Studies on the MUC1 promoter

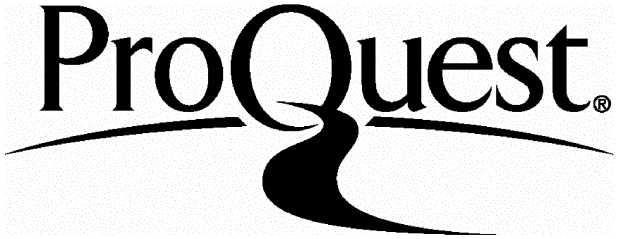
Joanna Ruth Morris

Thesis presented in partial fulfilment of the degree of Doctor of Philosophy at the University of London.

September 1999.

Breast Cancer Biology Laboratory & Molecular Endocrinology Laboratory,
Imperial Cancer Research Fund,
Lincoln's Inn Fields
London.

Department of Biology
University College London,
Gower Street,
London



All experiments included in this thesis were performed by myself, unless indicated otherwise in the text.

Abstract

The gene for human MUC1 is selectively expressed in simple epithelia lining ducts and glands of many organs and is highly expressed in many carcinomas. Results herein show that the *MUC1* 1.4 kb 5' region and cDNA are sufficient for faithful tissue specific expression, high expression levels in lactating mammary gland and in mammary tumours *in vivo*.

Transient transfection experiments show that promoter sequences -119/-62 are sufficient to direct epithelial specific transcription when cloned adjacent to the consitutively active HSV TK promoter. *In vitro* binding assays revealed that Sp1, Sp3, and a potentially novel factor, SpA, compete for DNA binding at the -97 GC box while USF and two less well characterised factors bind the sequence at -87/-75. The tissue-specific activity of the -119/-62 sequence is mediated by Sp1 and Sp3 binding activities, since a consensus Sp1 sequence cloned into the same TK-proximal site can substitute for the activity. Introduction of an Sp1 consensus elsewhere could not direct tissue specific activity, demonstrating that the position of the sequence is crucial.

Transcription from the 1.4 kb MUC1 promoter reporter plasmid and the endogenous gene could be activated in non-epithelial cells by co-transfection of Sp1, suggesting that this factor is limiting in these cells.

Functional analysis of a mutant promoter which lacked the ability to bind the SpA factor, showed an increase in transcription over wild-type. This was most marked in the epithelial cell line expressing least MUC1, MCF-7, suggesting that SpA may have a role in modulating MUC1 levels in epithelial cells. In support of this idea band-shift experiments showed that SpA levels were highest in MCF-7.

These studies suggest that the binding of Sp1 and Sp3 and SpA to the -97 GC box may have an important role in both tissue-specific transcription and regulation of MUC1 expression in carcinomas.

Acknowledgements

I would like to thank Joyce Taylor for her optimism and thoughtfulness throughout this work and Malcolm Parker for his critic and thoroughness. All the members of Joyce's and Malcolm's lab' deserve thanks for their support, resourcefulness and friendliness. I would like to thank, Dan, Pie-Juan, Ian, Ros, Caroline, Martin, Roger and Vera in particular. Thanks are also due to those who provided reagents and protocols and to the staff of the oligonucleotide synthesis department, the animal unit at Clare Hall and the photography and computing departments.

I wish to give my thanks also to my family and friends who gave encouragement, forgiveness and frequently finances throughout this work. In particular I wish to acknowledge the Herculean support of my parents, Sheila and Bruce, and Dan.

This thesis is dedicated to Mike Allinson, 1973-1998.

Table of Contents

Table of Figures.

Table of Tables

Abbreviations

β-gal	β-galactosidase
A	adenine or adenosine; one letter code for alanine
Ab	antibody
acetyl CoA	acetyl coenzyme A
Ag	antigen
AP	alkaline phosphatase
ATCC	American Type Culture Collection
ATP	adenosine 5'-triphosphate
BLAST	Basic Local Alignment Research Tool
bp	base pair
BSA	bovine serum albumin
C	cytosine or cytidine; one letter code for cysteine
C-terminal	carboxyl terminal
cAMP	adenosine 3'5'-cyclic-monophosphate
CAT	chloramphenicol acetyl transferase
cDNA	complementary DNA
Ci	Curie
CIP	calf intestine phosphatase
CMV	cytomegalovirus
cpm	counts per minute
CTD	carboxyl terminal domain
CTP	cytidine 5'-triphosphate
DAB	3'3'-diaminobenzidene
dATP	2'-adeoxyadenosine-5'-triphosphate
dCTP	2'-deoxycytidine-5'-triphosphate
DEAE	diethylaminoethylamine
dGTP	2'-deoxyguanosine-5'-triphosphate

DMSO	dimethyl sulfoxide
DNA	deoxyribonucleicacid
DNase	deoxyribonuclease
dNTP	deoxynucleoside triphosphate
ds	double stranded
DTT	dithiothreitol
dTTP	2'-deoxythymidine-5'-triphosphate
E4	Dulbecco's modified Eagle's medium
ECL	enhancedchemicalluminescence
EDTA	ethylenediaminetetraceticacid
EMBL	European Molecular Biology Laboratory
ER	oestrogen receptor
ERE	oestrogen response element
FCS	foetal calf serum
GR	glucocorticoid receptor
GTP	guanosine 5'-triphosphate
HBS	His-buffered saline
HEPES	N-2-hydroxyethylpiperazine N-2-ethansulphonic acid
HRE	hormone response element
HRP	horse radish peroxidase
HSV-tk	Herpes Simplex Virus thymidine Kinase gene
IPTG	isopropl-thio-galactopranoside
kb	kilo base
K_d	dissociation constant
kDa	kilo Dalton
LacZ	β-galactosidasegene
LTR	long terminal repeat
M	relative molecular weight
MAb	monoclonal antibody
mins	minutes

MMTV	mouse mammary tumour virus
mRNA	messenger RNA
N-terminal	amino terminal
ND	Not done
NEB	New England Biolabs
NP40	nonidet p40
O/N	over-night
OD_x	optical density at wavelength of x nm
OLB	Oligonucleotidelabellingbuffer
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
PMSF	phenylmethylsulphonyl fluoride
Poly-T	Polyomavirus middle-T antigen.
PPAR	peroxisome proliferator activated receptor
PVP	polyvinylprolidine
RAR	retinoic acid receptor
RARE	retinoic acid response element
RNA	ribonucleicacid
RNase	ribonuclease
RT	room temperature
RXR	retinoid X receptor
S	Svedberg units
SDS	sodium dodecyl sulphate
SV40	simian virus 40
T3	3,5,3'-triiodo-L-thyronine
TAF	TBP associated factor
TBE	Tris/borate buffer

TBP	TATA binding protein
TBS	Tris-buffered saline
TEMED	N'N'N'N'-tetramethylethylenediamine
TK	Thymidine Kinase promoter (from the Herpes Simplex Virus)
Tris	tris(hydroxymethyl)aminoethane
Triton X-100	octyl phenoxy polyethoxyethanol
Tween-20	polyoxyethylenesorbitan monoluarate
UTP	uridine 5'-triphosphate
UV	ultra violet

Chapter 1: Introduction

1.1 MUC1 : Genomic and protein structure.

MUC1[1] was initially identified as a component of fat globules in milk (8). It is expressed at low levels on the majority of simple secretory epithelial cells and is also highly expressed by most human carcinomas including breast, colon, ovary, pancreas, lung and stomach. There are 34,000 new cases of breast cancer a year in the UK, of which over 90% of primary and metastatic tumours express high levels of MUC1 (9, 10, 11, 12, 13). Indeed high expression of MUC1 has been reported to correlate with a poor prognosis (14). Already laboratories are experimenting with the MUC1 promoter to drive the expression of pro-drugs or toxic proteins within epithelial tumour cells (15, 16, 17). Also considerable effort is being put into evaluating formulations based on the MUC1 gene and product to prime the body's own immune system to reject tumours (18) and clinical trials are already in progress (19). Whether the expression of MUC1 in malignancy is coincidental or forms an integral part of malignant progression is unclear and further investigation into its expression is warranted.

1.1.1 Genome location and context

The *MUC1* locus is located in a gene rich area of chromosome $1q^{21}$ (20, 21). 2.6 kb upstream of the *MUC1* transcription start site is the polyadenylation site for the Thrombospondin-3 gene (22) and less than 1 kb downstream of the *MUC1* polyadenylation site lies another, unidentified gene, which gives a 13 kb transcript in most tissues (23) (see figure 1.1a).

---

[1] The MUC1 gene product has been identified and named more than once, examples include; peanut-lectin binding urinary mucin (PUM), epithelial membrane antigen (EMA), MAM-6, DF3 antigen, H23 antigen, non-penetrating glycoprotein (NPGP) polymorphic epithelial mucin (PEM) and episialin. In this thesis the Human Genome Mapping convention (1989) will be followed, and the protein referred to as MUC1 and the gene as *MUC1*. Similarly the mouse homologue is designated Muc-1 and *Muc-1*, respectively (7).

1.1.2 Structure and composition

The MUC1 cDNA was the first mucin to be cloned (24, 25, 12). Figure 1.1b and 1.1c illustrate the structure of the gene and full length protein respectively. The most striking feature is the large repeat domain, which in MUC1 is made up of variable numbers of a 20 amino acid repeat and which subsequently was shown to be a feature of all mucins. Unlike the extracellular gel-forming mucins, MUC1 lacks cysteine residues that allow the formation of disulphide bridges to generate a mucin network (reviewed in (23)). MUC1 is a type 1 transmembrane protein. The carboxy terminus of the molecule, 3' to the tandem repeat domain encodes degenerate repeats, the transmembrane domain and a cytoplasmic domain of 69 amino acids, while the amino terminal includes a signal peptide and degenerate repeats.

The tandem repeat sequence is rich in serine (10%), threonine (15%) and proline (25%) amino acid residues (see figure 1.1c) the serine and threonine residues being sites for O-glycosylation. The extent of glycosylation varies between tissue types, time in development, and tumour status. For example MUC1 in the mammary gland has a molecular weight of 250 -500 kDa, about half of which is carbohydrate, in contrast the molecular weight of MUC1 in the pancreas is greater than 1000 kDa and is about 80% carbohydrate (reviewed in (26)).

Both the sites of glycosylation and the composition of the O-glycans can vary. There are five potential sites of O-linked glycosylation per repeat. In MUC1 isolated from milk an average 2.6 sites were glycosylated per repeat (27), compared to all five glycosylated per repeat from the breast cancer cell line T47D (28). O-linked carbohydrates on breast tumour-associated MUC1 have reduced branching complexity, often terminating early with sialic acid (29). The glycosylation of MUC1 is catalysed by a family of enzymes and the composition of carbohydrate on the protein may be due to both an altered profile of the expression of glycosyl transferases and the position of the transferases in the Golgi relative to other enzymes which compete for the same substrate. In breast cancer cell lines α2,3 sialyltransferase which terminates carbohydrate chain extension by adding sialic acid to core 1 is elevated, whilst an

enzyme catalysing the formation of core 2 from core 1 leading to chain extension (core-2 β1,6 *N*-acetyl-D-glucosamine transferase), is decreased (30). Thus in breast cancer chain extension is reduced in favour of termination (illustrated in figure 1.2b). The role of α2,3 sialyltransferase in MUC1 carbohydrate shortening has been confirmed by transfection into immortalised 'normal' breast cells. An increase in sialylated structures was seen as well as a decrease in carbohydrate branching (31). As shown in figure 1.2a the reduced length of carbohydrate branches on tumour-associated MUC1 is thought to allow access of MUC1-core-specific antibodies, such as SM-3, to the underlying protein, thus providing the resulting differential detection of MUC1 between normal and tumour breast samples (32). Several novel carbohydrate epitopes are found expressed on carcinoma-associated MUC1, including the STn antigen which is being used as an immunogen in clinical trials (see section 1.4.1).

Nuclear magnetic resonance (NMR) studies have shown that the MUC1 tandem repeat protein core is present as a polyproline β-turn helix-type secondary structure. Each repeat has a hydrophilic region and two β-turn helices facing the surface of the molecule together with a hydrophobic core (33) (see figure 1.3a). Visualisation of MUC1 on cell membranes using an electron microscope has revealed that the molecule extends over 125 nm above the glycocalyx (10 nm) and most adhesion molecules (35 nm), (see figure 1.3b) (34, 35).

The GC content of the region coding for the tandem repeats is high, being composed of 82% G+C, compared to 40% observed overall in the mammalian genome. The size of the coding region can be between 4-7 kb depending on the numbers of repeats which vary between 25 to more than 125 per allele. Thus MUC1 is an expressed variable number tandem repeat (VNTR) and in Northern Europeans typical numbers of tandem repeats are between 41 and 85. The theoretical mass of the protein core is between 120 to 225 kDa, typically of between 800 to 1700 amino acids (24).

Although MUC1 is a transmembrane protein a soluble form exists in tissue culture supernatants and bodily fluids (36, 37). During intracellular processing it appears that proteolytic cleavage occurs yielding a short C-terminal peptide containing the cytoplasmic tail, transmembrane region and small section of extracellular domain and the

large tandem repeat-containing domain. However the small fragment and the rest of the protein do not dissociate. They remain together as a non-covalently bound bi-peptide structure and are expressed as such on the surface of the cell (figure 1.3b) (38). The cleavage site has been mapped to an 18 amino acid region, just 53 amino acids N-terminal of the transmembrane domain and deletion of the site results in a mutant protein which is not cleaved. The cleavage sequence corresponds to a putative substrate for the kallikrein family of serine proteases. What mechanism is responsible for the release of the ectodomain from its associated transmembrane domain to form a secreted product is not known, neither is the biological reason for such an unusual complex. Finally, several isoforms of MUC1 have been reported which are thought to arise from alternative splicing (see figure 1.3c) (39). From a limited number of studies it would appear that these forms are not often expressed since even when experimental conditions have allowed detection of these variants they have rarely been found (26, 13, 25). Work from Daniel Wreschner's group has shown that MUC1/Y protein lacking the tandem repeat domain is expressed in breast and ovarian cancer cell lines and breast cancer tissues. Like the full length protein MUC1/Y expression was low in normal breast tissues. However, the relative amounts of MUC1/Y to the full length is not clear (40, 39). The subtleties of MUC1 expression may have been missed by the majority of workers because of the way its expression has been studied. Both antibody and northern detection methods have frequently used reagents to the tandem repeat and even when antibodies to the cytoplasmic region have been used, SDS-PAGE gels have generally been run to the extent that small proteins, such as the 42 kDa size of MUC1/Y, are run off. The MUC1/SEC protein is similar is size and characteristics to the MUC1 ectodomain that it may have been mistaken for it. There are no data presently on variants of Muc-1 in the mouse. What contribution each variant makes to the MUC1 pool is not yet clear, neither is the expression pattern of these variants, nor their biological role.

1.1.3 MUC4 has structural similarities to MUC1

The genomic sequence of MUC4 has recently been determined (41). The gene product has striking similarities with MUC1, as well as the potentially O-glycosylated polymorphic tandem repeat it is a transmembrane molecule. MUC4 is homologous to

the rat protein Ascites Sialoglycoprotein-1 (ASGP-1), a glycoprotein which in a manner analogous to MUC1 has been shown to be cleaved and re-associated on the cell surface as a heterodimeric complex (42).

1.2 MUC1 gene expression.

1.2.1 MUC1 is found in many species

The antibody, CT-1, raised to the cytoplasmic domain of human MUC1 cross-reacts with mucin-1 homologues of other species. It has enabled immunohistochemical detection of the protein in many different mammals including rhesus monkey, cow, human, mouse, guinea pig, cat, dog, horse and goat, whereas Xenopus (amphibian) and Dictyostelium (protozoa) are negative for the protein. The presence / absence of a mucin-1 like gene was also confirmed at the genomic level (43).

Alignment of MUC1 tandem repeat sequences from human, cow and mouse indicate poor conservation overall, but show a persistence of serines, threonines and prolines. This suggests that selection pressure was for a repeated O-linked structure rather than the amino acid sequence. In all species examined, apart from rodents, the protein appeared polymorphic (43, 44).

The mouse homologue differs from human MUC1 in that it is smaller, with a fixed number of 16 tandem repeats (44). Homology with human amino acid sequence is only 34% which is largely due to the lack of conservation within the repeats themselves, over the transmembrane and cytoplasmic domains it is 87% homologous. 72% of nucleotides over the promoter region of MUC1 and Muc-1 (-659/+16) are identical and indeed the mouse and human genes show similar patterns of expression (see figure 1.6 and table 1) (45, 46, 47). The degree of homology over the protein and promoter suggests that the function of the protein has been conserved in the evolutionary time since the divergence of mice and men.

1.2.2 Expression Profile

MUC1 is found on the apical surface of simple epithelial cells of ducts and glands of many organs. Connective and fat tissue, smooth and striated muscle, blood vessels,

cartilage and bone have been reported to be negative with many antibodies to MUC1 (see table 1) (48). MUC1 cDNA clones have been obtained from mammary, pancreatic and ovarian mRNA. Once cells lose their polarity, as in the case of less differentiated tumour cells, MUC1 can be detected all over the surface of the cell (49).

In normal polarised uterine epithelial cells the half-life of total membrane-associated glycosylated MUC1 is 16.5±0.8 hours. 95% of MUC1 made finds its way to the apical surface and 34% of this is released as ectodomains in a 24 hour period (50).

In the human, mouse, and the transgenic mouse carrying a 10.6 kb MUC1 genomic fragment, MUC1 and Muc-1 show similar patterns of tissue expression (see table 1 and figure 1.1a). Importantly in all three systems MUC1/Muc-1 also showed high expression in the lactating mammary gland and in mammary tumours, suggesting that control of expression is also conserved (13, 43, 46, 51, 52).

Table 1. Expression profile of MUC1 and Muc-1

(ND = Not Done)

Tissue type	Endogenous human MUC1(48)	Endogenous mouse Muc-1(43)	Human MUC1 genomic fragment in mice (46)
Mammary gland			
Acini	+	+	+
Ducts	+	+	+
Epidermis	-	(ND)	(ND)
Salivary glands			
Serous acini	+	+	+
Mucous acini	+	+	+
Ducts	+	+	+
Esophagus			
Squamous epithelium	+	+	-
Stomach			
Mucus secreting cells	-	(ND)	+
Parietal cells	+	+	+
Peptic cells	+	+	+
Pancreas			
Acini	+	+	+ (male only)
Ducts	+	+	+
Islets of Langerhans	-	-	-
Liver			
Hepatocytes	-	-	-
Bile ducts	+	+	-
Large intestine			
Enterocytes	-	-	(ND)
Duodenum			
Enterocytes	-	-	-
Brunner's glands	-	-	(ND)
Thyroid			
Follicle epithelium	+	(ND)	(ND)
Lung			
Respiratory epithelium	+	(ND)	+
Ciliated epithelium	+	(ND)	+ (Clara cells)
Serous bronchial glands	+	+	+ (pneumocytes)
Mucous bronchial glands	+	+	+
Kidney			
Glomeruli	-	-	-
Proximal tubules	-	-	-
Distal tubules	+	+	+
Collecting ducts	+	+	+
Bladder			
Urothelium	+	(ND)	(ND)
Prostate gland	+	-	(ND)
Uterus			
Endometrium (resting)	+	+	+
Ovary			
Oocytes	-	-	-
Follicular epithelium	-	(ND)	-
Surface epithelium	-	-	-
Vagina	(ND)	+	(ND)
Cervix	(ND)	+	+
Fallopian tubes	(ND)	+	+
Testis			

Tubuli contorti	-	-	(ND)
Rete testis	-	-	(ND)

1.2.3 Developmental expression

The timing of Muc-1 expression in mice shows that it is closely associated with epithelial differentiation, however since mice lacking the gene develop normally it is not necessary for epithelial development (52). Muc-1 is first detected at gestational day 12 in embryonic stomach, pancreas, and lung. In all cases the onset of Muc-1 expression correlated with epithelial differentiation of each individual organ (see figure 1.4a). Indeed it can be detected lining the apical surfaces of the developing lumens when the epithelium is still undergoing folding and branching and glandular activity has not yet started, suggesting that Muc-1 may be an early event in epithelial sheet differentiation (47). In organs that are subject to active and inactive phases, such as the mammary gland and endometrium Muc-1 tends to increase during cellular proliferation (53).

1.2.4 Regulated expression of MUC1

Once organogenesis is complete MUC1 is thought to be constitutively expressed on the apical surface of simple epithelial cells. However there are examples of regulation during the normal course of life. In the lactating mammary gland and endometrium changes in MUC1 expression suggests a regulatory role for hormones. MUC1 is also expressed in the wound response, and in activated T-lymphocytes. While tissue culture techniques have revealed as yet unidentified factors in normal colon conditioned media (NCCM) and charcoal stripped calf serum that are able to increase levels of MUC1 mRNA. Such instances of increased or decreased MUC1 expression suggests another level of control over and above that of constitutive expression in epithelial cells.

a) In mice, pigs and rats high levels of uterine Muc-1 occur in estrus and pro-estrus when estrogen is at a maximum. Lowest levels of Muc-1 occur during implantation (apparently down regulated by progesterone) suggesting that Muc-1 acts as a barrier to implantation (54, 55). In humans the regulation of MUC1 protein expression in

the endometrium differs in that MUC1 mRNA levels are low in the proliferative phase and increase during the secretory phase, with most expression in the implantation phase, suggesting some regulatory role for progesterone and ruling out MUC1 expression as a barrier of implantation in humans (56). Changes in the type of carbohydrate attached to MUC1 in the luminal endometrium have been noted at the time of implantation, leading to the suggestion that these may encourage implantation or lessen the barrier function of MUC1 (57, 58). Two groups have found that the level of MUC1 mRNA is not regulated by progesterone or estrogen in various cell types including endometrial (59, 60) although a third study found that MUC1 in MCF-7 and ZR75 cells was increased by progesterone but not estrogen (61).

b) High expression of MUC1 at lactation suggests that lactogenic hormones such as prolactin may be involved in its regulation. The MUC1 and Muc-1 promoter sequences have a signal transducer and activator of transcription site (STAT site), and STAT-5 is one downstream effector of prolactin-receptor/JAK2 activation. MUC1 can also be regulated by IL-6 through STAT-3 (62). In mouse mammary cells insulin plus hydrocortisone or prolactin increased MUC1 mRNA (63). Furthermore, like milk proteins, increased expression of Muc-1 could be achieved by culture of cells on a basement membrane (63). It is likely therefore that MUC1 is regulated by prolactin. (It is interesting at this point to note that prolactin, the prolactin receptor and high levels of STAT-5 have been reported to be present in breast cancer cells (64, 65) raising the possibility that an autocrine loop exists and that MUC1 expression may be one outcome of it. However this idea has been challenged by the observation that the MUC1 promoter mutated in the STAT site is as active as wild type when transfected into the T47D breast carcinoma cell line. In these cells at least STAT-5 activation cannot explain the high levels of MUC1 expression (66)).

c) Although MUC1 is considered a marker for epithelial cells there are reports of MUC1 on non-epithelial cells. Normal epidermis is generally negative for MUC1. However it may become positive in cases of inflammation or tumour invasion (48) where it may serve to prevent infection of the site.

d) T-cell lymphomas and myelomas have been found to express MUC1 epitopes. 37% of lymphoid malignancies studied expressed MUC1 in a few cells (67). MUC1 has been reported to be expressed on T-cells after stimulation with phytohemagglutinin or infection with T-cell leukaemia virus or Epstein Barr virus (68). This expression is not accompanied by keratin expression so it is thought that in these cases the expression of MUC1 is a T-cell activation-related phenomenon, and does not suggest a relationship with epithelial cells (67, 48). This observation is important in the context of implementing MUC1 immunotherapies and gene therapies.

e) Botti *et al* (1997), noted that increasing amounts of charcoal stripped serum resulted in a corresponding increase in the levels of MUC1 mRNA. The authors propose that growth-factors such as Insulin-like Growth Factor-1 (IGF-1) maybe involved (60).

f) Colorectal carcinomas express high levels of MUC1 compared to normal epithelial cells of the intestine. A novel, soluble factor present in normal colon conditioned media (NCCM) can elevate MUC1 mRNA and the activity of chloramphenicol acetyltransferase (CAT) reporter gene driven by the MUC1 promoter region. The sequences between -531/-513 were found to confer this activity (this region corresponds to a DNAse I footprint seen by other authors (69), see figure 1.8). In band-shift analysis an oligonucleotide corresponding to this region binds factors whose properties are consistent with Sp1 and Sp3 transcription factors. No difference in binding of these factors was observed between extracts treated with NCCM, and those not (70, 71).

1.2.5 MUC1 expression in tumours

MUC1 is highly expressed in carcinomas of the pancreas, ovary, lung and colon and is particularly prominent in breast tumour cells (25, 14, 10). As measured by immunohistochemistry with several different MUC1 specific antibodies it is found in more than 90% of breast tumours where it is found expressed all over the surface of the cell, in the cytoplasm and on the apical surface of carcinoma cells (10, 12, 13, 72). By *in situ* hybridisation MUC1 mRNA is highly expressed in breast tumours compared to adjacent normal breast tissue (34) and breast cancer cell lines express high levels of

MUC1, compared to the SV40 large-T antigen-immortalised 'normal' luminal breast cell, MTSV1-7 (73). RNA analysis in three studies of tumour and normal breast tissue has shown that tumours are enriched an average of 18 fold more for MUC1 mRNA (± 14, n= 10 breast tissues + 13 tumour tissues) (13, 12, 74). This type of quantification is difficult because the fold increase in MUC1 mRNA observed may reflect the presence of non-epithelial cells in the normal tissue samples compared to the large number of epithelial cells in the tumour sample. There is a need for quantification that allows for the different amounts of luminal epithelia between normal and tumour samples

The chromosomal location of MUC1, $1q^{21}$, undergoes complex re-arrangements in tumours. An overall loss of heterozygosity (LOH) in primary tumours compared to normal tissue has been seen in some studies (75) whilst others report a gain of an allele (76, 77), and still others see an even loss and gain of the $1q^{21}$ region in different patients (78, 79). Gene dosage cannot therefore be ruled out as a possible cause of some of the increase of MUC1 seen in tumours.

1.3 The biological significance of MUC1.

While this thesis examines the regulation of the expression of MUC-1, it would not be complete without addressing the awkward question, "*What does MUC1 do ?*". The biological significance of MUC1 is poorly understood, nevertheless there is no shortage of tantalising observations that have led to theorising about its role in both normal cells and in tumour progression. These functions are in several areas; physical protection, adhesion, and signal transduction, both as a receptor and as a ligand.

1.3.1 The normal function of MUC1

a) A role that is generally agreed upon is providing physical protection to the cell (34, 80). MUC1 provides an initial defence to shear force and infection (see figure 1.4b), for example one would expect MUC1 to be a factor in preventing infection of the

mammary gland (mastitis). This role is thought to extend to the stomach of the new-born where mucins prevent *Escherichia coli* from binding to the epithelium of breast fed infants. MUC1 may also imitate receptor structures for numerous viruses (81). In Portugal a population study found a correlation between small MUC1 genotypes and helicobacter-pylori infection and gastric carcinoma (82).

b) Recently a body of evidence has built up to suggest that the MUC1 cytoplasmic domain may be part of a signalling cascade from outside to inside the cell,

- The cytoplasmic domain is highly conserved between mammalian species suggesting a conserved role (43).
- Evidence has recently been published that suggests interaction of MUC1/Y with MUC1/SEC may result in phosphorylation of the cytoplasmic domain and changes in cell morphology (83). These findings maybe significant if they can be shown to be the case for the more predominant full length protein.
- The tyrosine phosphorylated cytoplasmic region of the full length MUC1 protein can interact with the SH2 domain of Grb2 protein and through it to Sos, the exchange factor for Ras (84, 85) (see figure 1.5).
- Immunoprecipitation and competition experiments have demonstrated that the serine phosphorylated cytoplasmic region of the full length protein can interact with β-catenin and Glycogen Synthase Kinase 3β (GSK3β) (84, 86). Similar experiments showed that phosphorylation of the MUC1 cytoplasmic tail by GSK3β inhibits its binding to β-catenin. However, the role of MUC1 in the Wnt pathway is unclear because no changes in β-catenin bound to APC or LEF/TCF transcriptional ability were seen when MUC1 was over expressed (see figure 1.5). Instead it seems that MUC1 may act as a competitor with E-cadherin for β-catenin as MUC1 expression correlates with decreased binding of β-catenin to E-cadherin, whereas co-expression of MUC1 and GSK3β restores β-catenin to E-cadherin (figure 1.5).

c) Findings regarding the role of MUC1 in adhesion have sometimes appeared contradictory. MUC1 is a large rod-like molecule, extending 5-10 times further than

the majority of adhesion molecules. In general then researchers have found that MUC1 acts as an anti-adhesive, preventing contacts between cells or between cells and the underlying matrix. However, since particular carbohydrate residues can also act as ligands for certain lectin molecules, in specific situations, where the correct carbohydrate moiety is expressed MUC1 may also act to encourage adhesion. MUC1s anti-adhesive properties appear to be achieved by preventing the interaction of adhesion molecules. For example, cells transfected with MUC1 bearing more than eight tandem repeats could over-ride the cell:cell interactions of E-cadherin, whereas MUC1 bearing just three repeats could not. This effect was not significantly altered by the removal of negatively charged sialic acid groups, suggesting that it is the steric hindrance provided by the sheer size of MUC1 which is the dominant force (35, 87). In an elegant experiment Wesseling *et al.* showed that MUC1 could also inhibit integrin-mediated cell adhesion to the extracellular matrix. The inhibition could be relieved by redistributing MUC1 into patches or caps using monoclonal antibodies (88). Although no α-catenin was detected in MUC1/β-catenin complexes there is likely to be an interaction between MUC1 and the cytoskeleton. The evidence for this comes from Parry *et al.* who showed that MUC1 could be diffused by treating cells with cytochalasin D which depolymerizes actin microfilaments (89).

Given even these observations one might expect MUC1 to be vital in the development of epithelial sheets/lumen or perhaps be a significant part of a transduction pathway. Not so. Muc-1 null mice had very few defects, developed normally, were healthy and fertile and appeared phenotypically normal in all respects. Clearly Muc-1 is not a requirement for healthy development in the mouse (52). It is possible that other mucins or mucin-like genes compensate for the lack of Muc-1 in these mice, although no up-regulation of any known mucin gene or mucin-like gene examined in the study was seen. It must also be remembered that these mice were raised in a pathogen-free environment and any deleterious effect of Muc-1 absence on the immune system would not have been evaluated.

1.3.2 MUC1 is involved in tumour progression

The presence of high levels of MUC1 in many carcinomas may suggest that cells expressing it have a selective advantage over those that do not. Direct evidence for the importance of MUC1 in tumour progression has come from Muc-1 null mice crossed with Polyomavirus middle-T transgenic mice which develop spontaneous mammary tumours. The tumours that were induced in Muc-1 deficient mice grew more slowly than tumours growing in wild-type mice, demonstrating that Muc-1 is important in progression of the tumour. The Muc-1 negative mice also had fewer lung metastases, although this was not statistically significant (52).

Until recently there has been no evidence that MUC1 protein sequence differs between tumours and normal cells. However a proportion of tandem repeats obtained from MUC1 expressed in the breast cancer cell line, T47D differed in amino acid sequence from repeats obtaining in a similar fashion from human milk (28). The tandem repeats of other tumours and corresponding normal tissue must be adequately sequenced before it is clear if this is a feature of carcinoma-associated MUC1. Over expression of either MUC1 or MUC1/Y in normal epithelial cells is not transforming, suggesting that its effects are slight, require other events, or that its activity can only be seen in the context of the whole animal (40).

Why MUC1 should give an advantage to tumour cells is unclear; the possibilities are discussed below. If all these occur MUC1 could be a potent agent of tumour progression, metastasis, and immune suppression.

a) As argued in section 1.3.1b, MUC1 may promote a loss of contact between cells and cells and/or cells and the underlying matrix, especially when cells become de-polarised, by blocking the interaction of adhesion molecules such as E-cadherin and matrix integrins (34, 49, 72). Many tumour cells down-regulate E-cadherin to increase invasion ability (reviewed in (90)). Tumours may achieve the same effect by depolarising and up-regulating MUC1, thereby masking E-cadherin. In a similar, steric fashion, MUC1 may also prevent access of immune effector cells to tumours, (see 'c' below).

b) MUC1 acts as an adhesive by presenting carbohydrates on its surface as ligands to selectin-like molecules and in that way may aid metastatic progression of tumour cells. Carbohydrate epitopes, sialyl Lewis^{x} and sialyl Lewis^{a,} expressed on carcinoma-associated MUC1 can act as ligands for the endothelial molecules selectin-P and E respectively (91, 92). Tumours cells may also be aided in metastasis by binding to endothelia via interactions between the MUC1 protein core, and intercellular adhesion molecule 1 (ICAM-1) (93, 94).

c) The tumour associated MUC1 not only presents new antigens to the immune system, in the form of novel carbohydrates and a previously cryptic protein core, but since tumour cells lose their polarity it does so in a location the immune system finds accessible. It is perhaps therefore not surprising to learn that many cancer patients can mount an immune response against MUC1. MHC restricted and unrestricted cytotoxic T cells (CTLs) have been isolated from breast and pancreatic cancer patients which are able to kill MUC1 expressing cells (95, 96) although the results of unrestricted T-cells remain controversial because of the unusual mechanism of cross linking T-cell receptors by MUC1 which is thought to occur (97). Patients can also produce antibodies to MUC1, detected both in serum and by isolated and immortalised B-cells (98, 99). However, it is obvious that the immune system is often ineffective and that the tumours sometimes persist. One explanation is that MUC1 may disable the immune system. MUC1 may contribute to immuno-suppression by binding to ICAM-1 and E-selectin thereby blocking diapedesis of CTLs (93, 100), inhibiting natural killer (NK) cell mediated killing of target cells (100) and causing anergy of human T-cell proliferative response (101).

Despite these drawbacks the clinical trials of MUC1-based treatments proceeds since experiments in animals have shown that immune-tolerance can be overcome (see the following section). Immunotherapy represents the most promising new therapy to have arisen from the study of MUC1. The challenge to immunologists is to develop methods for presenting the MUC1 antigen in a manner that results in effective rejection of the tumour.

1.4 Applications of MUC1.

1.4.1 Immunotherapy and Clinical Trials

As described above, despite an attempt to target MUC1, in many cases the immune system is ineffective and the tumour grows. This may be due to suppressive effects of MUC1 itself or aspects of tumour biology, such as down-regulation of MHC molecules or the microenvironment of the tumour which is immunosuppressive ((102) and references therein).

Several groups are focusing on the use of potent antigen presenting cells, i.e. dendritic cells (DCs), to prime naive T-cells to MUC1 antigen in a MHC-restricted fashion. Human MHC restricted CTLs against MUC1 fragments have been produced which can lyse MUC1 expressing cells (103). Various methods are being used to induce expression of MUC1 antigen in DCs, sometimes in combination with the cytokine IL-2, which can reverse MUC1-induced immune suppression. Many of these studies have already succeeded in reducing tumour load in mice (104, 105).

MUC1 transgenic mice provide a good model to examine tumour rejection and tolerance. Modification of the MUC1 expressing tumour cells by fusion with dendritic cells or by expression of the co-stimulatory molecule B7.1 prior to injection into the transgenic animal resulted in a dramatic decrease in tumourigenicity (106, 107). Importantly these studies also showed that autoimmunity was not induced.

Several clinical trials have been undertaken for MUC1-based treatments, so far all of these have shown that no autoimmunity or toxicity has been associated with immunisation and in several cases a response was also detected in the treated patients. Peptide based vaccines tested under trial have included a mannan-MUC1 fusion which showed a predominantly humoral response in patients, and a 105 amino acid peptide (5 tandem repeats) together with the BCG adjuvant which resulted in T-cell infiltration of the tumour and improved T-cell function. Biomira has completed its Phase I clinical trial with BLP25 vaccine, a 25 amino acid sequence of the MUC-1 tandem repeat core encapsulated in a liposomal delivery system. Preliminary analysis of the data show that BLP25 is both safe and triggers a cytotoxic T lymphocyte immune response against

cancer cells (Chemical Business Newsbase, 9th April 1999). A phase I trial using humanised HMFG-1 in patients is now in progress (19). The phase I trial of MUC1 plus IL-2 expressed from a vaccinia virus vector has shown no toxicity in patients and an immune response was detected in some patients. It is now being tested in a phase II clinical trial (108). In a randomised trial patients receiving the MUC1 associated carbohydrate Tn and cyclophosphamide (given in order to prevent tolerance) had a significantly longer survival mean than those that did not. There was a negative correlation between the growth of the tumours and the antibody titre of Tn. In November 1998, a major Phase III multicentre trial with the Tn immunogen (Theratope) was launched in 900 metastatic breast cancer patients in North America and Europe. If proved successful the vaccine could be licensed in four to five years.

1.4.2 Gene therapy using the MUC1 promoter

The group of Donald Kufe has been most active in exploring the possibility of using the MUC1 promoter as part of a therapy. There have been three studies assessing the MUC1 promoter (-714/+33) and one assessing the enhancer region (-585/-476) to function as a carcinoma-specific switch (109, 17, 16, 110). The promoter has been used to transcribe toxic genes, such as the apoptotic gene BAX and herpes simplex thymidine kinase (HSV-tk) which can activate the non-lethal ganciclovir pro-drug to a lethal nucleoside, at the site of the tumour. HSV-tk has also been used in a retroviral vector using MUC1 promoter sequences together with c-erbB2 promoter sequences to drive expression (15). So far these experiments have been successful, with no apparent toxicity to normal parts of the animal.

Adenoviral vectors were used as a method of gene delivery by Kufe's group. Expression of β-gal using -714/+33 of the MUC1 promoter was restricted to breast cancer cells *in vitro* and *in vivo*. Although no β-gal was detected in liver parenchyma, muscle or lung tissue, the mice used were athymic, and so any effect on T-lymphocytes was not evaluated. Selective sensitivity to ganciclovir could be conferred to cells injected with

adenovirus particles containing the MUC1 promoter to drive HSV-tk in human breast cancer cell line tumours in athymic nude mice (109). In a similar study the reduction of ovarian carcinoma in the peritoneal cavity was achieved in athymic mice using the proapoptotic gene BAX, rather than HSV-tk, in MUC1-adenoviral constructs (17). Similarly sequences from the enhancer portion of MUC1 have been used to direct expression. Building on their earlier work with the MUC1 enhancer -585/-476 upstream of the HSV-thymidine kinase promoter (TK promoter) (see figure 1.8) Kufe's group made a MUC1(-585/-476) -TK-HSV-tk adenovirus construct that successfully conferred ganciclovir sensitivity to MCF-7 cells, but not to MUC1-negative PA1 cells (110). However, no positive control promoter was included in this study and it is possible that the results are due to differential susceptibility of adenovirus transduction between MCF-7 and PA1 cells. No *in vivo* work was carried out.

The MUC1 promoter has also been used in gene therapy systems as part of a hybrid promoter, consisting of the MUC1 promoter and c-erbB2 enhancer driving HSV-tk in a retroviral vector (15). The effect of the c-erbB2 sequence is to boost sensitisation to ganciclovir over that of cells transduced with the MUC1 construct alone. The construct has not yet been tested *in vivo.*

These studies demonstrate the feasibility of using the MUC1 promoter in gene therapy setting. The approach may be limited for the treatment of primary or metastatic lesions until a better delivery system can be found since mice injected intravenously by tail vein showed no therapeutic effect (109) and only the surface of large tumours are transduced if the injection is intra-peritoneal (17). An application that might be of use today is the purging of cancer cells from bone marrow. Presently high dose chemotherapy followed by autologous transplantation of bone marrow or peripheral blood is used as a treatment for breast cancer. Sadly relapses are frequent, and may occur due to contaminating tumour cells. Using the MUC1 promoter driving β-gal in adenovirus one cancer cell in $5x10^5$ bone marrow or peripheral blood cells could be detected and the MUC1-HSV-tk adenovirus could purge bone marrow and peripheral blood of contaminating breast cancer cells. The success of this approach not only relies on the high expression of HSV-tk from the promoter, but also the resistance of $CD43^+$ hematopoietic cells to

adenovirus, due to their lack of αvβ3 and αvβ5 integrins required for internalisation of the virus particle (109).

There is a real possibility that the MUC1 promoter might be part of a therapeutic for cancer treatment. Its success will also depend on efficient means of delivery, and perhaps targeted delivery, to ensure that MUC1 expressing normal cells are not affected. Nevertheless these preliminary observations are promising.

1.5 Control of MUC1 gene expression.

1.5.1 Evidence for transcriptional control

Gene expression can be controlled at all levels of the pathway leading to expression of the gene product; through gene rearrangements, at the levels of transcription, by mRNA processing, through transport to the cytoplasm, by mRNA stability, at the level of translation, and through post-translational modifications. Levels of MUC1 and Muc-1 protein are concordant with the levels of mRNA, thereby ruling out translational or post-translational control as a significant part of the process (54).

There is agreement that MUC1 expression is primarily controlled at the level of mRNA production. The evidence for this comes from two areas; the expression of transfected *MUC1* gene:reporter gene constructs into cells, and the study of endogenous MUC1 transcript in the presence of a transcription inhibitor together with compounds known to increase MUC1 levels. Methylation of CpGs in the tandem repeat has been found in non-MUC1-expressing cells, but the relationship with expression is not clear.

1) The *MUC1* gene 5' to the coding region (-1600/+33) has transcriptional activity. Expression of chloramphenicol acetyltransferase (CAT), β-galactosidase (β-gal), BAX, and herpes simplex virus thymidine kinase (HSV-tk) genes using MUC1 sequences as a promoter has been seen in transfected or transduced epithelial cells but not in non-epithelial cells (111, 112, 109, 16, 17). The ability to express the reporter gene correlates with the presence of endogenous MUC1. Furthermore the amount of

reporter expressed broadly correlates with the amount of MUC1 expressed suggesting that the main point of control is transcriptional (16, 111).

2) Treatment of MCF-7 cells with cycloheximide (a translational inhibitor) results in an increase in the steady-state levels of MUC1 transcript from both alleles. Abe and Kufe (1990) argued that these effects are dependent on transcription (113). Cells treated with actinomycin for 12 hours are transcriptionally inactive and contain approximately half the MUC1 transcript of untreated cells. Transcript levels in these cells could not be increased, or further degradation prevented, by cycloheximide. The effect of cycloheximide treatment on MUC1 transcript is therefore dependent on transcription and not translation or mRNA stability, although it is possible that a more stable protein may be involved in post transcriptional regulation. (That cycloheximide increases MUC1 transcription is intriguing. It suggests a labile protein is involved in transcriptional repression in MCF-7 cells). MUC1 transcript and protein can be increased by treatment with 12-0-tetradecanoylphorbol-13-acetate (TPA) a phorbol ester. This too is also thought to act primarily through transcription since TPA treatment of MCF-7 cells increased transcriptional activity in nuclear run-on assays (113).

3) Methylation status of genes, in the 5' region of CG islands of house keeping genes, in X–chromosome inactivation, and in parental imprinting is associated with inactivity of gene expression, (114, 115). In cells expressing MUC1 the CpG residues of the *MUC1* tandem repeat are hypomethylated, whereas it is completely methylated in cells that do not make MUC1. Regions outside the tandem repeat, including the 5' promoter region, show no differences in methylation between MUC1 expressing and non-expressing cells. This phenomenon appears to be related to whether the gene is on or off and not to quantitative differences since the level of MUC1 expression in epithelial cells has no correlation with the amount of methylation in the tandem repeat (116).

At this point it is worth noting that the mouse gene, which is expressed in an almost identical pattern to human MUC1 does not have a significant number of CpG dinucleotides in the tandem repeat region. It has just 0.75 CpGs per repeat compared to 6 CpG dinucleotides per repeat in human MUC1. However since the methylation status of the mouse gene has not been examined, methylation could be affecting the mouse gene in other regions.

These observations may mean that the MUC1 tandem repeat must be de-methylated before transcription can commence. Alternatively methylation may occur to stabilise inactivity of the gene or may be a consequence of transcriptional inactivity. Further work, perhaps by de-methylation by 5-azacytidine, is required before the significance of this finding can be related to MUC1 expression.

1.5.2 The MUC1 promoter

The compact size of the gene inferred by the observation of nearby flanking genes expressed at different sites and times was confirmed in transgenic mice. These mice carried a 10.6 kb genomic fragment of *MUC1*, containing 1.6 kb of 5' flanking sequence and 1.9 kb of 3' flanking sequence. They expressed human MUC1 in a pattern faithful to that of expression in the human (see table 1). MUC1 was highly expressed at lactation and in tumours induced by the mouse mammary tumour virus (MMTV) (46, 51). These studies also showed that transcription began from the same start site in lactating mammary gland of the human and transgenic mouse, suggesting the same promoter usage (46).

The MUC1 5' region is a TATA-box containing polymerase II promoter with transcriptional activity in reporter gene constructs used in transient transfection assays. Abe and Kufe (1993) (69) and Kovarik *et al.* (1993) (111) found that 5' deletions down to, –608/+33 bp and –585/+33 bp respectively had no deleterious effect on transcriptional ability. The mouse and human promoter is 78% conserved over the -659/+16 bp region. Since the mouse and human genes are expressed in the same cells and tissues, at lactation and in mammary tumours it is likely that the conserved motifs

are important for correct expression of the gene (45). Figure 1.6[2] shows the proximal section of the MUC1 promoter and illustrates the conserved nucleotides while figure 1.7 illustrates the many theoretical transcription factor binding sites proposed by various authors and the transcription factor search engine MatInspector (http://genomatix.gsf.de/cgi-bin/matinspector/matinspector.pl). The potential number of transcription factor binding sites in the promoter region is large and their interaction potentially complex. Analysing such a promoter *in vivo* would prove difficult and time-consuming, therefore specific regions of the promoter have been examined by the use of *in vitro* transfection of cell lines with promoter:reporter vectors. Such studies reveal that the promoter can be thought of as split into two functional regions; a −570 GC box which binds proteins of the Sp1 family and functions to boost transcription, and a more proximal region which has the dual activity of repressing transcription in non MUC1 expressing cells, and allowing it in MUC1 expressing cells.

