitation

with 0.1 vol of 3 M NaOAc and 2 vol of 100 % ethanol. The pellet was washed in 70 % ethanol and redissolved in TE or TlowE.

2.7.2.2 CsCl protocol II - "Triton protocol"

200ml of overnight culture (in rich medium) were centrifuged in a Beckman J2-21 centrifuge, and the cell pellet was resuspended in 3 ml of 15% sucrose, 50mM EDTA, 50mM Tris (pH 8.0), 3 mg/ml lysozyme. The cells were transferred to an Oakridge tube and incubated on ice for 15 min. 6 ml of triton mix were added and mixed gently, and left on ice for 20 min or until lysis was complete. Following centrifugation at 18,000 rpm for 45 min, the supernatant was added to 9.5g CsCl and 1ml of 10 mg/ml EtBr. The tubes were centrifuged for 10 min at 15,000 rpm. The supernatant was decanted from under the resultant pellicle, and loaded into 13 ml ultracentrifuge tubes. Following balancing and sealing, the tubes were centrifuged for 16h at 55,000 rpm followed by 40,000 rpm for 1h. The plasmid DNA band was removed and dialysed against 5l of TE for at least 2h, then phenol/chloroform extracted and precipitated with isopropanol. The DNA pellet was washed with 70% ethanol and redissolved in TE.

2.7.2.3 Qiagen maxiprep for plasmid DNA

This protocol followed the Qiagen protocol booklet, using the Qiafilter units and the proprietary reagents.

2.8 Genomic DNA preparation

2.8.1 From mouse tail biopsy

Mouse tail tissue was incubated in 3 ml Tail Buffer supplemented with 250µg/ml proteinase K for 16h at 55°C with vigorous shaking. 0.5 ml of the digested material was transferred to a microcentrifuge tube and extracted twice with 700 µl of TE-saturated phenol/chloroform (50:50), then once with chloroform/IAA (24:1). The DNA was precipitated with 700 µl of isopropanol and pelleted by centrifugation in a microcentrifuge. The DNA pellet was washed twice with ice-cold 70% ethanol and resuspended in TE.

2.8.2 From yeast – DNA in agarose plugs

Yeast cells were grown in rich medium overnight. 50 ml of cells were pelleted by centrifugation at 3000 rpm in a Beckman GPR benchtop centrifuge. The cell pellet was washed twice with 20 ml water and then twice with 20 ml SE, and then transferred to an Eppendorf tube by the addition of 300 µl SE. Zymolyase 20T (ICN Biomedicals) was added to a final concentration of 2mg/ml, and 1µl of β mercaptoethanol was added. Occasionally the Zymolyase treatment was omitted with little effect, but more usually it was found to be necessary for release of the DNA from the agarose plug. An equal volume of molten 1.6 % InCert agarose (FMC) was added at 50°C and the mixture was cast either into a BioRad plug former or by taking up into 1 ml syringes. Once set the agarose could be released from the syringe by cutting off the nozzle and extruding the agarose, slicing it into 2mm (~20 µl) slices with a backed razor blade. The plugs or slices were incubated in NDS with 2mg/ml proteinase K at 55°C for at least 48h, with gentle shaking. Following proteinase K digestion the plugs or slices were washed at least 4 times in a large volume (typically 50x the agarose volume) of 0.5M EDTA. Plugs were stored at 4°C in 0.5M EDTA for up to a year. Aliquots were equilibrated in TlowE for digestion and maintained in this at 4°C for a few months without visible degradation.

2.9 Ethanol precipitation of DNA

0.1 vol of 3 M NaOAc were added to the DNA (if necessary), followed by 2 to 2.5 vol of 100 % ethanol. If the quantity of DNA was very small (typically less than 1 µg) 10 µg glycogen were also added. Following incubation for at least 10 min on ice or at -20 °C, the precipitated DNA was recovered by centrifugation, typically at 13,500 rpm in a microfuge, but larger volumes were centrifuged for 30 min at 3000 rpm in a Beckman GPR centrifuge, or at 15,000 rpm for 30 min in a Beckman J2-21 centrifuge using the JA-20 rotor.

Precipitated DNA was typically washed once or twice with 70 % ethanol before brief (if any) air drying and redissolving in TE or TlowE, or occasionally water.

2.10 Quantification of DNA

DNA was usually quantified using the GeneQuant spectrophotometer from Pharmacia, using the

capillary cuvette and following the manufacturer's instructions.

YAC DNA was quantified by electrophoresis alongside samples of lambda bacteriophage (λ) DNA of known concentration, from a proprietary stock (Boehringer Mannheim).

2.11 Enzymatic modifications of DNA – restriction digests, ligations etc

Restriction digests of low MW DNA were carried out according to the supplier of the restriction enzyme and using the supplier's restriction buffer unless otherwise stated. Double digests were performed on a trial and error basis having consulted manufacturer's literature as to the most suitable restriction buffer.

Restriction digestion of yeast DNA in agarose plugs was essentially according to Hamvas *et al.* (1994). Restriction digests were supplemented with 0.5mg/ml BSA (NEB), 1.25μM DTT and 2.5μM spermidine. Reaction volumes were 200μl per 20μl plug, and plugs were pre-equilibrated in reaction buffer on ice once for 20 min before the enzyme was added, then again for 20 min on ice before transfer to 37°C. This was to allow the enzyme to enter the agarose; some enzymes have short half lives at 37°C. Partial digests were typically incubated for 30 min. Complete digests were incubated overnight. Digestion was stopped by the addition of 0.5 volumes 0.5M EDTA.

Ligations were performed using the Ready To Go Ligase lyophilised reaction mix tubes from Pharmacia (now Amersham Pharmacia Biotech Inc.) and according to the manufacturer's instructions.

Dephosphorylation was performed using shrimp alkaline phosphatase (SAP) from New England Biolabs according to the manufacturer's instructions except that a 10 fold excess of enzyme was used. SAP reactions were stopped by heating for 15 min at 75°C.

Filling in of 5' overhangs was performed with the Klenow enzyme (Boehringer Mannheim) according to the manufacturer's recommendations. 3' overhangs were removed with the same enzyme.

2.12 Gel-purification of small DNA fragments

This was done using the Prepagene kit from BioRad. The DNA band was excised from the agarose gel and incubated with 3 volumes of Prep-A-gene binding buffer (50mM Tris pH 7.4, 1mM EDTA, 6M $NaClO_4$) for 5min, at 55^{0}C. 5µl of Prep-A-gene matrix was added per µg of DNA and incubated for 5-10min at room temperature followed by centrifugation in a microfuge for 30sec. The pellet was washed 3 times in 50x volume of binding buffer, followed by three washes in 20mM Tris pH 7.4, 2mM EDTA, 0.4M NaCl, 50% Ethanol. After a final spin to dry the pellet, the DNA was eluted from the matrix in 1 pellet volume of either H_2O or TE, 5min, at 55°C. Two elutions were usually performed to improve yield. The eluted material was recovered by centrifugation.

2.13 DNA Sequencing

Double stranded DNA was sequenced using a Perkin Elmer ABI377 Automated DNA Sequencer with 48 cm well-to-read gel plates and a 36-well comb. Sequencing reactions were prepared using either the Dye Terminator, dRhodamine Dye Terminator or BigDye Ready Reaction sequencing kits from Perkin Elmer and following the manufacturer's instructions. Typically 1 to 2 µg of plasmid DNA was included with 4 pmol of primer. Reactions were cycled on an MJ Research PTC-200 thermal cycler without oil and using the heated lid as follows:

25 cycles: 94°C, 30 sec; 50 °C, 30 sec; 60°C, 1 min.

The block was preheated to 95°C before adding the reaction tubes and starting the reaction. Completed reactions were precipitated by the addition of 0.1 vol 3M NaOAc and 2.5 vol of 90%

ethanol and incubation on ice for 10 min. Following centrifugation at 13,500 rpm in a benchtop microcentrifuge the supernatant was removed and the pellet resuspended in 4 µl of ABI Loading Dye and stored frozen for up to two weeks before electrophoresis. Samples were denatured at 95°C in a thermal cycler for 2 min and snap cooled on ice immediately prior to loading onto the gel.

The sequencing gel was prepared according to the manufacturer's instructions, using the Long Ranger Singel single-pack gel kits from FMC. 1.5µl of each reaction was loaded onto the gel and run out for 10 h using the 48 cm standard conditions recommended by the manufacturer.

Sequencing data was generated from the raw data (ABI gel file) using Sequence Analysis, trimmed for ambiguous sequences or vector sequences using Factura; and aligned into contigs using AutoAssembler, all from Perkin Elmer. Other sequence analysis software was from Lasergene.

It was generally found that miniprep DNA prepared by alkaline lysis/phenol sequenced well, but Wizard miniprep DNA did not. This may be due to high salt carryover, or to resin fines which escape some batches of Wizard columns. In most cases good sequence could be obtained after a clean-up step on a proprietary column such as the High Pure kit from Behringer Mannheim. CsCl- and Qiagen-purified DNA gave very high quality sequence data.

2.14 Radiolabelling of double-stranded DNA

The "Random Primed DNA Labelling Kit" (Böehringer Mannheim) was employed for incorporation of ^{32}P into the randomly primed, and newly synthesised complementary DNA strand. The reaction was carried out as recommended by the supplier. Typically a 20µl reaction contained 20-50ng of DNA, 0.5 nmol each of dATP, dGTP and dTTP, and 0.5nmol of [α–^{32}P]dCTP (NEN). Unincorporated ^{32}P nucleotides were separated from the DNA using a Sephadex G-50 column according to the manufactures recommendations (Nick Column, Pharmacia). Specific activity of probes was measured by liquid scintillation using a Beckman LS500 CE. Typically $5x10^7$ cpm were added to each hybridisation. The labelled probe was denatured by a 10 min incubation at RT with 0.1 volumes of 5M NaOH.

2.15 Radiolabelling of oligonucleotides

Oligonucleotides were labelled with γ–^{32}P using polynucleotide kinase (PNK) from Boehringer Mannheim. 50 pmol of oligonucleotide in 19 µl of water were added to 3.0 µl of 10x PNK buffer (Boehringer Mannheim), and 5.0 µl of [γ–^{32}P]dATP (NEN). 2.0 µl of PNK were added and the reaction was incubated at 37°C for 30 min before being added directly to the prehybridised filter in hybridisation solution.

2.16 Polymerase chain reactions (PCR)

2.16.1 Amplification of murine β casein probes β3 and β5.

The reaction was carried out in a 100µl volume overlaid with silicone oil in a Hybaid Omnigene thermal cycler. The reaction comtained 1x PCR buffer (2.25mM $MgCl_2$, 50mM Tris.HCl pH 9.2, 160mM $(NH_4)_2SO_4$, Boehringer Mannheim), 200ng of murine genomic DNA, 200 pmol of dNTPs, 0.4 pmol of each oligonucleotide primer, and 2 U of *Taq* DNA polymerase. The cycling conditions were as follows:

1cycle: 94°C, 2min; 53°C, 30 sec; 70°C, 90 sec;
39 cycles: 94°C, 1 min; 53°C, 30 sec; 70°C, 90 sec.

2.16.2 Amplification of human β casein probe from human genomic DNA

The reaction was carried out as above except that the template was 200ng of human placental DNA, and the cycling conditions were as follows:

1cycle: 94°C, 2min; 46°C, 30 sec; 70°C, 50 sec;
39 cycles: 94°C, 1 min; 46°C, 30 sec; 70°C, 50 sec.

2.16.3 Amplification of TRP1 *for construction of pURACorp3*

The reaction conditions were as above, except that 1 ng of pRS314 (Sikorski and Hieter, 1989)

plasmid DNA was used as template. The completed PCR reaction was passed over a PCR Select III column (5-3 Prime Inc.) in accordance with the manufacturer's instructions to purify the amplified DNA before cloning.

2.16.4 Amplification of HIS3 *for retrofitting of yATS1*

HIS3 was amplified from pRS313 (Sikorski and Hieter, 1989) as described for *TRP1* above. Prior to transformation into yATS1 yeast cells (Section 5.4.2) the PCR product was passed over a Microcon column (Amicon) in accordance with the manufacturer's instructions.

2.17 Reverse transcription polymerase chain reactions (RTPCR)

2.17.1 Amplification of murine casein cDNA probes

Murine casein cDNA probes were amplified from lactating mammary gland total RNA using Superscript II RNaseH$^-$ reverse transcriptase (Life Technologies). The first strand reaction was carried out in a 0.5ml microcentrifuge tube containing 1 µg total RNA, 1x PCR buffer (2.25mM $MgCl_2$, 50mM Tris.HCl pH 9.2, 160mM $(NH_4)_2SO_4$, Boehringer Mannheim), 25 pmol 3' primer and 1mM dNTPs, in a 20 µl volume overlaid with silicone oil. The reaction mix was heated in an Omnigene thermal cycler (Hybaid) to 95°C for 5 min, followed by snap cooling on ice. 200U of Superscript II RnaseH$^-$ reverse transcriptase were added, and the reaction incubated at 42°C for 30 min. To the completed first strand reaction was then added 25 pmol of the 5' primer, 8 µl of 10x PCR buffer, 2U of *Taq* DNA polymerase, and the volume made up to 100 µl with water. The PCR cycles used were as follows:

το amplify α casein cDNA:
1 cycle: 94°C, 2 min; 62°C, 30 sec; 72°C, 1 min;
39 cycles: 94°C, 1 min; 62°C, 30 sec; 72°C, 1 min;

to amplify γ, ε or κ casein:
1 cycle: 94°C, 2 min; 62°C, 30 sec; 72°C, 42 sec;
39 cycles: 94°C, 1 min; 62°C, 30 sec; 72°C, 42 sec.

2.17.2 RTPCR to detect human casein gene expression in HP8 mice

RTPCR to detect human casein transcripts in HP8 mouse lactating mammary gland total RNA was performed using the Superscript II RNaseH$^-$ reverse transcriptase (Life Technologies) and following the manufacturer's protocol for random primed first strand generation and subsequent PCR amplification. The PCR reaction was carried out in 0.5ml thin-walled reaction tubes (Perkin Elmer) using an MJ Research PTC200 DNA Engine thermal cycler with a heated lid. A hot start was effected by omitting the *Taq* DNA polymerase until after the first 94°C step. The conditions were as follows:

94°C, 3 min, then hold at 80°C (add *Taq* polymerase);
40 cycles: 94°C, 20 sec; 55°C, 25 sec; 72°C, 30 sec.

2.18 Electrophoresis of DNA

Except where stated (for RNA gels and PFGE), DNA gels were poured and run in appropriate concentrations of agarose (MP agarose from Gibco BRL Life Technologies or FMC) according to standard methods. The running buffer was TAE. DNA was visualised in agarose gels by the inclusion of 10µg/ml EtBr in the gel and subsequent photography with a Polaroid land camera on a Srtatagene UV transilluminator. Where migration of the EtBr out of the gel necessitated it, the gel was restained in running buffer containing 1µg/ml EtBr.

2.19 Southern transfer of DNA

Agarose gels were treated with 120mJ/cm^2 of 254nm UV light in a Spectronics Spectrolinker UV crosslinker to nick the DNA. The DNA was denatured by a 30 min incubation in 0.5M NaOH/1.5M NaCl and neutralised in Neutralising solution. Transfer in 10xSSPE was by conventrional Southern blot or downward transfer using the TurboBlotter from Schleicher and Schuell. PFGE gels were transferred exclusively by downward transfer as the reduced pressure on the gel reduces crushing and results in more complete transfer. All transfers were carried out overnight. In conventional Southern blotting, a weight of no more than ~300g was used to avoid crushing the gel. Filters used were either Duralon UV (Stratagene), Nytran (Schleicher and Schuell) or Sartolon (Sartorius). DNA was crosslinked to the filter with 120mJ/cm^2 of 245nm

UV light.

2.20 Hybridisation of Southern blots and colony lift filters

Southern blots and colony lift filters were hyrbidised according to Church and Gilbert (1984). Prehybridisation and hybridisation were at 65°C in 0.5 M NaPi/7% SDS/1 mM EDTA. Filters were washed twice for 10 min at 65°C in 2x SSPE/0.1% SDS, followed by one or more washes of 10 min at 65°C in 0.2x SSPE/0.1% SDS.

2.21 Autoradiography

Autoradiography was carried out at –70°C with intensifying screens, using either Kodak X-OMAT-AR or Amersham Hyperfilm MP autoradiography. Of these Hyperfilm MP gave better results owing to its high transparency with sensitivity equal to that of X-OMAT-AR.

2.22 RNA preparation

Total RNA was prepared from tissue or cultured cells using either RNAzol (Invitrogen) or TRI reagent (Sigma). In each case the manufacturer's protocols were followed. RNA was stored at -70°C or, for short periods, at -20°C, in 1mM EDTA. Tissue was homogenised in an Ultraturrax homogeniser at 23,000 rpm in RNAzol or TRI reagent.

2.23 Northern blots

RNA was prepared at least 2.5 μg/μl. If more dilute, the RNA sample was concentrated by brief drying in a SpeedVac vacuum centrifuge (Savant) on low heat.

To 10μg of RNA in either 2 or 4μl of 1mM EDTA, two volumes of buffer A (2.9x HEPES buffer pH7.8, 71% v/v formamide) and 2.3 volumes of F/F (9.7% v/v formaldehyde, 74% v/v formamide) were added and the mixture heated for 10 min at 70°C then quenched on ice. Either 1.5 or 3.0μl of RNA loading buffer (1x HEPES, pH7.8, 50% v/v formamide, 8% v/v formaldehyde, 5mg/ml bromocresol green, 400mg/ml sucrose) was added and mixed well.

Denaturing agarose gels were prepared in 1x HEPES buffer with formamide at 6%. The gel was allowed to set for at least an hour in a fume hood wrapped in Saranwrap (Dow Corning). Electrophoresis was at 10 v/cm for 3 to 4 hours, using RNA markers from Gibco BRL Life Technologies to monitor migration. In some cases the gel was EtBr-stained to visualise the RNA, but in others it was prepared unstained. No difference in RNA integrity was observed between these two procedures. The gel was then transferred to Nytran (Schleicher and Schuell) membrane overnight in 10x SSPE using the TurboBlotter downward transfer system from Schleicher and Schuell. The filter was crosslinked by 120 mJ cm^{-2} of 254nm UV light in a Spectrolinker UV crosslinker from Spectronics Corp. Prehybridisation and hybridisation were performed according to Church and Gilbert (1984) at 42°C.

Blots probed with double stranded DNA were washed for 10 min at RT in 2x SSPE, followed by 2 washes at 55°C in 2xSSPE/0.1% SDS. High stringency washes were performed at 60°C with 0.2xSSPE./0.1% SDS

Blots probed with oligonucleotides were washed twice at RT for 10 min in 6x SSPE, 0.1% SDS, followed by a single 10 min wash in 6x SSPE, 0.1% SDS at 42°C.

2.24 Pulsed Field Gel Electrophoresis

PFGE was performed using a CHEF DRII contour-clamped homogenous electric field PFGE apparatus from BioRad. Gels were prepared in 18 MΩ water as 1% Rapid Agarose or 1% SeaKem Gold; Rapid Agarose was no longer available after 1996. SeaKem Gold gives very similar results but results in slightly reduced mobilities, such that gel conditions were adjusted slightly. Gels were prepared and run in 0.5x TBE buffer.

For a separation, of 50-500 kb fragments in Rapid agarose, such as is shown in Figure 4.2, the gel was run at 14°C for 22h at 200V with the switch time ramped from 5s to 30s. Preparative PFGE was performed with these conditions (see chapter 6). With SeaKem Gold the final switch time was raised to 35s (as in Figure 5.6).

Gel conditions for lower MW separation (as in Figure 4.5) also involved 1% Rapid Agarose cast and run in 0.5x TBE. The gel was run for 15h at 200V with switch time ramped from 0.5s to 2.5s.

Pulsed field gels were stained for 30 min in 1 μg/ml EtBr and photographed using either a CCD camera (Mitsubishi) or a Polaroid land camera, both from UVP, on a longwave UV transilluminator (Stratagene). Pulsed field gels were Southern blotted onto Nytran membrane (Schleicher and Schuell) by downward transfer using a TurboBlotter apparatus (Schleicher and Schuell) as for normal gels (Section 2.19).

2.25 Preparation of chemically competent E. coli for transformation.

A 10 ml overnight culture of bacteria was inoculated into 250 ml of prewarmed LB and grown until the OD_{600} reached 0.5 (for DH5α and XL1-Blue strains). This is indicative of the late-log phase of growth. The cells were then chilled on ice for 5 min and harvested in a precooled rotor in a Beckman J2-21 centrifuge for 5 min at 10,000 rpm. Subsequent steps were performed in a cold room (4 °C).

The cells were resuspended in 100 ml ice-cold TFBI (30mM potassium acetate, 100mM RbCl, 10mM $CaCl_2$, 50mM $McCl_2$, 15% w/v glycerol, pH 5.8 via acetic acid) and incubated on ice for 10 min before being harvested as before. The pellet was resuspended in 10 ml ice-cold TFBII (10mM MOPS (3-[N-Morpholino]butanesulfonic acid), 75mM $CaCl_2$, 10mM RbCl, 15% w/v glycerol, pH 6.5 via KOH) and incubated on ice for 15 min, then snap-frozen in liquid nitrogen as 200 μl aliquots in prechilled Eppendorf tubes. The tubes were stored at -70 °C and thawed on ice immediately before use. The transformation efficiency was tested by transformation with a known quantity of a test plasmid as below.

2.26 Transformation of chemically competent E. coli

Competent cells were thawed on ice and split into 100 μl aliquots. Up to 1 μg of transforming DNA was added and mixed gently, and the tube was incubated on ice for 20 min. Following a 90 sec heat shock at 42 °C, 400 μl LB or 2YT were added and the transformations were incubated at 37 °C with agitation for 30 min to 1 h, then plated at an appropriate density

(depending on postulated transformation efficiency) onto selective medium (typically LB plates supplemented with 50 μg/ml ampicillin).

2.27 Yeast colony lift protocol

Yeast colony lifts were performed from freshly grown plates. Yeast cells were patched or spotted onto a rich medium plate and grown for 1-2 d. However, it was found that overgrowth of the patched cells could inhibit lysis after transfer (Section 3.4). A 90 mm disk of Hybond N+ membrane (Stratagene) was laid onto the plate for 10 minutes, and marked for orientation. The filter was then laid yeast side up onto Whatman 3MM paper soaked in SCE supplemented with 8mg/ml Novozym 234 (Novo Nordisk) and 10mM DTT, and incubated in wrapped dishes for 4 h at 37°C to allow digestion of the yeast cell walls. The filter was then transferred to 3MM paper soaked in denaturant (0.5N NaOH, 1.5M NaCl) for 5 min, then air dried for 5 min on fresh 3MM paper. The filter was then floated on neutralising solution (1M Tris pH 7.5, 1.5M NaCl) for 5 min, then on 0.1M Tris (pH 7.5), 0.15M NaCl for 5 min, then submerged in this solution for 2 min. The filter was then incubated in fresh 0.1M Tris (pH 7.5), 0.15M NaCl supplemented with 0.25mg/ml Proteinase K at 55°C for 1 hour with gentle agitation. Any remaining colony debris was wiped off with a tissue soaked in 0.1M Tris (pH7.5), 0.15M NaCl and the filter was briefly rinsed in this solution before the DNA was UV-crosslinked onto the filter.

2.28 Yeast Strain Mating

2.28.1 Mating on solid medium

Haploid strains of opposite mating types (a and α) were patched onto the same region of a plate of suitable selective medium (*e.g.* histidine selection for mating of *his3* and *his5* strains). This plate was incubated at 30 °C for 3-4 h, after which the mated cells were inoculated into 100 μl YPDA and plated onto dropout medium (typically selecting for histidine prototrophy which would mark diploid cells).

After incubation overnight at 30 °C, single colonies were picked and patched onto selective plates to confirm the desired phenotype. Clones of this phenotype were then sporulated as follows.

2.28.2 Random sporulation

A large freshly grown diploid colony was inoculated into 5 ml of YPDA and grown at 30 °C for 2 d with shaking. The cells were then pelleted and resuspended in 2 ml sporulation medium, supplemented as required for the particular strain. The cells were incubated in this medium for 3 d with shaking.

The cells were then pelleted and resuspended in 0.5 ml water, to which was added 0.5 ml diethyl ether. The tube was vortexed for 1 min, and left at RT for 5 min. The aqueous phase was then removed into a fresh tube and 100 µl was spread onto a YPDA plate. 100 µl water was pippetted onto two more YPDA plates, and a rough dilution series was achieved by spreading this water with the spreader from the first plate, without flaming.

The plates were incubated for 7-10 d, after which colonies were patched onto dropout plates to screen for the desired phenotype. In some cases backcrosses were useful, such as to distinguish between *his3* and *his5*, and these were performed by mating on solid medium.

2.29 Yeast transformation

Yeast cells were transformed by lithium acetate transformation or spheroplast transformation.

2.29.1 Lithium acetate transformation

This is based on the method of Ito *et al.* (1983) with modifications based on those of Markie (1997). The host strain was grown overnight in YPD broth supplemented with adenine. The 20 ml overnight culture was diluted to an OD600 of 0.1 in YPDA and grown to an OD600 of approximately 1.0. The cells were then harvested by centrifugation at 2000 rpm for 5 min in a 50 ml Falcon tube in a Beckman GPR centrifuge, washed twice with UHP water, and transferred to an Eppendorf tube in 1 ml water. The cells were then washed with 1 ml of freshly made TE/LiOAc and resuspended in TE/LiOAc at about 2x109 cells/ml.

50 µl of cells were mixed with up to 1 µg of transforming DNA and 50 µg of sheared, single

stranded herring sperm DNA (average size 500 bp). 300 µg of PEG Solution were added and mixed well, and the cells were incubated at 30 °C for 30 min with agitation. Following a 15 min heatshock at 42 °C, the cells were spun down for 5 sec at top speed in a microfuge and resuspended in 1 ml TE, then plated onto selective medium, 200 µl/plate. Plating the cells more densely than this results in fewer colonies due to cell death. Transformation efficiencies of 10^3 to 10^4 cfu/µg were achieved by this method.

2.29.2 Yeast spheroplast transformation

A single freshly-grown yeast cell colony was picked from a rich medium plate, and used to inoculate a 30 °C overnight culture in YPDA. The following morning this was diluted for a starting OD_{600} of 0.5, and grown until the OD_{600} reached 2.0.

The cells were pelleted in 50 ml Falcon tubes by centrifugation in a GPR centrifuge (Beckman) or similar, at 1000g for 5 minutes. The cells were washed once with 20 ml/tube of UHP water and then 10 ml 1M sorbitol, then resuspended in 10 ml SCE plus β mercaptoethanol (20 µl/10 ml SCE).

25 µl of 10 mg/ml Zymolyase 20T (ICN Biomedicals Inc.) were added to each tube, and the tubes were incubated at 30 °C for about 20 mins. The percentage spheroplasting was assessed as follows (protocol kindly provided by Brenda Grimes):

a) 100µl cells (using cell saver tips) were added to 900µl water, mixing well.
b) 100µl cells were added to 900µl 1M sorbitol, mixing well.
c) The OD_{660} of cells in water (c.w) and of cells in sorbitol (c.s) was read, using a water blank.
d) The percentage spheroplasting is given by:

$$\%\text{age spheroplasting} = 100 - [\,(OD_{c.w} / OD_{c.w}) \times 100\,]$$

The digestion was continued until the percentage of spheroplasts reached about 50%. The cells were then pelleted at 400g for 3 mins, and resuspended gently in 10ml STC (1M Sorbitol, 10mM Tris pH7.4, 10mM $CaCl_2$). At this point the cells were usually examined under a

microscope to look for contamination and to check the spheroplasting by the addition of 10% SDS (the spheroplasts burst, whereas the walled cells do not).

The spheroplasted cells were then transferred to 1 ml STC, and either used in a transformation, or DMSO was added to the cells to 7%, and 100 µl aliquots of the cells were stored at -70 °C. Some loss of transformation efficiency was seen after freezing. It was found that snap-freezing in N_2(l) was detrimental to the transformation efficiency compared to simply transferring to a –70°C freezer.

2.29.2.1 Transformation of spheroplasts

100µl of spheroplasts were mixed with 1µg transforming DNA plus 50µg carrier DNA in no more than 10µl. After 10 min at RT 10 volumes of PEG/$CaCl_2$ (20% w/v PEG_{3350}, 10mM Tris pH 7.4, 10 mM $CaCl_2$) were added and mixed thoroughly. After a further 10 min at RT the cells were pelleted by gentle centrifugation (2000 rpm for 3 min). The cells were gently taken up in 0.5ml of SOS (1M Sorbitol, 6.5mM $CACL_2$, 0.25% yeast extract, 0.5% bactopeptone) and added to molten top agar (SD agar plus 1M Sorbitol) at 55°C. The top agar was immediately poured onto the surface of a regeneration plate (SD agar plus 1M Sorbitol) with the appropriate selection. Transformation efficiencies of 10^4 cfu/µg were achieved by this method.

3. Chapter 3 - Screening and clone characterisation

3.1 Summary

The C3H mouse YAC library generated by Larin *et al.* (1991) was screened by colony hybridisation using two murine βcasein genomic probes. Fifteen positive clones were identified. These 15 clones were screened with cDNA probes for murine α, γ, εand κ caseins by colony hybridisation and Southern blotting. Two clones which hybridised to all probes were selected for further characterisation.

The large insert size of YACs has led to some developments in the methods of maintenance and screening of YAC libraries relative to libraries of smaller insert size. YACs can achieve the same coverage as lambda (λ) or cosmid libraries with a tenth of the number of recombinants. This has resulted in YAC libraries being stored and screened as arrays of individual clones rather than as random collections of recombinants. Libraries of smaller inserts are randomly plated out and screened by hybridisation to replicas. In contrast, YAC libraries are stored as single clones in individual microtitre plate wells. Screening of these clones can be accomplished via PCR-based strategies first reported by Green and Olson (1990), where specific superpools of clones form multidimensional arrays, allowing the screen to isolate a single YAC clone in a manageable number of steps; the exact number of steps depends on the size of the library and the screening strategy. An alternative and often complementary strategy involves screening the library by colony filter hybridisation. The reliability of this strategy has been greatly enhanced by the use of specialised robotic equipment for gridding the clones out onto the filter in precisely-defined high-density arrays (Nizetic *et al.* 1991; Bentley *et al.* 1992; McKeown *et al.* 1993). Screening of these arrays can isolate a single clone in one step, and depending on the gridding pattern the use of duplicate clones on the same array can greatly reduce false positives. To obtain the clones used in this project the ICRF mouse YAC library (Larin *et al.* 1991) was screened by colony filter hybridisation. This work was done by M. Dalrymple and R. Cox.

3.2 Probes

3.2.1 Genomic probes

The genomic murine β casein probes were prepared from murine genomic DNA by polymerase chain reaction (PCR) as described in Chapter 2. The oligonucleotide primers used are given in Table 3.1. The PCR products were cloned into pGemT (Promega) and sequenced using an ABI377 automated DNA sequencer (chapter 2). No significant differences were found between the PCR products and the published sequence for murine β casein (Genbank accession numbers X13484 X15991 X15992). The extent of the murine β casein probes is shown in Figure 3.1. β5 includes exons 2, 3 and 4 and extends into intron 4; β3 includes part of intron 6, all of exon 7 and intron 7, and part of exon 8. Each of the two probes was hybridised to a Southern blot of genomic murine DNA to assess specificity (Figure 3.1). The bands obtained were as expected. In particular, the small portion of β3 which was 3' to the internal EcoRI site detected the expected 1.4 kb band, demonstrating the sensitivity of the blot.

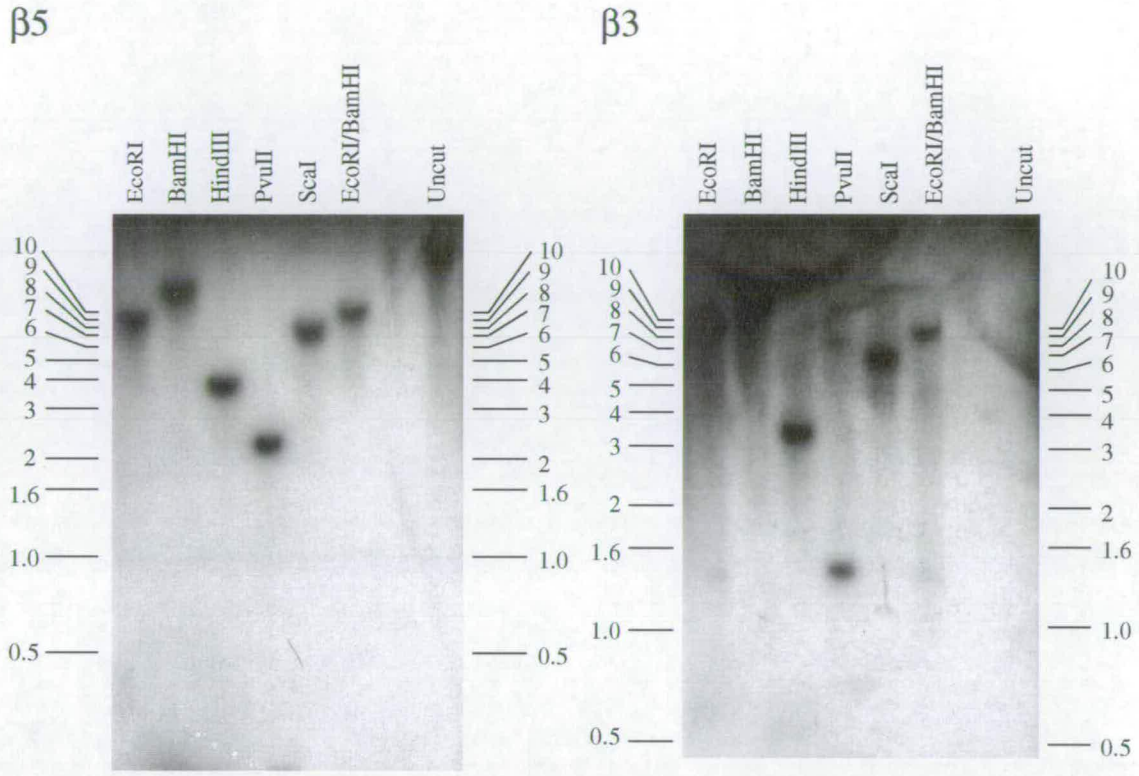
The murine α, γ, and κ casein probes were amplified from lactating mouse mammary gland total RNA using the SuperscriptII RTPCR kit from Life Technologies (See Chapter 2). The RTPCR primers used are shown in Table 3.1. The RTPCR products were cloned into pGemT (Promega) and sequenced as above. No significant changes from the published sequences were found.

Figure 3.1: Murine β casein probes

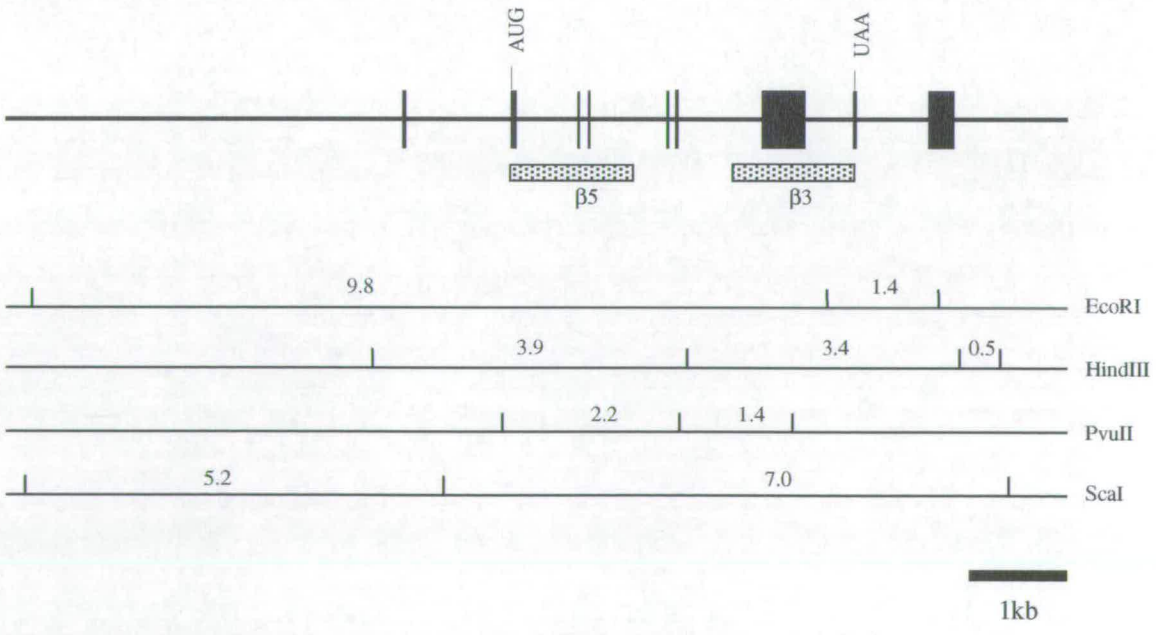
Lower panel: the murine β casein gene. Exons are represented by black boxes. The two screening probes β3 and β5 are shown below. Sizes of predicted restriction fragments are in kb.

Upper panel: β3 and β5 were used to probe two replica blots of mouse genomic DNA digested with various restriction enzymes. All of the bands were as expected (no data are available concerning BamHI sites within the mouse β casein gene).

Figure 3.1: Murine β casein genomic probes



Murine β casein gene



3.3 Primary library screening

The genomic murine β casein probes were used to screen the C3H (male) murine YAC library of Larin *et al.* (1991) by colony hybridisation (Church and Gilbert 1984). This work was carried out by M. Dalrymple and R. Cox at the ICRF by kind permission of R. Cox and H. Lehrach. The screen produced 15 positive clones. These were brought back as stabs in YPD agar.

3.4 Secondary screening

The 15 primary positive clones were renumbered MP08 to MP22 (Table 3.2). Initial secondary screening was by colony filter hybridisation (Church and Gilbert, 1984; Figure 3.2). However, colony lifts proved unreliable and gave false negatives. There was no evidence of false positives (after comparison of several lifts and later PFGE data), which suggests that transfer of DNA onto the membrane was inefficient. Supporting this was the observation that some colonies failed to hybridise to a pYAC4 probe. Therefore the 14 samples were screened by PFGE and Southern blotting. Later it was appreciated that it is important not to overgrow the yeast colonies before transfer onto the filters, as lysis of the yeast is inhibited in the centre of large colonies, resulting in a faint halo-shaped signal after hybridisation (Cole *et al.* 1996). For the small number of clones screened here PFGE and Southern blotting proved practicable and extremely robust. In addition, PFGE would be necessary in any case to check for clones with more than one YAC, and gives information about YAC size.

It can be seen from Figure 3.2 that MP12 and MP20 are positive for murine α, β, and κ caseins. MP18 hybridised to α, β and γ but not to κ casein. The bovine κ casein gene had by this time been mapped to one end of the bovine locus (Ferretti *et al.* 1990; Threadgill and Womack, 1990). The fact that MP18 hybridised to all casein probes except κ casein may indicate that the murine κ casein gene is organised similarly. However, given the variability of the colony hybridisation results in general the possibility that this is a false negative cannot be ruled out, although it seems unlikely given the strong hybridisation of the same MP18 DNA to the other probes in Figure 3.2, The number of samples not hybridising to the

Figure 3.2: Screening YAC clones by colony hybridisation

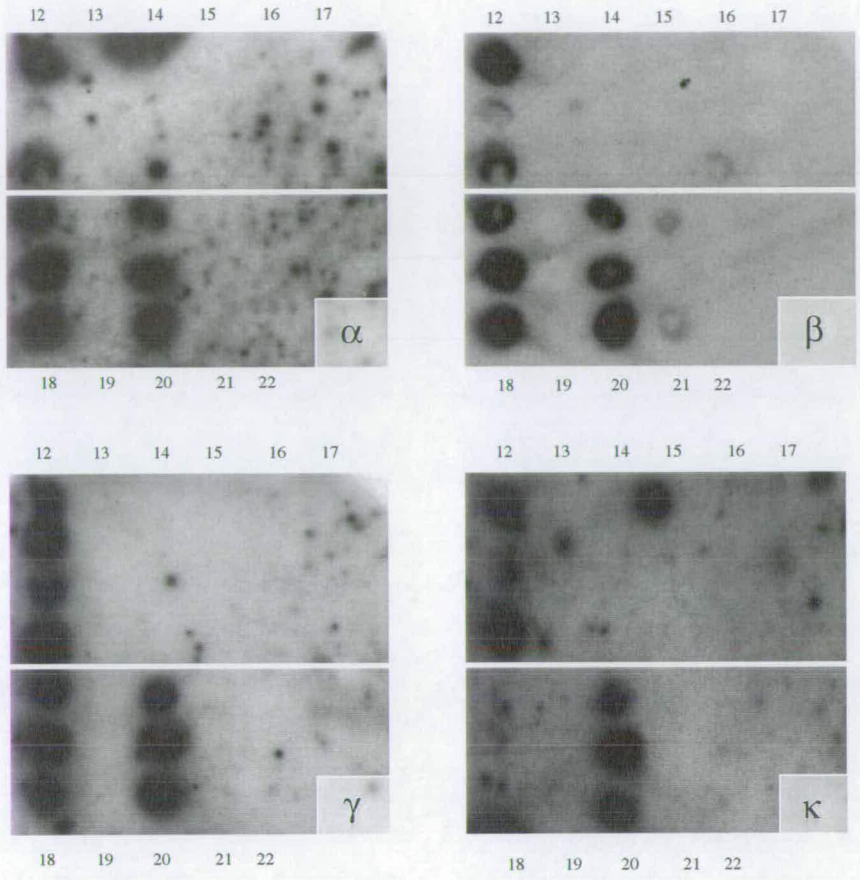


Figure 3.2: Screening of YAC clones by colony hybridisation

Four successive probings (not in order) of the same filter are shown. The probes were the cDNA probes for murine α, γ and κ casein (α, γ, κ panels) and β5 (β). Three colonies of MP clones 12 to 22 were screened such that each clone is represented by three vertically-aligned colonies; the clone numbers appear above and below each panel. MP12 and MP20 are positive for murine α, β, γ and κ casein. MP18 is positive for α, β and γ but appears to be negative for κ casein. It was not possible to screen this filter with ε casein due to buildup of noise.

Figure 3.3: Screening of YAC clones by PFGE and Southern blotting

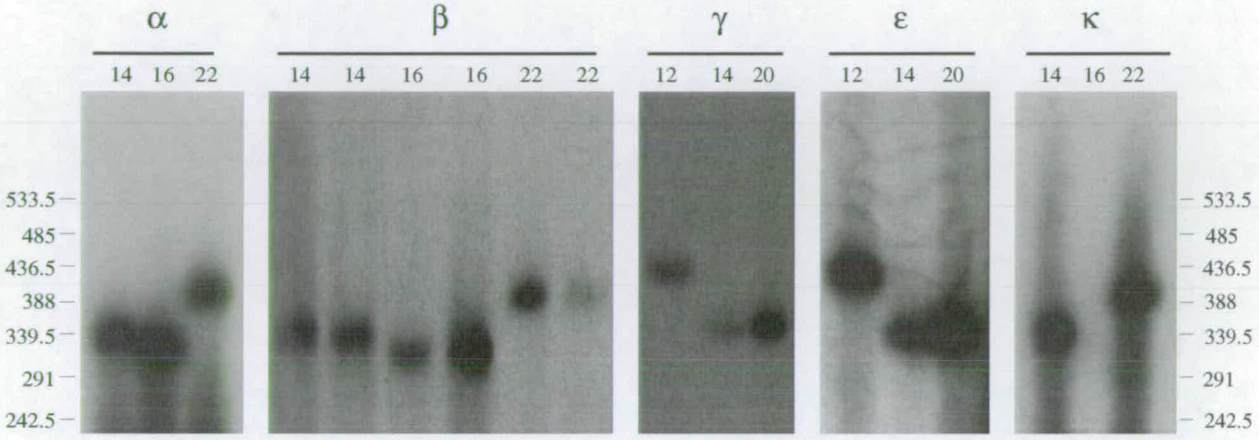


Figure 3.3: Screening of YAC clones by PFGE and Southern blotting

DNA was separated by PFGE, Southern blotted and hybridised to cDNA probes for α, γ, ε and κ caseins (panels α, γ, ε, κ) and β5 (β). MP12 spans approximately 430 kb, MP14 380 kb, MP16 360 kb, MP20 380 kb, and MP22 420 kb. Marker sizes are shown in kb. The markers used were Lammbda Ladder from NEB.

casein probes in these lifts is at odds with other lifts (not shown), while MP20 was negative for β casein in a previous lift.

Southern blotting was then used as a more robust screen of the MP clones. MP14, which appears negative in Figure 3.2, hybridised to all murine casein probes in subsequent Southern blots (Figure 3.3). MP12 hybridised to ε casein (Figure 3.3), which together with the data from Figure 3.2 mean that MP12 is positive for all five murine caseins, as is MP14. From these blots MP12 was estimated at 430 kb, and MP14 at 380 kb. MP16 (360 kb) hybridised to all casein probes except κ. This supports the inference from the MP18 results that κ casein is on one end of the gene locus.

In total four clones were identified which were positive for all casein genes (Table 3.2). At this stage the screening was halted and MP14 (380 kb) and MP12 (430 kb) were selected for further work as they were the first two clones to hybridise to all casein probes. An initial partial restriction digest (shown in Figure 4.4 panel A) probed with the α casein cDNA probe indicates that the two YACs overlapped extensively. The difference between the sizes of the two YACs demonstrates that MP14 and MP12 are independent recombinants. Mapping of the two in parallel would therefore convincingly demonstrate whether either was rearranged with respect to the genome in the region of overlap. No further screening was done.

Probe	Type	Size (kb)	Primer sequences (5'-3')	Accession numbers
α casein	cDNA	0.9	GCAGTTAGCAGTCAAACTCAG C CACTGACCTGGGAGGTAAGA GG	M36780
β casein 5' (β5)	genomic DNA	1.5	GACTTGACAGCCATGAAGGTC T CTTCCTCTGAGGTATTTGCAG C	X13484
β casein 3' (β3)	genomic DNA	1.5	GGAGACACTCCTTAAGAACA CT TAACCTGGAAATCCTCTTAGA C	X15991 X15992
γ casein	cDNA	0.7	CCTCCAGTGAGGAATCATCTG C CCTGACACCTCGGCATAATTG A	D10215
ε casein	cDNA	0.5	AGCAATACATCTCCAGTGAGG A TGACAGAAGTGAAGACGAGG GT	J00379
κ casein	cDNA	0.6	CAGATTCAAACTGCCGTGGTG A GTCTAGAAAGAGCAGAAGGG AA	M10114

Table 3.1 - Murine casein screening probes

Upper primers are displayed above lower primers for each probe.

