The Transcriptional Regulation of Maspin

A thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Biochemistry at Massey University, Palmerston North, New Zealand.

Andrew Hollings

2004

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr Kathryn Stowell who has been a fantastic supervisor. Her endless guidance, ideas, and enthusiasm for transcriptional regulation was infectious, and made working in the lab so much more enjoyable.

I would also like to thank Dr Richard Isaacs who has been a fountain of ideas for the project.

To my fellow twilight zoners, thank you for your help, friendship, and laughs. I would especially like to thank Amram, Carole, Henning, John, Kirsty, Mark, Mel, Natisha, and Robyn for all their help and support. The willingness to help and the friendly atmosphere made the twilight zone such a great environment to learn in.

Thanks also to Carolyn and Brendon from Scott Base, who generously gave me copious amounts of radiolabelled dCTP for footprinting trials.

I am indebted to my family, in more ways than one, for the encouragement and support they have given me, without which I wouldn't have made it this far. To my family and friends, you all made the last few years a great deal easier, and a lot of fun.

Finally, I would like to thank my best friend Tracey, our time apart is almost over!

ABSTRACT

Maspin (mammary serine protease inhibitor) is a tumour suppressing member of the serpin superfamily. Maspin is expressed in normal breast and prostate cells, but reportedly down regulated during progression of cancer in these tissues. Maspin has been shown to inhibit cellular migration and invasion *in vitro*; while *in vivo,* maspin has been shown to inhibit tumour growth, metastasis, and angiogenesis. Maspin also plays a role in the sensitisation of cells to induced apoptosis. These functions of maspin are independent of serine protease inhibition; however the cellular mobility function is dependent on an intact reactive site loop. Despite this knowledge, the molecular mechanisms for all reported functions of maspin are currently unknown.

Maspin is reported to be transcriptionally regulated; to date Ets, Ap1, and p53 transcription factors have been shown to activate transcription of maspin by binding directly to the promoter. Androgen is reported to be a negative regulator through the binding of the androgen receptor to a hormone response element within the promoter. This hormone response element is also responsible for an increase in maspin expression in response to tamoxifen, an anti-oestrogen drug. Transcriptional regulation of maspin has also been reported to be activated by other molecules, including gamma linolenic acid, manganese containing super-oxide dismutase, and nitric oxide, the mechanisms of regulation by these molecules is unknown. Loss of maspin expression in cancerous cells lines has been attributed to loss of one or more of the activating factors, and aberrant methylation of cytosine residues resulting in chromatin compaction.

This study investigated the transcriptional regulation of maspin, with the aim of identifying transcriptional effectors important to the regulation of the gene. Identification of such factors may help identify a pathway in which maspin exerts its tumour suppressor functions. To this end, the maspin promoter was cloned and functional assays carried out, identifying several putative regions of the maspin promoter which may be important for the regulation of the gene. To date, the precise activator/repressor binding sites and the cognate proteins responsible for this regulation are unidentified.

ABBREVIATIONS

Amp	Ampicillin
Ap1	Activating protein-1
AR	Androgen receptor
ATP	Adenosine triphosphate
bFGF	basic fibroblast growth factor
β-gal	β-galactosidase
BLAST	Basic local alignment search tool
bp	base-pairs (DNA)
BSA	Bovine serum albumin
CM	Conditioned media
CMV	Cytomegalovirus
Contig	Contiguous sequence
CpG	Cytosine-guanine di-nucleotide
DCIS	Ductal carcinoma *in situ*
DNaseI	DeoxyribonucleaseI
dNTP	deoxy-nucleotide tri-phosphate
DTT	Dithiothreitol
E4BP4	E4 Binding Protein 4
E.coli	*Escherichia coli*
ECM	Extracellular matrix
EDTA	Ethylene diamine tetra-acetic acid
EGF	Epidermal growth factor
EMSA	Electrophoresis mobility shift assay

eNOS	endothelial nitric oxide synthase
ER	Oestrogen receptor
Ets	E26 transformation-specific
FCS	Fetal calf serum
GCG	Genetics computer group sequence analysis software
GLA	Gamma linolenic acid
GR	Glucocorticoid receptor
HDAC	Histone deacetylase
HRE	Hormone response element
Hrs	Hours
iNOS	inducible nitric oxide synthase
kb	kilobase (DNA)
Klenow	*Klenow* fragment of *E.coli* DNA polymerase I
LB	Luria-Bertani
LMPCR	Ligation mediated PCR
MCS	Multiple cloning site
MeCP	Methyl-CpG binding protein
mMaspin	Mouse maspin
MnSOD	Manganese-containing super-oxide dismutase
mRNA	Messenger RNA
mt	Mutant
NO	Nitric oxide
ONPG	*o*-nitrophenyl-beta-D-galactopyranoside
ONP	*o*-nitrophenol
p53	53 kDa protein

PAI-1	Plasminogen activator inhibitor-1
PBS	Phosphate buffered saline
PC	Prostate cancer
PCR	Polymerase chain reaction
Pdef	Prostate derived Ets factor
Pen/Strep	Penicillin/streptomycin
PPAR	Peroxisome proliferator activated receptor
PR	Progesterone receptor
RNase	Ribonuclease
rpm	Revolutions per minute
RSL	Reactive site loop
sctPA	short chain tissue type plasminogen activator
SDS	Sodium dodecyl sulfate
SFM	Serum free medium
SNP	Single nucleotide polymorphisms
Sp1	Specificity protein 1
STS	Staurosporine
TAF	TBP-associated factor
TAg	SV-40 T-antigen
TBP	TATA binding protein
TESS	Transcriptional element search software
Tet	Tertacycline
TFIID	Transcription factor D for RNA polymerase II
tPA	tissue type plasminogen activator
TsAP	Thermostable alkaline phosphotase

TSS	Transcription start site
uPA	urokinase plasminogen activator
uPAR	urokinase plasminogen activator receptor
UTR	Un-translated region
UV	Ultra violet
V	Volts
VEGF	vascular epithelium growth factor
WAP	Whey acidic protein
wt	Wild type
YY1	Yin and Yang 1

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1 - Introduction

1.1 Cancer

Cells normally have a finite lifespan, and progress through development and differentiation to senescence in a regulated and defined manner. Cancer represents the situation where cells have lost this regulation and tend to proliferate in an uncontrolled manner. Cancer can be caused by multiple genetic and epigenetic changes, resulting in decreased expression or inactivation of tumour suppressor genes, and the increased expression or activation of oncogenes. The transformation from benign to malignant tumour is an important step in cancer progression. This step involves the gain of an invasive phenotype, i.e. the ability to invade neighbouring tissues. Malignant tumours can then become metastatic, characterised by the spreading of cancerous cells through the lymphatic or vascular system to other organs and tissues. Growth of tumours at distant sites follows, ultimately resulting in organ failure and death (DeVita et al., 2001).

In New Zealand, prostate cancer was the most common form of cancer diagnosed in 1999 at 98.2 diagnoses per 100,000 people per year, breast cancer followed with 87.6 diagnoses per 100,000 people per year. Breast and prostate cancer had the second and fourth highest mortality rates respectively in 1999 (NZHIS, 2002). Research world-wide is aimed at understanding the mechanisms of cancer development and progression, with the ultimate goal being the cure and prevention of cancer.

1.2 Maspin

1.2.1 Expression of maspin

Maspin (mammary serine protease inhibitor) was initially discovered by subtractive hybridisation between normal and cancerous breast cell lines, with maspin being expressed in normal breast cell lines, whilst down-regulated in breast cancer cell lines (Zou et al., 1994). Expression of maspin has since been found in cells of epithelial origin; in the mouth, airway, skin, breast, and prostate and also in testis, thymus, and small intestine (Zhang et al., 1997a; Futscher et al., 2002). In addition, expression of maspin has been reported to be progressively down-regulated as cancers of the breast and prostate progress (Zou et al., 1994; Hojo et al., 2001; Maass et al., 2001a; Maass et al., 2001c). Supporting this, the normal breast cell line MCF12a expresses maspin, whilst cancerous breast cell lines MDA-MB-231, MCF7 and T47D do not (figure 1.1) Interestingly, the prostate cancer cell lines PC3 and CA-HPV-10 both express maspin, as does the cervical cancer cell line HeLa (see figure 1.1). Hepatocyte carcinoma HepG2, lymphoblast K562, and colon cancer CaCO-2 cell lines do not express maspin (Bradbury, 2004).

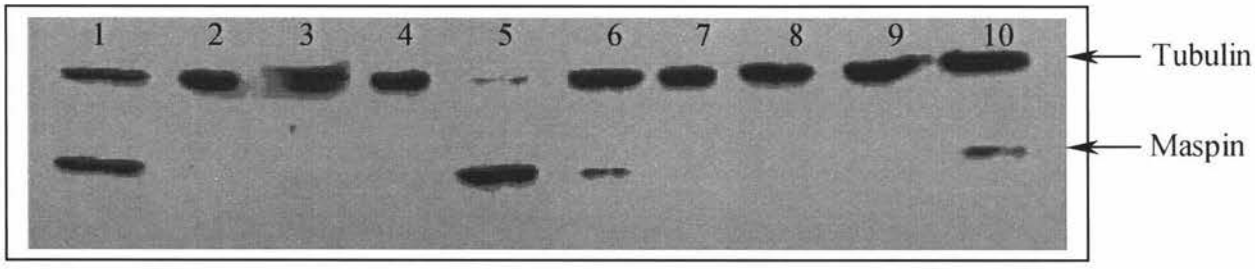


Figure 1.1. Maspin expression in human cell lines. Cells were grown to ~80% confluence and whole cell protein extracted. Protein extracts were then subjected to electrophoresis on an acrylamide gel in presence of SDS. Proteins were then transferred to nylon membrane and probed for maspin and tubulin expression. 1 = MCF12a, 2 = MCF7, 3 = MDA-MB-231, 4 = T-47D, 5 = CA-HPV-10, 6 = PC-3, 7 = HepG2, 8 = K562, 9 = CaCO-2, 10 = HeLa. Western blot courtesy of Dr Penny Bradbury (Bradbury, 2004)

1.2.2 Maspin as a tumour suppressor

Maspin has been reported to be a tumour suppressor gene, whose function is to inhibit cellular mobility, angiogenesis, and increase stress-induced apoptosis *in vitro* and *in vivo* (Zou et al., 1994; Pemberton et al., 1995; Sheng et al., 1996; Sternlicht et al., 1997; Seftor et al., 1998; Zhang et al., 1999; Zhang et al., 2000b; Zou et al., 2000; Jiang et al., 2002; Shi et al., 2002). Inhibition of cellular mobility can be affected by the addition of recombinant maspin to cultured cells, while apoptosis requires endogenously expressed maspin (Sheng et al., 1994; Zou et al., 1994; Sheng et al., 1996; Jiang et al., 2002).

Myoepithelial cells of the breast are considered to be a protective cell layer (Sternlicht et al., 1997; Reis-Filho et al., 2002). These cells surround the ductal cell layer (figure 1.2) and secrete high levels of many serpins, including maspin, which may act as paracrine inhibitors of tumour growth (Sternlicht et al., 1997).

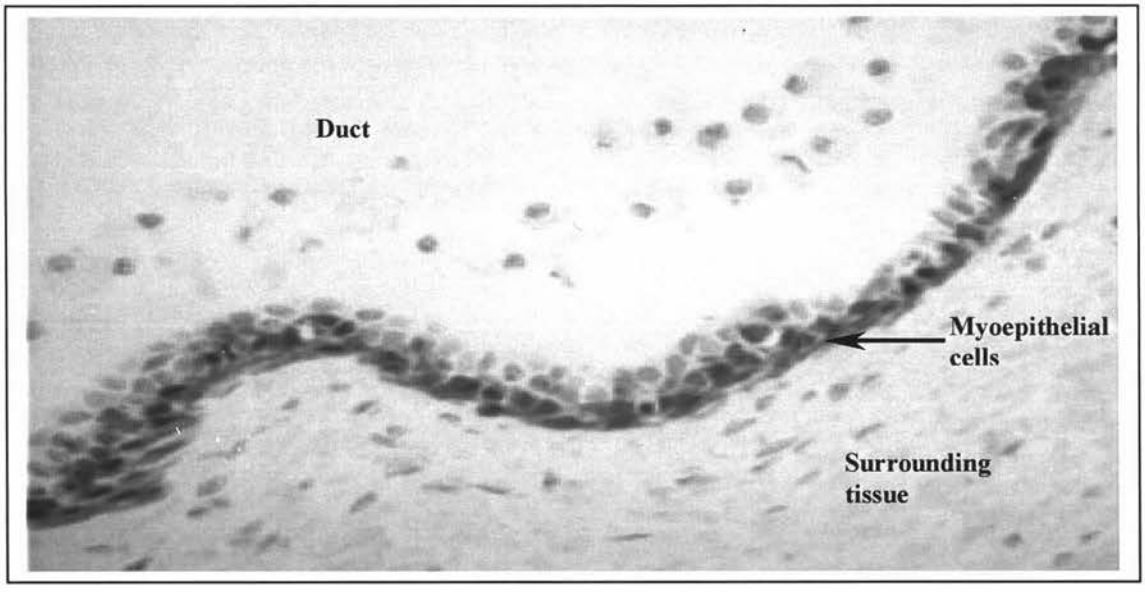


Figure 1.2. Myoepithelial cells of the breast express maspin. The myoepithelial cell layer expresses high levels of maspin as shown by brown staining with anti-maspin antibody. The myoepithelial cell layer separates the ducts of the breast from the surrounding tissue. Figure courtesy of Dr Penny Bradbury (Bradbury, 2004).

Maspin is mainly present in the cytosol, but also contains an N-terminal signal sequence which causes a small amount to be packaged into microsomes and secreted, where it acts to inhibit migration and invasion (Zou et al., 1994; Sheng et al., 1996; Pemberton et al., 1997; Sternlicht et al., 1997). Although controversial, nuclear localisation of maspin has also been demonstrated using immunohistochemistry and sub-cellular fractionation. Interestingly some tumour cells showed only nuclear localisation of maspin (Maass et al., 2002b; Odero-Marah et al., 2002; Reis-Filho et al., 2002; Sood et al., 2002; Khalkhali-Ellis and Hendrix, 2003; Kim et al., 2003; Mohsin et al., 2003; Khalkhali-Ellis et al., 2004). To date, the function of nuclear maspin is unknown.

Mouse and rat homologues of human maspin have been isolated, and at the amino acid level show 89% and 88% identity respectively to human maspin (Umekita et al., 1997; Zhang et al., 1997c). Mouse, rat, and human maspin also show similar expression patterns and are down-regulated in cancerous cells (Umekita et al., 1997; Zhang et al., 1997c). Mouse maspin (mMaspin) was shown to exhibit similar characteristics as human maspin: reactive site loop-dependent inhibition of cellular migration and invasion *in vitro* (discussed in 1.3.2) (Zhang et al., 1997c). Rat maspin, however did not show tumour suppressive properties *in vivo* (Umekita et al., 1997).

1.3 Function of maspin

1.3.1 Protease inhibition by maspin

As suggested by its name, maspin shares a high level of identity to the serine protease inhibitor (serpin) super-family (Hopkins and Whisstock, 1994; Zou et al., 1994). Serpins act by presenting the highly conserved reactive site loop (RSL) to the target protease, which recognises and cleaves the RSL. This cleavage results in the protease and serpin becoming trapped in a 1:1 heat and detergent resistant, covalently bound complex (Bass et al., 2002). The P_1 residue of the RSL determines specificity of the serpin for the type of serine protease. Maspin contains an arginine at this residue indicating that it is an inhibitor of trypsin-like proteases (Zou et al., 1994).

It has been suggested that maspin is not an inhibitory serpin i.e. does not inhibit serine proteases, as the hinge region of the RSL is highly divergent from known inhibitory serpins. The RSL also contains a deletion, leaving maspin with the shortest RSL of all serpins (Hopkins and Whisstock, 1994; Pemberton et al., 1995; Fitzpatrick et al., 1996; Bass et al., 2002). Maspin was shown to be unable to go through the stressed to relaxed transition, which is characteristic of inhibitory serpins, also indicating that maspin is a non-inhibitory serpin (Pemberton et al., 1995). Finally, maspin does not inhibit a range of trypsin-like proteases, in fact maspin acts as a substrate for these proteases (Pemberton et al., 1995; Bass et al., 2002).