Two groups have examined the more distal region of the MUC1 promoter and have made similar observations (summarised in figure 1.8). Abe *et al.* showed that when the promoter was deleted from −613 to −544 activity was reduced by 75% in the breast carcinoma cell line, MCF-7 (69). Kovarik *et al.* saw the same reduction in activity in breast and pancreatic carcinoma cells lines when the promoter was deleted from -590 to -403 (111). By cloning a 400 bp fragment of the promoter (-854/-403) into the enhancer site of a heterologous SV40 promoter reporter vector Kovarik *et al.* could enhance activity four fold in MUC1 expressing cells, compared to just two fold in non-expressing cells. Abe and Kufe also observed an enhancement of transcription in MCF-7 cells by cloning a smaller (-593/-481) fragment upstream of the mouse thymidine kinase (TK) promoter, no data were shown for non-MUC1-expressing cells.

---

[2] The sequence in figure 1.7 is taken from GenBank™ accession number X69118. This numbering system is used in all figures and discussion throughout this thesis. Other authors have used different numbering systems, for clarity when discussing their work their sequence has been aligned to this system. There are slight differences in sequence between authors, again for clarity the accession number sequence is used.

The region -593/-481 showed three protected areas in DNAse 1 foot-printing and south-western blots showed that the sequence protected in the most 3' footprint (-504/-484) bound a 45 kDa protein. In later studies the sequence present in both Abe and Kufe's deleted region (-613/-544) and enhancer region (-593/-481), was shown to form complexes with a factor immunologically related to Sp1. Furthermore a two bp mutation made in the GC box which eliminated Sp1 binding reduced activity of the promoter by 75% in MUC1-expressing cells (112). Therefore the majority, if not all, of the loss of activity seen in the initial deletion mutations appear to be due to a loss of a single GC box at -574.

The proximal region appears to be more complex (111, 112). In order to examine the activity of sequences nearer to the TATAA box without the complication of the rest of the promoter, reporter constructs were used that contained a heterologous SV40 enhancer. This allowed detection of activity in MUC1 promoter deletion mutations from -297/+33 to -62/+33. The combination of the SV40 enhancer and just -152/+33 bps of the MUC1 promoter is sufficient to maintain transcriptional activity in the MUC1 expressing breast carcinoma cell ZR75 and inactivity in non MUC1 expressing fibrosarcoma cells HT1080. The same enhancer with -62/+33 of the promoter does not have transcriptional activity in MUC1 expressing cells, suggesting that sequences important in directing tissue specificity lie in the 90 base pairs between -152 and -62 (see figure 1.9).

Two potential transcription factor binding sites were identified in this region, a GC box (@ -97) and a direct repeat sequence (@ -89/-77) named EMUC after a similar sequence, called E-pal, known to direct epithelial specificity of the E-cadherin promoter (117, 118). Band-shift experiments were used to validate mutations that might prevent factor binding. These mutations were then introduced into the full length promoter and the activity of the mutated promoter examined in transiently transfected cells. While individual or combined GC and EMUC mutations reduced activity in ZR75 cells, either mutation had a enhancing effect on transcription in HT1080 and both mutations were more than additive. These observations suggested both an activating and repressive role for the

GC box and EMUC sequences. DNAse I protection of the region showed no obvious differences between expressing and non expressing cells, and indeed showed no protection at all of the EMUC sequence.

A more detailed analysis of the -97 GC box revealed that several factors bind to it, some of which appear related to the Sp1 family, one of which is immunologically related to Sp1. One of the complexes appeared to be unrelated to Sp1, it was named SpA after the nucleotides that when mutated prevented the oligonucleotide from binding to it GGGAGGGGGCGGGGTT (112).

1.6 The aims of this thesis.

The aims of this thesis were:

1. to further define the *cis* acting sequences involved in;
    - the regulation of MUC1 expression in malignancy and lactation *in vivo* (Chapter 3).
    - the epithelial-specific expression of MUC1 using cell lines cultured *in vitro* (Chapter 4).
2. to attempt to identify and define factors interacting with functionally relevant MUC1 promoter sequences (Chapter 5).

1.7 Chapter 1 : Figures

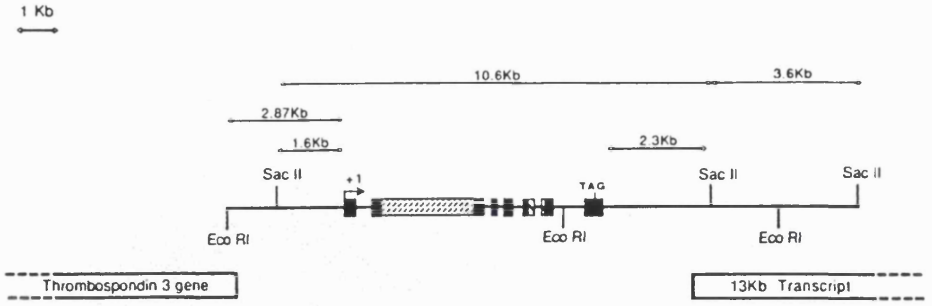
Figure 1.1. The structure of MUC1.

a) Representation of the genomic context of *MUC1* to scale. Exons are shown as closed boxes. The transmebrane domain is indicated as hatched in exons 5 and 6. kb distances are shown (from (111)). The tandem repeat is indicated by the scored-region in exon 2.

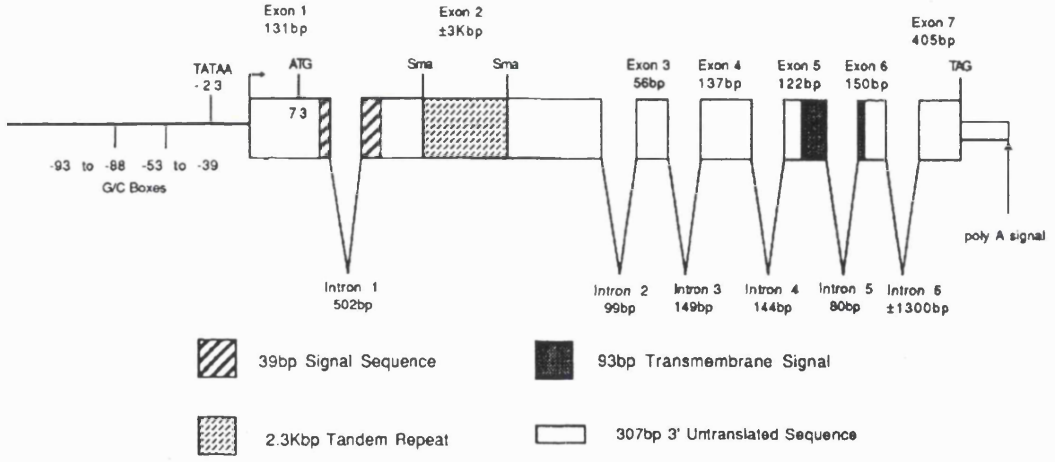
b) Exon and intron organisation of the *MUC1* gene. The tandem repeat sequence varies between 2 to 6 Kb depending on the allele (from (7)).

c) The MUC1 protein, containing 60 tandem repeats drawn to scale. Most of the protein consists of the repeat units. TM = transmembrane region, (from (7)).

a


1 Kb
10.6Kb
3.6Kb
2.87Kb
1.6Kb
2.3Kb
Sac II
+1
TAG
Sac II
Sac II
Eco RI
Eco RI
Eco RI
Thrombospondin 3 gene
13Kb Transcript

b


Exon 1
131bp
Exon 2
±3Kbp
Exon 7
405bp
TATAA
-23
ATG
Sma
Sma
Exon 3
56bp
Exon 4
137bp
Exon 5
122bp
Exon 6
150bp
TAG
73
-93 to -88 -53 to -39
G/C Boxes
poly A signal
Intron 1
502bp
Intron 2
99bp
Intron 3
149bp
Intron 4
144bp
Intron 5
80bp
Intron 6
±1300bp
39bp Signal Sequence
93bp Transmembrane Signal
2.3Kbp Tandem Repeat
307bp 3' Untranslated Sequence

c

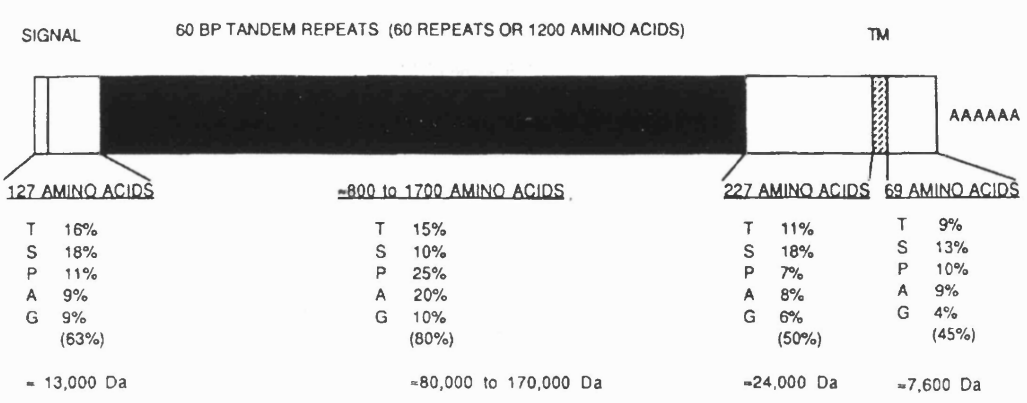

SIGNAL
60 BP TANDEM REPEATS (60 REPEATS OR 1200 AMINO ACIDS)
TM
AAAAAA
127 AMINO ACIDS
≈800 to 1700 AMINO ACIDS
227 AMINO ACIDS
69 AMINO ACIDS
T 16%
S 18%
P 11%
A 9%
G 9%
(63%)
T 15%
S 10%
P 25%
A 20%
G 10%
(80%)
T 11%
S 18%
P 7%
A 8%
G 6%
(50%)
T 9%
S 13%
P 10%
A 9%
G 4%
(45%)
≈ 13,000 Da
≈80,000 to 170,000 Da
≈24,000 Da
≈7,600 Da
CORE PROTEIN ESTIMATED MOL WT 120,000 TO 225,000 Da

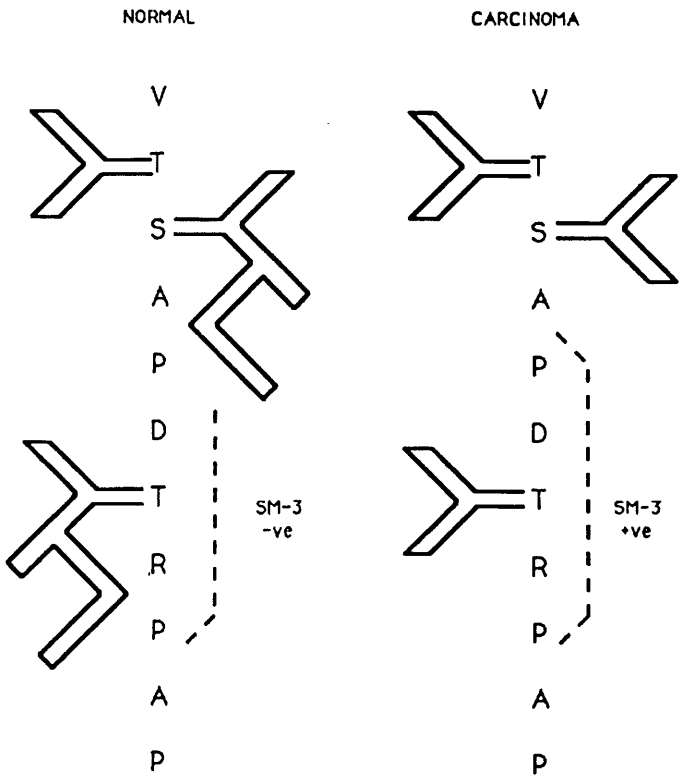
Figure 1.2. MUC1 Glycosylation.

a) Diagram of the MUC1 'PDTRP' epitope in the tandem repeat recognised by the monoclonal antibody SM-3. This antibody binds to MUC1 in breast tumours and cancer cell lines, but not normal breast tissue (32). Changes in glycosylation that shorten carbohydrate side chains may unmask the epitope in carcinomas (modified from (7)).

b) Flow diagram of alternative pathways of O-linked glycosylation of MUC1. Chain branching occurs via core-2 β1,6 GlcNAc-transferase while chain termination by adding sialic acid requires α2,3 sialyltransferase (from (31)). An elevation of α2,3 over core-2 in breast cancer favours chain termination and results in shorter, less complex carbohydrate chains.

Figure 1.2

a


NORMAL
CARCINOMA
V
T
S
A
P
D
T
R
P
A
P
SM-3
-ve
SM-3
+ve

b

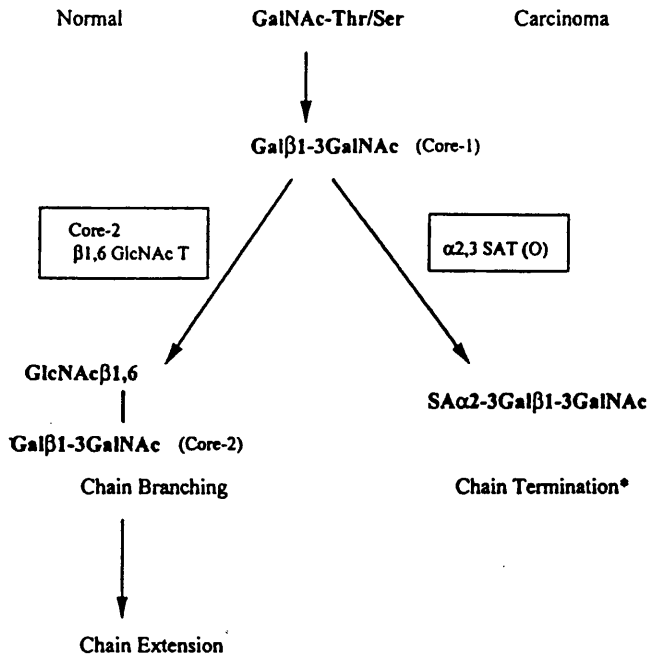

Normal
GalNAc-Thr/Ser
Carcinoma
Galβ1-3GalNAc (Core-1)
Core-2
β1,6 GlcNAc T
α2,3 SAT (O)
GlcNAcβ1,6
Galβ1-3GalNAc (Core-2)
Chain Branching
SAα2-3Galβ1-3GalNAc
Chain Termination*
Chain Extension

Figure 1.3. MUC1 secondary protein structure .

a) A computer model of a MUC1 60 amino acid peptide (3 tandem repeats) in the polyproline β-turn helix showing that the amino acid side chains radiate outward from an extended rod like structure (taken from (33)).

b) Diagrammatic structure of MUC1 in relation to the glycocalyx. The horizontal line indicates the length of a glycosylated repeat according to Jentoft (80). Lollipop sticks indicate N-linked carbohydrates, plain sticks indicate O-linked carbohydrates. The arrow indicates the site of cleavage that occurs in the endoplasmic reticulum. The resulting ectodomain and cytoplasmic cleaved product remain non-covalently associated (from (34)).

c) A comparison of proposed MUC1 isoforms with the full length protein, here called MUC1/REP. In MUC1/Y and MUC1/X (not shown) splice donor and splice acceptor sites (S.D. & S.A. respectively) are thought to result in the deletion of the tandem repeat array and flanking sequences. MUC1/SEC results from a stop codon in intron 5. The abbreviations SP, TM and CYT indicate the signal peptide, transmembrane domain and cytoplasmic domains respectively (from (40)).

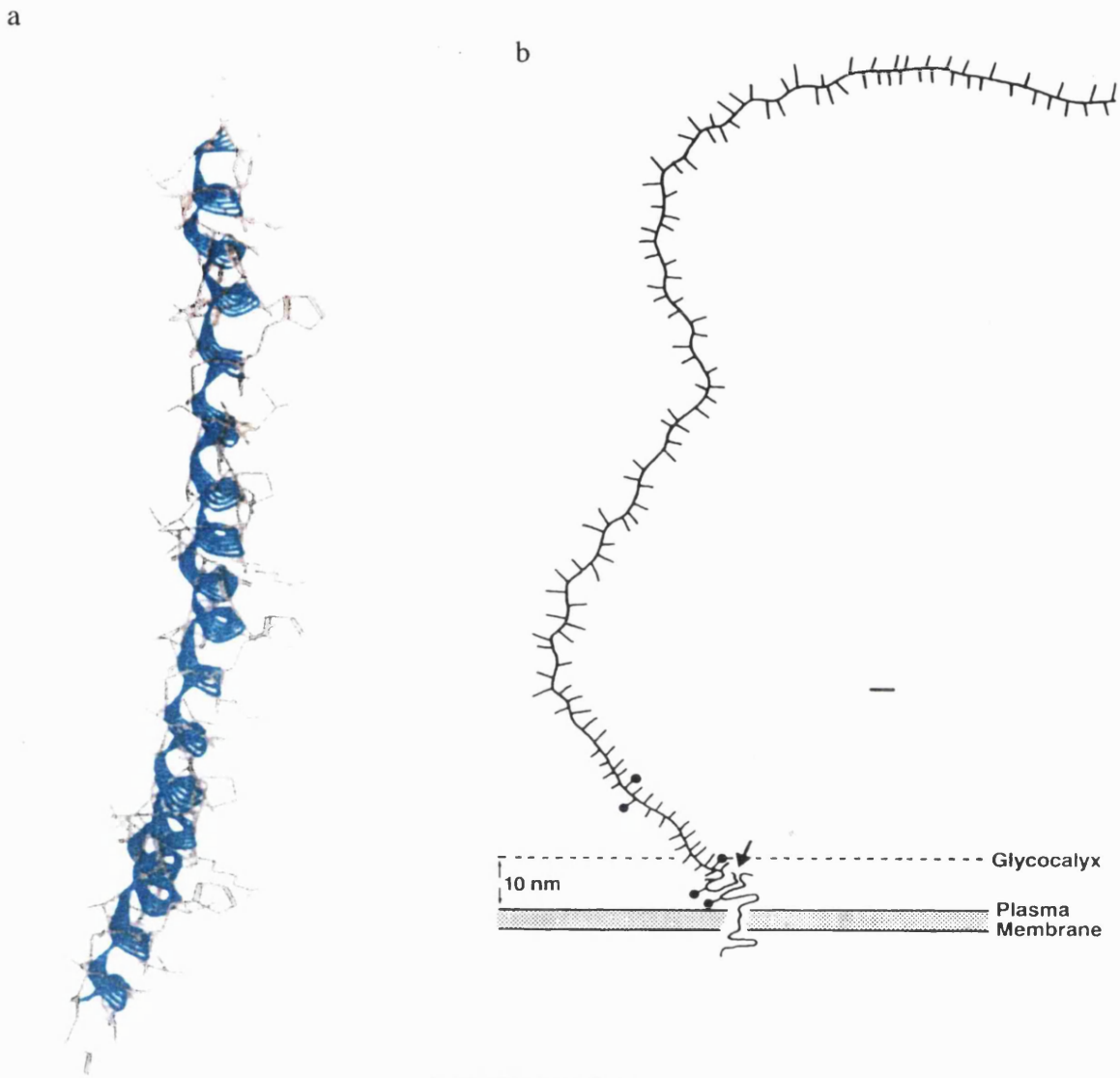
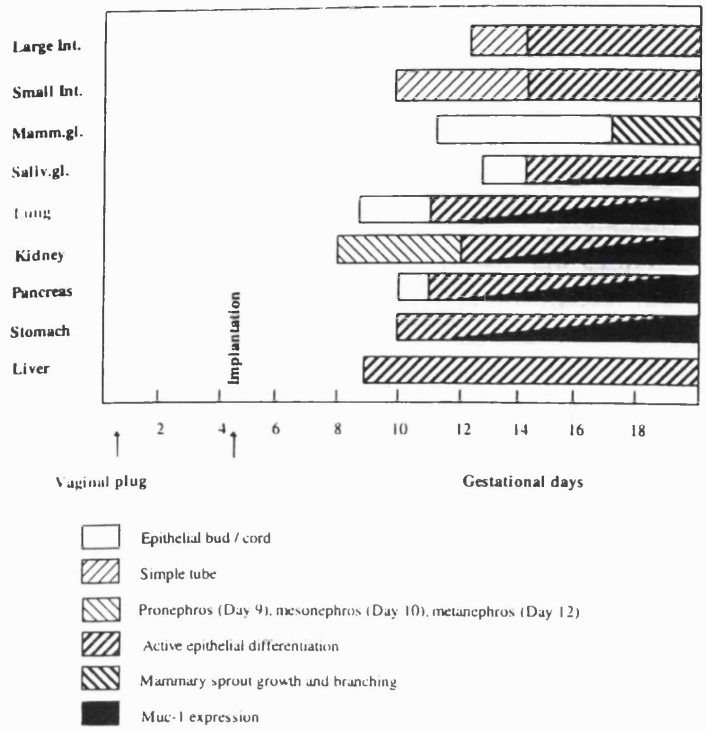

a
b
Glycocalyx
10 nm
Plasma
Membrane
A. MUC1 cDNA forms
c
1. MUC1/REP
2. MUC1/SEC
3. MUC1/Y
init. codon
stop codon
S.D.
S.A.
SP
TM
CYT
B. MUC1 protein isoforms
N
C
MUC1/REPex
cleavage prod.

Figure 1.4 Muc-1 in development, and protective function.

a) Emergence of Muc-1 expression in developing organs of the mouse, based on histochemical staining (from (47)).

b) MUC1 protects the cell from micro-organisms (from (80)).

a


Large Int.
Small Int.
Mamm.gl.
Saliv.gl.
Lung
Kidney
Pancreas
Stomach
Liver
Implantation
2
4
6
8
10
12
14
16
18
Vaginal plug
Gestational days
Epithelial bud / cord
Simple tube
Pronephros (Day 9), mesonephros (Day 10), metanephros (Day 12)
Active epithelial differentiation
Mammary sprout growth and branching
Muc-1 expression

b

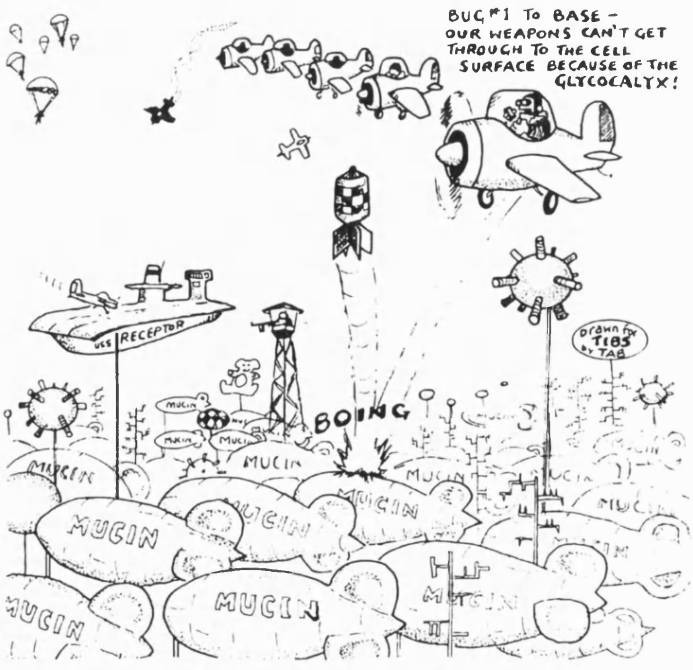
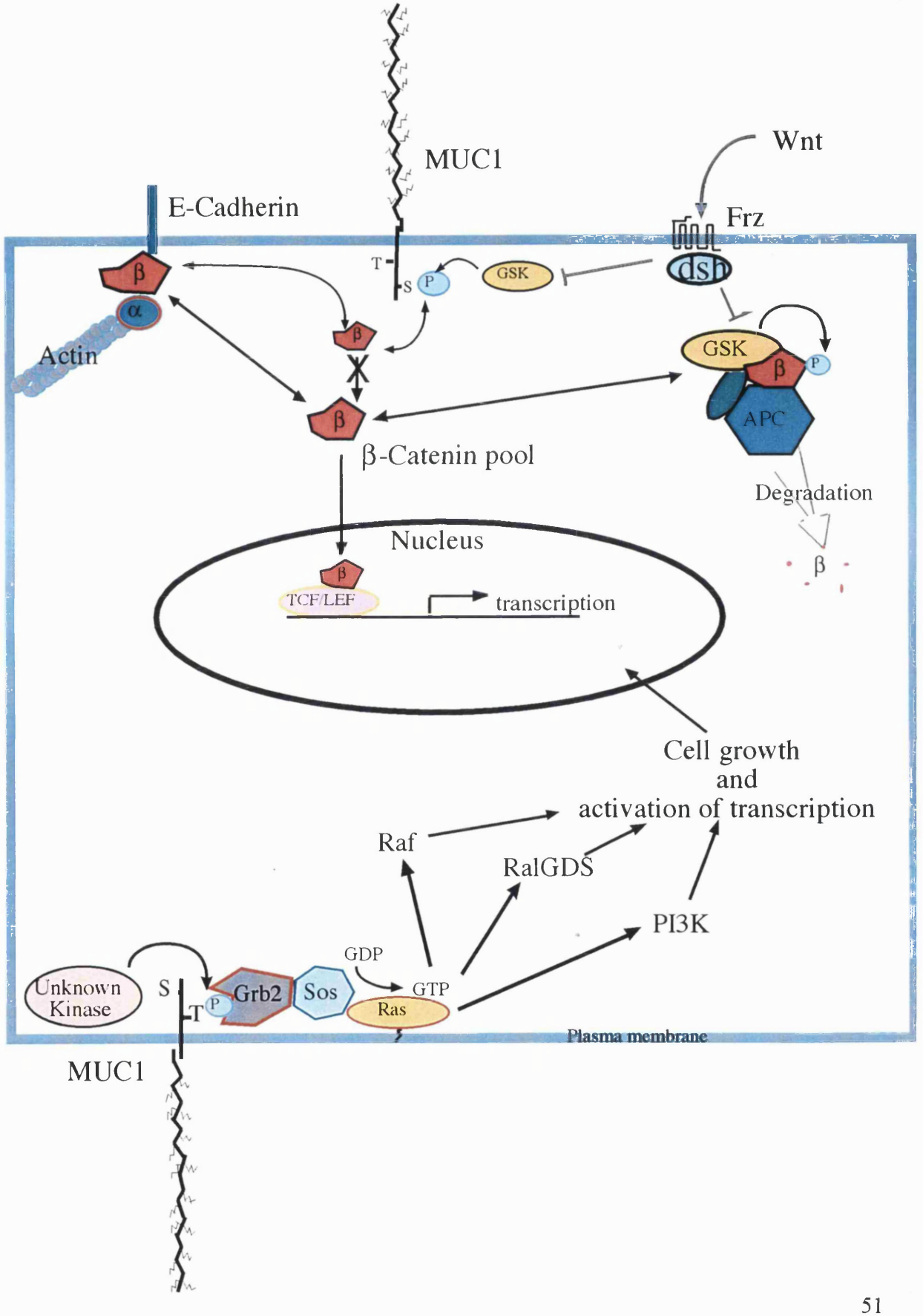

BUG #1 TO BASE - OUR WEAPONS CAN'T GET THROUGH TO THE CELL SURFACE BECAUSE OF THE GLYCOCALYX!
VEGF RECEPTOR
BOING
MUCIN
drawn for TIBS by TAB

Figure 1.5. Potential role of MUC1

a) The proposed interaction of MUC1 cytoplasmic domain in competition for β-catenin with E-cadherin is illustrated in the top half of the square. In classical Wnt signalling; by binding to the Frizzled (Frz) family of receptors, the dishevelled (dsh) protein is hyperphosporylated and recruited to the membrane. Activated dsh inhibits Glycogen Synthase 3β (GSK) action, which normally phosphorylates (P) β-catenin (β) and directs it together with adenomatous polyposis coli (APC) and axin family members to degradation by the ubiquitin-proteosome system. Decreased degradation of β–catenin leads to its accumulation, nuclear translocation and association with the lymphoid enhancer binding factor / T cell-specific factor (TCF/LEF) transcription factors, leading to transcriptional activation. β-catenin is also a component of the E-cadherin - catenin adhesion system that links E-cadherin via β- and α-catenin to the actin cytoskeleton. By altering the amounts of adhesion components the signalling role of β-catenin can be significantly affected (from (119, 120)). MUC1 can interact with β-catenin and GSKβ3. Phosphorylation of a serine in the MUC1 cytoplasmic domain by GSKβ3 is thought to dissociate β-catenin from the domain. β-catenin released in this way does not appear to be added to the general β-catenin pool, but can associate with E-cadherin. Inhibition of GSKβ3 by Wnt signalling may result in β-catenin accumulation on MUC1, as well as in the cellular pool, and a loss of β-catenin from E-cadherin adhesion complex (86, 121).

b) The lower half of the square illustrates the tyrosine phosphorylated MUC1 cytoplasmic tail which has been found bound to Grb2 and Sos and by implication may signal to Ras. Although the kinase responsible, and the ligand is unknown these observations suggest MUC1 may have a role in signalling (84).

a


MUC1
Wnt
E-Cadherin
Frz
dsh
GSK
T
S
P
β
α
Actin
β-Catenin pool
APC
Degradation
Nucleus
TCF/LEF
transcription
Cell growth
and
activation of transcription
Raf
RalGDS
PI3K
GDP
GTP
Ras
Grb2
Sos
Unknown
Kinase
Plasma membrane
MUC1

b

Figure 1.6. Promoter homology.

Homology of the human MUC1 promoter with the mouse sequence (from -789/+16) is indicated by a capital letter (45). The TATA box (boxed), the transcriptional start site (+1), and the first codon (underlined) are indicated.
Bases +17/+67 are not compared to the mouse sequence but are included for completeness.
(This sequence can be found in GenBank, accession number X69118).

The MUC1 promoter (X69118)

```
-789  cGaGCgGCCc ctCagcttCG gcGcCcAGcC cCGCaAgGCT CcCGGTGACC
-739  ACTAGAGGGC gGGaGGAgct cctGGCCaGT GGTGGAGAGT GGcaAGGAAG
-689  GACCcTAGGg TTCAtcGGAG CCCAgGTTTA cTCCCTTAAG TGGAAATTTC
-639  ttCCcCCaCt ccTCcttGGc tttctCCAAg GAGGGaACCC AGGCTgCTGG
-589  AAAGTCcGGC tGGGGCGGGG AcTGTGGgTT caggggagaa cggggtgtgg
-539  aAcgGgACaG GGAgCGGtTA GAAGGGtGGG GCTaTTCCGG GAAGTGGTGg
-489  GGGGAGGGAG cCcAAAACTA GCacCtaGTC CACTcATTAT CCAGCCCtCT
-439  TATttctcGg CCgCTCtgCT TcaGtGGACC CGGgGAGggc gGGgAAgtGG
-389  AgtGGGaGAC CTAGGGgTGG gCTTCCCgac CTTGCTGTaC AGGaCctcga
-339  cctagCTggC TtTGTTcCCc aTCCcCacgT TaGttGTtgC cCTGaGgCTA
-289  AAACtAGAGC CcaGGGGCCC CAaGTTcCag aCTgCcCcTc CcCcCtCccc
-239  cGgAGcCAgG GAGtGGTTGG TgAAAGgGGG AGGCCAGcTG GAGAAcAAAC
-189  gGGTaGTCAg ggGGTTGAGc GATTaGAGCC CttGTACCCT ACcCAGGAAT
-139  GgtTGGGGaG GAGGAGgaaG AGGtAggagg TAGGGGAGGG GGCGGgGTTT
 -89  TGTcACCTGT CACCTgCTGc CTGTGCCTAG GGCGGGcggg cggggagtgG
                                                    +1
 -39  GGGGGACcGG TATAAAGCgG tAGGCgCCtg tGCCCGcTCC ACCTCtCAag
 +13  CcaGCGCCTG CCTGAATCTG TTCTGCCCCC TCCCCACCAT TTCACCACCA
 +63  CCATG
```

Figure 1.7. Theoretical transcription factor binding sites on the MUC1 promoter.

This figure was compiled from several sources; (69, 111, 122) and by using the on-line transcription factor binding site engine 'MatInspector-Professional' at http://genomatrix.gsf.de/cgi-bin/matinspector/matinspector.pl.

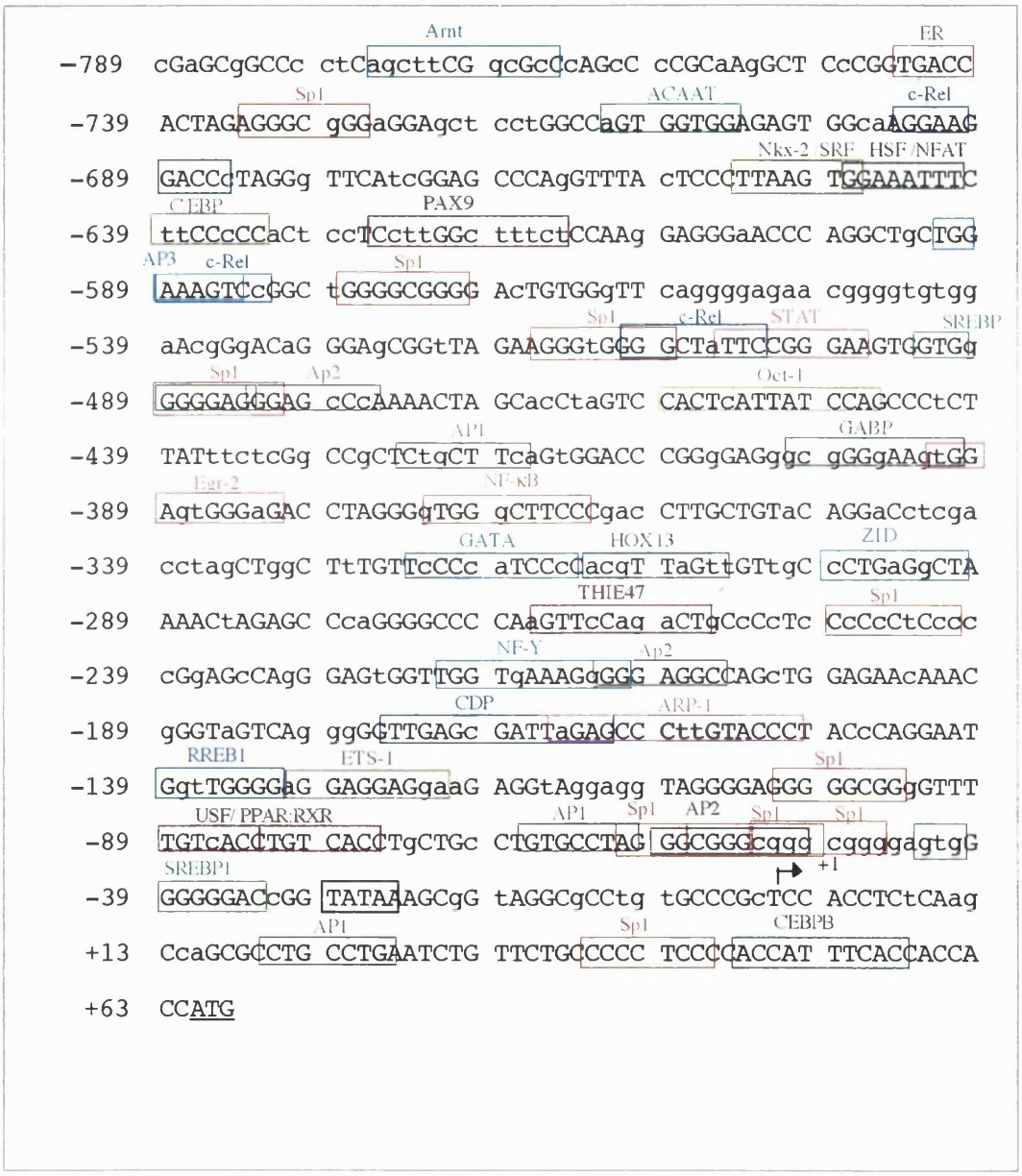
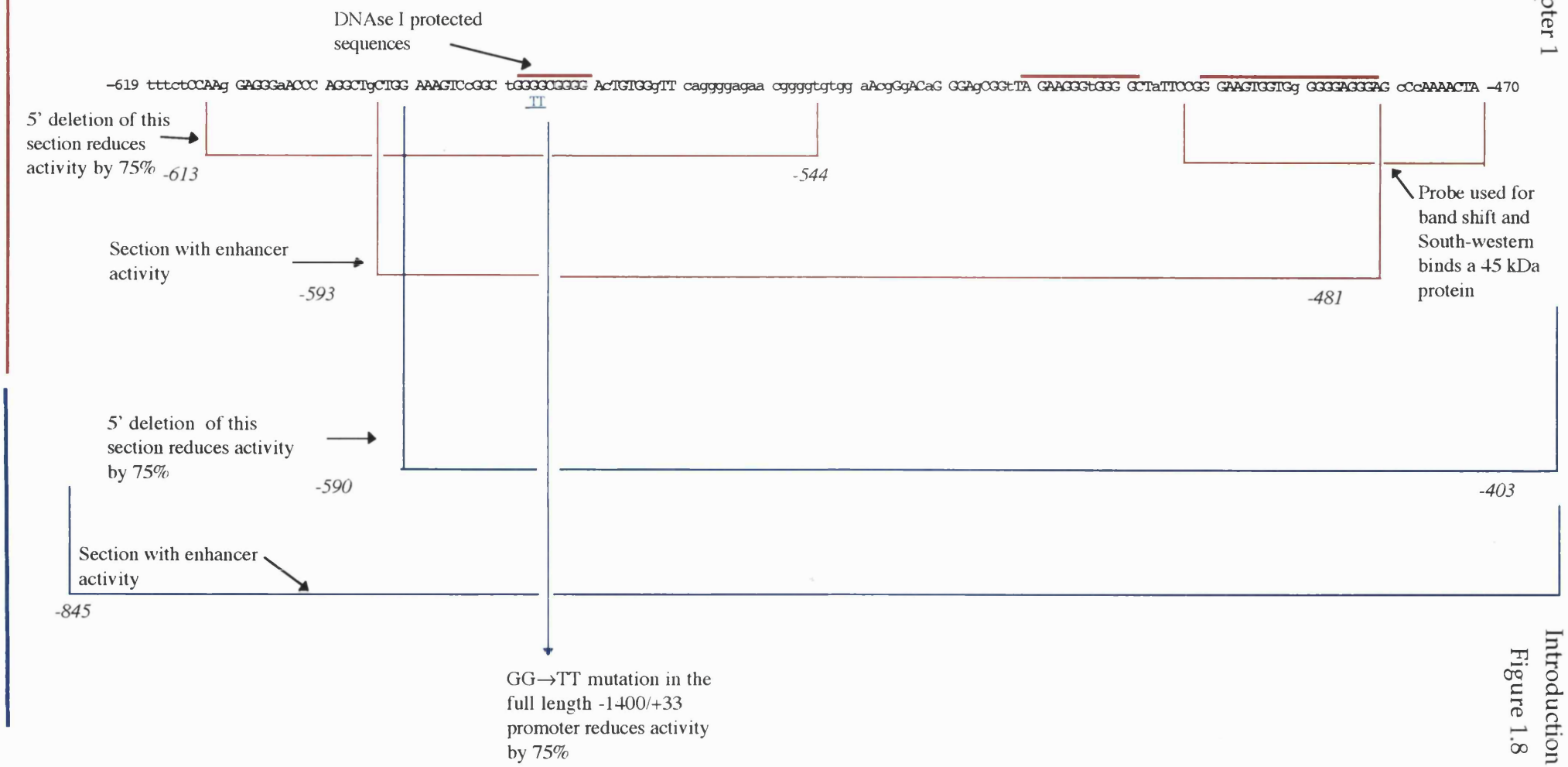

Arnt
ER
-789 cGaGCgGCCc ctCagcttCG qcGcCcAGcC cCGCaAgGCT CcCGGTGACC
Sp1
ACAAT
c-Rel
-739 ACTAGAGGGC gGGaGGAgct cctGGCCaGT GGTGGAGAGT GGcaAGGAAG
Nkx-2
SRF
HSF/NFAT
-689 GACCcTAGGg TTCAtcGGAG CCCAgGTTTA cTCCCTTAAG TGGAAATTTC
CEBP
PAX9
-639 ttCCcCCaCt ccTCcttGGc tttctCCAAg GAGGGaACCC AGGCTgCTGG
AP3
c-Rel
Sp1
-589 AAAGTCcGGC tGGGGCGGGG AcTGTGGgTT caggggagaa cggggtgtgg
Sp1
c-Rel
STAT
SREBP
-539 aAcgGgACaG GGAgCGGtTA GAAGGGtGGG GCTaTTCCGG GAAGTGGTGg
Sp1
Ap2
Oct-1
-489 GGGGAGGGAG cCcAAAACTA GCacCtaGTC CACTcATTAT CCAGCCCtCT
AP1
GABP
-439 TATttctcGg CCgCTCtqCT TcaGtGGACC CGGgGAGggc gGGgAAqtGG
Egr-2
NF-kB
-389 AqtGGGaGAC CTAGGGqTGG qCTTCCCgac CTTGCTGTaC AGGaCctcga
GATA
HOX13
ZID
-339 cctagCTggC TtTGTTcCCc aTCCcCacqT TaGttGTtgC cCTGaGgCTA
THIE47
Sp1
-289 AAACtAGAGC CcaGGGGCCC CAaGTTcCaq aCTqCcCcTc CcCcCtCccc
NF-Y
Ap2
-239 cGgAGcCAgG GAGtGGTTGG TqAAAGqGGG AGGCCAGcTG GAGAAcAAAC
CDP
ARP-1
-189 gGGTaGTCAg ggGGTTGAGc GATTaGAGCC CttGTACCCT ACcCAGGAAT
RREB1
ETS-1
Sp1
-139 GqtTGGGGaG GAGGAGgaaG AGGtAggagg TAGGGGAGGG GGCGGgGTTT
USF/PPAR:RXR
AP1
Sp1
AP2
Sp1
Sp1
-89 TGTcACCTGT CACCTgCTGc CTGTGCCTAG GGCGGGcqqq cqqqgagtgG
SREBP1
+1
-39 GGGGGACcGG TATAAAGCgG tAGGCgCCtg tGCCCGcTCC ACCTCtCAag
AP1
Sp1
CEBPB
+13 CcaGCGCCTG CCTGAATCTG TTCTGCCCCC TCCCCACCAT TTCACCACCA
+63 CCATG

Figure 1.8. The distal region of the MUC1 promoter (-1600 / -400).

The red lines indicate the work of Abe and Kufe (1993) (69). Sequential deletion of the full length promoter from the 5' end revealed that loss of the sequences -613 to -544 reduced activity of the promoter by 75% in MCF-7 breast carcinoma cells. An overlapping sequence of -593/-481 could enhance transcription when cloned upstream of the TK promoter. The region -593/-481 was protected in three positions from DNAse I. The third 3' protected region was examined in band-shift experiments and south-western analysis, which showed that the factor binding the sequence was not related to Sp1, and had a molecular weight of 45 kDa.

The Blue lines indicate the work done by Kovarik *et al.*,(1993 and 1996) (111, 112). A large deletion from -590 to -403 resulted in a similar 75% drop in activity of the MUC1 promoter in two MUC1-expressing cells. A large fragment encompassing this region also had enhancer activity (–845/–403). Factors which bind the –574 GC box in *in vitro* band-shift experiments could be competed by a consensus Sp1 sequence, but not by oligonucleotide probes containing a two base pair mutation. The same mutation in the context of the full length promoter (shown) reduces transcriptional activity by 75%, demonstrating that this site is likely to be responsible for the enhanced activity.



Introduction
Figure 1.8

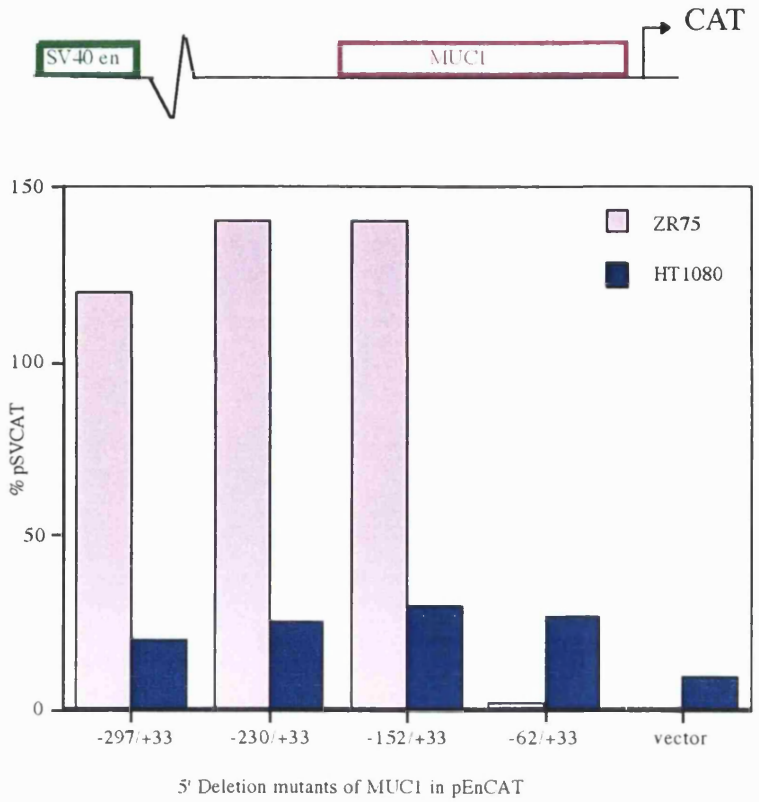
Figure 1.9. The proximal region of the MUC1 promoter.

a) 5' deletions of the region between -297 and -62 in the context of the SV40 enhancer CAT reporter vector showed that promoter sequences -152/+33 could support expression in MUC1-expressing cell line ZR75, but that sequences -62/+33 could not (111).

b) Introduction of mutations in two potential binding sites between -152 and -62 into the full-length promoter (-1400/+33) altered transcriptional ability. Transcription in ZR75 cells was reduced while in the non-MUC1-expressing cell line, HT1080, it was increased several fold, especially in the presence of both mutations, suggesting that these two sites may function to repress transcription in non expressing cells (111).