Library clone	In-house name	Positive for	Negative for	YAC size
ICRFy902C0438	MP08	α, β		
ICRFy902C0338	MP09	α, β		
ICRFy902H0629	MP10	α, β		
ICRFy902B1232	MP11	α, β		
ICRFy902G0781	MP12	α, β, γ, κ,ε		430 kb
ICRFy902E1115	MP13			
ICRFy902C11116	MP14	α, β, γ, κ, ε,		380 kb
ICRFy902B0621	MP15		α, β, γ, κ	
ICRFy902B0721	MP16	α, β, γ, ε,	κ	360 kb
ICRFy902E0540	MP17		α, β, γ, κ	
ICRFy902B0583	MP18	α, β, γ	κ	
ICRFy902A0587	MP19		α, β, γ, κ	
ICRFy902B0469	MP20	α, β, γ, κ,ε		380 kb
ICRFy902G0579	MP21	β	α, γ, κ	
ICRFy902H0556	MP22	α, β, γ, κ, ε		420 kb

Table 3.2 - Murine casein YAC clones

Empty cells indicate screens which were not performed or were inconclusive.

4. Chapter 4 - Mapping the mouse casein gene locus

4.1 Introduction

As discussed in the introduction to this thesis, YACs are often rearranged or deleted spontaneously. For this reason it is paramount to ascertain that any YAC represents a single contiguous genomic region. Some form of physical characterisation of the YAC and comparison to the genomic region is therefore indicated. Fluorescent *in situ* hybridisation (FISH) of the entire YAC to genomic metaphase DNA can indicate whether or not the YAC contains a single contiguous genomic region since a chimaeric YAC would be expected to hybridise to more than one chromosomal locus. However, FISH is technically demanding and expensive, and moreover was not an established technique within the group or facility. Screening for markers gives some indication, especially if several overlapping clones are screened, of the structure of each clone. However, restriction mapping using internal and end-probes gives the same information as marker scoring with additional information about the physical structure of each clone. Since four YACs had already been scored for all five casein genes in Chapter 3, restriction mapping was chosen to characterise them further.

Restriction mapping of YACs and comparison of the data with data from genomic DNA is the most direct method of comparing YAC and genome. However, mapping single-copy genes by PFGE and Southern blotting of large genomic DNA fragments is less straightforward than with yeast DNA, because the smaller size the yeast genome effectively concentrates the target sequence. A faster, more convenient method is to map two or more similar YACs and compare the maps. If the maps are essentially collinear within the region where the YACs overlap, it would be very unlikely that any YAC contained any deletions or rearrangements within that region. This is the approach that is reported here.

MP12 (ICRFy902G0781) and MP14 (ICRFy902C11116), which hybridised to all five of the known murine casein genes (chapter 3) were restriction mapped in parallel and compared to each other. Partial restriction digest and the *LA* and *RA1* probes were used to rapidly generate a restriction map from a single Southern blot (Hamvas *et al.* 1994; Figure 4.2). To further enhance the accuracy of the map, the same Southern blot was reprobed wherever possible. Although this leads to some diminution in signal quality from later probings it has the advantage

that gel to gel variation is minimised, and direct ordering of adjacent restriction sites is facilitated. The two YACs overlapped extensively and within the region of overlap (some 300kb) are collinear. The casein genes are arranged on approximately 250 kb of DNA in the order α–β–γ–ε–κ; α and β casein are convergently transcribed. This work constitutes the first report of the physical structure of the murine casein locus (Tomlinson *et al.* 1996; Appendix I). Two more reports have followed (Rijnkels *et al.* 1997a; George *et al.* 1997) which are broadly in agreement. The discrepancies are discussed.

4.2 Probes

The left and right arm probes (*LA* and *RA1* respectively) were derived from pBR322 (Genbank accession: 701749; Sutcliffe, 1979; Stratagene), which is the base vector used in the construction of the pYAC4 cloning vector which is the basis for the YACs in the ICRF library. pBR322 was cut with EcoRI and PvuII, which bisects the plasmid. The 2.3 kb fragment was purified as the *LA* probe, and the 2.1 kb fragment as the *RA1* probe. Unfortunately, it was noted after the completion of this work that the *RA1* probe in fact contains a small amount of sequence, between the EcoRI site and the BamHI site, which is homologous to the sequence between the centromere and cloning site on the right arm of pYAC4. This region will therefore map to the left of the insert in pYAC4-derived YACs and might be expected to yield some faint spurious bands (see section 4.5) However, comparison with other published maps suggests that this has not substantially compromised the results presented here.

The murine casein gene probes are described in more detail in Chapter 3. The probes for murine β casein are shown in Figure 4.2. β3 and β5 were generated by PCR from murine genomic DNA (Table 3.1). The probes for mouse α, γ, ε and κ casein were cDNA probes generated by RTPCR from total lactating mouse mammary gland RNA using the primers in Table 3.1.

Figure 4.1: Genomic YAC mapping probes

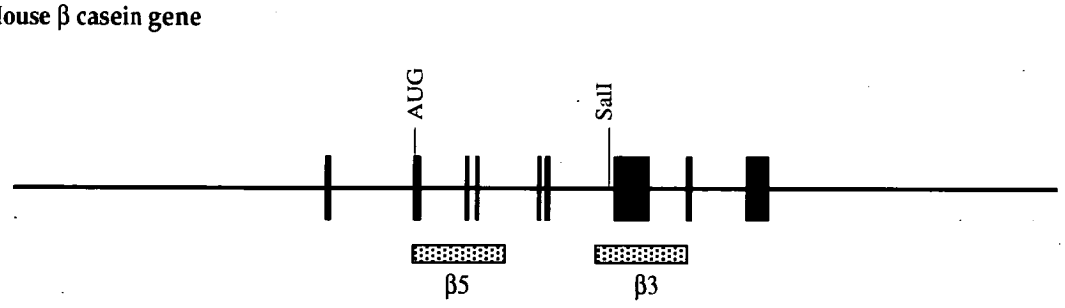


Figure 4.1: Genomic YAC mapping probes

Two genomic probes for murine β casein, β5 and β3, were amplified by PCR. Exons are shown as black boxes. Probes are shown as shaded boxes. The SalI site within the β3 probe is shown.

4.3 Calculation of restriction fragment sizes

Band mobilities were measured manually. A third order polynomial standard curve of bacteriophage lambda DNA (λ) concatemer mobility against size was generated using plotting software (Cricket Graph) on an Apple Macintosh. The fitted curve formula from this plot was used to calculate fragment sizes from mobilities of the sample bands. Due to slight differences between the standard curves from the λ ladders on the left and right sides of the gels used, the digests which ran on the left of the gel were calibrated using the standard curve from the left ladder while those on the right were calibrated using the right hand ladder standard curve. Table 4.1 shows the fragment sizes calculated from Figure 4.2.

4.4 A restriction map of the mouse casein gene locus

A single blot (Figure 4.2) was used to generate the restriction map of the locus. The filter was probed with the *LA* and *RA1* probes separately. The restriction fragments detected by *LA* and *RA1* are listed below. When the MP12 and MP14 PmeI band patterns are represented as maps (Figure 4.3), it is clear that the two maps can be aligned so that the 121 kb MP14 PmeI band is aligned with the 105 kb MP12 PmeI band. The alignment is almost perfect over the region of overlap, varying by less than 5 kb. By this analysis the 26 kb MP14 PmeI band should fall some 10 kb from the left telomere of MP12 and therefore give a band in the MP12 lanes, but, given that there are 6 kb of vector sequence at this end, it is possible, within the accuracy of the blot below 48.5 kb, that MP12 does not contain this PmeI site. Because of the gel conditions and the λ markers used in this gel, the accuracy of the blot is compromised below 48.5 kb.

The SalI, XhoI and ClaI patterns were treated the same way, to generate a four enzyme map of the locus from the *LA* autoradiograph. The same was done for the *RA1* band patterns, and then the two maps were aligned about a central site; the PmeI site 206 kb from the left arm of MP14 was chosen arbitrarily, and aligned with the PmeI site 249 kb from the right arm of MP12 (Figure 4.2). The agreement between the two maps is almost perfect with an error of about 5 - 10 kb (Figure 4.3). However, the following ambiguities remain: the XhoI site at 220 kb was positioned at 230 kb by the *RA1* map, but at 215 kb by the *LA* map (this was the only site that diverged to this extent between the two maps); the pattern of MP14 bands cut with XhoI was

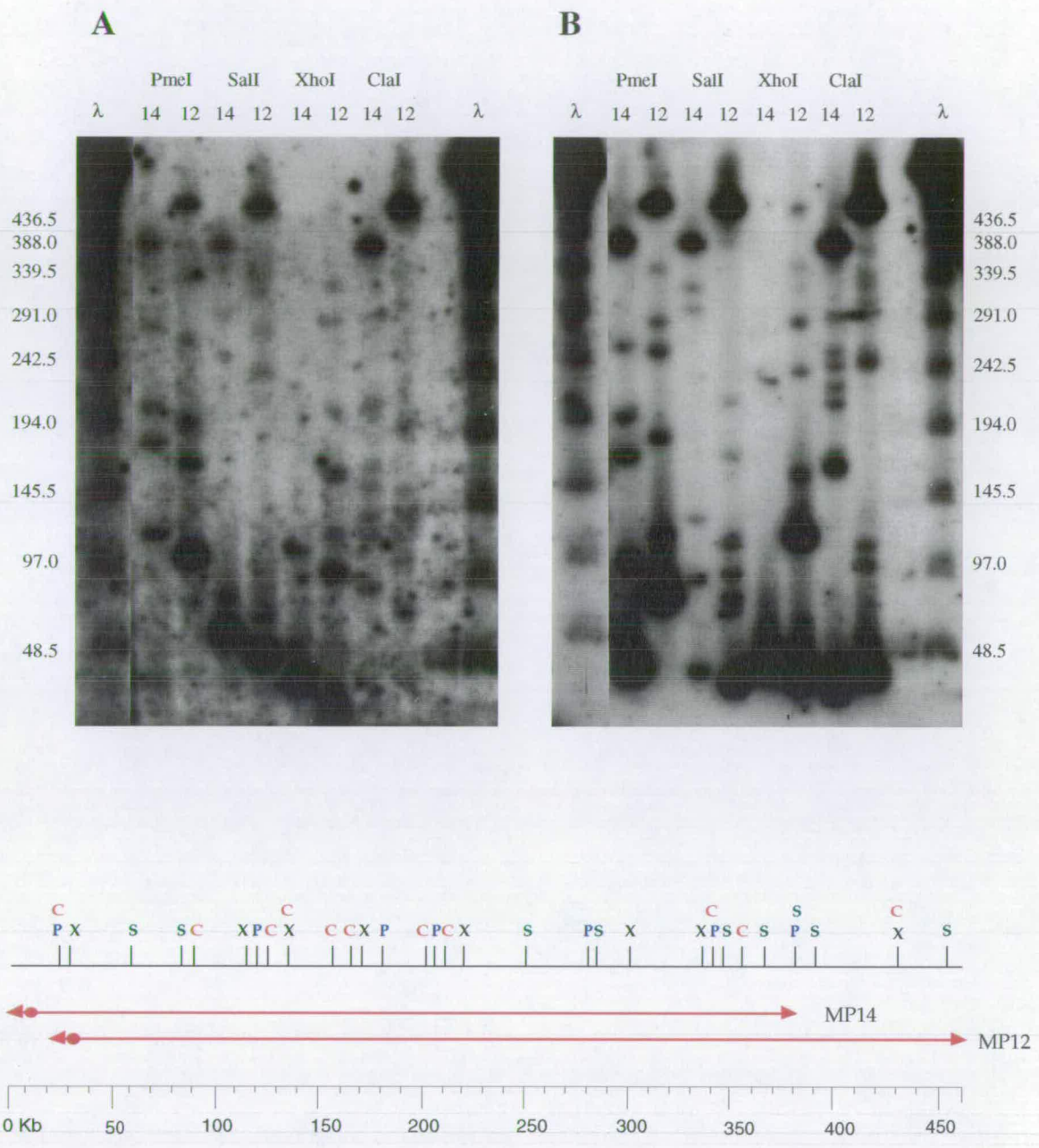
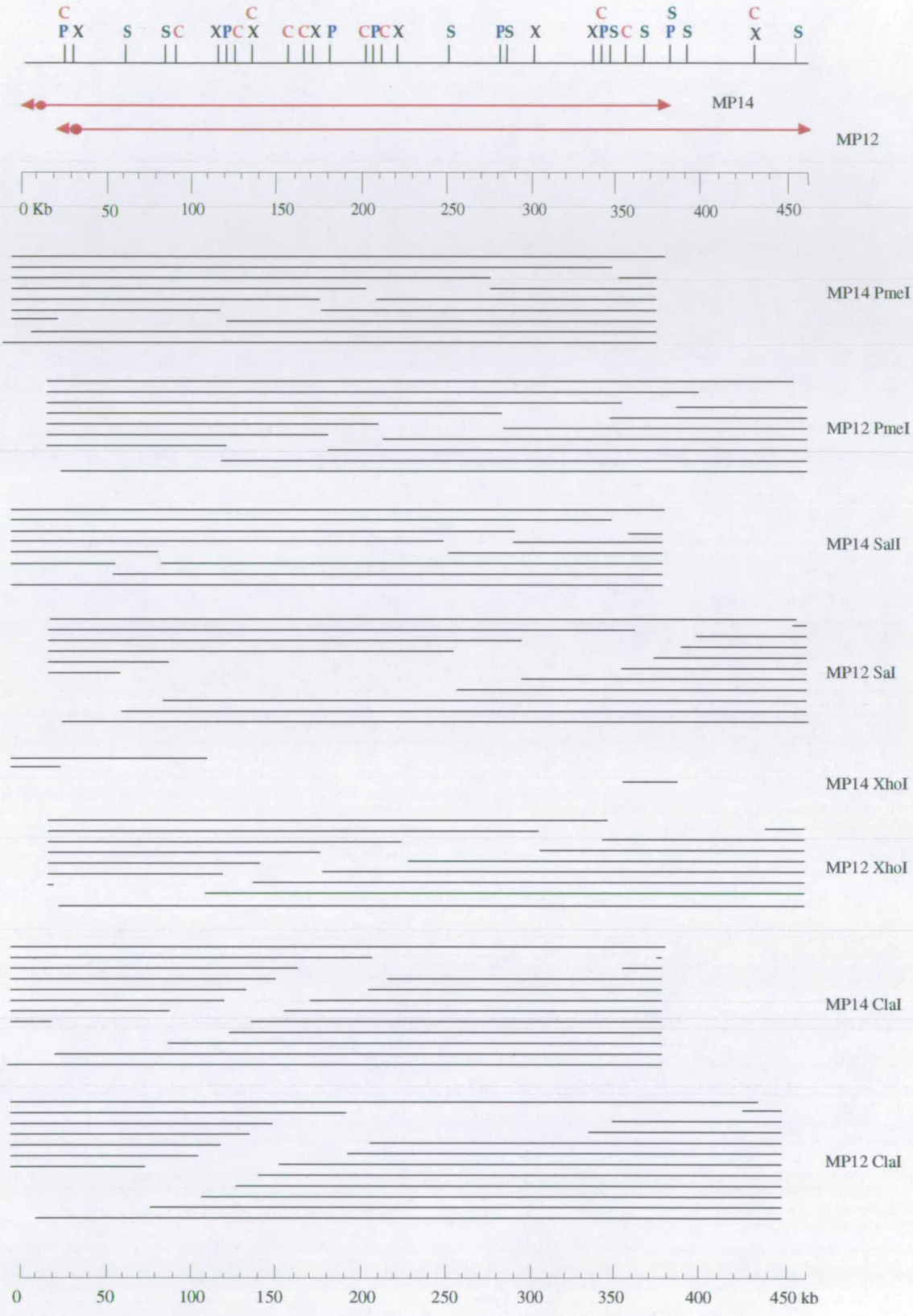


Figure 4.2: Restriction mapping of MP14 and MP12

MP14 (14) and MP12 (12) DNA was partially digested with PmeI (P), SalI (S), XhoI (X) and ClaI (C), separated by PFGE, Southern blotted and hybridised to the *LA* (A) and *RA* (B) YAC arm probes. The same blot was probed with *RA*, stripped and reprobed with *LA*. The sizes of the lambda DNA concatemers (λ) are shown. The derived restriction map of the two YACs is shown below (see also Figure 4.3).

Figure 4.3: Aligning restriction maps of MP14 and MP12



incomplete in this gel due to over-digestion, but the MP12 XhoI digest allowed the map to be generated, with good agreement between the *LA* and *RA1* band patterns; *RA1* detected a ClaI doublet in MP14 and MP12, at 200 kb and 210 kb respectively in Figure 4.2, but only a single diffuse band could be discerned in the *LA* pattern. The *RA1* autoradiograph is clearer than the *LA* autoradiograph, so this doublet is considered to be genuine.

Probe	YAC	PmeI	SalI	XhoI[b]	ClaI
LA	MP14	26[a], 121, 179, 206, 279, 350, 381	59, 85, 251, 293, 350, 381	28[a], 114	23[a], 92, 124, 137, 154, 167, 210, 382
	MP12	105, 164, 193, 265, 336, 381, 437	41[a], 70, 236, 276, 343, 437	3[a], 102, 124, 159, 206, 286, 327	78, 109, 122, 139, 152, 195, 437
RA1	MP14	22[a], 98, 164, 196, 251, 365, 381	20[a], 88, 126, 293, 322, 381	32[a]	3[a], 23, 161, 172, 206, 221, 240, 253, 289, 355, 382
	MP12	77, 116, 179, 249, 281, 343, 437	8[a], 65, 74, 90, 109, 168, 206, 377, 402, 437	23[a], 119, 156, 233, 283, 323, 351, 437	23[a], 100, 114, 242, 255, 295, 307, 320, 340, 370, 437

Table 4.1: Restriction fragment sizes calculated from Figure 4.2.

[a]Fragments denoted as less than 50 kb are subject to significant error (See text).

[b] The XhoI digest of MP14 was excessive, leading to an incomplete partial band pattern.

4.4.1 Organisation of the murine casein genes

The α casein and both β casein probes detected the same pattern of PmeI bands (Figure 4.4), which for MP12 was identical to that obtained with *LA*. In MP14, α and β probes detected PmeI bands identical to the *LA* bands, with the exception of the 26 kb PmeI band. α and β also detected another MP14 PmeI band some 25 kb below each *LA*-identical PmeI band. From these data α and β caseins were mapped to the same 95 kb PmeI fragment near the left arm of both

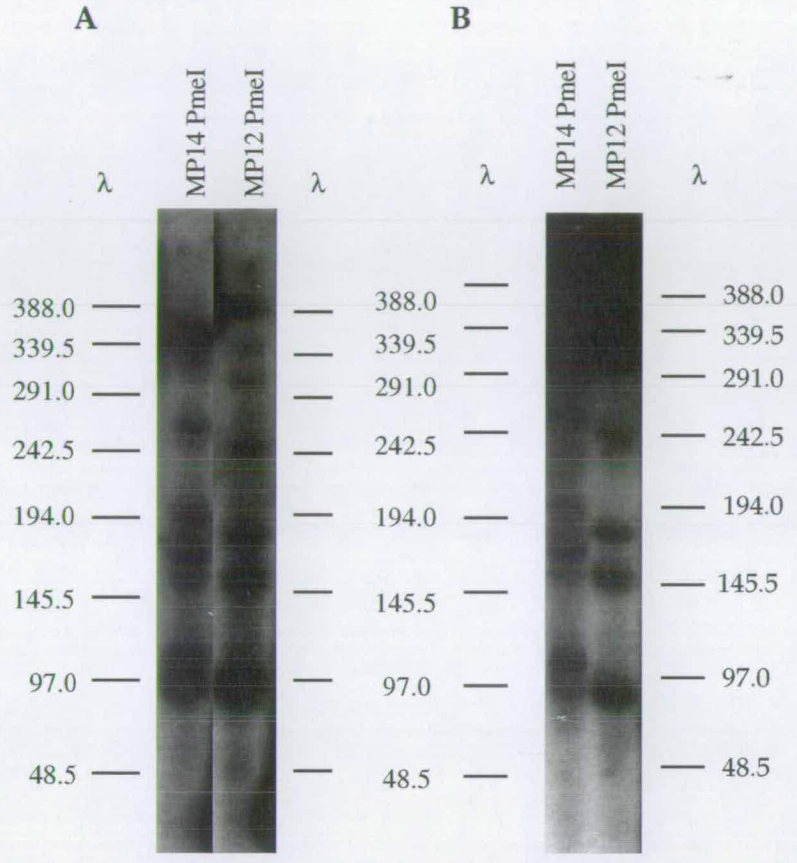


Figure 4.4: Mapping the murine α and β casein genes

Partial PmeI digests of MP14 and MP12 probed for α (panel A) and β (panel B) casein. The sizes in kb of the λ concatemer markers are shown. Note that two separate filters are represented

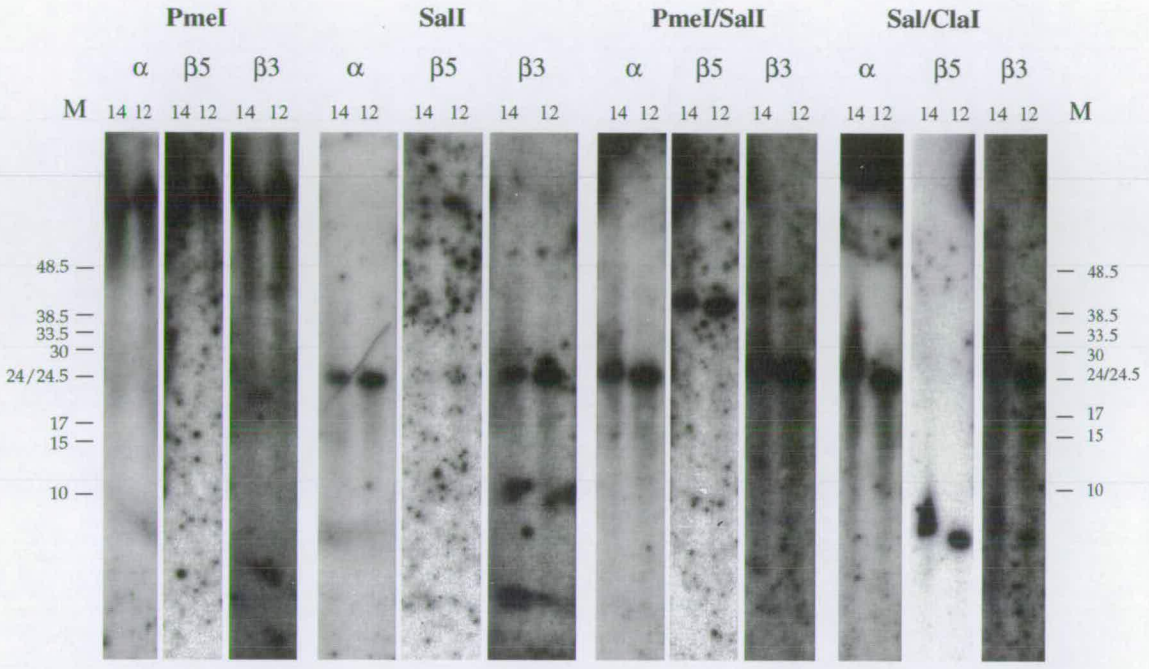


Figure 4.5: resolving the murine α and β casein genes.

The same filter was successively probed with one probe for α casein and two for β casein, denoted α, β5 and β3. Sizes of λ monocut markers (NEB) are shown in kilobases. Restriction enzymes used are displayed above each panel. 14 - MP14 DNA; 12 - MP12 DNA.

YACs, and subsequently to an 85 kb XhoI fragment and a 65 kb ClaI fragment within this PmeI fragment.

SalI proved useful in resolving α from β (Figure 4.5). In complete digests, α gave identical 25 kb SalI, PmeI/SalI, and SalI/ClaI fragments. The two genomic β casein probes, β5 and β3 were used separately. β3 detected the same bands as α, plus a very faint SalI band above 100 kb (the low MW gel used resolved up to 100 kb), a very faint PmeI/SalI band at 34 kb and a very faint SalI/ClaI band of much less than 10 kb. β5 detected this high MW SalI band, the 34 kb PmeI/SalI band, and the small SalI/ClaI band. Taking these results together and referring to the map in Figure 4.2, it appears that β casein lies to the right of α, that the SalI site at 85 kb on the map lies within the β casein gene, and that β is transcribed from right to left in Figure 4.2. In the published β casein genomic sequence (EMBL accession X13484) there is a SalI site some 200 bp from the 5' end of the 1.5 kb 3' probe; the very faint bands detected by β3 were presumably detected by the short portion of the probe that is 5' of this SalI site and map to the right of the SalI site at 85 kb, while the strong bands map to the left of this site. This suggests that the gene is transcribed from right to left in the map. This inference might not be valid since probe labelling intensity can vary sequence dependently and signal strength may therefore not reflect the proportion of the probe that is bound. However, MβC5' maps to the right of the SalI site, confirming the orientation of the gene.

From the pattern of partial PmeI bands detected by γ in each YAC (Figure 4.6), it was possible to locate the gene on one of two PmeI fragments, either that between the sites at 120 and 180 kb, or that between those at 205 and 280 kb (Figure 4.2). The data favour the former, but are not conclusive. However, the latter PmeI fragment contains a SalI site, so if γ was used to probe complete PmeI and PmeI/SalI digests, the PmeI/SalI band would be expected to be smaller than the PmeI band. As can be seen in Figure 4.7, PmeI/SalI and PmeI bands detected by γ were in fact identical at 65 kb, suggesting that there is no SalI site within the PmeI fragment containing γ and therefore that γ cannot be located between the PmeI sites at 205 and 280 kb. A similar experiment with XhoI/PmeI double digests suggested that γ is located on a 38 kb XhoI fragment within the PmeI fragment, again ruling out the latter PmeI fragment which contains no such XhoI fragment (Figures 4.6 and 4.7). XhoI partial digests suggest that γ is located between the XhoI sites at 135 and 170 kb in Figure 4.2. In complete digests, γ detected ClaI and ClaI/SalI

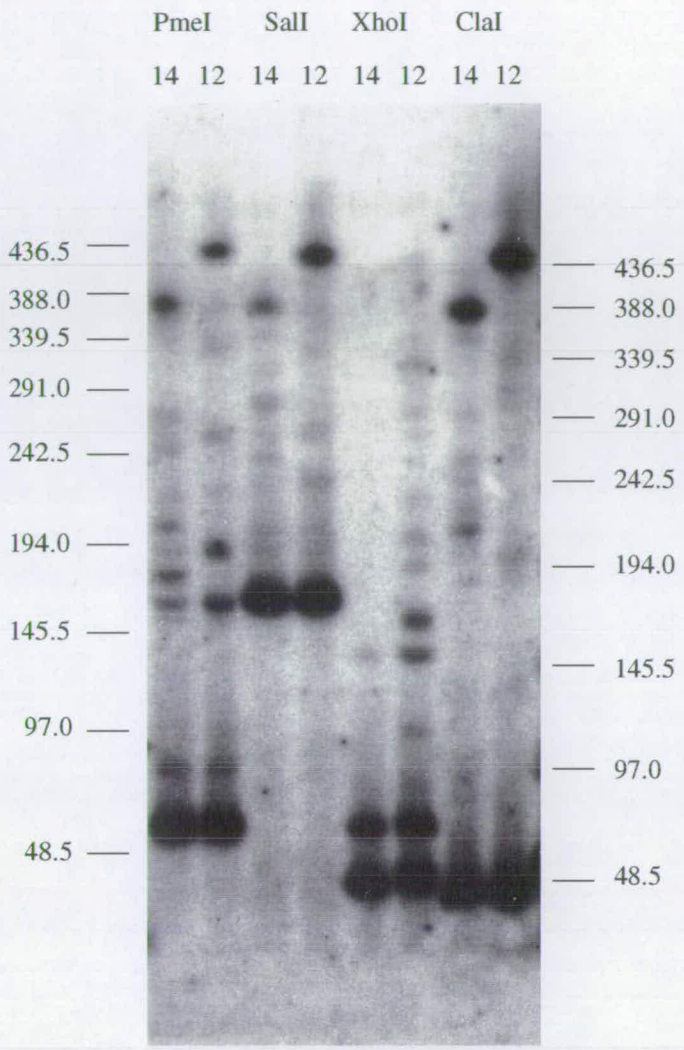


Figure 4.6: Mapping themurine γ casein gene

Partial restriction digests probed for γ casein. 14: MP14; 12: MP12. The sizes of λ concatemer markers (NEB) are shown in kb.

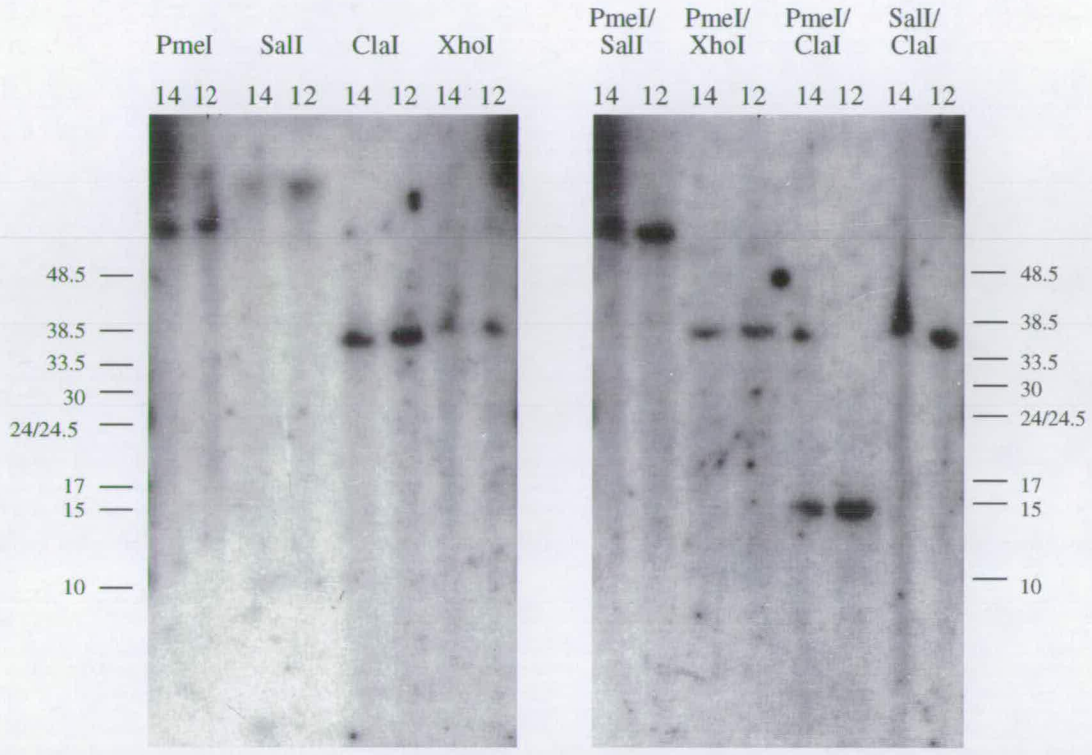


Figure 4.7: Mapping murine γ casein

Complete digests were probed for γ casein. The sizes of λ Monocut markers (NEB) are shown in kb. 14: MP14; 12: MP12. Single digests are on the left of the figure, double digests on the right.

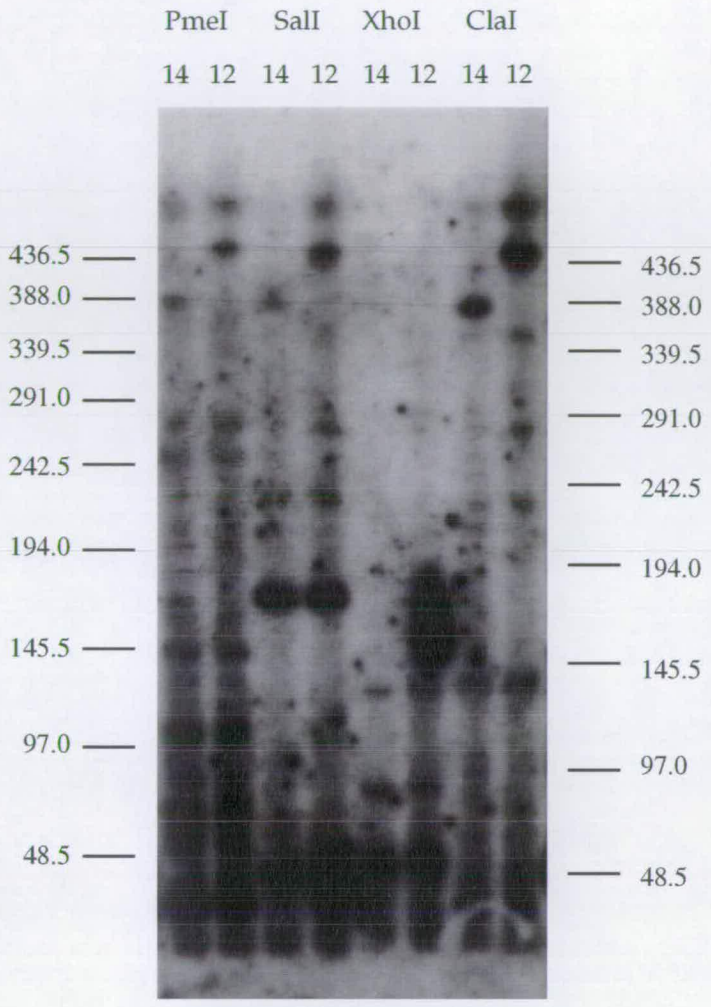


Figure 4.8: Mapping the murine ε casein gene

Partial restriction digests probed for ε casein. 14: MP14; 12: MP12. The sizes of λ concatemer markers are shown in kb. Little can be inferred from these data except that the gene is near the centre of the YAC, hence the profusion of bands.

Figure 4.9: Complete digests probed for ε casein.

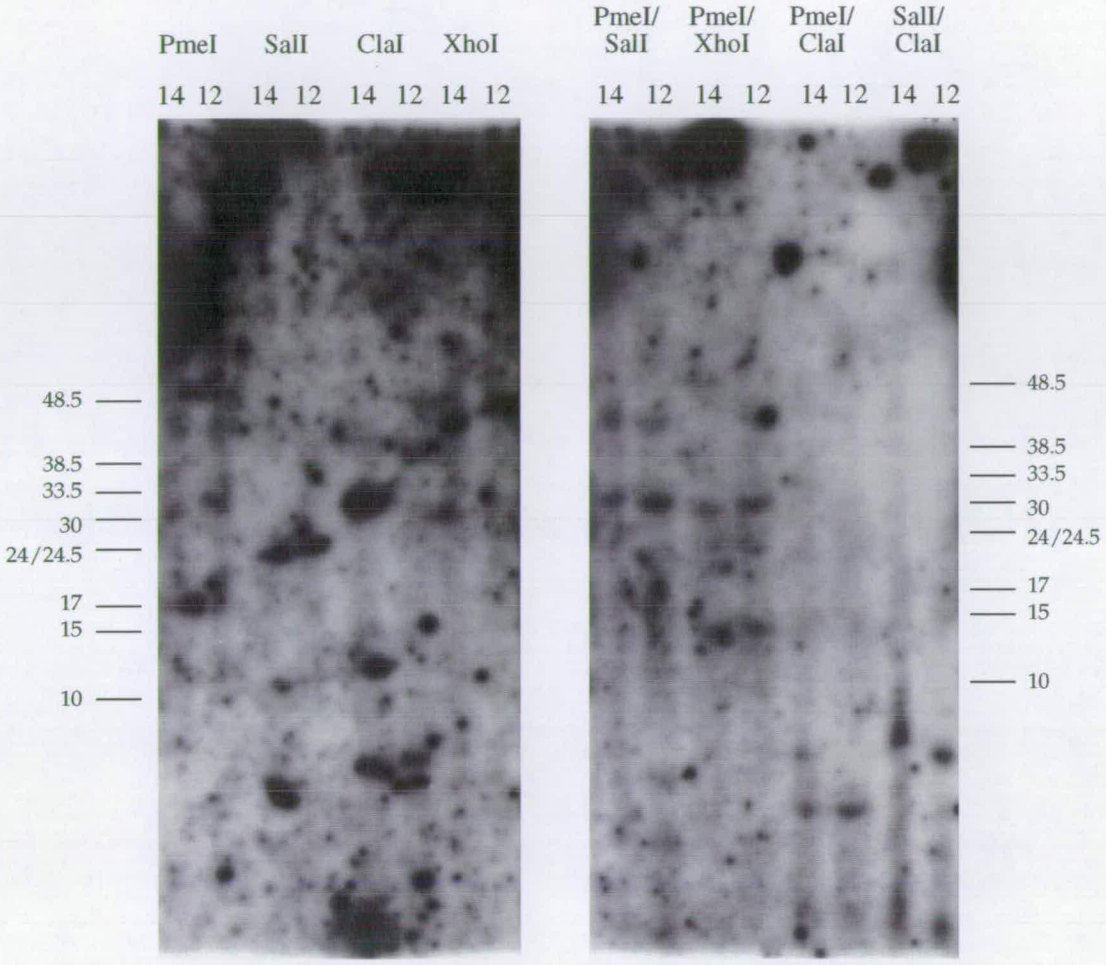


Figure 4.9: Complete digests probed for ε casein.

The sizes of λ Monocut markers (NEB) are shown in kb. 14: MP14; 12: MP12. Single digests are on the left of the figure, double digests on the right.

Figure 4.10: Partial restriction digests probed for κ casein.

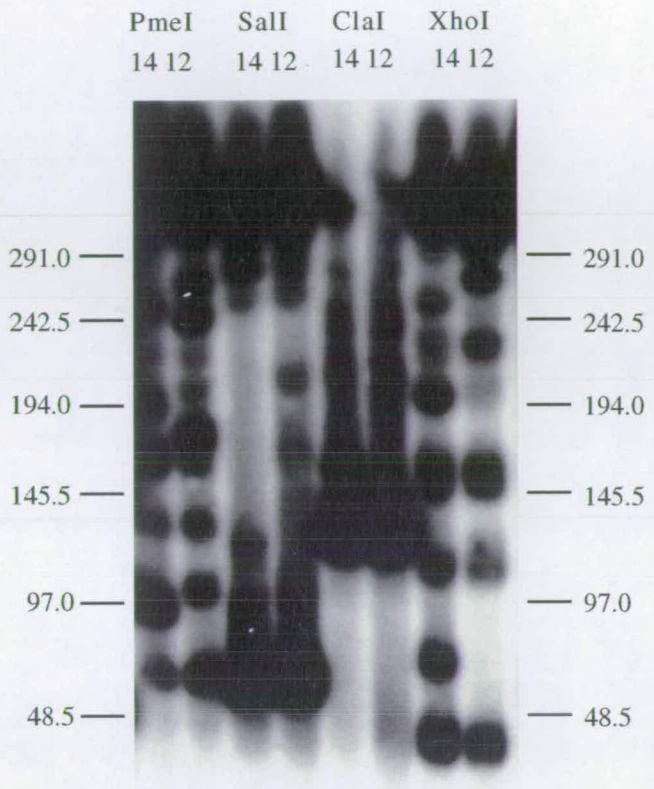


Figure 4.10: Partial restriction digests probed for κ casein.

Sizes of λ concateme (NEB are shown in kb. 14: MP14; 12: MP12.

Figure 4.11: Complete restriction digests probed for κ casein.

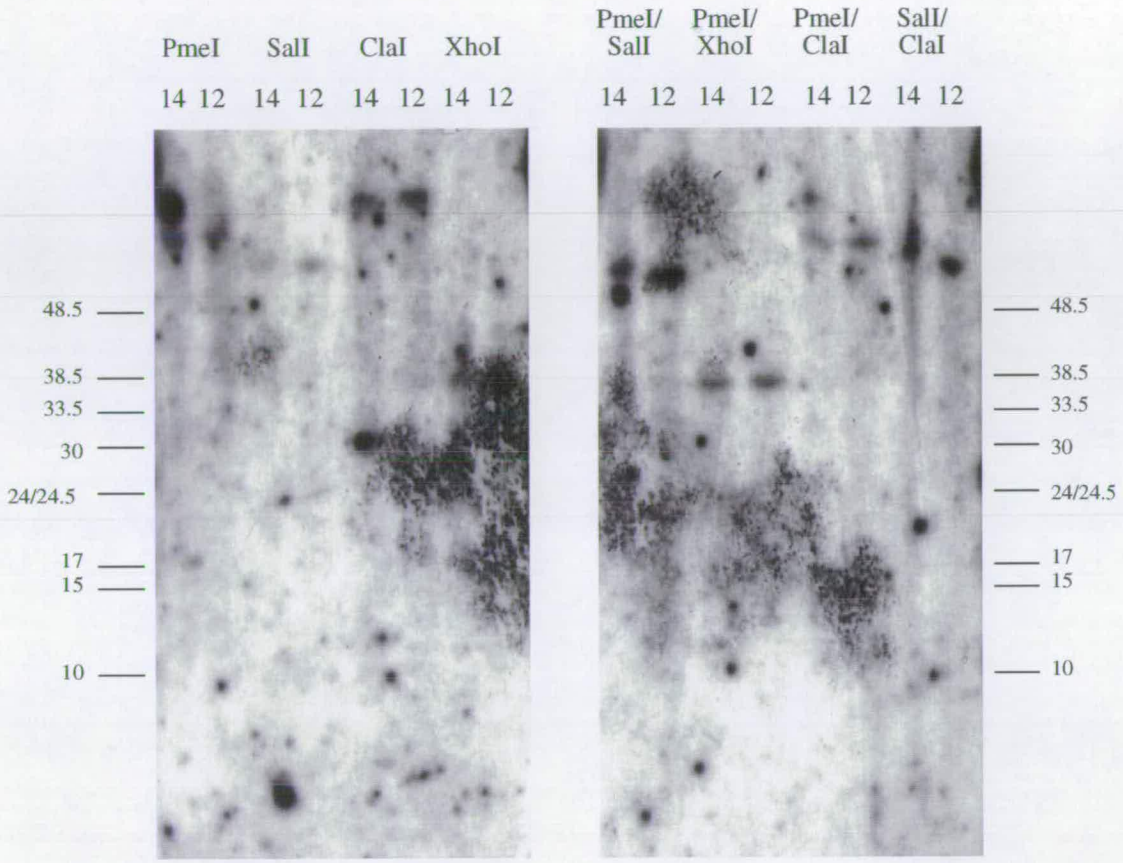


Figure 4.11: Complete restriction digests probed for κ casein.

Sizes of λ Monocut markers (NEB) are shown in kb. 14: MP14; 12: MP12. Single digests are on the left of the figure, double digests on the right.

bands of 35 kb, and a PmeI/ClaI band of 15 kb (Figure 4.7). This places the gene between the PmeI site at 180 kb and the ClaI site at 165 kb (Figure 4.2). The implied position of γ on a very small ClaI/XhoI fragment has not been confirmed.

The partial restriction fragment band pattern detected by the εcasein probe is too dense to be resolved by the gel conditions used (figure 4.8). However, the ε probe detected a 20 kb partial PmeI band in both YACs, suggesting that it is internal to both (Figure 4.8). The only PmeI fragment of that size within the map lies between the sites at 180 and 205 kb (Figure 4.2). The partial SalI pattern starting at 160 kb also suggests that ε lies near the centre of the map. The XhoI pattern is consistent with this position. However, there were bands in the PmeI profile that suggested that ε casein may span the site at 205 kb, in particular a band of 70 kb which may represent the fragment between the PmeI sites at 205 and 280 kb (Figure 4.2). In complete digests (Figure 4.9), ε detected identical PmeI and PmeI/SalI bands of 27 kb (though the PmeI data are poor), plus a PmeI/SalI band of 40 kb and PmeI/XhoI bands of 13 and 27 kb. ε also detected identical ClaI and ClaI/SalI bands of less than 10 kb, a PmeI/ClaI band of much less than 10 kb (smaller than the SalI/ClaI band) and an even smaller SalI/ClaI band (very close to the bottom of the gel); the lower limit of resolution on this gel was 10 kb. This very small ClaI/SalI band is difficult to account for unless there is another ClaI site very close to the site very close to the two at 200 and 210 kb in Figure 4.2. A double site of this nature would not be detected by partial digest. Nevertheless, these data suggest that ε lies around the small (less than 10 kb) ClaI fragment between the sites at 200 and 210 kb in Figure 4.2, and contains the PmeI site at 205 kb. Unfortunately the blot of the complete digests was rather poor (Figure 4.9) and the expected second PmeI band was not seen. Moreover, the ε gene might be expected to exceed 10 kb (as do the characterised casein genes) and therefore seems unlikely to reside entirely within the 10 kb ClaI fragment, though to date this has not been tested. The putative extra ClaI site has also not been tested for.

κ casein was located using PmeI and confirmed with XhoI and ClaI. The relationship between the κ pattern of PmeI bands and the *RA1* pattern was similar in MP14 to that between the α/β and *LA* patterns (Figure 4.10), suggesting that κ maps between the PmeI sites at 280 and 340 kb in Figure 4.2. In complete digests, κ detected identical XhoI and PmeI/XhoI bands of 34 kb, a SalI band of 60 kb and a 55 kb PmeI/SalI band, again consistent with this location (Figure 4.11).