Despite the weight of evidence that maspin is a non-inhibitory serpin, controversy still remains. Maspin was demonstrated to inhibit urokinase-plasminogen activator (uPA) and tissue-type plasminogen activator (tPA). uPA and tPA convert plasminogen into active plasmin, which in turn degrades the extracellular matrix (ECM) (figure 1.3). Thus uPA and tPA cause an increase in cellular mobility (Sheng et al., 1998; Biliran and Sheng, 2001). Interestingly, maspin activated the activity of tPA when a co-factor (fibrinogen or poly-lysine) was not bound however, maspin acted as a competitive inhibitor when a co-factor was bound to tPA (Sheng et al., 1998).

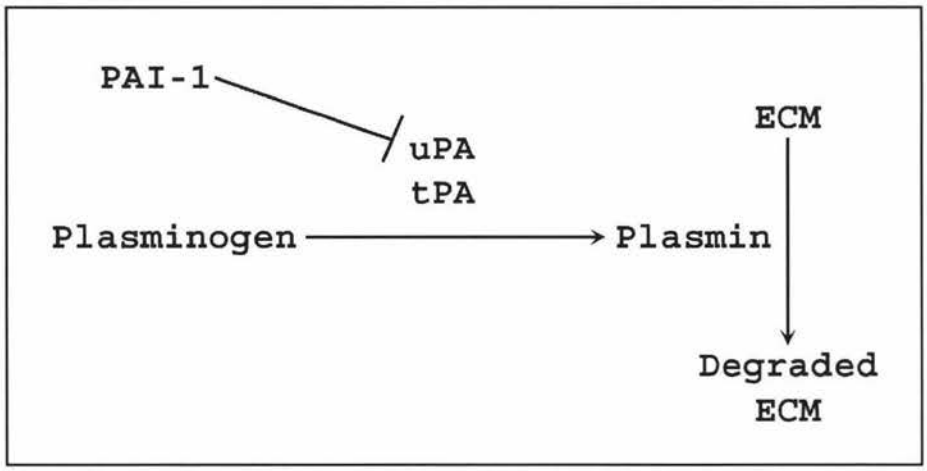


Figure 1.3. Schematic representation of plasmin activation pathway. Plasminogen is cleaved into active plasmin by uPA and tPA, which are inhibited by the serpin PAI-1. Active plasmin causes the degradation of the extracellular matrix, and thus an increase in cellular mobility.

uPA was also shown to be inhibited only when associated with the cell surface. The levels of uPA and uPA Receptor (uPAR) were shown to be decreased in cells expressing maspin. The authors suggested that this decrease was due to increased endocytosis and degradation of a maspin-uPA-uPAR complex (Sheng et al., 1998; Biliran and Sheng, 2001). The inhibition of uPA and tPA dependent proteolysis of plasminogen by maspin may decrease the amount of active plasmin, thus resulting in a decrease in ECM degradation and cellular migration (Sheng et al., 1998; Biliran and Sheng, 2001). uPAR is also known to modify the integrin profile of a cell, which may also play a role in inhibition of cellular mobility (discussed in 1.3.2) (Biliran and Sheng, 2001).

The inhibition of uPA and tPA, and the decrease in uPA and uPAR may explain how maspin inhibits cellular migration and invasion, however several publications dispute these findings (Zhang et al., 1999; Bass et al., 2002). Bass, et al., (2002) demonstrated that maspin does not inhibit cellular mobility by inhibiting proteolysis, and that maspin is a non-inhibitory serpin. Under a wide range of conditions, in which plasminogen activator inhibitor-1 (PAI-1) (figure 1.3), a serpin with a high level of identity to maspin, is an effective inhibitor, maspin showed no detectable inhibition of proteolysis (Bass et al., 2002). PAI-1 inhibits both cell-surface bound and free uPA with similar kinetics, indicating that the cell-surface is not required for the inhibitory activity of this serpin. Maspin then, would also be expected to show little difference between bound

and non-bound inhibition (Bass et al., 2002). Also, the competitive inhibition shown by Sheng, et al., (1998) is not a recognised mechanism of serpin action (Bass et al., 2002)

Taken together the data indicate that maspin is not an inhibitory serpin, and that the inhibition of migration and invasion is not dependent on protease inhibitory activity.

1.3.2 Cellular mobility

For a cell to gain the ability to invade, it must modify cell-cell and cell-ECM adhesions (DeVita et al., 2001). Proteolysis of the ECM and basement membrane is important for the invasion of tumour cells into surrounding tissues. This is carried out by many different proteases including metalloproteases and serine proteases (Seftor et al., 1998; DeVita et al., 2001; Liotta and Kohn, 2001). Modification of the integrin profile of a cell, which mediates the interactions with the ECM, changes the ability of the cell to interact with and perceive its environment; thus affecting signal transduction, and can lead to the acquisition of an invasive phenotype (Seftor et al., 1998; DeVita et al., 2001).

Maspin has been shown to inhibit the migration of cancerous cells across a membrane, and invasion through a membrane *in vitro* (Sheng et al., 1994; Zou et al., 1994; Sheng et al., 1996). Both endogenously expressed, and purified recombinant maspin have been shown to inhibit cellular mobility. Maspin exerts this effect at the plasma membrane, and cellular mobility can be restored to normal levels by addition of maspin monoclonal antibody (Sheng et al., 1994; Zou et al., 1994; Sheng et al., 1996). An antibody to the RSL of maspin prevents maspins inhibition of migration and invasion only when pre-incubated with, or added to the migration assay concurrently with recombinant maspin. This indicates that the maspin RSL is buried, and necessary for migration inhibition (Sheng et al., 1996). Cleavage at the P_1 position of the maspin RSL (thereby removing RSL) abrogates the migration inhibitory activity of maspin (Sheng et al., 1994; Pemberton et al., 1995; Sheng et al., 1996; Zhang et al., 2000b). Thus the RSL is necessary for inhibition of cellular mobility.

Myoepithelial cell lines have been shown to inhibit invasion and migration of cancerous cells. Myoepithelial cell conditioned media (CM) or co-culture has been shown to

reduce invasion and migration of cancerous cell lines *in vitro*. Immunoprecipitation of maspin from CM abrogates cancer cell mobility inhibition (Sternlicht et al., 1997; Shao et al., 2000). This suggests that myoepithelial cells secrete maspin, which inhibits the invasion of cancerous cells in a paracrine manner.

Inhibition of cellular migration and invasion has also been shown *in vivo* (Zou et al., 1994; Zhang et al., 1999; Zhang et al., 2000a; Zhang et al., 2000b; Shi et al., 2002). Maspin expressing cancerous cells implanted into nude mice, showed reduced tumour growth, and number of metastatic sites. This effect was also seen if maspin was delivered as a recombinant protein, or as an expression vector in liposome mediated gene therapy (Zou et al., 1994; Zhang et al., 2000b; Shi et al., 2002).

Transgenic mice containing maspin under the influence of the whey acidic protein (WAP) promoter, which is expressed exclusively in mammary epithelial cells during mid-pregnancy and lactation, showed disrupted mammary gland development (Zhang et al., 1999). This was due to the inability of the cells to migrate, and increased apoptosis (Zhang et al., 1999). Mice containing the SV-40 T-antigen under the control of the WAP promoter (WAP-TAg) produce mammary tumours with 100% efficiency. These mice were crossed with WAP-maspin mice (Zhang et al., 2000a). When compared to WAP-TAg mice, WAP-TAg:WAP-maspin bi-transgenic mice showed a reduction in tumour growth, and number of lung metastases, however maspin did not affect tumour initiation (Zhang et al., 2000a). Thus endogenously expressed maspin inhibits tumour growth and migration to metastatic sites *in vivo*.

Addition of recombinant maspin to non-expressing cancerous cells has been shown to change the integrin profile of these cells (Seftor et al., 1998). Cells expressing maspin had increased levels of α_3 and α_5 containing integrins, whilst decreased levels of other integrin monomers was observed. This addition of maspin inhibited cellular migration, which was reversed by the addition of $\alpha_5\beta_1$ integrin antibody (Seftor et al., 1998). Maspin has also been demonstrated to interact with collagen type I and III (Blacque and Worrall, 2002). Binding to collagen was within a region of the maspin protein which shares homology to the collagen binding protein colligin, and was independent of the RSL (Blacque and Worrall, 2002). These reports may help explain how maspin inhibits cellular migration and invasion; however the mechanism of action is unknown.

Taken together these data suggest that maspin can inhibit cellular migration and invasion *in vitro* and *in vivo*; and can act in an autocrine and paracrine manner. The molecular mechanisms responsible for the inhibition of cellular mobility are as yet unknown; however they may involve binding to collagen and/or modification of the cellular integrin profile. Inhibition of cellular mobility by maspin is dependent on an intact RSL, however does not involve inhibition of serine proteases (discussed in 1.3.1).

1.3.3 Role of maspin in apoptosis

Cells expressing maspin have been shown to be more sensitive to stress-induced apoptosis (Jiang et al., 2002; Khalkhali-Ellis and Hendrix, 2003). In breast cancer cells, increased levels of nitric oxide (NO) induced maspin expression and an increase in apoptosis. This increase in apoptosis was partially reduced by the addition of maspin antibody (Khalkhali-Ellis and Hendrix, 2003). NO pro-apoptotic effects are believed to be due to an increase in p53, activation of the caspase cascade, and DNA fragmentation (Khalkhali-Ellis and Hendrix, 2003). Supporting these suggestions Jiang, et al., (2002) demonstrated that maspin expressing cells were more sensitive to staurosporine (STS)-induced apoptosis, due to increased caspase activation. (Jiang et al., 2002). Protein kinase C prevents caspase-dependent apoptosis, STS induces apoptosis by inhibition of this enzyme. Maspin was shown to increase levels of caspase 3 and caspase 8 activity upon STS treatment; this was associated with increased DNA fragmentation and apoptosis. Increased apoptosis was not seen in untreated maspin expressing cells, indicating that maspin did not increase apoptosis directly, but sensitised cells to apoptotic inducers (Jiang et al., 2002).

PAI-1, a close relative of maspin, is an inhibitory serpin which inhibits the plasminogen activator-dependent proteolysis of plasminogen into active plasmin. This activation of plasmin results in the increased degradation of the ECM. Thus PAI-1 is an inhibitory serpin which inhibits cellular migration and invasion. Recombinant maspin, conditioned media, and N- and C-terminal maspin:PAI-1 fusion proteins did not sensitise cells to induced apoptosis, indicating that this is a function of intracellular maspin, and dependent on an intact protein (Jiang et al., 2002). Both N- and C-terminal maspin:PAI-1 fusion proteins inhibited cellular mobility, indicating that the fusion proteins were correctly folded and functional.

An increase in apoptosis in tumour cells when expressing maspin was observed in the WAP-TAg/WAP-maspin, bi-transgenic mouse model (see Chapter 1.3.2), when compared to the WAP-TAg single transgenic mouse (Zhang et al., 2000a). Also the WAP-maspin single transgenic mice were shown to have disrupted mammary gland development during pregnancy. This was due to the inappropriate expression of maspin causing a decreased rate of migration and increased apoptosis (Zhang et al., 1999). The cause of this increased rate of apoptosis in WAP-maspin mice is unknown, however it may be due to inappropriate integrin-ECM contacts/signalling due to inhibition of migration (Zhang et al., 1999). Maspin expressing cells may be more sensitive to apoptosis caused by such inappropriate integrin interactions.

Taken together, maspin sensitises cells to stress-induced apoptosis *in vitro* and *in vivo*. Maspin increases the activation of caspase 3 and 8, resulting in increased DNA fragmentation, and cell death. This function seems to be dependent on intact intracellular maspin. The molecular mechanism of the sensitisation of cells to induced apoptosis is currently unknown.

1.3.4 Maspin and angiogenesis

Angiogenesis is the development of new blood vessels from existing capillaries (Beecken et al., 2000). Tumour growth is angiogenesis-dependent, as cells in the centre of a tumour become hypoxic due to lack of nutrients and oxygen. This hypoxia causes proliferation and apoptosis of tumour cells to be in balance (Beecken et al., 2000). Metastasis is also angiogenesis-dependent as tumours require access to vessels to be able to spread to other organs (Beecken et al., 2000).

Maspin, independently of the RSL, inhibited the ability of prostate cancer cells to migrate toward vascular epithelium growth factor (VEGF), a stimulator of angiogenesis *in vitro* (Zhang et al., 2000b). Prostate cancer cells, grafted into nude mice, were supplied with exogenous maspin and the effect on tumour growth and vascularization measured. Maspin reduced the size of tumours and the density of vascularization compared to controls (Zhang et al., 2000b). In corneal neovascularization experiments, pellets containing basic fibroblast growth factor (bFGF), an angiogenesis stimulator, and maspin (mutant and normal) were implanted into rat corneas. Maspin, and maspin

lacking the RSL inhibited the formation of vessels, whilst maspin lacking the N-terminus had no effect (Zhang et al., 2000b). Thus as bFGF acts as a paracrine angiogenesis stimulator, maspin with an intact N-terminus, independent of the RSL, acts as a paracrine angiogenesis inhibitor. This report is supported by WAP-TAg/WAP-maspin bi-transgenic mice which showed a decrease in tumour vascular density (Zhang et al., 2000a). In addition, a clinical study carried out by Hojo, et at., (2001) demonstrated that vascularization was significantly inhibited in breast tumours expressing maspin.

These reports indicate that maspin acts as an inhibitor of angiogenesis *in vivo*. Maspin can exert this effect in a paracrine manner, and is independent of the RSL. Blacque and Worrall, (2002) showed that maspin could bind collagen I and III independently of the RSL. This binding may play a role in the inhibition of angiogenesis; however its significance is yet to be demonstrated.

1.3.5 Phosphorylation of maspin

Maspin has also been shown to be phosphorylated on tyrosine residues (Odero-Marah et al., 2002). Phosphorylation of tyrosine residues can be important for enzyme activity and/or signal transduction. Maspin was shown to be phosphorylated in normal and cancerous cell lines, and was phosphorylated by the tyrosine kinase domain of the epidermal growth factor receptor (Odero-Marah et al., 2002). The potential change of function upon phosphorylation of maspin was not studied. The phosphorylation of maspin may be involved in activation or inhibition of the protein; or may be part of a signal transduction pathway.

1.4 Clinical studies

1.4.1 Prognostic value

Maspin has been reported to be present in normal breast and prostate tissue but down-regulated, or lost in cancers of these tissues (Zou et al., 1994; Hojo et al., 2001; Maass et al., 2001a; Maass et al., 2001c). This, along with the reported tumour suppressor functions of maspin suggest that the loss of maspin may contribute to the progression of cancer. Initial studies indicated that major mutations/rearrangements of the maspin gene did not contribute to the loss of maspin, indicating that maspin was down-regulated rather than mutated (Zou et al., 1994). Many clinical studies have been carried out testing the prognostic value of maspin in several different cancer types. If maspin is not mutated, and loss of maspin is important for the progression of cancer, then maspin expression in cancers may be a prognostic marker for a less aggressive cancer.

1.4.2 Breast carcinomas

Many studies have been carried out on maspin expression in breast cancers. Several studies have shown that maspin is down-regulated during breast cancer progression (Zou et al., 1994; Maass et al., 2001c). Maass, et al., (2001c) showed a stepwise decrease in maspin expression as cancers progressed from ductal carcinoma *in situ* (DCIS) to invasive carcinoma to lymph node metastases; however a subset of invasive tumours were shown to express high levels of maspin. Another study by Maass, et al., (2001a) indicated that loss of maspin correlated with a higher risk of distant metastasis and shorter disease free survival. These results were supported by Hojo, et al., (2001).