Figure 1.9

a



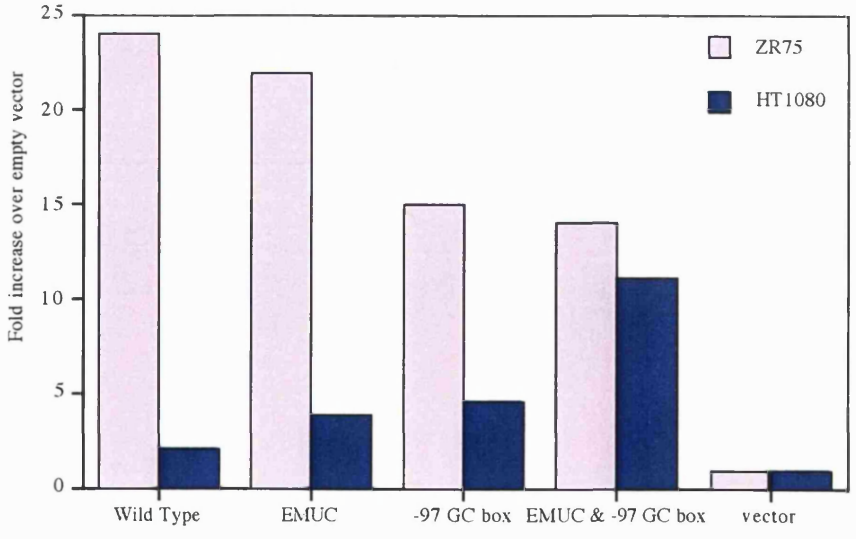
5' Deletion mutants of MUC1 in pEnCAT

b

GC box EMUC

Wild type −107 GGGGAGGGGGCGGGGTTTGTCACCTGTCACC −79

mutations GGGGAGGGTTCGGGGTTTGTCTTTAAACACC



Mutations in the -1400/+33 promoter

Chapter 2 Materials and Methods.

2.1 Materials

2.1.1 Chemicals and solvents

All chemicals were of analytical or tissue culture grade, and were obtained from Sigma Chemicals Ltd, Poole, UK, FSA Laboratory Supplies, Loughborough, UK, or BDH Chemicals Ltd, Poole, UK, except where otherwise stated.

2.1.2 Radiochemicals

All radiochemicals were obtained from Amersham International, Amersham UK.

Compound	**Specific activity**
(γ-^{32}P)ATP	3000 Ci/mmol, 10mCi/ml
(α-^{32}P)dCTP	3000 Ci/mmol, 10mCi/ml
(1-^{14}C)acetyl-coenzyme A	50-60 mCi/mmol
L-(^{35}S) methionine	>1000 Ci/mmol

2.1.3 Enzymes

Restriction endonucleases and their appropriate buffers were obtained from New England Biolabs, USA.

Calf Intestinal Alkaline Phosphatase (CIP)	Boehringer Mannheim
DNAse I (RNase free)	Boehringer Mannheim
Klenow fragment DNA polymerase I	NEB

Reverse Transcriptase (Superscript II™)	Gibco BRL
Sequenase ™ Version 2.0	USB
T4 DNA Ligase and buffer	NEB
T4 Polynucleotide Kinase (T4 PNK)	New England Biolabs.
Taq DNA polymerase	ICRF

2.1.4 Cell culture flasks and dishes

Tissue culture plastic-ware was obtained from Nunc or Falcon.

2.1.5 Cell Culture media

E4 (Dulbecco's Modified Eagle's medium)	ICRF media service.
E4 minus phenol red,	ICRF media service.
FCS (Foetal calf serum)	Gibco Life Sciences.
MEM-Bic (Minimal Essential Media + Bicarb)	ICRF media service.
RPMI	ICRF media service.
Trypsin	ICRF media service.
(0.8 (w/v) NaCl, 0.038% (w/v) KCl, 0.01% (w/v) U/ml penicillin, phenol red (stored at -20 °C)	
Versene (0.02 % (w/v) EDTA in PBS, phenol red)	ICRF media service.

2.1.6 Antibodies

Antibody	**Source and/or reference**
αRabbit-FITC	Secondary anti immunoglobulin antibodies (FITC or rhodamine conjugated) were obtained from DAKO Ltd, High Wycombe, UK.
αMouse-FITC	
αRabbit-peroxidase conjugated	DAKO Ltd, High Wycombe, UK.

αSp1 (PEP2 #sc-59)	From Santa Cruz Biotechnology (via Autogen Bioclear)
αSp2 (K-20 sc-643)	Santa Cruz
αSp3 (D-20 #sc-644)	Santa Cruz
αSp3 (s)	A gift of Guntam Suske (123).
αSp4 (V-20 # sc-6450	Santa Cruz
αUSF (C-20 #sc-862)	Santa Cruz
CT1	Rabbit polyclonal antibody raised against the cytoplasmic tail of MUC1. Developed in the laboratory of Dr. S. Gendler. (43)
HMFG-1,	Mouse monoclonal antibodies raised against MUC1. Developed in the laboratory of Dr. J. Taylor-Papadimitriou (124, 9). (0.7mg/ml)
HRP conjugated αhuman IgG	DAKO Ltd, High Wycombe, UK.
Humanised HMFG-1	Made by the ICRF Monoclonal Laboratory (Clare Hall),Ultrapure 5mg/ml.
LE61	Rabbit polyclonal antibody raised against cytokeratin 18. This antibody was used as a positive control for FACScan internal antigen staining (a gift of I. Lane, at the Royal London Hospital).

2.1.7 Plasmids

Reporter plasmids	**Description**	**Origin and/or reference**
CAT plasmids		
pGCAT-A	MCS vector that allows cloning of a promoter.	(125)
p1.4Kb MUC1-WT	MUC1 sequences -1400/+33 cloned 5' of and adjacent to the reporter CAT in pGCAT-A	(111)
p1.4KbGG(-97) MUC1	As WT, but with a two-base-pair mutation at -99/-98 GG→TT.	(111)
p1.4KbEMUC-MUC1	As WT, but with a six-base-pair mutation, at -95/-90→AAATTT.	(111)
p1.4Kb DM	As WT, but with both mutations above.	(111)
pTk	Based on the pBLCAT5 construct but with the putative AP1 site removed. It contains the minimal 105 bp Thymidine Kinase promoter (TK) adjacent to the CAT reporter gene. MCSs are adjacent (5') to the promoter and in the enhancer site (Shown in appendix A3).	A kind gift of R. Offringa,Lieden, The Netherlands (126, 127)
Luciferase plasmids	Luciferase reporter vectors	From Promega Ltd. Southhampton.
pGL3basic	MCS at the promoter site and enhancer site	"
pGL3enhancer	An SV40 enhancer at the enhancer site,	"

	MCS at the promoter site	
pGL3promoter	An SV40 promoter at the promoter site, a MCS at the enhancer	"
pGL3control	Both SV40 enhancer and promoters in the enhancer and promoter sites.	"
Expression plasmids		
pCMV-Sp1	Sp1 cDNA down stream of the CMV promoter in the pEVR2 vector.	A kind gift of G. Suske, Marburg, Germany. (128)
pCMV-Sp3	Sp3 cDNA cloned down stream of the CMV promoter in pRC.	"
pCMV	the CMV promoter alone in pRC	Generated by cleaving the Sp3 cDNA from the previous vector.
Others		
IMAGE clones	2106-H16 2399-A13	UK HGMP resource centre Cambridge UK (a kind Gift of R O'Shaughnessy)
Polyomavirus middle-T antigen plasmid.	contains the cDNA of the middle-T antigen.	A kind gift of V. Fantl, Viral Carcinogenesis, ICRF.
Polyomavirus transgenic mouse DNA.	Genomic DNA from a transgenic mouse carrying 154-1560 bp of the Polyomavirus middle T gene.	A gift from R. Treisman, Transcription Laboratory, ICRF (129).

2.1.8 Cell Lines

Name	Description	Growth Media	Subculture dilution	Reference or Source and ATCC designation
MCF-7	Cell line derived from human metastatic breast carcinoma cells isolated from a pleural effusion.	Grown in E4, 10% FCS, 10 µg/ml insulin.	1:5	(130). HTB22
T47D	Cell line derived from human metastatic breast carcinoma cells isolated from a pleural effusion.	Grown in E4, 10% FCS.	1:5	(131). HTB133
ZR75	Cell line derived from human metastatic breast carcinoma cells isolated from a pleural effusion.	Grown in E4, 10% FCS	1:10	(132). CRL 1500
HT1080	Cell line derived from human fibrosarcoma	Grown in E4, 10% FCS	1:10	(Cancer 33:1027-1033. 1974). CCL 121
HTB96 Also known as	Cell line derived from human osteosarcoma	Grown in E4, 10% FCS	1:10	(Int. J. Cancer 2: 434-447, 1967). HTB.96

U-2 OS				
COS-1	SV40 transformed African Green Monkey kidney cells derived from CV 1 cells	Grown in E4, 10% FCS.	1:10	(Gluzman, 1981)
DH5α		L-Broth		Derived from HB101
XL-1 (blue)		l-Broth (tetracycline resistance)		Strategene

2.1.9 Miscellaneous

1.5ml and 0.5ml sterile plastic tubes	Scotlab
13 mm screw cap tubes	Corning
Absolute alcohol	Hayman Ltd, Witham, UK
Acrylamide	Anachem
Agarose	Gibco BRL
Ammonium persulphate	Bio-Rad
Biotrap and biotrap membranes	Schleicher&Schuell
Bromophenol Blue	Bio-Rad
Centricon-10	Amicon, Beverley USA
Coomassie Brilliant Blue	Bio-Rad
DEAE Membranes	Schleicher & Schuell, Germany
Dithiothreitol	Bio-Rad

Dried skimmed milk powder (Marvel)	Boots plc, UK
ECL Detection Kit	Amersham
ECL western blotting detection reagents	Amersham
Filtration units (0.2 and 0.45 μm)	Millipore
GalactolightReagents	Trophix, Bedford, UK
GENE Pulser cuvettes	Bio-Rad
Glacto-light β-gal assay kit	Trophix, Bedford, UK
Hybond nylon membranes	Amersham
Hybond-N+ and C	Amersham
Liquid scintillation fluid (Ultima gold)	Amersham
LucLite™ Reagents	Pacard, Bioscience, Groningen, Netherlands
Luminometer cuvettes	Labsystems Group, UK
Microlite™ 96-well plates	Dynex
Mixed bed resin (AG501-X8)	Bio-Rad
Nucleotide triphosphates	PharmaciaBiotech
Oligonucleotides	Synthesised by Ian. Goldsmith's group at Clare Hall, ICRF.
Pre-stained SDS PAGE markers broad range (175-6.5 kDa)	NEB
Protease inhibitors (Cocktail)	BoehringerMannheim
Protein A-Sepharose	Pharmacia
Protein Assay (Bradford)	Bio-Rad
QiagenReagents	Qiagen
SequenasesequencingReagents	U.S. Biochemical Corporation, USA.
Sp1 Peptide (sc-644P)	Santa Cruz Biotechnology
Sp3 peptide (sc-59P)	Santa Cruz Biotechnology
TEMED	Bio-Rad
Trizol	Gibco BRL

Tween-20	Bio-Rad
X-ray film	Fuji or Kodak.

2.2 Buffers and solutions

All solutions were prepared using sterile de-ionised water and stored at room temperature unless otherwise stated. Solutions were sterilised when necessary by autoclaving, or by filtering through a 0.22μm filter unit as appropriate.

Band- shift buffer (4x)	40 mM Tris-HCl pH 7.6, 4 mM $MgCl_2$, 200 mM NaCl, 2 mM EDTA, 2 mM DTT, 16% (v/v) glycerol (stored at 4 °C).
Band-shift Sample buffer(4x)	1X Band-shift buffer, 10% (v/v) glycerol, 0.05% bromophenol blue (stored at 4°C)
$CaCl_2$ (2.5 M)	Prepared from a fresh unopened bottle and filter sterilised, (stored at 4°C)
CAT/Luciferase lysis buffer	0.65% NP-40 10 mM Tris-HCl pH 8.0 150 mM NaCl 1 mM EDTA pH 8.0 (stored at 4°C).
CIP buffer (10x)	0.5 M Tris-HCl pH 8.5, 1 mM EDTA (stored at 4°C)
Coomasie Stain	4.5:4.5:1 methanol : H_2O : Acetic acid,

	and 0.05% Coomasie Brilliant Blue
DeionisedFormamide:	Formamide was stirred with mixed bed resin for 1 hour and then filtered twice through Whatman N1 paper and stored in aliquots at 4°C.
Denhardt's (100x)	2% (w/v) BSA solution 2% (w/v) Ficoll 2% (w/v) polyvinylpyrolidine (PVP) Aliquots were stored at -20°C.
DEPC treated water	1µl DEPC (diethyl pyrocarbonate)/ml of H_2O, autoclaved and cooled.
DNA loading buffer (5x)	0.25% (w/v) bromophenol blue, 5x TBE, 25% (v/v) glycerol
DNA sample buffer (5x)	33% glycerol 3.75 X TBE 125 mM EDTA pH 8 0.275% SDS 0.008% bromophenol blue
HBS (2x)	40 mM HEPES, 275 mM NaCl, pH 7.1 (filter sterilised, stored at -20°C).
Klenow buffer (10x)	0.5 M Tris-HCl pH 7.4, 100 mM $MgCl_2$, 10 mM DTT (stored at -20°C)
L-Agar:	L-Broth with 15g/l Bacto Agar (produced

	by media production unit at the ICRF).
L-broth	1% (w/v) bacto tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 0.1% (w/v) glucose (produced by media production unit at the ICRF).
L-Broth:	10 g/l Bacto Tryptone,5 g/l Bacto yeast extract,10 g/l NaCl, (produced by media production unit at the ICRF)
Luciferase reaction buffer	25 mM glycylglycine pH 7.8 5 mM ATP pH 8.0 15 mM $MgSO_4$ Aliquot and store at -20^{O}C.
Luclite™ Lysis buffer	0.5 M HEPES pH 7.8, 2% Triton N101 1 mM $CaCl_2$, 1 mM $MgCl_2$
Luclite™ substrate	Add 5mls of Luclite lysis buffer to the lyophilized Luclite substrate.
Methylene blue stain	0.03% methylene blue 0.3 M sodium acetate pH
Mouse-Tail digestion buffer	50 mM Tris-HCl pH 7.5 5 mM $CaCl_2$ 1% SDS 35ng / ml Proteinase K (see proteinase K solution)

Mouse-Tail Injection buffer	10 mM Tris pH 1.4 0.1 mM EDTA (made with ultra-pure water, stored at -20 °C).
Nuclear extraction buffer (high salt)	10 mM HEPES pH 7.9, 0.1 mM EDTA, 0.1 mM EGTA, 1.5 mM $MgCl_2$, 420 mM NaCl, 0.5 mM DTT, 25% (v/v) glycerol.
Nuclear lysis NP40 buffer	0.6% NP40, 0.15 M NaCl, 10 mM Tris-HCl pH 7.9, 1 mM EDTA pH 8
Nuclear wash buffer	As Nuclear lysis NP40 buffer, but without NP40
NZY agar	NZY broth plus 1.5% bacto agar
NZY broth	0.5% yeast extract, 1% bacto tryptone, 0.5% NaCl, 0.2% $MgSO_4$.7H2). Adjusted to pH 7.5 with NaOH.
OLB is made by mixing solutions A:B:C in a ratio of 100:250:150 and stored in aliquots at -20°C	
OLB Solution A:	1 ml of solution O, 18 µl β-mercaptoethanol, 5 µl each of dATP, dTTP, dGTP, (each triphosphate dissolved in TE pH 7.0 at a concentration of 0.1 M), Aliquot and store at -20°C
OLB Solution B:	2 M Hepes pH 6.6,(pH with 4 M NaOH)

OLB Solution C:	Hexadeoxyribonucleotidesevenly suspended in TE at 90 OD units/ml (Pharmacia)
OLB Solution O:	1.2 M Tris-HCl pH 8,0.125 M $MgCl_2$, (store at 4°C)
PBS	All forms of PBS were produced by media production unit at the ICRF. Complete PBS was obtained by mixing 400 ml of PBSA with 10 ml each of PBSB and PBSC.
PBSA:	137 mM NaCl, 3.4 mM KCl, 20 mM Na_2HPO_4,1.8 mM KH_2PO4 pH 7.2
PBSB:	36 mM $CaCl_2$
PBSC:	20 mM $MgCl_2$
PCR buffer (10x)	500 mM KCl, 100 mM Tris-HCl pH 9, 1% Triton-X 100.
Phenol / chloroform	phenol in TE pH 8 in a 50:50 mix (v/v) with chloroform.
phosphates (70 mM)	separately: NaH_2PO_4 and Na_2HPO_4 filter sterilised and kept at -20°C.

Ponceau S solution	0.1% (w/v) Ponceau S, 5% (v/v) acetic acid
Proteinase-K Solution (100 µg/ml)	200 mg in 20 ml distilled water frozen in aliquots at -20 °C after snap-freezing on dry ice.
Qiagen P1	50 mM Tris-Cl, ph 8, 10 mM EDTA, 100 µg/ ml RNAse A. Store at 4 °C.
Qiagen P2	200 mM NaOH, 1% SDS.
Qiagen P3	3 M potassium acetate pH 5.5
Qiagen QBT	750 mM NaCl, 50 mM MOPS pH 7 15% isopropanol, 0.15% Triton X-100.
Qiagen QC	1 M NaCl 50 mM MOPS pH 7, 15% isopropanol
Qiagen QF	1.25 M NaCl, 50 mM Tris, Tris-Cl pH 8.5, 15% isopropanol.

Restriction enzyme buffer 1 (yellow)	10 mM Bis Tris Propane-HCl, 10 mM $MgCl_2$, 1 mM DTT pH 7
Restriction enzyme buffer 2 (blue)	10 mM Tris-HCl, 10 mM $MgCl_2$ 50 mM NaCl, 1 mM DTT, pH 7.9
Restriction enzyme buffer 3 (red)	50 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, pH 7.9
Restriction enzyme buffer 4 (green)	20 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM DTT pH 7.9
RIPA Buffer (whole cell lysis buffer)	20 mM sodium phosphate pH 7.2, 50 mM sodium fluoride, 5 mM EDTA, 1% Triton-X100, 1% deoxycholate, 25 µg/ml leupeptin, 10 µg/ml PMS, 10 µg/ml aprotinin. Protease inhibitors were added immediately prior to use.
RNA Dye Solution	7.5% Ficoll 400, Bromophenol Blue
RNA hybribisation buffer	0.2 M sodium phosphate buffer pH 7.2 7% SDS 45% formamide aliquots stored at -20°C.
RNA running buffer (10x)	0.2 M MOPS 50 mM sodium acetate pH to 7 with NaOH

	10 mM EDTA.
RNA Sample Buffer (10x)	12% (V/V) 10x RNA running buffer, 64% (v/v) deionised formamide, 23%(v/v) formaldehyde, 7.5% (w/v) Ficoll 400 plus bromophenol blue.
RNA wash buffer	0.1% SDS 2, 1 or 0.5 x SSC
RT-PCR first strand buffer	20 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$
SDS PAGE running buffer (10x)	0.25 M Trizma base 1.92 M glycine 1% SDS
SDS PAGE sample buffer (2x)	4% SDS 20% glycerol 0.16 M Tris-HCl pH 6.8 0.05% Bromophenol blue 10% β-mercaptoethanol
SM Buffer	0.1 M NaCl, 8 mM $MgSO_4.7H_2O$, 50 mM Tris-HCl pH 7.5%(w/v) gelatin.
Sodium Phosphate buffer (pH 7.2)	1M Na_2HPO_4: 1M $NaH_2PO_4.H_2O$
South-western binding buffer	25 mM NaCl 10 mM Tris pH 7.5

	1 mM $MgCl_2$*
	1 mM zinc acetate*
	5 mM EDTA
	1 mM DTT (fresh)*
	* see text for details, these may be omitted.
South-western blocking buffer (BLOTTO)	50 mM Tris pH 7.5
	50 mM NaCl
	1 mM EDTA
	1 mM DTT (fresh)
	5% Milk powder (Marvel)
South-western HBB buffer(10x)	250 mM HEPES-KOH pH 7.7
	250 mM NaCl
	50 mM $MgCl_2$*
	50 mM zinc acetate*
	1 mM DTT (fresh)*
	* see text for details, these maybe omitted
Southern blot Denaturing solution	1.5 M NaCl, 0.5 M NaOH
Southern blot Neutralising solution	1.5 M NaCl, 0.5 M Tris-HCl pH 7.2, 1 mM EDTA
Southern Prehybridisation buffer (1x)	5X SSPE
	5X Denhardt's solution
	0.5% SDS (w/v)
SSC:(20x)	3 M NaCl

	0.3 M Tri-Sodium Citrate
SSPE: (20X)	3.6 M NaCl 0.2 M Sodium phosphate buffer pH 7.2 0.02 M EDTA pH 8.0
T4 DNA ligase buffer	(NEB) 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 µg/ml BSA.
T4 PNK buffer: (10x)	0.5 M Tris-HCl pH 7.6 0.1 M $MgCl_2$ 50 mM DTT 1 mM spermidine 1 mM EDTA
TBE (10x)	0.89 M Trizma base, 0.89 M Boric acid,10 mM EDTA pH 8
TE buffer (1x)	10 mM Tris-HCl pH 8, 1 mM EDTA pH 8
TENS lysis buffer:	10 mM Trizma base, 1 mM EDTA, 0.1 M NaOH, 0.5% SDS
Tfb1	30 mM KOAc, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2.4H_20$, 15% (v/v) glycerol. Adjusted pH to 5.8 with acetic acid and filter sterilised (stored at 4°C)

Tfb2	10 mM MOPS, 75 mM $MgCl_2$, 10 mM RbCl, 15% (v/v) glycerol. Adjusted pH to 6.5 with KOH and filter sterilised (stored at 4°C)
TG buffer (5x)	20 mM Tris-HCl, 200 mM Glycine
Top agar	NZY broth plus 0.7% agarose
Western blotting transfer buffer	25 mM Trizma base 190 mM glycine 20% methanol

2.3 Methods

2.3.1 Cell Culture

Cells were grown as monolayer cultures in tissue culture flasks (75 cm^2, and 175 cm^2) at 37 °C in a humidified atmosphere kept at 10% (v/v) CO_2. Cell cultures were subcultured once or twice a week depending on the growth rate of the cells. To split cells, the media was removed and the cell monolayer washed with versene and incubated at room temperature for 3 - 5 minutes. The versene wash was removed and a covering amount of trypsin/versene (1:5, 0.05% trypsin) was incubated on the cells at 37 °C for a further 5 mins or until the cells could be detached by agitation. Two volumes of growth media were added to inhibit the trypsin. The media was pipetted up and down to ensure good separation of the cells and an aliquot transferred to a fresh culture flask where fresh growth media was added. Typical sub-culture dilutions are shown in the table above.

2.3.2 Cell Storage

Subconfluent cultures from a 175 cm^2 flask were trypsinised as described above and pelleted in a universal tube by spinning in a bench top centrifuge at 1200 rpm for 5 mins. The pellet was resuspended in 4.5 ml of E4 10% FCS plus 0.5 ml DMSO. 1 ml aliquots were placed in cryovials (Nunc) and the tubes wrapped in thick tissue then frozen at -20 °C for 2 hours. Thereafter they were kept at -70 °C overnight and then placed into liquid nitrogen for long-term storage.

2.3.3 Cell recovery

Cells were taken from liquid nitrogen and transferred to a 37 °C water bath via dry ice. Once there the cells were thawed rapidly and diluted into 10 mls of cold RPMI in a universal tube. They were spun for 5 minutes at 1200 rpm in a bench top centrifuge then resuspended in the appropriate growth medium and placed into a 75 cm^2 cell culture flask. Fresh growth media was added the next day to remove any residual DMSO.

2.3.4 Storage of bacteria

The *Escherichia coli* strains XL-1 Blue or DH5α were used to propagate and amplify plasmid DNA. Bacteria were stored in LB containing 50% (v/v) glycerol at -70 oC. All plasmids described carry the *β-lactamase* gene (Amp^R) and transformed colonies were selected on 100 µg/ml ampicillin.

2.3.5 Preparation of competent bacteria

Bacteria were streaked out from a frozen stock onto L-Agar plates and incubated overnight at 37°C. A single colony was picked, inoculated into 5 ml of L-Broth and grown with shaking at 37°C overnight. This culture was then subcultured 1:20 into 100 ml of pre-warmed L-Broth and grown at 37°C until the OD_{550} reached 0.48.
Cells were chilled on ice for 5 min before recovery by centrifugation in chilled Corex tubes at 6000 rpm (Sorval) at 4°C for 5 min. The pellet was resuspended 0.4 volumes of Tfb1 buffer and left on ice for 5 min, centrifuged as described before, and the pellet resuspended in 0.04 volumes of Tfb2 buffer. These were left on ice for 15 min before

aliquoting into freezing vials (200 µl of cells per vial), and snap frozen on dry ice. For long term storage, cells were stored under liquid nitrogen.

2.3.6 Transformation of competent bacteria

Competent XL-1 or DH5α cells were thawed on ice for 15 min. For transformation of an intact plasmid, 40 µl of cell suspension was added to 5-20 ng of DNA and the tubes placed on ice for ~30 min, mixing gently every 10 min. When transformation of a ligation reaction was performed 3 µl of ligation reaction was mixed well with 40 µl of bacteria and incubated on ice as above. The cells were then heat shocked at 42°C for 90 seconds and immediately returned to ice. 160 µl of L-Broth was added to the cell-DNA mixture and the tubes incubated at 37°C for 45 min. Transformed cells were then spread onto pre-dried Amp L-Agar plates and incubated overnight at 37°C. Competent cells typically gave 10^6-10^8 colonies per µg of supercoiled DNA.

(For blue/white selection of colonies transformed with recombinant pBluescript plasmid (Stratagene), 30 µl of 0.1M IPTG and 30 µl of 25 mg/ml X-gal was spread onto the plates before the addition of the cells).

2.3.7 Small scale preparation of plasmid DNA (miniprep)

Transformed colonies were analysed to ensure that the correct plasmid or ligation product was present. Single colonies were picked from a fresh plate, inoculated into 5 ml of L-Broth containing 100 µg/ml ampicillin and grown overnight at 37°C with shaking. 1.5 ml of culture was transferred to an eppendorf and spun in a microcentrifuge at 15,000 rpm for 20 sec to collect the cell pellet. The L-Broth was poured off, leaving 50-100 µl in the tube and then vortexed briefly to resuspend the cell pellet. 300 µl of TENS buffer was added and vortexed briefly, followed by the addition of 150 µl 3M sodium acetate pH 5.2 and vortexed briefly again. After centrifugation at top speed in a microcentrifuge for 2 min to remove the cell debris, the supernatant was removed to a fresh tube and nucleic acid precipitated by the addition of 0.9 ml ice cold absolute ethanol. This was then centrifuged as before, the supernatant removed, and the nucleic acid pellet washed with 70% ethanol. After a further centrifugation step the ethanol was

poured off and the nucleic acid pellet air-dried before dissolving in 30 µl water. This 'miniprep' DNA was then analysed by restriction digest.

2.3.8 Large scale preparation of plasmid DNA (Maxi-prep)

Large scale DNA preparations were made using Qiagen maxiprep kit. The protocol was as follows; Bacteria were pelleted from 400 mls O/N culture. The pellet was resuspended in 10 ml buffer P1. Then 10 ml of P2 (lysis buffer) was added, mixed and incubated for 5 mins at RT, after which 10 ml of P3 (neutralisation buffer) was added, mixed, and incubated on ice for a further 15 mins. Cell debris was pelleted by centrifugation at 5,000 rpm in a J6 centrifuge. The supernatant was filtered through 4 layers of gauze to remove particles. Qiagen columns were equilibrated by allowing 10 ml of buffer QBT to flow through prior to the addition of the bacterial supernatant. Subsequently two 10 ml additions of wash buffer QC followed to remove impurities and the DNA released by the addition to the column of 15 ml of QF. The elutant was collected in a 30 ml Corex tube. Plasmid DNA was precipitated by the addition of 0.6 % (v/v) isopropanol, vortexed and centrifuged at 10,000 rpm in a Sorval centrifuge at 4^0C for 30 mins. The pellet was washed with 20 ml of 75% ethanol, air dried and resuspended in 0.5 ml TE and stored at -20 °C.

2.3.9 Preparation of Mouse genomic DNA

Mouse tails snipped 0.5 cm from the tip were placed in Mouse tail digestion buffer with 35 ng / ml proteinase K, and incubated for 2 hours at 50 °C. The samples were centrifuged at top speed in a desk top centrifuge for 15 mins and the supernatant transferred to a fresh tube. One volume of RT isopropanol was added and the tube closed and shook by hand. The resulting fibrous DNA precipitation was looped out with a sterile plastic loop into a fresh tube and allowed to air dry for 5 mins. 30 µl of TE was added and the tube vortexed.

2.3.10 Restriction digest of plasmid DNA

Restriction endonuclease digestion of plasmid or genomic DNA was typically carried out in a volume of 20 µl. 10X buffers were supplied with the individual restriction

enzymes and used at a final 1X concentration. The optimal buffer for digests involving two restriction endonucleases was determined according to the supplier's instructions (NEB). If no optimal buffer could be used in which digest by both enzymes was possible simultaneously, the respective digests were carried out sequentially. Plasmid DNA (typically 1-2 μg) was digested for a ~3 hr at the optimal temperature with no more than 10% (v/v) enzyme in glycerol. Where star activity was possible, the period of incubation did not exceed 2 hr.

2.3.11 Restriction digest of oligonucleotide DNA

Oligonucleotides were synthesised with hydroxyl groups at both 5' and 3' ends, therefore ligation into plasmid DNA could be achieved either with addition of 5' phosphates or by restriction digestion to create cohesive ends. Typically the latter approach was taken and oligonucleotides were synthesised to include restriction enzyme sites at the ends and were digested as described above but for 4 hours rather than 3.

2.3.12 Agarose gel electrophoresis of DNA

Agarose gels were prepared by dissolving agarose at 0.8 -3.0% (w/v) in 1x TBE buffer in a microwave oven. The solution was allowed to cool before ethidium bromide was added to a final concentration of 0.2 μg/ml and the gels allowed to set at room temperature in a gel former with a comb. DNA samples were prepared by the addition of DNA sample buffer to 1/5th final volume, and loaded into the wells of the gel submerged in 1x TBE buffer. Electrophoresis was performed in1x TBE using a horizontal gel electrophoresis apparatus at 5 -7 V/cm at room temperature until the desired range of separation of the DNA fragments was achieved. DNA was visualised using a long wave (UVP) light box, and photographed with a video camera and printed on a Sony Graphic printer. The sizes of fragments were estimated by comparison of their mobility relative to molecular weight markers of known size (NEB). Molecular weight markers used were as follows:

1) HindIII restriction endonuclease digestion of bacteriophage λ DNA.

2) 1 kb DNA ladder containing bands from 1-12 repeats of a 1018 bp DNA fragment.

2.3.13 Isolation of DNA fragments from TBE-agarose gels and 8% native acrylamide gels

Unless otherwise stated, the purification of DNA fragments from TBE-agarose gels and acrylamide gels for use as probes or for ligation reactions was performed using the Bio-Trap apparatus according to the manufacturers instructions (Schleicher & Schuell). The band of interest was excised from the gel under long wave UV illumination as a thin slice, or in the case of radiolabelled oligonucleotides excised from 8% acrylamide, following auto-radiography. The slice was transferred to a well between two BT-2 membranes and electrophoresed at 5 -7 V/cm so that the direction of DNA movement was through the non-selective BT-2 membrane and on to the semi-permeable BT-1 membrane for 45 mins. The polarity was reversed briefly <60 seconds, and the free DNA (~200 µl) pipetted from the well created between the BT-1 and BT-2 membranes and placed into the top section of a Centricon-10 column. The volume in the top of the column was made up to 2 ml with H_2O and centrifuged at 8,000 rpm in a Sorval for 30 minutes. The flow through was discarded and water volume again made up to 2 ml and the centrifugation repeated. This procedure allows the removal of salt from the DNA and the concentration of the fragment into 40 µl at the top of the column.

2.3.14 Spectrophometric determination of nucleic acid concentration

For quantitation of RNA, DNA and oligonucleotides, OD readings were taken at 260 nm and 280 nm of 1:500-1:1000 dilution of nucleic acid. An OD of 1.0 at 260 nm corresponds to 50 µg/ml DNA, 40 µg/ml RNA and 20 µg/ml single stranded oligonucleotides. Estimates of purity were obtained by the ratio of OD_{260}/OD_{280}, where for pure preparations, the ratios were 1.8 (DNA) and 2.0 (RNA) respectively.

2.3.15 Polymerase chain reaction (PCR)

PCR reactions were used to amplify sequences from genomic DNA, plasmid DNA and RNA by RT-PCR. The protocol for genomic DNA and plasmid DNA was the same and is described below:

50 ng plasmid DNA and 100-500 ng genomic DNA were amplified in a 50 µl reaction containing 2% deionised formamide, 0.2 mM dNTP mix of dATP, dCTP, dGTP and

dTTP, 1X PCR buffer (Promega) 1 µM of forward and reverse oligonucleotide primers and 0.5 -2.5 mM $MgCl_2$, (typically 1.75M). PCR reactions were overlaid with 1-2 drops of mineral oil to prevent evaporation. Typical cycling conditions included a 'hot start' of 10 min at 94 °C before 1 µl (5 units/µl) Taq DNA polymerase was added and cycled according to the following schedule:

Denaturation	1 min at 94 °C
Anneal	1 min at oligonucleotide melting temperature - 5 °C
Extension	1 min at 72 °C.

repeat for 30 cycles, followed by a further incubation at 72 °C for 10 min. An aliquot, typically of 10 µl was examined by gel electrophoresis.

Primers used were:

For screening mouse tail snip DNA for the presence of MMTV Polyomavirus middle-T gene.

Forward(5'-3'): CCAGAACTCCTGTATCCAGAAGCG.

Reverse(5'-3'): GGATGAGCTGGGGTACTTGTTCCCC.

Annealing temperature used was 55 °C. The 130 bp product was amplified from the Polyomavirus middle T coding region, nucleotides 1110-1240.

For generating a DNA probe for detecting mouse cytokeratin 19 RNA.

Forward(5'-3') GGTCCCACTAAAACTTCCACCGCGG

Reverse(5'-3') CAGCAGTTCTCAGACCTGCGTCCC

Annealing temperature used was 60 °C. A 281 bp product of the 5' untranslated region of mouse Cytokeratin 19 (Accession No M28698) was amplified from IMAGE clones 2106-H16 and 2399-A13.

2.3.16 DNA sequencing

DNA sequencing was performed either using the Sequenase™ Version 2.0 DNA sequencing kit (United States Biochemical) or PRISM™ ready reaction dyedeoxy™ terminator cycle sequencing kit, both as directed by the manufacturers.

2.3.17 DNA ligation reaction

Vector and insert plasmid DNA was digested with the appropriate restriction endonucleases and gel-purified using the Bio-Trap as described previously. Approximately 100 ng vector DNA was incubated with insert DNA in molar ratios of 1:1, 1:3 and 1:5 in the presence of 1X T4 DNA ligase buffer (NEB) and 1 µl (1 U/µl) T4 DNA Ligase in a 10 µl reaction volume at 16 °C O/N. Ligation reactions were made up to a 40 µl volume with water before transformation into competent bacteria as described above.

2.3.18 Annealing oligonucleotides

To anneal complementary oligonucleotides 20 µM of each was added to 1 mM $MgCl_2$, 1 mM Tris-HCl pH 8 in a total volume of 100 µl and heated at 80 °C for 3 mins, then allowed to cool to RT on the bench top (for long term storage then frozen at -20 °C).

2.3.19 End Labelling of Oligonucleotide Probes

Oligonucleotide probes without 5' overhangs or unpaired guanines were 5' end-labelled by T4 PNK phosphorylation with γ^{32}P ATP. 2 µl of 20µM oligonucleotide in 12 µl water was incubated with 2µl 10X PNK buffer, 5 µl γ^{32}P ATP (10 µCi/µl) and 1µl T4 PNK (10 U/µl) at 37 °C for 30 min and the reaction stopped by the addition of 2 µl 0.5 M EDTA.

2.3.20 Klenow fill-in labelling of oligonucleotide probes

This method was used in preference to end-labelling since it allows greater radio-chemical incorporation and should not label single stranded DNA. Complementary oligonucleotides were generated to create 5' overhangs so that at least one guanine on each strand was unpaired. These overhangs were filled in using Klenow DNA polymerase in the presence of (α^{32}P)dCTP. 2µl of 20 µM double stranded oligonucleotide was incubated for 30 mins with 1X Klenow buffer, 5 µl of (α^{32}P)dCTP (10 µCi/µl), 1 µl (2 U) Klenow DNA polymerase and 2 µl BSA (10 mg/ml).

2.3.21 Purification of oligonucleotide probes

As many of the oligonucleotide probes used in this study are too small to be purified by Sephadex or MicroSpin columns, native acrylamide gels were used to separate the labelled oligonucleotide from unincorporated nucleotide. 8% gels were prepared from a 30% acrylamide/0.8% N,N'-bis-methylene acrylamide stock in 1X TG buffer and polymerised using 0.06% TEMED and 0.03% fresh ammonium persulphate. Radiolabelled reactions were mixed with 1 volume of Band-Shift Sample buffer and electrophoresed in 1x TG buffer at 150V until the blue dye was 3 cm from the bottom of the gel. The apparatus was then dismantled and one of the two plates removed from the gel and replaced with Saran wrap. A film was placed over the Saran and the position of the film relative to the gel marked with a pen. The gel was subject to autoradiography for 2-3 mins. After developing the film the position of the labelled oligonucleotide was determined and the appropriate slice cut from the gel. Purification of oligonucleotides from gel slices by Bio-Trap is described elsewhere.

2.3.22 Random prime labelling of DNA probes.

This method was used to label DNA probes larger than 200 bp. Appropriate restriction digests or PCR reactions were performed and the probe of interest was isolated as described. 50 ng of DNA was diluted in a volume of 32 µl distilled water and heated at 100 °C for 5 min before cooling on ice for 5 min. 10 µl OLB buffer, 2 µl BSA (10 mg/ml), 5 µl (α^{32}P)dCTP (10 µCi/µl) and 1 µl (2 U) Klenow DNA polymerase were added in that order, and incubated at room temperature for a minimum of 4 hr. Unincorporated radionucleotides were separated from the reaction by centrifugation through a Sephadex G-50 column equilibrated in TE pH 8.0, in which labelled DNA probe is eluted and unincorporated dCTP is retained in the column, or alternatively, using MicroSpin S-200 HR columns (Pharmacia) according to manufacturer's instructions.

2.3.23 Screening of bacterial colonies by labelled-oligonucleotide hybridisation (1)

Fresh bacterial plates were overlaid with Hybond N+ membrane and colonies transferred. The position of the membrane to the plate was marked by needle holes through the membrane and agar, to allow realignment. Plates were returned to the 37 °C incubator for several hours to regrow. The membrane was removed from the plate and placed cell-side up on Whatman I paper soaked in 2 X SSC/5 % SDS for 2 min. Cell lysis, DNA denaturation and DNA fixation to the membrane was achieved by microwaving for 2.5 min at full power. Membranes were prehybridised in 500 mM sodium phosphate buffer, 7 % SDS, 1 mM EDTA for 30 min at 65 °C, before labelled oligonucleotide, prepared as described (heated at 100 °C for 5 mins then cooled on ice for 5 min), was added and incubated at 65 °C for a further 4 hr. Membranes were washed three times at 65 °C, each for 30 min, in 1XSSC, 1 % SDS, 1 mM EDTA prewarmed to 65 °C. Membranes were wrapped in cling film and exposed for a short time to X-ray film, typically 5 min. Positive colonies were selected by realignment of the autoradiograph to the bacterial plate.

2.3.24 Preparation of total RNA.

All buffers and solutions used in the preparation and analysis of RNA were prepared using DEPC treated water. Total RNA was prepared by the guanidine isothiocyanate: phenol extraction using Trizol reagent (Gibco BRL). Cells were lysed directly on 10 cm diameter tissue culture dish by repetitive pipetting of 3 ml of Trizol reagent. Frozen tissues were homogenised with 1 ml Trizol per 100 mg of tissue in a Ulra-Turrax T25 homogeniser (Jencons). After transfer to a Corex tube the mix was incubated for 5 mins at RT, to ensure dissociation of the nucleoprotein complexes. 0.2 ml of chloroform/ ml Trizol was added and the sample shook vigorously for 15 seconds. Following a further incubation of 2 mins the samples were centrifuged at 10,000 rpm for 20 min at 4 °C to achieve phase separation and the top aqueous phase transferred to a fresh tube. 0.5 volumes of isopropanol per ml of Trizol was added and the RNA precipitated for 10 minutes at RT. The RNA pellet was precipitated by centrifugation at 10,000 rpm for 10 min at 4 °C. The supernatant was carefully pipetted and the glassy RNA pellet

dissolved in RNAse-free water. The RNA was also quantified using spectrophotometric measurement as described.

2.3.25 Size fractionation of RNA on formaldehyde agarose gels

RNA was run on in a 1.5 % agarose formaldehyde . Typically in a preliminary gel 2 µg of RNA was examined and 40 µg was run for a northern blot. 1.5g of agarose was added to 10 ml of 10X RNA Running Buffer with 72.1 ml DEPC water and microwaved until molten. Once cool 17.9 ml of 38% stock formaldehyde was added and the gel poured in a gel former in a fume hood. The appropriate amount of RNA was ethanol precipitated and redissolved in 4.4 µl of DEPC water, mixed with 15.6 µl of RNA sample buffer and heated to 65 °C for 5 mins and chilled on ice. Then 8 µl of RNA dye solution was added and the sample loaded and run in 1XRNA Running Buffer at 20 mA O/N or 5 V/cm for 2-3 hr. 2 µg RNA size markers (0.24-9.5 kb, Gibco BRL) were prepared as for an RNA sample. Ethidium bromide was not included in RNA gels because of the possibility of promoting secondary structure. Therefore after running preliminary gels or after transfer of RNA the gels were stained with 0.2 µg/ml ethidium bromide in 1x RNA running buffer. Intact 18S and 28S (double the intensity of the 18S) ribosomal RNA bands were taken as an indication of RNA stability in the sample.

2.3.26 Northern blotting

When electrophoresis was complete, the gel was rinsed briefly in DEPC treated water before washing in 0.05 M NaOH for 30 min followed by 20X SSC for 45 min. Sodium hydroxide partially hydrolyses the RNA and improves the efficiency of transfer. RNA was then transferred by capillary action overnight onto Hybond N+ membrane (Amersham) in the presence of 20X SSC. After transfer to the membrane, the RNA was immobilised by baking the membrane at 80 °C for 1 hr under vacuum. The quality and efficiency of transfer of the RNA after electrophoresis and blotting was assessed by cutting from the membrane lanes containing additional sample and markers and staining with methylene blue solution for 3 min and destaining in water for 2-5 min. Visualisation

of the 18S and 28S rRNA bands were indicative of good transfer and lack of RNA degradation.

2.3.27 Northern blot hybridisation and washing

The northern blot was prehybridised for 1-2 hr in northern hybridisation buffer at 42 °C. Hybridisation buffer was prepared by heating 1X RNA hybridisation buffer to 65 °C, and adding salmon sperm DNA, that had been previously heated for 5 min at 100 °C and cooled on ice, to a final concentration of 250µg/ml. After prehybridisation, the appropriate radiolabelled DNA probe (prepared and purified as described) was denatured at 100 °C for 5 min, chilled on ice for 5 min and added to the hybridisation buffer and blot. Hybridisation continued overnight at 42 °C. The following morning the blot was washed with buffers prewarmed to 65 °C at increasing stringency, starting with two washes at 2X SSC, 0.1 % SDS for 15 min, followed by two washes at 1X SSC, 0.1 % SDS. After the first four washes, background activity was monitored with a Geiger-counter and further higher stringency washes included where necessary (0.5X SSC, 0.1 % SDS and 0.1X SSC, 0.1 % SDS, each for 15 min). The blot was wrapped in cling film (Saran wrap) and an autoradiograph established by exposure to X-ray film at -70 °C. Signals were quantified on a LKB Ultroscan XL Laser Densitometer.

2.3.28 Transient transfection of Cultured Cells by Calcium phosphate (2)

Cells were plated out 6-24 hr prior to transfection at a density of between 10^4 - 10^5 cells per well in a 24 well plate tissue culture dish to give a confluency at time of transfection of ~60%. One hour before transfection cells were given fresh prewarmed E4 + 10% FCS medium. DNA precipitation was carried out per well as follows: Solution A (50 µl 2X HBS, 0.5 µl NaH_2PO_4, 0.5 µl Na_2HPO_4) was mixed at a constant rate with solution B (no more than 3 µg total plasmid DNA, water up to 50 µl and 6 µl 2.5 M $CaCl_2$) while bubbling air through. As a negative control, vector alone or no DNA was also used. The mixed solution was vortexed well and then allowed to stand at room temperature for 15-20 min before vortexing again. 100 µl of solution was added drop-wise to each well and incubated overnight at 37 °C. Cells were washed three times with E4 before replacing with fresh growth medium.

2.3.29 Transient transfection of Cultured Cells by Electroporation

COS cells at 70% confluence were typsinised as described, centrifuged and approximately $2x10^6$ cells resuspended in 900 µl PBSA per transfection. 15 to 20 µg of DNA was mixed with the cells which were then placed on ice for 10 mins. The cells were subjected to 450 V (250 µF) using a Biorad electroporater (time constant 4.7-6) after which the cells were placed on ice for a further 10 mins before resuspending in 30 mls of E4/10% FCS and plating. The media was changed the following morning to remove any dead cells.