Figure 4.12: Comparison of reported casein locus maps

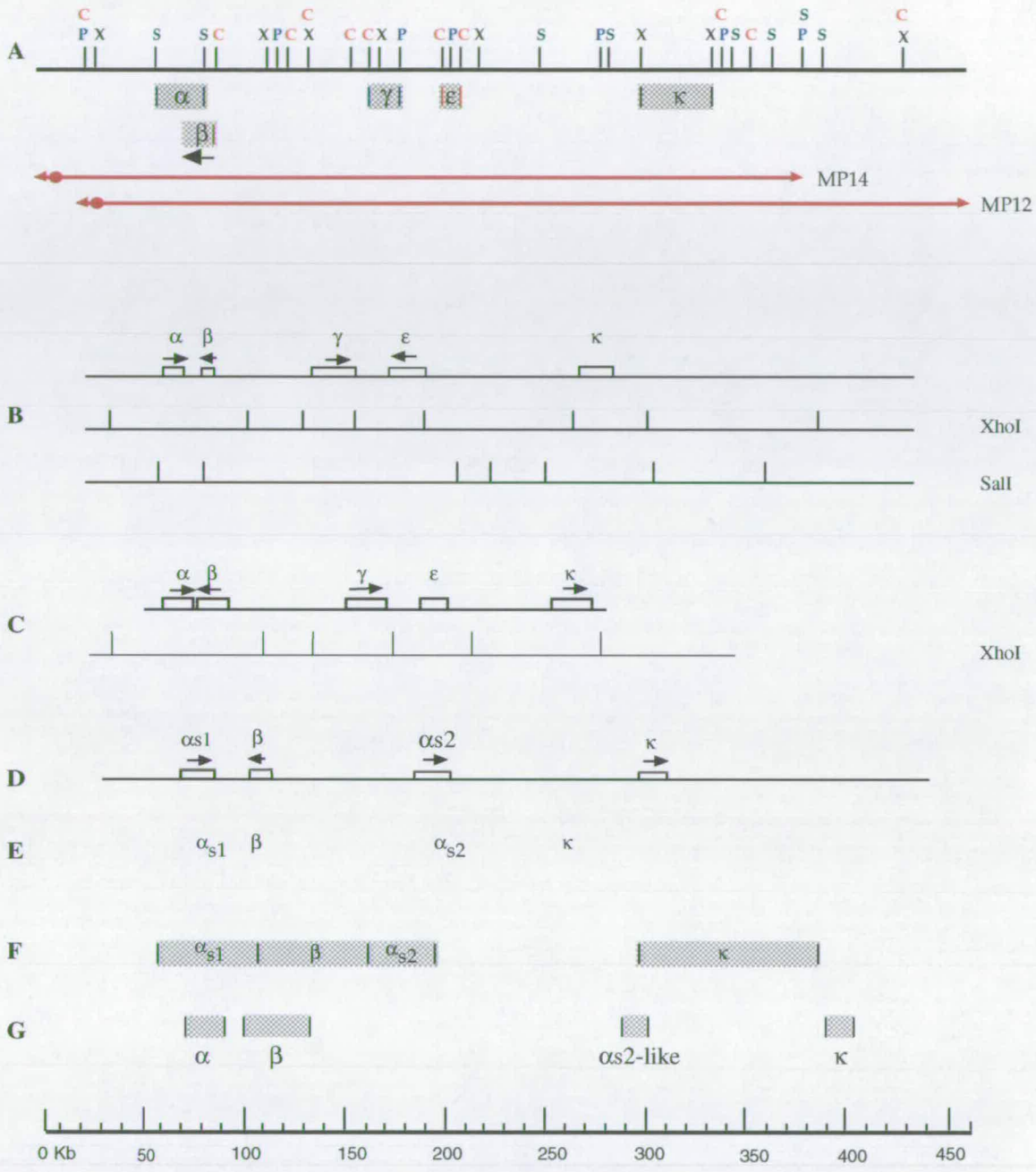


Figure 4.12: Comparison of reported casein locus maps

Seven casein locus maps are compared above. A: this thesis; B: murine, Rijnkels *et al*, 1997b; C: murine, George *et al*, 1997; D: bovine, Rijnkels *et al*, 1997a; E: bovine, Threadgill and Womack, 1990; F: bovine, Ferretti *et al*, 1990; G: human, Rijnkels *et al*, 1997c. The maps are aligned at the α–β gene pair; the mouse maps are specifically aligned at the SalI site within the β casein gene. The three maps published by Rijnkels and associates are aligned with each other as in Rijnkels *et al*, 1997c. S: SalI; P: PmeI; X: XhoI; C: ClaI. The genes are represented by Greek letters labelling either boxes or shaded minimal restriction fragments, depending on how they have been reported.

4.5 Discussion

In this chapter two YACs bearing the murine casein locus were mapped by a combination of partial and complete restriction digest. The two YACs overlap by approximately 350 kb, and are essentially collinear within the region of overlap. This precludes the possibility of either being substantially rearranged relative to the genome. The casein genes were also localised on this restriction map by combined partial and complete restriction digest. In maps general partial digests were only informative when the gene was close to one end of the YAC; internal probes simply hybridised to too many bands for unequivocal interpretation. The completed map, with the positions of the murine casein genes, is shown in Figure 4.12a. Beneath the map are representations of two other casein locus maps reported in the literature (Rijnkels *et al.* 1997a; George *et al.* 1997). A comparison of the three published maps (Figure 4.12) indicates broad agreement, but with some notable discrepancies.

Rijnkels *et al* (1997a) restriction mapped ICRFy902G0781, which corresponds to MP12 (Figure 4.12 B). Two enzymes, XhoI and SalI, are shared between the work reported here and that reported by Rijnkels *et al* (1997a). The XhoI digests agree, although in general Rijnkels *et al* found the locus to be somewhat smaller than was found in this project. Rijnkels *et al* report MP12 (mYAC9) to cover 460 kb, whereas here it was found to be approximately 430-440 kb in length (see for example, Figure 4.2). The resolution of the gels used here was not high at that Range however, and an error of up to 20-30 kb would not be surprising. The slightly smaller size of the murine locus is in agreement with the other reported map (George *et al.* 1997). Allowing for this discrepancy, the XhoI maps reported here and in Rijnkels *et al* (1997a) are in agreement. In contrast, George *et al* do not report the presence of a XhoI site close to the 5' end of the k casein gene. This site corresponds to that at approximately 300 kb in Figure 4.12A, and at approximately 270 kb in Figure 4.12B. This site was assigned based on the 156 kb XhoI fragment detected in MP12 DNA by *RA1* (Figure 4.2; Table 4.1). It is difficult to discount this band as an artefact. As noted in section 4.x, the *RA1* probe contains some sequence homologous to the YAC left arm, and so the possibility must be considered that some bands detected by *RA1* might in fact be spurious, belonging to the *LA* pattern instead. A 159 kb band was detected in the *LA* pattern. The two bands do not directly overlay each other; nevertheless, on this basis it is impossible to rule out the possibility that this band is indeed spurious. However, Rijnkels *et al*

(1997a) also report the existence of this site. Since both reports describe the same YAC clone there might be some possibility that this clone contains an extra site, but Rijnkels *et al* (1997a) also mapped genomic DNA. Although XhoI digests of genomic DNA were not performed, the YAC did not differ from genomic DNA in any digest that was performed on both. In addition, the XhoI and PmeI/XhoI bands detected by the k probe in MP12 and MP14 DNA in complete digests (Figure 4.11) are consistent with the existence of this XhoI site, and are inconsistent with its absence given the independent evidence for the location of the k casein gene (Figures 4.10, 4.11). Taken together, these results argue strongly for the presence of this XhoI site immediately 5' to the κ casein gene. Further work would utilise an uncontaminated right arm probe (such as the 1.7 kb BamHI-PvuII fragment of pBR322 – see chapter 6) to test the validity of the XhoI data presented here.

In Figure 4.12A four SalI sites are indicated 3' to the κ casein gene. However, Rijnkels *et al* (1997a) report only two SalI sites in this region. The four SalI sites in Figure 4.12A are indicated by the four SalI bands detected in MP12 DNA by *RA1* (Figure 4.2) at 65, 74, 90 and 109 kb (Table 4.1). Since a SalI band of 70 kb was detected by *LA* in MP12 DNA one of the *RA1* bands could be caused by the impure nature of the probe. In MP14 DNA digested with SalI, *RA1* detected a single band of an appropriate size (Figure 4.2), which argues against the existence of the SalI site at 365 kb in Figure 4.12A. Further work is required to resolve this discrepancy.

The bovine genomic casein region has been restriction mapped (Ferretti *et al.* 1990; Threadgill & Womack, 1990; Rijnkels *et al.* 1997d; Figure 4.12). The spatial arrangement of the genes is largely conserved between the two species, allowing for the presence of the second α_{s2}-like casein gene in mice. The bovine α_{s1}, β and κ casein genes are homologous to the murine α, β and κ genes, while bovine α_{s2} shows sequence homology with the murine ε and γ casein genes. If the presence of this extra gene in mice represents a deletion or duplication/insertion event within the locus, one might expect some difference in the spatial positions of the genes. Given the variation between the published bovine maps it is difficult to posit a significant difference between the two species. However, Rijnkels and associates have mapped both loci (Rijnkels *et al.* 1997a, d) and show the murine locus to be approximately 20 kb larger than the bovine locus.

The human casein gene locus has also recently been characterised (Rijnkels *et al.* 1997b). Human milk has been reported to contain three caseins, α, βand κ (Cavaletto *et al.* 1994; Rasmussen *et al.* 1995) and nucleic acid sequence is available for the human β (Hansson *et al.* 1993 unpublished)and κ (Edlund *et al.* 1996) casein genes and for the human αcasein cDNA (Johnsen *et al.* 1995). Rijnkels *et al* (1997b) found evidence for the existence of a fourth human casein gene by hybridisation to cDNA probes of bovine α_{s2} casein. This gene maps to a point approximately 80 kb from the κcasein gene, which is in keeping with its position in other species. However, there is approximately 100 kb of extra intergenic DNA between the β and α_{s2} genes relative to the other species analysed, possibly as a result of an insertion event (Rijnkels *et al.* 1997b).

It is interesting to note the spatial arrangement of the murine casein genes (Figure 4.12). α and β are both present on a 25 kb SalI fragment, although β contains the right hand SalI site, with approximately 2.5 kb of the published sequence falling within the 25 kb fragment on the map (Figure 4.12). The sensitivity of the low molecular weight blot has been demonstrated by the data obtained with β3', and it therefore seems likely that at least the majority of the α casein gene lies within the 25 kb SalI fragment. There are no data concerning the murine α casein gene structure, but the bovine αs1 gene covers 17.5 kb (Koczan *et al.* 1991) which suggests that there is no more than 5 - 10 kb between the mouse a and b genes. Leroux and Martin (1996) performed long-range PCR on goat DNA using primers from within the 3' ends of the αand β casein genes and isolated a 13.4 kb product. They estimate the distance between the α and β casein genes as 12 kb. By a similar method George *et al.* (1997) estimated the murine αand β casein genes to be 8 kb apart, which is in keeping with the results presented here.

Similarly, the murine ε and γcasein genes map to the same 45 kb ClaI fragment in Figure 4.12. There is no reported genomic sequence for murine γ casein, but the rat γ casein exon/intron structure has been mapped by restriction and R-loop analyses (Yu-Lee & Rosen, 1983) from which the primary transcript is estimated at 15 kb. Similarly, although there is no murine ε casein sequence, the bovine homologue α_{s2} has been sequenced (Groenen *et al.* 1993) and spans 18.5 kb from transcription start to polyadenylation signal. Thus these two genes are also likely to be very close to one another. The calcium sensitive caseins are proposed to have evolved from

a common ancestor by gene duplication followed by rapid sequence divergence and exon shuffling (Bonsing and Mackinlay 1987; Groenen *et al.* 1993). Sequence analysis supports this (Groenen *et al.* 1993; Rijnkels *et al.* 1997a,d). The calcium sensitive caseins are unusual in having few constraints on the primary structure of the protein (Mercier and Vilotte, 1993), and this is reflected in their rapid sequence divergence.

In contrast to the paired calcium sensitive caseins, the murine κ casein gene is situated 95 - 155 kb from ε, the nearest neighbouring casein gene. κ casein is believed to be evolutionarily unrelated to the other caseins, being evolutionarily and functionally related to γ fibrinogen (Alexander *et al.* 1988). The physical distance between the κ gene and the others has been said to reflect this phylogenetic distance (Ferretti *et al.* 1990), but it must be noted that the distance between the murine α/β pair and γ/ε pair in Figure 4.12 is some 75 kb, which is roughly comparable to the 95 - 130 kb separating κ from ε. Moreover, the distance between the four calcium sensitive casein genes and the κ gene is remarkably small given that they are evolutionarily unrelated. Interestingly, the γ fibrinogen gene and the casein genes are located on the same chromosome in pigs (SSC8) and humans (HSA4) but not in mice. In pigs the γ fibrinogen gene is approximately 5 cM away from the caseins, while in humans the distance is approximately 40 cM. In mice the caseins are on chromosome 5, while γ fibrinogen is on chromosome 3. It would appear, then, that κ casein and γ fibrinogen have arisen from a duplication of an ancestral gene. Subsequently the two genes became separated in some taxa, with one becoming recruited into the casein gene family. The subsequent co-segregation of κ with the other casein genes lends further support to the idea that their coordinate expression is controlled at the chromatin level.

The ratio of κ casein to the calcium sensitive caseins in milk is believed to play a significant role in determining micelle size, milk volume and milk quality (Sullivan *et al.* 1959; Dalgleish *et al.* 1989; Lin *et al.* 1989). Naturally occurring bovine genetic variants display a correlation between increased κ casein levels and decreased casein micelle diameter (reviewed in Gutiérrez-Adán *et al.* 1996) Transgenic mice expressing a bovine κ casein cDNA under the control of a caprine β casein promoter and 3' flank (Gutiérrez *et al.* 1996) showed a marked decrease in micelle size which was correlated with dosage of the transgenic locus (Gutiérrez-Adán *et al.* 1996). This is consistent with the prevailing view of casein micelle structure, where the micelle interior

consists of calcium sensitive caseins and the surface is made up of κ casein and calcium. An increase in κ casein concentration relative to that of the calcium sensitive caseins might therefore be expected to result in micelles with an increased ratio of surface area to volume, and therefore a decreased diameter.

As has been suggested (Alexander *et al.* 1988; Mercier & Vilotte, 1993; Rijnkels *et al.* 1997a) the presence of the κ casein gene close to the other casein genes may indicate that coordination of casein gene expression is effected at the chromatin level, and thus may be indicative of the existence of casein locus control. This view is strengthened by the finding that the structure of the locus as a whole is conserved between distant taxa, despite the variation in the number of calcium sensitive casein genes. It must be noted that the pattern of casein gene expression is not invariant between these taxa. For example, bovine milk contains a far greater proportion of α caseins than does human milk (Kunz and Lönnerdal, 1989). In addition, the calcium sensitive casein genes are highly divergent between taxa. However, the exon structure and the sequence of functional sequence motifs (especially the signal peptide) is highly conserved. In addition, since the caseins serve merely to raise the protein and mineral concentration of milk whilst maintaining a low viscosity through the colloid of micelles, the structure and even the number of the calcium sensitive caseins is open to considerable neutral variation. In contrast, the ratio of κ casein to the calcium sensitive caseins is critical for maintaining the colloid, and the cleavage of κ casein in the gut by rennin is critical for effective digestion of the milk proteins by the suckling infant. Therefore, while evolutionary divergence among the calcium sensitive casein genes would not be remarkable, the κcasein gene ought to be under considerable selective pressure. This appears to be the case (Ward *et al.* 1997).

In summary, the murine casein gene locus has been mapped in this project and by two other groups (Rijnkels *et al.* 1997a; George *et al.* 1997). The three maps are in broad agreement over the structure of the locus. The murine casein genes are arranged α–β–γ–ε–κ, and the α–β pair and the γ–ε pair are convergently transcribed. The close proximity of the structurally unrelated κ casein gene to the other members of the family is conserved in mammals. The κ casein gene appears to have arisen as a result of a duplication of the ancestral γ fibrinogen gene. The subsequent conservation of the structure of the casein locus may be related to the importance of coordinate casein expression in determining the characteristics of the milk.

5. Chapter 5 - Mouse YAC manipulations

5.1 Introduction

Chapters 3 and 4 descibe the isolation and mapping of two YACs bearing the murine casein locus, MP12 and MP14. Both carry all five murine genes and appear to be unrearranged with respect to the mouse genome. The principal aim of this project was to use one of these YACs to investigate the expression of the murine casein genes after transfer into mice or cultured cells, with a view to assaying for copy number dependent variation in gene expression between lines of animals or cells. MP14 was chosen over MP12 because it extends farther to the left of the casein locus.

The decision to use a mouse YAC was based on the desire to carry out as "clean" an experiment as possible. It was hoped that the use of murine DNA in a transgenic mouse would maximise the probability of detecting any transcriptional control functions, however subtle. However, detection of transgenic murine caseins in the lactating mouse mammary gland would be essentially impossible, and at this point casein knockout mice (Kumar *et al.* 1994) were unavailable. The YAC was therefore modified to carry a reporter gene. The protein C minigene construct pCorp3 (Figure 5.1) has been used extensively at PPL Therapeutics Ltd. It consists of a protein C cDNA inserted into an ovine β lactoglobulin (BLG) minigene containing the BLG transcription start sequence, part of exon 6 and all of intron VI and exon 7. This construct has given generally low levels of protein C expression in transgenic mouse milk, of the order of 10 μg/ml. Rescue of this transgene with a BLG genomic construct (Clark *et al.* 1992) raised these levels slightly (Table 5.1). There appears to be no correlation between transgene copy number and expression level. pCorp3 was chosen as a reporter for several reasons. Firstly, the wealth of previous data would provide a clear backdrop to this experiment so that any change in the pattern of pCorp3 expression would be apparent. Secondly, pCorp3 is generally expressed at low but detectable levels. This would facilitate detection of any improvement in expression (an important practical objective) while still allowing variation in expression between lines to be monitored. Thirdly, expression of pCorp3 appears to be position-dependent and thus sensitive to the influence of surrounding chromatin, which is important in an assay for functional chromatin elements. Fourthly, an ELISA assay for protein C in mouse milk has been developed and

validated at PPL Therapeutics by J. Percy (unpublished). Lastly, at the start of this project a parallel experiment was underway at PPL Therapeutics which aimed to target pCorp3 to the mouse β casein locus via ES cells (A. Kind, unpublished). The mouse β casein locus is highly transcribed during lactation; 50 % of mRNA in the mouse mammary gland is β casein mRNA (Mercier and Vilotte 1993). This locus therefore represents an attractive target for transgenesis. Replacement of the β casein coding region with that for an exogenous gene might cause the transgene to be expressed at levels close to those of β casein itself. However, the removal of the β casein gene from this locus might alternatively reduce the transcriptional activity of the locus if elements within the βcasein gene are required for full expression, or if the active transcriptional status of the β casein gene has an overall enhancing effect on the transcriptional activity of the locus as a whole (see chapter 1). Therefore two targeting constructs were planned (Figure 5.1). CPC-A is designed to target pCorp3 to within the β casein gene, replacing the β casein gene from transcriptional start to 300 bp 3' to the polyadenylation (polyA) site. CPC-D is designed to insert pCorp3 into a BclI site 300 bp 3' to the polyA of β casein. Both of these constructs were made and kindly donated by J. Howcroft. The success of this targeting experiment would also provide an excellent positive control for the YAC transgenic study. Expression of the same transgene in the same context via the two routes of gene targeting and YAC-mediated transgenesis would convincingly demonstrate whether the YAC contained all of the relevant elements present in the genome. For this reason, the same CPC-A and CPC-D targeting constructs were used to conduct modification of the mouse germline (A. Kind, unpublished) and the murine casein YAC MP14.

Construct	Copy number	Expression level of protein C in milk (μg/ml)	
		Total	Expression per copy
pCorp3 rescue	3	14.1	4.7
pCorp3 rescue	3	17.2	8.6
pCorp3 rescue	2	16.8	8.4
pCorp3 rescue	10	9.5	0.95
pCorp3 rescue	10	57	5.7

Table 5.1: pCorp3 expression data from previous work

Data reproduced with permission from PPL Therapeutics Ltd."Rescue" experiments involved co-injection of pCorp3 with gnomic ovine BLG in a 2:1 molar ratio. Copy numbers were estimated by Southern blot. Protein C levels were estimated by ELISA (J. Percy, unpublished).

In order to insert pCorp3 into MP14, a selection system was required. Since the aim was to replicate the murine gene targeting work in MP14 the permanent insertion of a yeast selectable marker was undesirable and a counterselectable marker was therefore sought. The yeast counterselectable marker of choice is *URA3,* which is present on the right arm of pYAC4-derived YACs. *URA3* can be selected for on growth medium lacking uracil (-ura) or against by inclusion of 5-fluoro-orotic acid (5-FOA; Boeke *et al.* 1984) in the growth medium. The first step in modifying MP14 was therefore to remove *URA3* from the right arm so that it could then be used in inserting pCorp3 within the β casein locus. Following this *URA3* could be reintroduced as part of a two-step gene replacement ending with the insertion of pCorp3. An alternative to two-step gene replacement in yeast is the so-called pop in-pop out (PIPO) strategy (Scherer and Davis 1979). This involves a single targeting construct containing the counterselectable marker flanked by direct repeats of the sequence to be modified. Yeast cells are transformed with this construct under positive selection (selection for the marker). The marker is then selected against, and cells in which the marker has been deleted by recombination between the flanking direct repeats are recovered. In this project both two-step and PIPO

Figure 5.1: Constructs used in MP14 manipulation

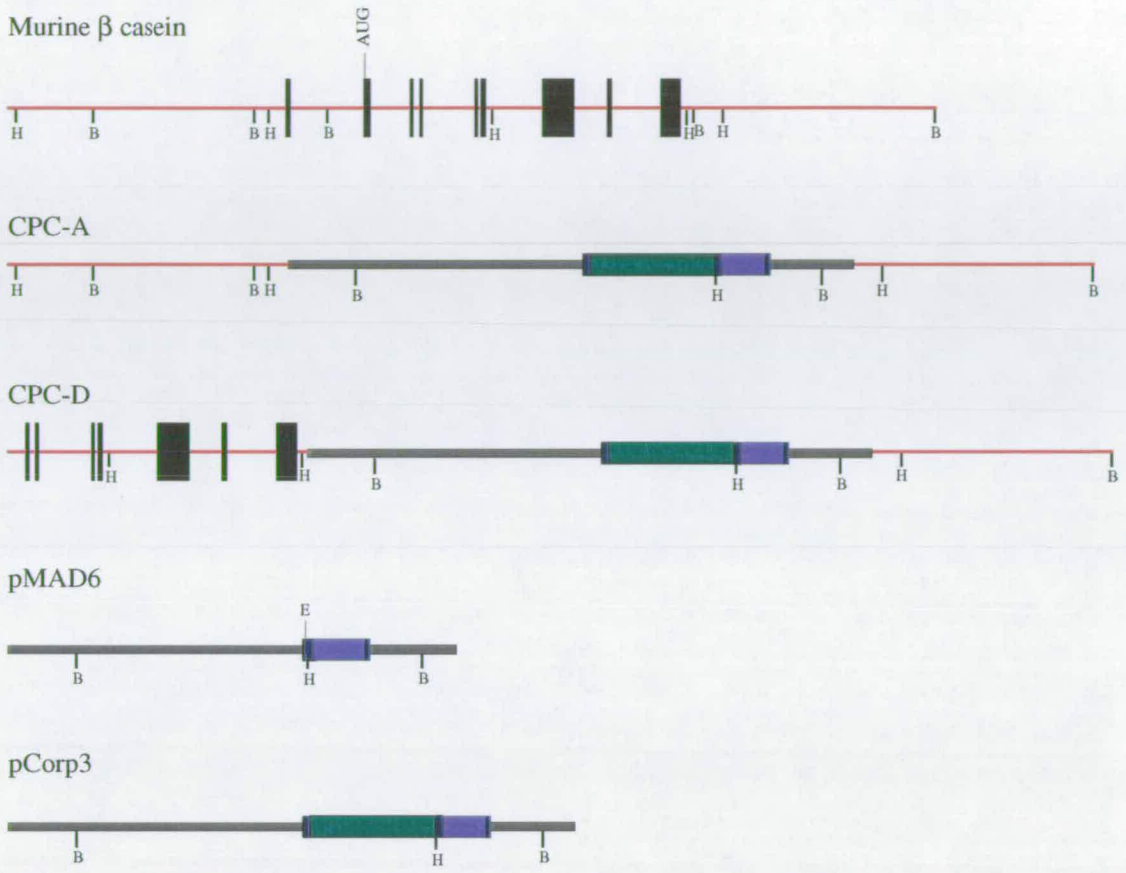


Figure 5.1: Constructs used in MP14 manipulation

The murine β casein locus is shown in red except for exons which are shown as black boxes. The protein C cDNA is shown in green. BLG exons are in dark blue; part of exon 1, part of exon 6 and all of exon 7 are present. BLG intron VI is pale blue. BLG non-transcribed sequence is grey. Restriction sites are provided for orientation. B: BclI; H: HindIII; E: EcoRV. CPC-A and CPC-D were constructed by J. Howcroft. CPC-D targets pCorp3 to a BclI site 300 bp 3' to the polyadenylation signal of the β casein gene. CPC-A targets pCorp3 to replace the β casein coding region from - 379 relative to the transcription start to the same BclI site. pMAD6 and pCorp3 were constructed by M. Dalrymple.

replacement strategies were attempted. However, the PIPO vector was more complicated to construct than the two-step vector and initial work therefore concentrated on the two-step strategy.

An additional hurdle to be overcome in the generation of YAC transgenic mice is the technical difficulty associated with purifying and handling YAC DNA (see chapter 6). YACs are necessarily present in single copy in the host cell, which facilitates genetic manipulation but hinders purification of YAC DNA in large amounts. Smith *et al.* (1990, 1992) have devised a system for inducibly amplifying YAC copy number, based on the use of a new YAC cloning vector, pCGS966 (Figure 5.2). The insertion of a *GAL1* promoter upstream of the *CEN1* centromere allows galactose-inducible transcription across the centromere, inactivating it. Selection for increased dose of a herpes simplex virus thymidine kinase (HSV-TK) gene present on the right arm of pCGS966 then selects for increased copy number of the YAC. A derivative of pCGS966 designed to retrofit the amplfication system to MP14 was constructed in this project. The construction and use of this plasmid, pAR1, is described below.

5.2 Retrofitting MP14 for copy number amplification

5.2.1 Construction of pAR1 retrofitting construct

The cloning strategy for construction of pAR1 is show in Figure 5.2. The finished plasmid constitutes a replacement left YAC arm containing the amplification components from pCGS966, the *HIS3* yeast selectable marker and the 550 bp ClaI-EcoRI fragment from the centromere to the cloning site of pYAC4 for homology. It was constructed in pBSIIsk+ (Stratagene),

Figure 5.2a: Construction of pAR1 retrofitting construct part I

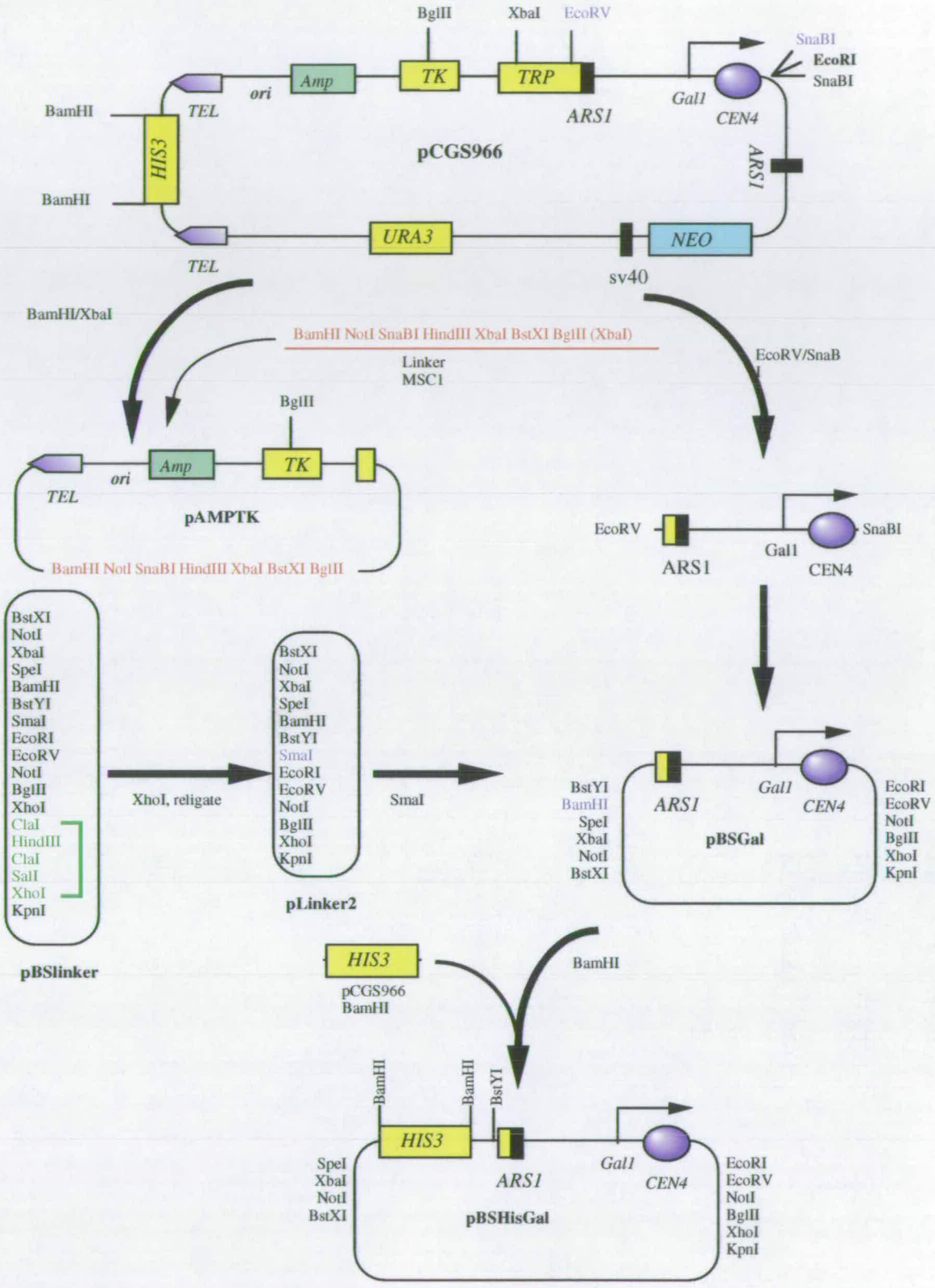
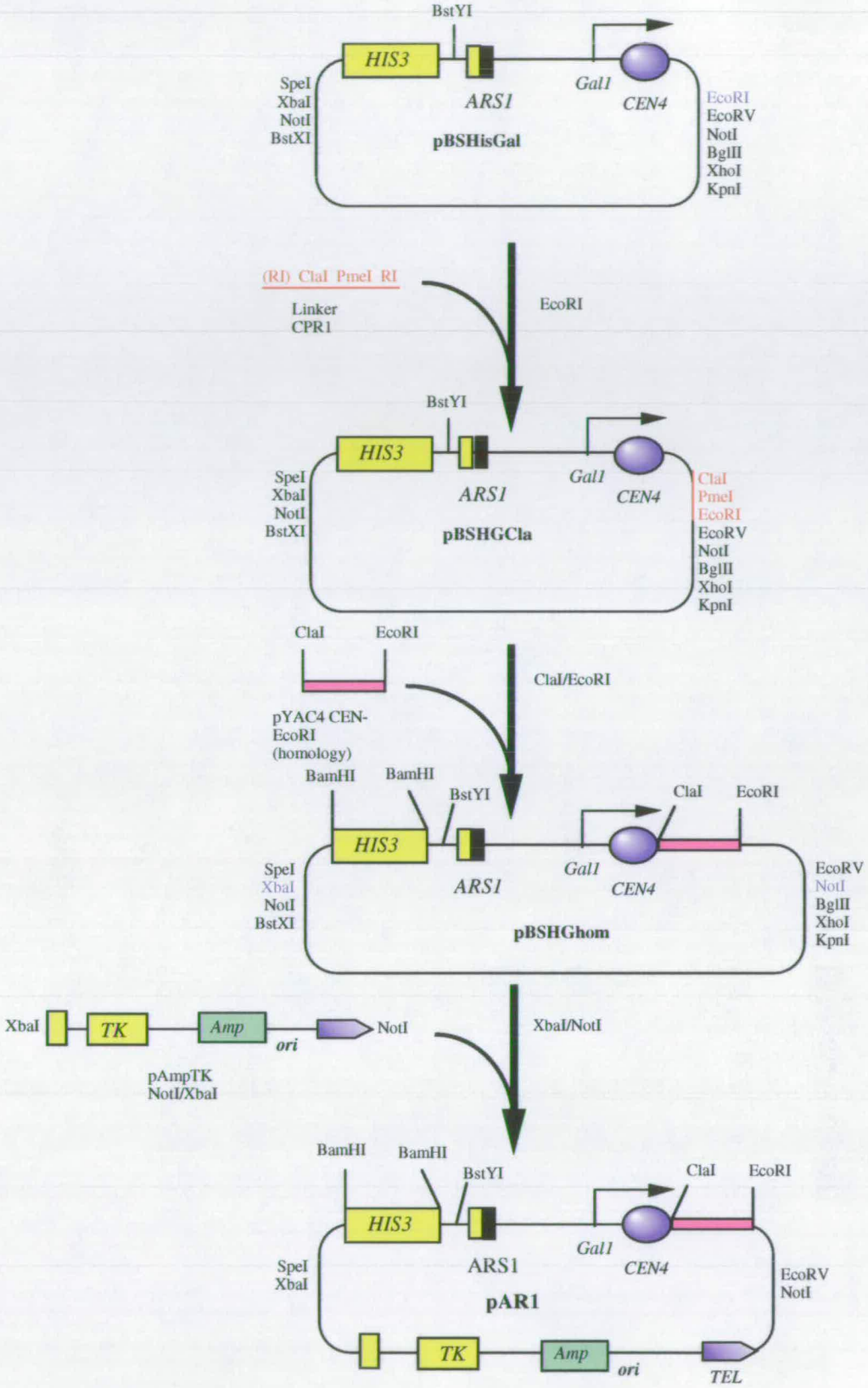


Figure 5.2b: Construction of pAR1 retrofitting construct part II



PCGS966 was digested with EcoRV and SnaBI and the 2.6 kb fragment from the EcoRV site within *TRP1* to the SnaBI site 3' to the centromere was purified. This was ligated to a derivative of pBSIIsk+ called pLinker2. pBSlinker, which was made by introducing new NotI, BglII, XhoI and ClaI sites between the EcoRV site and HindIII site of pBSIIsk+ on a synthetic linker, was kindly donated by E. Emslie. PBSlinker was modified by digestion with XhoI and religation, deleting the ClaI-HindIII-ClaI-SalI fragment between the two XhoI sites and generating pLinker2. Plinker2 was digested with SmaI and ligated to the 2.6 kb EcoRV-SnaBI fragment of pCGS966, generating pBSGal.

The *HIS3* marker from pCGS966 was cloned into the unique BamHI site of pBSGal as a 1.8 kb BamHI fragment, generating pBSHisGal.

The synthetic linker CPR1 (Table 5.2) was designed to insert a ClaI site and a PmeI site into the unique EcoRI site of pBSHisGal, regenerating the EcoRI site 3' to the linker. pBSHisGal was digested with EcoRI and ligated to CPR1, generating pBSHGCla. pBSHGCla was digested with ClaI and EcoRI and ligated to a 550 bp ClaI-EcoRI fragment of pYAC4, corresponding to the region between the centromere and the cloning site immediately to the left of the insert of pYAC-derived YACs. This 550 bp region was intended to serve as homology for targeting MP14. The resultant plasmid was named pBSHGhom.

Finally, pCGS966 (Smith *et al.* 1990) was digested with BamHI and XbaI (Figure 5.2a), and the 6.5 kb band from the left telomere to the XbaI site in *TRP1* was purified. The synthetic linker MCS1 was ligated to this fragment to generate pAMPTK. PAmpTK was then digested with XbaI and NotI and the 6.5 kb fragment was ligated to pBSHGhom digested with XbaI and NotI, generating pAR1 (Figure 5.2b). All cloning junctions and synthetic regions of pAR1 were sequenced.

Oligonucleotide	Sequence
RSPCR5'	GGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCAG ATT GTACTGAGAGTGCACC
RSPCR3'2	CAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCGCGCT GCGGTATTTCACACCGCATAG
URAMCS1	GTAGTCGACGGCGCGCCTCTAGA
URAMCS2	CGCGTCTAGAGGCGCGCCGTCGACTAC
CPR1a	AATTATCGATGTTTAAACG
CPR1b	AATTCGTTTAAACATCGAT
HisAmp1	GCCCTGTGGGCACGCGTGCAGATTGTACTGAGAGTGCACC
HisAmp2	GGCCCTGTGGGCCGTTTAAACGGCGCGCCTGCGGTATTTCACACCGCATA

Table 5.2: Oligonucleotides used in YAC manipulations

5.2.2 Change of MP14 yeast host strain

PAR1 was intended to replace the left arm of MP14, and carries a *HIS3* yeast selectable marker. *HIS3* was chosen because the available mutations in this gene are non-reverting, unlike the other obvious choices of selectable marker. This would greatly facilitate selection for *HIS3*. However, the AB1380 host strain of MP14 is not *his3*, it is *his5*. Therefore the host strain was changed to $his3^{\Delta 200}$ by mating. MP14 cells (*his5 MATa*) were mated to yPH500 ($his3^{\Delta 200}$ *MAT*α; Sikorski and Hieter, 1989) cells on solid medium as in chapter 2; diploids were selected for by complementation of the *his* alleles. His^+ clones were randomly sporulated. Haploid clones were screened for the presence of the *TRP1* and *URA3* YAC markers and the ability to complement MP14 on medium lacking histidine, demonstrating a *his3* phenotype. MP14 is *his5 MATa;* no *his5 Mat*α strain was available. Five clones complemented MP14 but did not complement either yPH499 ($his3^{\Delta 200}$ *MATa*) or yPH500 ($his3^{\Delta 200}$ *MAT*α) on histidine dropout medium, and so were classified as *his3 MAT*α. DNA from these five clones was tested by PFGE and Southern blotting for the continued presence of an intact YAC matching the size of MP14. Unfortunately, none matched the expected MP14 YAC size of 380 kb. Four gave a single band of approximately 440 kb; the other gave a single band of approximately 240 kb. Moreover, the

sample of MP14 DNA that was run alongside them also gave a single band of 440 kb. This DNA was prepared from the same frozen MP14 stock which was used to obtain material for mating with yPH500, so the most likely explanation for this result is that MP14 had somehow become perturbed prior to the mating, and this altered clone had been used both in the mating and to prepare DNA. A subsequent Southern blot of a PmeI digest of this "MP14" clone reveals no resemblance to MP14 when probed with *LA*. This sample did not hybridise to the murine β casein genomic probe β3. This indicates that this clone is not MP14, and indeed does not contain any murine β casein DNA. Therefore, rather than a perturbation it appears that some contamination of MP14 with another yeast strain has occurred at some point.

In light of the above result, further work with these *his3* clones was inappropriate. Moreover, by this time the functionality of pAR1 was in question.

5.2.3 Use of pAR1

The first use of the retrofitting vector pAR1 was in human YAC strains which had been transferred into a *his3* background. This work was done by M. Dalrymple. No amplification was observed in four attempts. The decision was therefore made to test the function of pAR1 by transforming the plasmid into yPH499 and testing for amplification under inducing conditions. No amplification was seen. In the same experiment pCGS966 plasmid DNA was transformed into yPH499 and successfully amplified, though only by a factor of approximately 4. This may suggest that pAR1 is not functional, but it seems equally likely that the experimental protocol was not correct as amplification of 20 fold or greater has been reported using this system (Schedl *et al.* 1993). However, the status of pAR1 was nevertheless uncertain. In addition, the derivation of *his3* MP14 strains had not so far been successful due to an unforeseen perturbation or contamination of the starting material (above). At this point work along these lines was abandoned in favour of an alternative strategy which did not require any YAC amplification. It had been suggested that purification of single copy YACs is in fact straightforward provided care is taken (personal communications from T. Umland, M. Antoniou and D. Markie). Two streamlined strategies were therefore devised to attempt to quickly prepare MP14 for generation of transgenic mice.

5.3 Insertion of pCorp3 into MP14 vector arm

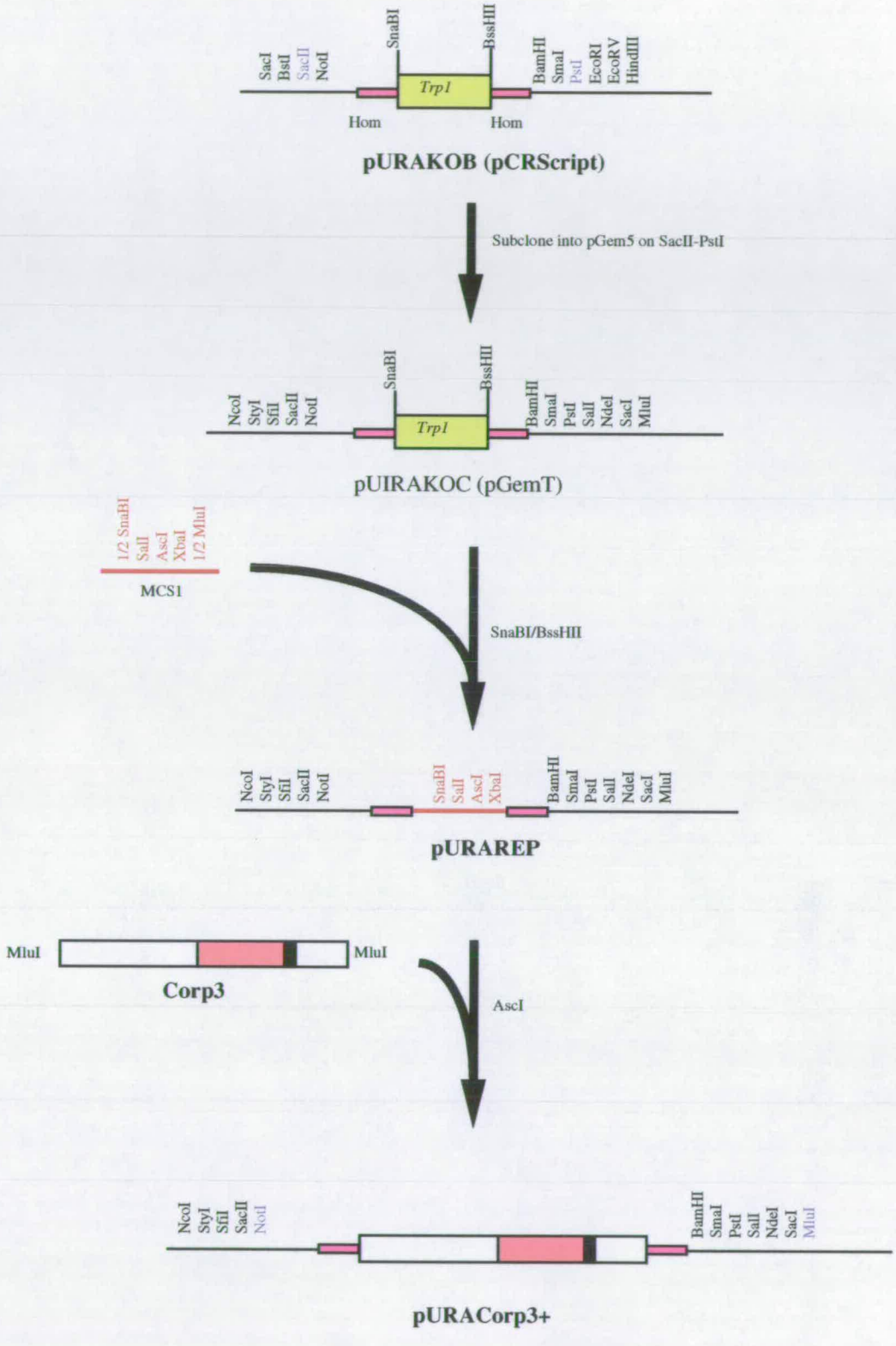
As a means of introducing the protein C reporter construct pCorp3 into MP14 as quickly as possible, a construct was made which targets pCorp3 to the right arm vector sequences of MP14, replacing the *URA3* marker. *URA3* can be selected against using 5-FOA, so the replacement of *URA3* with pCorp3 could proceed in one step.

5.3.1 Construction of pURACorp3

The cloning strategy for construct pURACorp3 is shown in Figure 5.3. pURACorp3 contains the pCorp3 construct flanked by sequences homologous to pYAC4 flanking the *URA3* marker, such that transformation of MP14 with linearised pURACorp3 should result in the replacement of *URA3* by pCorp3.

The *TRP1* gene was amplified from pRS314 (Sikorski and Hieter, 1989; purchased from Stratagene) using the primers RSPCR5' and RSPCR3'2 (Table 5.2). The PCR product was cloned into pCRScript (Invitrogen) and sequenced to confirm identity. This plasmid was named pURAKOB. In order to obtain the desired flanking restriction sites the PCR product was subcloned from pCRScript into pGem5 on a SacII-PstI fragment, generating pURAKOC. pURAKOC was digested with SnaBI and BssHII and ligated to the synthetic linker MCS1 which was made by annealing the oligonucleotides URAMCS1 and URAMCS2 (Table 5.2). The replacement of the *TRP1* insert by MCS1 in pURAKOC generated the plasmid pURAREP. This was digested with AscI and ligated to the 10.5 kb MluI fragment of pCorp3, containing the entire insert of pCorp3. The resultant plasmid was named pURACorp3. PURACorp3+ targets pCorp3 in a forwards direction within MP14; pURACorp3- targets pCorp3 in the opposite orientation.

Figure 5.3: Construction of the retrofitting vector pURACorp3



5.3.2 *Transformation of pURACorp3 into MP14*

MP14 yeast was transformed with linearised pURACorp3+ by the spheroplast method. Transformants were selected on 5-FOA medium. From several hundred colonies which were resistant to 5-FOA, 36 colonies were picked onto replica plates lacking tryptophan, lacking uracil, or supplemented with 5-FOA. Eight of these 36 clones had the desired Trp^+ Ura^- phenotype. These 8 clones were screened for YAC integrity by PFGE and Southern blotting. However, when hybridised to the murine β casein probe β5 none gave a band of the expected size (380 kb). Only two samples hybridised to the probe; in these cases the YAC was approximately 160 kb. The remaining 6 clones did not hybridise to the probe. This was not due to lack of DNA in the samples as evidenced by the EtBr-stained gel (not shown). A further 144 colonies were picked and screened for the appropriate marker phenotype. Eighteen Trp^+ Ura^- clones were identified and screened by PFGE and Southern blotting (Figure 5.4). Again, none gave a band consistent with an intact copy of MP14. Four YAC species were detected, at ~110 kb, ~155 kb or ~320 kb. One YAC had increased in size to ~420 kb, while in 10 clones there was no hybridisation to the probe.