In other reports, maspin expression has been correlated with poor prognosis (Umekita et al., 2002; Umekita and Yoshida, 2003). A step wise increase from DCIS to invasive cancer was seen (Umekita and Yoshida, 2003). Maspin expression correlated with larger tumour size, higher histological grade, and shorter disease free survival, and may have been related to an aggressive phenotype (Umekita et al., 2002; Umekita and Yoshida, 2003)

Mohsin, et al., (2003) studied nuclear and cytoplasmic maspin staining, and breast cancer progression; 96% of cancers stained for maspin showed nuclear staining, whilst 35% showed cytoplasmic staining. Nuclear staining was associated with good prognosis, and presence of oestrogen and progesterone receptor (ER and PR respectively). Cytoplasmic staining was correlated with the absence of ER and PR; and poor prognosis (Mohsin et al., 2003).

Another study showed that maspin was frequently associated with a higher histological grade, similar to the subset of invasive tumours found by Maass, et al., (2001c) which showed strong maspin staining (Kim et al., 2003). The histological grade of a tumour is influenced by the percent of tubular structures; size and shape of cells (uniformity); and rate of growth as determined by number of cells undergoing mitosis. Tumours are graded from 1 to 3, a tumour with a higher grade is more aggressive, faster growing, with poorly differentiated cells (www.imaginis.com). Kim et al., (2003) investigated 192 specimens, with 13% of tumours stained having high expression of maspin. Nuclear and cytoplasmic staining was noted. Maspin staining was not of prognostic value, and differed between types of breast cancer (Kim et al., 2003).

Maspin staining in breast cancers is controversial. Different methods of staining, quantification of staining, nuclear and cytoplasmic staining, and statistical analyses may be the cause of some of these reported differences. The contradictory nature of these results suggests that more large studies need to be carried out to determine if maspin expression in breast cancers is of any prognostic value.

1.4.3 Prostate carcinomas

Several clinical studies on prostate cancers (PC) have been carried out. Machtens, et al., (2001) showed that a decrease in maspin expression correlated with the loss of cellular differentiation, increased tumour grade, and shorter disease free survival. Zou, et al., (2002) also showed that a decrease in maspin expression correlates with dedifferentiation of the tumour.

Pierson et al., (2002) demonstrated that prostate epithelial cells express maspin; however the secretory cells from which PC arises did not. Interestingly the pre-

malignant prostate tumours, and surrounding cells, were shown to express maspin (Pierson et al., 2002). Thus maspin is up-regulated in the pre-cancerous stage of prostate cancer development. The pre-malignant stage can preceed PC by up to 10 years, and Pierson, et al., (2002) suggested that this up-regulation of maspin may prolong the pre-malignant stage, and may act as a marker for potential PC development. This report also demonstrated that once past the pre-malignant stage, maspin expression inversely correlated with prostate cancer tumour grade, and was down-regulated step-wise as cancer progresses (Pierson et al., 2002).

Maspin expression in prostate cancers is less controversial than in breast cancers, with all reports to date indicating that maspin expression correlates with higher grade PC. This may be due to fewer reports on maspin expression in prostate cancers. The up-regulation of maspin in pre-malignant prostate cancers may be a cellular response to tumourigenic changes, however whether maspin up-regulation in the pre-malignant to PC transition is a direct or indirect cellular response needs further investigation.

1.4.4 Other carcinomas

Studies in cancers of other tissues have also been conducted. In oral squamous cell carcinomas, maspin expression correlated with higher survival rates, and absence of lymph node metastases (Xia et al., 2000). This report also suggested that maspin expression may be a favourable prognostic marker in this type of cancer.

High levels of maspin expression have been shown in most pancreatic cancers, whilst normal pancreatic cells do not express maspin (Maass et al., 2001b; Ohike et al., 2003; Sato et al., 2003). In tumour specimens studies by Maass, et al., (2001b), 23 out of 24 expressed maspin; while 51 out of 57 tumours studied by Ohike et al., (2003) expressed maspin. Maspin expression was correlated with low grade (grade1) pancreatic cancers (Ohike et al., 2003). Maass, et al., (2001b) showed that maspin expression increased with the malignancy grade of the tumour.

A similar trend in maspin expression was seen in ovarian cancers (Sood et al., 2002). Maspin expression was not found in normal ovary tissue, whilst maspin was present in 71% of ovarian invasive cancers, and was correlated with a higher tumour grade.

Interestingly cancers with greater than 50% nuclear staining were associated with improved patient survival (Sood et al., 2002). Three mutations in maspin were reported in two ovarian cancerous cell lines. One of these mt-maspin cell lines was transfected with wt-maspin, resulting in inhibited cellular migration, which was abrogated by addition of maspin antibody (Sood et al., 2002). This suggests that the mt-maspin expressed in this cell line is not functional in cellular mobility inhibition. This raises the possibility that maspin may be mutated in ovarian cancers. Maspin mutations have also been reported in prostate cancer cell lines (Umekita et al., 1997)

The tumour suppressive function of maspin seems to be supported by its down-regulation in prostate and oral carcinomas, while the case for maspin down-regulation is not so clear cut in breast cancers. Up-regulation of maspin in ovarian and pancreatic cancers however, does not support maspin as a tumour suppressor. Pancreatic and ovarian cells may not be responsive to maspin i.e. interacting partners may not be present in these tissues; however maspin is up-regulated in a large proportion of pancreatic and ovarian cancers, which indicates expression of maspin is an advantage to these cancers. It is possible that maspin expressed in these cancers is mutated. This mutant form may carry out a function facilitating cancer progression. Few maspin mutations have been reported to date, and so little is known about the frequency of such mutations in cancers and their effect on cancer cells.

1.5 Regulation of maspin expression

1.5.1 Regulation of RNA polymerase II genes

Eukaryotes contain three RNA polymerase enzymes, of which only RNA polymerase II transcribes hn-RNA (heterogeneous-nuclear RNA), the precursor to mRNA. RNA polymerase II cannot directly recognise the promoter of target genes, and thus requires additional factors for the recognition of a promoter and subsequent transcription of the gene. Transcription factor II D (TFIID) is essential for transcription of RNA polymerase II genes. TFIID is a multi-subunit protein, made up of TATA box-binding protein (TBP) and TBP-associated factors (TAFs). TFIID is responsible for the recognition of

the TATA box element of a promoter. The TATA box is an important element in many promoters which binds TFIID resulting in the positioning of the polymerase over the transcription start site and activation of transcription. TFIID recognises and binds to the TATA box through the TBP subunit. Other essential factors including TFIIA and TFIIB are then recruited by the DNA bound TFIID. RNA polymerase then binds; this assembled complex, which is capable of transcriptional activation, is called the basal apparatus. The binding of RNA polymerase II to the assembled transcription factor complex positions the polymerase, allowing transcription to occur, with additional factors binding to help the polymerase clear the promoter.

Despite the importance of the TATA box, many promoters do not contain this element; such promoters are termed TATA-less promoters. In TATA-less promoters, the TBP sub-unit of TFIID cannot directly bind to the DNA. In these cases TFIID (or TFIIB) is recruited by transcriptional activators (e.g. Sp1) which are bound to the promoter at their corresponding element. Protein-protein interactions between activation domains of the transcription factor, or co-activator, and TAF sub-units anchor TFIID to the promoter, allowing the basal apparatus to assemble, and leading to the transcription of the gene. Thus, the transcription of TATA-less promoters relies upon TFIID sub-unit interactions with activating transcription factors (Weaver, 1999). The maspin promoter is a TATA-less promoter (Zhang et al., 1997a).

1.5.2 Transcriptional regulation of maspin

Maspin is regulated transcriptionally, and is classed as a class II tumour suppressor gene, because it is down-regulated rather than mutated (Lee et al., 1991; Zou et al., 1994). Several different factors have been shown to bind to the maspin promoter and activate/repress maspin transcription (figure 1.3). All regulatory factor sites reported to date are located within 500 base pairs (bp) of the transcription start site (TSS). Ets, Ap1 and p53 sites have been shown to be important for the activation of transcription, whilst androgen receptor (AR) binding to the hormone response element (HRE) represses transcription (Zhang et al., 1997a; Zhang et al., 1997b; Zou et al., 2000). Methylation of CpG di-nucleotides in this region of the promoter has also been shown to inhibit trans-activation of maspin (Domann et al., 2000; Futscher et al., 2002; Maass et al., 2002a; Oshiro et al., 2003; Sato et al., 2003).

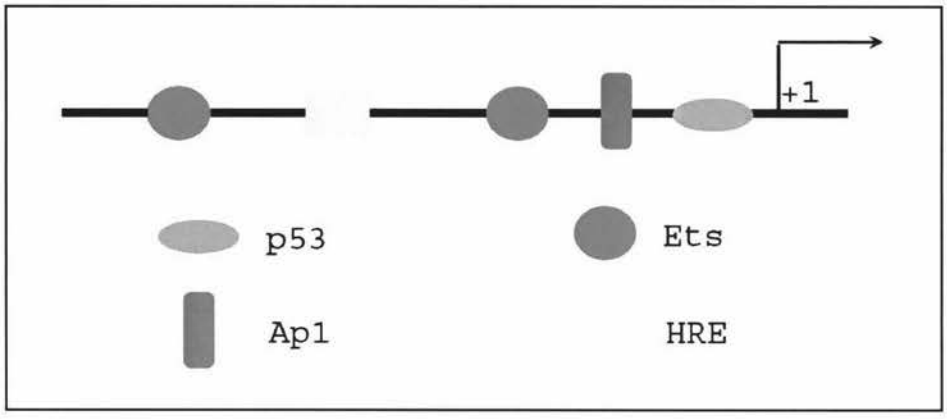


Figure 1.3. Reported regulatory elements contained in the maspin promoter. All sites are contained within 500 bp of the transcription start site. Thick black line indicates DNA, bent arrow indicates TSS (Zhang et al., 1997a; Zhang et al., 1997b; Zou et al., 2000).

1.5.3 Ets transcription factor

Two binding sites for the Ets family of transcription factors have been reported within 500 bp of the TSS (Zhang et al., 1997a; Zhang et al., 1997b) (figure 1.3). The proximal Ets site was sufficient for transcription of the reporter gene-maspin promoter construct in normal breast and prostate cell lines (Zhang et al., 1997a; Zhang et al., 1997b). Transient transfections in cancerous breast cell lines showed decreased trans-activation by the Ets site. This suggests that the decrease in maspin expression in breast cancer cells may be due in part to decreased transactivation by the Ets factors.

Ets (E26 transformation-specific) is a family of proteins which are known to play a role in apoptosis, cell proliferation, and differentiation. About 30 members of the Ets family have been identified to date, all containing a highly conserved, approximately 85 amino acid, Ets DNA binding domain (Verger and Duterque-Coquillaud, 2002; Oikawa and Yamada, 2003). The Ets domain confers specific binding to a sequence containing the core GGA(A/T) sequence, with the surrounding sequence and binding partners defining the specificity of the Ets protein for a particular site (Verger and Duterque-Coquillaud, 2002; Oikawa and Yamada, 2003). Some Ets proteins are ubiquitously expressed, while expression of others is developmentally regulated, or tissue-specific (Oikawa and Yamada, 2003). Some Ets transcription factors are targets of the Ras/MAP kinase signalling cascade, and have been implicated in breast cancer progression (Feldman et al., 2003; Oikawa and Yamada, 2003).

Pdef (prostate derived Ets factor) was shown to activate transcription of the maspin promoter in HeLa cell line transient transfections (Feldman et al., 2003). Pdef was found to be present in prostate, colon and breast tissue, and seemed to be down-regulated in invasive breast cancer cell lines. Pdef was not present in tissues of non-epithelial origin. The activation induced by Pdef, was reduced by co-transfection with Ets-1 (Feldman et al., 2003). This demonstrates that Ets factors are competing for binding sites, and so up-regulation of an Ets factor which does not activate maspin may cause a decrease in maspin expression by competing for the Ets binding site. The mouse Pdef (mPse) has also been shown to activate expression of the human maspin promoter in transient transfections using human kidney 293T cell line (Yamada et al., 2000). This indicates Ets factors have conserved recognition sequences, and possibly function.

1.5.4 Ap1

In breast cell lines, the transcription factor Ap1 (Activating protein-1) was shown to co-operate with Ets to activate maspin-reporter expression in transient transfections (Zhang et al., 1997a). Ap1 was not sufficient to activate transcription of the reporter gene, however with a functional Ets site a large increase in expression was seen compared to the Ets site alone (Zhang et al., 1997a).

Ap1 is a family of proteins that homo or heterodimerise to activate transcription of target genes (Shaulian and Karin, 2001). Ap1 proteins play a role in cell proliferation, neoplastic transformation, and apoptosis. Formation of active dimers by differential regulation of Ap1 monomers and phosphorylation of Ap1 contribute to the regulation of target genes (Shaulian and Karin, 2001). Ap1 proteins, like Ets transcription factors, are under control of the Ras/MAP kinase signalling cascade. Ap1 proteins can also be activated by UV, growth hormones, and alkylating agents (Shaulian and Karin, 2001). The identification of the Ap1 dimer which along with Ets is responsible for co-operative activation may help identify the stimulus which results in maspin expression.

1.5.4 p53

Another positive regulator of maspin expression, present within 500 bp of TSS, is the p53 element (Zou et al., 2000). Elevated levels of maspin expression were seen after exposure of cells containing wt-p53, to DNA-damaging agents. This effect was not seen in mt-p53 cells (Zou et al., 2000). p53 bound to the maspin promoter in electrophoretic mobility shift assays (EMSA), and p53 could not activate transcription when the p53 site was mutated (Zou et al., 2000).

Oshiro, et al., (2003) showed that transfection of MDA-MB-231 breast cancer cell lines with wt-p53 increased maspin transcription by modification of histones present at the promoter. The maspin promoter in MDA-MB-231 cell lines is reported to be methylated, and in a compact chromatin structure (see 'CpG methylation' below) (Maass et al., 2002a). Transfection of p53 expression vector did not effect the methylation status of the promoter, but allowed transcription by recruiting histone acetylases and opening up chromatin structure (Oshiro et al., 2003).

These studies were supported by a report by Umekita, et al., (2002), which showed that maspin expression positively correlated with p53 expression in breast cancer patients. Interestingly, a clinical study carried out on prostate cancer patients indicated that expression of p53 was correlated with lower maspin expression (Machtens et al., 2001). This result may be partially explained by a recent report which showed that missense p53 mutants could repress expression of target genes in a dominant negative manner (Willis et al., 2004). Missense mutations are the most common type of mutation in p53 (Gasco et al., 2002), and may cause repression of maspin transcription. This may be responsible for the correlation between p53 expression and the absence of maspin in prostate cancers. Interestingly only 20% of breast cancers express a mutant p53 allele, possibly explaining why p53 correlated with maspin expression in breast cancers (Gasco et al., 2002).

p53 is known to be involved in DNA repair, apoptosis and cell cycle arrest; the regulation of maspin by p53 may also indicate that p53 is involved in angiogenesis and inhibition of metastasis (Gasco et al., 2002).

1.5.5 Hormone response element

The presence of a hormone response element (HRE) in the maspin promoter has also been shown (Zhang et al., 1997b). This element was responsive to the androgen receptor (AR) in prostate cell lines, and caused transcriptional repression of a reporter gene in transient transfections. Interestingly this repression was mediated independently of the androgen ligand (Zhang et al., 1997b). This androgen-independence may be due to hormone analogues contained within the media, as charcoal stripped foetal calf serum (FCS) and phenol red free media were not used in these experiments. Supporting this, maspin-reporter gene expression was increased, in androgen responsive prostate cancer cell lines, when cultured in charcoal stripped FCS (Zou et al., 2002). This reporter gene expression was reduced upon growth with supplemented androgen. Attempts to reproduce the EMSA experiments performed by Zhang, et al., (1997b) showed weak non-specific binding of AR to the maspin HRE (van Dijk, 2003; Bradbury, 2004).

The role of androgen in the regulation of maspin in prostate cancers was supported by androgen ablation studies. Patients who underwent androgen ablation treatment before surgery showed significantly increased maspin expression (Zou et al., 2002). Prostate cancer cells showed greater maspin expression when implanted into castrated mice compared to non-castrated (Zou et al., 2002). It has also been shown that anti-androgen treatment of breast cancer cell lines (which over-express AR compared to normal breast cells) caused an increase in maspin expression (Khalkhali-Ellis et al., 2004). Interestingly no difference in prostate, kidney, or intestinal maspin expression was seen in normal and castrated rats, indicating rat maspin is not responsive to androgen (Umekita et al., 1997).