2.3.30 Transient transfection of Cultured Cells by FuGene™ 6 (Boehringer Mannheim)

$1x10^5$ HTB96 cells per well were plated 3 hours prior to transfection. For each well 3 µl of FuGENE reagent was added to 100 µl of E4 and incubated for 5 mins. In a second tube 1 µg of DNA was added and the E4 : FuGENE mix added drop-wise to it. After gently mixing the solution it was incubated at RT for a further 15 mins before the drop-wise addition to the wells.

2.3.31 Harvest of transfected cell cytosol

Three days post transfection (two days for COS cells) cells were washed three times at room temperature with cold PBSA. Cells were lysed in 80 µl LucLite buffer for 5 min on ice, before plates were tilted to allow the cell lysate to drain. Cell lysate was collected and removed to a pre-chilled 96 well plate

2.3.32 Luciferase reaction assay

Cells transfected with the pGL3 control plasmid (Promega) were assayed for luciferase activity as a control for transfection efficiency in experiments where CAT plasmids are used as reporters. Cells were transfected with 0.5 µg pGL3 plasmid together with the experimental plasmid as described above using the calcium phosphate protocol. 20 µl of

cell lysate was added to 20 µl LucLite substrate in a 96 well Microlite plate (Dynex). Counts were read using a Dynex plate reader .

2.3.33 Chloramphenicol acetyltransferase (CAT) activity assay.

20 µl of cell lysate was transferred to a safe-cap Eppendorf and heated to 68 ^{O}C for 5 min to remove de-acetylase activity. A cocktail of 20 µl 8 mM chloramphenicol (0.4 M stock diluted 1:50 immediately prior to use), 20 µl 0.5 mM acetyl CoA (made by diluting 54 mCi/mole ^{14}C acetyl CoA with cold 0.5 mM stock 1:10), 10 µl LucLite lysis buffer and 30 µl 250 mM Tris pH 7.8 was added to the cell lysate. The reaction was incubated at 37 ^{O}C for between 30 min and 2 hr, depending on the transfection efficiency as determined by luciferase activity. 100 µl cold ethyl acetate was added to each sample tube, capped tightly and vortexed twice. Samples were centrifuged at top speed in a bench-top microcentrifuge for 1 min and 80 µl of the top phase removed to a scintillation vial containing 2 ml of Ultima Gold scintillation fluid The aqueous phase was re-extracted by the addition of a further 100 µl ethyl acetate, vortexed and centrifuged as before, when 100 µl organic phase was removed and combined with the first. The emission from the samples was counted in a LS 1801 Beckman liquid scintillation counter.

2.3.34 Extraction of nuclear protein lysates from cultured cells (3)

Subconfluent cells grown in a 75 cm^2 tissue culture flask were harvested by typsinisation and washed in media containing 10% FCS to inactivate the trypsin. Following a wash with PBSA the cells were incubated on ice for 5 minutes with Nuclear lysis buffer containing fresh protease inhibitors and NP40 which is able to disperse the plasma membrane, but not the nuclear envelope. After centrifugation at 1250 g for 5 min at 4 °C the pellet was washed with Nuclearwashbuffer to remove any residual NP40 and re-centrifuged. The pellet was transferred to an Eppendorf and an equal volume of high salt extraction buffer added to give a salt concentration of 0.25>0.3 M. The sample was vortexed for 1-2 mins then spun in a bench top centrifuge for 2 mins at top speed. The supernatant was then aliquoted into tubes on dry ice and stored at -70 °C. One aliquot was retained for protein concentration assessment.

2.3.35 Bio-Rad (Bradford) micro protein concentration assay.

A serial dilution of protein of known concentration (0.5 - 10 mg/ml) was carried out in relevant buffer, typically nuclear extraction buffer, and 5 µl aliquots plated in duplicate in a 96 well microtitre plate. Buffer alone was plated as a reagent blank control. Dilution of samples of unknown protein concentration, were also carried out, and 5 µl of these samples were plated in duplicate in the 96 well microtitre plate. The Bio-Rad protein assay dye reagent concentrate was diluted 1/5 in distilled water and filtered through Whatman 2v paper to remove aggregates. 100 µl of this was added to each of the wells and the colour reaction allowed to develop. The absorption of samples was measured at an OD of 595 nm on a plate reader (Titertek Multiskan MC). The concentration of protein in the samples was calculated using the calibration curve (obtained by plotting OD 595 readings against the log protein concentrations of standards).

2.3.36 SDS-PAGE gel electrophoresis of proteins (4).

Proteins were fractionated by size on 10% acrylamide separating gel (prepared from a stock solution of 30 % acrylamide/0.8 % N, N' -bis-methylene acrylamide), containing 0.375 M Tris-HCl pH 8.8 and 0.1 % SDS and polymerised by the addition of 0.06 % TEMED and 0.03 % ammonium persulphate. Once polymerised a 6% stacking gel was poured on top of the separating gel and allowed to polymerise around a comb. The stacking gel contained 6 % acrylamide, 0.125 M Tris-HCl pH 6.8 and 0.1 % SDS and was polymerised as above. 30-50 µg of protein samples were mixed with an equal volume of 2x SDS-protein sample buffer plus 10% β-mercaptoethanol and boiled for 5 min before loading. 7µCi of ^{14}C methylated markers or 10 µl of rainbow molecular weight protein markers (Amersham International) were separated simultaneously to allow determination of molecular weight. Electrophoresis was performed in 1x SDS-PAGE running buffer at 100 V until the sample front had reached the running gel, when the voltage was increased to 150 V until the desired separation was achieved. After separation of radiolabelled proteins, gels were fixed in 20 % methanol/10 % acetic acid, dried, and exposed to film; or after separation of unlabelled proteins, gels were subjected to western blotting or Coomassiestained.

2.3.37 Coomasie staining of SDS-PAGE gels

After electrophoresis the gels were microwaved for 2 mins at a low setting in 1:4.5:4.5 Acetic acid: methanol: water with 0.05% Coomasie Brilliant Blue (Bio-Rad). Gels were de-stained for several hours in fresh washes of 1:4.5:4.5 Acetic acid: methanol: water, without the stain.

2.3.38 Western blotting

Western blotting was performed as described by (133) using a Bio-Rad transfer cell. Hybond-C extra nitro-cellulose membrane (Amersham International) and 4 sheets of Whatman 3MM filter paper were cut to the same size as the gel and pre-wetted in 1XTG buffer. In a tray of buffer the gel was laid on two sheets of 3MM paper and the membrane placed on top, followed by two more sheets of 3MM paper. In a second tray the cassette and two layers of plastic sponge were pre-wetted, taking care to remove all air bubbles from the sponge. The filter paper-gel-membrane-filter paper sandwich was placed between the two sponges and transferred to the transfer tank with the membrane closest to the positive electrode. Transfer was carried out at 30 V overnight in TG buffer at 4 °C. After transfer, the membrane was washed briefly in PBSA and the gel back-stained with Coomassie to check transfer.

2.3.39 Incubation with specific antibodies and detection of bound antibodies.

Non specific binding of antibodies to the membrane was blocked by pre-incubation with PBSA, 5% skimmed milk, 0.1 % Tween for 2 hr. All subsequent washes and antibody incubations were carried out in PBSA, 1 % skimmed milk and 0.1 % Tween unless stated otherwise. After blocking, the membrane was rinsed briefly and incubated with 0.1 μg/ml primary antibody. The membrane was then washed 4 times (1 x 15 min and 3 x 5 min). The secondary antibody (horse radish peroxidase conjugated anti rabbit) was incubated with the membrane for 1 hr at 1/2000 dilution. The membrane was subsequently washed 1 x 15 min, 1 x 5 min, and then 1 x 5 min with PBSA/ 0.1% Tween and 1 x 5 min with PBSA only. The ECL detection reaction was carried out according to the manufacturer's instructions and any signal detected by short exposure to X-ray film.

2.3.40 Fluorescence analysed cell scanning analysis (FACScan) for internal antigen (MUC1).

Monolayer cells were trypsinised, washed once in growth medium and a minimum of 5 x 10^5 cells were resuspended in 0.3 % saponin and incubated for 20 min at room temperature to make permeable the cell membrane. (0.1 % saponin was present in all incubations and washes to prevent reversal of cell permeability). The primary antibody CT-1 or Pre-immune sera (as a negative control) was diluted 1:2 in growth medium containing 10 % FCS, and 0.1% saponin and incubated with the cells for 1 hr before washing twice in growth medium containing 10 % FCS, and 0.1% saponin. Secondary antibody, for CT-1, fluorescein isothiocyanate (FITC) conjugated α rabbit was incubated with the cells at a concentration of 1:40 for 1 hour in growth medium containing 10 % FCS, and 0.1% saponin. The cells were washed twice in growth medium containing 10 % FCS, and 0.1% saponin and one in PBSA. The cells were resuspended in 0.5 ml PBSA

10000 cells were analysed by a Becton-Dickinson FACScan flow cytometer equipped with an argon ion laser tuned to 488 nm. FITC fluorescence was collected at 530 nm ±15 nm and cells were gated on scatter profiles to exclude dead cells.

2.3.41 Immunofluorescence of transfected cells

Adherent COS cells cultured in plastic chamber slides (Nunc) were fixed with paraformaldehyde (4%) for 10 minutes, treated with methanol for 2 minutes and stained by incubation with primary polyclonal antibody (1 μg/ml) against Sp1 or Sp3 (Santa Cruz Biotechnology) in 0.5 x FCS in PBS for 1 hour, followed by 3x 5 min washes with 0.5 x FCS in PBS and the addition of FITC-conjugated swine anti-rabbit IgG (DAKO) secondary at 1:500 dilution for 1 hour. Following 3 further washes of 2 x 5 min with PBS the slides were mounted with antifade and photographed using a Nikon microscope equipped with epiflourescence.

2.3.42 Preparation and Immunohistochemistry of Mouse tissue

Mouse organs and tissues were dissected and fixed in Methacarn (60% methanol, 30% chloroform, 10% acetic acid) for 2 hours then washed in 70% ethanol and processed for

wax embedding. (Embedding and section slicing was done by the Histopathology Department ICRF). 5 μm sections were de-waxed, and endogenous peroxide blocked by incubation for 30 minutes in 1% H_2O_2 in methanol. Non-specific binding of antibody was blocked by a 2 hour incubation in 50 % FCS in PBS. After washing in PBS a 1:40 dilution of directly biotinylated, affinity purified HMFG-1 monoclonal antibody or 1:80,000 dilution of humanised HMFG-1 was incubated on the section in 10% FCS in PBS for 1 hour. A control was included for each section by omitting the primary antibody and incubating with FCS alone. When the biotinylated antibody was used the sections were washed three times in PBS before being incubated in horseradish peroxidase (HRP) -conjugated avidin-biotin (ABC complex (DAKO)) for 30 min. When the humanised primary antibody was used a secondary peroxidase-conjugated anti-human antibody (DAKO) was incubated at 1:50 dilution for 1 hour before washing and incubation with the ABC complex. The complex was visualised by incubation of 3-3' diaminobenzidine tetrahydrochloride (or 0.5 mg/ml DAB + 0.15% H_2O_2) for 2-5 mins. Sections were counter stained in haematoxylin, dehydrated and mounted using DePeX (BDH).

2.3.43 Development and maintenance of Mouse lines

The *MUC1* 1.4 Kb promoter plus MUC1 cDNA (the *MUC1* mini-gene) homozygote transgenic mice were generated by Ros Graham and maintained by the ICRF animal unit Clare Hall. Mice bearing the MMTV LTR driving Polyomavirus-middle-T were made by Guy et al., (134) on a FVB background. We obtained four male heterozygous mice via the Beatson Institute. These too were maintained at Clare Hall by members of the ICRFs animal unit, namely Gary Saunders. Polyomavirus-Middle-T transgenic mice were crossed with *MUC1* mini-gene transgenic homozygote female mice. Progeny were screened by PCR analysis of tail-snip DNA.

2.3.44 Band-shift assays.

Double-stranded oligonucleotide probes were labelled and purified as described above. From the Centricon concentrate the volume was made up to 200 μl to give a concentration of approximately 0.2 μM of labelled oligonucleotide. In a final volume of

20 µl 1x Band-shift buffer was combined with 2 µg of poly(dI:dC)•(dI:dC) and 2-4 µg of nuclear proteins with or without unlabelled competitor double stranded oligonucleotide (typically 20-100 fold molar excess over the probe) or antibody (typically 2-5 µg). After gentle mixing the binding reaction was incubated on ice for 20 mins. Each reaction was mixed with band-shift sample buffer, and loaded onto a pre-run 6% native acrylamide gel. The reaction was electrophoretically separated at 75 V at room temperature for 15 mins, then 100 V for a further hour or until the probe was close to the base of the gel. (In cases where resolution of close running complexes was required the electrophoresis was carried out for an extended period). Gels were dried on a vacuum drier and exposed to X-ray film for 6-16 hours

2.3.45 South-westerns, (5, 6).

50 µg of nuclear proteins were subjected to reducing SDS-PAGE separation and western blotting as described above. Proteins on the membrane were denatured by incubation in 1x South-western HBB buffer plus 6M Guanidine-HCl, for 2x 10 mins at 4 °C. Proteins were allowed to renature by decreasing the Guanidine-HCl concentration gradually by 10 minute incubations at 4 °C through doubling dilutions of south-western HBB buffer five times, bringing the concentration of Guanidine-HCl to 3M, 1.5 M, 0.75 M, 0,375 M and 0.187 M. The membrane was incubated twice with 1x south-western HBB buffer alone at 4 °C before being incubated in blocking solution (BLOTTO) over night at 4 °C. The following morning the buffer was changed for 1x south-western binding buffer containing 2 x 10^6 cpm/ml labelled oligonucleotide probe and 5 µg/ml poly(dI-dC):poly(dI-dC) and incubated for 4 hours at 4 °C. Membranes were then washed four times for eight minutes each in south-western binding buffer alone, covered with Saran wrap and exposed to X-ray film for 24 hours at -70 °C.

2.3.46 DNA screening of Lamda-Zap cDNA library.

The cDNA library used was a λZAPII ZR-75-1 cDNA library (oligo(dT) primed) constructed by R. White (Molecular Endocrinology Laboratory ICRF) and screened according to the manufacturers instructions (Stratagene). The XL-1 blue host cells used

were prepared as follows. A colony from a fresh (1-2 day) L-agar/tetracycline plate was used to inoculate an L-broth culture (50 ml) supplemented with 0.2% maltose and 10 mM $MgSO_4$. The culture was grown O/N at 30 °C with vigorous shaking. The cells were spun at 1000g for 10 min, gently suspended in 0.5 volumes of 10 mM $MgSO_4$ to an OD_{600} of 0.5 and incubated with phage at 37 °C with shaking for 15 mins. To titre the library 400µl of XL-1 blue cells (infected with phage) and 3 ml of top agar were plated out on a 10 cm NZY plates. Once the titre had been established the library was plated out with 2 ml of host cells and 30 ml of top agar per 15 cm x 15 cm plate to 250,000 pfu's/plate (2 $x10^6$ plaques were screened in total). Plates were incubated for 3 hours at 42 °C. Hybond C+ membrane soaked in 10 mM IPTG and air-dried was carefully placed on the agar surface and a sterile needle used to orientate the filter. The plates were then returned to the incubator for a further 3 hours at 37 °C to induce protein production and then chilled for 1 hour to prevent the top agar from sticking to the filters After carefully removing the membrane a second was added to the agar and incubated overnight. Membranes were taken through the denature/renaturation procedure used in south-westerns in the presence of zinc. Spots in the same position on duplicate membranes were interpreted as positive. 30 putative positive plaques where cored and eluted into 1 ml of SM buffer containing 40 µl chloroform. Eluted phage were then rescreened as above but on smaller 10 cm plates. Under these conditions the plaques where allowed to grow over 3mm before induction with IPTC. Nine candidates were isolated and the pBluescript double-stranded phagemid with the cloned cDNA insert excised. 200 µl XL-1 blue cells, 200 µl UniZAP phage (>$1x10^5$ phage)and 1 µl of R408 helper phage ($1x10^6$ pfu/ml) were incubated at 37 °C for 15 mins. 1-100 µ1 of the mixture was plated out on L-agar/Amp' plates and incubated O/N at 37 °C. The next day colonies containing the pBluescript plasmid were selected for sequencing.

Chapter 3. MUC1 expression in vivo.

3.1 Aims.

Previous transgenic work showed that a 10.6 kb genomic fragment of the human *MUC1* gene was sufficient for tissue specific expression and high expression at lactation and in tumours of transgenic mice (table 1) (46, 51). Since just 1.4 kb of the 5' region can drive expression of reporter genes in transient transfection of epithelial cell lines we wished to know if this region was also sufficient for correct expression *in vivo*. Mice were generated that had 1.4 kb of the *MUC1* 5' region fused to the MUC1 cDNA, referred to in this thesis as the "*MUC1* mini-gene" (see the appendix A1 for details). These mice expressed the MUC1 protein in a tissue specific manner analogous to that in the human and the transgenic mice bearing the genomic *MUC1* fragment. The aims of the work presented in this chapter were to show whether mice carrying the *MUC1* mini-gene also expressed large amounts of MUC1 in mammary tumours.

3.2 Results

3.2.1 Tissue-specific expression of the *MUC1* mini-gene in FVB x (CBA x C57) mice.

In order to generate mammary tumours, mice bearing the *MUC1* mini-gene were crossed with transgenic mice carrying the Polyomavirus-Middle-T oncogene (Poly-T) transcribed from the MMTV promoter. 100% of female transgenic mice with this oncogene construct develop multifocal mammary adenocarcinomas which metastasise to the lung (134) and therefore transgenic females cannot be used for breeding. Males heterozygote for Poly-T were crossed with female mice homozygote for the *MUC1*-mini-gene. Progeny were identified as Poly-T positive or negative using a PCR-based method to analyse the DNA (see figure 3.2a). Since the Poly-T transgenic mouse is in the background FVB it was important to first confirm that the tissue specificity of the *MUC1* mini-gene previously observed in the CBA x C57 background was maintained. *MUC1* positive, Poly-T negative females were used for this analysis and several organs

were prepared and subjected to immunohistochemistry with biotinylated HMFG-1 monoclonal antibody. HMFG-1 is a mouse monoclonal antibody raised to the protein core of human MUC1. It is able to bind both heavily glycosylated MUC1, characteristic of lactating breast, and aberrantly glycosylated MUC1 found in many breast tumours (9, 32). Table 2 shows the results of staining various tissues from a female mouse negative for Poly-T resulting from a Poly-T heterozygote X *MUC1*-mini-gene homozygote cross. As expected the mini-gene showed epithelial specific expression as observed in the previous strain. The Poly-T parental line, negative for the *MUC1* mini-gene had no expression of the HMFG-1 epitope in any tissue tested (not shown).

Table 2. Tissue specific expression of MUC1 in transgenic mice bearing the *MUC1* mini-gene in the FVB x (CBA x C57) background.

Tissue	HMFG-1
Heart	-
Muscle	-
Salivary Gland.	++
Pancreas	++
Resting mammary	++
Lactating mammary	++

3.2.2 Indirect immunohistochemistry using the humanised HMFG-1 antibody.

The biotin/avidin detection system is a sensitive assay and the biotinylated antibody showed similar strong staining at all MUC1 positive sites. In order to gain a better differentiation between low and high levels of MUC1 expression while avoiding background staining the humanised HMFG-1 antibody was used. This antibody retains the specificity of the mouse monoclonal, but since the heavy and light chains of the immunologlobulin are of human origin anti-human secondary peroxidase-conjugated antibodies can be used which do not bind to endogenous mouse immunoglobulins. This indirect immunohistochemistry is less sensitive than the direct avidin/biotin assay and allows differences in the levels of MUC1 protein to be visualised. The antibody was titrated on sections from lactating mammary gland from mice containing the 10.6 kb genomic *MUC1* fragment to give a low to moderate detection of the MUC1 epitope (not shown). The results in figure 3.1 show that the humanised HMFG-1 antibody gives

weaker staining than the biotinylated HMFG-1 antibody. For a comparison of detection levels between the humanised antibody and the biotinylated antibody compare the salivary gland sections in figure 3.1 d and f. When this technique was employed minor strain differences were apparent in that the resting mammary gland and salivary gland showed apparently higher levels of MUC1 in the FVB x (CBA x C57) background compared to CBA x C57 background (compare the resting mammary gland tissue sections in a and c and the salivary gland tissue sections b and d in figure 3.1).

3.2.3 MUC1 expression in the resting mammary gland, and mammary tumours.

Satisfied that the new strain expressed the *MUC1* mini-gene in a tissue specific manner we wished to see whether mammary tumours brought about by Poly-T transformation expressed high levels of the gene. 28 offspring were generated from homozygote *MUC1* mini-gene x heterozygote Poly-T cross. 14 had the Poly-T transgene and nine of these were female. All nine females developed multiple tumours (>5 mm) in all mammary glands by the ninth week. Lungs of all nine also developed small metastases. A female negative for the Poly -T gene, but positive for the *MUC1* mini-gene from the same cross was included in the study as a control (a section from resting mammary gland of this mouse is shown in figure 3.2 b). One of the elderly Poly -T male parents (negative for the *MUC1* mini-gene) which had developed a mammary tumour was included as a negative control (a section from the tumour is shown in figure 3.2 c).

32 separate tumour nodules from nine female offspring of the cross showed strong expression of MUC1 as detected by immunohistochemistry using the humanised HMFG-1 antibody. As shown in figure 3.2 MUC1 in the resting mammary gland is found as deposits in the lumen of the duct and on the apical surface of luminal cells of the duct (b) (and figure 3.1c), whereas in the tumour there are examples of MUC1 expression in the cytoplasm and surface of invasive cells (d), in addition to apical expression in remaining luminal-type structures (e). The expression of MUC1 is widespread across the tumours yet heterogeneous, with both positive and negative areas (f). In some cells it appeared highly expressed (for example in panel d compare the area

of tumour morphology at the base which has higher deposits of diaminobenzidine than the apparently normal morphology at the top of the duct). All mice positive for the *MUC1* mini-gene, regardless of their Poly-T status, expressed the protein in a tissue specific manner (not shown).

From the immunohistochemistry it was clear that the expression of MUC1 in tumours was high. In order to discover if this was also the case at the level of RNA production, and therefore likely to be due to transcription from the promoter, a northern analysis of RNA from resting mammary glands and tumours, from a further 17 progeny resulting from the Poly-T x MUC1 mini-gene cross was undertaken. Lactating mammary gland from several Poly-T negative mice was also included to show whether the *MUC1* mini-gene was sufficient to direct high expression of MUC1 at lactation. As can be seen from the blot in figure 3.3a MUC1 mRNA levels are high in 9/10 tumours whereas it is low to undetectable in the resting mammary glands. The mammary glands from four lactating mice also expressed high levels of the gene. Phosphor-imager analysis was used to quantify the relative radiolabel per lane and the average fold increase of MUC1 in tumours and lactating mammary gland over resting mammary gland was 59 and 51 fold respectively (the values for each lane can be found in appendix A2).

3.3 Summary

These findings show that the 1.4 kb promoter plus the MUC1 cDNA are sufficient for several aspects of MUC1 gene expression including tissue specific expression and high expression at lactation and in mammary tumours. This pattern is similar to the expression pattern of the endogenous gene in humans and identical to the expression pattern from the 10.6 kb genomic *MUC1* fragment (135, 46) (and table 1.1). It is likely that the MUC1 promoter contributes to control of expression for two reasons; *in vitro* transient transfection experiments using plasmids containing the 1.4 region direct expression of reporter genes in epithelial cells, but not in non-epithelial cells (111, 113) and, importantly, transgenic animals using the 1.4 kb region to express *Ras* variants

express the gene in a tissue specific manner (N. Freshney, Signal Transduction Laboratory, ICRF, personal communication). Together with these results it can be concluded that sequences contained within the 1.4 kb region are responsible for directing tissue-specific transcription of the gene and are likely to be responsible for high expression in tumours and at lactation. Therefore the 1.4 kb region is a legitimate target for further study.

3.4 Chapter 3 : Figures

Figure 3.1. MUC1 HMFG-1 epitope immunolocalisation in histological sections of organs from *MUC1* mini-gene transgenic mice.

Sections from mice heterozygous for the *MUC1*-mini-gene on a CBA x C57 or (CBA x C57) x FVB background prepared and immunostained with either the biotinylated HMFG-1 antibody or the humanised HMFG-1 antibody (as described in the Materials and Methods).

Panel	Tissue section	background	HMFG-1 Antibody
a	Resting mammary gland,	CBA x C57,	humanised,
b	Salivary gland,	CBA x C57,	humanised,
c	Resting mammary gland,	(CBA x C57) x FVB,	humanised,
d	Salivary gland,	(CBA x C57) x FVB,	humanised,
e	Resting mammary gland,	(CBA x C57) x FVB,	biotinylated,
f	Salivary gland,	(CBA x C57) x FVB,	biotinylated.

Note that sections **c** & **e**, and **d** & **f** are from the same paraffin block (i.e. the same organ from the same mouse). The presence of the MUC1 HMFG-1 epitope is indicated by brown / reddish deposits of diaminobenzidine. Each section was counter stained with haematoxylin (blue). Controls where the primary antibody or the biotinylated antibody were omitted showed no staining (not shown).

Scale bar = 0.1 mm.

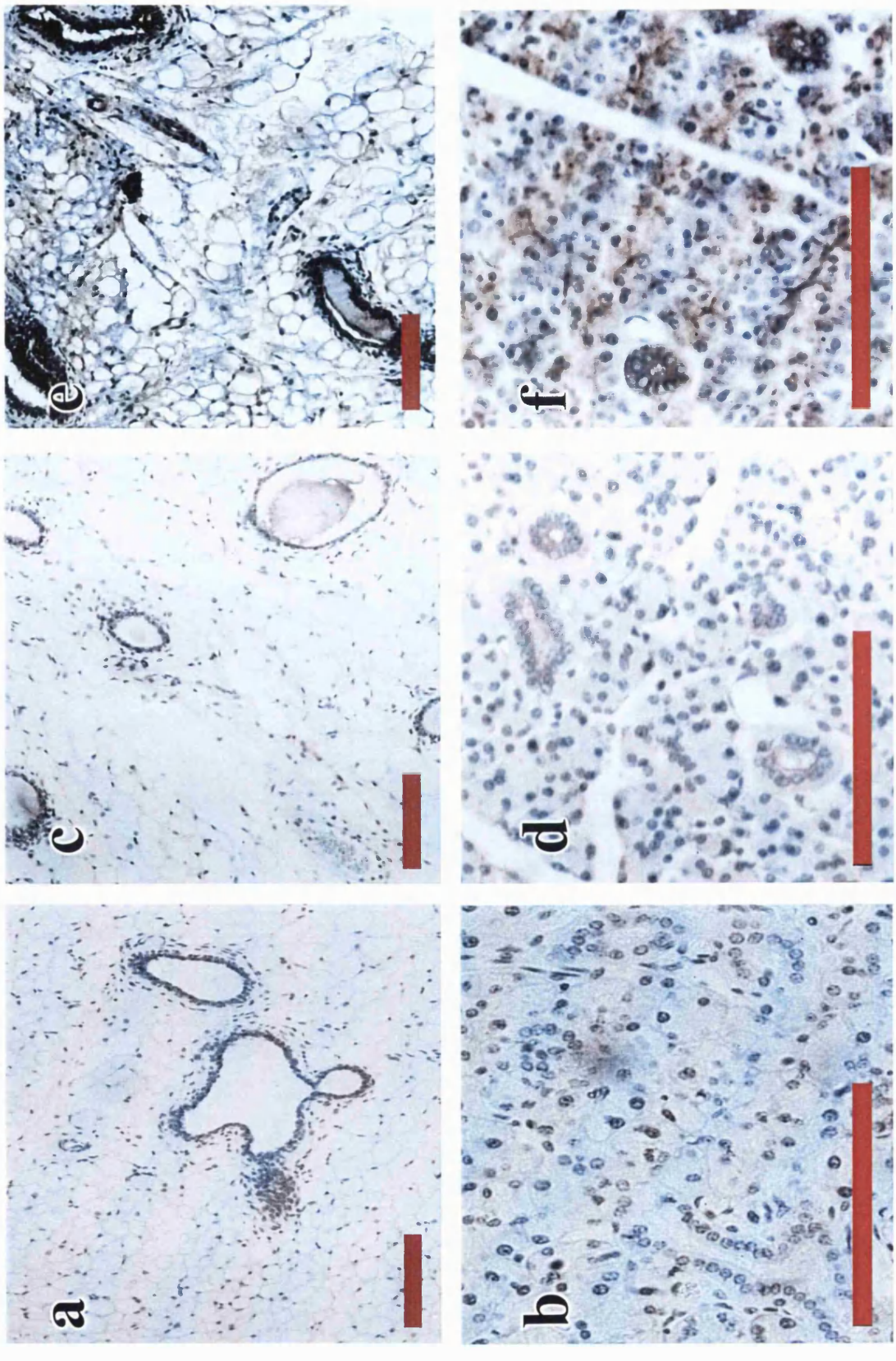

ab
f
c
d
a
b

Figure 3.2. MUC1 HMFG-1 epitope immunolocalisation in histological sections of mammary gland and mammary tumours from transgenic mice.

a) An example of the PCR screen for the Poly-T construct in mouse-tail DNA is shown. Template DNA was; a plasmid containing the Poly-T gene (lane 1), mouse genomic DNA containing the Poly-T gene (lane 2), no DNA (lane 8), or genomic DNA prepared from tail-snips from the offspring of the *MUC1* mini-gene homozygote x Poly-T heterozygote cross (lanes 3-7). Lanes 1 and 2 reveal the expected 130 bp product, as do lanes 3 and 7, indicating that these mice carry the Poly-T construct.

b-g) Indirect immunostaining of mammary gland sections with humanised HMFG-1 antibody.

Panel	Tissue section	*MUC1* mini-gene	Poly-T oncogene
b	Resting mammary gland,	+	-
c	Mammary tumour,	-	+
d	Mammary tumour,	+	+
e	Mammary tumour,	+	+
f	Mammary tumour,	+	+
g	Mammary tumour,	+	+

The presence of the MUC1 HMFG-1 epitope is indicated by brown / reddish deposits of diaminobenzidine. Each section was counter stained with haematoxylin (blue). Controls where the primary antibody was omitted showed no staining (not shown). Scale bar = 0.1 mm, except for **f**, where the scale bar = 0.5 mm.

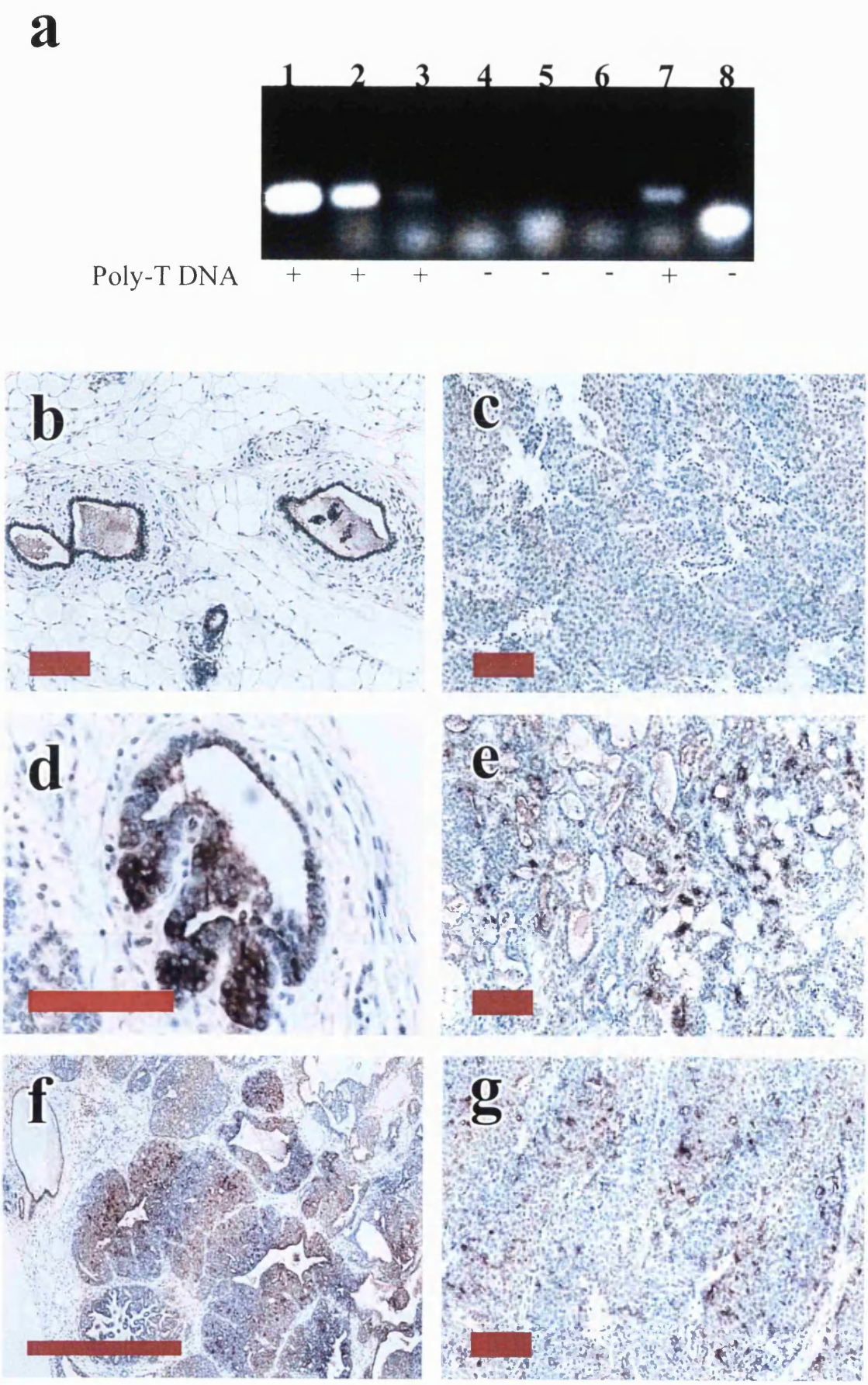

a
1 2 3 4 5 6 7 8
Poly-T DNA + + + - - - + -
b
c
d
e
f
g

Figure 3.3. Detection of MUC1 transcript in the mammary gland and tumours of *MUC1* mini-gene transgenic mice.

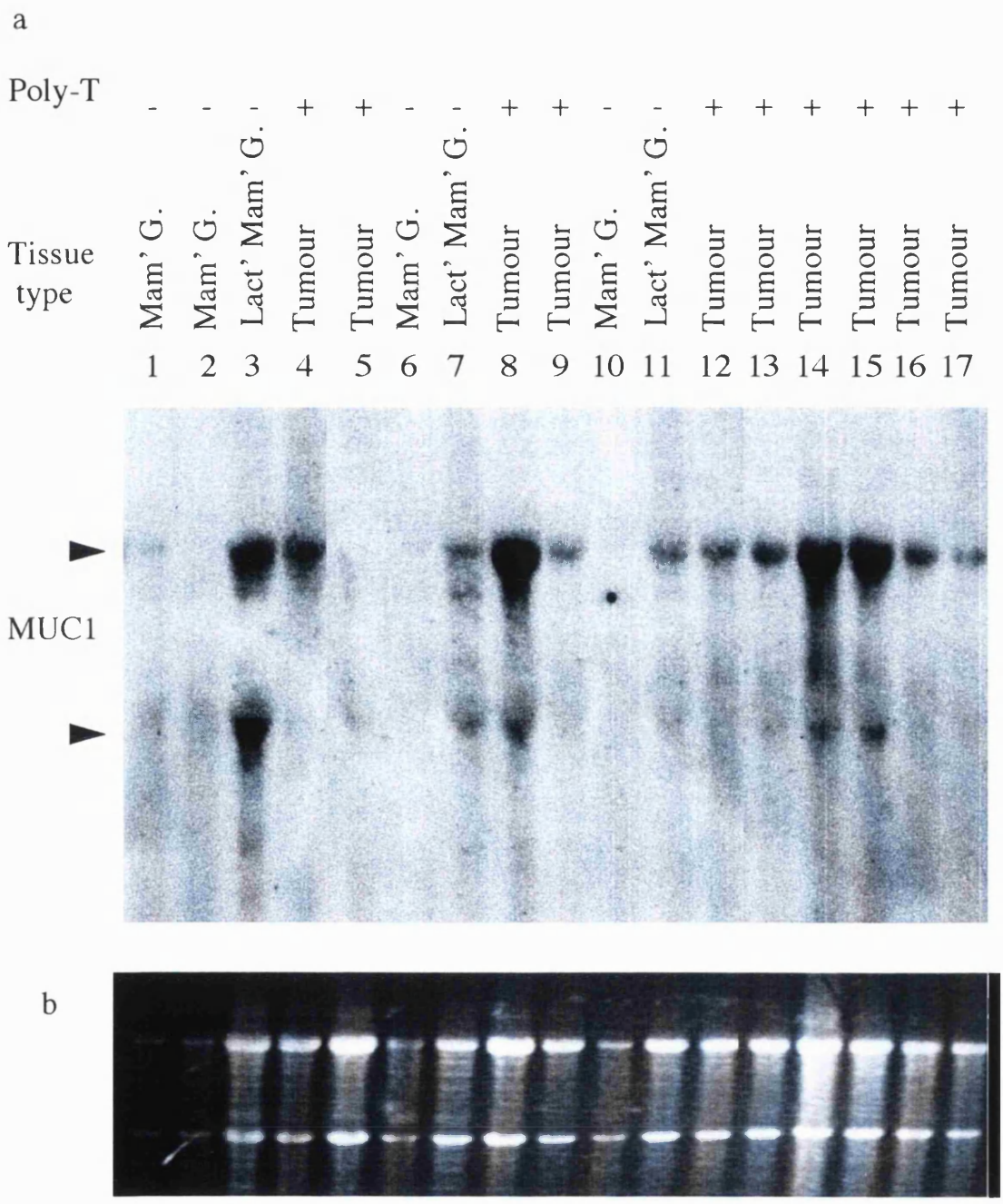
a) Northern blot analysis of 50 μg of mammary gland total RNA from the progeny of *MUC1* mini-gene homozygote x Poly-T heterozygote cross, the Poly-T status of each mouse is shown, all mice were positive for the *MUC1* mini-gene. Lactating mammary gland was processed at 9-10 days *postpartum* from Poly-T negative females. Once tumours from Poly-T positive females became palpable they were processed. The blot was probed with radiolabelled cDNA corresponding to the tandem repeat region of the *MUC1* gene.

Mam' G. = resting mammary gland,
Lact' Mam' G. = lactating mammary gland,
Tumours = mammary tumours .

The amount of radiolabel annealed to RNA in each sample lane was quantified with a phosphor-imager (the values for each lane are shown in appendix A2).

b) 1/10 of the RNA used in the northern blot was analysed simultaneously on a formaldehyde agarose gel to show loading.

Figure 3.3


a
Poly-T
- - - + + - - + + - - + + + + + +
Tissue type
Mam' G.
Mam' G.
Lact' Mam' G.
Tumour
Tumour
Mam' G.
Lact' Mam' G.
Tumour
Tumour
Mam' G.
Lact' Mam' G.
Tumour
Tumour
Tumour
Tumour
Tumour
Tumour
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
MUC1
b

Chapter 4. Functional analysis of the MUC1 promoter.

4.1 Aims

In previous work the region -152/-62 had been shown to be required for transcriptional activity of the MUC1 promoter. Mutations of sequences within that region, at the -97 GC box and -89/-77 EMUC sequence, altered the cell-type specificity of transcription from the promoter (figure 1.9) (111). The aims of the work shown in this chapter were to demonstrate whether *MUC1* sequences within -152/-62, were sufficient to confer epithelial-specific transcription, and if so to show which DNA binding factor(s) within that sequence were responsible for the activity.

4.2 Results

4.2.1 MUC1 is expressed in T47D and ZR75 adenocarcinoma cells but not HT1080 fibrosarcoma cells and HTB96 osteosarcoma cells.

The MUC1 status of four cell lines to be used in this study was checked by flow cytometry using the MUC1-cytoplasmic-domain specific CT-1 rabbit polyclonal antibody. Figure 4.1b shows that adenocarcinoma cell lines, T47D and ZR75, expressed high levels of the protein, whereas the fibrosarcoma cell line HT1080 and the osteosarcoma line HTB96 expressed none. These results confirm earlier northern analysis of MUC1 mRNA expression in these cells, shown in figure 4.1 a.

4.2.2 Mutations in the -97 GC box and EMUC sequences @ -89/-77 of the MUC1 -1400/+33 promoter alter the cell-type specificity of transcription.

The results of Kovarik *et* al., (1993) (111) were confirmed in the four cell lines by the transient transfection with four plasmids containing the wild-type promoter or a promoter mutated in the -97 GC box, EMUC sequence or both. In non-epithelial cells single mutations had little effect, while the double mutation increased activity synergistically (see figure 4.2). In epithelial cells however mutation of either sequence significantly decreased the activity of the promoter. Since the proximal deletions in the

-97 GC box and -89/-77 EMUC sequences together derepress transcription in non-epithelial cells and reduce it in epithelial cells these sites appear to have a dual function.

4.2.3 MUC1 sequences -119/-62 can direct an epithelial specific pattern of expression to the constitutively active thymidine kinase promoter.

The above results, together with the observation that -152/+33 is able to drive cell-type specific expression in the context of the pEn-CAT vector while -62/+33 cannot (figure 1.9) led to the idea that the -97 GC box and EMUC sequences may constitute a *cis*-acting element that might direct epithelial specificity of MUC1 transcription, suppressing expression in non-epithelial cells and allowing it in epithelial cells. Sequences -152/-117 were unable to form complexes with nuclear extracts *in vitro* (not shown) whereas sequences -119 to -62 could (see chapter 5), hence the sequences -119/-62 were cloned upstream and adjacent to the constitutively active HSV thymidine kinase promoter (TK promoter) to examine the effect they might have on transcriptional activity. This promoter was used for two reasons. Firstly its constitutive activity means that transcriptional repression as well as activation could be studied, secondly the small size of the promoter allowed the MUC1 sequences to be cloned in a position relative to the TATA box similar to that in the MUC1 promoter. An Sp1 consensus site was also cloned into the enhancer site (>1.8 kb from the TK promoter in either direction) to mimic the distal Sp1-site of MUC1 (@-574) which has been shown to be an important enhancer of MUC1 transcription (112, 69) (construct details can be found in appendix A3). Figure 4.3 shows that the effect of including MUC1 -119/-62 sequences adjacent to the heterologous TK promoter was to increase transcription in epithelial cells, T47D and ZR75, and to repress it in non-epithelial cells, HTB96 and HT1080 (4.3c). This effect was boosted slightly if the Sp1 consensus sequence at the enhancer position was included (4.3d), whereas the effect of Sp1 in the enhancer position alone had no correlation with MUC1 expression (4.3b). These results support the idea that MUC1 sequences -119/-62 are able to direct epithelial specific expression by boosting expression in epithelial cells and repressing it in non-epithelial cells.

4.2.4 Mutations in the -97 GC box and EMUC sequences eliminate the ability of the -119/-62 sequences to direct an epithelial pattern of expression.

To test whether these effects were due to the -97 GC box and EMUC sequences mutations were made in the -119/-62 sequence identical to those in the double mutant of the 1.4 kb promoter shown in figure 4.2 (also highlighted in red in the legend of figure 4.4). These mutations had already been shown to prevent complexes forming with corresponding oligonucleotides in band-shift experiments (111). In the epithelial cells these mutations eliminated the enhancing effect of the -119/-62 MUC1 sequences, whether with or without the Sp1 consensus sequence in the enhancer position. This observation is similar to that seen when the same mutations were made in the context of the 1.4 kb MUC1 promoter (figure 4.2). In non-epithelial cells the effect of the mutations was to further inhibit transcription from the TK promoter. These results, together with the observation that the mutation reduced transcription below 100% of the TK promoter in epithelial cells, suggests that the mutation itself was actively repressive or that a more repressive factor was able to bind in the absence of factors bound to the -97 GC box and EMUC sequences. When the same mutations were made in the 1.4 kb MUC1 promoter transcription was increased in non-epithelial cells, suggesting that other factors associated with the rest of the promoter may have over come the repressive effect of the mutation. Presumably the TK construct lacks sites for these factors and MUC1-expressing cells lack the factors themselves.

4.2.5 A consensus Sp1 sequence proximal to the TK promoter can substitute for the MUC1 -119/-62 sequences.

The -119/-62 MUC1 sequence contains potential sites for many factors, some of which are defined in the next chapter as being Sp1, Sp3 an E-box factor USF, a possibly new factor named SpA and two other unknown non-E-box factors. To assess the contribution of Sp1/Sp3 relative to the other factors in directing epithelial specific transcription a single consensus Sp1 site, which binds Sp1 and Sp3, was cloned

upstream of the TK promoter in place of the MUC1 sequences. (See figure 4.5a for a line up of the Sp1 consensus sequence and MUC1 sequences -119 to -62). Remarkably the Sp1 consensus sequence at the position proximal to the TK promoter could substitute for the MUC1 sequences and direct high expression in epithelial cells and reduced expression non epithelial cells. The differential was greater when a second Sp1 site was included at the enhancer position. The ability to direct an epithelial pattern of expression is dependent on the position of the site since an Sp1 consensus site in the enhancer region alone did not have this activity (figure 4.3b). These observations suggest that the majority of the epithelial specific transcription from TK promoter conferred by the MUC1 sequences -119/-62 was due to the activity of the Sp1-like sequence at the -97 GC box and not to the sequences outside it.