The number of Ura^- Trp^+ clones isolated here was very low compared to the number of Trp^+ clones isolated in the initial screen. In future work the selection against *URA3* would be combined with selection for *TRP1*, which would hopefully isolate Ura^- Trp^+ clones more efficiently. The inclusion of Trp selection is not usual because the *TRP1* gene in pYAC4 has a truncated promoter such that growth on medium lacking tryptophan is inhibited (Markie 1996). However, in this case the large percentage of Ura^- clones which were also Trp^- hindered the isolation of large numbers of clones for further analysis. Selection for the loss of a gene function (*URA3* in this case) can be satisfied in several different ways. Loss of the marker via the desired targeting event is one way. Loss of the marker via a non-specific YAC rearrangement is another. A third route is mutation of another gene in the same physiological pathway, in this case uracil biosynthesis. In this experiment no Ura^- colonies resulted from yeast transformed with carrier DNA only or from untransformed yeast, whereas over 200 colonies were seen in the experimental plates. This suggests that the presence of pURACopr3 is responsible for the change in resistance to 5-FOA in these cells. Since the structural changes to the YAC in those clones which were screened are various but mainly involve large deletions, it is reasonable to

assume that pURACorp3 has mediated large deletions in the YAC. The reasons for this remain unclear. There may be a repeat sequence within pURACorp3 which has recombined with similar sequences in the YAC. A similar mechanism has been exploited to create deletion series of YACs (Heard *et al.* 1994). The existence of repeat elements within pURACorp3 and MP14 has not been investigated. Further work would involve probing MP14 digests with repeat elements and also with pURACorp3 to investigate this possibility. In support of the idea that pURACorp3 has recombined at specific sites in the YAC is the observation that the resultant YACs are grouped into discrete sizes (Figure 5.4) as would be expected of a repeat-induced deletion series.

At this point a considerable effort had been made to modify MP14 for transgenesis, without success. However, two remaining strategies were pursued as far as possible in the remaining time. One strategy was designed to place pCorp3 into the β casein locus of MP14 by either two-step gene targeting or PIPO, whichever succeeded first. From work with pURACorp3 it appeared that pCorp3 could cause fragmentation of the MP14 YAC when introduced into yeast cells (Figure 5.4), which rendered pCorp3 an undesirable choice of reporter gene. However, by this time pCorp3 had been targeted to the β casein locus in transgenic mice (A. Kind, unpublished). The targeting construct CPC-A (Figure 5.1) was used to replace the β casein coding region with pCorp3 in murine ES cells by a two-step gene replacement strategy, and these ES cells had been successfully used to generate transgenic mice. The control for this project was therefore in place, and abandoning work with pCorp3 was not an attractive option, despite the apparent tendency of pCorp3-derived constructs to recombine with multiple sites within MP14.

The second remaining strategy was to insert a different reporter gene encoding enhanced green fluorescent protein (*eGFP1*; purchased from Stratagene) into MP14 and attempt to transfer the resultant YAC into cultured mammary epithelial cells. This work did not utilise pCorp3, and it was hoped that it would yield data more rapidly via tissue culture.

Figure 5.4: Clones produced by transforming MP14 with pURACorp3

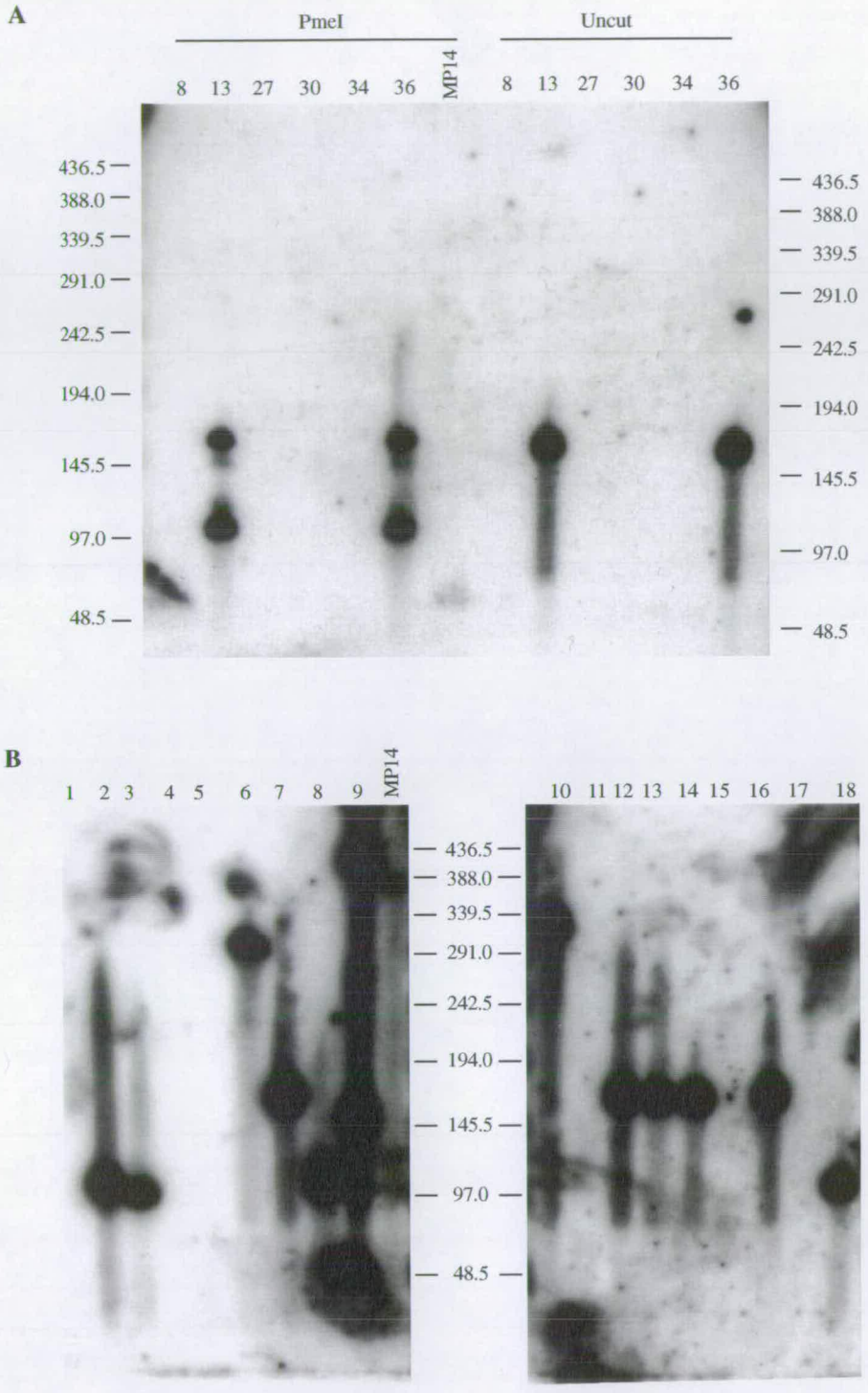


Figure 5.4: Clones produced by transforming MP14 with pURACorp3
Linearised pURACorp3 was transformed into MP14. Ura- Trp+ clones were analysed by PFGE. Panel A: first round of clones, digested with PmeI and uncut. Panel B: second round of clones, uncut. The probe in each case is β5. No clone gave a band consistent with the parental YAC. Sizes are shown in kb of the λ markers.

5.4 Two-step gene targeting of pCorp3 to the MP14 β casein locus

This strategy aimed to introduce the reporter construct pCorp3 into the β casein locus of MP14 via the counterselectable yeast marker *URA3*. Both two-step targeting and PIPO strategies were pursued. As the first step in both *URA3* was removed from the right arm vector sequences of MP14, replacing it with *HIS3*.

5.4.1 New his3 MP14 derivative strain

MP14 was mated to yPH500 as before. After sporulation, screening and backcrossing, two clones were identified. YATS1 was His^- Ura^+ Trp^+ and complemented histidine auxotrophy in another clone which had been designated *his5 MATα*. YATS1 therefore appears to be *his3 MATa*. YATS2 is His^- Ura^+ Trp^+ and complemented histidine auxotrophy in MP14, and is therefore designated *his3 MATa*. However, a Southern blot to assess the integrity of the YAC in these strains was poor and inconclusive, although it did suggest that the YAC was intact. This blot was to be repeated, but in the meantime yATS1 was transformed with *HIS3* to replace *URA3* on the right arm of the YAC.

5.4.2 Replacement of URA3 with HIS3 on yATS1

The *HIS3* gene was amplified from pRS313 (Sikorski and Hieter, 1989; Stratagene) using the primers pRSPCR5' and pRSPCR3'2 (Table 5.1). This PCR product was purified on a Microcon column (Amicon) and used to transform yATS1 cells. Very good stimulation over untransformed controls (>500 cfu vs. 0 cfu) was observed on medium lacking histidine. However, subsequent selection of 144 clones with 5-FOA revealed only 5 which were also Ura^-. These 5 clones were analysed by Southern blot. Initially uncut YACs were analysed (Figure 5.5a). yE32 and yE67 clearly contain two bands which hybridise to the *RA* probe, and therefore two YACs. These clones were not investigate further. Clones yE11, yE43 and yE92 gave single YAC bands, but of different sizes. The YAC in yE11 and yE43 appears larger than expected (~400 kb rather than ~380 kb) while yE92 gave a band of the expected size for MP14. These three clones were analysed by partial restriction digest (Figures 5.5b and 5.6). MP14 and yATS1 DNA were included as controls. The MP14 band profile was as expected (see chapter 4). However, the profile of yATS1 did not match that of MP14 (Figure 5.5b). In addition to each

MP14 band, yATS1 exhibits a second set of bands. The most logical explanation for this is that yATS1 carries two YACs. Since yE11 and yE43 contain the bands which are in yATS1 but not in MP14, these clones appear to have lost the MP14-identical YAC. Bands which are shared between yE11/yE43 and MP14 are presumably comigrating fragments in the two YACs.

In contrast, yE92 exhibited the same partial restriction fragment band pattern as MP14 (Figure 5.6a), suggesting that while yE43 and yE11 appear to have lost the murine casein YAC present in yATS1, yE92 has lost the spurious YAC. To further test this, DNA from yE92 and MP14 was digested with ClaI and probed with the whole MP14 YAC to generate a detailed fingerprint of the YAC (Figure 5.6b). The two band profiles appear to match within the limits of the blot. The small (~7 kb) ClaI band that appears to be present in MP14 (Figure 5.6 lane 10) but inconclusive in yE92 (Figure 5.6 lane 11) is discernible in the autoradiograph.

A standard Southern blot was performed to analyse the fine structure of the right arm of yE92 (Figure 5.6). This blot was unfortunately probed with *RA1* which carries a small region of the left arm of pYAC4, but the results are nevertheless consistent with a *bona fide* replacement event. The bands which are presumably detected within the left arm of the YAC by *RA1* are arrowed in Figure 5.7. All are consistent with the fragments expected to be detected within the left arm of the YAC by this small region of the probe.

The structure of the right arm of yE92 is therefore consistent with the expected recombination event replacing *URA3* with *HIS3*. On the basis of these results, it appears that although yATS1 is not as expected and contains a second YAC in addition to MP14, in yE92 the replacement of *URA3* with *HIS3* in the MP14 YAC has been accompanied by the loss of the second spurious YAC. Although this event appears serendipitous, it is perhaps not surprising. YE92 was isolated after a screen for $His^+ Trp^+ Ura^-$ clones among those which had incorporated the *HIS3* targeting construct. Out of 144 clones screened all were $Trp^+ His^+$, but only 5 were identified which were Ura^-. Among these 5 clones three carried only one YAC. In the presence of the *HIS3* PCR product, uracil auxotrophy could be achieved by targeting of both YACs, loss of both YACs, or targeting of one and loss of the other. Loss of both YACs would result in a negative screening result however, while the other two outcomes could both result in $Trp^+ His^+ Ura^-$ clones. It appears that in yE32 and yE67, both YACs have been targeted, although in yE32 the

spurious YAC appears deleted relative to yE11, yE43 and yE67 (Figure 5.5) which may indicate spontaneous loss of *URA3* rather than targeting. In yE11, yE43 and yE92, targeting of one YAC appears to have been accompanied by loss of the other YAC. In yE11 and yE43 the murine casein YAC was lost, while in yE92 the spurious YAC was lost. By using 5-FOA selection the loss of *URA3* has in fact been used to deliberately reduce the number of YACs in a clone (Heikoop *et al.* 1994). In contrast to the initial selection for histidine prototrophy described here, 5-FOA would apply the more stringent selection first and allow a far greater number of Ura$^-$ clones to be isolated. In this experiment only 144 clones were screened, which is consistent with the isolation of only 5 clones of the desired phenotype, since rather than a recombination event as expected the 5-FOA selection was for a spontaneous deletion event.

5.4.3 Two-step gene targeting

The clone yE92 appeared to be as desired, and was therefore transformed with a construct designed to target *URA3* to the β casein locus. A new plasmid, pBCURA3, was constructed to target *URA3* to the βcasein locus in MP14. The *URA3* gene was amplified from pRS306 (Sikorsky and Hieter, 1989; Stratagene) using the primers HisAmp1 and HisAmp2 (Table 5.2). The PCR product was cloned into pGemT (Promega) and sequenced. The PCR product was then excised from pGemT with MluI and AscI, which cut in the primer sequences, and cloned into MluI-digested pJH3; MluI and AscI are compatible. pJH3 is a derivative of CPCA carrying a multiple cloning site in place of Corp3 (Figure 5.1), and was kindly donated by J. Howcroft. This plasmid was called pBCURA3.

PBCURA3 was linearised with NotI and XhoI and transformed into yE92 cells by the LiAc method. The transformants were selected on medium lacking uracil and tryptophan. A large number of Ura$^+$ Trp$^+$ colonies developed, and screening of 18 of these 16 proved to carry a YAC of the same size as MP14 when analysed by Southern blot of undigested DNA probed with *URA3* (Figure 5.8). There was a second band in all lanes which did not resolve from the limiting mobility region of the gel. This band is presumed to result from hybridisation of the probe to the remainder of the endogenous copy of *URA3* in this yeast strain. This band was also seen in yE92 (Figure 5.8), further suggesting that it is not related to the presence of the targeting construct in the transformed clones. As expected, the yE92 YAC did not hybridise to the probe.

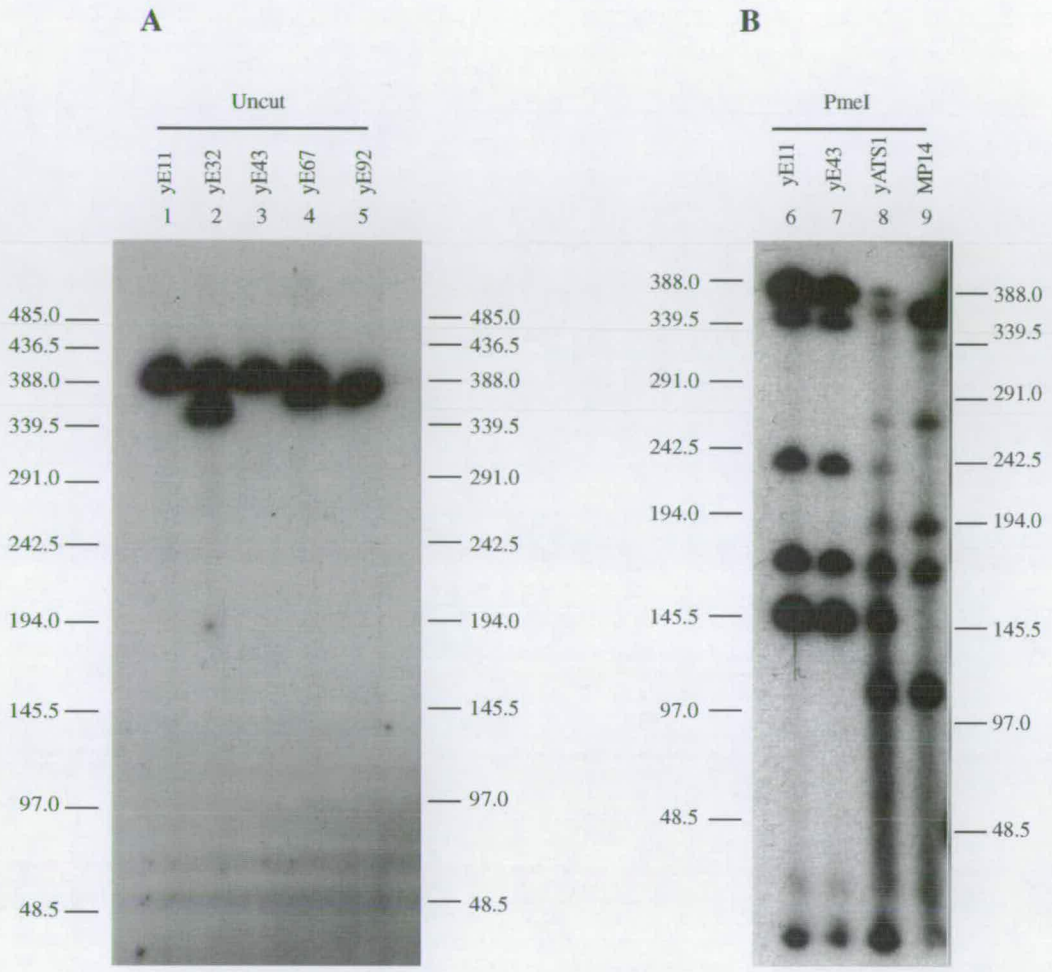


Figure 5.5: Structure of yE clones and yATS1

A *HIS3* PCR product targeted to replace *URA3* was transformed into the *his3* strain yATS1. Five clones were isolated following marker screening.

Panel A: The five clones were separated by PFGE and probed with *RA*. yE32 and yE 67 contain two YACs are were discarded.

Panel B: yE11 and yE43 were compared to yATS1 and MP14 by partial restriction digest and hybridisation to *LA*. Neither yE11 nor yE43 resemble MP14. yATS1 appears to harbour an additional YAC to MP14. The MP14 YAC appears to have been lost in yE11 and yE43, leaving the spurious YAC.

Sizes of lambda concatemer markers (NEB) are shown in kb.

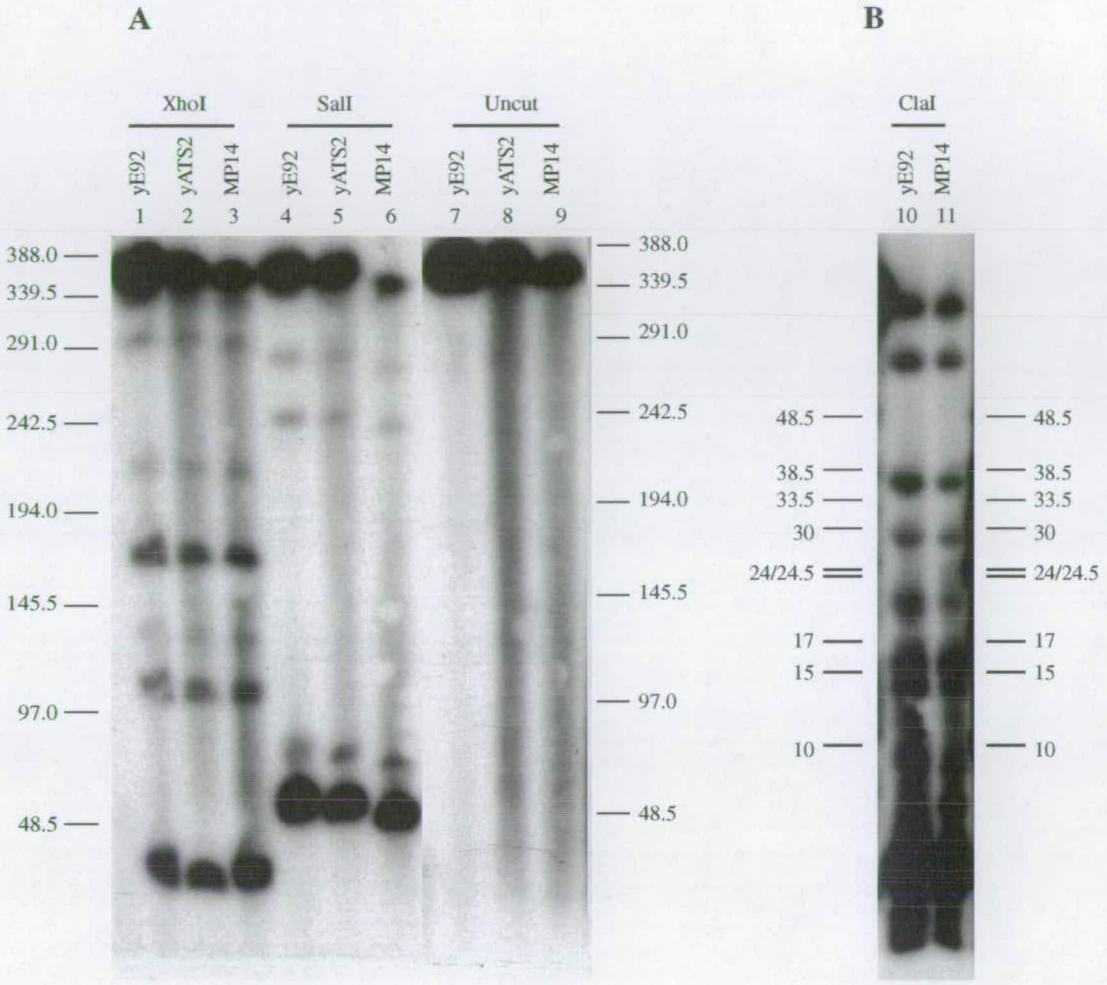


Figure 5.6: Structure of yE92

yE92 was generated by transforming a HIS3 PCR product targeted to replace URA3 into yATS1 (not shown above). The structure of yATS1 is shown in Figure 5.5.

Panel A: Partial digests of yE92 and MP14 (and yATS2) probed with *LA*.

Panel B: Complete digest of yE92 and MP14 probed with MP14.

Sizes of lambda markers are given in kb. With the exception of a small ClaI fragment (less than 10 kb), the structure of yE92 matches that of MP14. The small mobility difference seen in lane 6 compared to lanes 4 and 5 is a loading artefact.

Figure 5.7: Fine structure of right arm of yE92

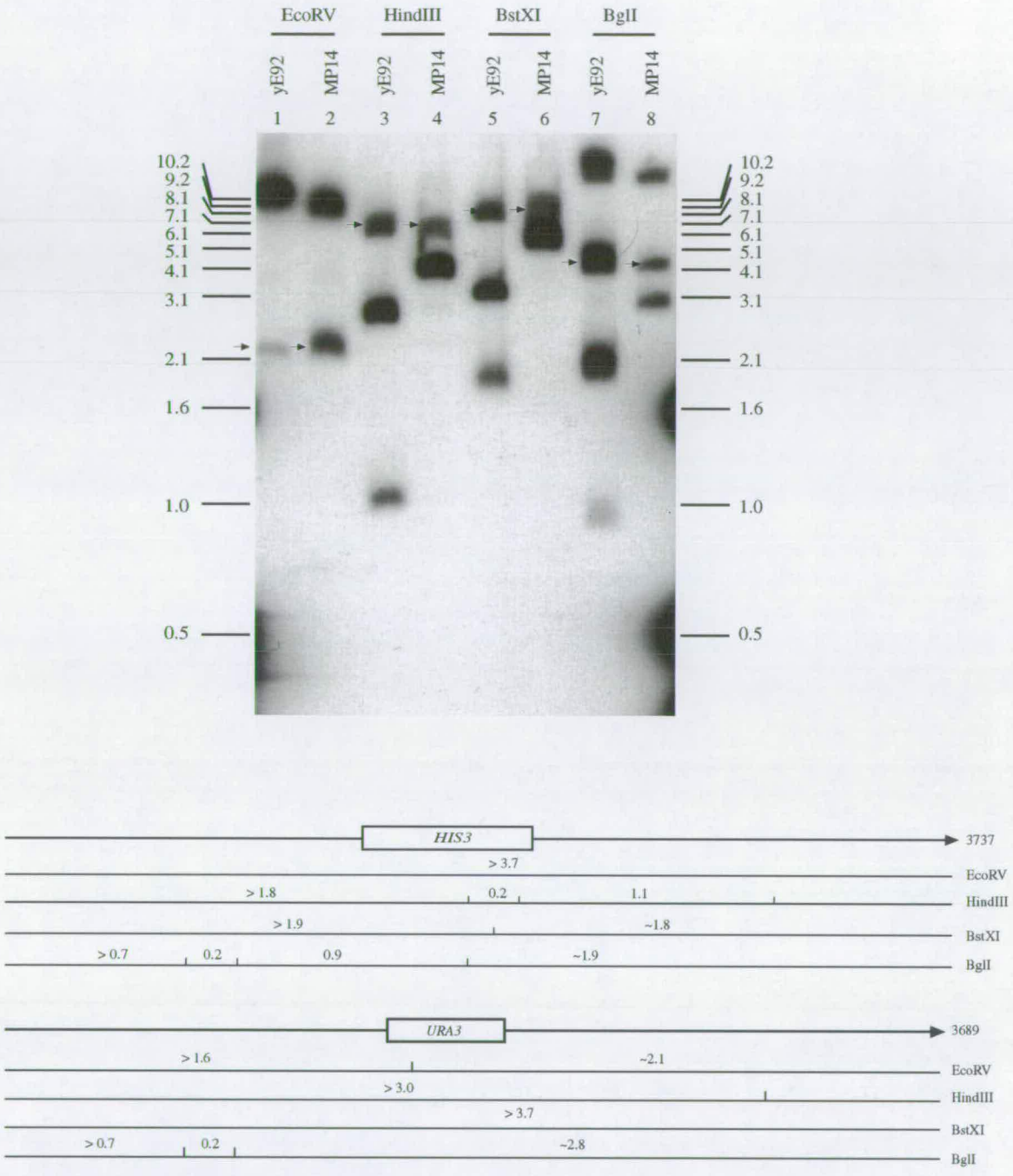


Figure 5.7: Fine structure of right arm of yE92

yE92 and MP14 DNA was digested completely and hybridised to *RA1* to compare right arm fine structure. The fragments detected are consistent with the replacement of *URA3* by *HIS3* in yE92. Thebands which are presumably detected in the left arm of the YAC by *RA1* are arrowed. See text.

The second step was to remove *URA3* from the β casein locus of the YAC and insert pCorp3, by means of the targeting constructs CPC-D or CPC-A. Work in mice had yielded a transgenic mouse carrying CPC-A targeted to its β casein locus, replacing the βcasein coding region. CPC-D chimaeric mice had yet to transmit the targeted locus to the germline. Therefore, it was attractive to pursue CPC-A in the YAC clones.

Clone 1 from Figure 5.8 was transformed with NotI-linearised CPC-A by electroporation. It was envisaged that spontaneous mutation of the Ura^+ phenotype to a Ura^- phenotype in both transformed and untransformed cells would give a high background in this experiment. Electroporation was therefore used to maximise the transformation efficiency and therefore the proportion of plated cells competent to carry out the desired recombination. No significant stimulation was seen over background (150 vs. 100 cfu). Given the number of cells plated (5×10^8) this background equates to a reversion rate of 2×10^{-7}. The transformation efficiency in this experiment was found to be approximately 10^3 cfu/μg, which is fairly low for electroporated yeast cells where 10^4 to 10^5 cfu/μg are routinely expected. Attempts were made to improve the transformation efficiency by using high purity DNA, plating fewer transformants per plate, and different transformation protocols, but the efficiency remained at approximately 10^3 cfu/μg. It is not known if this is due to the yeast strain used, which was generated here from MP14 and yPH500 and is largely uncharacterised, or if a technical problem is responsible. Transformation of a proprietary strain by the same protocol failed on the one occasion that it was attempted.

Figure 5.8: Insertion of URA3 into the β casein locus of yE92

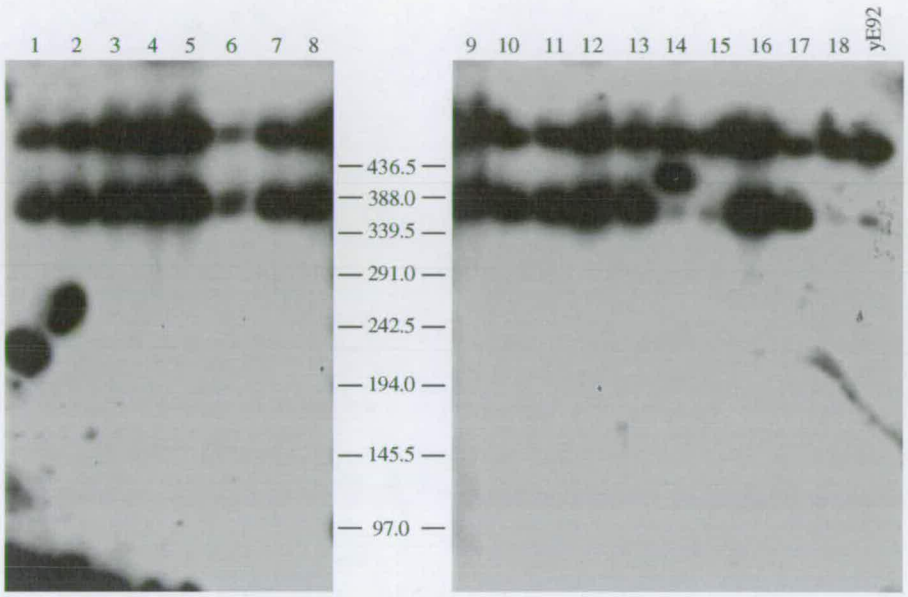


Figure 5.8: Insertion of URA3 into the β casein locus of yE92

pBCURA3 was used to target *URA3* to the β casein gene in yE92. Undigested DNA from Trp+Ura+ transformants was hybridised to a *URA3* probe. The probe hybridised to a band of the expected size of 380 kb in all but two clones, 14 and 18. The second band in all lanes is presumably the endogenous copy of URA3. Sizes of λ concatemer markers are shown in kb. Note that the 48.5 kb marker band is not present on this filter.

At this point in the project, there was no time available for further manipulations of the MP14 YAC. Had more time been available, there are several additional avenues which could be pursued. The second step of the targeting experiment could be optimised with respect to transformation efficiency with the aid of proprietary yeast strains to monitor progress. However, there is no guarantee that they would respond as would the in-house strain. Transfer of the *URA3*-targeted YAC to a defined highly transformable yeast strain would further facilitate this work. The *kar1* yeast transfer system (Hugerat *et al.* 1994; Spencer *et al.* 1994) allows straightforward transfer of YACs between host strains by exploiting the $kar1^{\Delta 15}$ mutation (Spencer and Simchen, 1996). This mutation prevents proper karyogamy during yeast mating, resulting in rare but selectable events in which a YAC is transferred during mating between two otherwise unaltered host strains. In retrospect this would have been a better method than traditional mating for changing the MP14 host strain to *his3*. However, *kar1* transfer was a newly-described technique at the start of this project and not widely tested.

Transfer of the *URA3* targeted YAC to a highly-transformable strain would not solve the inherent defect of two-step targeting with *URA3*. This is caused by the relatively high rate of spontaneous mutation to 5-FOA resistance. In addition to loss of *URA3*, 5-FOA resistance can occur by other loss-of-function mutations in the uracil biosynthesis pathway. These so-called "forward" mutations occur at a sufficiently high frequency to hinder efficient selection of the desired clone after transformation with a construct which removes or deletes *URA3* without also introducing a second means of selection. PIPO procedures are more efficient than two-step gene replacement when using *URA3*. This is because in a PIPO experiment, the clone carrying the PIPO construct is isolated before selection against *URA3*, so that every cell in the selective medium is competent to carry out the desired second-step recombination. In the second step of a two-step strategy, only those cells which have taken up the transforming DNA are competent to recombine as desired, while every cell is competent to acquire 5-FOA resistance by spontaneous mutation. In addition, the proximity of direct repeat DNA flanking *URA3* in the PIPO construct increases the probablility that *URA3* will excise, further increasing the proportion of 5-FOA-resistant clones which have recombined as intended. A PIPO vector was planned and constructed in this project, but insufficient time remained in which to characterise and use the completed construct.

In light of the failure of the work to manipulate the mouse casein YAC, alternative means of studying casein expression in transgenic animals was sought. The human casein locus is present on the 400 kb YAC HP8 (ICRFy90C0518) which was isolated from the ICRF human YAC library (T. Monaco, unpublished). This YAC was isolated and partly characterised by M. Dalrymple. A restriction map and comparison with genomic material has also been reported for this YAC (Rijnkels, 1997b), which concluded that it represents a single unrearranged region of the human genome harbouring the entire human casein gene locus. Detection of human casein RNA and protein in transgenic mice is straightforward. Thus, no additional characterisation or manipulations of HP8 were required. The emphasis of this project therefore switched from studying and modifying the MP14 murine casein YAC, to generating transgenic mice from HP8 by microinjection and analysing the expression of the human caseins as transgenes.

6. Chapter 6 - Human YAC transgenic mice

6.1 Introduction

As described in chapter 5, the mouse YAC MP14 proved refractory to rapid retrofitting and thus was unavailable for generating transgenic mice within the period of this project. As an alternative, the decision was taken to generate mice with a human casein YAC. No YAC modifications were required, since the expression of the human casein genes would be readily detectable in a transgenic mouse. The 400 kb human casein YAC HP8 (ICRFy901C0518) was a kind gift from M. Dalrymple. HP8 has been previously characterised by restriction mapping (M. Dalrymple and A. Tomlinson, unpublished). Rijnkels *et al.* (1997b) have also restriction mapped this clone and report that it is unrearranged relative to genomic DNA. The structure of our isolate of HP8 matches that reported by Rijnkels *et al.* (1997b). The HP8 YAC DNA was therefore purified and used to generate transgenic mice by pronuclear microinjection. Initially, no transgenic lines were obtained from screening of 255 founder (G_0) pups generated by microinjection of HP8 YAC DNA protected against shear by spermine and spermidine. A second round of microinjection used HP8 YAC DNA protected by hexamine-cobalt chloride (HCC). Seven transgenic lines were established after screening of 313 G_0 animals. The expression of human α, βand κcasein was investigated in five of these lines by northern blot and RTPCR. The expression of all three genes was below the detection threshold of the northern blot. This appears to be due to fragmentation of the YAC transgene. Expression of κ casein was detected by RTPCR in four lines.

In 1993 several reports appeared describing the generation of transgenic mice from YACs via embryonic stem (ES) cells (Choi *et al.* 1993; Davis *et al.* 1993; Jakobovits *et al.* 1993; Lamb *et al.* 1993; Pearson and Choi, 1993; Strauss *et al.* 1993) and pronuclear microinjection (Schedl *et al.* 1992; Peterson *et al.* 1993b; Schedl *et al.* 1993a, 1993b). The YAC transgenes appeared to be maintained without significant rearrangement although initial rearrangement before integration seemed higher than that of smaller constructs, and there were no deleterious effects due to the yeast CEN and TEL sequences (reviews: Forget, 1993; Peterson *et al.* 1997). Therefore YACs represent a workable means of introducing very large segments of genomic DNA into transgenic mice.

The introduction of YACs into murine ES cells by spheroblast fusion was first reported by Jakobovitz *et al* (1993) and has the advantage that the YAC is not subjected to any of the insults of DNA purification and handling, since the whole yeast cell is simply spheroblasted and fused to a murine ES cell. However, as well as being more time consuming than microinjection this method has the considerable disadvantage that the entire yeast genome tends to integrate with the YAC, which could affect expression of the YAC. Since this study was ultimately concerned with assessing the functionality of the YAC in a transgenic environment, the presence of other heterologous DNA was undesirable and spheroblast fusion was therefore not an attractive option. Lipofection of YAC DNA into ES or other cultured cells has also been reported (Choi *et al.* 1993; Lamb *et al.* 1993; Pearson and Choi, 1993; Strauss *et al.* 1993), but this method does not appear to increase the frequency of intact transgenes over that expected from microinjection (D Markie, personal communication; Vassilopoulos *et al.* 1999). Furthermore, both cell-mediated methods require G418 selection of the transfectants, which might select for cells in which the YAC had integrated into active chromatin, prejudicing the expression data. This would be undesirable in this study. Microinjection of purified YAC DNA was first reported by Schedl *et al* (1992). Compared to cell-mediated methods, microinjection of purified YAC DNA is faster and has the obvious advantage that gel-purified YAC DNA is relatively free of contaminating yeast sequences, unless the YAC comigrates with a yeast chromosome. However, purification and handling of appreciable quantities of intact, pure YAC DNA present several technical difficulties. Perhaps the greatest challenge is presented by the relative scarcity of the YAC in the yeast cell. YACs are present at single copy, which means in the case of HP8 that the YAC only makes up approximately 2.5% of the total yeast DNA. This may be overcome to an extent by the use of YACs containing inducible amplification cassettes which allow the copy number to be increased (Smith et al. 1990; Chapter 5), but this requires retrofitting of pYAC-based YACs. In addition to being time-consuming, this may result in unstable copy number which would render the YAC refractory to further manipulation by homologous recombination. Some experiments suggest that the amplification of YAC copy number by this method may be low and unpredictable (M Dalrymple, personal communication). Finally, YAC modification by homologous recombination (HR) involves a risk to the integrity of the insert. In this chapter purification of a single copy YAC is reported, which circumvented

the need for any YAC modification.

A second problem is caused by the size of YACs. DNA of several hundred kilobases is extremely sensitive to shear and other damage such as nicking by ethidium bromide (EtBr), damage by heavy metal ions often present in trace amounts in impure water and enzymatic digestion. To protect against shear, the buffers used to dissolve the purified YAC DNA were supplemented with polyamines which protect DNA from shear by intra- and intermolecular ionic interactions (Kaiser *et al.* 1963). Initially 70μM spermidine and 30μM spermine were used (Schedl *et al.* 1993b) to supplement the microinjection buffer. In later experiments hexamine cobalt chloride (HCC; Kovacic *et al.*, 1995) was used, as this has been shown to effect a similar protection against shear *in vitro* (Kovacic *et al.* 1995) and has no apparent effect on mouse embryo survival (P Shiels and Y Gibson, personal communication). Standard precautions were taken against heavy metal damage and enzymatic digestion such as using ultra high purity water (impedance at least 15 MΩ) and always wearing gloves. Efforts were also made to prevent EtBr coming into contact with the YAC preparation, by using washed electrophoresis equipment (the PFGE equipment has never been exposed to EtBr) and removing the gel regions to be stained from the unstained preparative section.

Finally, the risk of fragmentation associated with YAC transgenes was starting to become clear at the time of this study (Forget, 1993; see also Peterson *et al.* 1998a). In an effort to ensure an intact YAC transgene a large number of embryos were injected. The numbers were limited by convenience - more injections could only have been advantageous.

6.2 YAC DNA preparation

YAC DNA was prepared essentially according to Schedl *et al* (1993b, 1996; for details see Chapter 2), except that single copy YAC DNA was purified. This method involves purifying the YAC DNA by excising it as a band from a pulsed field gel, and then concentrating it by a second round of electrophoresis at 90° to the first (Figure 6.1). For clear resolution and high yield very high quality DNA is required. The best method for obtaining this was found to be that given in section 2.8.2, with the optional Zymolyase treatment included. DNA was prepared in agarose

plugs by this method and subjected to the initial round of PFGE in 1% Rapid agarose (Life Technologies). Several batches of YAC DNA were prepared and pooled for microinjection. The preparation of one such batch is shown in Figure 6.1. The edges of the PFG were cut off and stained, then reassembled once the YAC band had been excised from the stained region. EtBr nicks DNA in the presence of UV light, so every effort was made to keep EtBr away from the YAC DNA being prepared. With the stained region for reference the YAC band was cut out of the unstained portion of the gel. The same was done for a yeast chromosome which served as a marker during the procedure (Figure 6.1).

The PFG slice was then cast in a 4% NuSieve agarose (FMC) gel and subjected to the second round of electrophoresis at 90° to the first, to concentrate the DNA in the NuSieve agarose at the end of the Rapid agarose slice (Figure 6.1). The portion of the gel containing the marker chromosome was cut off and stained with EtBr to monitor this migration; the YAC DNA was excised from the NuSieve agarose without staining. After equilibration in 1x TE, 0.1mM HCC, or 1x TE, 70μM spermidine, 30μM spermine, the agarose was digested with agarase to release the YAC DNA and the solution was briefly spun to pellet any undigested agarose before microdialysis against injection buffer/0.1 mM HCC or microinjection buffer/70μM spermidine, 30μM spermine. An aliquot was analysed by electrophoresis to assess the concentration and integrity of the YAC before several batches were pooled to obtain sufficient material for microinjection.

6.3 Constructs and probes

The physical organisation of HP8 is illustrated in Figure 4.12. The structure of the pYAC4 arms and the *Trp1* and *RA* probes is shown in the same figure. The *Trp1* probe was excised from pURAKOB (Chapter 5) on a SnaBI-BssHII restriction endonuclease fragment. pURAKOB was constructed to replace *URA3* with *TRP1* on the right arm of pYAC4-derived YACs. The primers RSPCR5' and RSPCR3'2 (Table 6.1) were used to amplify *TRP1* from pRS314 (Sikorski and Hieter, 1989; Genbank accession U03440; purchased from Stratagene) by PCR. The PCR product was cloned into pGemT (Promega) creating pURAKOB. The insert was sequenced to confirm identity of the PCR product. SnaBI cuts within the PCR product just upstream of the *TRP1* coding region, while BssHII cuts within RSPCR3' between the pRS314 priming sequence

and the pYAC4-homologous sequence. Thus the SnaBI-BssHII fragment of pURAKOB contains the *TRP1* fragment from basepairs 320 to 1207 of pRS314.

The human β casein screening probe used was a PCR product amplified from intron 2 of the human β casein gene (Figure 6.2) The primers HBC1 and HBC2 (Table 6.1) were used to amplify a 770 bp fragment from basepairs 4899 to 5668 of the human β casein gene (Genbank accession L10615), using human placental DNA (a gift from A Scott) as a template. The PCR product was cloned into pGemT and sequenced to confirm identity.

The *RA* probe was the 1.7 kb BamHI-PvuII fragment of pBR322 (Genbank accession: 701749; Sutcliffe, 1979; Figure 6.2), corresponding to the cloning vector backbone of the right arm of pYAC4.

Probe	Type	Size	Primers/Comments
hβcas8	genomic	0.8 kb	HBC1: ACTATATTATTTGTAGAGA
			HBC2: ATTTTACTATTTTATGGTTCAT
Trp1	yeast marker	0.9 kb	RSPCR5': GGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCG
			CACGCGGCGCAGATTGTACTGAGAGTGCACC
			RSPCR3'2: CAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGA
			CCCAACGCGCGCTGCGGTATTTCACACCGCATAG
RA	pBR322	1.7 kb	BamHI-PvuII restriction fragment of pBR322
hβcas1	cDNA	1.1 kb	Human β casein cDNA from ATG to polyA
hαcas1	cDNA	0.9 kb	Human α casein cDNA from ATG to polyA
HACP2	oligo	23 bp	sequence: TCACCACTGTAGCATGACGTTAT
HKCP2	oligo	23 bp	sequence: CCAGGAGAGTGTGAAGTAGTAAT

Table 6.1: Probes for screening HP8 mice

The structure of the screening probes is shown in Figure 6.2. RSPCR5' and RSPCR3'2 consist of 22 nt of sequence which primes within pRS314 (Stratagene; Sikorski and Hieter, 1989) with a 5' tail of homology to pYAC4 flanking *URA3* (see text). Only the primer sequence of RSPCR3'2 was included in the probe.

Figure 6.1: HP8 YAC DNA purification

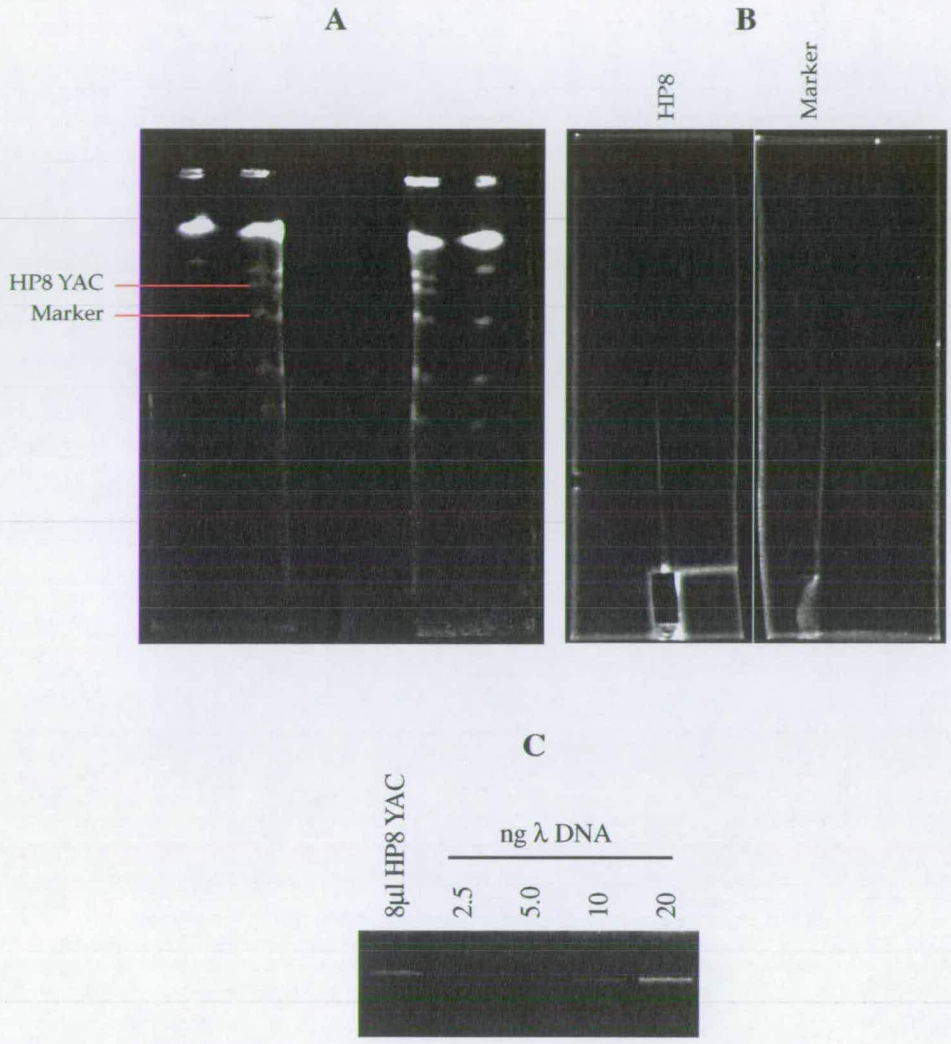


Figure 6.1 - HP8 YAC DNA purification

A: HP8 total yeast DNA was prepared in agarose plugs and separated by PFGE. The YAC band was excised from the unstained portion of the gel, as was a marker chromosome band (indicated above).