Androgen receptor plays a role in the development and differentiation of male sex organs, and also plays a role in prostate carcinogenesis (Cato and Peterziel, 1998; Bonaccorsi et al., 2000). The unliganded androgen receptor is normally present in the cytoplasm and upon binding to the androgen hormone, the receptor-ligand complex is translocated to the nucleus where it binds to DNA as a homodimer activating or repressing transcription (Cato and Peterziel, 1998).

Tamoxifen is an anti-oestrogen drug, used in cancer treatment, which competes with oestrogens for the oestrogen receptor (ER) (Khalkhali-Ellis et al., 2004). Treatment of breast cancer cell lines expressing ER with tamoxifen decreases invasion and migration, and increases apoptosis (Khalkhali-Ellis et al., 2004). Oestrogen, progesterone, and dihydrotestosterone had no effect on maspin expression, however treatment with tamoxifen increased expression in breast cancer cells, and in myoepithelial cells secretion of maspin was increased without increasing mRNA levels (Shao et al., 2000; Khalkhali-Ellis et al., 2004). Tamoxifen was shown to activate ERβ, but not ERα, and mutation of the maspin HRE abrogated the effect of tamoxifen on maspin expression (Shao et al., 2000; Khalkhali-Ellis et al., 2004). Increased maspin expression was not seen in MCF12a normal breast cells when treated with tamoxifen; however secreted maspin was not measured (Bradbury, 2004).

These data indicate that maspin is repressed by the androgen receptor, in response to androgen, in human and mice, but not rats. This suggests that maspin may have a role in development of male sex organs in mice and humans. The activation of maspin by anti-oestrogen suggests that other hormones may also play a role in regulation of maspin.

1.5.6 Other regulatory molecules

Maspin has been shown to be activated by several other mechanisms, however as yet the mechanism of action of these activators is not known. Manganese containing super-oxide dismutase (MnSOD) is a protein which converts oxygen radicals, a by-product of metabolism, into peroxide which is then converted to water by catalase. Oxygen radicals can react with proteins, lipids, and DNA potentially causing cellular and DNA damage (Li et al., 1998; Cullen et al., 2003). Super-oxide radicals may promote cellular proliferation, differentiation, and apoptosis by changing the redox potential of the cell, thus regulating redox-sensitive signal transduction pathways (Li et al., 1998; Cullen et al., 2003).

MnSOD is lost, or down-regulated in many cancer cell lines. Re-expression of MnSOD in MCF-7 breast cancer cell line, reduced cellular migration and invasion, concomitant with maspin re-expression (Li et al., 1998). Maspin re-expression was MnSOD dose dependent (Li et al., 1998). This indicates that maspin is activated by MnSOD, however

MnSOD is localised in mitochondria (Cullen et al., 2003), and so maspin induction due to MnSOD is likely to be indirect. Thus the re-expression of MnSOD in cancer cells may restore redox potential, thus restoring redox-sensitive signalling, increasing maspin expression.

Nitric oxide (NO) has also been shown to activate maspin expression (Shao et al., 2000; Khalkhali-Ellis and Hendrix, 2003). Increasing NO in media, or expression of endothelial nitric oxide synthase (eNOS) in MCF7 breast cancer cells caused re-expression of maspin (Khalkhali-Ellis and Hendrix, 2003). Increased levels of maspin correlated with decreased migration and invasion, and increased apoptosis (Khalkhali-Ellis and Hendrix, 2003).

Interestingly, increased NO levels have been shown to cause p53 phosphorylation and acetylation in a DNA damage-like response. This corresponded to increased levels of p53 target gene trans-activation (Hofseth et al., 2003). The MCF7 cell line used in the above experiments has been shown to express wt-p53 (Vojtesek and Lane, 1993). This may indicate NO treatment results in increased maspin expression by increasing p53 trans-activation; however this remains to be proven.

Gamma linolenic acid (GLA) has also been shown to induce maspin expression (Jiang et al., 1997; Jiang et al., 2000). The action of GLA was shown to be mediated through Perixosome proliferator activated receptor γ, which was translocated into the nucleus upon GLA treatment (Jiang et al., 2000). GLA is believed to be an anti-cancer fatty acid, with animals on a pure GLA diet showing a decrease in the occurrence of cancer (Jiang et al., 2000), however GLA treatments used by Jiang, et al., (1997 & 2000) in cell culture experiments seemed to be outside physiological levels (50 – 75 μM).

These reports show that other mechanisms can up-regulate the expression of maspin, however they seem to be indirect. The exact method of maspin regulation by these modulators is yet to be demonstrated.

1.5.7 CpG methylation

Methylation on the cytosine residue of CpG dinucleotides in the promoter region of genes has been demonstrated to prevent transcriptional activation of genes (Brown and Strathdee, 2002). Addition of a methyl group to the cytosine residue can prevent the binding of some transcription factors; however this is not believed to be the main mechanism by which methylation prevents trans-activation. Methylation leads to the binding of methyl-CpG binding proteins (MeCP), which recruit histone deacetylases, therefore causing compaction of chromatin in the methylated area (Brown and Strathdee, 2002). This compaction of chromatin prevents transcription factors gaining access to their respective binding sites, resulting in loss of trans-activation.

Methylation of the maspin promoter has been shown to prevent expression of maspin in some cancerous cells, and also in cells from non-expressing tissues (Domann et al., 2000; Futscher et al., 2002; Maass et al., 2002a; Oshiro et al., 2003; Sato et al., 2003). The maspin promoter has been reported to contain two CpG islands (Domann et al., 2000). A CpG island is defined as a region of DNA that has an unusually high cytosine (C), guanine (G), and CpG di-nucleotide content (Costello and Vertino, 2002). The proximal CpG island of the maspin promoter overlaps the transcription start site and positive effector sites reported to date. This island was shown to be highly methylated in cancerous cell lines, and tissues which do not normally express maspin, whilst maspin expressing cells had little methylation in this region. The distal CpG island is located approximately 120 bp upstream of the proximal island and was shown to be methylated in both expressing and non-expressing cells (Domann et al., 2000; Futscher et al., 2002; Maass et al., 2002a; Oshiro et al., 2003; Sato et al., 2003).

In cell lines which contained methylation of the proximal island, maspin could be re-expressed by addition of a histone deacetylase inhibitor, and/or methyltransferase inhibitors (Domann et al., 2000; Costello and Vertino, 2002; Maass et al., 2002a; Oshiro et al., 2003; Sato et al., 2003). Pancreatic tissue does not express maspin, however many pancreatic cancers do. Treatment of normal pancreatic cells with methyltransferase, and histone deacetylase inhibitor caused expression of maspin in these cells. Cancerous pancreatic cells were hypomethylated, whilst normal cells were found to be hypermethyalated (Maass et al., 2001b; Ohike et al., 2003; Sato et al., 2003). p53 was

able to induce maspin expression in cells with methylated promoters by recruiting histone acetylases, which in turn caused the opening of chromatin structure, thus allowing transcription factors to bind (Oshiro et al., 2003).

These findings indicate that at least some factors required for maspin transcription are present in many, if not all, tissues and cancers. This suggests that a major loss of maspin transcription may be due to methylation of the promoter, and concomitant histone compaction. The loss of trans-activation of maspin in methylated cell lines may be controlled by the balance of MeCP induced histone acetylation and p53 induced deacetylation. Thus the status and expression levels of p53 may be important.

1.6 Chapter summary

Maspin is a tumour suppressor gene whose expression has been observed in many tissues of epithelial origin, including prostate and breast. The tumour suppressive function of maspin is related to its ability to inhibit cellular mobility and invasion, apoptosis, and angiogenesis *in vitro* and *in vivo*. Whilst some interaction partners have been identified, no mechanism is known for the reported functions of maspin. Although maspin is a non-inhibitory serpin, inhibition of cellular migration and invasion is dependent on the RSL, however angiogenesis inhibition is reliant upon the N-terminus. This indicates that maspin contains at least two functional sites.

Maspin was initially reported to be down-regulated in breast cancers. Several reports support this hypothesis, however other reports disagree. Maspin down-regulation in prostate cancer however, is less controversial. The differing results seen in clinical studies may be due to the lack of standardised methods. The sub-cellular localisation of maspin also needs further investigation, as well as the role of maspin in these compartments. The up-regulation of maspin in pancreatic and ovarian cancers seems to oppose its reported tumour suppressive function; however these cells may be non-responsive to maspin, or may have high levels of mutation. The expression of mutant maspin may be an advantage to cancers of these tissues.

Expression of maspin is positively regulated by Ets, Ap1, and p53 transcription factors. The binding sites for these factors and the negative HRE element are contained within 500 bp of the transcription start site. Ets, Ap1 and p53 all play roles in cell proliferation and apoptosis. Ets and Ap1 also regulate cell differentiation. Androgen influences cellular differentiation and development. The development of tissues involves cellular migration, apoptosis, differentiation, and proliferation. The regulation of maspin by these factors suggests that maspin may be important for the development and differentiation of tissues. This hypothesis is supported by abnormal expression of maspin during pregnancy in WAP-maspin mice causing disrupted mammary gland development. These factors also support the functional evidence that maspin plays a role in apoptosis, as all activators of maspin transcription are involved in apoptosis regulation. Ets, Ap1 and AR play roles in cellular differentiation, and in prostate cancers the loss of maspin correlated with the loss of differentiation (higher tumour grade), however the role of maspin in cellular differentiation has not been studied.

Repression of maspin transcription by methylation of the promoter may be important for tissue-specific regulation. This silencing mechanism may also be used by cells to silence maspin as cancer progresses. The methylation of the proximal CpG island in the maspin promoter results in histone deacetylation and chromatin compaction in this area. This then prevents transcription factors from activating transcription, causing the loss of the tumour suppressor functions of maspin, and thus allowing cancer progression.

The mechanisms by which MnSOD, NO, and GLA stimulate maspin expression are yet to be deciphered, however NO may activate p53, and p53 in turn activate maspin expression.

1.7 Research aims

The elements that are involved in transcriptional regulation of maspin suggest that maspin is involved in apoptosis, cellular migration, and proliferation; hypotheses which are supported by functional studies. To date however, the molecular mechanisms of the function of maspin are unknown. The identification of further transcriptional elements which are involved in the transcriptional regulation of maspin may provide insights in the search for a molecular function of maspin. For example, if a transcriptional effector, responsible for activating transcription of genes which act in a particular apoptosis pathway, is identified as activating maspin transcription, then maspin may play a role in this pathway. This hypothesis can then be tested further.

The regulatory elements that have been functionally identified to date are all contained within 500 bp of the transcriptional start site. Transient transfection of deletion series has shown that transcriptional elements may be present in regions further upstream from the TSS (Zhang et al., 1997a; Zhang et al., 1997b), however these regions have not been investigated. The reported activation of maspin expression by MnSOD, GLA, and NO also suggest that other transcriptional regulators may play a role in maspin expression.

The aim of this study was to identify putative regulatory factors which are important for the regulation of the gene, by assessing the regulation of maspin transcription in cancerous and normal cell lines. To this end, the maspin gene was identified from the human genome sequence, and a fragment representing the promoter cloned and analysed for putative transcriptional elements. Transient transfection of cell lines with a series of maspin promoter deletion constructs was the first step in functional analysis. This was followed by footprinting of regions identified as being important for the regulation of the maspin gene. Footprinting has the potential to specifically identify protein binding motifs at the nucleotide level, enabling putative protein binding DNA elements to be identified. The DNA elements can then be used to predict the transcription factor which binds to this region.

CHAPTER 2 – Materials and Methods

2.1 Materials

2.1.1 DNA polymerases

Pwo, *Taq*, *Tgo*, T4 DNA Polymerase, as well as Expand™ DNA polymerase mix were all purchased from Roche Diagnostics, Mt Wellington, Auckland. Sequenase version 2.0 was obtained from USB Corporation, whose New Zealand distributor is Amersham Biosciences, Auckland.

2.1.2 Restriction endonucleases

The restriction enzymes used in this project (*Bgl* II, *Eco* RI, *Hae* III, *Hind* III, *Kpn* I, *Nco* I, *Not* I, *Pst* I, *Sac* I, *Sca* I, *Sma* I, *Sna* BI, *Sph* I, *Xho* I) were obtained from a variety of sources: New England Biolabs, MA, USA; Roche Diagnostics, Mt Wellington, Auckland; and Amersham Biosciences, Auckland.

2.1.3 Other enzymes

Klenow fragment of *E.coli* DNA polymerase I (*Klenow*), proteinase K, DNaseI (bovine pancreas grade II), and terminal transferase were all purchased from Roche Diagnostics, Mt Wellington, Auckland. T4 DNA ligase was purchased from Invitrogen Corporation, Auckland, NZ. RNase and lysozyme was obtained from Sigma Chemical Company, MO, USA. Thermosensitive alkaline phosphatase (TsAP) was purchased from Life Technologies Gibco BRL, Auckland.

2.1.4 Cell culture

All cell lines were obtained from ATCC, VA, USA. T75 cm^2, and T25 cm^2 culture flasks were purchased from Nunc Inc. IL, USA.

All media (Keratinocyte SFM, MEM (Eagle), DMEM, and Hams F12), foetal bovine serum, horse serum, trypsin, epidermal growth factor, and penicillin/streptomycin were purchased from Invitrogen Corporation, Auckland. FUGENE™ and bovine insulin were both purchased from Roche Diagnostics, Mt Wellington, Auckland. Hydrocortisone, soybean trypsin inhibitor, and DMSO were purchased from Sigma Chemical Company, MO, USA.

2.1.5 Footprinting reagents

Ready-to-go™ DNA labelling beads and Probe Quant™ G-50 micro columns were purchased from Amersham Biosciences, Auckland. Dimethyldichlorosilane solution, and acrylamide (Acrylogel 3 solution) 29.1 : 0.9 ratio were purchased from BDH Laboratory Supplies, Poole, England. Diagnostic X-ray film was made by Eastman Kodak Company, NY, USA. *E.coli* tRNA was purchased from Sigma Chemical Company, MO, USA.

Streptavidin M280 beads were obtained from Dynal Biotech, Victoria, Australia. Positively Charged Nylon Membrane was purchased from Roche Diagnostics, Auckland. 3MM paper was purchased from Whatmann, Maidstone, England. α-dCTP was purchased from PerkinElmer, MA, USA. Bovine Serum Albumin fraction V (BSA) was purchased from Gibco Laboratories, NY, USA. Wizard Genomic DNA Purification Kit for purification of 'naked' DNA was purchased from Promega Corporation, WI, USA. Sequenase version 2.0 DNA sequencing kit was purchased from USB Corporation, through Amersham Biosciences, Auckland.

2.1.6 Vectors

pGL3 Basic, and pSV-βgal vectors were purchased from Promega Corporation, WI, USA. pCMV-Sport-βgal was purchased from Invitrogen Corporation, Auckland. pBlue Script was purchased from Stratagene, LA, USA

2.1.7 *E.coli* strains

XL-1 Blue strain of *E.coli* was obtained from Stratagene, LA, USA. XL1-Blue strain has the genotype F'::Tn*10 pro*A$^+$B$^+$ *lacI*^{q} Δ(*lacZ*)*M15/recA1 endA1 gyrA96* (Nal^r) *thi hsdR17* (r_K^- m_K^+) *glnV44 relA1 lac* (NEB catalogue).