4.2.6 Increased expression of Sp1 can activate expression from the MUC1 promoter in non-epithelial cells.

These surprising results lead to the question of whether the relative activities of Sp1 and Sp3 are responsible for the activation and/or repression of MUC1 expression. Immunoblots of nuclear proteins from epithelial and non-epithelial cell lines showed no differences in the amounts of Sp1 (see chapter 5, figure 5.9). Sp3 appeared to be present at higher levels in the non-epithelial cells tested which maybe significant since Sp3 is able to repress the transcription of many genes (136). However immunoblots do not distinguish between the levels of total factor and the levels of factor available to the regulatory sequences of a gene since nuclear extracts contain proteins stripped from the DNA by high salt. To test whether Sp1 or Sp3 could effect transcription from the MUC1/TK promoter and Sp1 consensus/TK promoter chimeras Sp1 and Sp3 expression vectors or expression vector alone were co-transfected with the TK constructs. The results in figure 4.6 show that Sp1 activated transcription from the TK-constructs in the non-epithelial cell line particularly when the MUC1 (-119/-62) or Sp1 consensus sequences were adjacent to the TK promoter. In T47D cells the effect of Sp1 on transcription was seen only on the Sp1 consensus TK construct. Sp3 expression had little effect on these constructs whether in an expressing or non-expressing cell. These

experiments suggest Sp1 may be a limiting factor in MUC1 expression and may relate to the tissue specific expression directed by the -119/-62 sequences.
It was important to see if expression of Sp1 could influence the more complex wild-type 1.4 kb promoter. Figure 4.7 shows that Sp1 co-transfection was able to boost expression from the -1400/+33 MUC1 promoter. As in the previous experiment the effect of Sp1 was most dramatic in the non-epithelial cell type where MUC1 transcriptional activity was increased several fold. Expression of Sp1 in the epithelial cell line also increased transcription from MUC1 but the fold increase was less. The effect of co-transfection of Sp3 on transcription from MUC1 was most marked in the MUC1-expressing cell line T47D where it was halved. This experiment further supports the premise that Sp1 is of limited availability in non-epithelial cells.

To discover whether the endogenous gene could be regulated by increased amounts of Sp1 the CT-1 antibody binding activity in the non-epithelial cell line, HTB96, was measured by FACS analysis with and without Sp1 transfection. As in the previous experiment the MUC1 promoter has low activity in non-epithelial cells and transfection with the MUC1-minigene plasmid (see chapter 3 and appendix, A1) resulted in no change of luminescence over untransfected cells or cells incubated with the pre-immune sera (compare figure 4.8 a and b, the pre-immune sera is not shown). The addition of Sp1 resulted in an increase in fluorescence, suggesting a significant increase in transcriptional activity (d). Furthermore transfection of Sp1 expression vector alone resulted in readily detectable expression of the CT-1 epitope (e). This result suggests that the inactivity of the *MUC1* gene in this cell type is reversible and may depend on the activity of Sp1.

4.3 Summary

In previous studies the region between nucleotides -152 and -62 was shown to be required for transcription in epithelial cells (111). Data herein shows that the region -119/-62 conferred an epithelial specific expression pattern to the constitutively active TK promoter, increasing activity in epithelial cells and repressing it in non-epithelial

cells. *In vitro* experiments showed that Sp1 and Sp3 bind the region. Therefore by comparing the influence of a consensus Sp1 site (which binds Sp1 and Sp3 (137)), and the -119/-62 region on transcription from the TK promoter it was possible to show that a single Sp1 site could substitute for the activity of MUC1 sequences. These results suggest that within the -119/-62 sequence it is the -97 GC box that is responsible for the ability to direct a tissue specific expression pattern. The position of the GC box is important since a single Sp1 site in the enhancer region could not confer the activity. In the context of the TK promoter the contribution of factors binding outside the -97 GC box, such as those binding to the EMUC motif, appears to be slight.

Co-transfection experiments shown here provide evidence that the MUC1 promoter can be regulated by both Sp1 and Sp3. Co-transfection of the 1.4 kb promoter with Sp3 expression vector reduces expression by ~half in the epithelial cell line T47D, demonstrating that Sp3 represses MUC1 transcription and/or displaces more activating factors. Sp3 had little effect in the non-epithelial cell line, HT1080, presumably because the gene is already inactive. Band-shift experiments described in the next chapter have shown that Sp1 and Sp3 bind to the same nucleotides in the -97 GC box, suggesting that these factors are in competition for the same binding site, which has also been shown to be so on the Sp1 consensus sequence (137). While it is likely that the repressive effect of the MUC1 sequences -119/-62 and the Sp1 consensus site on transcription from the TK promoter is through Sp3, there is no direct evidence for this and another, unidentified factor may be involved. Most significantly MUC1 transcription can be activated in non-epithelial cells by over expression of Sp1. A ~7 fold increase in transcription is seen from the 1.4 kb promoter and the TK construct containing MUC1 sequences -119/-62 when Sp1 is cotransfected. No expression of MUC1 is detected in HTB96 osteosarcoma cells yet transfection of Sp1 leads to readily detectable levels of the MUC1 protein from the endogenous gene. Taken together these results suggest that the balance of Sp1 and Sp3 available to the promoter can modulate its activity. The availability and or activity of Sp1 in particular may be important in tissue-specific regulation of MUC1 gene expression.

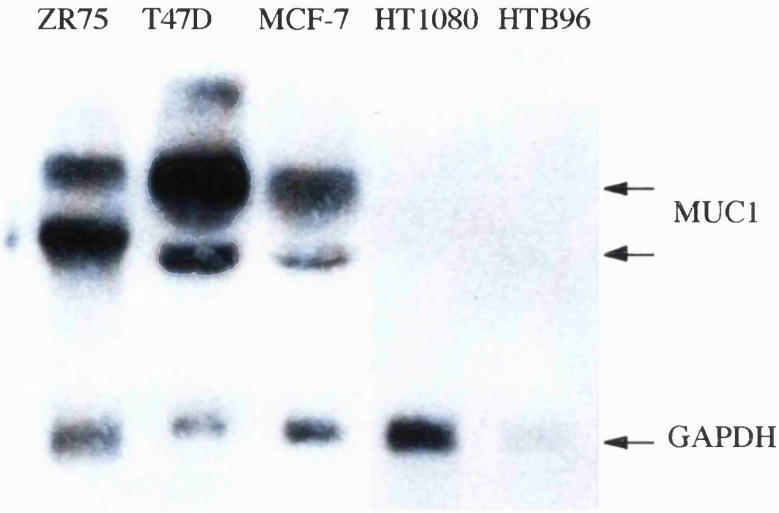
4.4 Chapter 4 figures

Figure 4.1. Expression of MUC1 in various human cell lines.

a) Northern analysis of MUC1 mRNA in various human cell lines. T47D, ZR75 & MCF-7 are derived from adenocarcinomas, HT1080 from a fibrosarcoma and HTB96 from an osteosarcoma. The blot was probed with labelled cDNA of MUC1 and glyceraldehyde-3-phosphate dehydrogenase (taken from Kovarik *et al.*, (111)).

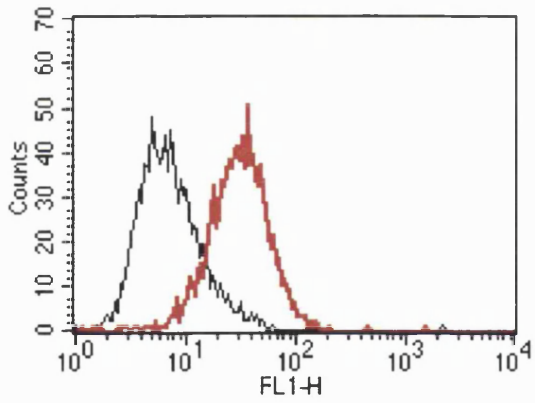
b) Profile of CT-1 epitopes in the four cell lines indicated. Cells were incubated with pre-immune sera (PI) or sera from rabbits immunised with the cytoplasmic region of MUC1 (CT-1), followed by FITC-conjugated swine anti-rabbit immunoglobulin (see Materials and Methods for details). Binding was analysed by FACSCalibur (Becton-Dickinson). Cells were gated by forward and side scatter to exclude dead cells and debris. The X-axis shows cell number, the y-axis log relative flourescence intensity. The geometric mean intensity of each population is indicated.

a



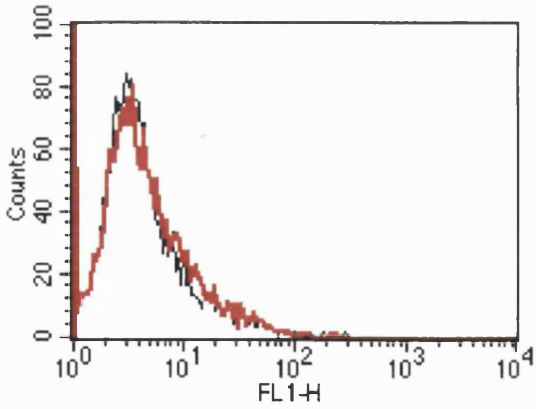
b

T47D.
Geometric Mean
PI : 6.8
CT-1 : 29.14



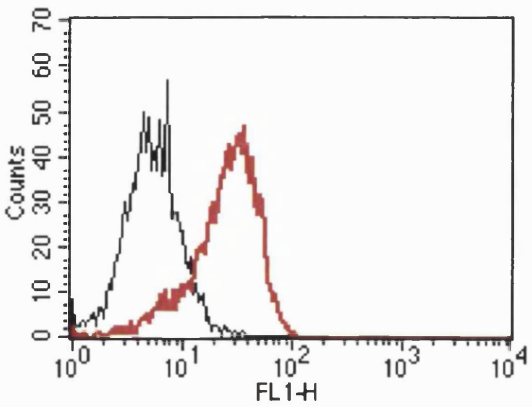
c

HTB96.
Geometric mean
PI : 4.2
CT-1 : 4



d

ZR75.
Geometric Mean
PI : 4.8
CT-1 : 22.5



e

HT1080.
Geometric Mean
PI : 2.4
CT-1 : 3.1

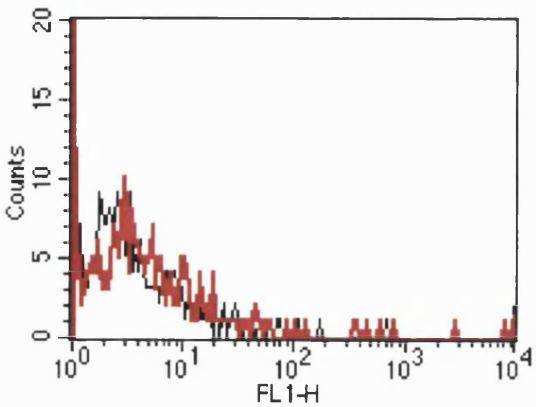


Figure 4.2. Mutations in the MUC1 promoter alter cell-type specific transcription.

Two epithelial lines, T47D (a) and ZR75 (b), and two non epithelial lines, HTB96 (c) and HT1080 (d), were transfected with wild-type and mutated forms of the -1400/+33 *MUC1* 5' region fused to the chloramphenicol acetyl transferase reporter gene (CAT) in the pGCAT-A vector.

The mutations made by Kovarik *et al.*, (111) and used in this experiment were as follows:

Wild-type	-109TAGGGGAGGGGGCGGGGTTTTGTCACCTGTCACCTGCTGC	-80
EMUC mutation	-109TAGGGGAGGGGGCGGGGTTTTGTCTTTAAACACCTGCTGC	-80
-97 GC box mutation	-109TAGGGGAGGGTTCGGGGTTTTGTCACCTGTCACCTGCTGC	-80
double mutation	-109TAGGGGAGGGTTCGGGGTTTTGTCTTTAAACACCTGCTGC	-80

CAT activities were normalised with respect to the activity of the pGL3 control vector (Promega). The results are expressed as percent activity from the wild-type 1.4 kb promoter. Error bars show the standard deviation about the mean of three separate experiments.

Figure 4.2

MUC1 expressing adenocarcinoma cell lines

MUC1 non-expressing cells, (osteosarcoma and fibrosarcoma, respectively).

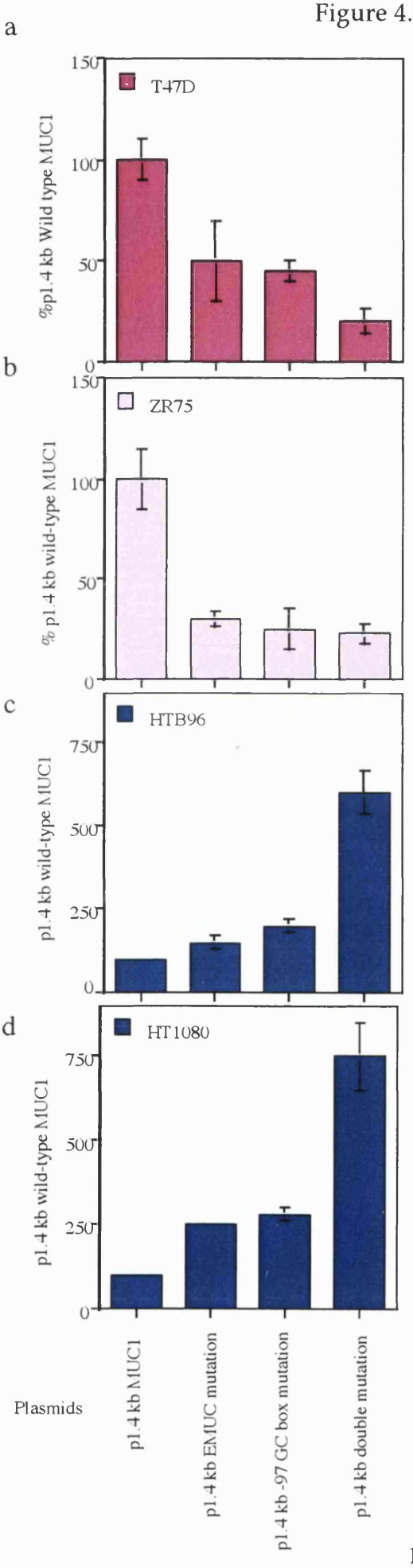


Figure 4.3. MUC1 sequences -119/-62 can direct a more epithelial specific pattern of expression to the constitutively active thymidine kinase promoter.

T47D, ZR75, HTB96 and HT1080 were transfected with engineered forms of the pBLCAT5 vector which contains the 200 bp thymidine kinase promoter driving the CAT gene (the plasmid map of pBLCAT5 can be found in the appendix A3). Constructs used were as follows and the sequences cloned into the vector are shown below:

a) the pBLCAT5 thymidine kinase promoter construct (pTK).
b) pTK with a single Sp1 consensus site cloned into the Kpn I site of the enhancer region.
c) pTK with an oligonucleotide corresponding to MUC1 sequences -119/-62 cloned into the Sal I and Xba I restriction enzyme sites 5' and adjacent to the TK promoter.
d) pTK with both the single Sp1 site in the enhancer region and the MUC1 sequences -119/-62 adjacent to the TK promoter.

Sequences (excluding the restriction enzyme sites).

Sp1 consensus ATTCGATCGGGGCGGGGCGAG

MUC1 -119/-62

AGGTAGGAGGTAGGGGAGGGGGCGGGGTTTTGTCACCTGTCACCTGCTGCCTGTGCCT

CAT activities were normalised with respect to the activity of the pGL3 control vector (Promega). The results are expressed as percent activity of the TK promoter in pBLCAT5 shown in (a). Error bars show the standard deviation about the mean of three separate experiments.

Figure 4.3

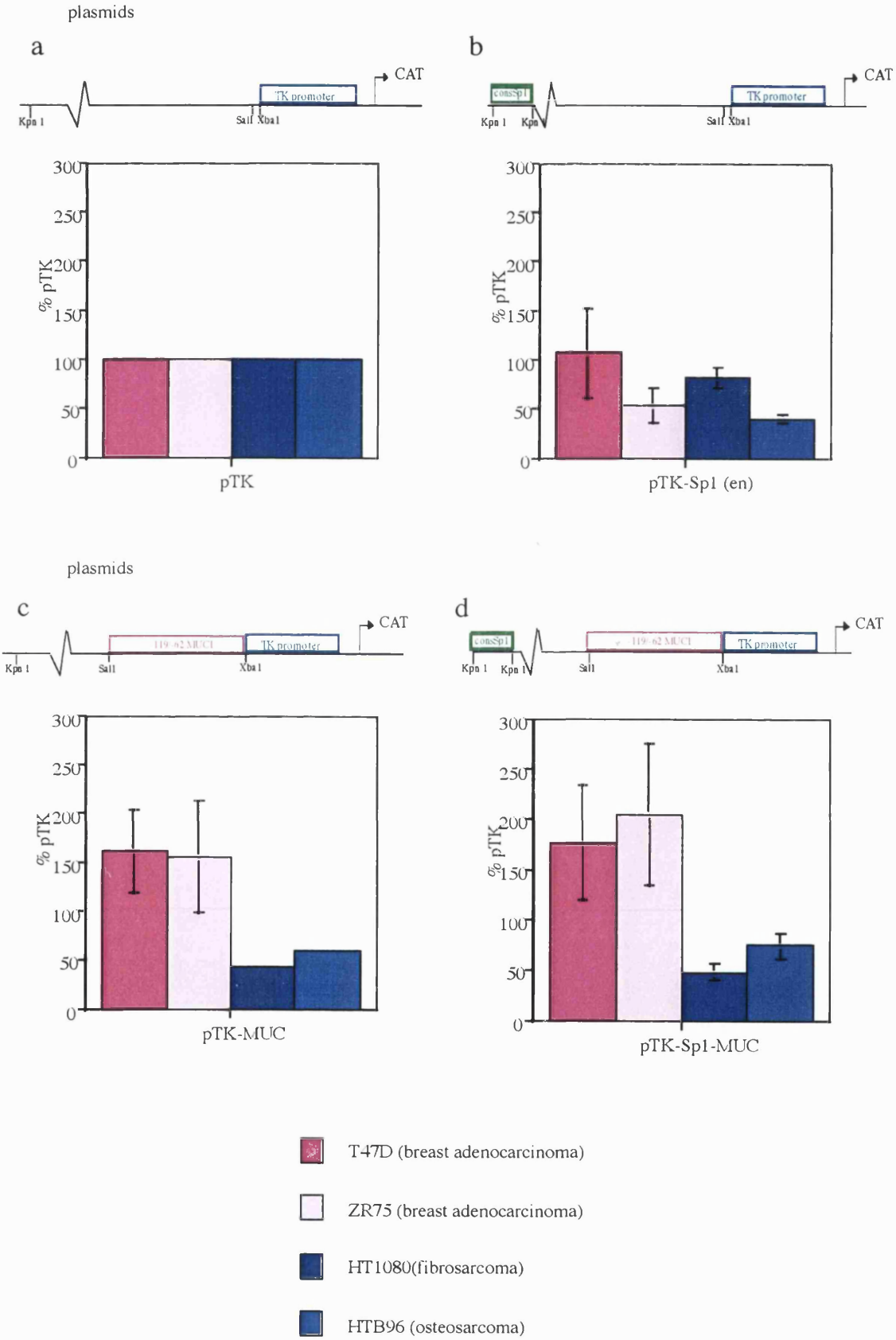

plasmids
a
CAT
TK promoter
Kpn 1
Sall Xba1
300
250
200
150
100
50
0
% pTK
pTK
b
CAT
consSp1
TK promoter
Kpn 1 Kpn
Sall Xba1
300
250
200
150
100
50
0
% pTK
pTK-Sp1 (en)
plasmids
c
CAT
TK promoter
Kpn 1
Sall
Xba1
300
250
200
150
100
50
0
% pTK
pTK-MUC
d
CAT
consSp1
TK promoter
Kpn 1 Kpn
Sall
Xba1
300
250
200
150
100
50
0
% pTK
pTK-Sp1-MUC
T47D (breast adenocarcinoma)
ZR75 (breast adenocarcinoma)
HT1080(fibrosarcoma)
HTB96 (osteosarcoma)

Figure 4.4. Mutations in the -97 GC box and EMUC sequence eliminate the ability of the -119/-62 sequences to direct an epithelial pattern of expression.

Again T47D, ZR75, HTB96 and HT1080 cells were transfected with engineered forms of the pBLCAT5-TK vector. A mutated -119/-62 MUC1 sequence was included between the Sal I and Xba I site both with and without a consensus Sp1 site in the enhancer position. The mutations made in the MUC1 sequence are the same as those included in the doubly mutated -1400/+33 bp promoter in figure 4.2, and shown below.

MUC1 -119/-62

AGGTAGGAGGTAGGGGAGGGGCGGGGTTTTGTCACCTGTCACCTGCTGCCTGTGCCT

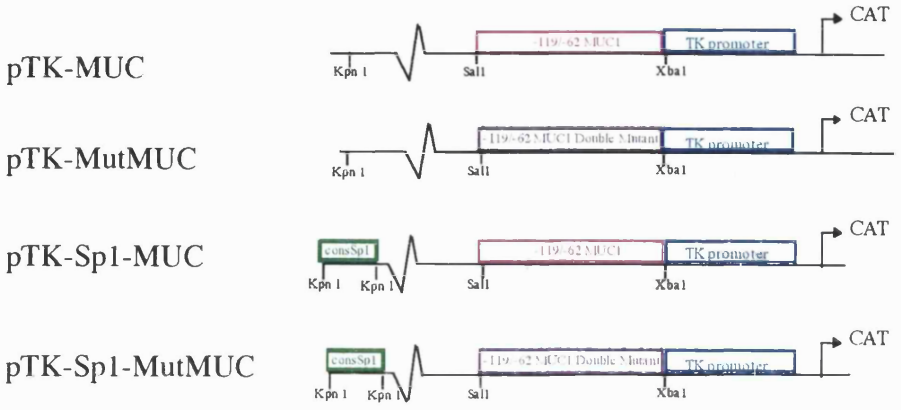
Double mutant

AGGTAGGAGGTAGGGGAGGGTTCGGGGTTTTGTCTTTAAACACCTGCTGCCTGTGCCT

The four constructs used are illustrated in (a), and the results of their transfection into the four cell types in (b).

CAT activities were normalised with respect to the activity of the pGL3 control vector (Promega). The results are expressed as percent activity of the TK promoter in pBLCAT5 (i.e. 100%, not shown). Error bars show the standard deviation about the mean of three separate experiments.

a


Plasmids
pTK-MUC
pTK-MutMUC
pTK-Sp1-MUC
pTK-Sp1-MutMUC
CAT
TK promoter
Kpn 1
Sal1
Xba1
consSp1

b

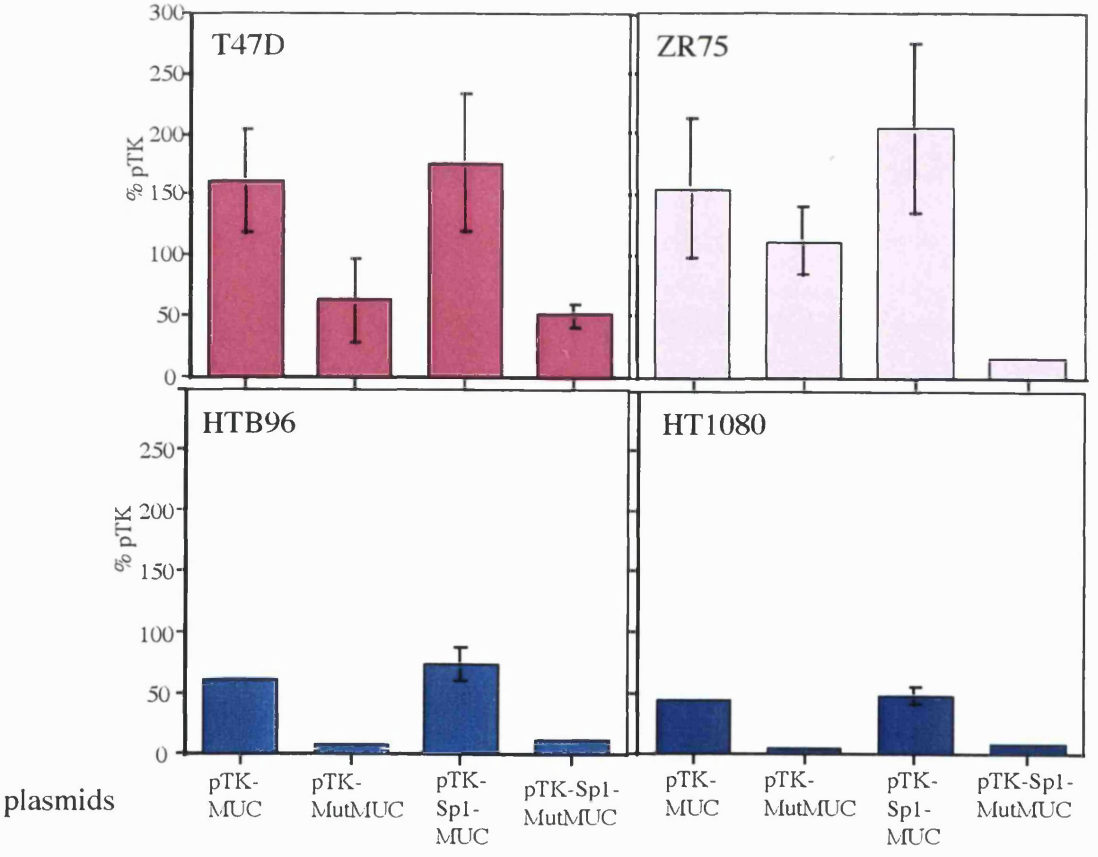

T47D
ZR75
HTB96
HT1080
% pTK
300
250
200
150
100
50
0
pTK-MUC
pTK-MutMUC
pTK-Sp1-MUC
pTK-Sp1-MutMUC
plasmids

Figure 4.5. A consensus Sp1 sequence proximal to the TK promoter can substitute for the MUC1 -119/-62 sequences.

a) A line up the Sp1 consensus sequence with the MUC1 -119/-62 to illustrate the similarities and differences between the two sequences.

b) An Sp1 consensus sequence was cloned into the restriction enzyme sites Sal I and Xba I adjacent to the TK promoter both with and without an Sp1 consensus site at the enhancer position. The activity of these constructs was then compared to that of the constructs bearing the -119/-62 sequences of MUC1 in the same position by transfection of the two epithelial cell lines T47D and ZR75 and two non-epithelial cells HT1080 and HTB96. An illustration of each construct used is shown above the results for that construct. The results are expressed as percent activity of the TK promoter in pBLCAT5 (i.e. 100%, not shown). Error bars show the standard deviation about the mean of three separate experiments. CAT activities were normalised with respect to the activity of the pGL3 control vector (Promega).

Figure 4.5

a

MUC1 -119/-62 AGGTAGGAGGTAGGGGAGGGGGCGGGGTTTTGTCACCTGTCACCTGCTGCCTGTGCCT
Sp1 consensus ATTCGATCGGGGCGGGGCGAG

b

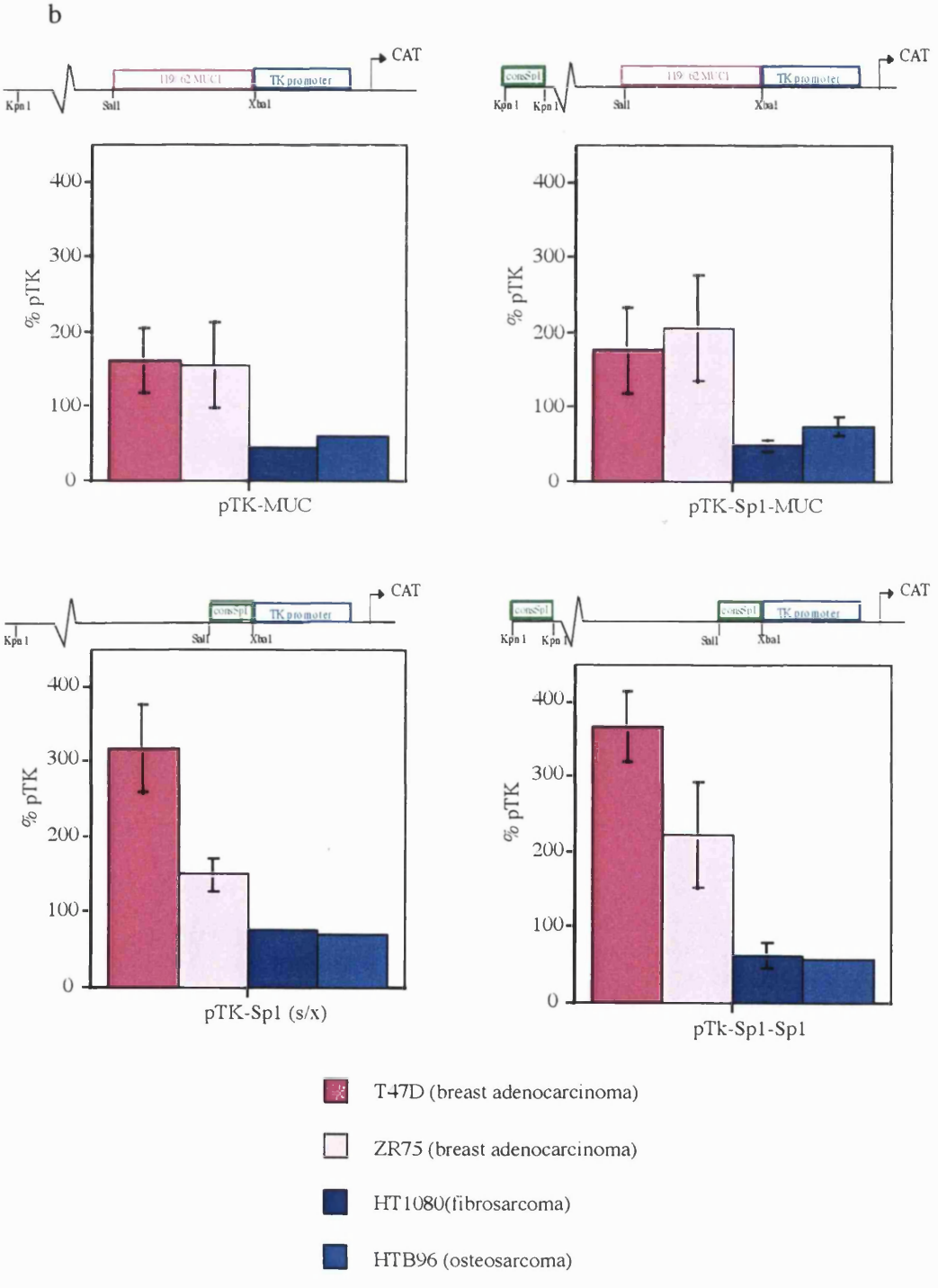
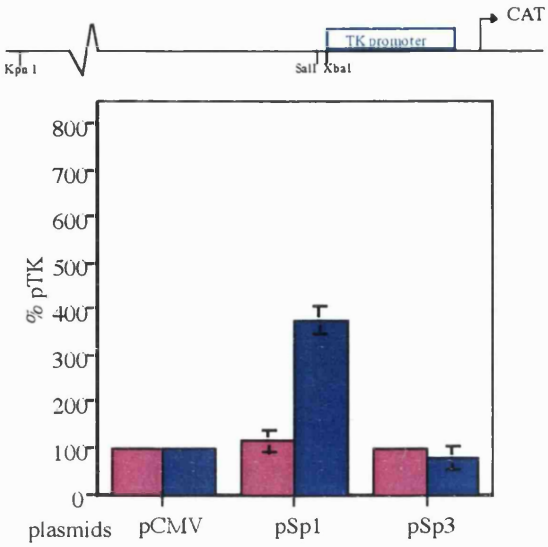


Figure 4.6. Increased expression of Sp1 can activate expression from the TK promoter constructs in non MUC1 expressing cells.

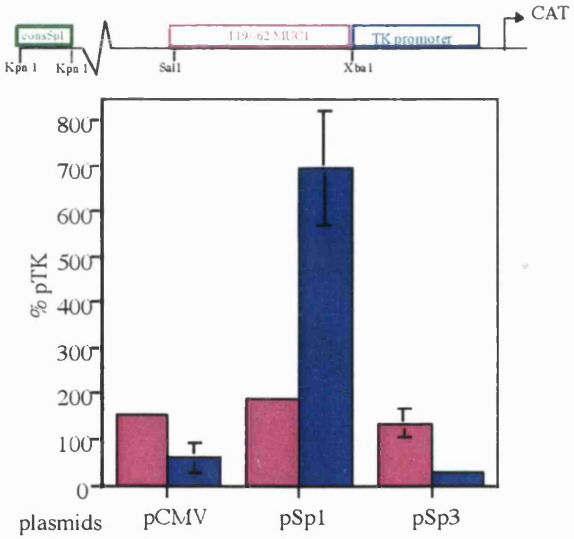
Sp1 (pSp1) and Sp3 (pSp3) expression plasmids, or the empty expression plasmid without a cDNA for either protein (here named pCMV because of the promoter used) were co-transfected with pBLCAT5-TK promoter based plasmids already described into T47D and HTB96. The results are expressed as percent activity of the TK promoter in pBLCAT5 cotransfected with pCMV. Error bars show the standard deviation about the mean of three separate experiments.
CAT activities were not normalised with respect to the activity of the pGL3 control vector since the SV40 sequences in this vector are sensitive to Sp1 and Sp3.

Expression of Sp1 and Sp3 from the expression vectors used in these experiments were checked by immunohistochemistry using specific anti-Sp1 and anti-Sp3 antibodies, followed by secondary swine anti-rabbit FITC conjugated antibody mounted in antifade and examined by fluorescence microscopy. Transfected COS cells showed brighter nuclei compared to un-transfected cells (not shown). The relative amount of each protein was not examined

a


CAT
TK promoter
Kpn 1
Sal1
Xba1
% pTK
800
700
600
500
400
300
200
100
0
plasmids
pCMV
pSp1
pSp3

b


CAT
consSp1
TK promoter
Kpn 1
Kpn 1
Sal1
Xba1
% pTK
800
700
600
500
400
300
200
100
0
plasmids
pCMV
pSp1
pSp3

c

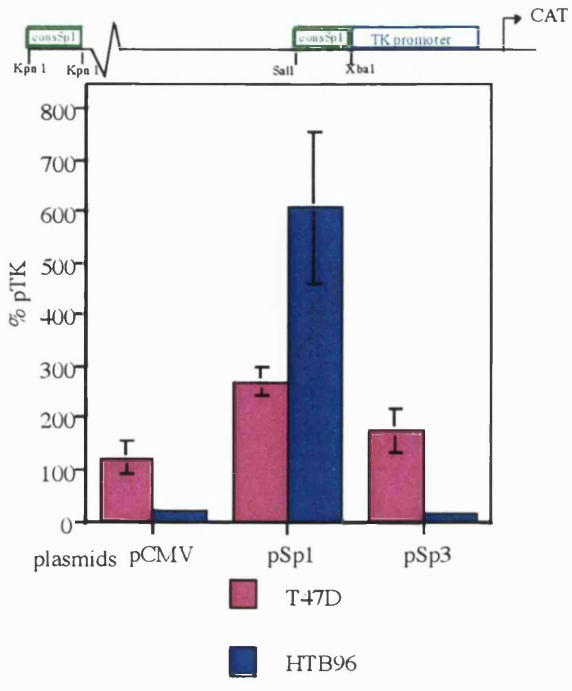

CAT
consSp1
consSp1
TK promoter
Kpn 1
Kpn 1
Sal1
Xba1
% pTK
800
700
600
500
400
300
200
100
0
plasmids
pCMV
pSp1
pSp3
T47D
HTB96

Figure 4.7. Sp1 and Sp3 expression can modulate transcription levels from the 1.4 kb MUC1 promoter.

Co-transfection of the -1400/+33 MUC1 promoter region fused to CAT with Sp1 or Sp3 expression vectors or with the empty expression vector, pCMV, into T47D and HT1080. The results are expressed as percent activity of the -1400/+33 bp MUC1-CAT construct cotransfected with pCMV. Error bars show the standard deviation about the mean of three separate experiments.

CAT activities were not normalised with respect to the activity of the pGL3 control vector since the SV40 sequences in this vector are sensitive to Sp1 and Sp3.

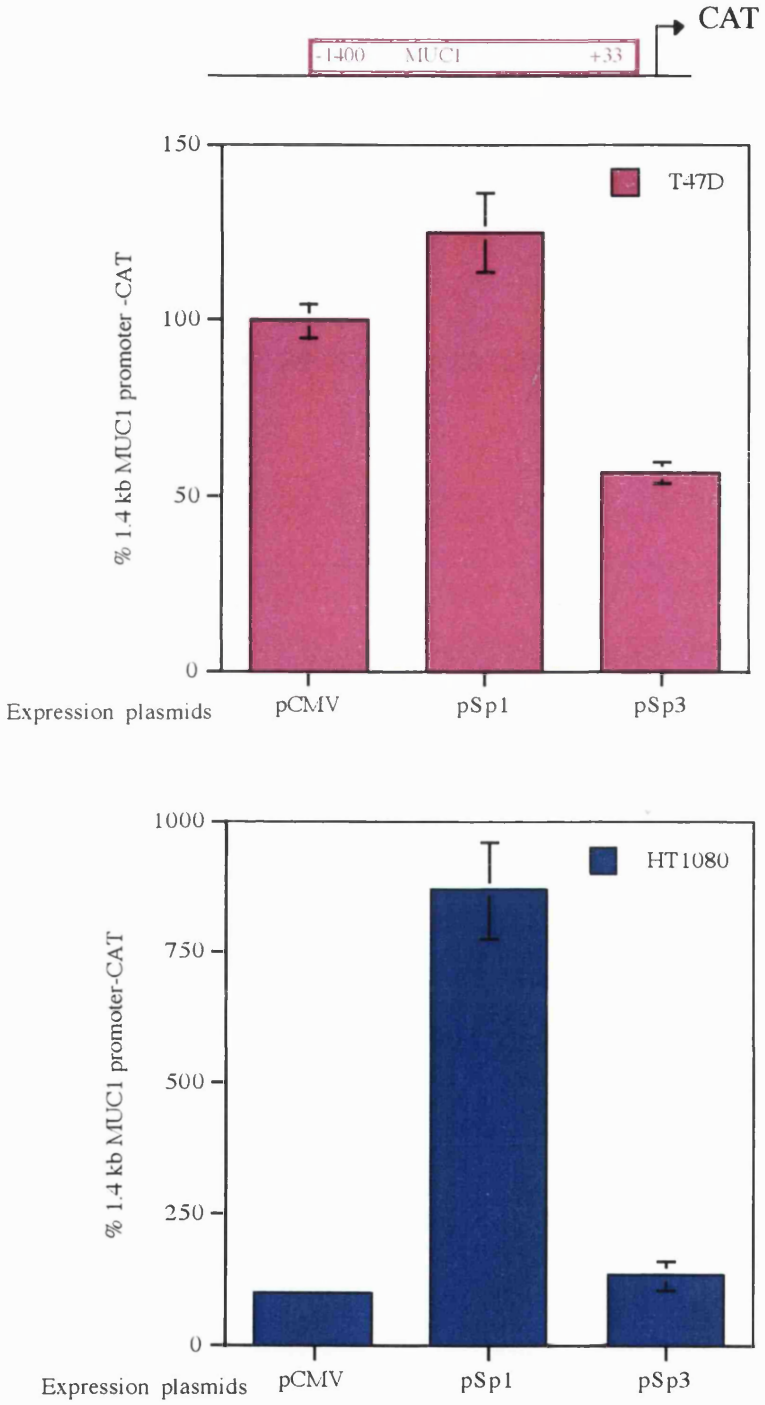
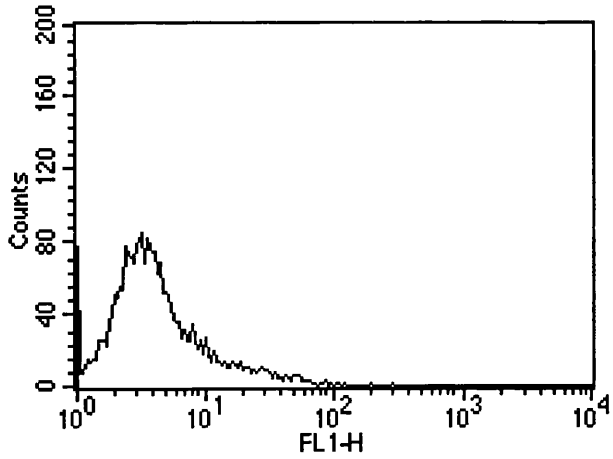

CAT
-1400
MUC1
+33
150
100
50
0
% 1.4 kb MUC1 promoter -CAT
T47D
Expression plasmids
pCMV
pSp1
pSp3
1000
750
500
250
0
% 1.4 kb MUC1 promoter-CAT
HT1080
Expression plasmids
pCMV
pSp1
pSp3

Figure 4.8. Expression of Sp1 can activate expression from the endogenous gene in cells that do not express MUC1.

The non MUC1 expressing cell line HTB96 was transfected with FuGene™ in order to increase the transfection efficiency together with no DNA (a), the MUC1 mini-gene and the cDNA-less expression vector pCMV (b), the MUC1 mini-gene and the expression vector for Sp1 (c), or the expression vector alone (d). After 48 hours the cells were harvested and subjected to the . protocol for internal antigen staining. Cells were incubated with pre-immune sera (not shown) or CT-1 sera (a-d) followed by FITC-conjugated swine anti-rabbit immunoglobulin (see Materials and Methods for details) and subjected to flow cytometry by FACSCalibur (Becton-Dickinson). Cells were gated by forward and side scatter to exclude dead cells and debris. The X-axis shows cell number, the y-axis log relative flourescence intensity. The geometric mean intensity of each population is indicated. The geometric mean intensity of cells incubated with pre-immune sera was equal to that of the cells transfected with no DNA in (a).

Figure 4.8

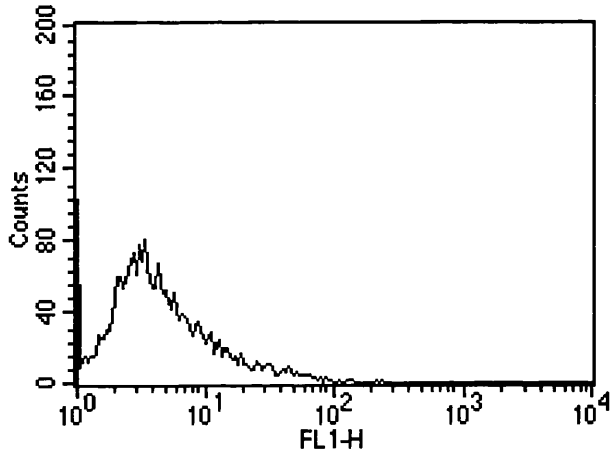
a



No DNA.

Geometric mean.
3.9

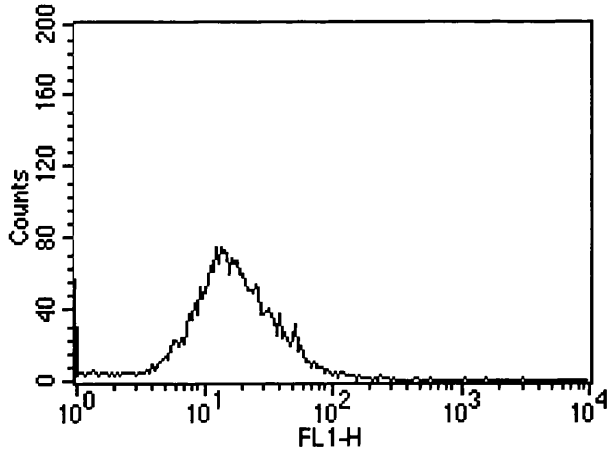
b



MUC1 mini-gene + pCMV

Geometric mean.
4.3

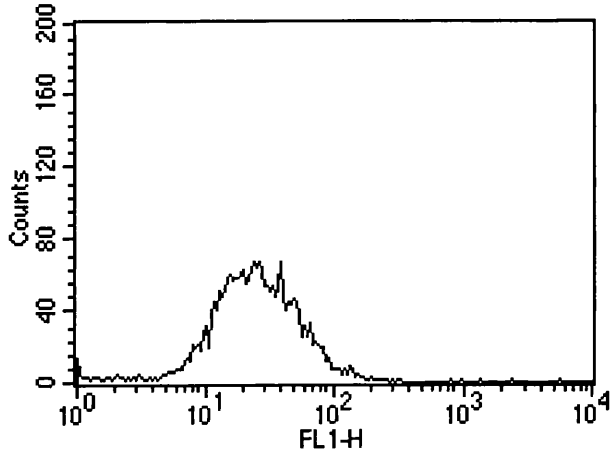
c



MUC1 mini-gene + pSp1

Geometric mean.
16.

d



pSp1 alone.

Geometric mean.
23

Chapter 5: Analysis of factors forming complexes with functional sites of the MUC1 promoter.

5.1 Aims.

Sequences -119/-62 were capable of directing an epithelial specific expression pattern onto the constitutively active TK promoter (see figure 4.3). Therefore characterisation of the factors in nuclear extracts that form complexes with these sequences was undertaken. The aim of the work presented in this chapter was to establish the biochemical features of these factors so that identification and/or cloning of the factors might be possible.

5.1.1 Formation of complexes between *MUC1* sequences and nuclear extracts.

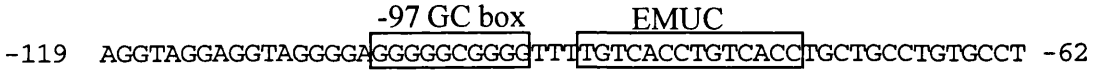


Figure 1.7 shows that there are several sites within the -119/-62 sequence that have high homology to DNA binding sites for many factors including Sp1, E-box factors, PPAR/RXR heterodimers and a potential API site at the -62 boundary. Oligonucleotides corresponding to each of the binding sites were tested in band-shift assays, in which DNA in a complex with protein is separated from free DNA. Although oligonucleotides corresponding to the *MUC1* API site formed a complex in band-shift assays, the complex was not specific, and complexes that formed with an API consensus oligonucleotide could not be competed by the *MUC1* API site (not shown). Only oligonucleotide sequences with homology to an Sp1 site at the -97 GC box and the sequence with E-box and PPAR/RXR homology, known as the EMUC sequence formed consistent complexes with nuclear extracts (sequence shown above). These complexes were examined in more detail and described below.