B: The PFG slices were cast in a 4% low melting point gel and subjected to electrophoresis at 90° to the PFGE. The side with the marker chromosome was stained to monitor the progress of the chromosome into the low melting point agarose. The YAC DNA was excised from the unstained portion of the gel, and the gel was then restained. The staining up the edge of the gel slice indicates DNA that has migrated out of the side of the slice when the orientation of the slice was not quite parallel to the electrophoretic field.

C: An aliquot of the purified YAC was run out alongside bacteriophage lambda (λ) DNA standards to measure concentration. This sample is at about 2 ng/µl.

Figure 6.2: Probes for screening HP8 mice

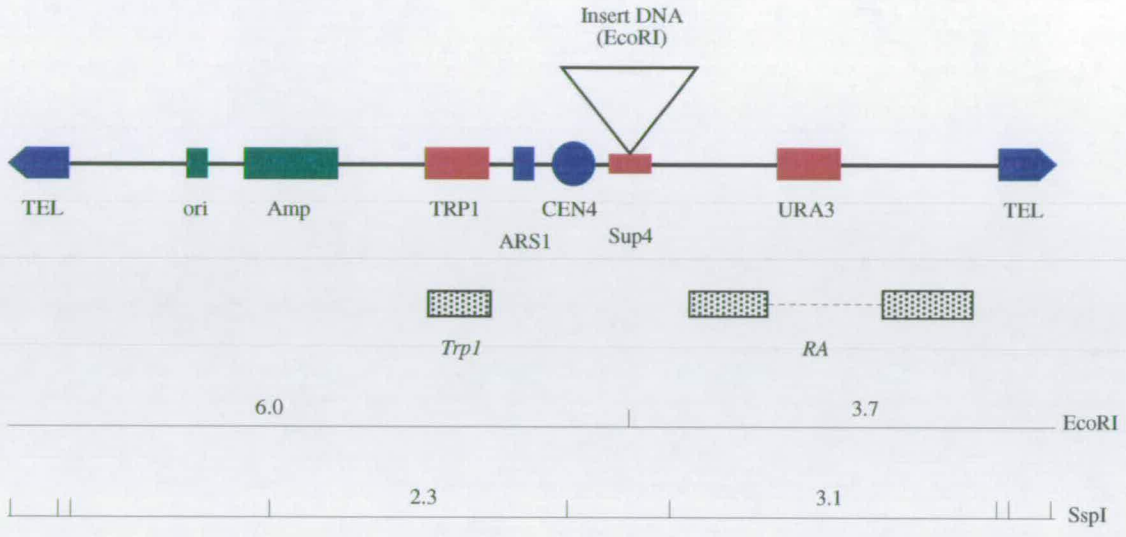
Panel A: A schematic representation of a pYAC4-derived YAC is shown, with the insert DNA removed to show the structure of the vector arms. TEL: telomere; ori: *E. coli* origin of replication; Amp: *E. coli* ampicillin resistance (β lactamase) gene; TRP1: *S. cerevisiae TRP1* selectable marker; ARS1: *S. cerevisiae ARS1* autonomous replication sequence; CEN4: *S. cerevisiae CEN4* centromere; Sup4: *S. cerevisiae* $Sup4^0$ repressor; URA3: *S. cerevisiae URA3* selectable marker. The *Trp1* and *RA* probes are shown below.

Panel B: Structure of the human β casein gene, including EcoRI restriction sites and hβcas8 probe. Exons are represented by black boxes. Probe sequences are represented by grey boxes.

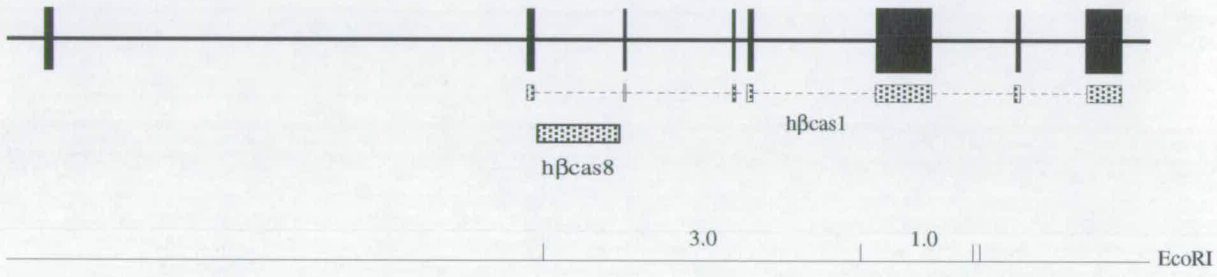
Panel C: Structure of the human κ casein gene, including EcoRI restriction sites. Exons are represented by black boxes. Probe sequences are represented by grey boxes.

Figure 6.2: Structure of probes for screening HP8 mice

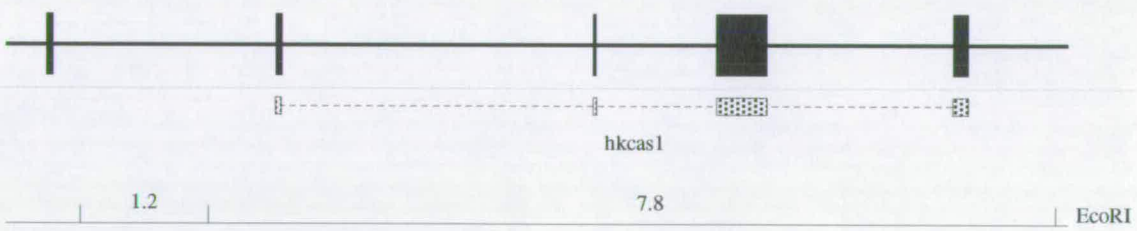
A: pYAC4-derived YACs



B: Human β casein gene



C: Human κ casein gene



1 kb

6.4 HP8 microinjection study I

HP8 YAC DNA was purified and dissolved in microinjection buffer supplemented with 70μM spermidine and 30μM spermine. Microinjection of HP8 YAC DNA into fertilised mouse eggs, embryo transfer and subsequent animal husbandry were undertaken by Yvonne Gibson and Scott Rhodes in the small animal husbandry unit at PPL Therapeutics. Tail biopsy DNA was prepared from pups and screened by Southern blotting for the presence of *TRP1* using the *Trp1* probe. The results are shown in Figure 6.3 and summarised in Table 6.2.

Injection of HP8 YAC DNA into 3399 fertilised eggs led to transfer of 1584 embryos into 54 pseudopregnant foster animals and the subsequent birth of 261 live pups, with an average litter size of 6. 255 pups survived to be biopsied and were screened by SspI digestion, Southern blotting and hybridisation to the *Trp1* probe. One clear transgenic animal was identified, along with two of uncertain status (Figure 6.3). 96/158.23 gave a band of the expected size though very weak (the positive control was loaded at approximately single copy), suggesting that this animal may be mosaic for the YAC, since the DNA loading was as expected. 96/150.33 gave an aberrant band pattern; in contrast to the expected SspI band of 2340 bp, bands of approximately 1.8, 3.1, 4.2 and 10kb were detected, along with a high molecular weight smear which was outside the sizing range of the gel (Figure 6.3). Given the presence of multiple bands, and especially of the high molecular weight smear, it might be that the digestion of 96/150.33 DNA was not complete. However, a band of 2340bp would still be expected and the band of 1.8kb cannot be explained by this hypothesis unless the YAC has become perturbed. Another possibility is that the signal in this lane is a results of an artefact such as contamination of the sample or electrophoresis apparatus. The leakage of signal into the adjacent lane suggests that whatever has hybridised to the probe was in the sample rather than contaminating the comb. Although it is possible that the DNA sample was contaminated, it is not likely. Plugged pipette tips were used at all times when pipetting DNA during this screen, and great care was taken not to bring control DNA (pYAC4 in this case) into contact with samples; the pYAC4 DNA for the control digests was diluted to single-copy concentration separately. Moreover, the bands are not consistent with pYAC4 contamination. 96/150.33 was therefore regarded as a putative transgenic. 96/150.34 was loaded alongside 96/150.33 and gave bands similar to that sample (Figure 6.3) suggesting that this signal was caused by leakage between the lanes.

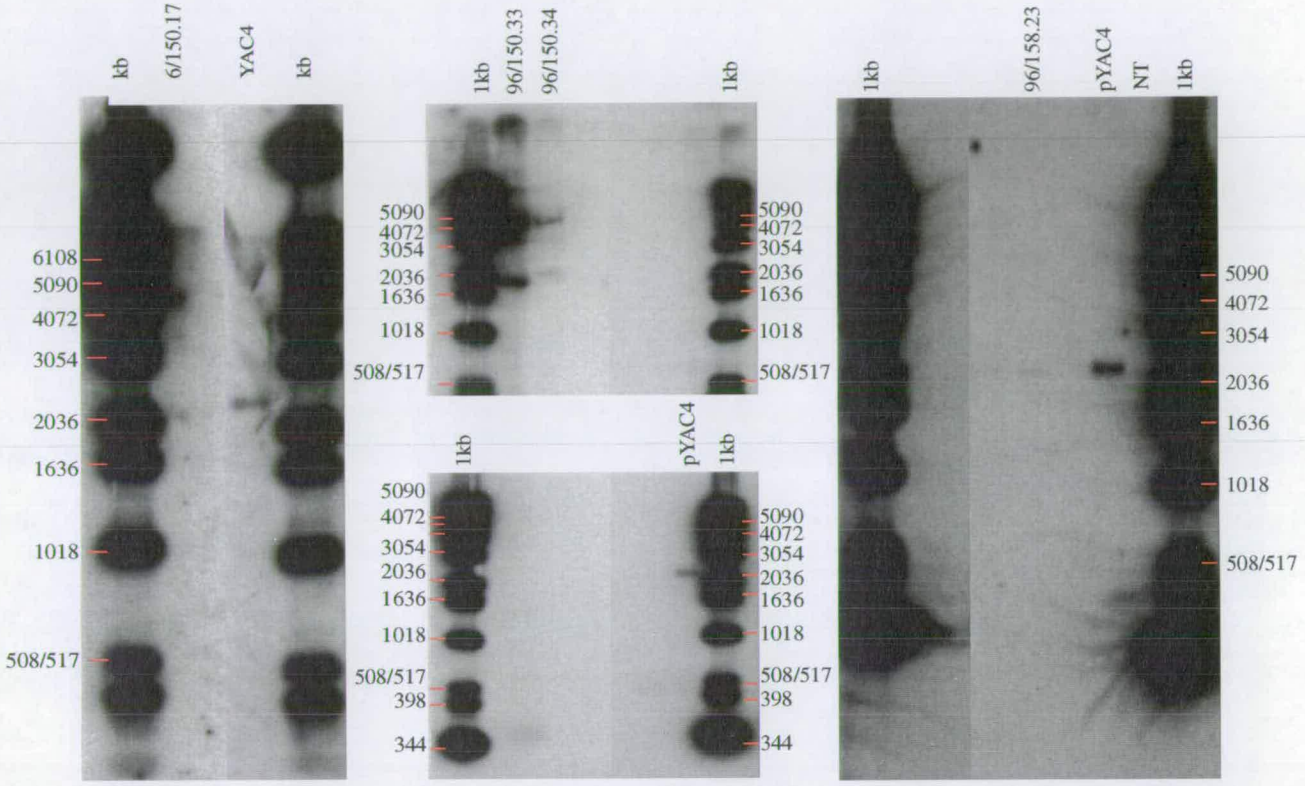


Figure 6.3: Initial screen for HP8 transgenic mice.

5µg of tail biopsy DNA was digested with SspI and probed with a Trp1 probe to detect the left arm of the YAC. Markers are 1kb ladder from Life Technologies; sizes are shown in bp. Irrelevant lanes have been omitted for clarity. PYAC4 - 20pg of pYAC4 added to an SspI digest of nontransgenic mouse genomic DNA.

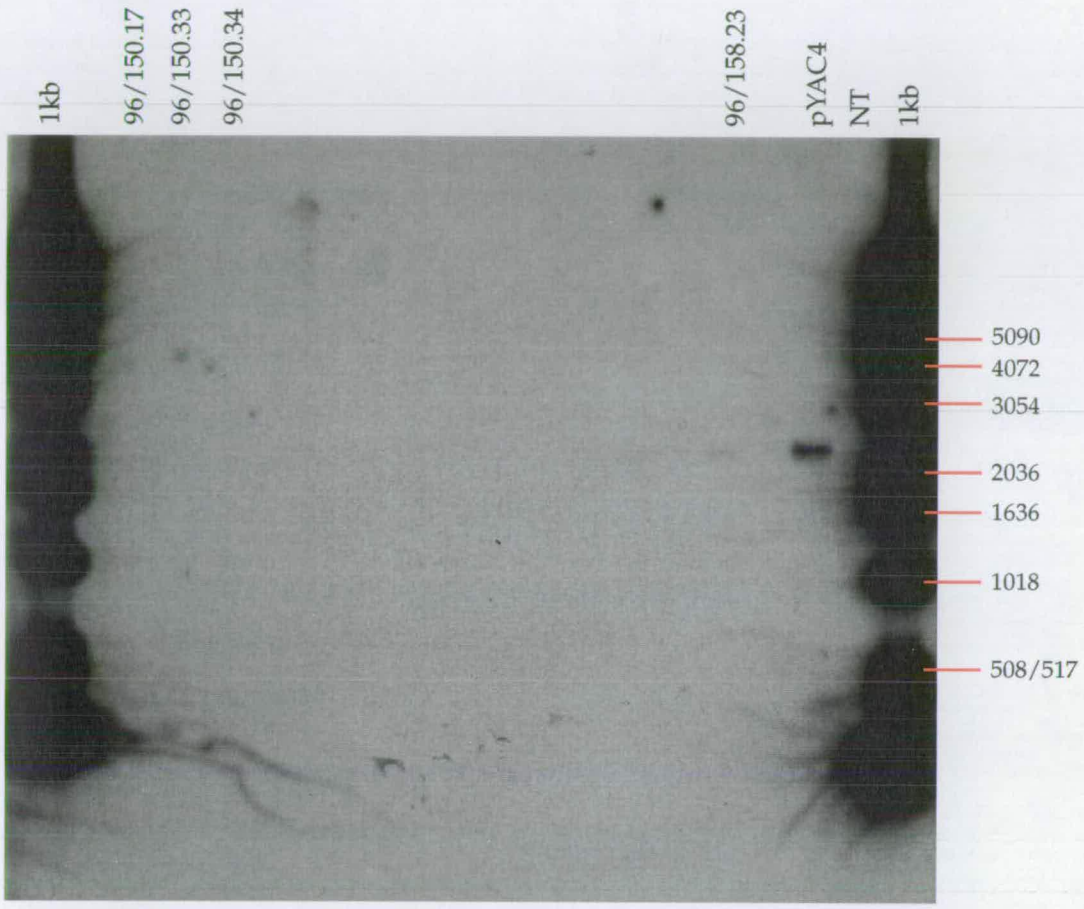


Figure 6.4 -Secondary screening of HP8 G0 animals

5µg of DNA from animals which were possible transgenics (Figure 6.4) were digested with SspI and probed with Trp1. 1kb - 1kb ladder from Life Technologies, with some sizes shown; NT - non-transgenic control. pYAC4 - 20pg pYAC4 added to 5µg of nontransgenic mouse DNA and digested with SspI, equating roughly to a single copy gene. Only 96/158.23 appears to be transgenic, and at less than single copy.

One other sample, 96/150.17, gave a signal when probed with the *Trp1* probe. This appeared to be an artefact as it was not an obvious band (Figure 6.3), but the proximity of the 1 kb marker DNA precludes definitive rejection of 96/150.17.

In light of these results 96/158.23, 96/150.33 and 96/150.17 were mated to produce G_1 animals. However, no hybridisation of hβcas8 was detected to DNA from a total of 68 offspring from these matings (Table 6.3), and in two subsequent Southern blots DNA from both 96/150.17 and 96/150.33 failed to hybridise to *Trp1* (one blot is shown in Figure 6.4). Taking all of these results into account it seems likely that only one positive animal was produced, 96/158.23, that this animal was mosaic and that the transgene was not present in the germline. Therefore a second study was initiated to attempt to generate animals harbouring the HP8 YAC transgene in the germline.

6.5 HP8 microinjection study II

A second preparation of HP8 YAC DNA was made and microinjected as before except that the shear-protectant used in this study was HCC. Kovacic *et al* (1995) reported that 0.1 mM of HCC in 0.5x TBE provided complete protection for megabase plant DNA when it was passed repeatedly through a standard yellow pipette tip. Subsequently P. Shiels and Y. Gibson (personal communication) found that up to 4mM HCC in microinjection buffer had no detrimental effect on embryo survival in murine recipient foster mothers, and that 0.1mM HCC in TE was sufficient to protect megabase DNA from passage through a yellow pipette tip. Since HCC is reported to give fewer problems with precipitation and needle blockage than spermine/spermidine (Y. Gibson, personal communication; Kovacic *et al.*, 1995) 0.1 mM HCC was added to all solutions used in the manipulation of YAC DNA in solution, to compare its effectiveness with that of spermine/spermidine and to assess any effect it might have on long-term embryo survival and transgenesis. However, another difference from Study I was that in this study, owing to time constraints and the low yield of purified YAC DNA from each preparation, not every batch was tested for YAC integrity before microinjection. Previous batches were tested and displayed degradation amounting to no more than 20%. Nevertheless, this remains an omission from this experiment which may have a bearing on the results (below).

Figure 6.5 - HP8 YAC transgenic mice

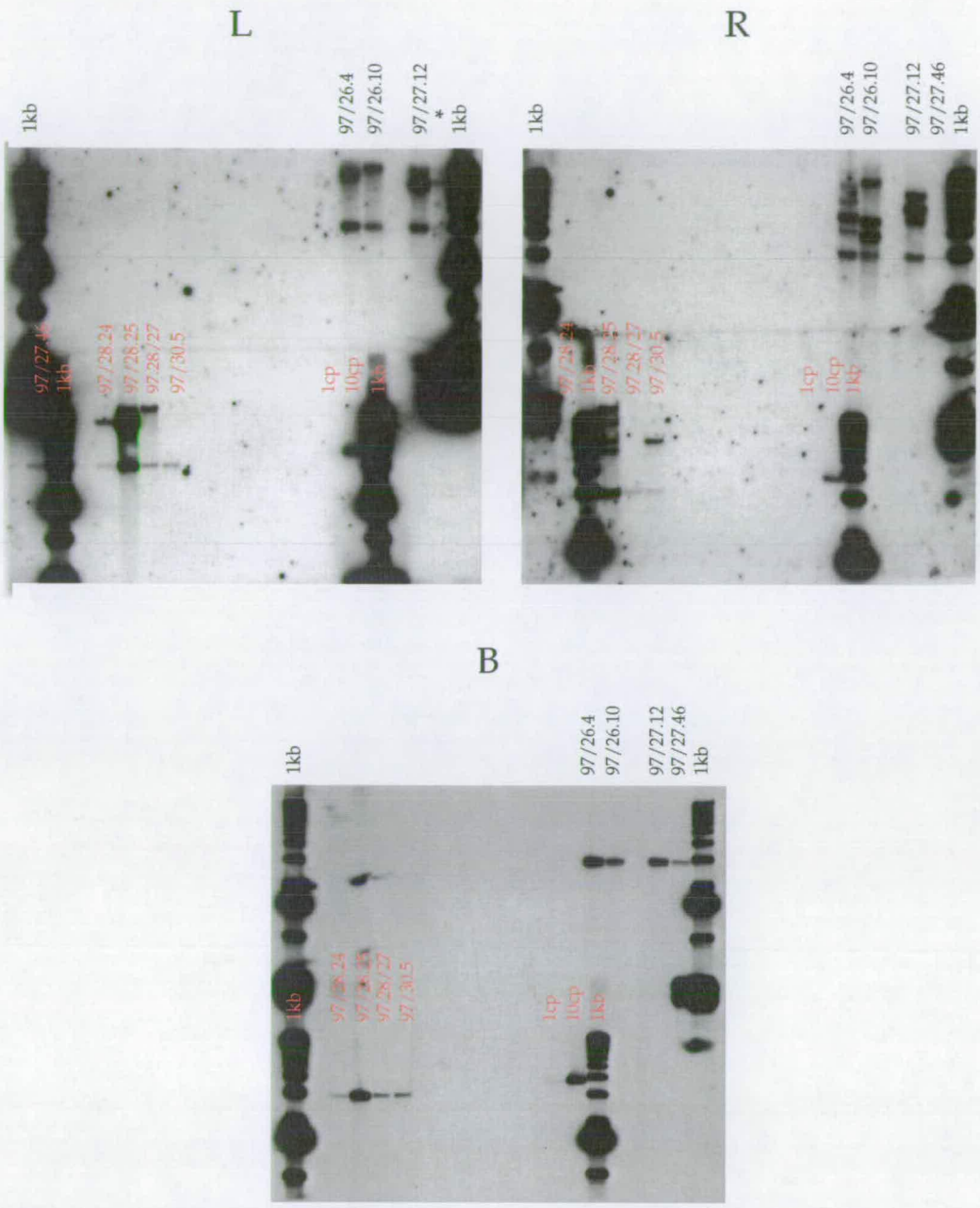


Figure 6.5 - HP8 YAC transgenic mice

Mouse tail biopsy DNA was digested with EcoRI and screened for three YAC markers by Southern blotting. See text for details. L - probed with hβcas8 and then *Trp1*. R - probed with hbcas8 and then *RA*. B - probed with hβcas8. Markers are 1kb ladder from Life Technologies as before. 1cp, 10cp - copy number controls of pYAC4 DNA in nontransgenic mouse DNA, digested with EcoRI. Only positive samples are labelled. * - empty lane, showing some well leakage from adjacent transgenic sample.

The results of screening the G_0 animals thus produced are shown in Figure 6.5 and summarised in Tables 6.2 and 6.3. Note that Figure 6.5 represents a secondary screen of the putative transgenics and any samples of uncertain status from the first screen (not shown). Of 2964 injected embryos 1890 were transferred to 63 pseudopregnant foster mothers, yielding 327 live births from 60 pregnancies. Tail biopsy DNA from 313 animals was screened by Southern blot as before, except that the DNA was digested with EcoRI restriction endonuclease. Three gels were prepared by splitting a triple digest of each sample, and the blots were hybridised to hβcas8. This eliminated any of the ambiguity caused by gel-specific artefacts in the first study. Two of the blots were then reprobed with either *Trp1* or *RA*. The hβcas8 probe was not removed before reprobing (Figure 6.5).

Eight G_0 samples hybridised to the human β casein probe hβcas8 (Figure 6.5), giving a transgenic frequency of 2.4% of live births. A ninth band was seen in the hβcas8 blot (Figure 6.5, bottom panel), but in only one of the three blots. Also, this apparent band was irregular, unlike the bands in the gel, and was not located a discrete number of lanes away from the marker lane. Therefore this is assumed to be an artefact. Of the eight positive samples 97/26.4, 97/26.10, 97/27.12 and 97/28.25 also hybridised to both YAC arm probes, while the other four samples hybridised to one arm or the other. This suggests that considerable fragmentation of the YAC has occurred during generation of the animals. Also significant is the finding that the animals which screened positive for all three YAC markers were apparently present in multiple copy, while the other four animals harboured fewer copies. While copy number determination is unreliable in G_0 animals due to mosaicism and inherently variable due to problems in estimation, autoradiography (small linear response range) and manipulating very low concentration DNA, this result nevertheless undermines any inference from the data that the multiple copy YACs may be intact. It may be that the multiple copy animals simply contain multiple YAC fragments, including one or more copies of each marker.

Further evidence that the YACs are not intact in these animals is provided by the band patterns detected by the YAC arm probes in Figure 6.5. The gels were not run far enough for accurate size determination in the desired range, but some inferences about the structure of the YAC transgene arrays can be drawn from them. Although most of the bands discussed below are discernible in Figure 6.5, some are only apparent in the original autoradiographs. Work is

underway to repeat these blots to obtain clearer data concerning the structure of the YAC in these mice.

As there is a single EcoRI site in each arm of the YAC where it abuts the insert DNA, three results are expected when YAC-transgenic mouse DNA is cut with EcoRI and probed for the YAC arms, depending on the structure of the array. The left arm of pYAC4 covers 6kb, the right 3.7kb (Figure 6.2). A head-head junction of two YAC copies would therefore give an EcoRI band of 12kb, a head- tail junction would give a band of 9.7kb and a tail-tail junction would give a band of 7.4kb. An intact copy of the YAC juxtaposed with genomic DNA would give an EcoRI band of at least 6kb when probed with *Trp1* and a band of at least 3.7kb when probed with *RA*.

When probed with *Trp1* (Figure 6.5 panel L), 97/26.4 gives two bands which might be 9.7kb and 12kb, consistent with the existence of at least one head-head and one head-tail junction in this animal. There is also a very faint band above 12kb which may represent a head-genome junction, and another very faint band below the strong 9.7kb band. The 9.7kb band is considerably more intense than the other bands. If the faint band above 12kb does represent a head-genome junction, then the relatively intensities of these three bands are consistent with the existence of a single head-genome junction and several internal head junctions in this animal, which has been estimated at 10-copy based on the human β casein signal (Figure 6.5 panel B). The *RA* pattern also seems to contain a band at 9.7kb (Figure 6.5 panel R), indicating a head-tail junction. However, the relative strength of the 9.7kb band detected by *Trp1* is not matched by that detected by *RA*. In addition, *RA* detected bands at approximately 9kb, 7kb, 4.5kb and 5.5kb. The 9kb band is also detected by *Trp1*, suggesting that this may represent a head-tail junction in which some deletion has occurred. The 5.5kb band detected by *RA* is considerably more intense than the other bands, suggesting that this fragment is present in higher copy within the array than the other fragments detected. However, this is hard to understand given that this 5.5kb band would most likely represent a tail-genome junction of which only one would be expected, and which relies on the nearest genomic EcoRI site so that even if there were more than one tail-genome junction, multiple bands of the same size would still be highly unlikely. Similarly, the presence of the 4.5kb and 7kb bands cannot be explained. Neither was detected by *Trp1*, and the 7kb band is clearly too small to represent a 7.4kb junction of two intact right arms (Figure 6.5).

Thus the structure of this 10-copy array is complex and uncertain.

97/26.10 appears to carry approximately 2 copies of human β casein (Figure 6.5), and gave a single large band when probed with *Trp1* which may represent a head-head junction. *RA* again detected several different bands in this animal. A band of the appropriate size for a head-tail junction (expected size 9.7kb) was detected by *RA*, and there were also bands of approximately 4kb and 5kb. However, the putative head-tail band is not detected by *Trp1*, suggesting that this does not represent a head-tail junction.

97/27.12 carries approximately 5 copies of human β casein (Figure 6.5). Trp1 detected what appears to be a 9.7kb head-tail junction in this animal, but also a very strong band at approximately 7kb. This could be a head-genome junction, but the intensity is unexpected. This band is clearly too small to represent a 9.7kb head-tail junction. In this animal *RA* detected bands of approximately 5, 6 and 7.5 kb. The largest band is consistent with a tail-tail junction, but might also be the same size as the 7kb band detected by *Trp1,* and thus represent a deleted head-tail junction. The other two bands cannot be accounted for, but either could represent a tail-genome junction.

Little can be inferred from the pattern obtained from 97/27.46 when probed with *Trp1*. The marker ladder signal obscured any bands which may have been present (Figure 6.5). *RA* failed to detect any signal for this animal. The human β casein signal suggests that there is a single copy of the gene present (Figure 6.5 panel B).

In 97/28.24 the human β casein gene is estimated to be present at single copy (Figure 6.5 panel B). *Trp1* detected a single band while *RA* failed to detect any signal. The *Trp1* band appears to be 8-9kb, and thus may represent a head-genome junction. These data are consistent with the presence of a single copy of HP8 truncated at the right arm.

97/28.25 is estimated to carry 10 copies of the human β casein gene (Figure 6.5 panel B). *Trp1* detected bands consistent with head-head and head-tail junctions. The putative head-tail band is much more intense than the putative head-head band. However, there is also a band of approximately 7kb which is of comparable intensity to the head-tail band. The 7kb band might

represent a head-genome junction, but the intensity cannot be explained in that context. *RA* detected bands of approximately 5.5kb, 6.5kb, 9.5-10kb and over 12kb, all of approximately equal intensity. The 9.5-10kb band probably represents a head-tail junction, as implied by the similar band detected by *Trp1*. The 6.5kb band may represent a deleted head-tail junction since *Trp1* detected a band of a similar size.

97/28.27 has approximately 1 copy of the human β casein gene (Figure 6.5). *Trp1* detected a single band of approximately 12kb in this animal, which could be indicative of a head-head junction. However, given that the animal is apparently single-copy, this could alternatively represent a head-genome junction. *RA* failed to detect any bands in this animal. Similarly to 97/28.24, this animal may thus carry a single copy of HP8 truncated at the right arm.

97/30.5 is apparently single-copy with respect to human β casein (Figure 6.5). *Trp1* failed to detect any bands from this animal. *RA* detected a single band of approximately 6.5kb, which may represent a tail-genome junction fragment. Thus 97/30.5 may carry a single copy of HP8 truncated at the left arm.

In summary, the EcoRI digest data convey a confused picture of the organisation of the transgene array in most of the transgenic mice. Although the data are not of high quality, it nevertheless seems likely that the arrays are complex and not made up of uniformly-intact YAC copies in these animals. Further analysis would be useful in order to elucidate the structure of the YAC copies so that any expression data could be viewed in an appropriate context. To ascertain the number of copies of the YAC arms, digests which closely bracket the probe sequences would be desirable, so that a single band would be expected for each animal. To determine more closely the structure of the junctions within the array, EcoRI digests are useful. The data in Figure 6.5 are of low quality; a repeat of these blots should yield clearer information about the structure of the ends of the YAC copies. Finally, PFGE is an obvious choice when investigating YAC structure. Several reports in the literature detail methods and results using PFGE to analyse YAC transgenic arrays (Peterson *et al.* 1998). In general the frequency of intact YAC copies as determined by this method is low. The results presented here would concur with this view.

6.6 Summary of HP8 transgenic founders

Table 6.2 summarises the results of the two HP8 microinjection studies. A cursory review of these suggests that of the two injection buffers, HCC leads to a higher transgenic frequency than s/s. There also appears to be a significant increase in embryo survival, both in terms of embryos implanted and in live births. However, it must be borne in mind that microinjection of YACs into mouse eggs is a new technique at this facility, having been performed only once before this study, and the HCC study was done soon after the s/s study. It is therefore likely that some improvement in technique may have occurred between the two. YAC DNA is difficult to microinject, causing significant needle blockage and increased visible damage to the egg relative to plasmid DNA (Y. Gibson, personal communication). Nevertheless, these data are intriguing. Further work with HCC suggests that transgenic frequency is at least as good as that obtained with s/s (A. Tomlinson and P. Shiels, unpublished), although transgene integrity remains to be investigated.

Buffer	Inj	2 cell	Transf	Preg	Births	Ave/F	Tailed	TG	% TG
s/s	3399	1770	1584	47/54	261	6	255	1	0.4
HCC	2964	2024	1890	60/63	327	5	313	8	2.4

Table 6.2 - Summary of HP8 transgenic mouse studies
s/s - spermidine/spermine microinjection buffer; HCC - hexamine cobalt chloride microinjection buffer; Inj - number of injections; 2 cell - number of embryos surviving to 2 cell stage; Transf - number of embryos transferred to foster mothers; Preg - number of pregnancies achieved; Ave/F - mean litter size; Tailed - number of pups surviving to tail biopsy at two weeks post parturition; TG - number of transgenics detected; % TG - number of transgenics as a percentage of live births (Births).

Nine transgenic founders were identified in this study, 8 of which were generated using HCC and 1 using s/s. The single animal generated from s/s did not transmit the YAC and was discarded. Of the remaining 8 animals, 4 possessed all three screening markers. These 4 animals were apparently multicopy. This correlation together with the unexpected data obtained from digests probed for terminal YAC restriction fragments imply that the YAC has become fragmented in these animals. Nevertheless, they were bred and used to analyse the expression of the human casein transgenes.

Founder	Injection Buffer	Sex	Screening Left arm	Screening β casein	Screening Right arm	Copy no. (approx.)	G_1	TG G_1
96/150.17	s/s	F	?	nd	nd	nd	21	0
96/150.33	s/s	M	?	nd	nd	nd	21	0
96/158.23	s/s	F	+	nd	nd	<1	26	0
97/26.4	HCC	F	+	+	+	10	11	3
97/26.10	HCC	F	+	+	+	2	7	5
97/27.12	HCC	F	+	+	+	5	7	3
97/27.46	HCC	M	nd	+	-	1	17	5
97/28.24	HCC	M	+	+	-	1	15	0
97/28.25	HCC	M	+	+	+	10	14	5
97/28.27	HCC	M	+	+	-	1	15	3
97/30.5	HCC	F	-	+	+	1	14	3

Table 6.3 - HP8 transgenic mice

See text for details. s/s - microinjection buffer/ 70μM spermidine, 30μM spermine. HCC - microinjection buffer/0.1mM hexamine cobalt chloride. nd - not determined; ? - uncertain. G_1 - live offspring surviving to be biopsied. TG G_1 - number of transgenic G_1 animals

6.7 Analysis of human casein gene expression in HP8 transgenic mice

In order to establish stable transgenic lines and to obtain animals for tissue samples, the nine transgenic founders (Table 6.2) were mated to wild-type mice. As discussed above, 96/158.23 failed to transmit the YAC. 97/28.24 also failed to give rise to any transgenic offspring. The other founders transmitted the YAC successfully, so that seven transgenic lines were established (Table 6.3). These animals were mated to wild-type mice as necessary to obtain lactating mammary tissue.

6.7.1 Analysis by northern blot

Day 10-lactating mice were killed and mammary glands recovered and snap frozen in liquid nitrogen (N_2(l)) by Y. Gibson and D. Morrison in the small animal husbandry facility at PPL. Total RNA was prepared by the guanidium thyocyanate method (Chomczynski and Sacchi, 1987) from these glands, using the RNAZol kit (Invitrogen) or the TRI reagent kit (Sigma). Five

Figure 6.6: Northern blot analysis of human casein transgene expression

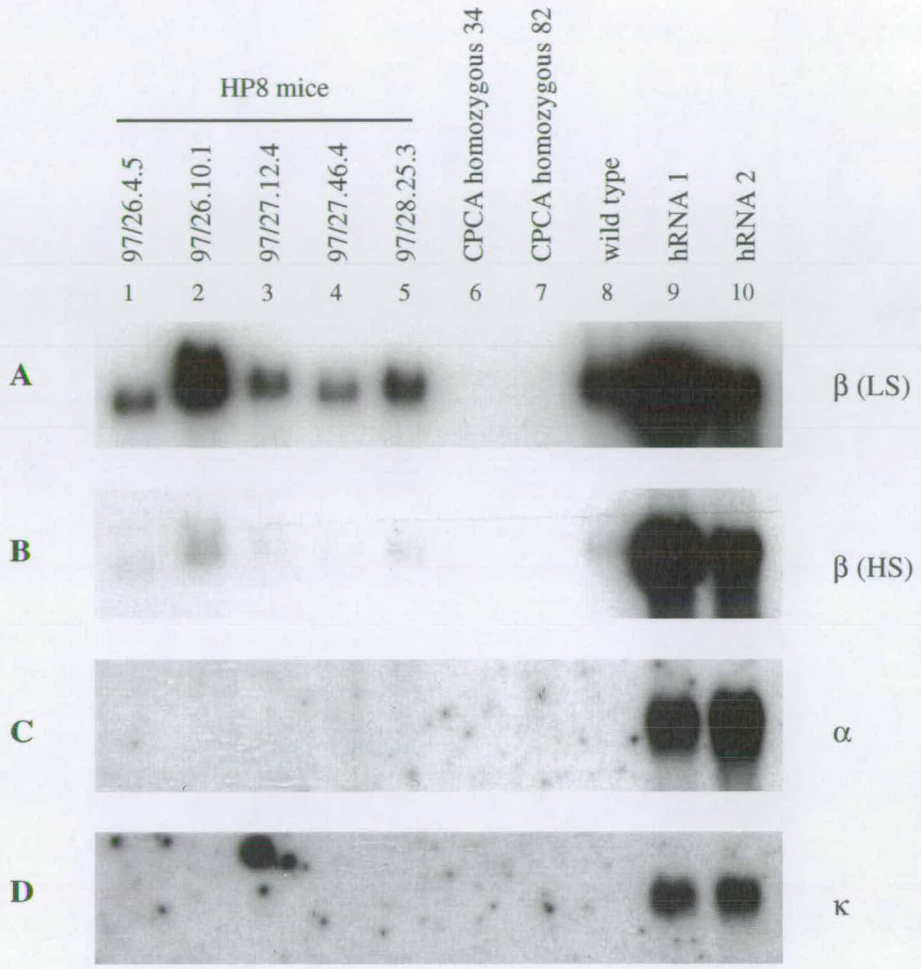


Figure 6.6: Northern blot analysis of human casein transgene expression

Total RNA from HP8 transgenic mice was analysed for human casein expression by northern blot. Lactating mammary gland total RNA was prepared from five lines of HP8 transgenic mice (lanes 1-5), two lines of homozygous β casein knockout mice (lanes 6 and 7) and a wild-type mouse (lane 8). RNA was also prepared from the cell pellet of two samples of human breast milk (lanes 9 and 10). A human β casein cDNA probe (hβcas1) strongly hybridised to the human control RNA and to transgenic and control mouse RNA under low stringency (Panel A). A high stringency wash of the same filter removed nearly all of the mouse signal (Panel B). No hybridisation of a human α casein oligonucleotide probe (HACP2) to transgenic samples was observed (Panel C). Similarly, no hybridisation of the transgenic samples to a human κ casein cDNA probe was observed (Panel D).

or 10μg of total RNA was used to assay for the presence of human casein RNAs by northern blotting.

Five lines produced lactating mammary tissue in time to be assayed for human casein gene expression (Figure 6.6). Total lactating mammary gland RNA from these lines was northern blotted and hybridised to hβcas1, which is a human β casein cDNA fragment from translation start to polyadenylation signal and was a kind gift from M. Dalrymple (Figure 6.2). A sample of lactating mammary gland total RNA from a wild-type mouse was included as a negative control. To control for cross-hybridisation to mouse β casein, two samples of lactating mammary gland total RNA from homozygous β casein knockout mice (a kind gift from A. Kind) was included. In addition, two samples of total RNA prepared from human milk were included. HRNA2 was a kind gift from M. Dalrymple. HRNA1 was prepared from human breast milk kindly donated by J. White. The milk was briefly centrifuged to pellet cells and total RNA was extracted from the pellet.

As can be seen in Figure 6.6a, the human β casein cDNA probe hβcas1 hybridised strongly to both human RNA samples, although hRNA2 is clearly degraded. There is also some signal from the transgenic samples and from the wild-type mouse RNA. The relatively intense signal from the 97/26.10 sample correlated with an excessive RNA load onto the gel as shown by a later probing with murine κcasein. No hybridisation was detected to the knockout samples. This result suggests that the probe has cross-hybridised to an extent to the mouse βcasein RNA. Stringent washing of this filter removed the majority of the signal from the transgenic and wild-type samples, while the signal from the human RNA samples was undiminished (Figure 6.6b). This argues that the signal emanating from the experimental samples is due to cross-hybridisation of the mouse β casein RNA to the probe, rather than the presence of human β casein RNA in these samples. A rough examination of this figure would suggest that the signal from the HP8 mouse samples is approximately 3 orders of magnitude weaker than that from the human controls. Therefore the transgenic mice do not appear to express significant quantities of human β casein.

Human α casein expression was assayed for by Northern blotting as above, except that in this case the probe was an end-labelled oligonucleotide, HACP2 (Table 6.1). As can be seen from

Figure 6.6c, no human αcasein RNA was detected in samples from five transgenic lines. However, the sensitivity of this Northern blot was not high. This remained the case in a repeat experiment. Therefore an RTPCR assay was performed (below).

As before, total lactating mammary gland RNA was analysed by Northern blot to assay for human κ casein expression. In this case the probe used was *hkcas1,* a human κcasein cDNA fragment from the translation start site to the polyadenylation site, and which was kindly donated by M. Dalrymple (Figure 6.2). No κcasein RNA was detected (Figure 6.6d). However, once more the sensitivity of the assay was insufficient to rule out very low levels of κcasein expression. A longer exposure of the same blot failed to show any bands but the background was fairly high in this case and may have obscured otherwise detectable κ casein hybridisation. κ casein expression was therefore analysed by RTPCR along with α casein expression.

6.7.2 Analysis by RTPCR

RTPCR was performed to detect human α and κcasein RNA (Figure 6.7). The RTPCR primers and expected product sizes are shown in Table 6.4. In addition to human αand κcasein primers, primers designed to amplify either mouse or human bile salt-stimulated lipase (BSSL; Lidmer et al. 1995; Genbank accession X54457) RNA were also included. BSSL is a milk protein and so this amplification served as a positive control of the first strand synthesis reaction. A water blank and RNA from a wild-type mouse were included as negative controls. Human breast milk RNA was included as a positive control. A 100-fold dilution of the normal human first strand aliquot was also amplified to test the sensitivity of the assay. A band was obtained from the human samples in all cases, indicating that the assay would detect RNA at least down to 1% of the normal level in humans. No human α casein RNA was detected in the mouse samples (Figure 6.7).

In contrast, an RTPCR product of the expected size was amplified from four out of five of the HP8 mouse samples by the human κ casein primers (Figure 6.7). To confirm the identity of the product these RTPCR products were gel-purified, cloned into pGemT and sequenced using an ABI377 automated DNA sequencer (Perkin Elmer). The sequences were aligned using MegAlign software (Lasergene). The alignment was almost perfect although with some apparently PCR-generated mismatches (Figure 6.8). The 97/28.25.3 sequence contains an A at

Figure 6.7: RTPCR on HP8 mammary gland total RNA

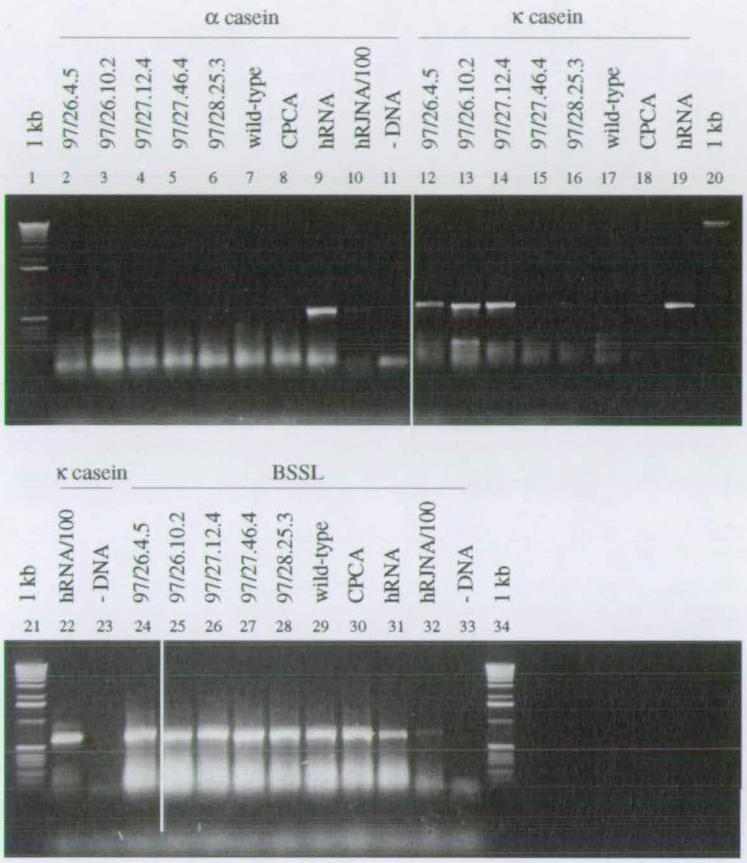


Figure 6.7: RTPCR on HP8 mammary gland total RNA

RTPCR was performed on lactating mammary gland total RNA from five HP8-transgenic lines (lanes 2-6, 12-16, 24-28), one wild-type mouse line (lanes 7, 17, 29) and one homozygous β casein knockout mouse line (lanes 8, 18, 30), and on human breast milk total RNA (lanes 9,10, 19, 22, 31, 32). First strand synthesis was carried out and 5% of this reaction was used in the subsequent PCR. A PCR reaction with 0.05% of the human RNA first strand reaction was also performed to test the sensititvity of the PCR (lanes 10, 22 and 32). A water blank was included (lanes 11, 23, 33). Primers are given in Table 6.4 and were designed to amplify human α casein, human κ casein or human and mouse bile salt-stimulated lipase (BSSL) products. The BSSL reaction serves as a positive control of the first strand

Figure 6.8 - RTPCR product sequences

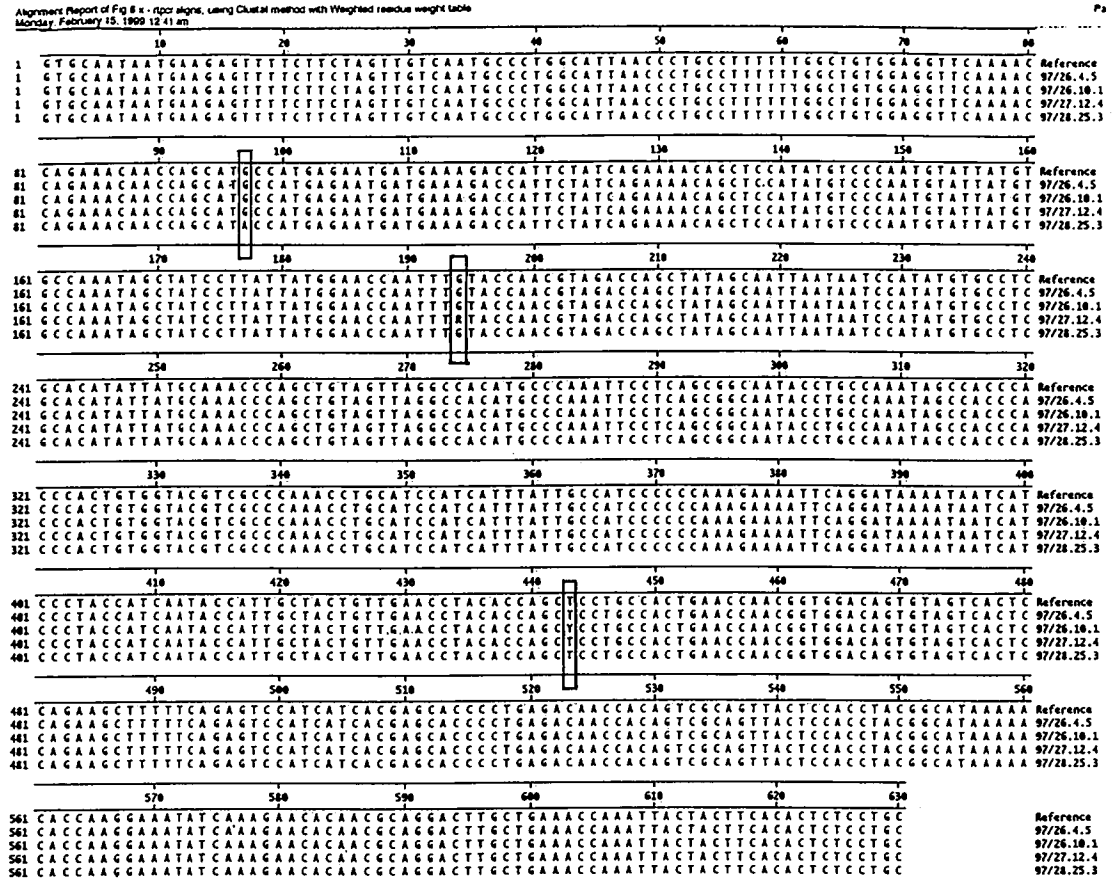


Figure 6.8 - RTPCR product sequences

κ casein RTPCR products (Figure 6.7) were cloned and sequenced and the sequences aligned to the published sequence of the human κcasein cDNA ("Reference"). The alignment software was MegAlign (Lasergene). The alignment is perfect except for three apparent amplification errors (boxed).