The DH5α strain of *E.coli* was purchased from Life Technologies Gibco BRL, Auckland. DH5α strain of *E.coli* has the genotype F'/*endA1 hsdR17*(r_K^- m_K^+) *glnV44thi-1 recA1 gyrA* (Nal^r) *relA* Δ(*lacIZYA-argF*)*U169 deoR* (Φ80*dlac*Δ(*lacZ*)*M15*) (NEB catalogue)

2.1.8 Other specialist materials

Acrocap Filter Unit (0.2 μm) sterilisation filters for cell culture media, and Acrodisk Syringe Filters (0.45 μm and 0.2 μm) for general solutions were purchased from Pall Corporation, MI, USA. QIAquick PCR purification and Maxi Prep kits were purchased from Qiagen, Victoria, Australia. DNA primers (including biotinylated) were purchased from Sigma Genosys Australia Pty. Ltd., NSW, Australia.

dNTPs and tris-equilibrated phenol were purchased from Amersham Biosciences, Auckland. Agarose, and 1kb Plus DNA Ladder were purchased from Invitrogen Corporation, Auckland. Ampicillin, tetracycline, PEG-4000, Luria Bertani (LB) base, and TEMED (N,N,N',N'-Tetramethylethylenediamine) were purchased from Sigma Chemical Company, MO, USA.

NuSeive agarose was obtained from BioWhittaker Molecular Applications, ME, USA. Luciferase Reagent was purchased from Promega Corporation, WI, USA. 5-azacytidine was purchased from Fluka, Switzerland. Bacteriological agar was purchased from Oxoid LTD, Hampshire, England. ONPG (o-Nitrophenyl-β-D-Galactopyranoside) was purchased from Roche Diagnostics, Auckland. Parafilm M® was purchased from Pechiney Plastic Packaging, WI, USA.

All other reagents were of analytical grade or better.

2.2 DNA modification and purification

2.2.1 PCR

All PCR was carried out in 0.2 mL tubes on the GeneAmp 2700 PCR system (Applied Biosystems). In general three steps were used in each cycle: denaturing, annealing, and extension. At the end of the cycles, a 5 minute incubation at the extension temperature was added to ensure all products were completely extended.

2.2.2 Restriction endonuclease digests

Restriction digests were carried out in 1 x digestion buffer (as recommended by manufacturer of individual enzymes) in a 30 μL total volume. For complete digestion, 1 unit of enzyme was used per 1 μg of DNA, and samples incubated at the optimal temperature as recommended by manufacturer for 1 hour.

2.2.3 Creation of blunt ends

DNA digested with restriction enzymes which left 3' overhangs were end filled with *Klenow* enzyme. In these cases, 1 μg of digested DNA in 1 x digestion buffer was incubated with 2 nmol mixed dNTPs in the presence of 2 units of *Klenow* enzyme for 15 minutes at 30°C. The enzyme was heat denatured at 75°C for 15 minutes and the required DNA fragment gel purified.

2.2.4 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out using Mini Sub™ DNA Cell, or Wide Mini Sub™ Cell (BioRad) gel boxes, in 1 x TAE (0.04 M Tris-Acetate, 1 mM EDTA pH 8.3). Agarose was dissolved in 1 x TAE, and 2 μL of ethidium bromide solution (10 mg/mL) per 50 mL of gel was added once cooled. Agarose was then poured into a gel tray with combs in place and left to set. DNA loading dye (40% (w/v) sucrose, 0.25% (w/v) bromophenol blue) was added to samples which were applied to the wells of gel submerged in 1 x TAE. Current was applied to the gel, with DNA migrating toward

positive terminal. When the bromophenol dye band was at the bottom of the gel, the current was removed, and DNA visualised under UV light.

2.2.4 Removal of phosphate groups

Phosphate groups on the 5' end of 1 μg of digested DNA were removed by incubation with 0.2 units of thermosensitive alkaline phosphatase (TsAP), in 1 x TsAP buffer, at 65°C for 15 minutes. The reaction was stopped by the addition of 4 μL of stop buffer (200 mM EDTA, pH 8.0) and incubation at 65°C for a further 15 minutes.

2.2.5 Gel extraction

DNA fragments were purified from gels using the freeze-squeeze method (Thuring et al., 1975). The required band was cut out of an agarose gel, wrapped in parafilm and placed at -20°C for at least 1 hour. The frozen agarose fragment was then squeezed and the liquid collected in 1.5 mL micro-centrifuge tubes. To increase the yield of purified DNA, 50 μL of 1 x TAE buffer was added to the agarose and re-frozen. The agarose was then re-squeezed, and the liquid pooled with that from the first squeeze. DNA was then purified from the collected liquid by ethanol precipitation.

2.2.6 Column purification of DNA

DNA was purified from PCR and other enzymatic procedures using the QIAquick PCR purification kit (Qiagen). This kit relies upon the ability of DNA to bind to a silica column in the presence of high salt, and pH of less than 7.5. The samples were mixed with buffer to provide the optimum conditions for binding, and applied to the column. DNA binds to the silica, whilst contaminants pass through. The column was then washed with ethanol containing buffer, and DNA eluted in low salt buffer. Manufacturers' directions were followed and resulting samples were gel quantified.

2.2.7 Phenol chloroform extraction

Phenol chloroform DNA purifications were carried out using 1 sample volume of tris-equilibrated phenol and 1 volume chloroform. The samples were mixed thoroughly and centrifuged at 13,000 rpm for 5 minutes in a micro-centrifuge. The aqueous layer was removed, and 1 volume of chloroform was added, mixed, and re-centrifuged for a further 5 minutes. The aqueous layer was again collected and the DNA was ethanol precipitated.

2.2.8 Ethanol precipitation

DNA was precipitated from samples by adding 1/10 sample volume of 3 M sodium acetate pH 5.2, and 3 volumes of 100% ethanol. Samples were then incubated at -20°C for at least 1 hour, followed by centrifugation at 13,000 rpm for 15 minutes at 4°C. The DNA pellet was then washed in 70% ethanol and centrifuged at 13,000 rpm for 5 minutes at 4°C. The pellet was dried upside down, and resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0).

2.2.9 Plasmid DNA Maxi Prep

Large amounts of pure plasmid DNA was extracted using the Maxi Prep Kit (Qiagen). *E.coli* cells containing the plasmid of interest were grown overnight with shaking, in 500 mL LB (Luria-Bertani) broth containing the selectable marker. The culture was divided into two GS3 centrifuge bottles, and centrifuged at 2,000 rpm for 5 minutes at 4°C. The culture medium was removed, and the cell pellet was then used for plasmid DNA extraction. The Maxi Prep Kit relies on an alkaline lysis procedure, where cells were lysed, releasing contents, with genomic DNA and cellular debris being collected in the pellet. Plasmid DNA was then extracted using an anion-exchange resin, to which DNA binds under low-salt conditions. After washing the resin, DNA was eluted by high-salt buffer. Salt was removed by isopropanol precipitation, followed by ethanol washes. The Maxi Prep was carried out following manufacturers' directions, and the resulting DNA resuspended in 300 µL TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0).

2.2.10 DNA quantification

DNA extracted using nuclei isolation and DNaseI digestion; Wizard Prep Kit (see footprinting); or Maxi Prep Kit was quantified using an Ultrospec 300 Spectrophotometer (Pharmacia Biotech). A 100 fold dilution of the sample DNA was made in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0), and the absorbance of the sample measured using quartz curvettes. The spectrophotometer measured the absorbance of the samples at 260, 280 and 320 nm. The ratio of 260/280 indicates the purity of the DNA – 1.8 being pure DNA; whilst a lower ratio indicates protein contamination, and a higher ratio indicates RNA contamination. The A_{260nm} measures concentration; a DNA solution with $A_{260nm} = 1.0$ has a concentration of 50 μg/mL. A 'background' measurement for solvent contamination was taken at 320 nm, and was subtracted from the 260 nm reading before multiplying by 50 μg/mL to give the final concentration.

DNA which was purified using freeze squeeze or micro-centrifuge column based purification kits was gel quantified. Samples were analysed by agarose gel electrophoresis with 100, 50, 20, and 10 ng DNA standards. The intensity of the band of interest was then compared visually to that of the quantification standards.

2.2.11 DNA sequencing

DNA to be sequenced was pre-mixed with 3.2 pmol of primer in a 15 μL total volume. For sequencing of PCR products 1 ng of DNA was used per 100 bp. For plasmids 300 ng of DNA was used. Pre-mixed DNA and primer samples were then submitted to the Allan Wilson Centre Genome Service, where the DNA was sequenced using the BigDye® Terminator v3.1 Kit (Applied Biosystems) and samples analysed using the ABI3730 Genetic Analyzer (Applied Biosystems). The sequence data was edited manually and used in alignments.

2.3 Transformation of *E.coli*

Aseptic technique was used for all *E.coli* manipulations.

2.3.1 *E.coli* culture

E.coli strains were streaked from frozen stocks onto LB (Luria-Bertani) agar (1% tryptone, 0.5% yeast extract, 0.085 M NaCl, 1.5% agar) with selection antibiotic (XL-1 Blue – tetracycline, DH5α – no antibiotic, pGL3 transformants - ampicillin) and incubated overnight at 37°C. The next day single colonies were chosen, and re-streaked. From the resulting streak, 5 mL broths containing antibiotic were inoculated with single colonies and grown overnight at 37°C with shaking (220 rpm).

2.3.2 'Growth on the day' competent cells

To create cells capable of taking up ligated DNA, competent cells were prepared according to Chung et al., (1989). A fresh 5 mL LB broth containing the required antibiotic was inoculated with 100 µL of an overnight culture, and incubated with shaking until cells reached an A_{600} of 0.4-0.5. Aliquots of the culture were then pelleted in micro-centrifuge tubes, and resuspended in 1/10 volume of TSS buffer (10 g/L Tryptone, 5 g/L yeast extract, 100 g/L PEG-4000, 0.085 M NaCl, 5% DMSO, 0.05 M $MgCl_2$, pH 6.5). Resulting competent cells were stored on ice until transformation, and were viable for up to 3 hours.

2.3.3 Ligation

Ligations were carried out in 1 x DNA ligase buffer (50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 1 mM ATP, 1 M DTT, 5% polyethylene glycol), with 1 unit of T4 DNA ligase, in a total volume of 10 µL. Insert and vector ratios were varied, with no more than 100 ng of total DNA. Ligations were left at room temperature for 30 minutes, and then 5 µL used to transform cells. Remaining ligation reactions were stored in the refrigerator door overnight.

2.3.4 Transformation

Transformations were carried out by addition of 5 μL of ligation reaction to 100 μL of competent cells, and incubation on ice for 5 minutes. Transformed cells were then plated onto LB agar containing selective antibiotic. Plates were incubated overnight at 37°C. The following day, 5 mL LB broths containing the appropriate selective marker were inoculated with single colonies, and incubated overnight with shaking.

2.3.5 Rapid boil DNA preparation

To identify transformed *E.coli* cultures plasmid DNA was extracted using the rapid boil preparation (Holmes and Quigley, 1981). In a micro-centrifuge tube, 1.5 mL of an overnight *E.coli* culture was pelleted and resuspended in 350 μL of STET buffer (0.1 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 5% Triton X-100). Cells were lysed by addition of 25 μL of 10 mg/mL lysozyme and placing in boiling water for 40 seconds. Cellular debris was then pelleted by centrifugation at 12,000 rpm for 10 minutes in a micro-centrifuge. The resulting pellet was removed, and 400 μL of isopropanol added to supernatant to precipitate DNA. Samples were incubated at -20°C for 30 minutes, and DNA pelleted by centrifugation at 12,000 rpm for 5 minutes. The pellet was washed in ice cold 95% ethanol, and centrifuged for 1 minute at 12,000 rpm. The DNA pellet was then air dried and resuspended in distilled water, or TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0).

2.3.6 Preparation of glycerol stocks

Glycerol stocks were prepared by addition of 800 μL of an *E.coli* overnight culture containing the plasmid of interest and 200 μL of 100% glycerol into a cryo-tube. The sample was mixed and placed at -70°C for long term storage.

2.4 Mammalian cell culture

2.4.1 Maintenance of cell lines

All cell culture manipulations except for protein extractions were carried out in a laminar flow hood. Cells were grown under humid conditions at 37°C with 5% CO_2. Media was changed on slow growing cells every 2-3 days.

2.4.2 Media

Media was prepared by dissolving powder in milliQ water, and addition of sodium bicarbonate as directed by manufacturers. The amount of sodium bicarbonate added varied according to type of media, and was not dependent on the cell line to be grown in the media. The pH was altered to 6.8, and media filter sterilised using Acrocap Filter Units 0.2 μm (Pall). Media was stored at 4°C and supplements added to media when ready for use; table 2.1 details media and additives used for each cell line.

Cell Line	Base media	Supplements
MCF12a	DMEM:Hams F12 (1:1)	5% Horse Serum 1% Pen/Strep 20 ng/mL EGF 500 ng/mL hydrocortisone 0.01 mg/mL bovine insulin
MDA-MB-231	MEM (Eagle)	10% FCS 1% Pen/Strep
HeLa	MEM (Eagle)	10% FCS 1% Pen/Strep
CA-HPV-10	Keratinocyte SFM	1% Pen/Strep 5 ng/mL human recombinant EGF 0.05 mg/mL bovine pituitary extract
PC3	Hams F12	10% FCS 1% Pen/Strep

Table 2.1. Media and additives used for growth of cells. Media was prepared by addition of powder and sodium bicarbonate to manufacturers' directions, followed by changing the pH to approximately 6.8, and filter sterilisation into autoclaved bottles. Media was stored incomplete until ready to use. All additives were stored at -20°C, and thawed before use. All media was stored at 4°C.

2.4.3 Establishing cells from frozen stocks

Cell line stocks stored in gaseous phase of liquid nitrogen were thawed quickly, transferred to 15 mL tubes, and cells pelleted by centrifugation. Cells were resuspended in complete media, and dispensed into a flask with media covering the base. T75 flasks contained 14 mL of complete media, whilst T25 flasks contained 9 mL of media.

2.4.4 Passage of cells

Cells grown to at least 80% confluence were passaged into new flasks by removal of media, and rinsing in 1 x trypsin PBSE (1 mL trypsin (Invitrogen) in 9 mL PBSE (0.14 M NaCl, 4.3 mM $NaHPO_4$ pH 7.2, 2.7 mM KCl, 0.5 mM EDTA). The solution was then removed and the cells rinsed again. Trypsin was removed, and cells were left to detach at room temperature. The flask was hit sharply to dislodge and separate the cells, which were then resuspended in complete media. Cells were then dispensed into new flasks containing complete media. The ratio of sub-culturing depended on both the growth rate of the cell line, and when they were required to reach confluence.

CA-HPV-10 cells required the use of soybean trypsin inhibitor (Sigma). For these cells, the second rinse of trypsin was not removed. Flasks were placed at 37°C for 2-5 minutes, until cells were detached. Detached cells were then dispensed into 15 mL tubes containing trypsin inhibitor, mixed, pelleted by centrifugation, resuspended in complete media, and dispensed into flasks.

2.4.5 Freezing cells

Cells which were at least 80% confluent in a T75 flask were passaged, and resuspended in 2 mL 95% FCS, 5% DMSO. Resuspended cells were dispensed into cryo-tubes, wrapped in laboratory tissue, and placed at -70°C overnight. The next day the tubes were placed in the gaseous phase of a liquid nitrogen store for long term storage.

2.4.6 Transient transfection

Cells were passaged as normal, and resuspended in approximately 5-8 mL complete media. A drop of cell suspension was then added to each well of a 12 well plate containing 800 µL of complete media. Plates were then incubated overnight at 37°C. Cells were seeded so that by the next day each well was ~50% confluent.

Cells were transfected on the day following passage. Tubes were set up with the appropriate mix of maspin-pGL3 construct and internal control vector. A serum free media:FUGENE™ mix was made, added to the tubes containing plasmid prep of interest, and incubated at room temperature for 15 minutes. FUGENE™ is a lipid based transfection reagent, and as such is believed to form liposomes containing the DNA to be transfected. These liposomes are then believed to fuse with the cell membrane, thus transfecting the cells. Transfections were set up in all cell lines using a 3:2 ratio of FUGENE:DNA i.e. 3 µL FUGENE for every 2 µg of DNA per well. Each tube contained enough FUGENE:DNA mix to transfect triplicate wells, with 100 µL being added drop wise to each well.