5.2 GC box results.

5.2.1 The relation of complexes formed with the -97 GC box to the Sp1 family.

Previous work had established that three consistent complexes were formed with an oligonucleotide corresponding to the -97 GC box of *MUC1* (112). The complexes formed with the -97 GC box were specific since competition with excess unlabelled -97 GC box prevented detection of the complex (figure 5.1, lane 2). Excess Sp1 consensus oligonucleotide competed all but one of the complexes formed (lanes 3,4 & 5), the uncompeted complex was named SpA since it also bound poorly to a mutant -97 GC box in which sequences 'AG' 5' of the box were changed to 'CT' (lanes 6,7 & 8).
The possibility that C1 and C3 are members of the Sp1-family, implied from the observation that they could be competed by an Sp1 consensus sequence, was tested using polyclonal antibodies specific to members of the Sp1 family. Incubation of the DNA: protein complex with anti Sp1 antibody resulted in a reduction in the motility of the C1 complex (figure 5.2, lanes 4 &5). This interaction was specific since addition of the peptide to which the antibody was raised blocked the interaction (lane 6). C2 (SpA) is not affected by the addition of the antibody or by the peptide (lanes 7-9). Similarly a polyclonal antibody raised to an Sp3 peptide was able to further retard the motility of complex C3, (figure 5.3, lane 2). This dramatic shift could be inhibited by the addition of peptide to which the antibody was raised (lane 3), demonstrating that the protein : antibody interaction is specific. Again C2 could not be shifted by antibodies to Sp3 peptide (lanes 6-9), nor by polyclonal antiserum raised to the full length Sp3 (lane 10). Antibodies specific to factors Sp2 and Sp4 had no effect on complexes formed with the -97 GC box (not shown). These results suggest that C1 and C3 are closely related to, or are, Sp1 and Sp3.

5.2.2 The nucleotide requirements of complexes formed with the -97 GC box.

That complex C2 (SpA) was not competed either by Sp1 consensus nor by the AG to CT mutant -97 GC box contrasts with the complexes C1 and C3 (see figure 5.1) showing that it has different DNA binding requirements. These were assessed further by

testing the ability of oligonucleotides containing point mutations across the -97 GC box region to compete factors binding to the wild type -97 GC box sequence. This experiment was done to gain information about the binding specificities of each of the factors and to identify mutants that bound C2 and not C1 or C3. Such an oligonucleotide would allow examination of C2 without the competition of C1 and C3. Figure 5.4 shows that mutation of any of the 'G's in the GC box 'GGGCG' which failed to compete C1 also failed to compete C3, confirming the view that these complexes are closely related. Mutants of the first three 'G's in this central GC box similarly failed to compete C2, hence C1, C3 and C2 have overlapping DNA binding requirements. However the requirements for C2 binding covered a larger area than those of C1 and C3 and mutation of either of the 'G's 5' of the box (mutants B and C) prevented competition of C2, while C1 and C3 were competed. While less striking, the same was true for the oligonucleotide mutated at the second 'G' 3' of the central 'C' (mutant I). Only one mutation, that of the first G after the central 'C', i.e. mutant H, appeared to compete C2 better than C1 or C3. A summary of the effect that each nucleotide mutation made has on the ability to bind the three complexes based on data shown in figure 5.4 and data from Kovarik *et al.*, 1996 (112) is shown below.

Underlined in red are nucleotides that when mutated prevented competition of complexes C1 and C3, in blue are nucleotides that when mutated prevented competition of the complex C2 (SpA), and in green are nucleotides that when mutated did not prevent competition of complex C2 (SpA). The central 'C' is shown in grey to help orientation and the Sp1 consensus, which competes C1 and C3 only is shown for comparison.

AGGTAGGAGGTAGGGGAGGGGCGGGGTTTTGTCACC
AGGTAGGAGGTAGGGGAGGGGGGGGGTTTTGTCACC
AGGTAGGAGGTAGGGGAGGGGGCGGGTTTTGTCACC
ATTCGATCGGGGCGGGGCGAG

Only one mutant, mutant H appeared to compete C2 (SpA) better than C1 and C3. To test whether this mutant might differentiate between C2 and C1 and C3 the mutant

oligonucleotide was labelled and tested for its ability to form complexes C1-C3. Figure 5.5 shows a comparison of complexes formed with the wild type oligonucleotide and complexes formed with the mutant. C2 was formed adequately with the mutant (lanes 9-11) while less of complexes C1 and C3 were formed (contrast lanes 1 and 6). While these data supported the observed behaviour of this mutant in the previous experiment, the ability to form C2 over the ability to form C1 and C3 was not great enough to be of use in further study.

5.2.3 Complex formation between the -97 GC box and separated proteins.

Based on the idea that competition with Sp1 consensus oligonucleotide or the -97 GC box oligonucleotide bearing specific mutations should allow differentiation of C2 (SpA) from C1 and C3, proteins capable of binding to the -97 GC box were detected by south-western blotting. Breast cancer cell line ZR75 nuclear extract was separated by reducing SDS-PAGE, blotted, denaturated, and after slow renaturation incubated with labelled oligonucleotide -97 GC box probe (see Materials and Methods for details). Two areas of immobilised probe were identified, labelled SW1 at 97 kDa and SW2 at 66 kDa (figure 5.6a). The formation of SW1 and SW2 were competed by fold excesses of unlabelled wild-type -97 GC box (lanes 5,6,& 7) suggesting both are specific. SW2 was poorly competed by increasing amounts of mutated oligonucleotide corresponding to the mutant 'C' (lanes 2,3 & 4), which poorly competed complex C2 in band-shifts (figure 5.4). When the mutated oligonucleotide was itself labelled and used as a probe it formed SW2 poorly compared to SW1 (compare lane 1 with lane 8). The intensity of each of the complexes with increasing amounts of each competitor was measured by a phosphor-imager and shown in figure 5.6b.

To test whether south-western complexes correspond to Sp1 or Sp3, strips from the same blot were incubated with antibodies specific to Sp1 and Sp3. The proteins corresponding to complex SW1 had a similar mobility to that of the 106 kDa variant of Sp1 and the 97 kDa variant of Sp3 (figure 5.7). The other variants of Sp1 and Sp3 had mobilities similar to that of weak forming complexes that could be resolved on the

south-western film after a long exposure to autoradiography (not shown). The second complex, SW2 had a faster mobility than any Sp1 or Sp3 variant.

These results suggest that Sp1 and Sp3 can bind to the -97 GC box and that a 66 kDa protein may be a candidate for C2. In this assay the amount of SW2 formed was greater than SW1 (figure 5.6a, lane 1), which suggests that C2 may have been under-represented in band-shifts when using competitors to reveal it. Also since the SW2 complex could be resolved in this manner it is likely that the protein is a single polypeptide. Since the conditions used for south-western analysis are similar to those used for screening expression libraries using labelled DNA it may be feasible to isolate the cDNA corresponding to C2 (SpA) using this approach.

5.2.4 Ion requirement of -97 GC box complexes.

Sp1 and Sp3 require zinc as part of their DNA-binding motif (138). While C2 (SpA) requires the central GGGCG sequence to bind DNA (figure 5.4), there was no reason to suppose that it was also a zinc-finger protein. Indeed, if C2 does not require zinc it would be possible to exclude Sp1 and Sp3 from analysis by excluding zinc. To test the ion requirements for the formation of complexes both band-shift and south-western analyses were carried out in the presence or absence of zinc acetate. Addition of EDTA resulted in the abolition of complexes C1-C3 in band shift assays (figure 5.8a, lanes 2 and 3) these could be restored by the addition of zinc ions (lane 4). Similarly in the presence of Sp1 consensus complex C2 formation was restored by the addition of zinc ions after EDTA treatment (figure 5.8a, lane 8).

In south-western blot analysis the proteins were renatured and then incubated in buffers containing sodium, magnesium or zinc ions. Formation of a complex between the probe and the 66 kDa protein, was greatest in the presence of sodium and zinc ions, indicating that complex formation requires physiological levels of salts in addition to zinc ions (figure 5.8b lane 3 and 6).

5.2.5 Functional analysis of the MUC1 promoter with reduced C2 (SpA) binding ability.

In order to asses the functional importance of C2 (SpA) in transcription the AG to CT mutation 5' of the -97 GC was made in the context of the promoter. In band-shift analysis such a mutant fails to compete complex C2, yet competes C1 and C3 well (figure 5.1). In reporter gene assays the mutant promoter had increased transcriptional activity in all cell types tested, suggesting that the factor acts to repress MUC1 expression in both MUC1-expressing and non-expressing cells. The largest fold increase was seen in MCF-7 cells, which within the group of epithelial cell lines tested expresses the least MUC1 (figure 4.1). These observations suggested that reduced C2 binding resulted in a release of transcriptional repression that may specifically affect the level of MUC1 expression within epithelial cells.

Table 3. Effect of the AG to CT mutation on transcription from the MUC1 promoter.

Cellline	Relative MUC1 expression	Fold increase in transcription from the mutant promoter over wild-type promoter.
T47D	8	2
ZR75	2.4	2.6
MCF-7	1	6.6
HTB96	0	2
HT1080	0	1.5

(This table is taken from data in Kovarik *et al.*, (1996) (112)).

5.2.6 Attempts to identify C2 (SpA).

Initially transcription factors with similar DNA binding properties to C2 were assessed by testing the ability of specific antibodies to retard complex C2. Factors WT-1 and HTLM have similar nucleotide requirements to C2 (139, 140, 141) but introduction of antibodies to either factor had no effect on the motility of C2 (not shown).

At this point an attempt was made to clone the C2 factor using an oligomerised wild type -97 GC box oligonucleotide to probe a ZR75 cDNA library in lambda-zap™ . This library consists of cDNA sequences cloned into the *lac Z* gene which can be induced by IPTG. $1x10^6$ phage were plated and allowed to grow in a bacterial lawn after which IPTG was added and the cDNAs expressed. After blotting, denaturation and renaturation the proteins were incubated with a -95 GC box labelled probe. Ideally this approach should identify Sp1-related proteins, Sp1 and Sp3, and the C2 (SpA) protein. Positive reproducible plaques were detected, however they were not dependent on IPTG induction and plates of positive plaques proved to be heterogeneous on PCR amplification of the inserts. Subsequent rescue of the plasmids and sequencing of some of these clones confirmed that they were expressed sequences, three had 98-100% identity with known genes (Radixin, amyloid precursor and GA17) two had high homology to ESTs at the protein and nucleotide levels and one had no identified homology. Importantly south-western analysis of bacterial lysate expressing the selected clones showed no DNA binding activity (with or without IPTG). Similarly *in vitro* transcribed translated products from rescued plasmids were not able to retard the *MUC1* -97 GC box oligonucleotide in a band-shift assays. Taken as a whole these observations suggest that the candidate clones were false positives. Since attempts were unsuccessful the approach was not pursued further.

5.2.7 Assessment of the levels of Sp1, Sp3 and C2 (SpA) in epithelial and non-epithelial cells.

Given the potential importance of Sp1, Sp3 and C2 (SpA) in the transcriptional control of MUC1 it was important to examine the relative amounts of each protein in epithelial and non-epithelial cells. Immunoblots of nuclear extracts from various cell lines in figure 5.9a show that there was little difference in Sp1 amounts between epithelial cells, T47D, ZR75 and MCF-7 and non-epithelial cells, HTB96 and HT1080. The non-epithelial cell lines appeared to have more Sp3 when compared to the epithelial cells (figure 5.9b), which may be a significant factor in regulation of the gene. It was important to establish whether this was also true for the binding activity of complexes

formed with the -97 GC box. Figure 5.10 shows a representative band-shift of complexes C1-C3 formed with nuclear extracts of MUC1-expressing and non-expressing cells. There are variations in C1, C2 and C3, however no consistent difference in C1 or C3 levels was observed between band-shift experiments. The levels of C2 detected when Sp1 consensus or mutant -97 GC box sequence were used as competitors were consistent in that MCF-7 cells revealed highest levels of the complex. This is consistent with the data from section 5.2.5 and with the hypothesis that SpA is involved in repressing expression of MUC1 in epithelial cells, most dramatically in MCF-7.

5.3 EMUC results.

5.3.1 Competition of factors forming complexes with the EMUC sequence.

Previous *in vitro* experiments had shown that three complexes were formed with an oligonucleotide encompassing the EMUC sequence. (The complexes formed with the EMUC sequence will be called E1-E3 in this thesis). The identity and characteristics of these complexes was unknown (111). A reminder of the EMUC sequence is shown below.

EMUC

MUC1 -95 GGGTTTTGTCACCTGTCACCTGCTGCCTGT -66

The EMUC sequence contains a direct repeat and the sequence has similarities to several known transcription factor binding elements including; RXR:DR1 type sequence, which binds heterodimers of nuclear receptors RXR, RAR and PPAR and E-box sequences with the consensus 'CANNTG' which binds proteins of the helix-loop-helix family. To determine whether any of these factors bind to the EMUC sequence the ability of a consensus RXR DR1, two types of E-box sequence and the E-cadherin E-Pal sequence, which also has similarities to EMUC, were assessed for the ability to compete complexes formed with the EMUC sequence (figure 5.11). A consensus RXR:DR1 sequence failed to compete EMUC complexes (lane 4), the E-Pal sequences slightly competed complexes E1-3 (lane 6), whereas E-box type sequences, USF and myc:max consensus (not shown), competed E1 only (lane 5). Furthermore antibody to USF was able to retard complex E1 (not shown). These results suggest that E-box helix-loop-helix

factors are able to bind the EMUC sequence and that one of these is closely related to, or is, USF.

5.3.2 The nucleotide requirements of complexes formed with the EMUC sequence.

The identity and characteristics of complexes E2 and E3 remained unknown. To gain information about their DNA binding requirements point mutations were made across the first repeat of the EMUC sequence. This region was chosen because previous work has shown that a 6 bp mutation across the first repeat inhibited the formation of all three EMUC complexes (111). Competition of EMUC complexes with mutant oligonucleotides revealed that any one base pair change did not prevent the competition of E1 (figure 5.12). Complexes E2 and E3 were poorly competed by oligonucleotides in which the 'TGTC' sequence at the centre of the direct repeats was mutated, the most effective mutation being of the second 'T', mutant Ef. This nucleotide is outside the two E-box 'CANNTG' sequences confirming that these factors are unrelated to E-box binding proteins.

A summary of the competing sequences for EMUC complexes E1-E3 can be found in the table overleaf.

Table 4. Summary of oligonucleotide competition of EMUC-bound complexes.

Sequencename	Sequence (similarities to the EMUC sequence are in magenta, mutated nucleotides are lower case and in blue)	Competes complex? E1	E2	E3
EMUC (WT)	TTTGTCACCTGTCACCTGCTCG	Y	Y	Y
E-Pal	TCCGGCTGCCACCTGCAGG	Y/N	Y/N	Y/N
USF	TAGGCCACGTGACCGGGTG	Y	N	N
USF	TAGGCCACGTGACCGGGTG	Y	N	N
myc:max	CAGACCACGTGGTCTGCTT	Y	N	N
myc:max	CAGACCACGTGGTCTGCTT	Y	N	N
RXR: DR1	TCTGACCTCTGACCTCAAGCT	N	N	N
EMUC-Ef	TTTGTCACCTGgCACCTGCTCG	Y	N	N

5.3.3 Complex formation between the EMUC sequence and separated proteins.

In order to gain information about the molecular weights of complexes forming with the EMUC sequence south-western blotting was carried out using the labelled EMUC sequence in the presence and absence of wild-type EMUC competitor and a six base-pair mutant, which has been shown previously to compete EMUC factors poorly (111). The result is shown in figure 5.13, unlike the south-western analysis using the -97 GC box DNA it was only possible to resolve immobilised EMUC DNA if the reducing agent was omitted from the renaturation and binding buffer, suggesting that di-sulphide bonds are required to form a DNA binding complex. Although the complex can be competed by EMUC wild-type oligonucleotide it can also be competed by the 6 bp mutant. This suggests that the complex formed in the south-western study was not a specific EMUC factor. USF which has a molecular weight of 43 kDa (USF-2) or 44 kDa (USF-1) binds DNA as homo- or heterodimer (142), and would be unlikely to be detected by south-western blotting. The lack of any specific binding may suggest that the E2 and E3 factors also bind DNA as oligomers.

5.4 Summary.

5.4.1 Sp1 family members.

The evidence that -95 GC box complexes C1 and C3 are closely related to the Sp1-family of zinc-finger transcription factors is strong;

- C1 and C3 could be competed by Sp1 consensus oligonucleotide,
- the use of specific antibodies to Sp1 and Sp3 could retard the motility of C1 and C3 respectively,
- specific nucleotides required for binding C1 and C3 across the GC box were consistent with the reported binding site for Sp1 and Sp3 'GGGcG' (143),
- binding of C1 and C3 to the -97 GC box is dependent on zinc,
- and proteins that bind the -97 GC box in south-western blots had a mobility similar to variants of Sp1 and Sp3 identified by immunoblotting.

5.4.2 The C2 (SpA) complex.

Evidence is also presented that shows complex C2, spuriously named SpA, is not related to the Sp1-family and may be a repressive transcription factor of 66 kDa which is involved in regulation of MUC1 in epithelial cells;

- it did not bind the Sp1 consensus sequence,
- it was not shifted by antibodies to known Sp1-family members,
- it bound the -97 GC box with overlapping but broader nucleotide requirements than Sp1 and Sp3, 'GGGGGNNG',
- south-western analysis revealed a 66 kDa protein with DNA binding properties consistent with C2.
- both the C2 complex in band-shift analysis and the 66 kDa protein in south-westerns required zinc to bind the -97 GC box.
- transient transfection studies in which the AG 5' of the -97 GC box had been mutated to a CT in the promoter, (reducing the ability of SpA but not Sp1 or Sp3 to bind) resulted in an increase in transcription.
- within the epithelial cells tested this increase was inversely proportional to the amount of MUC1 produced. In support of the idea that the C2 factor may be a

modulator of MUC1 expression in epithelial cells C2 was consistently highest in MCF-7 cells, which produce least MUC1.

5.4.3 The EMUC factors.

The factors which complex with the EMUC sequence have been further characterised and complex E1 appears to be related to the helix-loop-helix family of transcription factors;

- it could be competed by oligonucleotides that contained an E-box motif,
- its mobility could be retarded by incubation with an antibody to USF.

The other two complexes remain unidentified, they are unlikely to be helix-loop-helix factors or nuclear receptors RXR, RAR or PPAR, but they may be related to each other;

- point mutations which affected the binding of E2 also affected the binding of E3,
- they could not be competed by E-box containing sequence,
- a mutation outside the two E-box repeats inhibited competition of the factors,
- competition with RXR:DR1 consensus sequence or the E-Pal sequence did not compete E2 and E3 significantly.

5.5 Chapter 5: Figures.

Figure 5.1. Band-shift showing the complexes formed with the wild-type *MUC1* -97 GC box sequence.

2-3 µg of ZR75 nuclear extract was pre-incubated with indicated amounts of unlabelled -97 GC box oligonucleotide called 'Self' (lane 2), Sp1 consensus oligonucleotide (lanes 3-5) or a mutant -119/-84 oligonucleotide in which the AG has been changed to CT (lanes 6-8) before the addition of radio-labelled -97 GC box probe and separation on a 6% native gel. Fold excess refers to fold molar excess over the labelled probe. The sequences of oligonucleotides used in this experiment can be found below. Complexes indicated by an asterisk are inconsistent between experiments, they may represent degradation products of C1 and C3 as they can be competed by self and Sp1 consensus.

Oligonucleotidesused:

-97 GC box (wild type)	-119 AGGTAGGAGGTAGGGGAGGGGGCGGGGTTTTGTCACC -84
-97 GC box (AG→CT mutant)	-119 AGGTAGGAGGTAGGGGCTGGGGCGGGGTTTTGTCACC -84
Sp1 consensus	ATTCGATCGGGGCGGGGCGAG

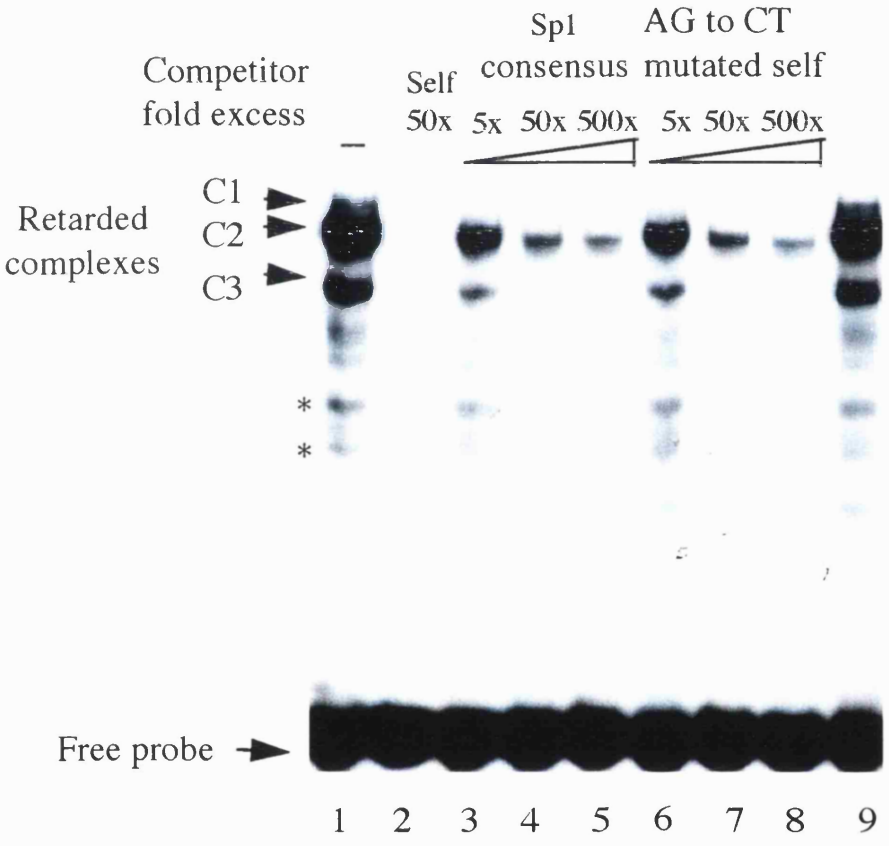

Competitor
fold excess
Self
Sp1
consensus
AG to CT
mutated self
—
50x
5x
50x
500x
5x
50x
500x
Retarded
complexes
C1
C2
C3
*
*
Free probe
1 2 3 4 5 6 7 8 9

Figure 5.2. Band-shift showing that Sp1 antibody further retards complex C1.

2-3 μg of ZR75 nuclear extract was pre-incubated with indicated μg of polyclonal Sp1 antibody (lanes 4-9) and/or 40 μg of Sp1 blocking-peptide (lanes 6 and 9) before the addition of a 200x molar excess over the labelled probe of unlabelled Sp1 consensus oligonucleotide (lanes 3 and 7-9) or unlabelled 'Self' (wild-type -97 GC box oligonucleotide) (lane 2). Following incubation with radiolabelled wild-type -97 GC box oligonucleotide the reactions were separated on a 6% native gel, (see Materials and Methods for details of incubations). Supershift denotes further retardation of a complex.

The sequences of oligonucleotides used can be found in the previous legend.
The Sp1 antibody was raised to a peptide corresponding to human Sp1 amino acids, 436-454 and cross reacts with the 95 kDa and 106 kDa forms.
The blocking peptide corresponds to Sp1 amino acids 436-454 also.

Figure 5.2

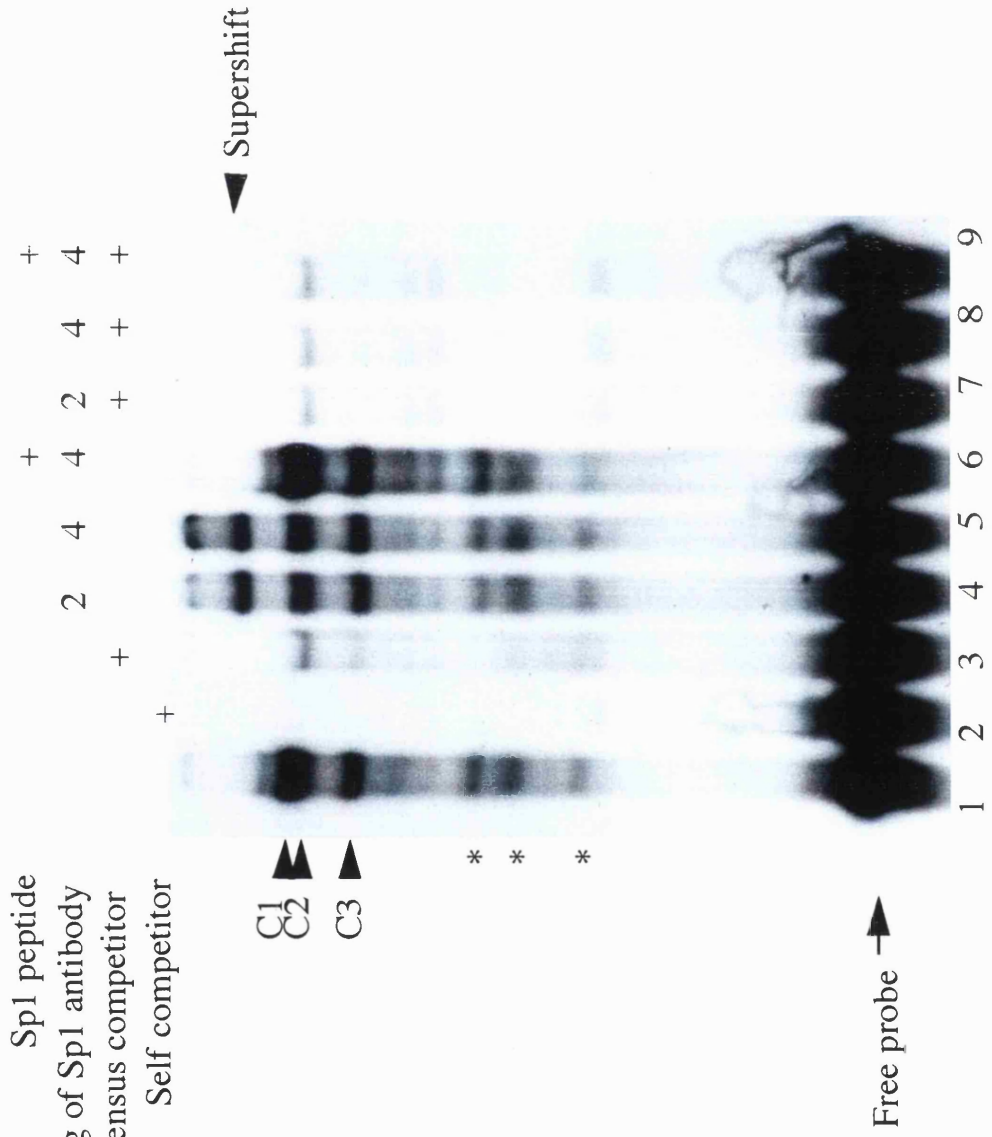

Supershift
Sp1 peptide
ug of Sp1 antibody
Sp1 consensus competitor
Self competitor
C1
C2
C3
Free probe
1 2 3 4 5 6 7 8 9

Figure 5.3. Band-shift showing Sp3 antibody further retards complex C3.

2-3 μg of ZR75 nuclear extract was pre-incubated withindicatedμg of polyclonal Sp3 antibody (lanes 2,3 and 6-10) and/or 40 μg of Sp3 blocking-peptide (lanes 3,4 and 9) before the addition of an indicated molar excess over the labelled probe of unlabelled Sp1 consensus oligonucleotide(lanes 5-10). Following incubation with radiolabelled wild-type -97 GC box oligonucleotide the reactions were separated on a 6% native gel, (see Materials and Methods for details of incubations). Supershift denotes further retardation of a complex.

The sequences of oligonucleotides used can be found in the legend of figure 5.1.
The Sp3 polyclonal antibody used was raised to human Sp3, amino acids 676-695.
The blocking peptide corresponds to the same amino acids.
$_S$: refers to a polyclonal anti-Sp3 antibody raised to the full length bacterially expressed Sp3, a kind gift of G. Suske (136) this antibody also further retards complex C3 (not shown).

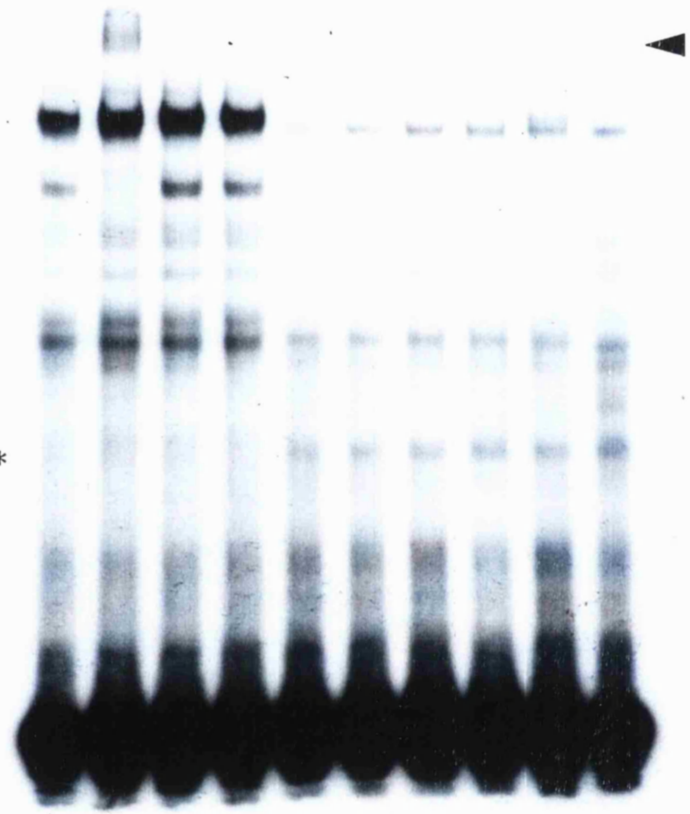

Sp3 peptide + + +
ug of Sp3 antibody 2 2 2 4 6 6 6s
200 fold excess Sp1 consensus competitor + + + + + +
Supershift
Retarded complexes
C1
C2
C3
*
*
Free probe
1 2 3 4 5 6 7 8 9 10

Figure 5.3

Figure 5.4. Band-shift showing the nucleotide binding specificities of complexes C1, C2 and C3 within the -97 GC box.

2-3 µg of ZR75 nuclear extract was pre-incubated with unlabelled mutant -97 GC box oligonucleotides followed by addition of the wild type -97 GC box radiolabelled oligonucleotide before separation on a gel. The code for each mutant is indicated and the exact sequence of the mutant oligonucleotide is shown below (the mutated nucleotide is underlined and in red, the central 'C' is green to help alignment). Across the top of the figure the wild type nucleotide is indicated by a capital, and the mutation made at that position is shown in lower case type. The reactions were run on a 6% native gel for an extended period to allow greater separation of complexes C1 and C2.
Where a specific base pair was required for complex formation the oligonucleotide mutated at that position poorly competed for the factors. The nucleotides required for specific binding to complexes C1-C3 are indicated in the box in the figure. The AG to CT -97 GC box mutant that allowed visualisation of C2 in figure 5.1 and after which the complex was named SpA is underlined.

AGGTAGGAGGTAGGGGAGGGGGCGGGGTTTTGTCAC	**wild type -97 GC box**
AGGTAGGAGGTAGGGGCGGGGGCGGGGTTTTGTCAC	**A**
AGGTAGGAGGTAGGGGATGGGGCGGGGTTTTGTCAC	**B**
AGGTAGGAGGTAGGGGAGTGGGCGGGGTTTTGTCAC	**C**
AGGTAGGAGGTAGGGGAGGTGGCGGGGTTTTGTCAC	**D**
AGGTAGGAGGTAGGGGAGGGTGCGGGGTTTTGTCAC	**E**
AGGTAGGAGGTAGGGGAGGGGTCGGGGTTTTGTCAC	**F**
AGGTAGGAGGTAGGGGAGGGGGAGGGGTTTTGTCAC	**G**
AGGTAGGAGGTAGGGGAGGGGGCTGGGTTTTGTCAC	**H**
AGGTAGGAGGTAGGGGAGGGGGCGTGGTTTTGTCAC	**I**
AGGTAGGAGGTAGGGGAGGGGGCGGTGTTTTGTCAC	**J**
AGGTAGGAGGTAGGGGAGGGGGCGGGTTTTTGTCAC	**K**
AGGTAGGAGGTAGGGGAGGGGGCGGGGCTTTGTCAC	**L**

Figure 5.4

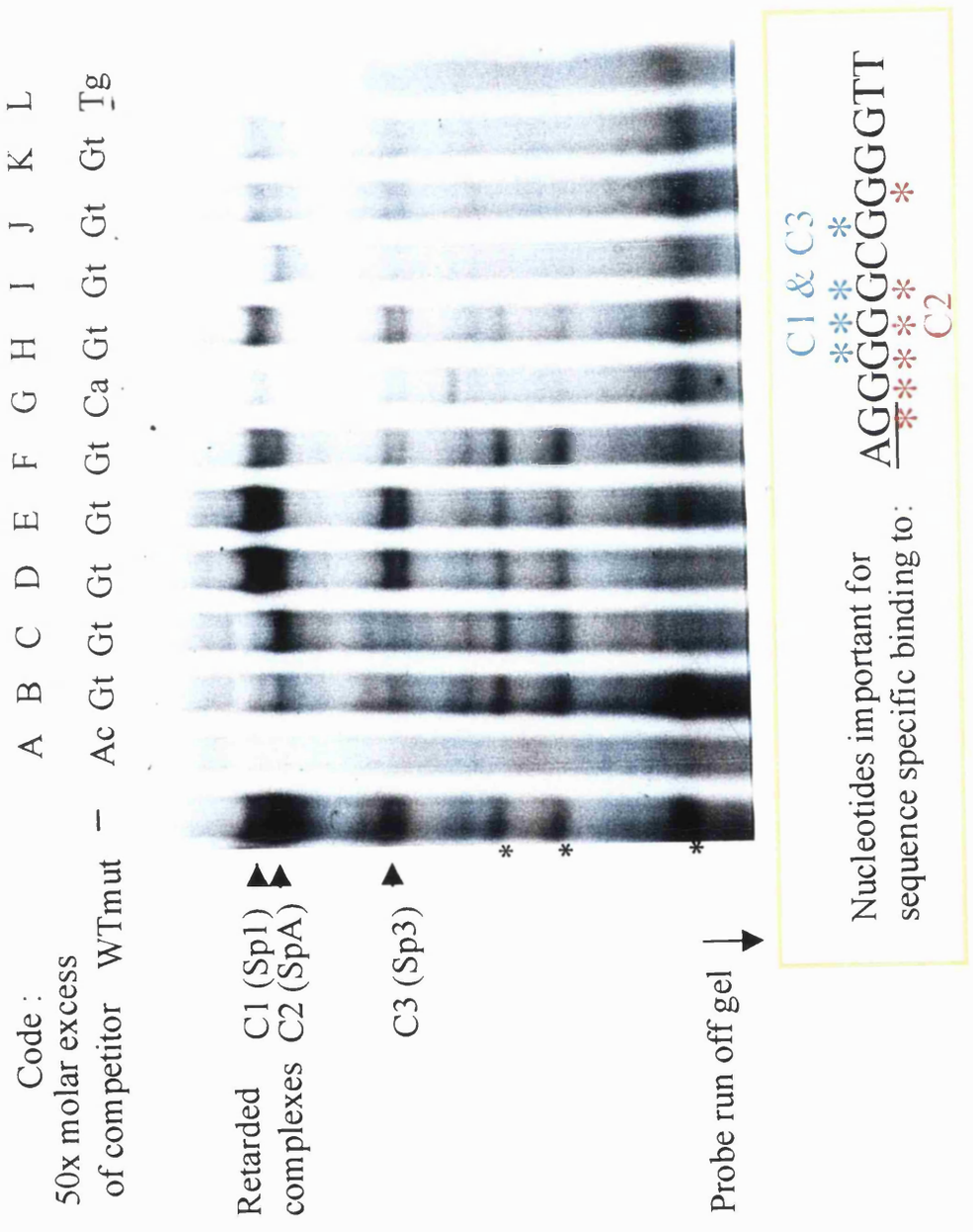

Code :
A B C D E F G H I J K L
50x molar excess
of competitor – Ac Gt Gt Gt Gt Gt Ca Gt Gt Gt Gt Tg
WTmut
Retarded complexes
C1 (Sp1)
C2 (SpA)
C3 (Sp3)
Probe run off gel
C1 & C3
C2
Nucleotides important for sequence specific binding to:
AGGGGGCGGGGTT

Figure 5.5. Band-shift showing that mutant H binds C1 and C3 less well than wild type, yet still binds C2 (SpA).

2-3 μg of ZR75 nuclear extracts were pre-incubated with unlabelled Sp1 consensus oligonucleotide (lanes 4,5 and 9-11), unlabelled mutant H (lanes 3 and 8) or wild type -97 GC box (lanes 2 and 7) before addition of radiolabelled wild type oligonucleotide (a) or radiolabelled mutant H oligonucleotide (b) and then run on a native 6% gel. Again the reactions were run for an extended period to allow greater separation of complexes C1 and C2.

The sequence of mutant 'H' can be found in the legend for figure 5.4.

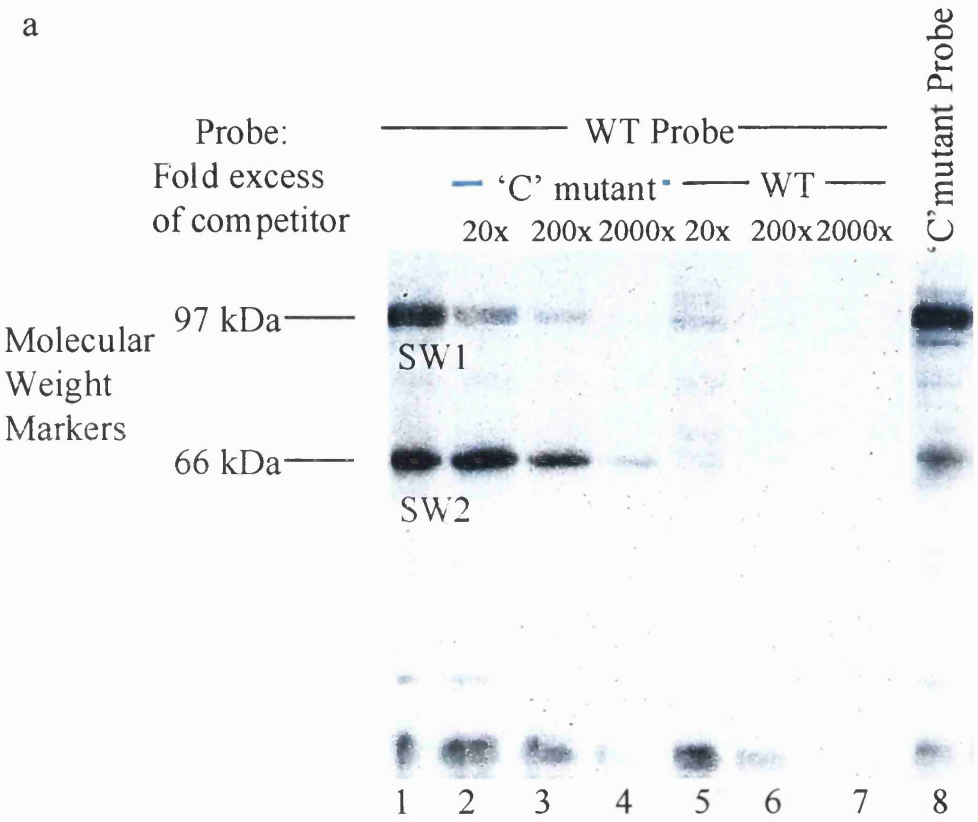
Analysis of Factors
Figure 5.5

Figure 5.6. South-western showing complexes formed between *MUC1* -97 GC box and separated proteins.

a) Nuclear extract from ZR75 cells was separated by reducing SDS-PAGE. The gel was blotted onto a membrane and the proteins denatured, then renatured through a Guanidium hydrochloride gradient. The membrane was then cut into strips and incubated with south-western binding buffer and radiolabelled wild type -97 GC box probe (lane 1 to 7) or the 'C'' mutant oligonucleotide (lane 8, see the legend of figure 5.4 for the sequence of mutant 'C'). Lanes 2 to 7 were also incubated with indicated fold molar excesses of unlabelled wild type -97 GC box or the 'C' mutant as indicated. After washing with south-western binding buffer the strips were re-united before autoradiography.

b) Graphical representation of the radioactivity of complexes SW1 and SW2 formed with the wild-type oligonucleotide with increasing amount of competitors. The relative radiolabel in complexes SW1 and SW2 formed was measured by phosphor-imager analysis and the results presented as relative arbitrary units.

a


Probe:
Fold excess
of competitor
WT Probe
'C' mutant
WT
20x 200x 2000x 20x 200x 2000x
'C' mutant Probe
Molecular
Weight
Markers
97 kDa
66 kDa
SW1
SW2
1 2 3 4 5 6 7 8

b

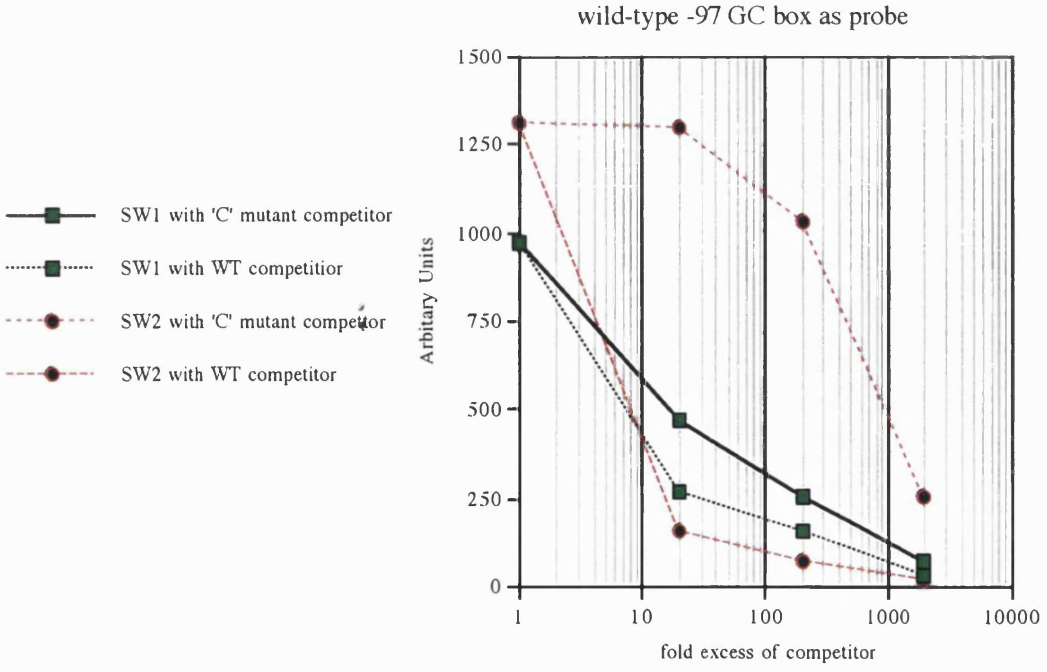

wild-type -97 GC box as probe
SW1 with 'C' mutant competitor
SW1 with WT competitior
SW2 with 'C' mutant competitor
SW2 with WT competitor
Arbitary Units
1500
1250
1000
750
500
250
0
1
10
100
1000
10000
fold excess of competitor

Figure 5.7. Immunoblot and south-western to compare the mobility of proteins Sp1 and Sp3 with proteins binding the -97 GC box in south-western analysis.

As in the previous figure nuclear extract from ZR75 cells was separated by reducing SDS-PAGE. After blotting onto a membrane, followed by denaturation and renaturation of the proteins, strips of membrane were either incubated with south-western binding buffer and radiolabelled -97 GC box probe (lane 2) or subjected to immunoblot conditions with anti Sp1 (lane 1) or anti Sp3 antibody (lane 3). After autoradiography or ECL detection and exposure to film, the X-ray film of each strip was aligned, so that the top of the strip was level.