0position 97 in Figure 6.8. Only one RTPCR product was cloned from this animal so it is impossible to definitively ascribe this to PCR-introduced error, but it seems highly likely. G/A substitutions are the most common misincorporations introduced by Taq polymerase. The 97/27.12.4 product contains an R at position 194 in Figure 6.8. Out of three clones sequenced, two contained the expected G at this position and one contained an A. Thus this is a misincorporation. Similarly, 97/26.10.1 shows á Y at position 443 in Figure 6.8. Out of three clones sequenced, one contained a C at this position and the other two contained the expected T. Again, this is therefore a misincorporation. The human κ casein RTPCR products were also sequenced and were found to be genuine products (not shown).

Target RNA	**Predicted product**	**Primers**
Human α casein	558 bp	HACP1: ATGAGGCTTCTCATTCTCACCTG
		HACP2: TCACCACTGTAGCATGACGTTAT
Human κ casein	623 bp	HKCP1: GTGCAATAATGAAGAGTTTTCTT
		HKCP2: CCAGGAGAGTGTGAAGTAGTAAT
Human/Mouse BSSL	701 bp	BSSL1: TCTTCGGGGAGTCTGCTGGAGGTGCCAGC
		BSSL2: CCGAGAGGGATGGGAAAACAGGTAG

Table 6.4: RTPCR primers

The upper (forwards) primer is numbered 1 in each case. BSSL: bile salt-stimulated lipase. BSSL1 and BSSL2 anneal to both human and mouse BSSL transcripts. They were used as a positive RTPCR control, indicating that the RTPCR was successful. No primer pair anneal within the same exon.

Thus, the cloned RTPCR products from all four lines matched the published sequence for the human κ casein cDNA (Figure 6.8), indicating that they are all genuine amplified transcripts and therefore that human κcasein RNA is present in the lactating mammary glands of these mice. The sample which failed to give a band was from 97/27.46.4. This mouse apparently has one copy of the human β casein gene but has no right YAC arm (Figure 6.5). Since the human κ casein gene lies near the right arm of HP8, it is conceivable that this animal lacks an intact copy of the gene. To test this, a Southern blot was probed for each of the three human casein

genes, using hβcas8 and the cDNAs of human α and κ casein.

6.7.3 Southern analysis of G1 mice

HP8-transgenic mouse tail biopsy genomic DNA was digested with EcoRI and analysed for the presence of the three human casein genes by Southern blot. Hβcas8 was used to detect the human β casein gene. The human α and κ casein genes were detected with the cDNA probes hαcas1 and hκcas1 (Table 6.1; Figure 6.2).

All five transgenic mouse samples hybridised to the human β casein probe (Figure 6.9), confirming the presence of the gene. However, there was some variation in the apparent size of the detected EcoRI fragment. This variation has not been observed before (*c.f.* Figure 6.5), and so might be an artefact of this gel. The relative intensities of the β casein bands in this blot are in agreement with those in Figure 6.5 and serve to confirm the copy number estimates of these lines.

The same filter was stripped and reprobed with the human α casein cDNA probe hαcas1, corresponding to the published sequence (Genbank accession X78416) from translation start to polyadenylation site (a kind gift from M. Dalrymple). All five samples hybridised to α casein, and similar mobility differences to those observed with the β casein probe were seen, indicating that this phenomenon probably is due to variation within the gel or samples rather than differences in the transgenes themselves. The major band detected by hαcas1 was of approximately 7 kb, and this was also detected in human genomic DNA on the same filter (Figure 6.10). In addition, the two high-copy samples, 97/26.4.5 and 97/28.25.3, also gave bands at 4.2 and 5 kb approximately. The structure of the human α casein gene has not yet been reported, but these extra bands presumably represent internal EcoRI sites within the gene which are spanned by the cDNA probe. The absence of these bands from the other HP8 mouse lanes and the human DNA lane is probably due to the lower abundance of the target sequence in these lanes such that minor signals are undetectable.

The filter was stripped again and hybridised to the hkcas1 probe corresponding to the human κ casein cDNA from translation start to polyadenylation (Figure 6.2). Four of the five samples hybridised to the κ casein probe and gave a band of 7.8 kb as expected (Figure 6.11). However,

Figure 6.9: Scoring HP8 G1 mice for the human β casein gene

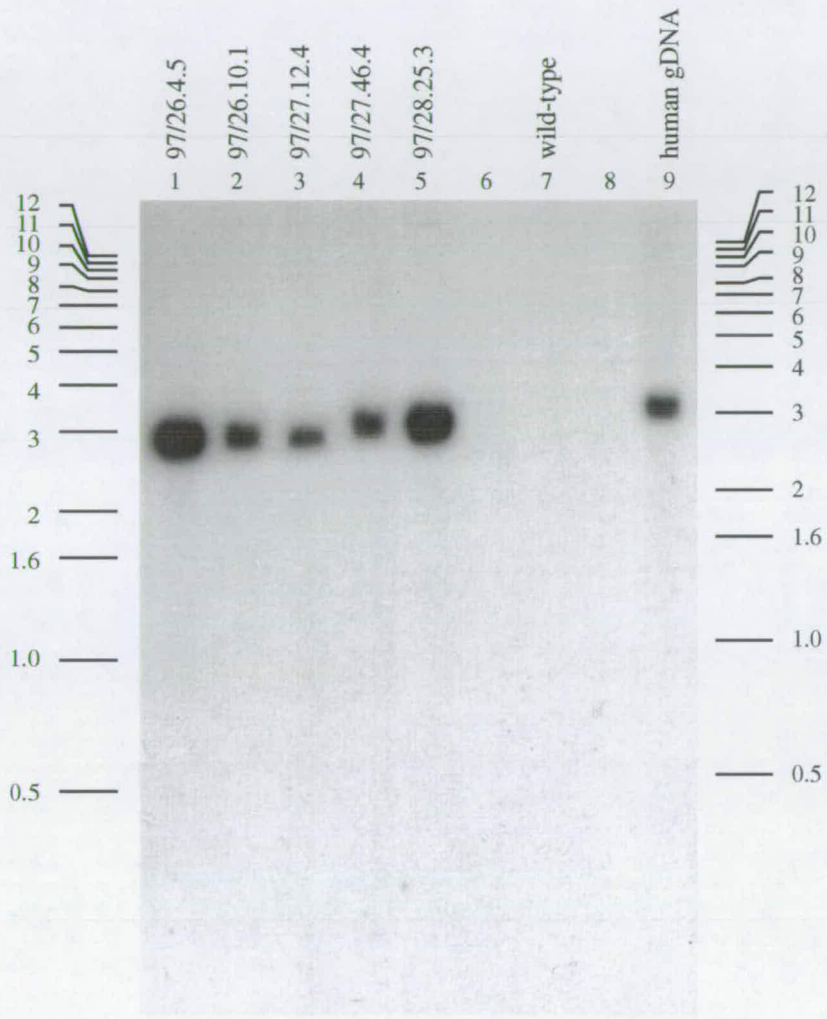


Figure 6.9: Scoring HP8 G1 mice for the human β casein gene

EcoRI restriction digestion was performed on genomic DNA from five lines of HP8 mice (lanes 1-5) and one wild-type mouse (lane 7), and on a sample of human placental DNA (lane 9). The restriction digests were separated on a 1% agarose gel, Southern blotted and hybridised to a human β casein intronic probe (hβcas8). All of the HP8 samples hybridised to the probe. The variation in the band sizes appears to be an artifact.

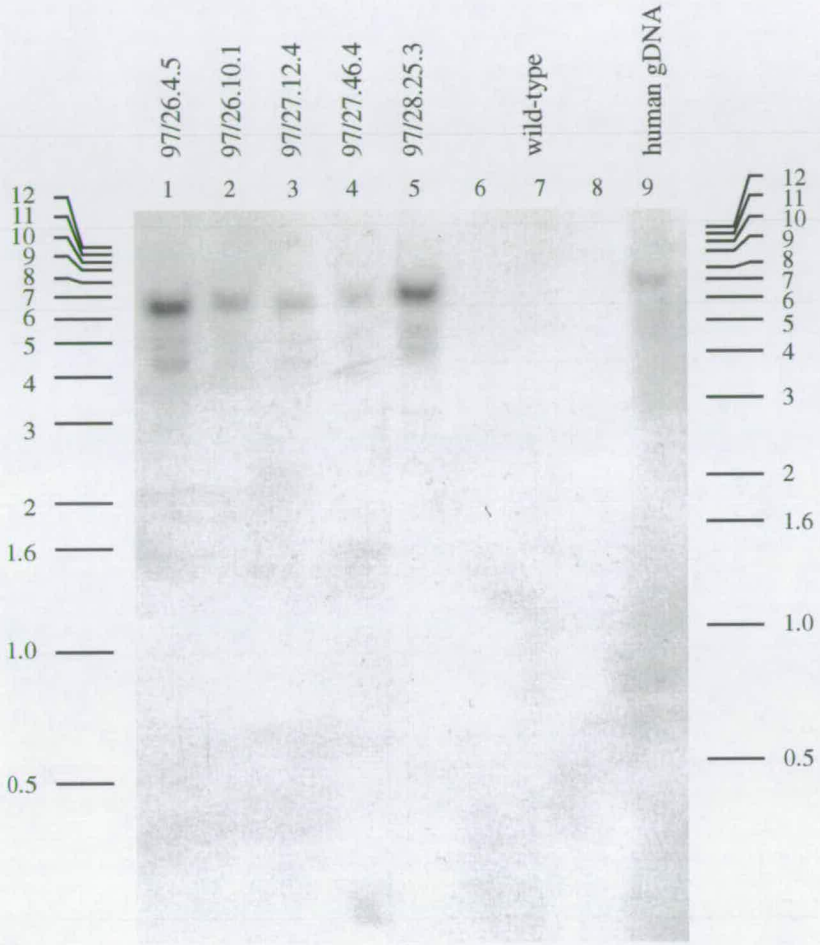


Figure 6.10: Scoring HP8 mice for the human α casein gene

The filter from Figure 6.x was stripped and rehybridised to a human α casein cDNA probe. All of the HP8 samples hybridised to the probe, giving a band of approximately 6.5 kb. Two additional bands are visible in lanes 1 and 5, which probably represent internal EcoRI sites which are spanned by the α casein cDNA probe.

Figure 6.11: Scoring HP8 mice for the human κ casein gene

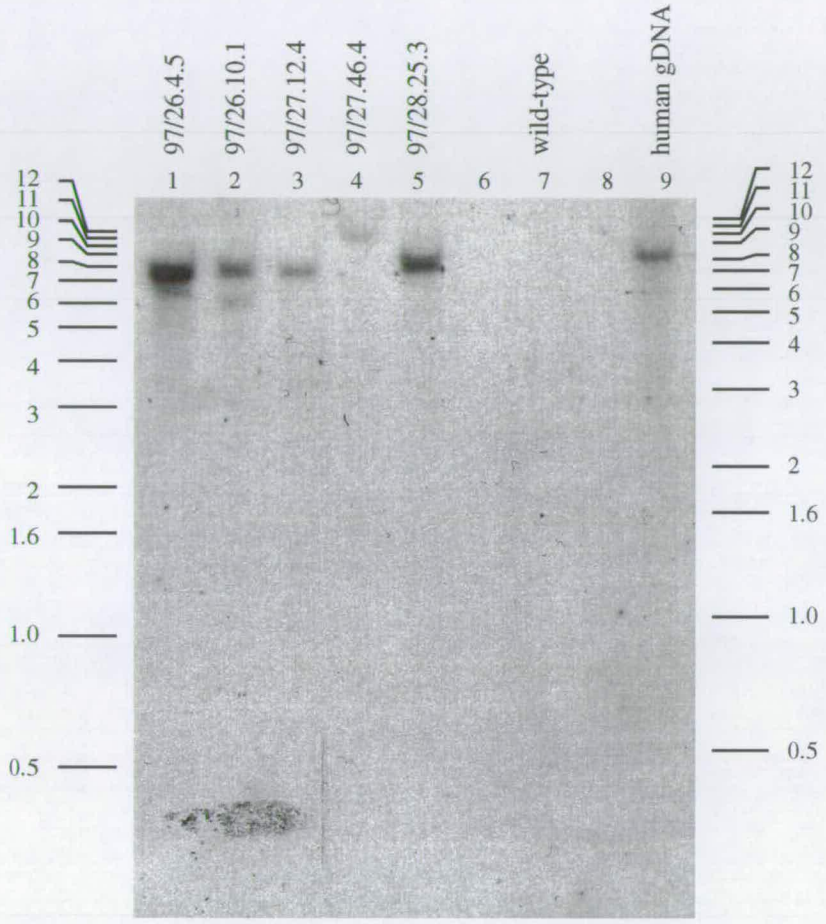


Figure 6.11: Scoring HP8 mice for the human κ casein gene

The filter from Figure 6.x was stripped and rehybridised to a human κ casein cDNA probe. The expected band is 7.8 kb. 97/27.46.4, which failed to express detectable human κ casein RNA, gave an aberrant EcoRI restriction fragment (lane 4). In lane 2 (97/26.10.1) both expected and aberrant bands are visible.

in the sample from 97/27.46.4, which did not give detectable human κ casein expression, the κ casein probe detected a band of approximately 10 kb. The simplest explanation for this is that the human κ casein coding region has become perturbed in this animal, and that the 10 kb band spans from the internal EcoRI site at position 1870 within the gene (Genbank accession U51899; Figure 6.2) to the next EcoRI site within the transgenic locus. This could be within the YAC or within flanking mouse DNA. In addition to the expected 7.8 kb band, 97/26.4.5 gave a very faint band of approximately 7 kb. This may reflect a deleted copy of human κ casein in this animal. Similarly, the κcasein probe hybridised to an additional band from 97/26.10.1, at approximately 6.5 kb. This also presumably indicates the presence of a deleted copy of the gene in the 97/26.10 line, which is approximately 2-copy.

6.8 Discussion

In this chapter a YAC carrying the human casein gene locus was introduced into transgenic mice by pronuclear microinjection. Subsequent analysis of these mice indicated that there was no significant expression of the human casein genes. The northern blot data for human β casein expression convincingly demonstrate that any expression of the transgenic β casein is approximately three orders of magnitude below that of β casein in the human lactating mammary gland. Similarly, expression of human α and κ casein in the HP8 mice appears to be less than 1% of that in the human. However, since the sensitivity of the northern blots was less for α and κ casein than for β casein, the expression of α and κ casein was also investigated by RTPCR. No expression of human α casein was detected, but in four of five animals some κcasein expression was detected. RTPCR can detect extremely rare transcripts, and indeed the positive human RNA controls suggested that the κ casein RTPCR was more sensitive than the α casein RTPCR. Further work would attempt to quantify this expression and to evaluate more fully the sensitivity of the RTPCR. At present, it can be concluded from the northern blot data that human α and κ casein RNA abundance is at least 100 fold lower in the HP8 mice than in humans. There is some κcasein expression in four of the five transgenic lines, but quantification is impossible without further work.

Thus expression from the YAC transgenes is negligible in these animals. There appear to be at least two possible explanations for this. The YAC might be perturbed in all transgenic mice to

the extent that no intact casein gene loci remain. Alternatively, sequences which are essential for human casein gene expression might be absent from the HP8 YAC.

The generation of transgenic mice from YACs is a relatively new technique, having first been reported in 1993. At the time of this work, initial reports of YAC transgenes suggested that the YAC could be stably maintained and that the functional yeast sequences had no detrimental effect on stability in mitosis or meiosis. Furthermore, expression of previously intractable genes was possible from YAC transgenes, demonstrating the power of the method of introducing a gene embedded within its normal chromatin environment on a large construct. However, a growing body of literature suggests that integrated YAC copies are rarely intact. The work reported here would seem to offer some support for this view. Five transgenic lines were examined with three gene-specific probes and a probe specific for each arm. Only high-copy animals possessed all markers. In addition, the ends of the YAC copies were not joined in a simple tandem array. Indeed, the data obtained by probing for junction copies within the YAC arrays are very confused. The data are most clear in low or single-copy animals and here none appear to carry all three markers, indicating that the YACs are perturbed. However, there are currently very few data regarding the extent of this perturbation. Transgene arrays in general rarely demonstrate perfect tandem repeats (*e.g.* Clark *et al.* 1992, various unpublished work at PPL Therapeutics); in particular the ends of the transgene copies are frequently deleted to a small extent. Therefore, it might not be significant in terms of the overall structure of the YAC copies that the transgene-transgene junctions are not as expected (Figure 6.5). Evidence that the YAC copies might be intact is similarly inadequate, however. The low-level expression of the κ casein gene indicates that at this gene is present, although important control sequences appear not to be. Further evidence that the YAC is intact in these animals comes from the Southern blots probed with the human casein gene probes (Figures 6.9, 6.10 and 6.11), but the fragments detected are small and internal to the three genes. Nevertheless, this is evidence that at least some internal YAC structure remains in these animals, although in some cases aberrant bands are also apparent (Figure 6.11). More detailed Southern blots, especially incorporating PFGE to resolve large fragments, would reveal more concerning the structure of the transgene array in these animals.

There are several means by which the HP8 transgenes may have become rearranged. YAC DNA

is extremely sensitive to shear, and though polyamines successfully protect large DNA *in vitro,* some shear must still occur. Data referring to the fate of large DNA upon passage through a microinjection needle is absent from the literature. In addition, the small volume of DNA solution injected into an egg would mean that attached polyamines would very rapidly be diluted out. Simple traffic through the cell might then cause shear of the DNA. Several hundred kb of naked DNA would represent a large, unusually fragile macromolecule even within the cell, and uptake into the nucleus for example might be expected to involve significant shear forces. Alternatively, enzymatic action might cause perturbations in YAC sequences. YAC DNA, by its very size, can usually be expected to carry multiple repeat sequences such as Alu sequences. Recombination between these sequences would fragment the YAC within the cell. Since YACs are so much larger than previously introduced transgenes, the risk of inter- and intramolecular recombination within the YAC copies must be correspondingly greater. Finally, although the mechanism by which exogenous DNA is integrated into the genome of the zygote is not understood, it is conceivable that the integration of very large pieces of DNA into a single locus might carry with it an increased risk of breakage and repair at that locus.

In this project there is a further possibility, however. Although in general few problems were experienced in preliminary YAC preparations, the batches of YAC DNA which were used to generate the mouse lines examined here were not tested by PFGE, due to time constraints and the amount of material available. Thus the possibility cannot be ruled out that the YAC was fragmented before it was microinjected. Microinjection characteristics of this DNA did not differ markedly from those of other YAC preparations in terms of needle blockage and viscosity of material (Y. Gibson, personal communication). Nevertheless, it is impossible to state with certainty that this YAC DNA was injected in an intact state.

Thus there are several means by which the HP8 YAC transgenes might become rearranged relative to the YAC clone, and this might account for the negligible levels of transgene expression in the mice. However, another possible explanation for the expression data reported here is that the YAC is essentially intact in the transgenic mice, but does not contain sequences which are essential for human casein expression. Although there is evidence of some rearrangement of the YAC in the transgenic mice, the evidence is inconclusive and where it directly relates to the human casein genes themselves, the majority of the fragments detected are

intact (Figures 6.9, 6.10 and 6.11). It is therefore possible that the YAC copies are substantially intact but lack vital human sequences. It is also possible that the overall structure of the locus is important for human casein expression, such that even large fragments might not express if rearranged relative to each other. However, although there are no published reports of human casein transgenes, the casein genes of other species are able to direct detectable expression in small constructs, albeit at varying levels. As discussed in chapter 7, casein promoters have been used to drive expression of both cognate and heterologous transgenes, though expression has generally been highly position-dependent and lower than that of the endogenous casein genes. This has led to the assumption that there are elements involved in controlling casein gene expression which have not been included in the truncated constructs employed to date.

Rijnkels and colleagues analysed the expression of the bovine casein genes in transgenic mice (Rijnkels *et al.* 1995, 1997d) using cosmids. The β casein construct contained 16 kb of 5' and 8 kb of 3' flank, the α_{s2} construct contained 8kb of 5' and 1.5 kb of 3' flank, while the κ gene was flanked by 5 kb of 5' and 19 kb of 3' genomic DNA. Two α_{s1} constructs were used, with 10 kb of 3' flanking DNA and either 5.4 or 14.2 kb of 5' flank. Expression of the larger α_{s1} construct was assayed at up to 160% of the endogenous bovine gene at the mRNA level, although the smaller construct expressed at less than 2% of endogenous bovine mRNA. β casein gene expression was detected at between 0.5 and 40% of endogenous bovine levels at the RNA level, while the α_{s2} and κ constructs failed to give detectable levels of expression.

Comparing this work with the results reported in this thesis, if the assumption is made that the human caseins are similar in structure and control to the bovine genes then it is surprising that no significant expression was detected in the HP8 mice. The possibility that the HP8 mice harbour no human α or β casein gene fragments as large as the bovine cosmids used by Rijnkels *et al.* (1997d) seems unlikely, especially given that the expected restriction fragments were detected by the internal gene probes (Figures 6.9 to 6.11). Moreover, the detection of low level human κ casein expression clearly indicates that intact κcasein genes are present in these animals. Indeed, the fact that the bovine κ casein cosmid employed by Rijnkels and co-workers failed to express might indicate that quite large gene fragments are present in the HP8 mice, though differences between the human and bovine genes might also account for this result.

Thus the failure of the HP8 mice to demonstrate any significant human casein gene expression remains unexplained. The possibility that the YAC transgenes are perturbed to the extent that no sizeable casein gene fragments remain seems small, and yet there is negligible transgene expression. In the light of the existing data regarding casein transgenes from other species, this is difficult to account for. However, in the absence of published data concerning the human casein locus, the possibility that the human locus is inherently refractory to expression in other systems cannot be excluded. In support of this, Rijnkels and co-workers (M. Rijnkels, personal communication) have attempted to express the human caseins in HC11 cells (Ball *et al.* 1988), using the same YAC clone, HP8, as that used here to make transgenic mice. Although the experiment was preliminary, no human casein expression was detected in HC11 cells. This is despite the finding (Rijnkels *et al.* 1995) that HC11 cells express casein constructs which are not expressed by transgenic mice, indicating that HC11 cells might represent a more permissive casein expression system. If this result is repeatable, it indicates either that human casein expression is not readily achievable in other systems, or, more likely, that HP8 lacks important sequences. Whether this is because HP8 does not contain an unperturbed genomic region or because HP8 does not span sufficient human DNA could be tested by the use of other YAC clones.

Further work with these mice would involve quantitative RTPCR and a more detailed analysis of the structure of the transgenic arrays. Although technically difficult, this could yield useful data and might offer an explanation for the expression data. Peterson *et al* (1995) were able to capitalise on spontaneously-deleted YAC transgenes to study the β globin locus in transgenic mice. They isolated a set of transgenic animals in which the YAC had fortuitously become deleted such that it mimicked disease genotypes and studied globin expression in these animals. It is possible that detailed analysis of the HP8 mice might yield some information about which regions of the transgenic casein locus are intact and which are not. If YAC perturbation is responsible for the lack of transgene expression, such data might be very informative in terms of studying human casein gene expression.

If YAC perturbation is the reason for the lack of expression in the HP8 mice, a natural progression might be to make more transgenic animals in order to obtain a line with intact YAC

copies. Two lines of animals generated here remain to be analysed; time constraints prevented detailed analysis of all lines. However, if none of the seven HP8 lines proves to harbour an intact YAC copy, then in terms of analysing casein gene expression this study will be fruitless. Since the generation of transgenic animals represents a significant investment of time and resources, some consideration of alternatives is merited here. Transfer of the YAC into cultured mammary cells is one option. HC11 cells express the endogenous murine βcasein gene following induction with prolactin and dexamethasone, and are easily transfected (Ball *et al.* 1988). Therefore HC11 cells are an attractive alternative to transgenic animals. At the outset the decision was made to pursue transgenic animals as a model system because the full range of developmental controls of casein gene expression are not accurately mimicked by cultured cells. This view is still valid but must now be weighed against the difficulty of obtaining cells or animals bearing intact YAC transgenes. In this context cell culture begins to look more attractive as multiple transfected lines are more easily and quickly established and screened. However, the selection which is necessary to isolate transfectants remains a problem as this might prejudice expression data by skewing the transfected population towards those which have integrated the YAC into open chromatin. The marker system of choice is *neo* (Southern and Berg, 1982). Very low levels of the *neo* gene product are required to provide resistance against the aminoglycoside G418 and thus even basal expression of the transgene should be sufficient for selection. This, coupled to the relative difficulty of obtaining enough transgenic animals to ensure a line with an intact YAC, argues in favour of cultured cells as testbeds for YAC expression. However, the control of mammary gland gene expression is highly complex and relies heavily on intercellular interactions *in vivo,* which are necessarily absent from the HC11 cell line. Neither do HC11 cells express all of the milk genes at appropriate levels in response to hormonal induction, further underlining the difference between HC11 cells and the mammary gland in terms of gene expression. More recently another mammary epithelial cell line has been derived, Kim-2 (Gordon *et al.* 1999), but the case here is much the same. Thus any study of the control of milk gene expression in these cells would be subject to significant inaccuracy as an indication of the likely behaviour of such genes *in vivo.* Nevertheless, the speed with which large numbers of transfected cell lines may be generated and characterised argues in favour of cultured mammary cells as a "first-pass" system for YAC transgene analysis.

Cell-mediated transgenesis is another option. YACs may be transferred into ES cells and thus give rise to transgenic mice (Jakobovits *et al.* 1993). As discussed in the introduction to this chapter, the DNA may be transfected into the ES cells or introduced by spheroblast fusion. Since the incorporation of additional DNA may prejudice this study, spheroblast fusion is not an attractive option. Transfection of purified YAC DNA would be preferable. Compared to microinjection, this method does not increase the chance of integrating an intact YAC. However, characterisation of the transgene array in transfectants prior to implantation would facilitate isolation of lines carrying intact YAC transgenes.

In the event that the YAC was found to be substantially intact in the HP8 transgenic mice, the alternative explanation, that HP8 does not contain all of the sequences necessary for human casein transgene expression, would be explored. Generation of transgenic mice or cell lines harbouring other YAC clones spanning the human casein locus would be a first step in this process. Detailed analysis of HP8 itself might also be informative, since current data indicate that it represents a genuine genomic region (M. Dalrymple and A. Tomlinson, unpublished; Rijnkels *et al.* 1997d). If a minor rearrangement of HP8 relative to the human genome has resulted in the low expression seen here, any data regarding the nature of the change would be highly informative.

7. Chapter 7 – Discussion

7.1 Casein gene expression

In this project an attempt was made to study the expression of the mammalian casein genes in the context of the intact locus on a YAC. The mammalian milk protein genes have long attracted attention as a model system for studying the developmental control of gene expression. The transition from virgin to lactating mammary gland is accompanied by considerable tissue remodelling and changes in gene expression in response to multiple endocrine signals. In particular, the milk proteins are largely undetectable in the virgin gland but are stimulated by two to three orders of magnitude over the course of pregnancy and lactation (Rosen 1987) before disappearing below the threshold of detection again at involution, which is accompanied by another round of tissue remodelling. Thus the mammary gland provides an opportunity to study profound developmental changes in gene expression in postnatal animals.

The caseins are the major milk proteins, making up approximately 80% of the protein in bovine milk. In mice β casein mRNA makes up approximately 50% of the mRNA in the lactating mammary gland (Mercier and Vilotte, 1993). This represents a considerable potential for gene expression. Since the mammalian mammary gland is a popular target for the expression of proteins in transgenic animals, there is considerable interest in the milk protein genes as vehicles for directing high levels of expression of heterologous proteins to the lactating mammary gland. As the most active transcription unit in the mouse lactating mammary gland, the β casein gene has been studied extensively.

In mice β casein mRNA is undetectable in the virgin mammary gland, reaching the threshold of detection by mid-pregnancy and increasing to a maximum at mid-lactation (Rijnkels *et al.* 1997a). The levels of casein protein have previously been shown to follow those of mRNA, and the changes in steady-state mRNA levels to result from changes in transcription rates, demonstrating that the primary control of casein gene expression is vested at the level of transcription (reviewed by Mercier and Vilotte, 1993), although there is also some evidence of post-transcriptional control (Eisenstein *et al.* 1988; Altiok and Groner, 1993, 1994).

Transcription of the calcium-sensitive casein genes is regulated by a synergistic combination of

peptide hormones, growth factors and the basement membrane (Topper and Freeman, 1980; Taketani and Oka, 1983; Bissel and Hall, 1987; Streuli *et al.* 1991; Roskelly *et al.* 1994; reviewed by Mercier and Vilotte, 1993). Work with milk protein gene promoters in transgenic and tissue culture systems has elucidated some of the molecular mechanisms of this regulation. A large number of deletion studies involving chimaeric constructs have resulted in a broad consensus concerning the structure of the calcium sensitive casein gene promoters. The MGF consensus binding site was identified for the so-called mammary gland factor (MGF; Schmitt-Ney *et al.* 1991; Watson *et al.* 1991; Wakao *et al.* 1992) which was later included in the cytokine-regulated transcription factor family and renamed stat5 (Gouilleux *et al.* 1994; Wakao *et al.* 1994). This binding site is essential for the induction of transcription in response to prolactin (Wakao *et al.* 1994), but this function also requires a second, more distal element (Pierre *et al.* 1994).

A highly conserved region in milk gene promoters is the "milk box" (Laird *et al.* 1988) which is located at −140 to −110 of the mouse βcasein gene. This region is necessary for hormonal induction of milk gene expression by an apparent relief of repression (Schmitt-Ney *et al.* 1991). By a combination of tissue culture and gel-retardation assays, it was determined that the ubiquitous transcription factor YYI (Shi *et al.* 1991) is involved in this repression, and that in response to lactogenic hormones YYI is released, leading to the formation of a activating transcription factor complex at the milk box, which has been termed the lactation associated complex (LAC; Raught *et al.* 1994).

The glucocorticoid response of the βcasein promoter appears to involve synthesis of a the C/EBP protein and binding of C/EBP to an overlapping MGF/C/EBP consensus binding site within the milk box (Raught *et al.* 1995). This region has been named the composite response element or CoRE and is conserved in casein promoters of all species (Raught *et al.* 1995) although the exact structure (overlap) of the composite elements varies between casein genes.

Thus the proximal promoter of the calcium-sensitive casein genes is highly conserved and contains CoRE regions which allow synergistic interaction of hormone signals to control the transcription of the genes during pregnancy and lactation. However, over the course of these studies, and in attempts to use casein promoters to drive heterologous protein production in

transgenic animals (see chapter 1), it has been noted that in general expression from all casein promoters has been low compared to the endogenous genes. Tissue and developmental-specificity of casein gene expression is conferred by the proximal (~1.5 kb) promoter, but high-level position-independent expression has not been achieved. This has led to the assumption that there are elements involved in controlling casein gene expression which have not been included in the truncated constructs employed to date. It has been shown that introns within the β casein gene enhance transcription (see chapter 1). Studies at PPL Therapeutics involving a BLG expression vector including intron 1 of the bovine β casein gene have shown a considerable increase in the expression and the frequency of high expression of linked transgenes (unpublished results). However, expression levels are still below those of the endogenous genes and position-dependent.

In an attempt to study the expression of casein genes in a more natural context than that employed in experiments involving chimaeric constructs and cultured cells, entire casein genes have been introduced into the germline of other species. Rijnkels and colleagues analysed the expression of the bovine casein genes in transgenic mice (Rijnkels *et al.* 1995, 1997d; see Section 6.8), and obtained significant expression (up to 160% of endogenous bovine at the mRNA level) from two of the four genes using relatively large constructs (30-40 kb) in cosmid vectors. However, expression was again highly variable and not related to copy number, although higher than previously reported for chimaeric casein gene constructs. Nevertheless, the expression remains position-dependent even in the presence of large genomic constructs.

In summary, casein promoters have been used with some success to direct cognate and heterologous gene expression to the transgenic mammary gland and in cultured cells. Correct developmental regulation of casein gene expression has usually been achieved, and seems to follow that of the source rather than host species. However, in transgenic animals expression is always position dependent, even with large genomic constructs. The simplest explanation for this is that there are functional elements at the endogenous casein gene locus that have not been included in individual gene constructs.

Since the regulatory elements responsible for high-level expression of the casein genes have so far eluded discovery via plasmid and cosmid constructs, a YAC-based strategy was indicated.

YAC inserts are commonly 400-2000 kb in size, and YACs of up to 800 kb have been introduced into transgenic mice. Therefore the decision was taken in this project to generate transgenic mice carrying an entire casein locus on a YAC. This approach has been fruitful in the discovery and study of long-range control elements in other loci (Schedl *et al.* 1993a; Montoliu *et al.* 1995; Peterson *et al.* 1997). To maximise the probability of introducing functional elements on the YAC, a mouse YAC was sought. It was envisaged that this would present problems in detecting expression from the transgenic YAC, since the transgenic and endogenous casein transcripts and proteins would be indistinguishable. The most subtle change to a YAC would have been to introduce a small mutation such that RNA could be detected against a murine background. However, a practical goal in this project was to enhance expression of heterologous constructs in the transgenic mammary gland. cDNAs are often introduced into transgenic animals owing to their ease of isolation and small size, or to the occasionally prohibitive size of the corresponding gene. However, cDNA expression is generally very low in transgenic mice, and is moreover subject to severe position effects and PEV. In the mammary gland some rescue of this expression is possible if the cDNA construct incorporates introns from another gene or is co-injected with the ovine BLG gene (Brinster *et al.* 1988; Clark *et al.* 1992), but in general cDNA expression is still very poor compared to that of genomic constructs. It was hoped that the active β casein locus would potentiate expression of an inserted cDNA or minigene construct, and for this reason the protein C/BLG minigene construct pCorp3 was chosen as a reporter to be used to modify the mouse casein YAC. In a parallel experiment pCorp3 was targeted to the endogenous β casein locus via ES cells (A. Kind, unpublished) which was anticipated as a control for the YAC experiment.

7.2 Mapping and manipulation of the murine casein YACs

The ICRF YAC library (Larin *et al.* 1991) was screened with two genomic mouse β casein probes, and after screening of 15 β casein YAC clones four were identified which also hybridised to cDNA probes for murine α, γ, ε and κ casein. Two of these, MP14 (380 kb) and MP12 (430 kb), were restriction mapped. This was important to assess the integrity of the YACs in these two clones as well as to characterise them before further work. YAC clones are often chimaeric or otherwise rearranged relative to the genome; comparison of a YAC clone to genomic DNA is therefore a common first step in characterising the YAC. However, a more

rapid method was devised based on the YAC restriction mapping protocol of Hamvas *et al.* (1994). It was reasoned that if two independent YAC clones overlap extensively, and within the region of overlap are collinear, the likelihood that either is rearranged must be very small. This approach was taken with MP12 and MP14 in Chapter 4. The two YACs were restriction mapped by partial restriction digest and hybridisation to YAC end probes. The YACs overlap in a central region of approximately 360 kb, and within this region are collinear. The possibility of gross rearrangements within either YAC was thus discounted.

The casein genes were positioned on the map of these two YACs by a combination of partial and complete restriction digest. Where possible the same Southern blot was reprobed in an attempt to avoid gel-to-gel variation, which in PFGE can be considerable. However, the accumulation of noise on the filters reduced the quality of the data somewhat, although the map could still be generated with confidence. Further work would address the quality of these data by repeating some of the blots.

The bovine casein locus had been mapped at the start of this project (Ferretti *et al.* 1990; Threadgill and Womack, 1990). Since the mapping work described here (this thesis; Tomlinson *et al.* 1996) further reports of the human (Rijnkels *et al.* 1997b) mouse (George *et al.* 1997; Rijnkels *et al.* 1997a) and bovine (Rijnkels *et al.* 1997d) have been published. A comparison of these maps is made in chapter 6. In general the structure of the locus appears conserved across the taxa. However, in humans there are three casein genes, while cattle have four and mice have five. This is not indicative of a high level of conservation of the structure of the locus. Nevertheless, the basic spatial organisation of the genes is maintained, and no species has been found to harbour more than one κ–like casein gene. It has been suggested that the casein genes are under locus control (Alexander *et al.* 1988). This is based on the conserved structure of the locus but also on the linkage of κ casein. κcasein is structurally unrelated to the calcium-sensitive caseins of the α and β types. It appears that κ casein is related to γ fibrinogen, to which it is genetically linked in pigs. In other taxa the linkage with γfibrinogen is much reduced, while the linkage to the caseins is conserved. Thus κ casein appears to share a common ancestor with γ fibrinogen, but to have been recruited to the casein family. The physical proximity of the κ gene to the other casein genes lends credence to the notion of a casein LCR.

Once the murine casein YACs had been characterised and found to represent unrearranged regions of the mouse genome, MP14 was chosen for further work. As discussed in chapter 5, pCorp3 was chosen as a reporter gene to be inserted into MP14 in order to follow expression from the YAC in transgenic animals. This necessitated a succession of yeast manipulations. Unfortunately, these manipulations did not proceed smoothly. The MP14 YAC repeatedly rearranged when manipulated. In particular, targeting of the pCorp3 construct to the right arm of MP14 resulted in fragmentation of the YAC such that no full-length clones were recovered. This may have been due to a repeat element within pCorp3 recombining within the YAC. A second attempt to insert pCorp3 into the YAC was aborted due to lack of time. This second attempt was based on a two-step gene targeting strategy, which was conceptually flawed, resulting in delays to the project. In retrospect a pop in-pop out (PIPO) strategy would be far more efficient. This highlights what must be considered a recurrent lesson from this project, which is that work with YACs is technically demanding and requires a background of yeast expertise. At the start of this project YACs were chosen as the large cloning system of choice, and it was felt that their ease of manipulation, particularly by homologous recombination, was a great advantage. This is certainly true, but the corollary is that YAC clones tend to recombine spontaneously and in unexpected ways, necessitating vigilance and care in their maintenance and manipulation. The acquisition of the knowledge and skills required for effective work with YACs took longer than anticipated. Despite this, MP14 was modified to carry *HIS3* instead of *URA3* on the right arm of the YAC following transfer into a *his3* host strain by classical mating and random sporulation. The *URA3* counterselectable marker was then removed from the right arm of the YAC (replaced by *HIS3*) and reintroduced 3' to the polyadenylation site of the β casein gene in MP14. It was at this point that the second targeting step, replacement of *URA3* with pCorp3, foundered. Had this targeting been achieved, MP14 would have been microinjected to generate transgenic mice, and also transfected into a murine mammary epithelial cell line, to study expression of the reporter construct from within the YAC.

7.3 Human casein YAC transgenic mice

Since the work with the murine casein YACs did not allow generation of transgenic animals within the period of this project, an alternative was sought. The human casein YAC HP8 was kindly provided by M. Dalrymple. This clone had been obtained from the ICRF human YAC

library (clone number ICRFy901C0518) and had been shown to represent an unrearranged portion of the human genome (Rijnkels *et al.* 1997b; M. Dalrymple and A. Tomlinson, unpublished). This YAC was therefore gel-purified and microinjected into fertilised mouse oocytes. By screening with either a *TRP1* or β casein probe 9 transgenic founders were obtained, of which 7 transmitted the YAC markers in the germline. Five lines were analysed for human casein expression by northern blot. No expression was observed within the limits of the assay, which admittedly was only sensitive to about 1% of endogenous mouse casein expression. α and κcasein expression was then investigated by RTPCR, which is more sensitive. No α casein RNA was detected. In contrast, κ casein gene RNA was detected in 4 out of the 5 transgenic lines. A Southern blot revealed the κ casein gene in the non-expressing line to be rearranged.

The structure of the transgene array in the mouse lines was not studied in detail. A preliminary Southern blot (Figure 6.5) suggests the structure of the YAC junctions within the multi-copy arrays to be complex and not as expected. However, this experiment was not primarily intended to study the transgene junctions and the gel conditions were inappropriate for resolution of the relevant bands. Additionally, rearrangement and deletion at the ends of transgenes is common, and does not necessarily indicate perturbation elsewhere within the array. Nevertheless it is possible that the YAC was rearranged in these lines. YAC transgene fragmentation can result from a number of possible mechanisms which in this case must include fragmentation during preparation (see chapter 6). Due to pressures of time and availability of purified material, the YAC DNA which was used to generate the mice analysed in this thesis was not tested for integrity. In general, little fragmentation has been observed to occur during YAC preparation, and during injection this material exhibited characteristics of high molecular weight DNA. Nevertheless the integrity of this YAC preparation prior to injection was not defined, an omission which may have prejudiced the outcome of the experiment. It could also explain the rather low transgenic frequency observed, which could reflect the presence of other lines carrying different YAC portions from those assayed for.

No significant expression of the human casein genes was observed in these mice. Although κ casein mRNA was detected by RTPCR the northern blot suggests that this is some two orders of magnitude less abundant than the endogenous murine κ casein mRNA. One explanation for

this would be fragmentation of the YAC. However, significant levels of casein gene expression have been obtained by others with small constructs, and the κ genes at least are intact in four out of five of the HP8 mouse lines. There is only one report of a human casein transgenes, and this describes the expression of human βcasein in transgenic potato plants under the control of a mannopine synthase promoter (Chong *et al.* 1997). However, the expression of the caseins of other species has been studied in cell culture and transgenic animals (Sections 6.8 and 7.1). In general, although expression has been low compared to the endogenous genes it has been detectable, and should have been within the limits of detection of the northern blots described in Figure 6.6, especially in the case of β casein. Thus casein gene expression has been seen from constructs of a few tens of kilobases. That disruption of the HP8 YAC DNA could have produced no calcium-sensitive casein gene fragments in excess of this seems unlikely. Other possible explanations for the expression data might include small changes in the YAC relative to genomic human DNA that are not detected by PFGE but disrupt gene expression. Alternatively the YAC might not contain sequences which are crucial to expression of the human caseins. Although from work with other casein genes this would be unexpected it is conceivable, and no human casein transgenes have been reported in animal systems. As described in Section 6.8, Rijnkels and co-workers have introduced HP8 into HC11 cells and could detect no expression, but this experiment was preliminary and has not been repeated.

7.4 Further work

The mouse lines generated in this project do not express significant levels of human casein mRNA. In light of this further work with these lines would be confined to analysing the transgene arrays and quantifying the low levels of human casein mRNA which were observed.

The RTPCR performed here was not quantitative. A quantitative or semi-quantitative RTPCR would yield more informative data concerning the level of human κ casein mRNA in these mice. Truly quantitative RTPCR is difficult to achieve without access to specialised equipment, but semi-quantitative RTPCR merely requires careful optimisation and control. Although the levels of human κcasein RNA detected here were certainly below 1% of that of the endogenous murine mRNA, they may still be significant and amenable to quantification relative to copy number. In addition, the expression level of the β and α casein genes will be reinvestigated by

RTPCR, with some quantification as to the sensitivity of the assay.

Further restriction analysis of the HP8 transgenic array is also warranted. Initially the junctions of the YAC copies would be examined in more detail to clarify the results presented in Chapter 6. PFGE and Southern blotting could yield further information about the structure of the array as a whole. However, interpretation of such data is problematic. It should also be feasible in principle to observe the transgene array directly by extended fibre FISH, which is able to resolve to 10 kb or more. In combination with standard FISH of interphase nuclei this might be expected to yield detailed information about the structure of the YAC copies in these animals. However, the considerable cost and effort involved in this type of work would render it inappropriate for study of these mice. A more reasonable approach might be to simply generate additional lines of mice from intact purified HP8 YAC DNA. In addition, the introduction of rare-cleavage restriction enzyme sites into the ends of the YAC (Peterson *et al.* 1995) would allow testing for intact YAC copies by PFGE and Southern blotting. This procedure seems to facilitate YAC transgenic experiments to the point where its inclusion in future work would be a certainty.