Transfected cells were incubated at 37°C for 40-48 hours before harvesting. Cells were rinsed twice in 1 mL PBS (0.14 M NaCl, 4.3 mM $NaHPO_4$ pH 7.2, 2.7 mM KCl) and then lysed with 80 µL cell lysis buffer (25 mM Tris-phosphate pH 7.8, 2 mM DTT, 2 mM EDTA, 10% glycerol, 10% triton X-100) for 15 minutes at room temperature. The wells were then scraped, and lysates placed in individual wells of a cone-shaped 96 well plate. The plate was then centrifuged at 2,000 rpm for 5 minutes to pellet debris.

Luciferase assays were set up in an opaque, flat bottom, 96 well plate by adding 5 µL of cell lysate per well. The Fluostar Galaxy plate reader (BMG Technologies), on well mode, was set up to inject 20 µL of luciferase reagent (Promega), and measure the levels of luminescence every second for 140 seconds. Wells were injected and measured sequentially.

β-galactosidase assays were set up in clear, flat bottom, 96 well plates by the addition of 5 µL of cell lysate into each well containing 100 µL of β-galactosidase buffer (60 mM NaH_2PO_4, 40 mM Na_2HPO_4, 10 mM KCl, 1 mM $MgCl_2$) and 50 µL of ONPG buffer

(60 mM NaH_2PO_4, 40 mM Na_2HPO_4, 2 mg/mL ONPG). The plate was incubated at 37°C until a yellow colour developed, this was usually 45 minutes to 1 ½ hours. The A_{405} of the plates was then measured, and the reading used to normalise the luciferase values.

2.4.7 Statistical analysis

Transient transfection results were analysed using Microsoft Excel 98. The averages of triplicate wells were analysed using the Students T-test. This test returns the probability that the values come from populations with the same mean i.e. a low probability indicates that the values are not likely to be populations with the same mean. Therefore a low probability indicates that the values are likely to be from populations with different means, and so are classed as significantly different. A 1 tailed, 2 sample equal variance test was used, and values below 0.05 or 0.01 considered significant at 95%, or 99% confidence interval, respectively.

2.5 *In vivo* DNaseI footprinting

DNaseI footprinting was adapted from Riggs et al., (1998) and Carey and Smale, (2000) by Dr Kirsty Allen (Allen, 2003). Only minor changes to the protocol were made.

2.5.1 Nuclei isolation

Cells (80% confluent, T75 flask) were harvested, resuspended in 4 mL of complete media, and aliquoted equally into four micro-centrifuge tubes. The cells were then pelleted at 2,000 rpm for 10 minutes at 4°C, washed in cold PBS (0.14 M NaCl, 4.3 mM $NaHPO_4$ pH 7.2, 2.7 mM KCl), and re-centrifuged. To isolate the nuclei, the cells were lysed in 500 µL NP-40 lysis buffer (10 mM Tris-HCl pH 7.7, 10 mM NaCl, 3 mM $MgCl_2$, 0.5% NP-40, 0.15 mM spermine, 0.5 mM spermidine) for 5 minutes on ice. The nuclei were collected by centrifugation at 2,000 rpm for 10 minutes at 4°C, and washed in Buffer A (100 mM NaCl, 50 mM Tris-HCl pH 8.0, 3 mM $MgCl_2$, 0.15 mM spermine, 0.5 mM spermidine) followed by re-centrifugation. Isolated nuclei were immediately digested using DNaseI.

2.5.2 DNaseI digestion

Isolated nuclei were resuspended in 100 µL Buffer A + $CaCl_2$ (100 mM NaCl, 50 mM Tris-HCl pH 8.0, 3 mM $MgCl_2$, 0.15 mM spermine, 0.5 mM spermidine, 1 mM $CaCl_2$). The samples were then exposed to DNaseI (4.5 µg) and incubated at 37°C. Digests were stopped by addition of 2 µL 0.5 M EDTA, pH 8.0; 100 µL Buffer A; 3 µL 25 mg/mL proteinase K; and 10 µL 20% SDS. Samples were incubated at 37°C overnight and DNA was purified by sequential phenol:chloroform extraction, RNase digestion, phenol:chloroform extraction, and ethanol precipitation. The resulting DNA was resuspended in 30 µL TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0).

DNaseI footprinting requires 'naked' DNA controls, and so DNA was extracted from cell lines using the Wizard Genomic DNA Purification Kit (Promega) following manufacturers' directions. Briefly, cells at least 80% confluent in a T80 flask were passaged, and dispensed into four micro-centrifuge tubes. Cells were lysed, and RNA

digested, followed by precipitation of protein. The supernatant, containing genomic DNA was then precipitated by addition of isopropanol, and washed in ethanol. DNA was resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0).

For naked DNA controls, extracted DNA was quantified, and 20 µg of DNA was digested with 1 µg DNaseI. Samples were incubated at 37°C for 30 seconds in the presence of 1 mM Ca^{2+}. Digestion was stopped by addition of 2 µL 0.5 M EDTA and the enzyme was heat denatured at 75°C for 20 minutes. Digested DNA was then purified by phenol:chloroform extraction and ethanol precipitated.

2.5.3 Alkaline gel

To ensure DNA was appropriately digested (single-stranded fragments between 300-600 bases), samples were run on an alkaline gel along with a 450 base PCR product as a size marker. To prepare samples, EDTA was added to 15 µL of DNaseI digested DNA, to a final concentration of 10 mM, followed by 4 µL of 6 x loading dye (300 mM NaOH, 6 mM EDTA, 18% ficoll (type 400), 0.15% bromocresol green, 0.25% xylene cyanol). Samples were then electrophoresed under alkaline conditions (50 mM NaOH, 1 mM EDTA) on a 1.5% agarose gel, at 10 V overnight. The gel was then neutralised in 1 M Tris-HCl pH 7.7, 1.5 M NaCl for 1 hr, stained with ethidium bromide, and DNA visualised with UV light.

2.5.4 Primer extension

Primer extension was carried out using 0.6 pmol of biotinylated primer and 1.5 µg of DNaseI digested DNA. DNA and primer mixes were set up in 0.2 mL PCR tubes in 1 x Sequenase buffer in a total volume of 15 µL. Primer was annealed by denaturation at 95°C for 3 minutes followed by 45°C for 30 minutes. The samples were then chilled on ice, and primer extension carried out by addition of 7.5 µL Mg-dNTP stock (20 mM $MgCl_2$, 20 mM DTT, 1 mM mixed dNTPs), and 1.5 µL of 3.25 units/µL Sequenase version 2.0 (4 fold dilution in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0)) and incubation at 48°C for 15 minutes. The enzyme was denatured by addition of 6 µL 300 mM Tris-HCl pH 7.7 and incubation at 67°C for 15 minutes. The samples were

placed on ice until used in ligation reactions, which were carried out immediately after primer extension.

2.5.5 Linker preparation

Double-stranded linker was prepared by annealing complementary 11 base (11 mer), and 25 base (25 mer) oligonucleotides (refer Appendix 4). Primers were annealed in 100 µL linker mix (20 pmol/µL 11 mer and 25 mer primers, 250 mM Tris-HCl pH 7.7) by heating to 95°C for 3 minutes, then 70°C for 1 minute followed by decreasing the temperature by 2°C every 4 minutes until at 4°C. Annealed linker was then stored at -20°C for up to 3 months.

2.5.6 Ligation

Primer extended sample (from 2.5.4) was ligated to linker by addition of 45 µL ligation mix (13.33 mM $MgCl_2$, 30 mM DTT, 1.66 mM ATP, 83.3 µg/mL BSA, 100 pmol/45 µL linker, 3 units/45 µL *T4* DNA ligase) and incubation at 20°C overnight.

2.5.7 Product capture

Primer extended and ligated products (from 2.5.6) were captured using the biotin group attached to the 5' end of the extended primer. Products were captured with 37.5 µL Streptavidin M-280 beads (Dynal Biotech). Beads were washed twice in 2 x BW buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl), and resuspended in 75 µL 2 x BW buffer. The entire ligated sample (75 µL) was added to beads, and incubated at room temperature for 15 minutes. Non-bound DNA was removed by two bead washes with 2 x BW buffer. DNA was then melted by resuspension of the beads in 37 µL 0.15 M NaOH (made fresh) and incubation at 37°C for 10 minutes. The supernatant, containing the melted DNA strand, was transferred to a new tube, and neutralised with 3.75 µL 2 M Tris-HCl pH 7.7 and 37 µL 0.15 M HCl. Samples were then ethanol precipitated by addition of 9.4 µL ligation stop mix (2.7 M sodium acetate, 1 mg/mL *E.coli* tRNA see below) and 220 µL 100% ethanol, and placing on dry ice for 20 minutes. DNA was pelleted by centrifugation at 13,000 rpm for 15 minutes at 4°C. The pellet was washed in 70% ethanol, re-centrifuged, dried, and resuspended in 50 µL

water. *E.coli* tRNA was prepared by phenol:chloroform extraction and ethanol precipitation of commercially available tRNA.

2.5.8 PCR

Single-stranded captured products (from 2.5.7) were then subjected to PCR by addition of 50 µL 2 x *Taq* mix (2 x *Taq* buffer (20 mM Tris-HCl pH 8.3, 100 mM KCl), 4 mM $MgCl_2$, 1.6 mM mixed dNTPs, 10 pmol of both primer #2 and linker primer, 3 units *Taq* DNA polymerase). PCR was carried out with 20 cycles of 94°C for 1 minute, 55°C for 2 minutes, and 74°C for 3 minutes. This was followed by addition of 10 µL *Taq* booster mix (1 x *Taq* buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl), 1 unit *Taq*/10 µL) and incubation at 74°C for 10 minutes. DNA was purified by addition of 15.5 µL of *Taq* stop mix (2.4 M sodium acetate, 0.08 M EDTA, 0.32mg/mL *E.coli* tRNA), followed by phenol:chloroform extraction, and ethanol precipitated as described in 2.5.7. Samples were stored dry at -20°C until gel electrophoresis.

2.5.9 Capture of sequencing template

PCR was carried out using a biotinylated forward primer and appropriate reverse primer to give an at least 400 bp product. The product was then captured using 20 µL Streptavidin M-280 beads (Dynal Biotech). The beads were washed once in 20 µL 1 x BW buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl), resuspended in 40 µL 2 x BW buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl), and 40 µL of PCR product was added. The samples were incubated at room temperature for 15 minutes. Beads were washed with 40 µL 1 x BW buffer and resuspended in 8 µL 0.1 M NaOH. After 10 minutes at room temperature the supernatant was removed and beads washed with 50 µL 0.1 M NaOH, followed by 40 µL 1 x BW buffer, 50 µL TE buffer, and resuspension in 7 µL of water.

2.5.10 Sequencing ladder production

A sequencing ladder was produced using the dGTP reagents of the Sequenase version 2.0 DNA sequencing kit (USB), and the captured product and relative footprinting primer 2, according to the manufacturers' instructions (note: primer 2 binds to the

opposite strand to which the biotinylated primer, used to create PCR product, binds. Thus primer 2 sequences the reverse strand to the biotinylated primer).

2.5.11 Gel electrophoresis

A 0.4 mm thick, 42 mm x 33 mm sequencing gel was prepared (8% (w/v) 29.1:0.9 acrylamide:bis-acrylamide ratio, 7 M urea, 1 x TBE (0.8 M Tris-Borate, 1 mM EDTA pH 8.3)), and left to set overnight. The gel was warmed to ~55°C by applying 1500 V at max 75 W for 1 hour. Footprinting samples were resuspended in 2 μL water, and 4 μL of formamide loading dye (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) was added. Samples were then heated to 95°C for 2 minutes and chilled on ice. Sequencing samples, containing formamide, were heated to 75°C and also cooled on ice. Samples were applied to the warmed gel, and electrophoresis was carried out until the bromophenol blue dye band had reached the bottom of the gel (~4 hours). The gel apparatus was disassembled and left to cool for approximately 10 minutes before continuing.

2.5.12 Transfer

DNA was transferred from the gel onto a membrane using the Panther™ Semidry Electroblotter model HEP-3 (Owl Separation Systems) according to manufacturers' directions. The gel plates were carefully disassembled, and the gel placed on a dry piece of Whatman 3MM paper. Three layers of Whatman 3MM paper, soaked in 1 x TBE, were placed on the apparatus removing air bubbles each time. This was followed by 50 mL 1 x TBE poured evenly over the Whatman 3MM layers. The dry Whatman 3MM with adherent gel was then placed over top (gel up), avoiding any air bubbles. The pre-wet (in 1 x TBE) positively charged nylon membrane (Roche) was then placed on top of the footprinting gel, followed by three more layers of soaked Whatman 3MM paper, removing bubbles each time. The electroblotter was assembled and transfer carried out at ~1.3 amps (~12V) for 45 minutes. The apparatus was then disassembled, and the membrane allowed to dry. DNA was then UV crosslinked to the membrane using the auto crosslink setting of the UV Stratalinker 2400 (Stratagene).

2.5.13 Probe synthesis

Template for the labelled probe was prepared by PCR using forward and reverse primer 3's (schematic diagram shown in figure 5.4). The product was then column purified (QIAquick PCR purification kit, Qiagen), and gel quantified. The template DNA was denatured in boiling water for 2 minutes, cooled on ice, and 50 ng added to Ready To Go™ DNA labelling beads (Amersham), along with 5 µL (α^{32}P) dCTP in a 50 µL total volume. Labelling beads were incubated at 37°C for 30 minutes. DNA was labelled by random priming of the template DNA causing incorporation of radio-labelled nucleotide into short complementary oligonucleotides. Labelled probe was purified using Probe Quant™ G-50 micro columns (Amersham) following manufacturers' directions.

2.5.14 Hybridisation

Crosslinked footprinting membrane, wet in 1 x TBE, was pre-hybridised in 15 mL hybridisation buffer (0.25 M sodium phosphate pH 7.2, 1 mM EDTA, 7% (w/v) SDS, 1% (w/v) BSA) for two hours at 60°C. Buffer was then removed, and a further 5 mL of hybridisation buffer along with labelled probe was added. Hybridisation was carried out at 60°C overnight.

2.5.15 Washes and detection

Hybridised membrane was washed for 5 minutes in 250 mL pre-warmed (60°C) buffer A (20 mM sodium phosphate pH 7.2, 1 mM EDTA, 2.5% (w/v) SDS, 0.25% BSA). Five 7 minute washes in 400 mL pre-warmed buffer B (20 mM sodium phosphate pH 7.2, 1mM EDTA, 1% (w/v) SDS) followed. The membrane was allowed to air dry before incubation with phospho-imager plate (FujiFilm) or X-ray film (Kodak). Membrane was left with phospho-imager plate for 4-18 hours before reading with the Phospho-imager FLA 5000 (FujiFilm). X-ray films were incubated at -70°C for 7-14 days before developing.

CHAPTER 3 – Cloning of the Maspin Promoter

3.1 Identification of the human maspin promoter

The aim of this project was to identify regions of the maspin promoter which may be important for the regulation of gene expression. Since a clone representing the human maspin promoter was not readily available, the human genome sequence was used to identify the putative upstream regulatory region of the maspin gene. Human maspin mRNA (accession number U04313), and a genomic contig (accession number AC036176) containing the maspin gene were obtained from NCBI and aligned using BLAST (Altschul et al., 1990). This alignment located the maspin gene within the contig and predicted the intron-exon boundaries (Appendix 1). The last nine bases of the maspin mRNA did not align with the genomic contig and so another mRNA sequence (accession number BC020713) was obtained and aligned. While this sequence was shorter, it contained an extended 5' sequence. This second alignment predicted the 5' flanking region to be approximately 90 bp 3' to that experimentally determined by Zhang et al. (1997a). This suggests that the mRNA sequences available from GenBank may have truncated 5' ends. The Zhang et al. (1997a) transcription start site was used for all further analyses.

3.1.1 Analysis of maspin promoter

The maspin gene was found to be located on chromosome 18q21.3 within a cluster of serpin genes. The mRNA is encoded by 7 exons spread across approximately 30 kb of genomic DNA, with 7 kb separating the first and second exons. A translation start site (ATG) is located in the second exon.