Figure 5.7

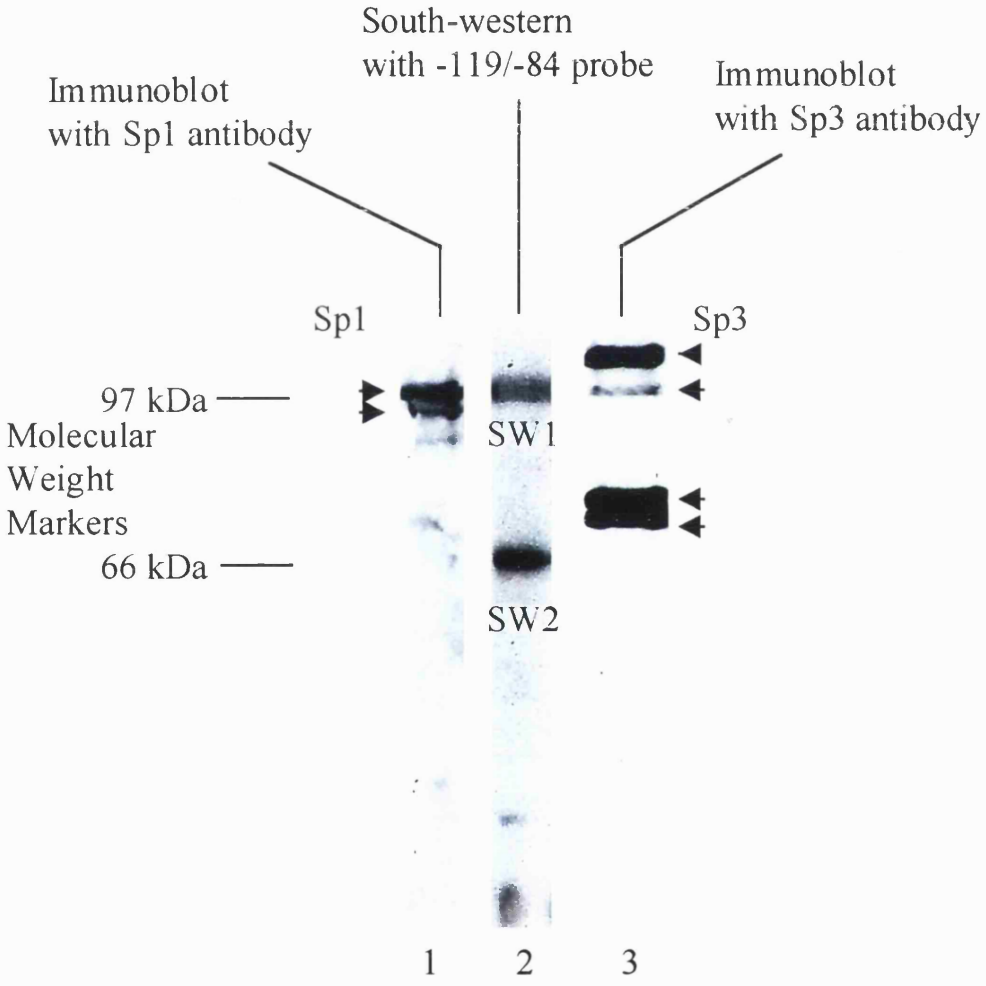
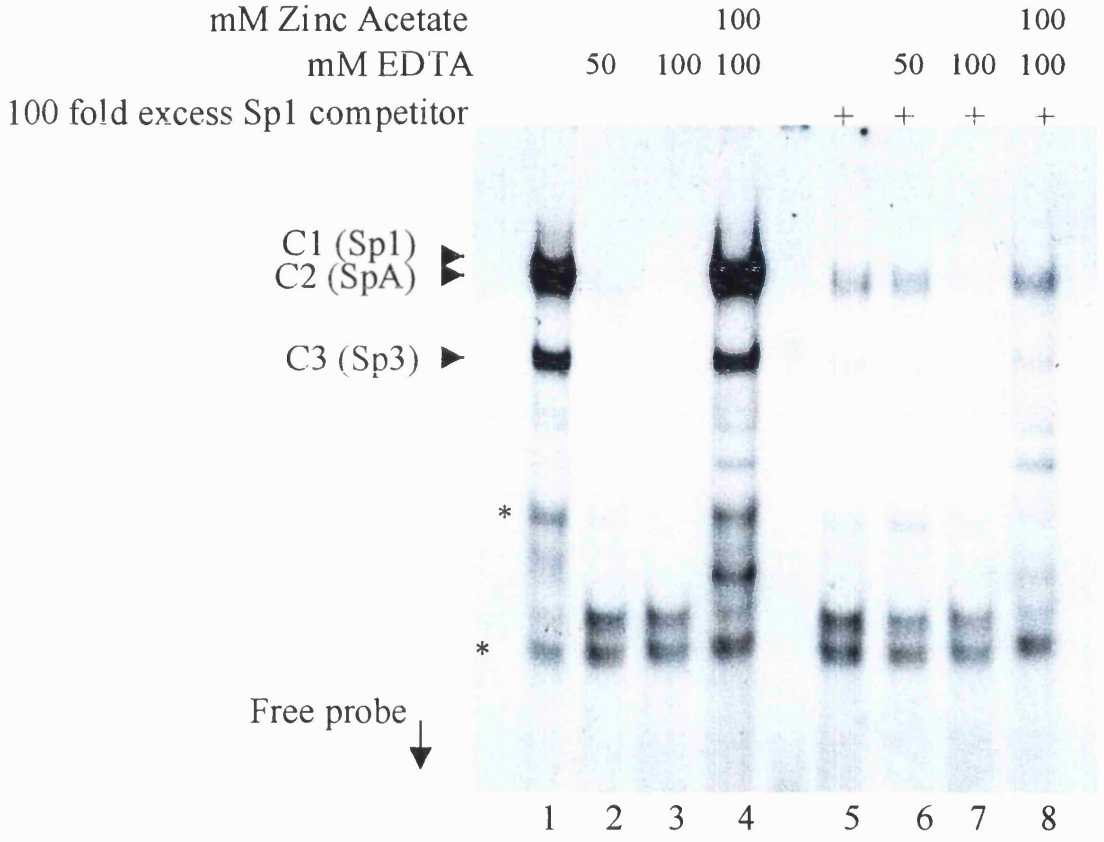

South-western
with -119/-84 probe
Immunoblot
with Sp1 antibody
Immunoblot
with Sp3 antibody
Sp1
Sp3
97 kDa
Molecular
Weight
Markers
66 kDa
SW1
SW2
1
2
3

Figure 5.8. South-western and band-shift assays showing the ion requirements of complexes formed by the -97 GC box.

a) ZR75 nuclear extracts were pre-incubated with EDTA (lanes 2-4 and 6-8), zinc acetate (lanes 4 and 8) and unlabelled Sp1 consensus oligonucleotide (lanes 5-8) before the addition of radiolabelled -97 GC box oligonucleotide and separation on a 6% native gel.

b) ZR75 nuclear extracts were separated by reducing SDS-PAGE and subjected to denaturation, and renaturation through a Guanidium hydrochloride gradient before incubation in south-western binding buffer with labelled -97 GC box oligonucleotide. Where indicated both the renaturation and south-western DNA binding buffer contained 25 mM NaCl,1 mM Magnesium Chloride, and 1 mM Zinc Acetate. The 66 kDa SW2 band is shown.

a


mM Zinc Acetate
100
100
mM EDTA
50
100
100
50
100
100
100 fold excess Sp1 competitor
+
+
+
+
C1 (Sp1)
C2 (SpA)
C3 (Sp3)
*
*
Free probe
1
2
3
4
5
6
7
8

b

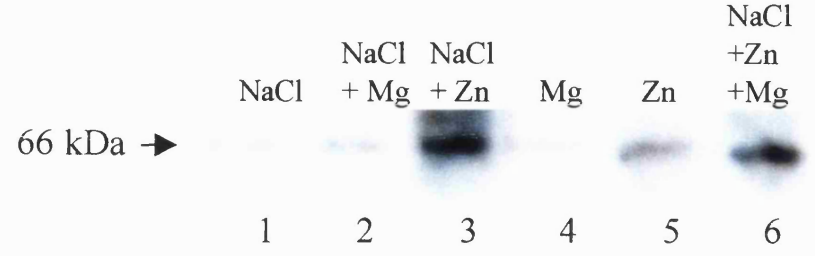

NaCl
NaCl + Mg
NaCl + Zn
Mg
Zn
NaCl +Zn +Mg
66 kDa
1
2
3
4
5
6

Figure 5.9. Assessment of the levels of Sp1 and Sp3 in epithelial and non-epithelial cells.

50 μg of nuclear extract from epithelial cells (T47D, ZR75 & MCF-7) and non-epithelial cells (HT1080 & HTB96) were separated by reducing SDS-PAGE, blotted and incubated with Sp1 antibody (a) then stripped and incubated with Sp3 antibody (b). The remaining proteins on the blotted gel were Coomassiestained to indicate loading (c).

Figure 5.9

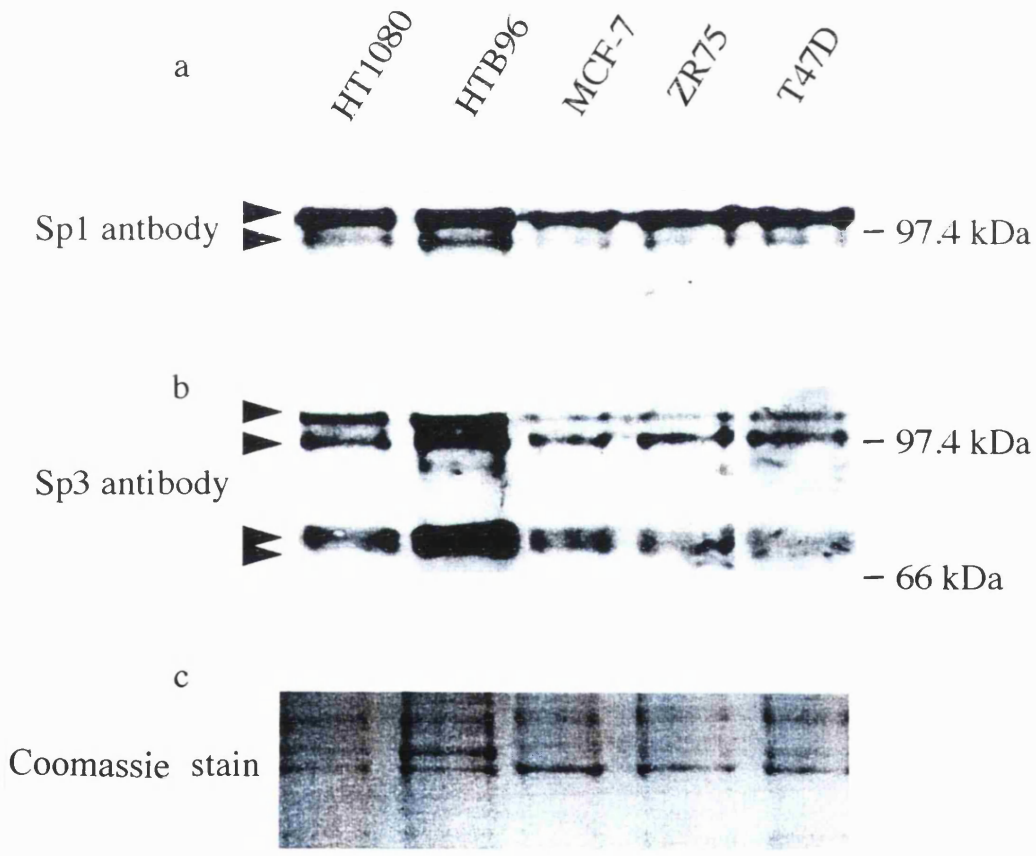

a
HT1080
HTB96
MCF-7
ZR75
T47D
Sp1 antbody
– 97.4 kDa
b
Sp3 antibody
– 97.4 kDa
– 66 kDa
c
Coomassie stain

Figure 5.10. Assessment of the levels of Sp1, Sp3 and SpA (C2) DNA binding in epithelial and non-epithelial cells.

A representative band-shift assay of 2-3μg of nuclear extracts from epithelial and non-epithelial cells pre-incubated with 100 fold over the labelled probe of unlabelled Sp1 consensus oligonucleotide (lanes 7-12) before the addition of wild-type -97 GC box radiolabelled oligonucleotide and separation on a 6% native gel (d).

Figure 5.10

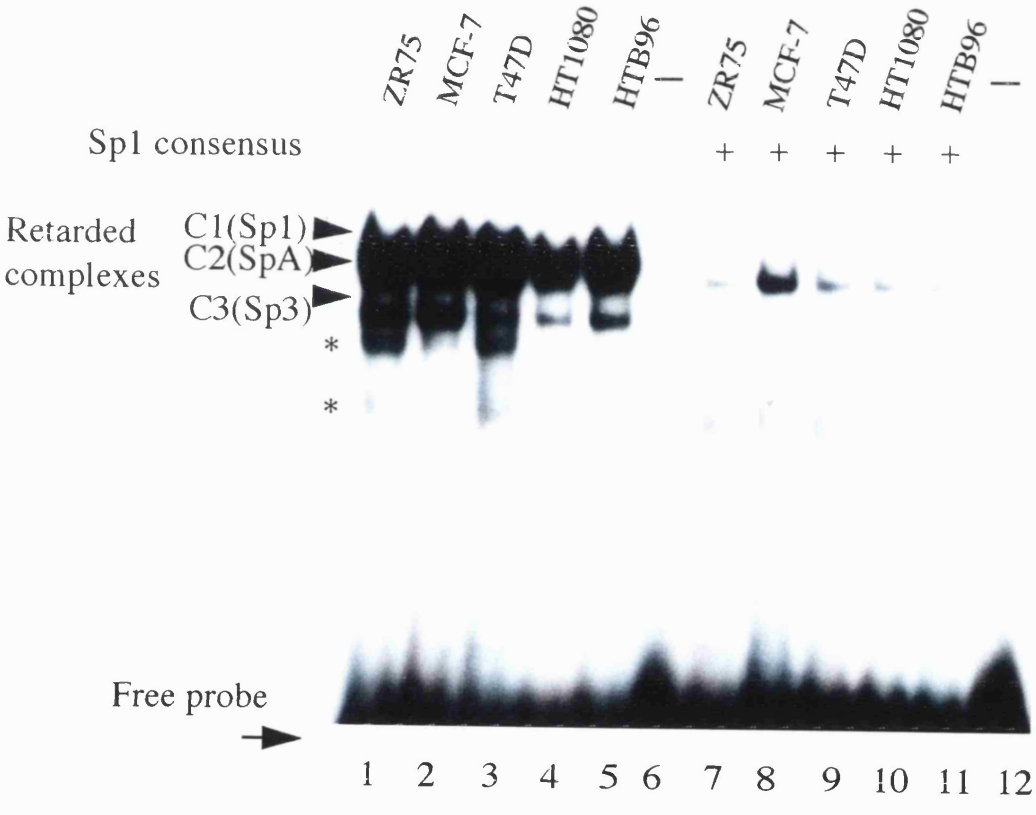

ZR75
MCF-7
T47D
HT1080
HTB96
–
ZR75
MCF-7
T47D
HT1080
HTB96
–
Sp1 consensus
+ + + + +
Retarded complexes
C1(Sp1)
C2(SpA)
C3(Sp3)
*
*
Free probe
1 2 3 4 5 6 7 8 9 10 11 12

Figure 5.11. Band-shift showing competition of complexes formed with the EMUC sequence by consensus sequences with high homology to EMUC.

2-3 μg of ZR75 nuclear extract was pre-incubated with an indicated molar excess of unlabelled oligonucleotides, self (lane 3), RXR DR1 (lane 4), USF-1 (lane 5) and E-Pal (lane 6) before the addition of labelled oligonucleotide corresponding to -95/-66 of the MUC1 promoter, the exact sequence of which is shown below. Table 4 in the text indicates the sequences of the competitors.

EMUC-95 GGGTTTTGTCACCTGTCACCTGCTGCCTGT -66

Figure 5.11

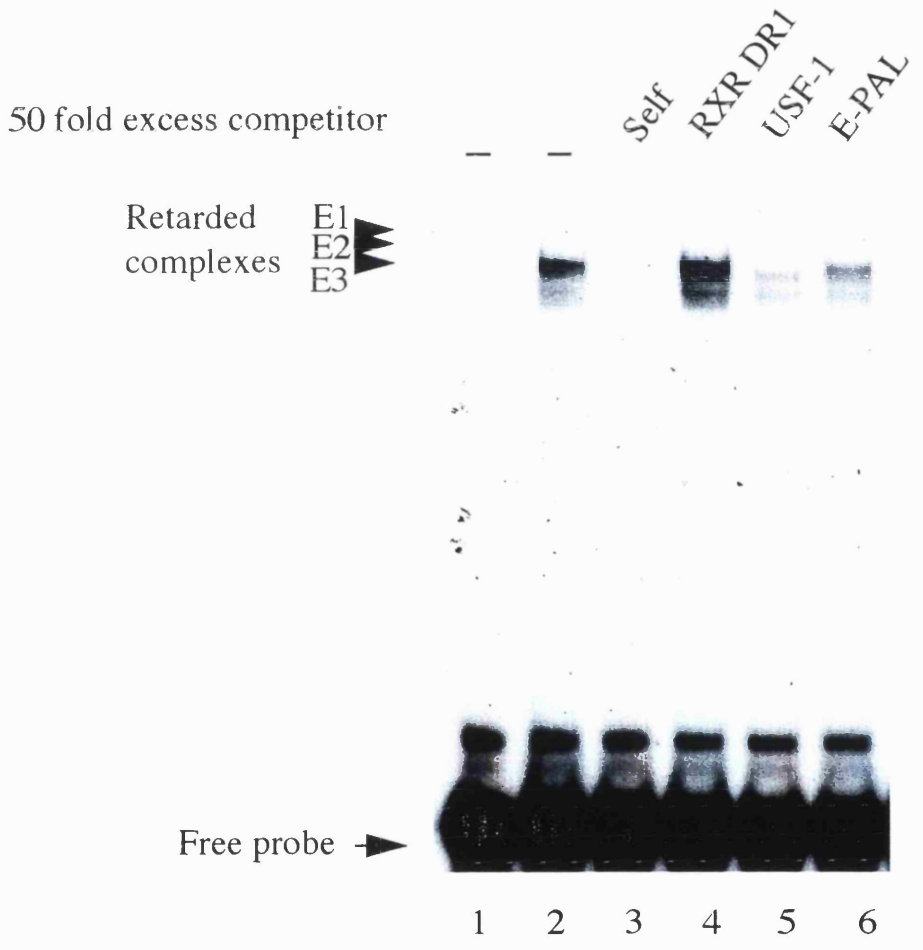

50 fold excess competitor
Self
RXR DR1
USF-1
E-PAL
Retarded complexes
E1
E2
E3
Free probe
1
2
3
4
5
6

Figure 5.12. Band-shift showing nucleotide specificity of complexes forming with the EMUC sequence.

2-3 µg of ZR75 nuclear extract was pre-incubated with an indicated excess of mutant EMUC oligonucleotide before the addition of labelled -95/-66 wild-type EMUC oligonucleotide and separation on a 6% native gel. Each mutant oligonucleotide used is represented by a code in the figure which corresponds to the sequence below (the mutation is underlined and in red). Across the top of the figure the wild type nucleotide is indicated by a capital, and the mutation made at that position is shown in lower case type. The reactions were run on a 6% native gel for an extended period to allow greater separation of complexes .

-95 GGGTTTTGTCACCTGTCACCTGCTGCCTGT -66	EMUC (wild-type)
-95 GGGTTTTGTCCCCTGTCACCTGCTGCCTGT -66	Ea
-95 GGGTTTTGTCAACTGTCACCTGCTGCCTGT -66	Eb
-95 GGGTTTTGTCACATGTCACCTGCTGCCTGT -66	Ec
-95 GGGTTTTGTCACCGGTCACCTGCTGCCTGT -66	Ed
-95 GGGTTTTGTCACCTTTCACCTGCTGCCTGT -66	Ee
-95 GGGTTTTGTCACCTGGCACCTGCTGCCTGT -66	Ef
-95 GGGTTTTGTCACCTGTAACCTGCTGCCTGT -66	Eg

The asterisks over the EMUC sequence indicate by their size the mutations which were most deleterious for competing factors E2 and E3.

Analysis of Factors
Figure 5.12

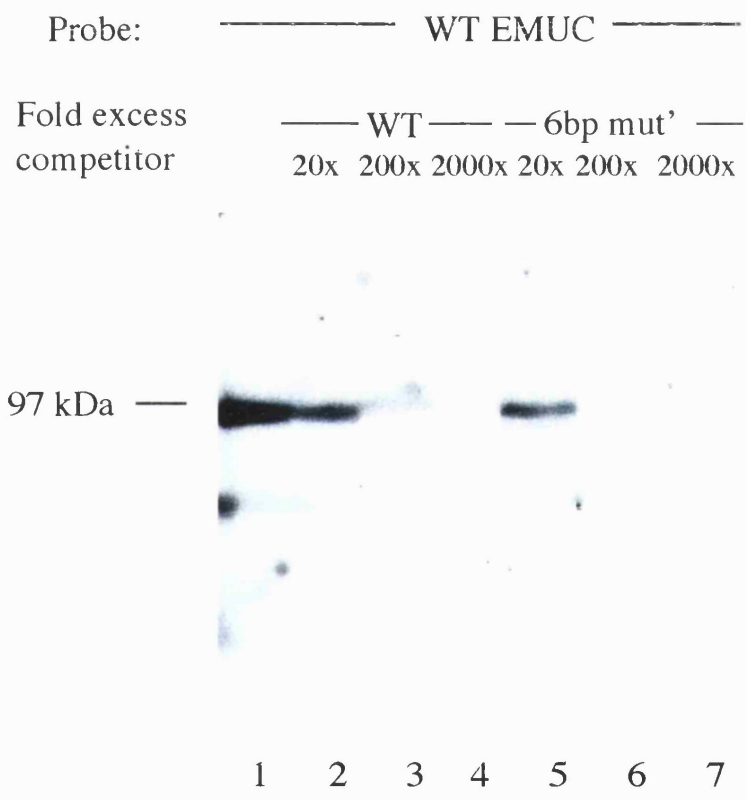
Figure 5.13. South-western showing the binding of *MUC1* EMUC sequence by separated proteins.

Nuclear extract from ZR75 cells was run on a reducing SDS-PAGE slab gel. The gel was blotted onto a membrane and the proteins denatured, then renatured through a Guanidium hydrochloride gradient without the reducing agent DDT. The membrane was cut into strips and incubated with south-western binding buffer and radiolabelled -95/-66 wild type EMUC probe (lane 1 to 7). Lanes 2 to 7 were also incubated with indicated fold excess molar amounts of unlabelled wild type EMUC sequence (lanes 2-4) or the 6 bp mutant (lanes 5-7) the sequence of which is shown below (the sequence of the mutated nucleotides is underlined and in red). After washing with south-western binding buffer the strips were re-united before autoradiography.

-95 GGGTTTTGTCACCTGTCACCTGCTGCCTGT -66 EMUC (wild-type)

-95 GGGTTTTGTCTTTAAACACCTGCTGCCTGT -66 EMUC (6 bp mutant)

Figure 5.13


Probe:
WT EMUC
Fold excess competitor
WT
6bp mut'
20x 200x 2000x 20x 200x 2000x
97 kDa
1 2 3 4 5 6 7

Chapter 6. Discussion

6.1 MUC1 expression from the *MUC1*-mini-gene *in vivo.*

The 10.6 kb genomic fragment of human *MUC1* had been shown previously to be capable of directing faithful expression in transgenic mice (46). Findings in this thesis show that just 1.4 kb of the 5' region plus the cDNA of the *MUC1* gene (the *MUC1*-mini-gene) are sufficient for tissue specific expression *in vivo*. In these mice MUC1 is expressed at high levels in mammary tumours and at lactation suggesting that all information necessary for transcriptional control are present within the *MUC1* mini-gene construct. Introns and the 3' region are not required.Theselata supports previous findings that showed the region 3' of the MUC1 coding region had no regulative activity on transcription (111). While it is possible that the cDNA contributes to the regulation of the gene it is unlikely because in mice transgenic for variants of the *Ras* oncogene transcribed from the 1.4 kb MUC1promoter, the *Ras* genes were expressed in MUC1 expressing cells (N. Freshney, Signal Transduction Laboratory, ICRF, personal communication) and the 1.4 kb region alone is sufficient to direct cell-type specific expression of reporter genes in transient transfection analysis (144, 111). Transcriptional control by the 5' region is also consistent with the observation that *MUC1* and the mouse homologue, *Muc-1*, which are expressed in a similar pattern, are 78% conserved at the 5' region (47, 135, 45).

Many authors have claimed that MUC1 is 'over-expressed' in mammary tumours (12, 145). While it can be said that total RNA from tumours in the *MUC1* mini-gene transgenic mice is enriched for MUC1 mRNA 51 fold over that of the resting mammary glands, it cannot yet be unequivocally stated that those mammary tumour cells 'over-express' MUC1. Tumours which develop in the mammary gland have a phenotype consistent with that of luminal epithelial cells, which expresses MUC1 normally (146). Quantification of MUC1 mRNA should therefore take into account the relative contribution of RNA from luminal epithelial cells to the whole cell population, which is likely to be higher in the tumour sample than in the resting mammary gland. Cytokeratin

19 is characteristic of luminal epithelial cells and is not expressed by surrounding stroma or basal epithelial cells (myoepithelia). In a continuation of this work, mRNA for cytokeratin 19 will be assessed as a measure of the luminal epithelial cell composition in tumours and resting breast. After analysis of the relative amounts of cytokeratin 19 and MUC1 mRNA it should be possible to state whether high levels of MUC1 in mammary tumours represents over-expression in individual mammary carcinoma cells, or reflects a proliferation of luminal epithelia.

6.2 The Sp1 and Sp3 binding activity of a single GC box is important in tissue-specific transcription of MUC1.

The promoter region -152/-62 had been shown to be required for epithelial transcription in transient transfection analysis (111). Experiments reported herein show that when sequences between -119/-62 within this region were cloned upstream of and adjacent to the constitutively active TK promoter they were capable of directing epithelial specific transcription of the reporter in transfected cells. Detailed analysis of complexes formed between nuclear extracts and the -119/-62 sequence revealed that factors which are, or are related to Sp1 and Sp3, together with a possibly novel factor named SpA, bind the -97 GC box. The EMUC sequence (-87/-75) on the other hand binds a factor closely related to USF and forms two complexes with non-E-box related factors. When a single Sp1 consensus GC box sequence, which binds Sp1 and Sp3, was cloned into the TK-proximal site it could substitute for the activity of the -119/-62 sequences. Therefore only the Sp1 and Sp3 binding activities are required to confer tissue specific activity to the TK promoter, and the other factors that bind the -119/-62 sequence are not required. While surprising, these findings are supported by previous data which showed that;

- a mutation in the -97 GC box of the 1.4 MUC1 promoter resulted in reduced transcription in epithelial cells (by an average of 77%), whereas in non-epithelial cells transcription from the mutant promoter was increased over the activity of the wild type (by an average of 41%) (112), suggesting that the site has both a repressive and activating activity.

- the -95 GC box is protected in DNAse I foot-printing, whereas the EMUC sequence is not (111).

The finding that EMUC sequences are not required to direct tissue-specific transcription is in contrast with previous data which showed that a six base-pair mutation (ACCTGT→TTTAAA) across the EMUC sequence resulted in a mutant promoter with increased transcriptional activity in non-epithelial cells and reduced transcriptional activity in epithelial cells (111). It must be noted that an identical mutation, made in the context of the -119/-62-TK construct was repressive in all cell types tested (figure 4.4). The six bp mutation potentially forms a secondary cruciform structure (147), and results obtained using this sequence may therefore be artefactual.

6.2.1 Activation and repression of the MUC1 promoter by Sp1 and Sp3

Sp1 over-expression caused high levels of transcription from the 1.4 kb promoter-reporter gene construct in non-epithelial cells, compared to a more modest increase in epithelial cells. Over-expression of Sp1 was also capable of activating transcription from the endogenous *MUC1* gene in non-epithelial cells to readily detectable levels. The observation that transcription can occur despite the probably methylated MUC1 tandem repeat is not at odds with earlier reports which have shown that methylation of the coding region of a gene does not prevent transcriptional elongation (reviewed in Jones, 1999 (115)). These observations show that an increase in Sp1 can release repression of MUC1, suggesting that Sp1 may be limiting in non-epithelial cell types.

In vitro analysis of the factors that complex with the -97 GC box showed that Sp1 and Sp3 factors and SpA bound the box with overlapping nucleotide requirements, suggesting that the three factors compete for the site. Sp3 has been reported both as a strong repressor and weak activator of transcription (148). Sp3 over-expression appeared to repress transcription from the MUC1 promoter, however this may reflect the weak activator function replacing bound Sp1. Since both the -119/-62 sequence and the consensus Sp1 GC box were able to repress transcription from the TK promoter below the level of TK alone in non-epithelial cells suggests that Sp3 is acting as a

repressor in these cell types. (This evidence is indirect and it is theoretically possible that another GC box binding protein is responsible). Sp3 is slightly higher in non-epithelial cells than in epithelial cells, as detected by immunoblot analysis, and therefore may represent the factor bound to the -97 GC box and other GC boxes more frequently than Sp1. Repression of transcription by bound Sp3 may partially explain the inactivity of the MUC1 promoter in non-epithelial cells. However the difference between Sp3 and Sp1 levels is not great and may not represent the full explanation.

6.2.2 Tissue-specific genes critically regulated by Sp1.

Sp1 is a ubiquitously expressed zinc finger protein that binds specifically to GC boxes and stimulates the initiation of transcription (149). It seems contradictory that a ubiquitous factor might mediate tissue-specific expression. However the MUC1 promoter is not the first to be shown to be regulated in this manner. The endogenous involucrin gene, expressed in squamous epithelial cells, can be activated in non-epithelial cells (fibroblasts and embryonic kidney cells) by transfection of Sp1. This activity is mediated by a 52 bp region containing an Sp1 and an AP1 site (150).

Other examples of epithelial specific genes whose tissue-specific expression has been shown to be critically regulated by Sp1 include the promoters for HPV-E6, E-cadherin, cytokeratin 18, IL-1 and Transglutaminase-3 (151, 152, 153). The α_2 integrin promoter is regulated in epithelial cells in response to c-erbB2 activation through Ras and Sp1 (154). Interestingly mutant derivatives of the Moloney murine sarcoma virus carrying an Sp1 site have transcriptional ability in epithelial cells (155).

However, transcriptional control by Sp1 is not limited to epithelial-specific genes and the tissue-specific expression of the Wilms' tumour-1 gene in kidney development (156) and CD11b and CD14 promoters in monocytes (157, 158) have been shown to be critically regulated by Sp1. It is possible that Sp1-mediated activation of epithelial specific promoters is a common mechanism of control but that Sp1 expression is unlikely to be the sole regulator of tissue-specific expression.

6.2.3 Differential activity of Sp1.

The results presented in this thesis are consistent with a mechanism in which Sp1 or Sp3 factors are unavailable to the promoter in non-epithelial and epithelial cells respectively and/or Sp3 concentration is higher in non-epithelial cells and therefore present at the site more often. When Sp1 is over-expressed it displaces Sp3 and activates transcription.

Sp1 was the first transcription factor to be cloned and over the years the complexity of potential interactions has grown. Sp1 activates transcription by interacting with the TBP-associated factor $TAF_{II}110$, which in turn binds to $TAF_{II}250$ (149, 159). More recently a 700 kDa multi-subunit cofactor, called CRSP (cofactor required for Sp1 activity), containing proteins with homology to yeast mediator subunits, nuclear receptor co-activators and elongation factors has been shown to be required for activity of Sp1 (160). Activation and repression of transcription by co-factors is thought to involve the combinatorial interaction of several subunits that in turn interact with the DNA bound transcription factor, nucleosomes, other co-factors, the core promoter sequences and the basal machinery (161, 160). Therefore subtle differences between the availability, composition and interaction of CRSP and other co-factors between cell types may determine the specificity of transcription. Nevertheless even if Sp1 transcriptional co-factors such as CRSP are limiting in non-epithelial cells the need for them must be overcome by Sp1 over-expression alone.

Alternatively Sp1 and Sp3 maybe differentially activated between MUC1-expressing and non-expressing cells. While at the present time there is little evidence for the regulation of Sp3 there is growing evidence for regulation of the activity and availability of Sp1, described below;

1. Complexes between cellular proteins and Sp1 may render Sp1 transcriptionally inactive. Two such complexes have been described in fibroblast cell lines. Sp1 can be super-activated by the retinoblastoma gene product (pRB) by its competitive interaction with a 20 kDa inhibitor of Sp1 called Sp1-I (162, 163, 164, 165, 166, 167). Negative regulation of Sp1 has also been correlated with the formation of an

Sp1 complex with a 75 kDa cellular protein (168). It is not yet known if such complexes are tissue-specific.

2. Differential expression between epithelial and non-epithelial cells of effectors for activators of Sp1 have been reported. Sp1 transcriptional activity has been reported to be increased by IGF-1 (169) and INT-γ (170), while the level of Sp1 mRNA has been shown to be increased by TNF-α, (171) and by viral infection (172).

6.3 Position effects of the -97 GC box on transcription.

The position of the introduced GC box had an influence on transcription since the consensus Sp1 GC box cloned adjacent to the TK promoter (-105 bp from the TK transcriptional start site) was capable of directing epithelial specific transcription, while the consensus Sp1 GC box cloned into the enhancer region (>1.8 kb from the TK promoter) was not. The position of GC boxes within the MUC1 promoter are similarly important since mutations in the distal -574 GC box reduced the activity of the promoter, but did not affect the tissue specificity, whereas mutation of the -97 GC box resulted in a loss of tissue specific expression (112). These observations are also consistent with an earlier report in which the distal *MUC1* GC box, within the sequence -595/-491 was capable of directing tissue-specific transcription when cloned adjacent to the HSV TK promoter (110).

The reason for the differential activity of the distal and proximal sites in both the MUC1 promoter and the TK system is not clear and at the present time can only be conjecture. Sp1 activation is a complex mechanism involving not only protein:DNA interactions but also the interaction of multiple areas of Sp1 with co-activator proteins (177, 149) and the higher-order interaction of Sp1 tetramers causing the DNA to loop (178). It is conceivable that the distance between the transcription machinery the CG box allows optimal access of Sp1 and Sp3 to co-activator proteins or basal machinery and the factors bound to such a box would be more influential than those bound elsewhere.

The effect of GC box context on transcriptional outcome is not without precedent. Using *Drosophila* SL2 cells which lack any endogenous Sp1 or Sp3 Birnbaum et al showed that only cellular promoters containing multiple Sp1 consensus GC boxes were repressed by Sp3. In contrast promoters containing a single Sp1 site were activated by Sp3 (173, 174). Sp3 has both activating and repressive domains and the repressive domain functions when Sp3 is bound to multimeric sites, presumably by interaction with specific co-factors (174, 175). It is possible therefore that the introduction of a further GC box adjacent to the TK promoter allows the repressive function of Sp3 to be realised in cells where Sp3 levels are higher. However such a mechanism would be inconsistent with the reported behaviour of Sp3 in which a single GC box allows activation, whereas two or three results in repression (173, 174) as two Sp1 GC box elements are already present in the TK promoter (176). Clearly promoter context can influence the activity of Sp1-family members and the possibility that the position of the Sp1 site may be critical to tissue-specific expression needs to be explored further.

6.4 MUC1 gene expression in carcinomas.

6.4.1 The role of SpA (C2).

Several groups have reported a complex formed in band-shifts with GC box sequences that has a mobility slightly faster than the Sp1 complex which is similar to that of C2 (SpA). In almost all cases the groups have referred to this complex as Sp3 (179, 151, 167). *In vitro* evidence shown in this thesis argues that SpA is not a member of the Sp1 family. It is not bound by antibodies specific to Sp1 family members, including an antibody raised to full length Sp3, and importantly both its DNA binding sequence and its size in SDS-PAGE gels differs from that of known Sp1-family members (137, 123, 179).

Initial work has suggested that the SpA factor, may function as a repressor of transcription in these cells. Mutations in the -97 GC box that prevent SpA binding but that did not affect Sp1 or Sp3 binding resulted in a mutant promoter with increased

transcriptional activity, particularly in the epithelial cell line expressing least MUC1, MCF-7. *In vitro* data showed that SpA levels were highest in the MCF-7 cell line, suggesting that SpA represses MUC1 expression. Since *in vitro* data also demonstrated that SpA could compete for DNA binding with Sp1 and Sp3 at the -97 GC box site, it may play an important part in the regulation of MUC1 transcription particularly within the epithelial cell lineage. Data herein shows that SpA is likely to be a single polypeptide of 66 kDa which requires zinc to bind DNA with a core sequence 'GGGGGNGN'. Other known GC box binding proteins that may be candidates for SpA include the zinc finger containing proteins ZBP-89 (180), RREB-1(181) BTEB2 (182), EKLF (183) and RN ZIF (184). As soon as antibodies to these factors become available they should be checked for the ability to bind C2 (SpA). However as none of these factors have the same reported size, expression pattern and activity as that shown in this thesis for SpA further attempts to clone the factor may be warranted.

6.4.2 Sp1 activity may be altered in malignancy.

Reports of an influence on Sp1 transactivation ability brought about by oncogenes and tumour suppressor genes are numerous and suggest a role for Sp1 in cellular growth. Sp1 and p53 heterocomplexes, formed when p53 levels are high in proliferating cells, show an increased DNA binding over Sp1alone (185). Many carcinomas have mutations that cause p53 to accumulate, however it is not yet known if the Sp1:p53 heterocomplexes form with mutated p53 which often has conformational changes (reviewed in (186)).

Sp1 has been reported to be regulated by three pivotal proteins of the cell cycle, pRB Cyclin D1 and E2F by various mechanisms. As only some of these effects have been mappedthrough the cell cycle and because of the complex interactions between these factors it is not yet clear whether activation or repression of Sp1- mediated transcription is the over all outcome at particular points of the cell cycle. pRB activates Sp1-mediated transcription in several ways, via direct interaction with Sp1 (187), through competition with an inhibitor of Sp1 (165) and via the interaction of pRB with $TAF_{II}250$ (188).

Cyclin D1 represses Sp1-mediated transcription independently of pRB, via $TAF_{II}250$ (189, 188) while E2F activates Sp1 via a direct interaction (190, 191).
The ability of Sp1 to activate transcription is increased by the expression of several viral oncogenes, v-Raf, v-Rel, v-Ras and v-Src (but not v-myb) (192, 193, 194). Downstream of Src, Ras and Raf the Ras-Erk pathway has been reported to increase Sp1 phosphorylation and Sp1:DNA binding 10-fold (195). Indeed support for a MAPK control of MUC1 gene expression comes from our own work which shows that the Mek-1 kinase inhibitor, PD98059, reduces endogenous MUC1 expression and expression form the 1.4 kb promoter in transient transfection analysis of mammary epithelial cells (data not shown, and A. Scibetta, Breast Cancer Biology Laboratory, ICRF, personal communication.). Together these observations suggest that a common aspect of transformation may involve an increase in expression from genes regulated by Sp1, such as MUC1. This mechanism may go some way in explaining why MUC1 expression is a common feature of carcinomas from different organs with presumably different genetic insults.

The work presented in this thesis establishes the promoter region as the agent responsible for many aspects of MUC1 expression *in vivo*. While *in vitro* experiments have shown that the -119/-62 section of the promoter is sufficient to direct tissue specific expression and that the activity is dependent on the Sp1/Sp3 binding activities of the -97 GC box. Sp3 over-expression reduced transcription from the MUC1 promoter whereas Sp1 over-expression increased it and was able to de-repress the endogenous gene in cells that do not normally express MUC1. Within the epithelial cell lines tested high levels of a third -97 GC box binding factor, SpA, correlated with increased transcription from a mutant promoter unable to bind the factor, suggesting that it acts to repress transcription.
The results of this thesis suggest the availability of Sp1 to a GC box at a specific position is important for tissue-specific expression of MUC1. Since Sp1 activity may be increased during malignancy the competition between SpA, Sp3 and Sp1 at this site may be of crucial importance to expression of the gene in carcinomas.

References

1. Ruluwela, L., Forster, A., Boehm, T., and Rabbitts, T. H. A rapid procedure for colony screening using nylon filters., Nucleic Acids Research. *17:* 452, 1989.

2. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Stuhl, K. Current Protocols in Molecular Biology, 1994.

3. Hoppe-Seyler, F., Butz, K., Rittmuller, C., and Doeberitz, M. V. K. A rapid microscale procedure for the simultaneous preparation of cytoplasmic RNA, nuclear DNA binding proteins and enzymatically active luciferase extracts., Nucleic Acids Research. *19:* 5080, 1991.

4. Sambrook, J., Fritsch, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual., 1988.

5. Vinson, C. R., LaMarco, K. L., Johnson, P. F., Landschulz, W. H., and McKnight, S. L. In situ detection of sequence specific DNA binding activity specified by a recombinant bacteriophage., Genes and Development. *2:* 801-806, 1988.

6. Dufort, D. and Nepveu, A. The human Cut homeodomain potein represses transcription from the c-myc promoter., Molecular and Cellular Biology. *14:* 4251-4257, 1994.

7. Taylor-Papadimitriou, J. and Gendler, S. J. Structure, biology and possible clinical applications of carcinoma-associated mucins (Review), International Journal of Oncology. *1:* 9-16, 1992.

8. Shimizu, M. and Yamauchi, K. Isolation and characterization of Mucin-like glycoprotein in human Milk Fat Globule Membrane, Journal of Biochemistry. *91:* 515-524, 1982.

9. Burchell, J., Gendler, S., Taylor-Papadimitriou, J., Girling, A., Lewis, A., Millis, R., and Lamport, D. Development and characterization of breast cancer reactive monoclonal antibodies directed to the core protein of the human milk mucin, Cancer Research. *47:* 5476-82, 1987.

10. Girling, A., Bartkova, J., Burchell, J., Gendler, S., Gillett, C., and Taylor-Papadimitriou, J. A core protein epitope of the polymorphic epithelial mucin detected by the monoclonal antibody SM-3 is selectively exposed in a range of primary carcinomas, International Journal of Cancer. *43:* 1072-6, 1989.

11. Hilkens, J., Buijs, F., and Ligtenberg, M. Complexity of MAM-6, an epithelial sialomucin associated with carcinomas, Cancer Research. *49:* 786-93, 1989.

12. Hareuveni, M., Tsarfaty, I., Zaretsky, J., Kotkes, P., Horev, J., Zrihan, S., Weiss, M., Green, S., Lathe, R., Keydar, I., and et al. A transcribed gene, containing a variable number of tandem repeats, codes for a human epithelial tumor antigen. cDNA cloning, expression of the transfected gene and over-expression in breast cancer tissue, European Journal of Biochemistry. *189:* 475-86, 1990.

13. Ho, S. B., Niehans, G. A., Lyftogt, C., Yan, P. S., Cherwitz, D. L., Gum, E. T., Dahiya, R., and Kim, Y. S. Heterogeneity of mucin gene expression in normal and neoplastic tissues, Cancer Research. *53:* 641-51, 1993.

14. Dong, Y., Walsh, M. D., Cummings, M. C., Wright, R. G., Khoo, S. K., Parsons, P. G., and McGuckin, M. A. Expression of MUC1 and MUC2 mucins in epithelial ovarian tumours, Journal of Pathology. *183:* 311-7, 1997.

15. Ring, C. J., Blouin, P., Martin, L. A., Hurst, H. C., and Lemoine, N. R. Use of transcriptional regulatory elements of the MUC1 and ERBB2 genes to drive tumour-selective expression of a prodrug activating enzyme, Gene Therapy. *4:* 1045-52, 1997.

16. Chen, L., Chen, D., Manome, Y., Dong, Y., Fine, H. A., and Kufe, D. W. Breast cancer selective gene expression and therapy mediated by recombinant adenoviruses containing the DF3/MUC1 promoter., Journal of Clinical Investigation. *96:* 2775-2782, 1995.

17. Tai, Y.-T., Strobel, T., Kufe, D., and Cannistra, S. A. In vivo cytotoxicity of ovarian cancer cells through tumour-selective expression of the BAX gene., Cancer Research. *59:* 2121-2126, 1999.

18. Burchell, J., Graham, R., and Taylor-Papadimitriou, J. Active specific immunotherapy: PEM as a potential target molecule. [Review] [52 refs], Cancer Surveys. *18:* 135-48, 1993.

19. Miles, D. W. and Taylor-Papadimitriou, J. Mucin based breast cancer vaccines., Exp. Opin. Invest. Drugs. *7:* 1865-1877, 1998.

20. Swallow, D. M., Gendler, S., Griffiths, B., Kearney, A., Povey, S., Sheer, D., Palmer, R. W., and Taylor-Papadimitriou, J. The hypervariable gene locus PUM, which

codes for the tumour associated epithelial mucins, is located on chromosome 1, within the region 1q21-24, Annals of Human Genetics. *51:* 289-94, 1987.

21. Middleton-Price, H., Gendler S., and Malcolm, S. Close linkage of PUM and SPTA within chromosome band 1q21, Ann, Hum, Genet. *52:* 273-278, 1988.

22. Vos, H. L., Devarayalus, S., Vries, Y. D., and Bornstein, P. Thrombospondin 3 (Thbs3), a new member of the Thrombospondin gene family, Journal of Biological Chemistry. *267:* 12192-12196, 1992.

23. Gendler, S. J. and Spicer, A. Epithelial Mucin Genes, Annual Review of Physiology. *57:* 607-34, 1995.

24. Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E. N., and Wilson, D. Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin, Journal of Biological Chemistry. *265:* 15286-93, 1990.

25. Lan, S. M., Batra, S. K., Qi, W.-N., Metzgar, R. S., and Hollingsworth, M. A. Cloning and sequencing of a human pancreatic tumor mucin cDNA, Journal of Biological Chemistry. *25:* 15294-15299, 1990.

26. Gendler, S. J., Spicer, A. P., Lalani, E. N., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Boshell, M., and Taylor-Papadimitriou, J. Structure and biology of a carcinoma-associated mucin, MUC1. [Review] [54 refs], American Review of Respiratory Disease. *144:* S42-7, 1991.

27. Muller, S., Goletz, S., Packer, N., Gooley, A., Lawson, A. M., and Hanisch, F.-G. Localization of O-Glycosylation sites on glycopeptide fragments from lactation-associated MUC1., Journal of Biologicla Chemistry. *272:* 24780-24793, 1997.

28. Muller, S., Alving, K., Peter-Katalinic, J., Zachara, N., Gooley, A., and Hanisch, F.-G. High density O-Glycosylation on the tandem repeat peptide from secretory MUC1 of T47D breast cencer cells., Journal of Biological Chemistry. *274:* 18165-18172, 1999.

29. Lloyd, K. O., Burchell, J., Kudryashov, V., Yin, B. W. T., and Taylor-Papadimitriou, J. Comparison of O-linked carbohydrate chains in MUC1 mucin from

normal breast epithelial cell lines and breast carcinoma cell lines., Journal of Biological Chemistry. *271:* 33325-33334, 1996.

30. Brockhausen, I., Yang, J. M., Burchell, J., Whitehouse, C., and Taylor-Papadimitriou, J. Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells, European Journal of Biochemistry. *233:* 607-17, 1995.