Given the problems associated with YAC transgenes, it might also be prudent to examine alternative systems in which to characterise the expression of the caseins on a YAC. In general, work with cultured cells provides an alternative or complement to work with transgenic animals. However, as discussed in chapter 1 and above, the behaviour of cultured cells does not accurately mimic that of the source tissue, particularly in cases where intercellular interactions are important factors *in vivo; in vitro* these interactions are necessarily disrupted. HC11 cells were derived from less readily-cultured mammary cell lines by selection for cells which could grow without interaction with extracellular substrata or other cell types (Ball *et al.* 1988). Further to this, the observation by Rijnkels *et al.* (1995) that bovine α_{s2} and κcasein cosmid constructs were expressed in HC11 cells but not in transgenic animals suggests that some levels of casein gene control are not active in cultured mammary cells.

These results and prior observations of the discrepancies between cultured cells and transgenic animals in the behaviour of specific constructs (notably cDNAs) have recently been added to by suggestions that developmentally-regulated chromatin structures are not faithfully reproduced in

differentiated cell lines. Skarpidi *et al.* (1998) report the transfection of a γ globin gene linked to the β globin miniconstruct (the so-called micro LCR, μLCR) into MEL cells. This construct had previously been expressed position-independently in transgenic mice (Stamatoyannopoulis *et al.* 1993, 1997). Although pools containing large numbers of clones exhibited average gene expression patterns similar to those of transgenic mice, they saw no correlation between copy number and globin gene expression when small pools or individual clones were examined.

Peterson *et al.* (1993a) introduced a 248 kb βglobin YAC into MEL cells. To overcome technical difficulties experienced when attempting to directly transfer the YAC into MEL cells, it was first transferred into non-committed L cells (Gnirke *et al.* 1993) and then these cells were fused with MEL cells (Peterson *et al.* 1993a). Once in MEL cells the globin genes were faithfully expressed independent of the position of integration, and showed a developmental switch from the foetal genes expressed in the L cells to adult genes in the MEL/L hybrids. In contrast, a subsequent experiment which successfully transfected a 155 kb βglobin YAC into MEL cells directly resulted in widely varying globin gene expression levels per YAC copy and a random pattern of activation of the genes over time (Vassilopoulos *et al.* 1999). However, when the 155 kb YAC was transfected into L cells as before and then introduced into MEL cells by cell fusion, expression was once again position independent and properly regulated over time. This result is striking in that it casts serious doubt over the validity of differentiated cells lines to study developmentally-controlled gene expression *per se.* It may be that the globin locus is unique in this requirement for an undifferentiated chromatin environment in order to develop normally. However, the globin locus is perhaps the most intensively studied in terms of gene expression, and it is conceivable that new insights gained here will be applicable at other loci. At the very least, this result and that of Skarpidi *et al.* (1998) underscore the caution with which results gained *in vitro* must be interpreted.

One considerable advantage of tissue culture is that it allows the relatively rapid generation of numerous lines harbouring a YAC, which may then be screened for an intact YAC copy. One way to combine this advantage with the undoubted superiority of transgenic animals as a model system would be to introduce a YAC into mice by cell-mediated transgenesis. In mice this is readily achievable via lipofection of ES cells (Choi *et al.* 1993; Lamb *et al.* 1993; Pearson and Choi, 1993; Strauss *et al.* 1993; Lee *et al.* 1996). If the YAC was modified to carry rare

restriction sites at each end (above), screening of transfected clones by restriction digest would allow mice to be generated only from those clones harbouring intact YAC copies.

The mapping data reported in this thesis are largely in agreement with those reported elsewhere (Rijnkels *et al.* 1997a; George *et al.* 1997). However, there are some discrepancies. Further work would be required to resolve these. In particular, probing partial restriction digests of MP14 and MP12 with the *RA* probe would be desirable to ask if any of the discrepancies between the maps are due to the small region of the *RA1* probe used here that hybridises to the left arm of the YAC. Additional digests would also aim to more accurately define the limits of the ε gene and confirm the orientation of the genes reported elsewhere.

With more time, the targeting of MP14 would be attempted again, but with a PIPO construct. A PIPO construct has already been made, but it did not prove possible to use it within the time allowed. The construct targets the reporter *eGFP1* (Clontech) to the mouse β casein locus using *URA3* as the selectable marker. The insertion of *eGFP1* into the β casein locus would facilitate the rapid study of gene expression from the YAC in the HC11 mammary epithelial cell line. However, given the limits of cell culture described above, a PIPO vector targeting pCorp3 or another suitable reporter would also be constructed, so that the original goal of introducing the murine casein gene locus into mice might be fulfilled. The use of pCorp3 was based on a parallel experiment targeting pCorp3 to the endogenous murine casein locus. This experiment successfully replaced the murine βcasein gene coding region with the pCorp3 construct. However, expression of the transgene was poor, leading to the detection of 17μg/ml of human protein C in the milk of these animals (A. Kind, unpublished results). Although this represents a two-fold improvement in per-copy expression level over the best seen in transgene rescue experiments (Table 5.1), it is considerably lower than the best total protein level achieved by rescue, which was approximately 57 μg/ml. Thus the high rate of transcription of the β casein gene is not necessarily acquired by a heterologous replacement in the β casein locus. Unfortunately the second targeting experiment which aimed to insert pCorp3 downstream of the β casein gene was unsuccessful. The attraction of replacing the β casein gene with pCorp3 in MP14 is diminished by these results. The insertion of pCorp3 downstream of the β casein gene might still yield interesting data, but a more definitive test of the YACs ability to behave as an intact locus would be to modify it as little as possible, so that any difference between the YAC

expression pattern and that of the endogenous murine locus could not be attributed to the structural changes made to the YAC. Point mutations allowing discrimination of the endogenous and transgenic mRNAs would be the most attractive prospect.

7.5 Summary

Two YACs harbouring all five murine casein genes were isolated from the ICRF genomic mouse YAC library. These two clones were mapped by restriction digest and Southern blot, generating the first reported physical map of the mouse casein locus. The locus covers 260 kb and contains the murine casein genes in the order α–β–γ–ε–κ. The β casein gene is transcribed towards the α gene. The map generated here broadly agrees with maps generated elsewhere (Rijnkels *et al.* 1997a; George *et al.* 1997) and shows the gross structure of the casein locus to be conserved across mammals.

The collinear maps of the two mouse casein YACs preclude the possibility of gross rearrangement within either. One, MP14, was manipulated to attempt to place a protein C reporter construct within the β casein locus. The YAC was transferred to a *his3* strain by mating with yPH500 (Sikorski and Hieter, 1989) and random sporulation. The *URA3* gene was removed from the right arm of the YAC and reintroduced 3' to the β casein gene as the first step in a two-step gene replacement strategy. However, the second step was unsuccessful, and was realised to be flawed. The selection against *URA3* in the second step would be expected to yield large numbers of spontaneous revertants such that those clones arising as a result of the desired replacement event would be difficult to isolate. Within the time allowed it did not prove possible to make any further progress with this strategy. A better strategy would have been to use the more efficient pop in-pop out (PIPO) method.

An attempt was also made to directly insert the protein C reporter into the right arm vector sequence of MP14. However, the transformation of the yeast cells with the targeting construct resulted in gross deletions within the YAC, probably due to recombination of the targeting construct with the YAC DNA at several internal sites which may be repeat sequences.

The human casein YAC HP8 had been previously characterised and shown to contain an

unrearranged region of the human genome including 400 kb of the human casein locus, containing the three human casein genes α, β and κ, in a similar organisation to the murine and bovine genes. This YAC was gel-purified and used to generate transgenic mice by pronuclear microinjection. Of five transgenic lines analysed, none expressed detectable human α or β casein mRNA. Four out of five lines expressed human κ casein mRNA which was detectable by RTPCR. The non-expressing line appeared to harbour a rearranged copy of the κ casein gene. Southern analysis suggested that the YAC transgene arrays might be rearranged, but was inconclusive. The expression of κ casein suggested that some significant YAC fragments are intact in the HP8 transgenic mice.

Further work would include more detailed analysis of the murine casein locus by Southern blot of conventional and pulsed field gels to resolve certain discrepancies between the three published maps of the locus. The manipulations of the mouse YAC would be repeated but by the PIPO method, and possibly to make a more subtle change to the locus, since the incorporation of the protein C reporter construct into the endogenous mouse β casein gene by gene targeting resulted in low expression levels (A. Kind, unpublished).

The transgenic mice will be more thoroughly characterised to attempt to gain information about the structure of the transgene array and ask if the lack of appreciable gene expression can be attributed to the fragmentation of the transgene in all lines. To continue the investigation of this locus in transgenic mice, more animal lines would be generated using a modified version of HP8 bearing rare restriction enzyme sites on each end, such that screening for animals with intact YAC copies would be simplified. The feasibility of cell-mediated transgenesis would be investigated, as screening of ES cells prior to their transfer into blastocysts would allow intact YAC transgenes to be identified before the generation of mice.

Bibliography

SCRIP (1996). First transgenic goat clinical trials. Scrip 2192(3): 23.

Alexander, L, Stewart, AF, MacKinlay, AG, Kapelinskaya, TV, Tkach, TM, and Gorodetsky, SI (1988). Isolation and characterization of the bovine κ-casein gene. Eur J Biochem 178: 395-401.

Ali, S, and Clark, AJ (1988). Characterisation of the gene encoding ovine β-lactoglobulin. J Mol Biol 199: 415-426.

Altiok, S, and Groner, B (1994). β-casein mRNA sequesters a single-stranded nucleic acid-binding protein which negatively regulates the beta-casein gene promoter. Mol Cell Biol 14: 6004-6012.

Altiok, S, and Groner, B (1993). Interaction of two sequence-specific single-stranded DNA-binding proteins with an essential region of the β-casein gene promotor is regulated by lactogenic hormones. Mol Cell Biol 13: 7203-7310.

Archibald, A, McClenaghan, M, Hornsey, V, Simons, JP, and Clark, AJ (1990). High-level expression of biologically active human alpha 1-antitrypsin in the milk of transgenic mice. Proc Natl Acad Sci U S A 87(13): 5178-5182.

Aronow, B, Silbiger, RN, Dusing, MR, Stock, JL, Yager, KL, Potter, SS, Hutton, JJ, and Wiginton, DA (1992). Functional analysis of the human adenosine deaminase gene thymic regulatory region and its ability to generate position-independent transgene expression. Mol Cell Biol 12(9): 4170-4185.

Ball, R, Friis, RR, Schoenenberger, CA, Doppler, W, and Groner, B (1988). Prolactin regulation of β-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. EMBO 7, 2089-2095.

Bayna, E and Rosen, JM (1990). Tissue-specific, high level expression of the rat whey acidic protein gene in transgenic mice. Nucleic Acids Res 18(10): 2977-2985.

Bean, T, and Ney, PA (1997). Multiple regions of p45 NF-E2 are required for β-globin gene expression in erythroid cells. Nuc Acids Res 25(12): 2509-2515.

Behringer RR, Ryan, TM, Palmiter RD, Brinster RL, Townes TM (1990). Human gamma- to beta-globin gene switching in transgenic mice. Genes Dev 4(3): 380-389.

Bender, M, Reik, A, Close, J, Telling, A, Epner, E, Fiering, S, Hardison, R, and Groudine, M (1998). Description and targeted deletion of 5' hypersensitive site 5 and 6 of the mouse beta-globin locus control region. Blood 92(11): 4394-4403.

Bentley, D, Todd, C, Collins, J, Holland, J, Dunham, I, Hassock, S, Bankier, A, and Giannelli, F (1992). The development and application of automated gridding for efficient screening of yeast and bacterial ordered libraries. Genomics 12(3): 534-541.

Berry, M, Grosveld, F, and Dillon, N (1992). A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. Nature 358(6386): 499-502.

Bijvoet, A, Kroos, M, Pieper, FR, Van der Vliet, M, De Boer, HA, Van der Ploeg, AT, Verbeet, MP, and Reuser, AJJ (1998). Recombinant human acid alpha-glucosidase: high level production in mouse milk, biochemical characteristics, correction of enzyme deficiency in GSDII KO mice. Hum Mol Genet 7(11): 1815-1824.

Bischoff, R, Degryse, E, Perraud, F, Dalemans, W, Ali-Hadji, D, Thepot, D, Devinoy, E, Houdebine, LM, and Pavirani, A (1992). A 17.6 kbp region located upstream of the rabbit WAP gene directs high level expression of a functional human protein variant in transgenic mouse milk. FEBS Lett 305(3): 265-268.

Bishop, J and Smith, P (1989). Mechanism of chromosomal integration of microinjected DNA. Mol Biol Med 6: 283-298.

Blackburn, E (1990). Telomeres: structure and synthesis. J Biol Chem 265: 5919-5921.

Blom van Assendelft, G, Hanscombe, O, Grosveld, F, and Greaves, DR (1989). The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. Cell 56: 969-977.

Boeke, J, LaCroute, F, and Fink, GR (1984). A positive selection system for mutants lacking orotidine-5'-phosphate decarboxylase in yeast: 5-fluoro-orotic acid resistance. Mol Gen Genet 197: 345-346.

Bonifer, C, Huber, MC, Bosch, FZ, and Sippel, AE (1994). Dissection of the locus control function located on the chicken lysozyme gene domain in transgenic mice. Nucl Acids Res 22, 4202-4210.

Bonifer, C, Vidal, M, Grosveld, F, and Sippel, AE (1990). Tissue specific and position independent expression of the complete gene domain for the chicken lysozyme in transgenic mice. EMBO J 9: 2843-2848.

Bonsing, J, Ring, JM, Stewart, AF, and Mackinlay, AG (1988). Complete nucleotide sequence of the bovine beta-casein gene. Aust J Biol Sci 41: 528-537.

Brandon, E, Idzerda, RL, and McKnight, GS (1995). Targeting the mouse genome: a compendium of knockouts. Curr Biol 5(8): 873-881.

Brem, G, Hartl, P, Besenfelde,r U, Wolf, E, Zinovieva, N, and Pfaller, R (1994). Expression of synthetic cDNA sequences encoding human insulin-like growth factor-1 (IGF-1) in the mammary gland of transgenic rabbits. Gene 149(2): 351-355.

Brinster, R, Allen, JM, Behringer, RR, Gelinas, RE, and Palmiter, RD (1988). Introns enhance transcriptional efficiency in transgenic mice. Proc Natl Acad Sci USA 85: 836-840.

Bühler, T, Bruyere, T, Went, DF, Stranzinger, G, and Bürki, K (1990). Rabbit β-casein promotor directs secretion of human interleukin-2 into the milk of transgenic rabbits. Bio/Technology *8*, 140-143.

Burgers, P, and Percival, KJ (1987). Transformation of yeast spheroplasts without cell fusion. Anal Biochem 163: 391-397.

Burke, D, Carle, GF, Olson, MV (1987). Cloning of large segments of exogenous DNA into yeast by means of artifical chromosome vectors. Science 236(4803): 806-812.

Butler, S, van Cott, K, Subrumanian, A, Gwazduaskas, FC, and Velander, WH (1997). Current progress in the production of recombinant human fibrinogen in the milk of transgenic animals. Thromb Haemost 78(1): 537-542.

Campbell, K, McWhir, J, Ritchie, WA, and Wilmut, I (1996). Sheep cloned by nuclear transfer from a cultured cell line. Nature 380(6569): 64-66.

Capecchi, M (1989). Altering the genome by homologous recombination. Science 244: 1288-1292.

Carson, S, and Wiles, M (1993). Far upstream regions of class II MHC Ea are necessary for position independent copy number dependent expression of Ea transgenesis. Nucleic Acids Res 21: 2065-2072.

Carver, A, Dalrymple, MA, Wright, G, Cottom, DS, Reeves, DB, Gibson, YH, Keenan, JL, Barrass, JD, Scott, AR, Colman, A, and Garner, I (1993). Transgenic livestock as bioreactors: Stable expression of human alpha-1-antitrypsin by a flock of sheep. Bio/Technology 11(Nov): 1263-1270.

Carver A, Wright, G, Cottom D, Cooper J, Dalrymple M, Temperley S, Udell M, Reeves D, Percy J, Scott A, Garner, I, and Colman, A (1992). Expression of human alpha 1 antitrypsin in transgenic sheep. Cytotechnology 9: 77-84.

Castilla, J, Pintado, B, Sola, I, Sanchez-Morgado, JM, and Enjuanes, L (1998). Engineering passive immunity in transgenic mice secreting virus-neutralizing antibodies in milk. Nat Biotechnol 16(4): 349-354.

Caterina, J, Ciavatta, DJ, Donze, D, Behringer, RR, and Townes ,TM (1994). Multiple elements in human beta-globin locus control region 5' HS 2 are involved in enhancer activity and position-independent, transgene expression. Nucleic Acids Res 22(6): 1006-1011.

Caterina, J, Ryan, TM, Pawlik, KM, Palmiter, RD, Brinster, RL, Behringer, RR, and Townes, TM (1991). Human beta-globin locus control region: analysis of the 5' DNase I hypersensitive site HS 2 in transgenic mice. Proc Natl Acad Sci U S A 88(5): 1626-1630.

Cavaletto, M, Cantisani, A, Giuffrida, G, Napolitano, L, and Conti, A (1994). Human alpha S1-casein like protein: purification and N-terminal sequence determination. Biol Chem Hoppe Seyler 375(2): 149-151.

Cerdan , M, Young, JI, Zino, E, Falzone, TL, Otero, V, Torres ,HN, and Rubinstein, M (1998). Accurate spatial and temporal transgene expression driven by a 3.8-kilobase promoter of the bovine beta-casein gene in the lactating mouse mammary gland. Mol Reprod Dev 49(3): 236-245.

Chamberlain, J, Vasavada, HA, Ganguly, S, and Weissman, SM (1991). Identification of cis sequences controlling efficient position-independent tissue-specific expression of human major histocompatibility complex class I genes in transgenic mice. Mol Cell Biol 11(7): 3564-3572.

Chamberlain, J, Nolan, JA, Conrad, PJ, Vasavada, HA, Vasavada, HH, Ploegh, HL, Ganguly, S, Janeway, CA Jr, and Weissman, SM (1988). Tissue-specific and cell surface expression of human major histocompatibility complex class I heavy (HLA-B7) and light (beta 2-microglobulin) chain genes in transgenic mice. Proc Natl Acad Sci U S A 85(20): 7690-7694.

Chen, L-H, and Bissell, MJ (1989). A novel regulatory mechanism for whey acidic protein gene expression. Cell Regulation 1: 45-54.

Choi, B-K, Bleck, GT, Wheeler, MB and Jimenez-Flores, R (1996). Genetic modification of bovine β-casein and its expression in the milk of transgenic mice. J Agric Food Chem 44: 953-960.

Choi, O, and Engel, JD (1988). Developmental regulation of beta-globin gene switching. Cell 55(1): 17-26.

Choi, T, Hollenbach, PW, Pearson, BE, Ueda, RM, Weddell, GN, Kurahara, CG, Woodhouse, CS, Kay, RM, and Loring, JF (1993). Transgenic mice containing a human heavy chain immunoglobulin gene fragment cloned in a yeast artificial chromosome. Nature Genet 4: 117-123.

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 guanidium thiocyanate-phenol-chloroform extraction. Anal Biochem 162: 156-159.

Chong, D, Roberts, W, Arakawa, T, Illes, K, Bagi, G, Slattery, CW, and Langridge, WHR (1997). Expression of the human milk casein β-casein in transgenic potato plants. Transgenic Res 6: 289-296.

Chung, J, Whiteley, M, and Felsenfeld, G (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell 74(3): 505-514.

Cibelli, J, Stice, SL, Golueke, PJ, Kane, JL, Jerry, J, Blackwell, C, Ponce de León, FA, and Robl, JM

(1998). Cloned transgenic calves produced from nonquiescent fetal fibroblasts. Science 280: 1256-1258.

Clark, A, Cowper, A, Wallace, R, Wright, G, and Simons, JP (1992). Rescuing transgene expression by co-integration. BioTechnology 10(11): 1450-1454.

Clark, A (1992). Prospects for the genetic engineering of milk. J Cell Biochem 49(2): 121-127.

Clark, A, Bissinger, P, Bullock, DW, Damak, S, Wallace, R, Whitelaw, CBA and Yull, F (1994). Chromosomal position effects and the modulation of transgene expression. Reprod Fertil Dev 6: 589-598.

Clark, A, and Harold, G, and Yull, FE (1997a). Mammalian cDNA and prokaryotic reporter sequences silence adjacent transgenes in transgenic mice. Nuc Acids Res 25(5): 1009-1014.

Clark, A, Iwobi, M, Cui, W, Crompton, M, Harold, G, Hobbs, S, Kamalati, T, Knox, R, Neil, C, Yull, F, and Gusterson, B (1997b). Selective cell ablation in transgenic mice expressing E. coli nitroreductase. Gene Ther 4: 101-110.

Clark, A, Bessos, H, Bishop, JO, Brown, P, Harris, S, Lathe, R, McClenaghan, M, Prowse, C, Simons, JP, Whitelaw, CBA, and Wilmut, I (1989). Expression of human anti-hemophilic factor IX in the milk of transgenic sheep. Bio/Tech 7: 487-492.

Cole, C, Collins, JE, and Dunham, I (1996). YAC library screening I. YAC Protocols. D. Markie. Totowa, Humana Inc. 54: 23-31.

Cook, P (1989). The nucleoskeleton and the topology of transcription. Eur J Biochem 185(3): 487-501.

Cox, R, Meier-Ewart, S, Ross, M, Larin, Z, Monaco, AP, and Lehrach, H (1993). Genome mapping and cloning of mutations using yeast artificial chromosomes. Methods Enzym 225: 637-653.

Dalgleish, D, Horse, DS, and Law, AJR (1989). Size-related differences in bovine casein micelles. Biochim Biophys Acta 991: 383.

Darnell, JJ, Kerr, IM, and Stark, GR (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. Science 264(5164): 1415-1421.

Davies, N, Rosewell, IR, Richardson, JC, Cook, GP, Neuberger, MS, Brownstein, BH, Norris, ML, Brüggemann, M (1993). Creation of mice expressing human antibody light chains by introduction of a yeast artificial chromosome containing the core region of the human immunoglobulin κ locus. Bio/Technology 11: 911-914.

Davis, B, and MacDonald, RJ, (1988). Limited transcription of rat elastase I transgene repeats in transgenic mice. Genes Dev 2(1): 13-22.

Demmer, J, Burdon, TG, Djiane, J, Watson, CJ, and Clark, AJ (1995). The proximal milk protein binding

factor binding site is required for the prolactin responsiveness of the sheep beta-lactoglobulin promoter in Chinese hamster ovary cells. Mol Cell Endocrinol 107(1): 113-121.

Devinoy, E, Thepot, D, Stinnakre, MG, Fontaine, ML, Grabowski, H, Puissant, C, Pavirani, A, and Houdebine, LM (1994). High level production of human growth hormone in the milk of transgenic mice: the upstream region of the rabbit whey acidic protein (WAP) gene targets transgene expression to the mammary gland. Transgenic Res 3(2): 79-89.

Diaz, P, Cado, D, Winoto, A (1994). A locus control region in the T cell receptor alpha/delta locus. Immunity 1(3): 207-217.

Dillon, N, and Grosveld, F (1993). Transcriptional regulation of multigene loci: multilevel control. Trends Genet 9(4): 134-137.

Dillon, N, Trimborn, T, Strouboulis, J, Fraser, P, and Grosveld, F (1997). The effect of distance on long range chromatin interactions. Mol Cell 1: 131-138.

Dillon, N (1993). Regulating gene expression in gene therapy. Trends Biotechnol 11(5): 167-173.

DiTullio, P, Cheng, SH, Marshall, J, Gregory, RJ, Ebert, KM, Meade, HM, and Smith, A (1992). Production of cystic fibrosis transmembrane conductance regulator in the milk of transgenic mice. Bio/Technology *10*, 74-77.

Dobie, K, Lee, M, Fantes, JA, Graham, E, Clark, AJ, Springbett, A, Lathe, R, and McClenaghan, M (1996). Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. Proc Natl Acad Sci USA 93: 6659-6664.

Dobrovolsky, V, Lagutin, OV, Vinogradova, TV, Frolova, IS, Kuznetsov, VP, and Larionov, OA (1993). Human gamma-interferon expression in the mammary gland of transgenic mice. FEBS Lett 19: 181-184.

Dorer, D (1997). Do transgene arrays form heterochromatin in vertebrates? Transgenic Res 6(1): 3-10.

Dorer, D, Henikoff, S (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. Cell 77(7): 993-1002.

Eberl, D, Duyf, BJ, and Hilliker, AJ (1993). The role of heterochromatin in the expression of a heterochromatic gene, the *rolled* locus of *Drosophila melanogaster*. Genetics 125: 141-154.

Ebert, K, DiTullio, P, Barry, CA, Schindler, JE, Ayres, SL, Smith, TE, Pellerin, LJ, Meade, HM, Denman, J, and Roberts, B (1994). Induction of human tissue plasminogen activator in the mammary gland of transgenic goats. Biotechnology 12(7): 699-702.

Ebert, K, Selgrath, JP, DiTullio, P, Denman, J, Smith, TE, Memon, MA, Schindler, JE, Monastersky, GM, Vitale, JA, and Gordon, K (1991). Transgenic production of a variant of human tissue-type plasminogen

activator in goat milk: generation of transgenic goats and analysis of expression. Biotechnology 9(9): 835-838.

Edlund, A, Johansson, T, Leidvik, B, and Hansson, L (1996). Structure of the human kappa-casein gene. Gene 174(1): 65-69.

Edmunds, T, Van Patten, SM, Pollock, J, Hanson, E, Bernasconi, R, Higgins, E, Manavalan, P, Ziomek, C, Meade, H, McPherson, JM, and Cole, ES (1998). Transgenically produced human antithrombin: structural and functional comparison to human plasma-derived antithrombin. Blood 91(12): 4561-4571.

Eisenstein, R, and Rosen, JM (1988). Both cell substratum regulation and hormonal regulation of milk protein gene expression are exerted primarily at the posttranscriptional level. Mol Cell Biol 8, 3183-3190.

Eissenberg, J, and Elgin, SC (1991). Boundary functions in the control of gene expression. Trends Genet 7(10): 335-340.

Elliott, J, Festenstein, R, Tolaini, M, and Kioussis, D (1995). Random activation of a transgene under the control of a hybrid hCD2 locus control region/Ig enhancer regulatory element. EMBO J 14(3): 575-584.

Ellis, J, Tan-Un, KC, 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(3): 562-568.

Ellis, J, Pasceri, P, Tan-Un, KC, Wu, X, Harper, A, Fraser, P, and Grosveld, F (1997). Evaluation of β-globin gene therapy constructs in single copy transgenic mice. Nuc Acids Res 25(6): 1296-1302.

Ellis, J, Talbot, D, Dillon, N, and Grosveld, F (1993). Synthetic human beta-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. EMBO J 12(1): 127-134.

Engel, J (1993). Developmental regulation of human beta-globin gene transcription: a switch of loyalties? Trends Genet 9(9): 304-309.

Enver, T, Raich, N, Ebens, AJ, Papayannopoulou, T, Costantini, F, and Stamatoyannopoulos, G (1990). Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. Nature 344(6264): 309-313.

Epner, E, Reik, A, Cimbora, D, Telling, A, Bender, MA, Fiering, S, Enver, T, Martin, DIK, Kennedy, M, Keller, G, and Groudine, M (1998). The β-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse β-globin locus. Mol Cell 2: 447-455.

Ferretti, L, Leone, P, and Sgaramella, V (1990). Long range analysis of the bovine casein genes. Nucleic Acids Res 18(23): 6829-6833.

Festenstein, R, Tolaini, M, Corbella, P, Mamalaki, C, Parrington, J, Fox, M, Miliou, A, Jones, M, and

Kioussis, D (1996). Locus control region function and heterochromatin-induced position effect variegation. Science 271: 1123-1125.

Fiering, S, Epner, E, Robinson, K, Zhuang, Y, Telling, A, Hu, M, Martin, DIK, Enver, T, Ley, T, and Groudine, M (1995). Targeted deletion of 5'HS2 of the murine β-globin locus reveals that it is not essential for the proper regulation of the β-globin locus. Genes Dev 9: 2203-2213.

Forget, B (1993). YAC transgenes: Bigger is probably better. Proc Natl Acad Sci USA 90: 7909-7911.

Forrester, W, Takegawa, S, Papayannopoulou, T, Stamatoyannopoulos, G, and Groudine, M (1987). Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. Nucleic Acids Res 15(24): 10159-10177.

Forrester, W, Epner, E, Driscoll, MC, Enver, T, Brice, M, Papayannopoulou, T, and Groudine, M (1990). A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. Genes Dev 4(10): 1637-1649.

Forrester, W, Thompson, C, Elder, JT, and Groudine, M (1986). A developmentally stable chromatin structure in the human beta-globin gene cluster. Proc Natl Acad Sci U S A 83(5): 1359-1363.

Fraser, P, Pruzina, S, Antoniou, 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 & Dev 7: 106-113.

Ganss, R, Montoliu, L, Monaghan, AP, and Schütz, G (1994). A cell-specific enhancer far upstream of the mouse tyrosinase gene confers high level and copy number-related expression in transgenic mice. EMBO J 13(13): 3083-3093.

George, S, Clark, AJ, and Archibald, AL (1997). Physical mapping of the murine casein locus reveals the gene order as α-β-γ-ε-κ. DNA Cell Biol 16(4): 477-484.

Geyer, P (1997). The role of insulator elements in defining domains of gene expression. Curr Opin Genet Dev 7: 242-248.

Gnirke, A, Huxley, C, Peterson, K, and Olson, MV (1993). Microinjection of intact 200- to 500-kb fragments of YAC DNA into mammalian cells. Genomics 15(3): 659-667.

Gordon, J, Chesa, PG, Nishimura, H, Rettig, WJ, Maccari, JE, Endo, T, Seravalli, E, Seki, T, and Silver, J (1987). Regulation of Thy-1 gene expression in transgenic mice. Cell 50(3): 445-452.

Gordon KE, Chapman, RS, Kurian, KM, Clark, AJ, Lane, EB, and Watson, CJ (1999). Induction of differentiation and apoptosis in a novel conditionally immortal mouse mammary epithelial cell line, KIM-2. In press.

Gouilleux, F, Wakao, H, Mundt, M, and Groner, B (1994). Prolactin induces phosphorylation of Tyr694 of Stat5 (MGF), a prerequisite for DNA binding and induction of transcription. EMBO J 13(18): 4361-4369.

Greaves, D, Wilson, FD, 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.

Green, E, and Olson, MV (1990). Systematic screening of yeast artificial-chromosome libraries using the polymerase reaction. Proc Natl Acad Sci USA 87: 1213-1217.

Greenberg, N, Reding, TV, Duffy, T, and Rosen, JM (1991). A heterologous hormone response element enhances expression of rat β-casein promotor-driven chloramphenicol acetyltransferase fusion genes in the mammary gland of transgenic mice. Mol Endocrinol *5*, 1504-1512.

Gribnau, J, de Boer, E, Trimborn, T, Wijgerde, M, Milot, E, Grosveld, F and Fraser, P (1998). Chromatin interaction mechanism of transcriptional control *in vivo*. EMBO J 17(20): 6020-6027.

Groenen, M, Dijkhof, RJM, Verstege, AJM, and Van der Poel, JJ (1993). The complete nucleotide sequence of the bovine α_{s2}-casein gene. Gene *123*, 187.

Grosveld, F, Antoniou, M, Berry, M, de Boer, E, Dillon, N, Ellis, J, Fraser, P, Hurst, J, Imam, A, Meijer, D, et al (1993). Regulation of human globin gene switching. Cold Spring Harb Symp Quant Biol 58: 7-13.

Grosveld, F, de Boer, E, Dillon, N, Fraser, P, Gribnau, J, Milot, E, Trimborn, T, and Wijgerde, M (1998). The dynamics of globin gene expression and gene therapy vectors. Semin Hematol 35(2): 105-111.

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

Groudine, M, Kohwi-Shigematsu, T, Gelinas, R, Stamatoyannopoulos, G, and Papayannopoulou, T (1983). Human fetal to adult hemoglobin switching: changes in chromatin structure of the beta-globin gene locus. Proc Natl Acad Sci U S A 80(24): 7551-7555.

Groudine, M, and Weintraub, H (1982). Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. Cell 30(1): 131-139.

Gutiérrez, A, Meade, HM, Ditullio, P, Pollock, D, Harvey, M, Jiménez-Flores, R, Anderson, GB, Murray, JD, and Medrano, JF (1996). Expression of a bovine κ-CN cDNA in the mammary gland of transgenic mice utilizing a genomic milk protein gene as an expression cassette. Transgenic Res 5(4): 271-279.

Gutiérrez-Adan, A, Maga, EA, Meade, H, Shoemaker, CF, Medrano, JF, Anderson, GB, and Murray, JD (1996). Alterations of the physical characteristics of milk from transgenic mice producing bovine κ-casein. J Dairy Sci 79: 791-799.

Hamvas, R, Francis, FF, Cox, RD, Nizetic, D, Goldsworthy, ME, Brown, SDM , and Lehrach, HR (1994). Rapid restriction analysis of YAC clones. Nuc Acids Res 22(7): 1318-1319.

Hanahan, D (1983). Studies on transformation of *Escherischia coli* with plasmids. J Mol Biol 166: 447-589.

Hanscombe, O, Whyat,t D, Fraser, P, Yannoutsos, N, Greaves, D, Dillon, N, and Grosveld, F (1991). Importance of globin gene order for correct developmental expression. Genes Dev 5(8): 1387-1394.

Hanson, I, and van Heyningen, V (1995). Pax6: more than meets the eye. Trends Genet 11: 268-272.

Hansson, L, Edlund, M, Edlund, A, Johansson, T, Marklund, SL, Fromm, S, Stromquist, M, and Tornell, J (1994). Expression and characterisation of biologically active human extracellular superoxide dismutase in milk of transgenic mice. J Biol Chem 269: 5358-5363.

Harris, S, McClenaghan,M, Simons, JP, Ali, S, Clark, AJ (1991). Developmental regulation of the sheep beta-lactoglobulin gene in the mammary gland of transgenic mice. Dev Genet 12(4): 299-307.

Hart, C, and Laemmli, UK (1998). Facilitation of chromatin dynamics by SARs. Curr Opin Genet Dev 8(5): 519-525.

Heard, E, Avner, P and Rothstein, R (1994). Creation of a deletion series of mouse YACs covering a 500kb region around *Xist* . Nucleic Acids Res 22(10): 1830-1837.

Heikoop, J, Steensma, Y, van Ommen, G-JB, and den Dunnen, JT (1994). A simple nd rapid method for separating co-cloned YACs. Trends Genet 10: 40.

Hieter, P, Connelly, C, Shero, J, McCormick, M. K, Antonarakis, S, Pavan, W, and Reeves, R (1990). Yeast artificial chromosomes: promises kept and pending. In Genome Analysis Vol. I: Genetic and Physical Mapping. Cold Spring Harbour Laboratory Press.: 83-120.

Higginson, D, Perez, G, Lopez-Canovas, L, and Riveron, AM (1994). Non-enzymatic deproteinization of immobilized DNA suitable for pulsed field gel electrophoresis. Analytical Letters 27(7): 1255-1264.

Higgs, D (1998). Do LCRs open chromatin domains? Cell 95: 299-302.

Higgs DR, W. W, Jarman AP, Sharpe J, Lida J, Pretorius IM, Ayyub H (1990). A major positive regulatory region located far upstream of the human alpha-globin gene locus. Genes Dev 4(9): 1588-1601.

Hogan, B, Constantini, F, and Lacy, E (1986). Manipulating the mouse genome: A laboratory manual, Cold Spring Harbour.

Holdridge, C, and Dorsett, D (1991). Repression of *hsp70* heat shock gene transcription by the supressor of Hairy-wing protein of *Drosophila melanogaster*. Mol Cell Biol 11: 1894-1900.

Hugerat, Y, Spencer, F, Zenvirth, D, and Simchen, G (1994). A versatile method for efficient YAC transfer between any two strains. Genomics 22: 108-117.

Hurwitz DR, N. M, Barash I, Ilan N, Shani M (1994). Specific combinations of human serum albumin introns direct high level expression of albumin in transfected COS cells and in the milk of transgenic mice. Transgenic Res 3(6): 365-375.

Hyttinen JM, K. V, Hiltunen MO, Myohanen S, Janne J (1998). High-level expression of bovine beta-lactoglobulin gene in transgenic mice. J Biotechnol 61(3): 191-198.

Iannaccone PM, T. G, Garton RL, Caplice MD, Brenin DR (1994). Pluripotent embryonic stem cells from the rat are capable of producing chimeras. Dev Biol 163(1): 288-292.

Ito H, F. Y, Murata K, Kimura A (1983). Transformation of intact yeast cells treated with alkali cations. J Bacteriol 153(1): 163-168.

Jaenisch, R (1988). Transgenic animals. Science 240: 1468-1474.

Jakobovits, A, Moore, AL, Green, LL, Vergara, GJ, Maynard-Currie, CE, Austin, HA, and Klapholz, S (1993). Germ-line transmission and expression of a human-derived yeast artificial chromosome. Nature 362: 255-258.

Jantzen K, F. H, Igo-Kemenes T (1986). The DNase I sensitive domain of the chicken lysozyme gene spans 24 kb. Nucleic Acids Res 14(15): 6085-6099.

Johnsen LB, R. L, Petersen TE, Berglund L (1995). Characterization of three types of human alpha s1-casein mRNA transcripts. Biochem J 309(Pt 1): 237-242.

Jolles J, S. F, Hermann J, Alais C, Jolles P (1974). The sequence of sheep kappa-casein: primary structure of para-kappa A-casein. Eur J Biochem 46(1): 127-132.

Jones, B, Monks, BR, Leibhaber, SA, and Cooke, NE (1995). The human growth hormone gene is regulated by a multicomponent locus control region. Mol Cell Biol 15: 7010-7021.

Jones, W, Yu-Lee, L, Cliffs, SM, Brow, TL, and Rosen, JM (1985). The rat casein multigene family. Fine structure and evolution of the beta-casein gene. J Biol Chem *260*, 7042-7050.

Kaiser, D, Tabor, H, and Tabor, CW (1963). J Mol Biol 6(141-147).

Kamakaka, R (1997). Silencers and locus control regions: opposite sides of the same coin. TIBS 22: 124-127.

Karpen, G (1994). Position-effect variegation and the new biology of heterochromatin. Curr Opin Genet Dev 4: 281-291.

Kellum, R, and Elgin, SCR (1998). Chromatin boundaries: Punctuating the genome. Curr Biol 8: R521-R524.

Kellum, R, and Schedl, P (1991). A position-effect assay for boundaries of higher order chromosomal domains. Cell 64: 941-950.

Kim CG, Epner, E, Forrester, WC, and Groudine, M (1992). Inactivation of the human beta-globin gene by targeted insertion into the beta-globin locus control region. Genes Dev 6(6): 928-938.

Kioussis, D, and Festenstein, R (1997). Locus control regions: overcoming heterochromatin-induced gene inactivation in mammals. Curr Opin Genet Dev 7: 614-619.

Kioussis D, V. E, deLange T, Flavell RA, Grosveld FG (1983). Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. Nature 306(5944): 662-666.

Kleinjan, D, and van Heyningen, V (1998). Position effect in human genetic disease. Hum Mol Genet 7(10): 1611-1618.

Koczan, D, Hobom, G, and Seyfert, H (1991). Genomic organisation of the bovine alpha-S1 casein gene. Nuc Acids Res *10*, 5519-5596.

Kollias, G, and Grosveld, F (1992). The study of gene regulation in transgenic mice. Transgenic Animals. G. Kollias, and Grosveld, F. London, Academic Press: 79-98.

Kollias G, W. N, Hurst J, Grosveld F (1986). Regulated expression of human A gamma-, beta-, and hybrid gamma beta-globin genes in transgenic mice: manipulation of the developmental expression patterns. Cell 46(1): 89-94.

Korhonen VP, T. M, Hyttinen JM, Uusi-Oukari M, Sinervirta R, Alhonen L, Jauhiainen M, Janne OA, Janne J (1997). Expression of bovine beta-lactoglobulin/human erythropoietin fusion protein in the milk of transgenic mice and rabbits. Eur J Biochem 245(2): 482-489.

Kovacic, R, Comai, L, and Bendich, AJ (1995). Protection of megabase DNA from shearing. Nuc Acids Res 23(19): 3999-4000.

Krimpenfort, P, Rademakers, A, Eyestone, W, van der Schans, A, van den Broek, S, Kooiman, P, Kootwijk, E, Platenburg, G, Pieper, F, Strijker, R, and de Boer, H (1991). Generation of transgenic dairy cattle using '*in vitro*' embryo production. Biotechnology (NY) *9*, 844-847.

Kumar, S, Clarke, AL, Hooper, ML, Horne, DS, Law, AJR, Leaver, J, Springbett, A, Stevenson, E, and Simons, JP (1994). Milk composition and lactation of β-casein-deficient mice. Proc Natl Acad Sci USA 91(3): 6138-6142.

Kunz C, and Lonnerdal, B (1989). Human milk proteins: separation of whey proteins and their analysis by polyacrylamide gel electrophoresis, fast protein liquid chromatography (FPLC) gel filtration, and anion-exchange chromatography. Am J Clin Nutr 49(3): 464-470.

L'Huillier PJ, Soulier, S, Stinnakre MG, Lepourry L, Davis SR, Mercier JC, Vilotte JL (1996). Efficient and specific ribozyme-mediated reduction of bovine alpha-lactalbumin expression in double transgenic mice. Proc Natl Acad Sci U S A 93(13): 6698-6703.

Laird, J, Jack, L, Hall, AP, Boulton, D, Parker, and Craig, RK (1988). Structure and expression of the guinea-pig alpha-lactalbumin gene. Biochem J *254*, 85-94.

Lamb, B, Sisodia, SS, Lawler, AM, Slunt, HH, Kitt, CA, Kearns, WG, Pearson, PL, Price, DL, and Gearhart, JD (1993). Introduction and expression of the 400 kilobase *amyloid precursor protein* gene in transgenic mice. Nature Genet 5: 22-30.

Lang, G, Mamalaki, C, Greenberg, D, Yannoutsos, N and Kioussis, D (1991). Deletion analysis of the human CD2 locus control region in transgenic mice. Nucleic Acids Res 19(21): 5851-5856.

Lang, G, Wotton, D, Owen, MJ, Sewell, WA, Brown, MH, Mason, DY, Crumpton, MJ, and Kioussis, D (1988). The structure of the human *CD2* gene and its expression in transgenic mice. EMBO J 7: 1675-1682.

Langley B, V. J, Stinnakre MG, Whitelaw CB, L'Huillier PJ (1998). Rescue of an MMTV transgene by co-integration reveals novel locus control properties of the ovine beta-lactoglobulin gene that confer locus commitment to heterogeneous tissues. Transgenic Res 7(3): 205-212.

Larin, Z, Monaco, AP, and Lerhach, H (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc Natl Acad Sci USA 88: 4123-4127.

Larin, Z, Monaco, AP, and Lehrach, H (1996). Generation of large insert YAC libraries. Methods in Molecular Biology: YAC protocols. D. Markie. Totawa, New Jersey, Humana Press. 54: 1-12.

Larin, Z, Monaco, AP, Meier-Ewert, S, and Lehrach, H (1987). Construction and characterization of yeast artificial chromosome libraries from the mouse genome. Methods Enzym 225: 623-637.

Larionov V, Graves J, Kouprina N, and Resnick MA (1994). Recombination during transformation as a source of chimeric mammalian artificial chromosomes in yeast (YACs). Nucleic Acids Res 22(20): 4234-4241.

Lee, J, Strauss, WM, Dausman, JA, and Jaenisch, R (1996). A 450kb transgene displays properties of the mammalian X-inactivation centre. Cell 83: 83-94.

Lee, K, Atiee, SH, and Rosen, JM (1989). Differential regulation of rat β-casein chloramphenicol acetyl-transferase gene expression in transgenic mice. Mol Cell Biol *8*, 3183-3190.

Lee, K, Atiee, SH, Henning, SJ, and Rosen, JM (1989). Relative contribution of promotor and intragenic sequences in the hormonal regulation of rat β-casein transgenes. Mol Endocrinol *3*, 447-453.

Lee, CS, and Oka, T (1992). Progesterone regulation of a pregnancy-specific transcription repressor to beta-casein gene promoter in mouse mammary gland. Endocrinology 131(5): 2257-2262.

Lee TH, Kim, SJ, Han YM, Yu DY, Lee CS, Choi YJ, Moon HB, Baik MG, Lee KK (1998). Matrix attachment region sequences enhanced the expression frequency of a whey acidic protein/human lactoferrin fusion gene in the mammary gland of transgenic mice. Mol Cells 8(5): 530-536.

Leroux, C, and Martin, P (1996). The caprine α-s1- and β-casein genes are 12-kb apart and convergently transcribed. Anim Genet 27((Suppl 2)): 93.

Ley, T, Hug, B, Fiering, S, Epner, E, Bender, MA, and Groudine, M (1998). Reduced β-globin gene expression in adult mice containing deletions of locus control region 5' HS-2 or 5' HS-3[a-] Ann NY Acad Sci 850: 45-53.

Li, Q, and Stamatoyannopoulos, JA (1994). Position independence and proper developmental control of γ-globin gene expression require both a 5' locus control region and a downstream sequence element. Mol Cell Biol 14: 6087-6096.

Li S, and Rosen, JM (1994). Glucocorticoid regulation of rat whey acidic protein gene expression involves hormone-induced alterations of chromatin structure in the distal promoter region. Mol Endocrinol 8(10): 1328-1335.

Lin, C, McAllister, AJ, Ng-Kwai-Hang, KF, Hayes, JF, Batra, TR, Lee, AJ, Roy, GL, Vesely, JA, Wauthy, JM, and Winter, KA (1989). Relationships of milk protein types to lifetime performance. J Dairy Sci *72*, 3085-3090.

Loc PV, and Strätling, WH (1988). The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. EMBO J 7(3): 655-664.

Lui, K, Sandgren, EP, Palmiter, RD, and Stein, A (1995). Rat growth hormone gene introns stimulate nuclesome alignment *in vitro* and in transgenic mice. Proc Natl Acad Sci USA 92(August): 7724-7728.

Maga, E, Anderson, GB, Huang, MC, and Murray, JD (1994). Expression of human lysozyme mRNA in the mammary gland of transgenic mice. Trans Res *3*, 36-42.

Magram J, Chada, K, and Costantini F (1985). Developmental regulation of a cloned adult beta-globin gene in transgenic mice. Nature 15(6017): 338-340.