Approximately 8 kb of maspin 5' flanking region was analysed by Transcription Element Search Software (TESS) to identify putative transcription factor binding sites (Schug and Overton, 1997). TESS predicts transcriptional effectors which could bind to a region of DNA by aligning the consensus sequence which the factors bind to, with the DNA sequence of interest. Thus the binding sites are used to predict potential transacting effectors. TESS identified an extensive range of potential protein binding

sites on the maspin promoter, including the previously reported Ap1 and Ets sites (Zhang et al., 1997a), whilst failing to predict the p53 site (Zou et al., 2000) and hormone response element (Zhang et al., 1997b) (Reported sites are shown in figure 3.1.). TESS analysis was carried out with the intention of identifying important regulatory sites, however as the sites were too numerous these results were set aside to use subsequently in the analysis of transient transfection experiments (Chapter 4, complete TESS results in Appendix 2)

To narrow down the search for regulatory sites, the rat and mouse maspin orthologues were identified, their promoters mapped, and aligned with the human maspin promoter by BLAST analysis (Altschul et al., 1997) (Appendix 3). At the amino acid level, the mouse and rat maspin proteins show 89%, and 88% identity respectively to human maspin, and show similar expression patterns and functions (Umekita et al., 1997; Zhang et al., 1997c). Thus the maspin protein may have evolutionarily conserved functions in mice, rats, and humans; thus expression may be regulated similarly. If a region of the promoter is conserved between human and either or both mouse and rat, then the region may be important for the regulation of maspin in these species. Regions that showed conservation of sequence and were similarly positioned relative to respective transcription start site (TSS) (+/- 500 bp) were considered to be conserved. As shown by green lettering in figure 3.1, conservation of most of the p53, the entire proximal Ap1 site, and the region between, was found in all three species. The proximal Ets site was also conserved. This indicates that these three sites are likely to be important for the regulation of gene expression, and could play conserved roles in all three species. Other sites at approximately -2430 bp, -2120 bp, and -1950 bp showed complete conservation between all three species; however no transcription factors were predicted by TESS to bind these regions. Several other sites were found to be conserved between human and either mouse (figure 3.1, light blue) or rat (red); TESS predicted few transcription factors to bind to these regions. Figure 3.1 shows regions found to be conserved and the sites predicted by TESS to be located within these conserved regions of the maspin promoter; previously reported sites are also indicated, for complete TESS results see Appendix 2.


-2551 ctccatcctc tcccactttt ctgttttcta gtgtcttttt aggtagaatt
-2501 ctaaagaact taaaattttg gctgccgtgg aaaagaatat tactgggaaa
-2451 tgggtcttct cagtcaaata gaatcacagt tttaatactc tttccactga
-2401 atgggaacca ctttttttt tcctgcaatt ttgcctaaat caacagtgtg
-2351 ttttgtcaat gatgttacac ctgagacttg tgagtccttg ccttacttgg
-2301 gttcaatgag agggtgagaa cattcacgta cagggtctga cctccagaga
-2251 tacttcaaca agtgaaagct tttgacagga aatgttctgt ttaatctttt
-2201 gttaacatat ctgtccaata taccttgtat catcccagct ctggacctcc
-2151 gttctcaggc ttaaatggtc tcctgatgtc caatgatgct gcctccatga
-2101 tccttcccca gtcatagcca gactgaaaag gtcacaccca tctctgaact
-2051 acttttgtca tttgcacctc tcttaagtcc cttattacat gcctccagat
Ap2
-2001 actttctttt gtttcttgtc ttcctactgt acccttggg gagagggaac
-1951 atgtcttata gccccagtcg aatggtgcct ctaattttca cccagttaga
-1901 atcacccaaa ggcatttcac tttgctagaa cttccttggg tattgttgaa
-1851 ttcctgtcat tggaaatgtt tctgccccgg ttgagcaagc actgaacagg
-1801 agtgttttag agggaattaa aacttcaaaa aggagaatag agaggctcta
-1751 ccatttaaat gcaagtcgct tgaaatctgt agaccacagt agcaagcacg
-1701 aagcaggtat tcaataaatg tggtgaattc aatgtcaata acatgctcca
CP1A/CP1B
-1651 gtcttttgtc cactattccc cgagcctacc taaccaatgg gcaaagtggt
USF
-1601 actaattgcc atcattagga agcacatgct atgtgtcaga aagtatgtaa
-1551 gaattgtctc ttcatccatc tctgtctata tctcatttta tcgatcttat
-1501 tgacaaatga gaactatata tagatctcat tttacagatg aggaatctga
-1451 agataaaggg attaagcaat ttgctcagga tttctcaagc aatgagtggt
-1401 tagacttagt ttgcaaattt atgactctgc tcccaaattc tcattttccc
-1351 tcctacagct ggataataga atttttaaga agcatccttc cacacctcct
C/EBPα
F2F
-1301 cacctatgac cttccctcca cttgctccac tgatgcctaa aattctgttt
-1251 ttggagttca cttagtttat gactttgatt ctggttaaca ttaaaatacc
-1201 caggacatgg gggaggggaa gctgttatca atcttagttt agtcagttaa
-1151 tatagcgagt tgctcttatt cattactcaa caagcactta ttgactgcat
-1101 actgtacatc aggtctgtac caagctttgg tgttataaaa atgaatgaag
-1051 tttggcactt acctttaaca ttgttgtagt cacatatagg agaatgtaag
-1001 agattattat gaaacaatgt agtaaatgta caatagggaa ttctagataa

Figure 3.1. The human maspin promoter. Sequence of ~2.7 kb of the human maspin promoter including 180 bp of 5' untranslated region (UTR). Transcription start site (+1) is indicated by the bent arrow. Red sequences indicate conservation between rat and human; light blue sequence indicates conservation between mouse and human; green indicates conservation between all three species. Black boxes indicate putative sites identified by TESS within the conserved regions, and dark blue boxes indicate transcription element sites that have been reported to be functionally relevant to the regulation of maspin expression.

Interestingly, a region upstream of the maspin promoter was shown to be conserved in all three species (Appendix 3). This region, 4.1 kb upstream of the human transcription start site showed approximately 130 bp of conservation in mouse, and 210 bp of conservation in rat. Whilst not completely conserved, both rat and mouse showed 80% conservation with human. This region was not within the 500 bp positional conservation, however the high conservation in all three species, and its position relative to the transcription start site suggests that this region may be an enhancer. Enhancers are capable of functioning even if moved hundreds or thousands of base pairs away from the original site, indicating that position is not relevant to function. This lack of pressure on the positioning of such a region may explain why this putative enhancer region was not found to be positionally conserved. Factors which were predicted by TESS to bind to this region included NF-κB and Sp1. Both of these factors activate transcription and are ubiquitous. NF-κB is known to have reduced activity under oxidizing conditions caused by reactive oxygen intermediates (www.cbil.upenn.edu/tess). Sp1 interacts with NF-κB, and TFIID (www.cbil.upenn.edu/tess). The binding of NF-κB to this conserved region suggests a mechanism by which manganese containing super-oxide dismutase (MnSOD) may induce maspin expression. Increased expression of MnSOD would cause a decrease in reactive oxygen species, and thus an increase in transcriptional activation of maspin by NF-κB. This activation may be further enhanced by interaction with Sp1.

3.2 Cloning of the maspin promoter

3.2.1 Primer design

The promoter region had to be cloned in order to identify functionally important regions. The promoter sequence, identified from the human genome sequence, was used to design primers which amplify ~2.7 kb, including 180 bp of 5' untranslated region, and 2551 bp of putative maspin promoter (sequence shown in figure 3.1). This region was chosen as it contains the first exon, and a large portion of upstream sequence. Primers were designed manually, to have approximately 50-60% G-C content, and not to contain an A or T at the 3' end (primer sequences shown in Appendix 4). After amplification, the promoter was cloned directly upstream of the luciferase reporter gene in the pGL3 Basic vector (Promega, Appendix 5).

To aid cloning of the PCR fragment into pGL3 Basic, the region to be amplified was analysed for restriction sites contained within the pGL3 Basic multiple cloning site (MCS) using the 'map' function of GCG (GCG, 1997) (Appendix 6). Both *Xho* I and *Sac* I restriction endonucleases cut the MCS but not the maspin promoter fragment to be amplified. Recognition sites for these enzymes were then added to the ends of the primers to allow directional cloning into pGL3 Basic.

3.2.2 Amplification of promoter

Using the primers designed to the GenBank derived maspin promoter sequence, a 2.7 kb fragment was amplified from human genomic DNA. After trialling several DNA polymerases including *Taq*, *Tgo* and Expand™, it was found that *Pwo* DNA polymerase was the most efficient, giving the cleanest PCR product (figure 3.2A). *Pwo* DNA polymerase contains a 3'-5' exonuclease activity, also known as a proof-reading function, and so would be expected to incorporate fewer errors than *Taq* DNA polymerase which lacks this function. The PCR products were pooled and purified using QIAquick PCR purification kit (Qiagen). The resulting purified PCR product was digested with *Bgl* II to confirm that it represented the expected fragment of the maspin promoter. Digestion of the 2.7 kb fragment should yield approximately 1.1 kb and 1.7 kb fragments (figure 3.2B); this is what is observed in the gel (figure 3.2C), and thus the product was considered to be correct.

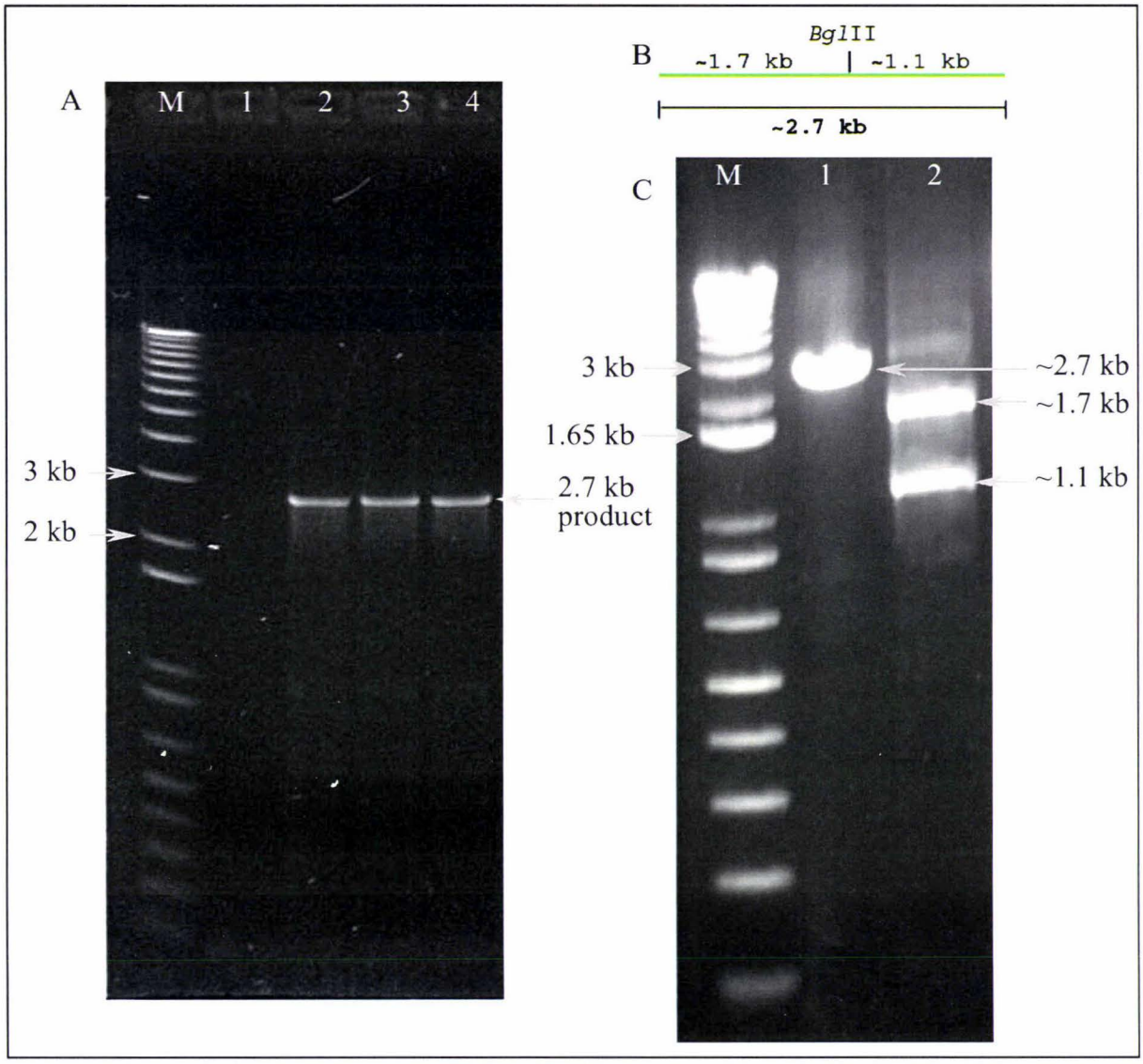


Figure 3.2. A. 2.7 kb maspin promoter PCR product. *Pwo* DNA polymerase was used to amplify the maspin promoter fragment from human genomic DNA. PCR was carried out in 1 x *Pwo* PCR buffer with 250 ng of forward and reverse primers, 0.3 mM mixed dNTPs and 2 units of polymerase in a 50 µL volume. 40 cycles were performed – denaturing for 30 seconds at 94°C, annealing for 30 seconds at 58°C, and extension at 72°C for 2 minutes and 30 seconds. Samples were then analysed by electrophoresing 5 µL of 50 µL PCR reaction on a 1% agarose gel in 1 x TAE, stained with ethidium bromide. M = 1 kb plus DNA ladder, 1 = negative PCR control. 2-4 = replicate PCR reactions.

B. Schematic diagram of the *Bgl* II digested 2.7 kb maspin promoter fragment. *Bgl* II cuts the PCR fragment once resulting in approximately 1.1 kb and 1.7 kb fragments. Green line indicates PCR fragment, vertical black line indicates *Bgl* II cut site.

C. Diagnostic digest of the 2.7 kb PCR product. Pooled, purified PCR product was digested with 1 unit of *Bgl* II, in the manufacturers recommended buffer for 1 hour at 37°C. The resulting sample was then electrophoresed on a 1% agarose gel as described above. M = 1 kb plus DNA ladder, 1 = Pooled 2.7 kb PCR fragment, 3 = *Bgl* II digested PCR fragment.

3.2.3 Cloning

A schematic representation of the cloning strategy is shown in figure 3.3. Briefly, pGL3 Basic was prepared to clone the PCR product by digesting with *Xho* I and *Sac* I restriction endonucleases. For some reason, *Xho* I did not completely cut the vector, whilst the pBlueScript vector (Stratagene) was cut to completion under identical conditions, indicating a possible problem with the *Xho* I restriction site in the pGL3 Basic vector. Because of this *Sma* I, which generates blunt-ends, was used to cut the vector along with *Sac* I thereby still allowing directional cloning. Removal of 5' phosphate groups from the cut vector was carried out using thermosensitive alkaline phosphatase (TsAP), and the linearised vector was gel purified.