31. Whitehouse, C., Burchell, J., Gschmeissner, S., Brockhausen, I., Lloyd, K. O., and Taylor-Papadimitriou, J. A transfected sialyltransferase that is elevated in breast cancer and localizes to the medial/trans-Golgi apparatus inhibits the development of core-2-based O-glycans, Journal of Cell Biology. *137:* 1229-41, 1997.

32. Burchell, J. and Taylor-Papadimitriou, J. Effect of modification of carbohydrate side chains on the reactivity of antibodies with core-protein epitopes of the MUC1 gene product, Epithelial Cell Biology. *2:* 155-62, 1993.

33. Fontenot, J. D., Tjandra, N., Bu, D., Ho, C., Montelaro, R. C., and Finn, O. J. Biophysical characterization of one-, two-, and three-tandem repeats of human mucin (muc-1) protein core, Cancer Research. *53:* 5386-94, 1993.

34. Hilkens, J., Ligtenberg, M. J., Vos, H. L., and Litvinov, S. V. Cell membrane-associated mucins and their adhesion-modulating property. [Review] [38 refs], Trends in Biochemical Sciences. *17:* 359-63, 1992.

35. Wesseling, J., van der Valk, S. W., and Hilkens, J. A mechanism for inhibition of E-cadherin-mediated cell-cell adhesion by the membrane-associated mucin episialin/MUC1, Molecular Biology of the Cell. *7:* 565-77, 1996.

36. Burchell, J., Wang, D., and Taylor-Papadimitriou, J. Detection of the tumour-associated antigens recognized by the monoclonal antibodies HMFG-1 and 2 in serum from patients with breast cancer, International Journal of Cancer. *34:* 763-8, 1984.

37. Boshell, M., Lalani, E. N., Pemberton, L., Burchell, J., Gendler, S., and Taylor-Papadimitriou, J. The product of the human MUC1 gene when secreted by mouse cells transfected with the full-length cDNA lacks the cytoplasmic tail, Biochemical & Biophysical Research Communications. *185:* 1-8, 1992.

38. Ligtenberg, M. J., Kruijshaar, L., Buijs, F., van Meijer, M., Litvinov, S. V., and Hilkens, J. Cell-associated episialin is a complex containing two proteins derived from a common precursor, Journal of Biological Chemistry. *267:* 6171-7, 1992.

39. Zrihan-Licht, S., Vos, H. L., Baruch, A., Elroy-Stein, O., Sagiv, D., Keydar, I., Hilkens, J., and Wreschner, D. H. Characterization and molecular cloning of a novel MUC1 protein, devoid of tandem repeats, expressed in human breast cancer tissue, European Journal of Biochemistry. *224:* 787-95, 1994.

40. Baruch, A., Hartmann, M., Zrihan-Licht, S., Greenstein, S., Burstein, M., Keydar, I., Weiss, M., Smorodinsky, N., and Wreschner, D. H. Preferential expression of novel MUC1 tumor antigen isoforms in human epithelial tumors and their tumor-potentiating function, International Journal of Cancer. *71:* 741-9, 1997.

41. Nollet, S., Moniaux, N., Maury, N., Petitprez, J., Degand, D., Laine, A., Porchet, N., and Aubert, J.-P. Human mucin gene *MUC4*: organization of its 5'-region and polymorphism of its central tandem repeat array., Biochemical Journal. *332:* 739-748, 1998.

42. Carraway, K. L., Fregien, N., Carraway III, K. L., and Carothers-Carraway, C. A. Tumour sialomucin complexes as tumour antigens and modulators of cellular interactions and proliferation., Journal of Cell Science. *103:* 299-307, 1992.

43. Pemberton, L., Taylor-Papadimitriou, J., and Gendler, S. J. Antibodies to the cytoplasmic domain of the MUC1 mucin show conservation throughout mammals, Biochemical & Biophysical Research Communications. *185:* 167-75, 1992.

44. Spicer, A. P., Parry, G., Patton, S., and Gendler, S. J. Molecular cloning and analysis of the mouse homologue of the tumor-associated mucin, MUC1, reveals conservation of potential O-glycosylation sites, transmembrane, and cytoplasmic domains and a loss of minisatellite-like polymorphism, Journal of Biological Chemistry. *266:* 15099-109, 1991.

45. Vos, H. L., de Vries, Y., and Hilkens, J. The mouse episialin (Muc1) gene and its promoter: rapid evolution of the repetitive domain in the protein, Biochemical & Biophysical Research Communications. *181:* 121-30, 1991.

46. Peat, N., Gendler, S. J., Lalani, N., Duhig, T., and Taylor-Papadimitriou, J. Tissue-specific expression of a human polymorphic epithelial mucin (MUC1) in transgenic mice, Cancer Research. *52:* 1954-60, 1992.

47. Braga, V. M., Pemberton, L. F., Duhig, T., and Gendler, S. J. Spatial and temporal expression of an epithelial mucin, Muc-1, during mouse development, Development. *115:* 427-37, 1992.

48. Zotter, S., Hageman, P. C., Lossnitzer, A., van den Tweel, J., Hilkens, J., Mooi, W. J., and Hilgers, J. Monoclonal antibodies to epithelial sialomucins recognize epitopes at different cellular sites in adenolymphomas of the parotid gland, International Journal of Cancer - Supplement. *3:* 38-44, 1988.

49. Hilkens, J., Wesseling, J., Vos, H. L., Storm, J., Boer, B., van der Valk, S. W., and Maas, M. C. Involvement of the cell surface-bound mucin, episialin/MUC1, in progression of human carcinomas, Biochemical Society Transactions. *23:* 822-6, 1995.

50. Pimental, R. A., Julian, J., Gendler, S. J., and Carson, D. D. Synthesis and intracellular trafficking of Muc-1 and mucins by polarized mouse uterine epithelial cells, Journal of Biological Chemistry. *271:* 28128-37, 1996.

51. Graham, R. A., Stewart, L. S., Peat, N. P., Beverley, P., and Taylor-Papadimitriou, J. MUC1-based immunogens for tumor therapy: development of murine model systems., Tumor Targeting. *1:* 211-221, 1995.

52. Spicer, A. P., Rowse, G. J., Lidner, T. K., and Gendler, S. J. Delayed mammary tumor progression in Muc-1 null mice, Journal of Biological Chemistry. *270:* 30093-101, 1995.

53. Seregni, E., Botti, C., Massaron, S., Lombardo, C., Capobianco, A., Bogni, A., and Bombardieri, E. Structure, function and gene expression of epithelial mucins. [Review] [74 refs], Tumori. *83:* 625-32, 1997.

54. Surveyor, G. A., Gendler, S. J., Pemberton, L., Das, S. K., Chakraborty, I., Julian, J., Pimental, R. A., Wegner, C. C., Dey, S. K., and Carson, D. D. Expression and steroid hormonal control of Muc-1 in the mouse uterus, Endocrinology. *136:* 3639-47, 1995.

55. DeSouza, M. M., Mani, S. K., Julian, J., and Carson, D. D. Reduction of mucin-1 expression during the receptive phase in the rat uterus, Biology of Reproduction. *58:* 1503-7, 1998.

56. Hey, N. A., Graham, R. A., Seif, M. W., and Aplin, J. D. The polymorphic epithelial mucin MUC1 in human endometrium is regulated with maximal expression in the implantation phase, Journal of Clinical Endocrinology & Metabolism. *78:* 337-42, 1994.

57. DeLoia, J. A., Krasnow, J. S., Brekosky, J., Babaknia, A., Julian, J., and Carson, D. D. Regional specialization of the cell membrane-associated, polymorphic mucin (MUC1) in human uterine epithelia, Human Reproduction. *13:* 2902-9, 1998.

58. Aplin, J. D., Hey, N. A., and Graham, R. A. Human endometrial MUC1 carries keratan sulfate: characteristic glycoforms in the luminal epithelium at receptivity, Glycobiology. *8:* 269-76, 1998.

59. Gollub, E. G., Waksman, H., Goswami, S., and Marom, Z. Mucin genes are regulated by estrogen and dexamethasone, Biochemical & Biophysical Research Communications. *217:* 1006-14, 1995.

60. Botti, C., Seregni, E., Lombardo, C., Massaron, S., and Bombardieri, E. Effects of steroid-freefoetalserum and steroid supplementation on MUC1 gene expression in human breast cancer cell line MCF7, Anticancer Research. *17:* 205-8, 1997.

61. McGuckin, M. A., Quin, R. J., and Ward, B. G. Progesterone stimulates production and secretion of MUC1 epithelial mucin in steroid-responsive breast cancer cell lines, International Journal of Oncology. *12:* 939-45, 1998.

62. Gaemers, I., Volders, H., Vos, H., Valk, v. S., and Hilkens, J. Regulation of episialin/MUC1 expression in breast carcinomas: cross-talk between the STAT- and Glucocorticoid receptor signalling pathways., 5th International Workshop on Carcinoma-Associated Mucins. *5:*, 1998.

63. Parry, G., Li, J., Stubbs, J., Bissell, M. J., Schmidhauser, C., Spicer, A. P., and Gendler, S. J. Studies of Muc-1 mucin expression and polarity in the mouse mammary gland demonstrate developmental regulation of Muc-1 glycosylation and establish the hormonal basis for mRNA expression, Journal of Cell Science. *101:* 191-9, 1992.

64. Ginsburg, E. and Vonderhaar, B. K. Prolactin Synthesis and Secretion by human breast cancer cells, Cancer Research. *55:* 2591-2595, 1995.

65. Watson, C. J. and Miller, W. R. Elevated levels of members of the STAT family of transcription factors in breast carcinoma nuclear extracts, British Journal of Cancer. *71:* 840-844, 1995.

66. Vos, H. L., Van der Valk, S. W., Prinsenberg, T., Mooi, W. J., and Hilkens, J. IL-6 induction of the MUC1/episialin promoter in T47D cells, 4th International Workshop on Carcinoma-Associated Mucins 30, 1996.

67. Delsol, G., Gatter, K. C., Stein, H., Erber, W. N., Pulford, K. A., Zinne, K., and Mason, D. Y. Human lymphoid cells express epithelial membrane antigen. Implications for diagnosis of human neoplasms, Lancet. *2:* 1124-9, 1984.

68. Agrawal, B., Krantz, M. J., Parker, J., and Longenecker, B. M. Expression of MUC1 mucin on activated human T cells: implications for a role of MUC1 in normal immune regulation, Cancer Research. *58:* 4079-81, 1998.

69. Abe, M. and Kufe, D. Characterization of cis-acting elements regulating transcription of the human DF3 breast carcinoma-associated antigen (MUC1) gene, Proceedings of the National Academy of Sciences of the United States of America. *90:* 282-6, 1993.

70. Shirotani, K., Taylor-Papadimitriou, J., Gendler, S. J., and Irimura, T. Transcriptional regulation of the MUC1 mucin gene in colon carcinoma cells by a soluble factor. Identification of a regulatory element, Journal of Biological Chemistry. *269:* 15030-5, 1994.

71. Shirotani, K. and Irimura, T. Purification of nuclear proteins that potentially regulate transcription of the MUC1 mucin gene induced by a soluble factor., Journal of Biochemistry. *124:* 585-590, 1998.

72. Hilkens, J., Vos, H. L., Wesseling, J., Boer, M., Storm, J., van der Valk, S., Calafat, J., and Patriarca, C. Is episialin/MUC1 involved in breast cancer progression?. [Review] [22 refs], Cancer Letters. *90:* 27-33, 1995.

73. Bartek, J., Bartkova, J., Kyprianou, N., Lalani, E. N., Staskova, Z., Shearer, M., Chang, S., and Taylor-Papadimitriou, J. Efficient immortalization of luminal epithelial

cells from human mammary gland by introduction of simian virus 40 large tumor antigen with a recombinant retrovirus, Proceedings of the National Academy of Sciences of the United States of America. *88:* 3520-4, 1991.

74. Zaretsky, J. Z., Weiss, M., Tsarfaty, I., Hareuveni, M., Wreschner, D. H., and Keydar, I. Expression of genes coding for pS2, c-erbB2, estrogen receptor and the H23 breast tumor-associated antigen. A comparative analysis in breast cancer, FEBS Letters. *265:* 46-50, 1990.

75. Gendler, S. J., Cohen, E. P., Craston, A., Duhig, T., Johnstone, G., and Barnes, D. The locus of the polymorphic epithelial mucin (PEM) tumour antigen on chromosome 1q21 shows a high frequency of alteration in primary human breast tumours, International Journal of Cancer. *45:* 431-5, 1990.

76. Bieche, I., Ruffet, E., Zweibaum, A., Vilde, F., Lidereau, R., and Franc, B. MUC1 mucin gene, transcripts, and protein in adenomas and papillary carcinomas of the thyroid, Thyroid. *7:* 725-31, 1997.

77. Devilee, P., van Vilet, M., Bardoel, A., Kievits, T., Kuipers-Dijkshoorn, N., Pearson, P. L., and Cornelisse, C. J. Frequent somaticimbalance ofmarker alleles for chromosome 1 in human primary breast carcinoma, Cancer Research. *51:* 1020-1025, 1991.

78. Bieche, v., Champeme, M.-H., and Lidereau, R. Loss and gain of distinct regions of chromosome 1q in primary breast cancer, Clinical Cancer Research. *1:* 123-127, 1995.

79. Merlo, G. R., Siddiqui, J., Cropp, C. S., Liscia, D. S., Lidereau, R., Callahan, R., and Kufe, D. W. Frequent alteration of the DF3 tumor-associated antigen gene in primary human breast carcinomas., Cancer Research. *49:* 6966-6971, 1989.

80. Jentoft, N. Why are proteins O-glycosylated?, Trends in Biochemical Science. *15:* 291-294, 1990.

81. Schroten, H. The benefits of human milk fat globule against infection, 14 52-53, 1998.

82. Carvalho, F., Garcia, E., and Sobrinho-Simoes, D. M. Small MUC1 genotypes predispose to Helicobacter from northern Portugal, Epidemiology and Transmission. *3:* A28 (03/141), 1997.

83. Baruch, A., Hartmann, M.-L., Yoeli, M., Adereth, Y., Greenstein, S., Stadler, Y., Skornik, Y., Zaretsky, J., Smorodinsky, N. I., Keydar, I., and Wreschner, D. H. The breast cancer-associated MUC1 gene generates both a receptor and its cognate binding protein., Cancer Research. *59:* 1552-1561, 1999.

84. Pandey, P., Kharbanda, S., and Kufe, D. Association of the DF3/MUC1 breast cancer antigen with Grb2 and the Sos/Ras exchange protein, Cancer Research. *55:* 4000-3, 1995.

85. Zrihan-Licht, S., Baruch, A., Elroy-Stein, O., Keydar, I., and Wreschner, D. H. Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine receptor-like molecules, FEBS Letters. *356:* 130-136, 1994.

86. Yongqing, L., Bharti, A., Chen, D., Gong, J., and Kufe, D. Interaction of Glycogen Synthase Kinase 3B with the DF3/MUC1 carcinoma-associated antigen and B-catenin., Molecular and Cellular Biology. *18:* 7216-7224, 1998.

87. Kondo, K., Kohno, N., Yokoyama, A., and Hiwada, K. Decreased MUC1 expression induces E-Cadherin-mediated cell adhesion of breast cancer cells, Cancer Research. *58:* 2014-2019, 1998.

88. Wesseling, J., van der Valk, S. W., Vos, H. L., Sonnenberg, A., and Hilkens, J. Episialin (MUC1)over-expressioninhibits integrin-mediated cell adhesion to extracellular matrix components, Journal of Cell Biology. *129:* 255-65, 1995.

89. Parry, G., Beck, J. C., Moss, L., Bartley, J., and Ojakian, G. K. Determination of apical membrane polarity in mammary epithelial cell cultures: the role of cell-cell, cell-substratum, and membrane-cytoskeleton interactions., Experimental Cell Research. *188:* 302-311, 1990.

90. Birchmeier, W. and Behrens, J. Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness. [Review] [180 refs], Biochemica et Biophysica Acta. *1198:* 11-26, 1994.

91. Baeckstrom, D., Hansson, G. C., Nilsson, O., Johansson, C., Gendler, S. J., and Lindholm, L. Purification and characterization of a membrane-bound and a secreted mucin-type glycoprotein carrying the carcinoma-associated sialyl-Le^a epitope on distinct core proteins, Journal of Biological Chemistry. *266:* 21537-47, 1991.

92. Hanski, C., Drechsler, K., Hanisch, F.-G., Sheehan, J., Manske, M., Ogorek, D., Klussmann, E., Hanski, M.-L., Blank, M., Xing, P.-X., McKenzie, I. F. C., Devine, P. L., and Riecken, E.-O. Altered glycosylation of the MUC-1 protein core contributes to the colon carcinoma-associated incerase of mucin-bound sialyl-lewisX expression., Cancer Research. *53:* 4082-4088, 1993.

93. Regimbald, L. H., Pilarski, L. M., Longenecker, B. M., Reddish, M. A., Zimmermann, G., and Hugh, J. C. The breast mucin MUC1 as a novel adhesion ligand for endothelial intercellular adhesion molecule 1 in breast cancer., Cancer Research. *56:* 4244-4249, 1996.

94. Kam, J. L., Regimbald, L. H., Hilgers, J. H. M., Hoffman, P., Krantz, M. J., Longenecker, B. M., and Hugh, J. C. MUC1 synthetic peptide inhibition of intercellular adhesion molecule 1 and MUC1 binding requires six tandem repeats, Cancer Research. *58:* 5577-5581, 1998.

95. Barnd, D. L., Lan, M. S., Metzgar, R. S., and Finn, O. J. Specific, major histocompatibility complex-unrestricted recognition of tumour-associated mucins by human cytotoxic T cells., Proceedings of the National Acadamy of Sciences USA. *86:* 7159-7163, 1989.

96. Jerome, K. R., Barnd, D. L., Bendt, K. M., Boyer, C. M., Taylor-Papadimitriou, J., McKenzie, I. F., Bast, R. C., Jr., and Finn, O. J. Cytotoxic T-lymphocytes derived from patients with breast adenocarcinoma recognize an epitope present on the protein core of a mucin molecule preferentially expressed by malignant cells, Cancer Research. *51:* 2908-16, 1991.

97. Taylor-Papadimitriou, J., Peat, N., Graham, R., and Burchell, J. Aberrant glycoslation of the MUC1 gene product in adenocarcinomas: Implications for immunotherapy., Complex Carbohydrates and Drug Research. *36:* 382-393, 1984.

98. von Mensdorff-Pouilly, S., Gourevitch, M. M., Kenemans, P., Verstraeten, A. A., Litvinov, S. V., van Kamp, G. J., Meijer, S., Vermorken, J., and Hilgers, J. Humoral immune response to polymorphic epithelial mucin (MUC-1) in patients with benign and malignant breast tumours, European Journal of Cancer. *32A:* 1325-31, 1996.

99. Gourevitch, M. M., von Mensdorff-Pouilly, S., Litvinov, S. V., Kenemans, P., van Kamp, G. J., Verstraeten, A. A., and Hilgers, J. Polymorphic epithelial mucin (MUC-1)-containing circulating immune complexes in carcinoma patients, British Journal of Cancer. *72:* 934-8, 1995.

100. Hansson, G. Sialyl-Lewis a carrying MUC1 and CD43 mucin-type glycoproteins expressed and secreted from colon carcinoma cells - a role in tumor progression?, Trends in Glycosciences and Glycotechnology. *9:* 211-222, 1997.

101. Agrawal, B., Krantz, M. J., Reddish, M. A., and Longenecker, B. M. Cancer-associated MUC1 mucin inhibits human T-cell proliferation, which is reversible by IL-2., Nature Medicine. *4:* 43-49, 1998.

102. Lee, J., Fenton, B. M., Koch, C. J., Frelinger, J. G., and Lord, E. M. Interleukin 2 expression by tumor cells alters both the immune response and the tumor microenvironment., Cancer Research. *58:* 1478-1485, 1998.

103. Apostolopoulos, V., Lofthouse, S. A., Popovski, V., Chelvanayagam, G., Sandrin, M. S., and McKenzie, I. F. Peptide mimics of a tumor antigen induce functional cytotoxic T cells, Nature Biotechnology. *16:* 276-80, 1998.

104. Gong, J., Chen, D., Kashiwaba, M., and Kufe, D. Induction of antitumor activity by immunization with fusions of dendritic and carcinoma cells, Nature Medicine. *3:* 558-61, 1997.

105. Graham, R. A., Burchell, J. M., Beverley, P., and Taylor-Papadimitriou, J. Intramuscular immunisation with MUC1 cDNA can protect C57 mice challenged with MUC1-expressing syngeneic mouse tumour cells, International Journal of Cancer. *65:* 664-70, 1996.

106. Gong, J., Chen, D., Kashiwaba, M., Li, Y., Chen, L., Takeuchi, H., Qu, H., Rowse, G. J., Gendler, S. J., and Kufe, D. Reversal of tolerance to human MUC1 antigen in MUC1 transgenic mice immunized with fusions of dendritic and carcinoma cells, Proceedings of the National Academy of Sciences of the United States of America. *95:* 6279-83, 1998.

107. Smith, M., Burchell, J. M., Graham, R., Cohen, E., and Taylor-Papadimitriou, J. Expression of B7.1 in a MUC1-expressing mouse mammary epithelial tumour cell line

inhibits tumorigenicity, but does not induce auto-immunity in MUC1 transgenic mice., J. Immunology. *97:* 648-55, 1999.

108. Scholl, S., Acres, B., and Scatz, C. The polymorphic epithelial mucin (MUC1): A phase I clinical trial testing the tollerance and immunogenicity of a vaccinia virus-MUC1-IL2 construct in breast cancer., Breast Cancer Research and Treatment. *97:* Abstract 269, 1997.

109. Chen, L., Pulsipher, M., Chen, D., Sieff, C., Elias, A., Fine, H. A., and Kufe, D. W. Selective transgene expression for detection and elimination of contaminating carcinoma cells in hematopoietic stem cell sources., Journal of Clinical Investigation. *98:* 2539-2548, 1996.

110. Manome, Y., Abe, M., Hagen, M. F., Fine, H. A., and Kufe, D. W. Enhancer sequences of the DF3 gene regulate expression of the herpes simplex virus thymidine kinase gene and confer sensitivity of human breast cancer cells to ganciclovir., Cancer Research. *54:* 5408-5413, 1994.

111. Kovarik, A., Peat, N., Wilson, D., Gendler, S. J., and Taylor-Papadimitriou, J. Analysis of the tissue-specific promoter of the MUC1 gene, Journal of Biological Chemistry. *268:* 9917-26, 1993.

112. Kovarik, A., Lu, P. J., Peat, N., Morris, J., and Taylor-Papadimitriou, J. Two GC boxes (Sp1 sites) are involved in regulation of the activity of the epithelium-specific MUC1 promoter, Journal of Biological Chemistry. *271:* 18140-7, 1996.

113. Abe, M. and Kufe, D. Transcriptional regulation of DF3 gene expression in human MCF-7 breast carcinoma cells, Journal of Cellular Physiology. *143:* 226-31, 1990.

114. Singer, M. and Berg, P. Global Influences on Gene Expression, Genes and Genomes 601-613, 1991.

115. Jones, P. A. The DNA methylation Paradox, Trends in Genetics. *15:* 34-37, 1999.

116. Zrihan-Licht, S., Weiss, M., Keydar, I., and Wreschner, D. H. DNA methylation status of the MUC1 gene coding for a breast-cancer-associated protein, International Journal of Cancer. *62:* 245-51, 1995.

117. Hennig, G., Lowrick, O., Birchmeier, W., and Behrens, J. Mechanisms identified in the transcriptional control of epithelial gene expression, Journal of Biological Chemistry. *271:* 595-602, 1996.

118. Behrens, J., Lowrick, O., Klein-Hitpass, L., and Birchmeier, W. The E-cadherin promoter: functional analysis of a G.C-rich region and an epithelial cell-specific palindromic regulatory element, Proceedings of the National Academy of Sciences of the United States of America. *88:* 11495-9, 1991.

119. Ben-Ze'ev, A. and Geiger, B. Differential molecular interactions of B-catenin and plakoglobin in adhesion., Current Biology. *10:* 629-639, 1998.

120. Eastman, Q. and Grosschedl, R. Regulation of LEF-1/TCF transcription factors and by Wnt and other signals., Current Opinion in Cell Biology. *11:*, 1999.

121. Yamamoto, M., Bharti, A., Li, Y., and Kufe, D. Interaction of the DF3/MUC1 breast carcinoma-associated antigen and beta-catenin in cell adhesion, Journal of Biological Chemistry. *272:* 12492-4, 1997.

122. Lancaster, C. A., Peat, N., Duhig, T., Wilson, D., Taylor-Papadimitriou, J., and Gendler, S. J. Structure and expression of the human polymorphic epithelial mucin gene: an expressed VNTR unit, Biochemical & Biophysical Research Communications. *173:* 1019-29, 1990.

123. Hagen, G., Muller, S., Beato, M., and Suske, G. Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes., Nucleic Acids Research. *20:* 5519-5525, 1992.

124. Taylor-Papadimitriou, J., Peterson, J. A., Arklie, J., Burchell, J., Ceriani, R. L., and Bodmer, W. F. Monoclonal antibodies to epithelium-specific components of the human milk fat globule membrane: production and reaction with cells in culture, International Journal of Cancer. *28:* 17-21, 1981.

125. Frebourg, T. and Brison, O. Plasmid vectors with multiple cloning sites and cat-reporter gene for promoter cloning and analysis in animal cells, Gene. *65:* 315-8, 1988.

126. Stein, B., Rahmsdorf, H. J., Steffen, A., Litfin, M., and Herrlich, P. UV-induced DNA damage is anintermediate step in UV-induced expression of Human

Immunodeficiency Virus Type 1,Collagenase, c-fos and metallothionein., Molecular and Cellular Biology. *9:* 5169-5181, 1989.

127. Lucklow, B. and Schutz, G. CAT constructs with multiplue unique restriction sites for the functional eukaryotic promoters and regulatory elements., Nucleic Acids Research. *15:* 5490, 1987.

128. Hagen, G., Muller, S., Beato, M., and Suske, G. Sp1-mediated transcriptional activation is repressed by Sp3., EMBO Journal. *13:* 3843-3851, 1994.

129. Treisman, R., Novak, U., Favaloro, J., and Kamen, R. Transformation of rat cellsby an altered polyoma virus genome expressing only the middle-T protein., Nature. *292:* 595-600, 1981.

130. Brooks, S. C., Locke, E. R., and Soule, H. D. Estrogen receptor in a human cell line (MCF-7) from breast carcinoma, Journal of Biological Chemistry. *248:* 6251-6253, 1973.

131. Keydar, I., Chen, L., Karby, S., Weiss, F. R., Delarea, J., Radu, M., Chaitcik, S., and Brenner, H. J. Establishment and characterization of a cell line of human breast carcinoma origin, European Journal of Cancer. *15:* 659-670, 1979.

132. Engel, L. W., Young, N. A., Tralka, T. S., Lippman, M. E., O'Brien, S. J., and Joyce, M. J. Establishment and characterisation of three new continuos cell lines derived from human breast carcinomas., Cancer Research. *38:* 3352-3364, 1978.

133. Harlow, E. and Lane, D. P. Antibodies: A laboratory Manual, 1988.

134. Guy, C. T., Cardiff, R. D., and Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease, Molecular & Cellular Biology. *12:* 954-61, 1992.

135. Zotter, S., Lossnitzer, A., Kunze, K. D., Muller, M., Hilkens, J., Hilgers, J., and Hageman, P. Epithelial markers for paraffin-embedded human tissues. Immunohistochemistry with monoclonal antibodies against milk fat globule antigens, Virchows Archiv - A, Pathological Anatomy & Histopathology. *406:* 237-51, 1985.

136. Hagen, G., Muller, S., Beato, M., and Suske, G. Sp1-mediated transcription activiation is repressed by Sp3., EMBO Journal. *13:* 3843-3851, 1994.

137. Kingsley, C. and Winoto, A. Cloning of GT box binding proteins: A novel Sp1 multigene family regulating T-cell receptor gene expression, Molecular and Cellular Biology. *12:* 4251-4261, 1992.

138. Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain, Cell. *51:* 1079-90, 1987.

139. Moshier, J. A., Skunca, M., Wu, W., Boppana, S. M., Rauscher III, F. J., and Dosescu, J. Regulation of ornithine decarboxylase gene expression by the Wilms' tumour suppressor WT1., Nucleic Acids Research. *24:* 1149-1157, 1996.

140. Coppes, M. J., Campbell, C. E., and Williams, B. R. G. The role of WT1 in Wilms' tumorigenesis., FASEB Journal. *7:* 886-895, 1993.

141. Ding, H., Descheemaeker, K., Marynen, P., Nelles, L., Carvalho, T., Carmo-Fonseca, M., Collen, D., and Belayew, A. Characterization of a helicase-like transcription factor involved in the expression of the human plasminogen activator., DNA and Cell Biology. *15:* 429-442, 1996.

142. Sirito, M., Lin, Q., Maity, T., and Sawadogo, M. Ubiquitious expression of the 43- and 44- kDa forms of transcription factor USF in mammalian cells, Nucleic Acids Research. *22:* 427-433, 1994.

143. Kadonaga, J. T., Jones, K. A., and Tjian, R. Promoter-speicic activation of RNA polymerase II transcription by Sp1., TIBS. *11:* 20-23, 1986.

144. Abe, M., Siddiqui, J., and Kufe, D. Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated antigen gene, Biochemical & Biophysical Research Communications. *165:* 644-9, 1989.

145. Hilgers, J., von Mensdorff-Pouilly, S., Verstraeten, A. A., and Kenemans, P. Quantitation of polymorphic epithelial mucin: a challenge for biochemists and immunologists. [Review] [49 refs], Scandinavian Journal of Clinical & Laboratory Investigation - Supplement. *221:* 81-6, 1995.

146. Taylor-Papadimitriou, J., Stampfer, M., Bartek, J., Lewis, A., Boshell, M., Lane, E. B., and Leigh, I. M. Keratin expression in human mammary epithelial cells

cultured from normal and malignant tissue: relation to in vivo phenotypes and influence of medium, Journal of Cell Science. *94:* 403-13, 1989.

147. Travers, A. DNA-Protein Interactions., 1993.

148. Prowse, D. M., Bolgan, L., Molnar, A., and Dotto, G. P. Involvement of the Sp3 transcription facotor in induction of p21 (Cip1/WAF1) in keratinocyte differentiation., Journal of Biological Chemistry. *272:* 1308-1314, 1997.

149. Gill, G., Pascal, E., Tseng, Z. H., and Tjian, R. A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation, Proceedings of the National Academy of Sciences of the United States of America. *91:* 192-6, 1994.

150. Banks, E. B., Crish, J. F., and Eckert, R. L. Transcription factor Sp1 activates involucrin promoter activity in non-epithelial cells., Biochemical Journal. *337:* 507-512, 1999.

151. Apt, D., Watts, R. M., Suske, G., and Bernard, H. U. High Sp1/Sp3 ratios in epithelial cells during epithelial differentiation and cellular transformation correlate with the activation of the HPV-16 promoter., Virology. *224:* 281-291, 1996.

152. Husmann, M., Jehnichen, P., Jahn, B., Schlosshan, D., Romahn, E., and Marx, J. A novel Sp1 site in the human interleukin-1 beta promoter confers preferential transcriptional activity in keratinocytes., European Journal of Immunology. *26:* 3008-3014, 1996.

153. Lee, J.-H., Jang, S.-I., Yang, J.-M., Markova, N. G., and Steinert, P. M. The proximal promoter of the human Transglutaminase 3 gene, Journal of Biological Chemistry. *271:* 4561-4568, 1996.

154. Ye, J., Xu, R. H., Taylor-Papadimitriou, J., and Pitha, P. M. Sp1 binding plays a critical role in Erb-B2- and v-ras-mediated downregulation of alpha2-integrin expression in human mammary epithelial cells, Molecular & Cellular Biology. *16:* 6178-89, 1996.

155. Prince, V. E. and Rigby, P. W. Derivatives of Moloney Murine Sarcoma virus capable of being transcribed in Embryonal Carcinoma Stem Cells have gained a functional Sp1 binding site., Journal of Virology. *65:* 1803-1811, 1991.

156. Cohen, H. T., Bossone, S. A., Zhu, G., McDonald, G. A., and Sukhatme, V. P. Sp1 is a critical regulator of the Wilms' tumor-1 gene, Journal of Biological Chemistry. *272:* 2901-13, 1997.

157. Chen, H. M., Pahl, H. L., Scheibe, R. J., Zhang, D. E., and Tenen, D. G. The Sp1 transcription factor binds the CD11b promoter specifically in myeloid cells in vivo and is essential for myeloid-specific promoter activity, Journal of Biological Chemistry. *268:* 8230-9, 1993.

158. Zhang, D. E., Hetherington, C. J., Gonzalez, D. A., Chen, H. M., and Tenen, D. G. Regulation of CD14 expression during monocytic differentiation induced with 1 alpha,25-dihydroxyvitamin D3, Journal of Immunology. *153:* 3276-84, 1994.

159. Chen, J. L., Attardi, L. D., Verrijzer, C. P., Yokomori, K., and Tjian, R. Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators, Cell. *79:* 93-105, 1994.

160. Ryu, S., Zhou, S., Ladurner, A. G., and Tjian, R. The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1, Nature. *397:* 446-50, 1999.

161. Torchia, J., Glass, C., and Rosenfeld, M. G. Co-activators and co-repressors in the integration of transcriptional responses., Current Opinion in Cell Biology. *10:* 373-383, 1998.

162. Kim, S.-J., Onwuta, U. S., Lee, Y. I., Li, R., Botchan, M. R., and Robbins, P. D. The retinoblastoma gene product regulates Sp1-mediated transcription., Molecular and Cellular Biology. *12:* 2455-2463, 1992.

163. Udvadia, A. J., Rogers, K. T., Higgins, P. D. R., Murata, Y., Martin, K. H., Humphrey, P. A., and Horowitz, J. M. Sp1 binds promoter elements regulated by the RB protein and Sp1-mediated transcription is stimulated by RB coexpression., Proceedings of the National Academy of Sciences USA. *90:* 3265-3269, 1993.

164. Udvadia, A. J., Templeton, D. J., and Horowitz, J.M. Functional interactions between the retinoblastoma (Rb) protein and Sp-family members: Superactivation by Rb requires amino acids necessary for growth suppression., Proceedings of the National Academy of Sciences USA. *92:* 3953-3957, 1995.

165. Chen, L. I., Nishinaka, T., Kwan, K., Kitabayashi, I., Yokoyama, K., Florence, Y.-H., Grunwald, S., and Chiu, R. The retinoblastoma gene product RB stimulates Sp1-mediated transcription by liberating Sp1 from a negative regulator., Molecular and Cellular Biology. *14:* 4380-4389, 1994.

166. Jensen, D. E., Rich, C. B., Terpstra, A. J., Farmer, S. R., and Foster, J. A. Transcriptional regulation of the elastin gene by insulin-like growth factor-1 involves disruption of Sp1 binding., Journal of Biological Chemistry. *270:* 6555-6563, 1995.

167. Lopez-Rodriguez, C., Chen, H. M., Tenen, D. G., and Corbi, A. L. Identification of Sp1-binding sites in the CD11c (p150,95 alpha) and CD11a (LFA-1 alpha) integrin subunit promoters and their involvement in the tissue-specific expression of CD11c, European Journal of Immunology. *25:* 3496-503, 1995.

168. Murata, Y., Kim, H. G., Rogers, K. T., Udvadia, A. J., and Horowitz, J. M. Negative regulation of Sp1 transactivation is correlated with the binding of cellular proteins to the amino-terminous of the Sp1 transactivation domain, Journal of Biological Chemistry. *269:* 20674-20681, 1994.

169. Alemany, J., Borras, T., and DePablo, F. Transcriptional stimulation of the delta-1 -crystallin gene by insulin-like growth factor 1 and insulin requires DNA cis elements in chicken, Proceedings of the National Academy of Sciences USA. *87:* 3353-3357, 1990.

170. Sanceau, J., Kaisho, T., Hirano, T., and Wietzerbin, J. Triggering of the Human Interleukin-6 Gene by Interferon-gamma and Tumour Necrosis Factor-alpha in monocytic cells involves co-operation between Interferon regulatory factor-1 and NFKB and Sp1 transcription factors., Journal of Biological Chemistry. *270:* 27920-27931, 1995.

171. Hamanaka, R., Kohno, K., Seguchi, T., Okamura, K., Morimoto, A., Ono, M., Ogata, J., and Kuwano, M. Induction of Low Density Lipoprotein Receptor and a transcription factor Sp1 by Tumour Necrosis Factor in Human Microvascular endothelial cells, 267 13160-13165, 1992.

172. Strayer, D. S. Effect of virus infection on levels of transcripts for cellular genes, Viral Research. *30:* 259-269, 1993.

173. Birnbaum, M. J., van Wijnen, A. J., Odgren, P. R., Last, T. J., Suske, G., Stein, G. S., and Stein, J. L. Sp1 transactivation of cell cycle regulated promoters is selectively repressed by Sp3, Biochemistry. *34:* 16503-16508, 1995.

174. Majello, B., De Luca, P., and Lania, L. Sp3 is a bifunctional transcription regulator with modular independent activation and repression domains., Journal of Biological Chemistry. *272:* 4021-4026, 1997.

175. De Luca, P., Majello, B., and Lania, L. Sp3 represses transcription when tethered to promoter DNA or targeted to promoter proximal RNA., Journal of Biological Chemistry. *271:* 8533-8536, 1996.

176. Jones, K. A., Yamamoto, K. R., and Tjian, R. Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro, Cell. *42:* 559-72, 1985.

177. Courey, A. J., Holtzman, D. A., Jackson, S. P., and Tjian, R. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1, Cell. *59:* 827-36, 1989.

178. Mastrangelo, I. A., Courey, A. J., Wall, J. S., Jackson, S. P., and Hough, P. V. C. DNA looping and Sp1multimer links: A mechanism for transcriptional synergism and enhancement., Proceedngs of the National Academy of Sciences USA. *8:* 5670-5674, 1991.

179. Kennet, S. B., Udivadia, A. J., and Horowitz, J. M. Sp3 encodes multiple proteins that differ in their capacity to stimulate or repress transcription., Nucleic Acids Research. *25:* 3110-3117, 1997.

180. Merchant, J. L., Iyer, G. R., Taylor, B. R., Kitchen, J. R., Mortensen, E. R., Wang, Z., Flintoft, R. J., Michel, J. B., and Bassel-Duby, R. ZBP-89, a Kruppel-like zinc finger protein, inhibits epidermal growth factor induction of the gastrin promoter., Molecular and Cellular Biology. *16:* 6644-6653, 1996.

181. Thiagalingam, A., Bustros, A., Borges, M., Jasti, R., Compton, D., Diamond, L., Mabry, M., Ball, D. W., Baylin, S. B., and Nelkin, B. D. RREB-1 a novel zinc finger protein, is involved in the differentiation response to Ras in human medullary thyroid Carcinoms., Molecular and Cellular Biology. *16:* 5335-5345, 1996.

182. Sogawa, K., Imataka, H., Yamasaki, Y., Kusume, H., Abe, H., and Fujoo-Kuriyama, Y. cDNA cloning and transcriptional properties of a novel GC box-binding protein, BTEB2., Nucleic Acids Research. *21:* 1527-1532, 1993.

183. Miller, I. J. and Bieker, J. J. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins, Molecular and Cellular Biology. *13:* 2776-2786, 1993.

184. Tillotson, L. G. RIN ZF, a novel zinc finger gene, encodes proteins that bind to the CACC element of the gastrin promoter, Journal of Biological Chemistry. *274:* 8123-8, 1999.

185. Borellini, F. and Glazer, R. I. Induction of Sp1-p53 DNA-binding heterocomplexes during Granulocyte/Macrophage colony-stimulating factor-dependent proliferation in human Erythroleukemia cell line TF-1., Journal of Biological Chemistry. *268:* 7923-7928, 1993.

186. Blagosklonny, M. V. Loss of function and p53 protein stabilization., Oncogene. *15:* 1889-93, 1997.

187. Noe, V., Alemany, C., Chasin, L. A., and Ciudad, C. J. Retinoblastoma protein associates with Sp1 and activates the hamster *dihydrofolate reductase* promoter., Oncogene. *16:* 1931-1938, 1998.

188. Adnane, J., Shao, Z., and Robbins, P. D. Cyclin D1 associates with the TBP-associated factor $TAF_{II}250$ to regulate Sp1-mediated transcription., Oncogene. *18:* 239-247, 1999.

189. Shao, Z. and Robbins, P. D. Differential regulation of E2F and Sp1-mediated transcription of G1 cyclins., Oncogene. *10:* 221-228, 1995.

190. Karlseder, J., Rotheneder, H., and Wintersberger, E. Interaction of Sp1 with the growth- and cell cycle-regulator transcription factor E2F, Molecular and Cellular Biology. *16:* 1659-1667, 1996.

191. Lin, S.-Y., Rhys Black, A., Kostic, D., Pajovic, S., Hoover, C. N., and Azizkhan, J. C. Cell cycle regulated association of E2F1 and Sp1 is related to their functional interaction., Molecular and cellular Biology. *16:* 1668-1675, 1996.

192. Sif, S., Capobianco, A. J., and Gilmore, T. D. The v-Rel oncoprotein increases expression from Sp1 site-containing promoters in chicken embryo fibroblasts, Oncogene. *8:* 2501-2509, 1993.

193. Miltenberger, R. J., Farnham, P. J., D.E., S., Stommel, J. M., and Cornwell, M. M. v-Raf activates transcription of growth-responsive promoters via GC-rich sequences that bind the transcriion factor Sp1, 6, 1995.

194. Saffer, J. D., Jackson, S. P., and Thurston, S. J. SV40 stimulates expression of the transacting factor Sp1 at the mRNA level., Genes and Development. *4:* 659-666, 1990.

195. Merchant, J. L., Du, M., and Todisco, A. Sp1 phosphorylation by Erk 2 stimulates DNA binding, Biochemical and Biophysical Research Communications. *254:* 454-461, 1999.

Appendices

Appendix A1

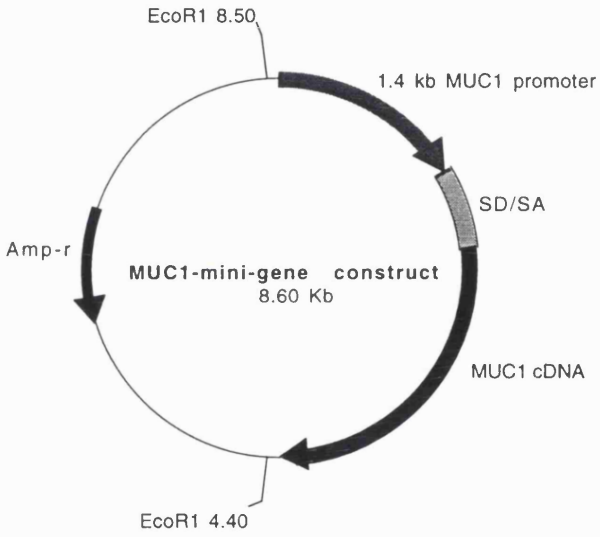


Illustration of the *MUC1*-mini-gene construct used in the generation of *MUC1*-minigene transgenic mice and in transfection experiments. Before injection into mouse oocytes the fragment containing the MUC1 promoter and MUC1 cDNA was cleaved with EcoR1 restriction enzyme.

Appendix A2

Phosphor-imager data for radio-activity per lane in the northern blot shown in figure 3.3.

Lane #	Tissue type	Poly-T	*MUC1*	Volume
1	Mammary gland	-	+	0
2	Mammary gland	-	+	421
3	Lactating mammary gland	-	+	9780
4	Tumour	+	+	4425
5	Tumour	+	+	179
6	Mammary gland	-	+	28
7	Lactating mammary gland	-	+	3283
8	Tumour	+	+	16450
9	Tumour	+	+	2621
10	Mammary gland	-	+	25
11	Lactating mammary gland	-	+	4903
12	Tumour	+	+	4778
13	Tumour	+	+	3449
14	Tumour	+	+	15140
15	Tumour	+	+	14437
16	Tumour	+	+	5516
17	Tumour	+	+	3137

Appendix 3

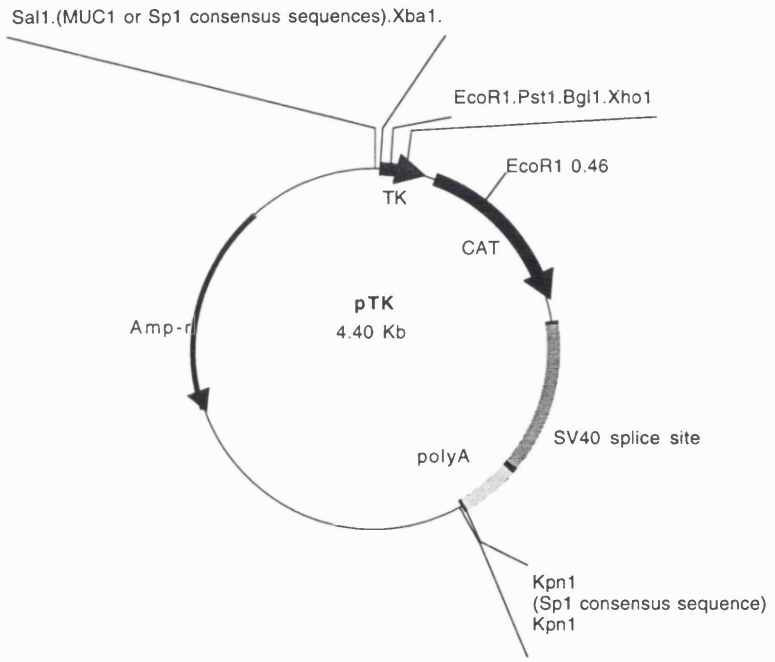


Illustration of the pTK-CAT vector showing, to scale the positions of the cloned oligonucleotides, TK promoter and CAT gene in the construct. The construct is based on the pBLCAT-5 vector with the AP1 -like sequence between the Eco0191 and Nde1 sites removed (126.127).