Maniatis, T, Fritsch, EF, Lauer, J, and Lawn, R. M (1981). Molecular genetics of human hemoglobins.

Ann Rev Genet 14: 145-178.

Markie, D (1997). Yeast Transformation. Course Notes: Wellcome Trust 19th Summer School, London.

Markie, D (1996). Markers, selection and media in yeast cloning. YAC Protocols. D. Markie, Humana Press Inc. 54: 359-371.

Martin, D, Fiering, S and Groudine, M (1996). Regulation of β-globin gene expression: straightening out the locus. Curr Opin Genet Dev 6: 488-495.

Matsuura, S, Episkopou, V, Hamvas, R, and Brown, SDM (1996). *Xist* expression from an *XIst* YAC transgene carried on the mouse Y chromosome. Hum Mol Genet 5(4): 451-459.

May, G, and Enver, T (1995). Targeting gene expression to haemopoetic stem cells: a chromatin-dependent upstream element mediates cell type-specific expression of the stem cell antigen CD34. EMBO J 14: 564-574.

McFarlane, M, and Wilson, JB (1996). A model for the mechanism of precise integration of a microinjected transgene. Transgenic Res 5: 171-177.

McKee C, Gibson, A, Dalrymple M, Emslie L, Garner I, and Cottingham I (1998). Production of biologically active salmon calcitonin in the milk of transgenic rabbits. Nat Biotechnol 16(7): 647-651.

McKeown, G, Watson, A, Karunaratne, K, and Bentley, D (1993). High throughput filter preparation robot. Program and abstracts genome sequencing and analysis conference V.
Abstract C21

McKnight, R, Wall, RJ, and Hennighausen, L (1995). Expression of genomic and cDNA constructs after co-integration in transgenic mice. Transgenic Res 4: 39-43.

McKnight, R, Shamay, A, Sankaran, L, Wall, RJ, and Hennighausen, L (1992). Matrix-attachment regions impart position independent regulation of a tissue specific gene in transgenic mice. Proc Natl Acad Sci USA 89: 6943-6947.

Meade, H, Gates, L, Lacy, E, and Lonberg, N (1990). Bovine alpha$_{s1}$-casein gene sequences direct high level expression of active human urokinase in mouse milk. Bio/Technology *8*, 443-446.

Mercier, J.-C, and Vilotte, J-L (1993). Structure and function of milk protein genes. J Dairy Sci 76: 3079-3098.

Milot, E, Stouboulis, J, Trimborn, T, Wijgerde, M, de Boer, E, Langeveld, A, Tan-Un, K, Vergeer, W, Yannoutosos, N, Grosveld, F, and Fraser, P (1996). Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. Cell 87: 105-114.

Milot, E, Fraser, P, and Grosveld, F (1996). Position effects and genetic disease. Trends Genet 12(4): 123-126.

Montoliu, L, Umland, T, and Schutz, G (1995). A locus control region at -12kb of the tyrosine gene. EMBO J 15: 6026-6034.

Moreadith, R, and Radford, NB (1997). Gene targeting in embryonic stem cells: the new physiology and metabolism. J Mol Med 75: 208-216.

Newlon, C (1988). Yeast chromosome replication and segregation. Micorbiol Rev 52(4): 568-601.

Neznanov, N, Thorey, IS, Ceceña, G, and Oshima, RG (1993). Transcriptional insulation of the human keratin 18 gene in transgenic mice. Mol Cell Biol 13: 2214-2223.

Nishio H, T. Y, Narita N, Yanagawa H, Suzuki Y, Ishikawa Y, Ishikawa Y, Minami R, Nakamura H, Matsuo M (1994). Identification of a novel first exon in the human dystrophin gene and of a new promoter located more than 500 kb upstream of the nearest known promoter. J Clin Invest 94(3): 1037-1042.

Nizetic, D, Zehetner, G, Monaco, AP, Gellen, L, Young, BD, and Lehrach, H (1991). Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. Proc Natl Acad Sci U S A 88(8): 3233-3237.

Paleyanda, R, Zhang, DW, Hennighausen, L, McKnight, RA, and Lubon, H (1994). Regulation of human protein C gene expression by the mouse WAP promoter. Transgenic Res 3(6): 335-343.

Palmiter, R, Sandgren, EP, Avarbock, M, Allen, DD, and Brinster, RL (1991). Heterologous introns can enhance expression of transgenes in mice. Proc Natl Acad Sci USA 88: 478-482.

Palmiter, R, and Brinster, RL (1986). Germ-line transformation of mice. Annu Rev Genet 20: 465-499.

Palmiter RD, Sandgren, E, Koeller DM, and Brinster RL (1993). Distal regulatory elements from the mouse metallothionein locus stimulate gene expression in transgenic mice. Mol Cell Biol 13(9): 5266-5275.

Pearson, B, and Choi, TK (1993). Expression of the human β-amyloid precursor protein gene from a yeast artificial chromosome in transgenic mice. Proc Natl Acad Sci USA 90: 10578-10582.

Peterson, K, Zitnik, G, Huxley, C, Lowrey, CH, Gnike, A, Leppig, KA, Papayannopoulou, T, and Stamatoyannopoulos, G (1993a). Use of yeast artificial chomosomes (YACs) for studying control of gene expression: correct regulation of the genes of a human β-gobin locus YAC following transfer to mouse erythroleukemia cell lines. Proc Natl Acad Sci USA 90: 11207-11211.

Peterson, K, Li, QL, Clegg, CH, Furukawa, T, Navas, PA, Norton, EJ, Kimbrough, TG, and Stamatoyannopoulos, G (1995). Use of yeast artificial chromosomes (YACS) in studies of mammalian

development: Production of β-globin locus YAC mice carrying human globin developmental mutants. Proc Natl Acad Sci USA 92: 5655-5659.

Peterson, K, Clegg, CH, Li, Q, and Stamatoyannopoulos, G (1997). Production of transgenic mice with yeast artificial chromosomes. Trends Genet 13: 61-66.

Peterson, K, Navas, PA, Li, Q, and Stamatoyonnopoulos, G (1998a). LCR-dependent gene expression in β-globin YAC transgenics: detailed structural studies validate functional analysis even in the presence of fragmented YACs. Hum Mol Genet 7(13): 2079-2088.

Peterson, K, Navas, PA, and Stamatoyannopoulos, G (1998b). β-YAC transgenic mice for studying LCR function. Ann NY Acad Sci 850: 28-37.

Peterson, K, Clegg, CH, Huxley, C, Josephson, BM, Haugen, HS, Furukawa, T, and and G. Stamatoyannopoulos (1993b). Transgenic mice containing a 248-kb yeast artificial chromosome carrying the human beta-globin locus display proper developmental control of human globin genes. Proc Natl Acad Sci USA 90(16): 7593-7597.

Pierre, S, Jolivet, G, Devinoy, E, and Houdebine, L (1994). A combination of distal and proximal regions is required for efficient prolactin regulation of transfected rabbit α_{s1}-casein chloramphenicol acetyltransferase constructs. Mol Endocrinol *8*, 1720-1730.

Pittius CW, H. L, Lee E, Westphal H, Nicols E, Vitale J, Gordon K (1988). A milk protein gene promoter directs the expression of human tissue plasminogen activator cDNA to the mammary gland in transgenic mice. Proc Natl Acad Sci U S A 85(16): 5874-5878.

Platenburg GJ, Kootwijk, E, Kooiman PM, Woloshuk SL, Nuijens JH, Krimpenfort PJ, Pieper FR, de Boer HA, and Strijker R (1994). Expression of human lactoferrin in milk of transgenic mice. Transgenic Res 3(2): 99-108.

Polach, K, and Widom, J (1995). Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. J Mol Biol 254: 130-149.

Pruzina S, Antoniou, M, Hurst J, Grosveld F, and Philipsen S (1994). Transcriptional activation by hypersensitive site three of the human beta-globin locus control region in murine erythroleukemia cells. Biochim Biophys Acta 19(2): 351-360.

Pursel, V, and Rexroad, CE Jr (1993). Status of research with transgenic farm animals. J Anim Sci 71 Suppl 3: 10-19.

Ragoussis, J, Trowsdale, J, and Markie, D (1992). Mitotic recombination of yeast artificial chromosomes. Nuc Acids Res 20(12): 3135-3138.

Raich N, Enver, T, Nakamoto B, Josephson B, Papayannopoulou T, and Stamatoyannopoulos G (1990).

Autonomous developmental control of human embryonic globin gene switching in transgenic mice. Science 250(4984): 1147-1149.

Rasmussen LK, Due, HA, and Petersen TE (1995). Human alpha s1-casein: purification and characterization. Comp Biochem Physiol B Biochem Mol Bio 111(1): 75-81.

Rathjen, P, Lake, J, Whyatt, LM, Bettes, MD, and Rathjen, J (1998). Properties and uses of embryonic stem cells: prospects for application to human biology and gene therapy. Reprod Fertil Dev 10: 31-47.

Raught, B, Khursheed, B, Kazansky, A, and Rosen J, (1994). YY1 represses β-casein gene expression by preventing the formation of a lactation-associated complex. Mol Cell Biol *14*, 1752-1763.

Raught, B, Liao, WS, and Rosen, JM (1995). Developmentally and hormonally regulated CCAAT/enhancer-binding protein isoforms influence β-casein gene expression. Mol Endocrinol *9*, 1223-1232.

Reid LH, Gregg RG, Smithies O, Koller BH (1990). Regulatory elements in the introns of the human HPRT gene are necessary for its expression in embryonic stem cells. Proc Natl Acad Sci U S A 87(11): 4299-4303.

Reik, A, Telling, A, Zitnik, G, Cimbora, D, Epner, E and Groudine, M (1998). The locus control region is necessary for gene expression in the human b-globin locus but not the maintenance of an open chromatin structure in erythroid cells. Mol Cell Biol 18(10): 5992-6000.

Reitman M, Lee E, Westphal H, Felsenfeld G. Mol Cell Biol 13(7): 3990-3998.

Rijnkels, M, Kooiman, PM, Platenburg, GJ, van Dixhhoorn, M, Nuijens, JH, de Boer, HA, and Pieper, Fr (1997). High-level expression of bovine α_{s1}-casein in milk of transgenic mice. Transgenic Res *7*, 5-14.

Rijnkels, M, Kooiman, PM, de Boer, HA, and Pieper, FR (1997). Organization of the bovine casein gene locus. Mamm Genome *8*, 148-152.

Rijnkels, M, Kooiman, PM, Krimpenfort, PJA, de Boer, HA, and Pieper, FR (1995). Expression analysis of the individual bovine β-, α_{s2}- and κ-casein genes in transgenic mice. Biochem J *311*, 929-937.

Rijnkels, M, Meershoek, E, de Boer, HA, and Pieper, FR (1997b). Physical map and localization of the human casein gene locus. Mamm Genome 8: 285-286.

Rijnkels, M, Wheeler, DA, de Boer, HA, and Pieper, FR (1997a). Structure and expression of the mouse casein gene locus. Mamm Genome 8: 9-15.

Riley, J, Morten, JEN, Anand, R (1992). Targeted integration of neomycin into yeast artificial chromosomes (YACS) for transfection into mammalian cells. Nuc Acids Res 20(12): 2971-2976.

Roberts, N, Sloane-Stanley, JA, Sharpe, JA, Stanworth, SJ, and Wood, WG (1997). Globing gene switching in transgenic mice carrying HS2-globin gene constructs. Blood 89(2): 713-723.

Robertson G, Garrick, D, Wu W, Kearns M, Martin D, and Whitelaw E (1995). Position-dependent variegation of globin transgene expression in mice. Proc Natl Acad Sci U S A 92(12): 5371-5375.

Rollema, H (1992). Casein association and micelle formation. In Advanced Dairy Chemistry, P. Fox, ed (London: Elsevier), pp. 111-140.

Roseman, R, Pirrotta, V and Geyer, PK (1993). The su(Hw) protein insulates expression of the *Drosophila melanogaster* white gene from chromosomal position-effects. EMBO J 12: 435-442.

Rosen, J (1987). Milk protein gene structure and expression. In The Mammary Gland: Development, regulation and function, M. Neville, and Daniel, C, ed (New York: Plenum Publishing), pp. 301-322.

Roskelley, C, Desprez, PY, and Bissell, MJ (1994). Extracellular matrix-dependent tissue-specific gene expression in mammary epithelial cells requires both physical and biochemical signal transduction. Proc Natl Acad Sci USA *91*, 12378-12382.

Ryan TM, Behringer, RR, Martin NC, Townes TM, Palmiter RD, and Brinster, RL (1989). A single erythroid-specific DNase I super-hypersensitive site activates high levels of human beta-globin gene expression in transgenic mice. Genes Dev 3(3): 314-323.

Sawyer L, and Holt, C (1993). The secondary structure of milk proteins and their biological function. J Dairy Sci 76(10): 3062-3078.

Schedl, A, Montolui, L, Kelsey, G, and Schütz, G (1993). A yeast artificial chromosome covering the tyrosine gene confers copy number-dependent expression in transgenic mice. Nature 362: 258.

Schedl, A, Larin, Z, Montolui, L, Thies, E, Kelsey, G, Lehrach, H, and Schütz, G (1993). A method for the generation of YAC transgenic mice by pronuclear microinjection. Nuc Acids Res 21(20): 4783-4787.

Schedl, A, Ross, A, Lee, M, Engelkamp, D, Rashbass, P, van Heyningen, V, and Hastie, ND (1996). Influence of *PAX6* gene dosage on development: overexpression causes severe eye abnormalities. Cell 86: 71-82.

Schedl, A, Grimes, B, and Montolui, L (1996). YAC transfer by microinjection. Methods in Molecular Biology: YAC protocols. D. Markie. Totowa, New Jersey, Humana Press. 54: 293-306.

Schedl, A, Beermann, F, Thies, E, Montoliu, L, Kelsey, G, and Schütz, G (1992). Transgenic mice generated by pronuclear injection of a yeast artificial chromosome. Nucleic Acids Res 20: 3073-3077.

Scherer, S, and Davis, RW (1979). Replacement of chromosome segments with altered DNA sequences

constructed *in vitro*. Proc Natl Acad Sci USA 76: 4951-4955.

Schiestl, R, and Gietz, RD (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Curr Genet 16: 339-346.

Schmidhauser, C, Bissell, MJ, Myers, CA, and Casperson, GF (1990). Extracellular matrix and hormones transcriptionally regulate bovine β-casein 5' sequences in stably transfected mouse mammary cells. Proc Natl Acad Sci USA *87*, 9118-9122.

Schmidt, D (1982). Association of caseins and casein micelle structure. Developments in Dairy Chemistry. P. Fox. New York, Appl Sci. 1: 61.

Schnieke AE, Kind, A, Ritchie WA, Mycock K, Scott AR, Ritchie M, Wilmut I, Colman A, and Campbell KH (1997). Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. Science 278(5346): 2130-2133.

Shani M, Barash, I, Nathan M, Ricca G, Searfoss GH, Dekel I, Faerman A, Givol D, Hurwitz DR (1992). Expression of human serum albumin in the milk of transgenic mice. Transgenic Res 1(5): 195-208.

Shesely, EG, Kim, HS, Shehee, WR, Papayannopoulou, T, Smithies, O, and Popovich, BW (1991). Correction of a human βS-globin gene by gene targeting. Proc Natl Acad Sci USA 88: 4294-4298.

Shi, Y, Seto, E, Chang, S, and Shenk, T (1991). Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. Cell *67*, 377-388.

Shih DM, Wall, RJ, and Shapiro, SG (1990). Developmentally regulated and erythroid-specific expression of the human embryonic beta-globin gene in transgenic mice. Nucleic Acids Res 18(18): 5465-5472.

Sikoski, R, and Hieter, P (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics 122: 19-27.

Simons, JP, McClenaghan, M, and Clark, AJ (1987). Alteration of the quality of milk by expression of sheep beta-lactoglobulin in transgenic mice. Nature 328(6130): 530-532.

Simpson, K, McGuigan, A, and Huxley, C (1996). Stable episomal maintenance of yeast artificial chromosomes in human cells. Mol Cel Biol 16(9): 5117-5126.

Sippel, AE, Saueressig, H, Winter, D, Grewal, T, Faust, N, Hecht, A, and Bonifer, C (1992). The regulatory domain organization of eukaryotic genomes: implications for stable gene transfer. Transgenic Animals. G. Kollias, and Grosveld, F. London, Academic Press: 1-26.

Sippel, AE, Schafer, G, Faust, N, Saueressig, H, Hecht, A, and Bonifer, C (1993). Chromatin domains constitute regulatory units for the control of eukaryotic genes. Cold Spring Harb Symp Quant Biol 58: 37-44.

Skarpidi, E, Vassilopoulos, G, and Stamatoyannopoulos, G (1998). Comparison of expression of human globin genes transferred into mouse erythroleukemia cells and in transgenic mice. Blood 92, 3416-3421.

Smith, D, Smyth, AP, and Moir, DT (1992). Copy number amplification of yeast artificial chromosomes. Methods in Enzymology. R. Wu. Totowa, NJ, Humana. 216: 603-614.

Smith, D, Smyth, AP, and Moir, DT (1990). Amplification of large artificial chromosomes. Proc Natl Acad Sci USA 87: 8242-8246.

Sola, I, Castilla, J, Pintado, B, Sanchez-Morgado, JM, Whitelaw, CB, Clark, AJ, and Enjuanes, L (1998). Transgenic mice secreting coronavirus neutralizing antibodies into the milk. J Virol 72(5): 3762-3772.

Southern, P, and Berg, P (1982). Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J Mol Appl Gen 1: 327-341.

Spencer, F, and Simchen, G (1996). Transfer of YAC clones to new yeast hosts. YAC protocols. D. Markie. Totowa, NJ, Humana Press Inc. 54: 239-252.

Spencer, F, Hugerat, Y, Simchen, Hurko, O, Connelly, C, and Hieter, P (1994). Yeast *kar1* mutants provide an effective method for YAC transfer to new hosts. Genomics 22: 118-126.

Stamatoyannopoulos, G, Josephson, B, Zhang, J-W, Li, Q (1993). Developmental regulation of human γ-globin genes in transgenic mice. Mol Cell Biol 13: 7636-7644.

Stamatoyannopoulos, J, Clegg, CH, and Li, Q (1997). Sheltering of γ-globin expression from position effects requires both an upstream locus control region and a regulatory element 3' to the $^{A}\gamma$-globin gene. Mol Cell Biol 17(1): 240-247.

Stamatoyannopoulos JA, G. A, Joyce T, Lowrey CH (1995). NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region. EMBO J 14(1): 106-116.

Stanworth, S, Roberts, NA, Sharpe, JA, Sloane-Stanley, JA, and Wood, WG (1995). Established epigenetic modifications determine the expression of developmentally regulated globin genes in somatic hybrids. Mol Cel Biol 15(8): 3969-3978.

Starck, J, Sarkar, R, Romana, M, Bhargava, A, Scarpa, AL, Tanaka, M, Chamberlain, JW, Weissman, SM, and Forget, BG (1994). Developmental regulation of human γ – and β–globin genes in the absence of the locus control region. Blood 84: 1656-1665.

Stief A, W. D, Stratling WH, Sippel AE (1989). A nuclear DNA attachment element mediates elevated and position-independent gene activity. Nature 341(6240): 343-345.

Still, I, Vince, P, and Cowell, JK (1997). Direct isolation of human transcribed sequences from yeast artificial chromosomes through the application of RNA fingerprinting. Proc Natl Acad Sci USA 94: 10373-10378.

Stinnakre, MG, Vilotte, JL, Soulier, S, and Mercier, JC (1994). Creation and phenotypic analysis of alpha-lactalbumin-deficient mice. Proc Natl Acad Sci U S A 91(14): 6544-6548.

Stratling WH, Dolle, A, and Sippel, AE (1986). Chromatin structure of the chicken lysozyme gene domain as determined by chromatin fractionation and micrococcal nuclease digestion. Biochemistry 25(2): 495-502.

Strauss, W, Dausman, J, Beard, C, Johnson, C, Lawrence, JB, and Jaenisch, R (1993). Germ line transmission of a yeast artificial chromosome spanning the murine α_1(I) collagen locus. Science 259: 1904-1907.

Streuli, C, Bailey, N, and Bissell, MJ (1991). Control of mammary epithelial differentiation: basement membrane induces tissue-specific gene expression in the absence of cell-cell interaction and morphological polarity. J Cell Biol *115*, 1383-1395.

Strouboulis, J, Dillon, N, and Grosveld, F (1992). Developmental regulation of a complete 70-kb human β-globin locus in transgenic mice. Genes Dev 6: 1857-1864.

Sullivan, R, Fitzpatrick, MM, and Stanton, EK (1959). Distribution of kappa-casein in skim milk. Nature (Lond) 183: 616.

Sutcliffe, J (1979). Complete nucleotide sequence of the *Escherischia coli* plasmid pBR322. Cold Spring Harb. Symp. Quant. Biol. 43: 77-90.

Takeda, H, Yamakuchi, H, Ihara, N, Hara, K, Watanabe, T, Sugimoto, Y, Oshiro, T, Kishine, H, Kano, Y, and Kohno, K (1998). Construction of a bovine yeast artificial chromosome (YAC) library. Animal Genetics 29: 216-219.

Taketani, Y, and Oka. T (1983). Epidermal growth factor stimulates cell proliferation and inhibits functional differentiation of mouse mammary epithelial cells in culture. Endocrinology 113(3): 871-877.

Talbot, D, Decombes, P, and Schibler, U (1994). The 5' flanking region of the rat LAP gene can direct high-level, position independent, copy number dependent expression in multiple tissues in transgenic mice. Nucleic Acids Res 22: 756-766.

Thepot, D, Devinoy, E, Fontaine, ML, Stinnakre, MG, Massoud, M, Kann, G, and Houdebine, LM (1995). Rabbit whey acidic protein gene upstream region controls high-level expression of bovine growth hormone in the mammary gland of transgenic mice. Mol Reprod Dev 42(3): 261-267.

Thompson, M, Dave, JR, and Nakhasi, HL (1985). Molecular cloning of mouse mammary gland κ-casein: Comparison with rat κ-casein and rat and human γ-fibrinogen. DNA *4*, 263-271.

Thompson, S, Clark, AR, Pow, AM, Hooper, ML, and Melton, DW (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell 56(2): 313-321.

Thorey, I, Ceceña, G, Reynolds, W, and Oshima, RG (1993). *Alu* sequence involvement in transcriptional insulation of the keratin 18 gene in transgenic mice. Mol Cell Biol 13(11): 6742-6751.

Threadgill, D, and Womack, JE (1990). Genomic analysis of the major bovine milk protein genes. Nucleic Acids Res 18(23): 6935-6942.

Tomlinson, A, Cox, RD, Lehrach, HR, and Dalrymple, MA (1996). Restriction map of two yeast artificial chromosomes spanning the murine casein locus. Mamm Genome *7*, 542-544.

Topper, Y, and Freeman, CS (1980). Multiple hormone interactions in the developmental biology of the mammary gland. Physiol Rev *60*, 1049-1106.

Trimborn, T, Gribnau, J, Grosveld, F, and Fraser, P (1999). Mechanisms of developmental control of transcription in the murine α- and β-globin loci. Genes & Dev 13: 112-124.

Tuan DY, S. W, London IM, Lee DP (1989). An erythroid-specific, developmental-stage-independent enhancer far upstream of the human beta-like globin genes. Proc Natl Acad Sci U S A 86(8): 2554-2558.

Tucker, R, and Burke, DT (1997). Two vectors for the insertion of mammalian selectable genes into yeast artificial chromosome cloned DNA. Gene 199: 25-30.

Udvardy, A (1999). Dividing the empire: boundary chromatin elements delimit the territory of enhancers. EMBO J 18(1): 1-8.

Uusi-Oukari, M, Hyttinen, J, Korhonen, V, Västi, A, Alhonen, L, Jänne, OA, and Jänne, J (1997). Bovine α_{s1}-casein gene sequences direct high level expression of human granulocyte-macrophage colony-stimulating factor in the milk of transgenic mice. Transgenic Res *6*, 75-84.

Vassilopoulos, G, Navas, PA, Skarpidi, E, Peterson, KR, Lowrey, CH, Papayannopoulou, T, and Stamatoyannopoulos, G (1999). Correct function of the LCR may require passage through a non-erythroid cellular environment. Blood 93(2): 703-712.

Velander, WH, Johnson, JL, Page, RL, Russell, CG, Subramanian, A, Wilkins, TD, Gwazdauskas, FC, Pittius, C, and Drohan, WN (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 U S A 89(24): 12003-12007.

Velmala, R, Vilkki, J, Elo, K, and Mäki-Tanila, A (1995). Casein haplotypes and their association with milk production traits in the Finnish Ayrshire cattle. Anim Genet 26: 419-425.

Vilotte, J, Soulier, S, Stinnakre, MG, Massoud, M, and Mercier, JC (1989). Efficient tissue-specific expression of bovine alpha-lactalbumin in transgenic mice. Eur J Biochem 186: 43-48.

Wakao, H, Gouilleux, F, and Groner, B (1994). Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor family and confers the prolactin response. EMBO J 13(9): 182-191.

Wakao, H, Schmitt-Ney, M and Groner, B (1992). Mammary gland-specific nuclear factor is present in lactating rodent and bovine mammary tissue and composed of a single polypeptide of 89 kDa. J Biol Chem *267*, 16365-16370.

Wakayama T, P. A, Zuccotti M, Johnson KR, Yanagimachi R (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. Nature 394(6691): 369-374.

Wakimoto, B, and Hearn, MG (1990). The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of *Drosophila melanogaster*. Genetics 125: 141-154.

Wall, R, Pursel, VG, Shamay, A, McKnight, RA, Pittius, CW, and Hennighausen, L (1991). High-level synthesis of a heterologous milk protein in the mammary glands of transgenic swine. Proc Natl Acad Sci USA *88*, 1696-1700.

Wall, R. J, and Seidel, G.E. Jr (1992). Transgenic farm animals - a critical analysis. Theriogenology 38: 337-357.

Walters, M, Fiering, S, Eidemiller, J, Magis, W, Groudine, M and Martin, DIK (1995). Enhancers increase the probability but not the level of gene expression. Proc Natl Acad Sci USA 92: 7125-7129.

Walters, MC, Magis, W, Fiering, S, Eidemiller, J, Scalzo, D, Groudine, M, and Martin, DI (1996). Transcriptional enhancers act in cis to suppress position-effect variegation. Genes Dev 10(2): 185-195.

Ward, T, Honeycutt, RL, and Derr, JN (1997). Nucleotide sequence evolution at the κ-casein locus: Evidence for positive selection within the family bovidae. Genetics 147: 1863-1872.

Watson, C, Gordon, KE, Robertson, M, and Clark, AJ (1991). Interaction of DNA-binding proteins with a milk protein gene promotor in vitro: identification of a mammary gland-specific factor. Nucl Acids Rev *19*, 6603-6610.

Webster, J, Wallace, RM, Clark, AJ, and Whitelaw, CBA (1995). Tissue-specific, temporally regulated expression mediated by the proximal ovine β-lactoglobulin promoter in transgenic mice. Cell Mol Biol Res 41(1): 11-15.

Wells DN, M. P, Day TA, Tervit HR (19971). Production of cloned lambs from an established embryonic cell line: a comparison between in vivo- and in vitro-matured cytoplasts. Biol Reprod 57(2): 385-393.

Wheeler, M (1994). Development and validation of swine embryonic stem cells: a review. Reprod Fertil Dev 6(5): 563-568.

Whitelaw, C, Harris, S, McClenaghan, M, Simons, JP, and Clark, AJ (1992). Position-independent expression of the ovine β-lactoglobulin gene in transgenic mice. Biochem J 286: 31-39.

Whitelaw, C, Archibald, AL, Harris, S, McClenaghan, M, Simons, JP, and Clark, AJ (1991). Targeting expression to the mammary gland: intronic sequences can enhance the efficiency of gene expression in transgenic mice. Transgenic Res 1: 3-13.

Wijgerde M, G. F, Fraser P (1995). Transcription complex stability and chromatin dynamics in vivo. Nature 377(6546): 209-213.

Willadsen, SM (1979). A method for culture of micromanipulated sheep embryos and its use to produce monozygotic twins. Nature 277(5694): 298-300.

Wilmut I, S. A, McWhir J, Kind AJ, Campbell KH (1997). Viable offspring derived from fetal and adult mammalian cells. Nature 385(6619): 810-813.

Wilson, C, Belen, HJ, and Gehring, WJ (1990). Position effects on eukaryotic gene expression. Annu Rev Cell Biol 6: 679-714.

Wolf, E, Zakhartchenko, V, and Brem, G (1998). Nuclear transfer in animals: Recent developments and future perspectives. J Biotechnol 65: 99-110.

Wood, W (1996). The complexities of β globin gene regulation. TIG 12(6): 204-206.

Wright G, C. A, Cottom D, Reeves D, Scott A, Simons P, Wilmut I, Garner I, Colman A (1991). High level expression of active human alpha-1-antitrypsin in the milk of transgenic sheep. Biotechnology (NY) 9(9): 830-834.

Yu-Lee, L, Richter-Mann, L, Couch, CH, Stewart, AF, Mackinlay, AG, and Rosen, JM (1986). Evolution of the casein multigene family: conserved sequences in the 5' flanking and exon regions. Nuc Acids Res *14*, 1883-1902.

Yu-Lee, L, and Rosens, JM (1983). The rat casein multigene family. I. Fine structure of the gamma-casein gene. J Biol Chem *258*, 10794-10804.

Yu SH, D. K, Lee E, Hennighausen L, Sweet RW, Rosenberg M, Westphal H (1989). Functional human CD4 protein produced in milk of transgenic mice. Mol Biol Med 6(4): 255-261.

Yull, F, Binas, B, Harold, G, Wallace, R, and Clark, AJ (1997). Transgene rescue in the mammalian gland is associated with transcription but not require translation of BLG transgenes. Trans Res 6: 11-17.

Yull FE, W. R, Clark AJ (1995). Restricted tissue-specific but correct developmental expression mediated by a short human alpha 1AT promoter fragment in transgenic mice. Transgenic Res 4(1): 70-74.

Zhang, Y, Cao, H, and Ferretti, L (1998). IDVGA75 (D15S24) a polymorphic microsatellite locus physically mapped to bovine chromosome BTA 15q25. Anim Genet 29(6): 472.

Zhao, K, Hart, CM, and Laemmli, UK (1995). Visualization of chromosomal domains with boundary element-associated factor BEAF-32A. Cell 81: 879-889.

Appendix I

Publication in the journal Mammalian Genome

Mammalian Genome 7, 542–544 (1996).

Mammalian Genome

Restriction map of two yeast artificial chromosomes spanning the murine casein locus

A.M. Tomlinson,[1] R.D. Cox,[2] H.R. Lehrach,[3] M.A. Dalrymple[1]

[1]PPL Therapeutics (Scotland) Ltd., Roslin, Midlothian, EH25 9PP, UK
[2]Physical Mapping and Gene Identification, University of Oxford, the Wellcome Trust Centre for Human Genetics, Windmill Road, Headington Oxford OX3 7BN, UK
[3]Genome Analysis, Imperial Cancer Research Fund, Lincoln's Inn Fields, London, WC2A 3PX, UK

Received: 19 January 1996 / Accepted: 1 March 1996

The caseins are the major mammalian milk proteins (reviewed by Mercier and Vilotte 1993), constituting a dietary supply of amino acids and calcium to the infant. In the presence of calcium, the so-called calcium-sensitive caseins of the α and β types form loose, non-crystalline aggregates termed micelles, which are stabilized by calcium-insensitive κ casein. In the gut the site-specific cleavage of κ casein by rennin causes the milk to clot and remain in the stomach, facilitating digestion. The caseins are encoded by a small gene family, which in cows and sheep consists of four members, α_{s1}, β, α_{s2}, and κ, and in mice and rats five, α, β, γ, ϵ, and κ (Yu-Lee and Rosen 1983; Jones *et al.* 1985; Thompson *et al.* 1985). The casein genes all map to a single chromosome in rodents, sheep, cows, humans, and pigs (reviewed by Mercier and Villotte, 1993) and are very tightly linked genetically. All four bovine casein genes have been mapped to a single 250-kb locus (Ferretti *et al.* 1990; Threadgill and Womack 1990). This close linkage might be expected in the case of the evolutionarily closely related calcium-sensitive caseins, but there is no evidence that κ casein is evolutionarily related to the other caseins. Both in sequence homology and protein function it appears to be related to γ fibrinogen (Jolles *et al.* 1974; Thompson *et al.* 1985; Alexander *et al.* 1988), which performs a cleavage-induced clotting function in blood similar to the clotting function of κ casein in the stomach. Therefore, the proximity of the bovine κ casein gene to the other casein genes is noteworthy.

Two yeast artificial chromosome (YAC) clones bearing the five murine caseins were obtained by screening the Imperial Cancer Research Fund mouse genomic YAC library, primarily with two genomic probes for murine β casein as described (Cox *et al.* 1993) and secondarily by screening β casein-positive clones with cDNA probes for murine α, γ, ϵ, and κ casein, by colony hybridization and pulsed field gel electrophoresis (PFGE). The two clones are ICRFy902G0781, renamed MP12, and ICRFy902C11116, renamed MP14. MP14 is approximately 380 kb in size, and MP12 is approximately 435 kb in size, as determined by PFGE and Southern blotting (data not shown).

A restriction map of the two YACs was generated by partial digestion with *Pme*I, *Sal*I, *Xho*I and *Cla*I, and probing a single filter with probes specific to the left and right arms of pYAC4 (LA and RA; Fig. 1). The restriction fragments detected by LA and RA are listed in Table 1. When the MP12 and MP14 *Pme*I band patterns are represented as maps, the two maps can be aligned so that the 121-kb MP14 band is lined up with the 105-kb MP12 band. The alignment is almost perfect over the region of overlap, varying by less than 5 kb. By this analysis, the 26-kb MP14 band should fall some 10 kb from the left telomere of MP12 and therefore give a band in the MP12 lanes, but, given that there are 6 kb of vector sequence at this end, it is possible, within the accuracy of the blot below 48.5 kb, that MP12 does not contain this *Pme*I site. Because of the gel conditions and the markers used in this gel (λ concatemers), the accuracy of the blot is compromised below 48.5 kb.

The *Sal*I, *Xho*I and *Cla*I patterns were treated the same way, to

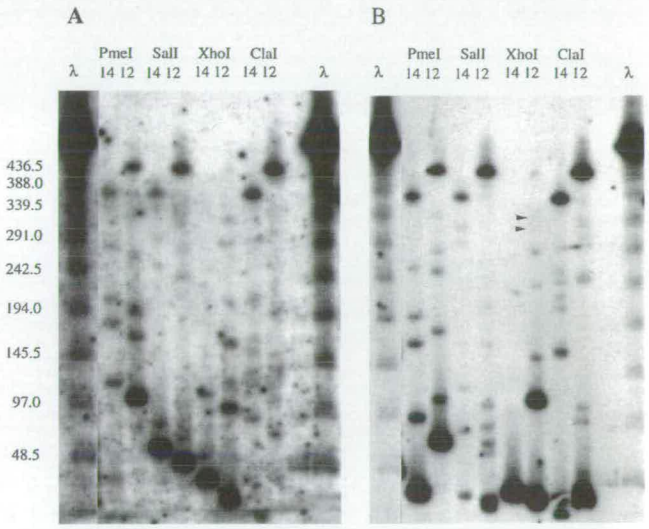


Fig. 1. Southern blot analysis of MP14 and MP12. Yeast genomic DNA was prepared in 0.8% InCert agarose and digested with restriction enzymes essentially according to Hamvas *et al.* (1994), with the exception that 20 units of each enzyme were used in a volume of 200 μl, reactions were carried out for 30 min for partial digests and overnight for complete digests, and reactions were stopped by placing on ice. PFGE was carried out in a CHEF DRII apparatus (Bio-Rad) with 1% Rapid Agarose (Life Technologies) cast and run in 0.5× TBE buffer. Running conditions were 200 V, 14°C, with switching ramped linearly from 5.0 s to 30.0 s. Gels were stained, photographed, blotted, probed, and stripped according to standard procedures. 1200 mJ/cm^2 UV (254 nm) light were used to nick the DNA before blotting. The probes for the left (LA; panel A) and right (RA; panel B) arms of the YACs were the two fragments generated by digesting pBR322 with *Pvu*II and *Eco*RI. β casein probes were generated by polymerase chain reaction (PCR) from genomic murine DNA, with the following primers (upper, then lower): 5′ (MβC5′): GACTTGACAGCCATGAAGGTCT and CGACGTTTATGGAGTCTCCTTC; 3′ (MβC3′), GGAGACACTCCTTAAGAACACT and CAGATTCTCCTAAAGGTCCAAT. Other casein probes were generated from murine lactating mammary gland RNA by reverse transcriptase PCR (RTPCR) with the following primers: α casein, GCAGTTCGCAGTCAAACTCAGC and CACTGACCTGGGAGGTAAGAGG; γ casein, AGCAATACATCTCCAGTGAGGA and TGACAGAAGTGAAGACGAGGGT; κ casein, CAGATTCAAACTGCCGTGGTGA and GTCTAGAAAGAGCAGAAGGGAA.

Correspondence to: M. Dalrymple

Table 1. Restriction fragments detected by LA and RA probes.

Probe	YAC	*Pme*I	*Sal*I	*Xho*I[b]	*Cla*I
LA	MP14	26[a], 121, 179, 206, 279, 350[c], 381	59, 85, 251, 293, 350, 381	28[a], 114	23[ac], 92, 124, 137, 154, 167, 210, 382
	MP12	105, 164, 193, 265, 336, 381, 437	41[a], 70, 236, 276, 343, 437	3[a], 102, 124, 159, 206, 286, 327	78, 109, 122, 139, 152, 195, 437
RA	MP14	22[a], 98, 164, 196, 251, 365[c], 381	20[a], 88, 126, 293, 322, 381	32[a]	3[a], 23, 161, 172, 206, 221, 240, 253, 289, 355[c], 382
	MP12	77, 116, 179, 249, 281, 343, 437	8[a], 65, 74, 90, 109, 168, 206, 377[c], 402[c], 437	23[a], 119, 156, 233, 283, 323[c], 351[c], 437	23[a], 100, 114, 242, 255, 295, 307, 320, 340, 370[c], 437

[a] Fragments denoted as less than 50 kb are below the limit of resolution of the gel.
[b] *Xho*I digest of MP14 was excessive, leading to incomplete partial band pattern.
[c] Bands that are faint in Fig. 1 but clearly discernible on the original autoradiograph.

generate a four-enzyme map of the locus from the LA autoradiograph. The same was done for the RA band patterns, and then the two maps were aligned about a central site; the *Pme*I site 206 kb from the left arm of MP14 was chosen arbitrarily and aligned with the *Pme*I site 249 kb from the right arm of MP12. The agreement between the two maps is almost perfect with an error of about 5–10 kb, but with the following ambiguities: the *Xho*I site at 220 kb was positioned at 230 kb by the RA map, but at 215 kb by the LA map (this was the only site that diverged to this extent between the two maps); the pattern of MP14 bands cut with *Xho*I was incomplete in this gel owing to over-digestion, but the MP12 *Xho*I digest allowed the map to be generated, with good agreement between the LA and RA band patterns; RA detected a *Cla*I doublet in MP14 and MP12, at 200 kb and 210 kb in Fig. 2, but only a single diffuse band could be discerned in the LA pattern. The RA autoradiograph is clearer than the LA autoradiograph, so this doublet is considered to be genuine. The collinearity of MP12 and MP14 precludes the possibility that either YAC is rearranged or chimeric.

The same filter was then probed with the probes for the other caseins, in order to map their locations within the YACs. Another blot was generated (not shown) from complete single and double digests to confirm the partial digest results.

α and β casein probes map to the same 65-kb *Cla*I fragment in partial digests. In complete digests, α casein maps to a 25-kb *Sal*I fragment, while MβC5′ maps to a 165-kb *Sal*I fragment. MβC3′ detects both of these fragments. Taken together, these data indicate that α casein lies between the *Sal*I sites at 60 and 85 kb in Fig. 2 and that β casein spans the *Sal*I site at 85 kb and is transcribed from right to left in the YAC.

γ casein was positioned primarily on the basis of partial *Pme*I digests, but these data alone were inconclusive. Confirmation was provided by the detection by the γ probe of identical *Pme*I and *Pme*I/*Sal*I complete digestion fragments of 65 kb; the minimal fragment detected by γ is a 15-kb *Cla*I–*Pme*I fragment which maps between the sites at 165 and 180 kb in Fig. 2.

The partial mapping data were inconclusive with respect to ϵ casein, but did allow assignment in the middle of the map. Complete *Pme*I, *Cla*I, *Pme*I/*Cla*I, PmeI/*Sal*I and *Pme*I/*Xho*I digests position ϵ on a 10-kb *Cla*I fragment between the sites at 200 and 210 kb in Fig. 2, spanning the *Pme*I site at 205 kb.

On the basis of partial *Pme*I data and complete *Xho*I, *Pme*I/*Xho*I, and *Sal*I restriction fragments, κ casein maps to a 34-kb fragment between the *Cla*I sites at 300 and 335 in Fig. 2.

The bovine genomic casein region has been restriction mapped (Ferretti *et al.* 1990; Threadgill and Womack 1990; Fig. 2). The spatial arrangement of the genes is largely conserved between the two species (bovine α_{s1}, β, α_{s2}, and κ caseins are homologous to murine α, β, ϵ, and κ caseins respectively; cattle have no known γ casein). The murine calcium-sensitive caseins are present as two pairs, separated by approximately 75 kb of intergenic DNA. The κ casein gene, although some 90–155 kb from the other genes, is nevertheless remarkably close considering its evolutionary unrelatedness. The ratio of κ to the calcium-sensitive caseins in milk is believed to play a significant role in determining micelle size, milk volume, and milk quality (Lin *et al.* 1989). As has been suggested (Alexander *et al.* 1988; Mercier and Vilotte 1993), the presence of the κ casein gene close to the other casein genes may indicate that coordination of casein gene expression is effected at the transcription level, and thus may be indicative of the existence of casein locus control.

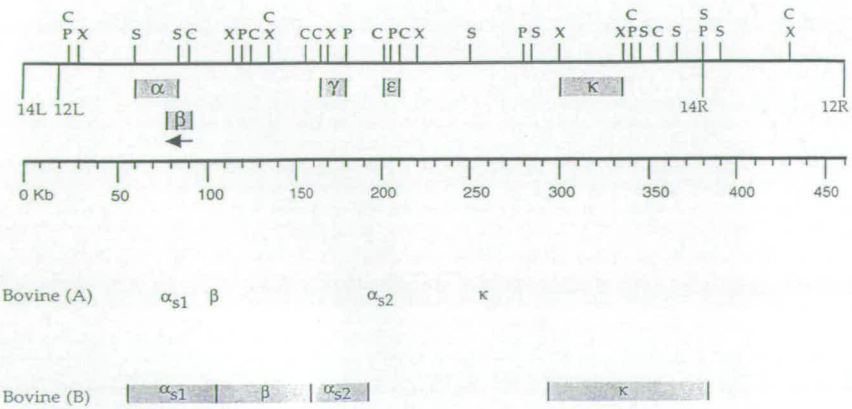


Fig. 2. Restriction map of the murine casein locus. Scale in kb is shown below the map. C, *Cla*I; P, *Pme*I; S, *Sal*I; X, *Xho*I; 14L and 14R, left and right ends of MP14; 12L and 12R, left and right ends of MP12. Casein genes are represented by Greek letters; shaded boxes delimit the minimal restriction fragment that each gene has been mapped to. Below the murine map are represented the two published maps of the bovine casein locus, to scale and aligned at α/α_{s1}. **A** from Threadgill and Womack 1990; **B** from Ferretti *et al.* 1990.

Acknowledgements. The authors would like to thank Ian Garner and Alan Colman for their support. This work was partly funded by the E.U.

References

Alexander, L.J., Stewart, A.F., Mackinlay, A.G., Kapelinskaya, T.V., Tkack, T.M., Gorodetsky, S.I. (1988). Isolation and characterization of the bovine κ-casein gene. Eur. J. Biochem. 178, 395–401.

Cox, R.D., Meier-Ewert, S., Ross, M., Larin, Z., Monaco, A.P., Lehrach, H. (1993). Genomic mapping and cloning of mutations using yeast artificial chromosomes. Methods Enzymol. 225, 637–653.

Ferretti, L., Paolo, L., Sgamarella, V. (1990). Long range restriction analysis of the bovine casein genes. Nucleic Acids Res. 18, 6829–6833.

Hamvas, R.J., Francis, F., Cox, R.D., Nizetic, D., Goldsworthy, M.E., Brown, S.M., Lehrach, H.R. (1994). Rapid restriction analysis of YAC clones. Nucleic Acids Res. 22, 1318–1319.

Jolles, J., Fiat, A.M., Schoentgen, F., Alais, C., Jolles, P. (1974). The amino acid sequence of sheep κ-A-casein. II. Sequence studies concerning the κA-casein glycopeptide and establishment of the complete primary sequence of the protein. Biochim. Biophys. Acta 365, 335.

Jones, W.K., Yu-Lee, L.Y., Clift, S.M., Brown, T.L., Rosen, J.M. (1985). The rat casein multigene family. Fine structure and evolution of the β-casein gene. J. Biol. Chem. 260, 7042–7048.

Lin, C.Y., McAllister, A.J., Ng-Kwai-Hang, K.F., Hayes, J.F., Batra, T.R., Lee, A.J., Roy, G.L., Vesely, J.A., Wauthy, J.M., Winter, K.A. (1989). Relationships of milk protein types to lifetime performance. J. Dairy Sci. 72, 3085–3090.

Mercier, J.-C., Vilotte, J.-L. (1993). Structure and function of milk protein genes. J. Diary Sci. 76, 3079–3098.

Thompson, M.D., Dave, J.R., Nakhasi, H.L. (1985). Molecular cloning of mouse mammary gland κ-casein: comparison with rat κ-casein and rat and human γ-fibrinogen. DNA 4, 263–271.

Threadgill, D.W., Womack, J.E. (1990). Genomic analysis of the major bovine milk protein genes. Nucleic Acids Res. 18, 6935–6942.

Yu-Lee, L.Y., Rosen, J. (1983). The rat casein multigene family 1. Fine structure of the γ casein gene. J. Biol. Chem. 258, 10794–10804.