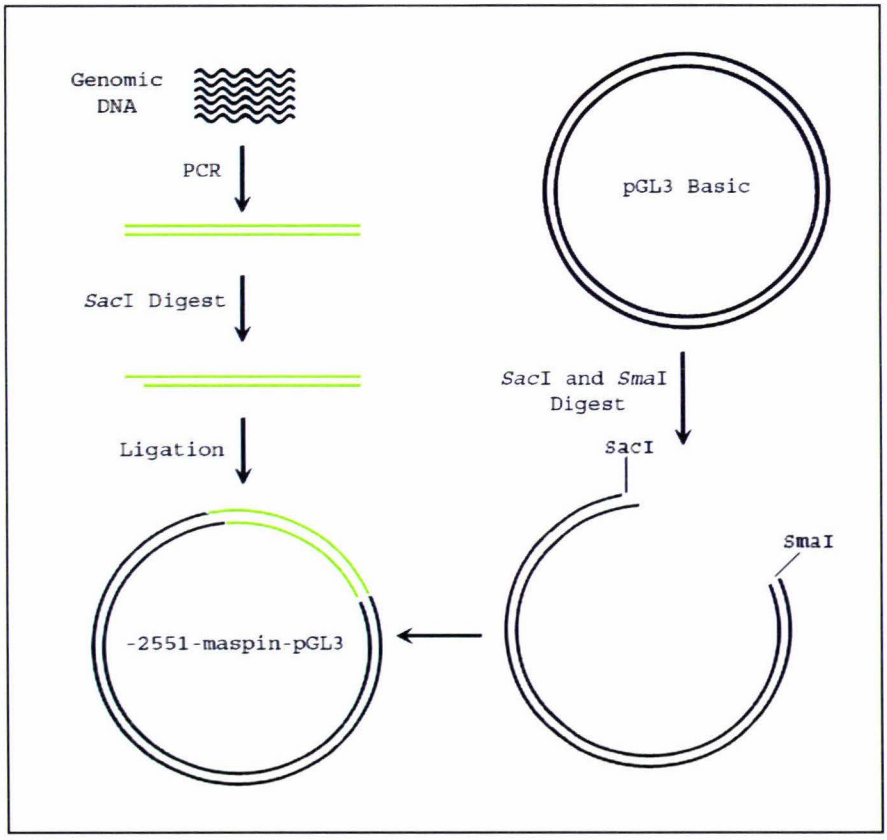


Figure 3.3. Schematic representation of cloning strategy. A 2.7 kb fragment of the maspin promoter was amplified from human genomic DNA using *Pwo* DNA polymerase. The product was then digested with *Sac* I and column purified. pGL3 Basic vector was prepared by digestion with *Sma* I and *Sac* I restriction endonucleases, followed by removal of phosphate groups with TsAP, gel purification, and gel quantification. *E.coli* XL-1 Blue strain was then transformed with insert-vector after ligation using T4 DNA ligase.

As *Pwo* DNA polymerase leaves blunt-ends, the maspin promoter PCR product was digested with *Sac* I only; this was then column purified (Qiagen) and gel quantified along with the purified vector (figure 3.4).

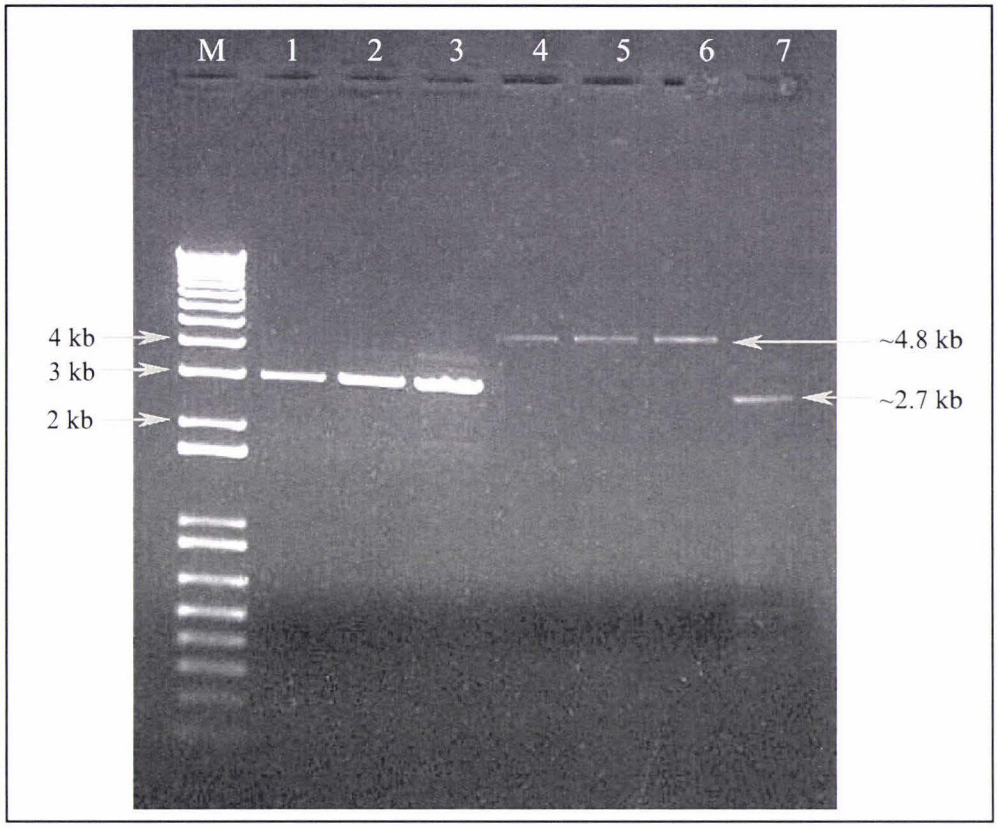


Figure 3.4. Gel quantification of insert and vector. Digested insert was purified using QIAquick PCR purification kit (Qiagen) and vector was purified using QIAquick gel extraction kit (Qiagen). Insert and vector were then quantified by electrophoresing 1 µL of each sample with 20, 50, and 100 ng quantification standards on a 1% agarose gel, in 1 x TAE, stained with ethidium bromide. M = 1kb plus DNA ladder, 1 = 20 ng standard, 2 = 50 ng standard, 3 = 100 ng standard. 4 – 6 = *Sma* I and *Sac* I cut vector, 7 = *Sac* I digested PCR product (insert).

Figure 3.4 shows a quantification gel from which both insert (lane 7) and vector (lanes 4-6) were quantified. From this gel, the DNA concentration of both insert and vector were estimated to be approximately ~10 ng/µL. The digested PCR product was purified using a column based kit, thus some non-specific bands from the PCR were present in this purification. The non-specific PCR bands are very faint compared to the 2.7 kb product indicating a much lower concentration. Consequently the PCR product was not purified further, and used directly in subsequent ligations. The quantification markers were created by dilution of a known concentration of cut vector. The faint bands present in the quantification marker lanes suggest that the plasmid was not digested to completion.

Ligation reactions were prepared in a 10 µL volume containing less than 100 ng total DNA with a 3:1 insert:vector ratio, and incubated at room temperature for 30 minutes. Half of each ligation reaction was then used to transform competent *E.coli* XL-1 Blue cells (see Chapter 2.3.2-2.3.4), which were plated on LB agar containing ampicillin and incubated at 37°C overnight. The next day growth on the plates was counted and tabulated (table 3.1)

Treatment	Media	Result
Untransformed	LB	Lawn
	LB amp	3
Uncut vector	LB amp	>1000
Cut vector	LB amp	9
Cut vector + ligase	LB amp	13
Ligation 1:3	LB amp	25

Table 3.1. Results of ligated DNA transformation. Ligations were carried out in a 10 µL volume, using 1:3 vector:insert ratio. Ligations were incubated at room temperature for 30 minutes. Competent cells were created using 'growth on the day' technique, transformed with 5 µL of ligation mix, and incubated on ice for 5 minutes before being plated onto stated agar. Cells were then grown overnight at 37°C, and the number of colonies counted the next day.

Table 3.1 shows results of a transformation. Several controls were carried out with transformations; untransformed cells were plated on LB media without antibiotic to ensure cell viability, and also plated on LB amp media to ensure that antibiotic restricts growth. Transforming with uncut vector demonstrates the capability of the cells to take up DNA. If growth is seen on cut vector controls this indicates that there is some uncut plasmid remaining, and cut vector plus DNA ligase indicates whether the vector can be re-ligated to itself.

The untransformed cells plated onto antibiotic media control showed growth of 3 colonies, this may be because the antibiotic was old and partially broken down, or because 3 spontaneous mutants prevented the toxicity of ampicillin, allowing growth on ampicillin containing media. Both cut vector, and cut vector plus DNA ligase showed increased numbers of colonies on ampicillin media. This was likely to be due to incomplete cutting of vector, and remaining 5' phosphate groups allowing re-ligation to occur. As both enzymes used to cut the vector cut within the MCS it was impossible to ensure both reactions went to completion. This means that compatible ends may be present enabling re-ligation of linearised vectors with 5' phosphate groups.

Unfortunately, transformations seemed to form only a few more colonies than the negative controls; this may be due to inefficient ligation of the insert and vector.

LB amp broths were inoculated with single colonies from transformation plates, and incubated with shaking at 37°C overnight. DNA was extracted from the resulting cultures using the rapid boil technique (Chapter 2.3.5), and digested with *Bgl* II to identify plasmids which had taken up the insert in the correct orientation. *Bgl* II cuts the vector and the insert each once (figure 3.5A), thus if the PCR product had been ligated into the vector in the correct orientation 1.7 kb and 5.9 kb fragments would be produced. Figure 3.5B shows a gel of this diagnostic digest; the empty pGL3 Basic vector has been linearised giving a 4.8 kb fragment (lane 2), whilst the extracted plasmid has released the predicted 1.7 kb and 5.9 kb products (lane 4). The uncut extracted plasmid (lane 3) has a lower mobility than the uncut pGL3 Basic vector (lane1) indicating that DNA has been ligated into the vector. Interestingly, a band in the uncut pGL3 Basic vector lane (lane 1) co-migrates with cut pGL3 Basic (lane 2), indicating that this plasmid prep has been partially linearised. Glycerol stocks were made from the cells containing the maspin promoter construct shown in lanes 3-4, and pure plasmid for use in transfections was isolated using the Qiagen Maxi Prep Kit.

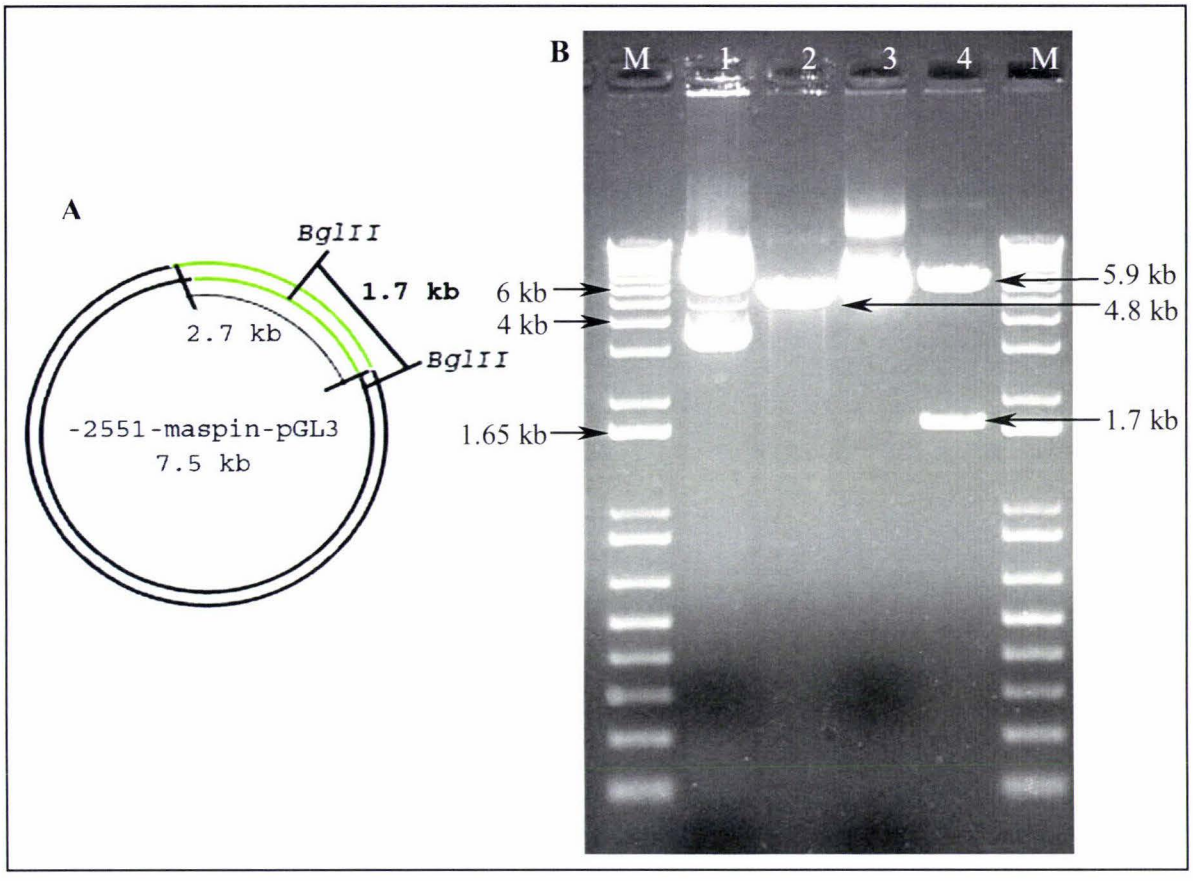


Figure 3.5. A: Schematic diagram of diagnostic digest. *Bgl* II digests -2551-maspin-pGL3 twice, once in insert, and once in the MCS. The inserted DNA is indicated in green and is 2.7 kb, the pGL3 Basic vector is black and 4.8 kb, thus total size of this plasmid is ~7.5 kb. When digested with *Bgl* II, -2551-maspin-pGL3 construct would release 1.7 kb and 5.9 kb fragments.

B: *Bgl* II digest of pGL3 Basic and a rapid boil sample. *E.coli* containing potential clones were grown in 5 mL LB broth containing 0.01 mg/mL ampicillin. Plasmids were then harvested from the cells using the rapid boil technique. Resulting DNA was digested with *Bgl* II and RNase for 1 hour at 37°C. Uncut plasmid DNA and 5 µL of digested DNA was electrophoresed on a 1% agarose 1 x TAE gel at 120 V for approximately 1 hour. M = 1 kb plus DNA ladder, 1 = undigested pGL3 Basic, 2 = digested pGL3 Basic, 3 = undigested extracted DNA,
4 = digested extracted DNA.

3.2.4 Sequencing of the maspin promoter

Sequencing primers were designed, using the maspin promoter from the human genome sequence, to enable sequencing of the full length construct. Primers were spaced approximately 400 bp apart, and designed to sequence forward and reverse strands (figure 3.6) (primer sequences are listed in Appendix 4). Sequencing was carried out by Allan Wilson Centre Genome Service, using an ABI3730 Genetic Analyzer (Applied Biosystems) (an example sequence is included in Appendix 7). Individual sequences were edited manually and aligned with the human genome derived maspin promoter sequence using the Bestfit function of GCG (GCG, 1997).

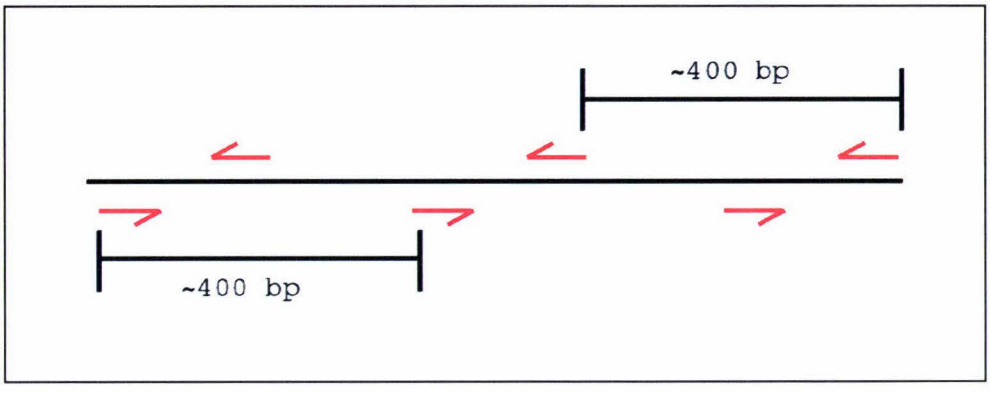


Figure 3.6 Schematic diagram of sequencing primer positioning. Primers were spaced approximately 400 bp apart, and forward and reverse primers were overlapped so that every position was sequenced in both directions. Black line indicates maspin promoter, red arrows indicate primers and direction of extension.

The full length construct was completely sequenced in both directions, and sequence data was assembled into a contiguous sequence (contig) using the Gelmerge program of GCG (GCG, 1997). The resulting contig was then aligned (using Bestfit) with the human genome derived sequence for comparison, and discrepancies between this and the reported sequence were noted (Appendix 8). There were 8 mismatches, and 2 single base deletions in the cloned sequence, when aligned with the human genome sequence. Four of these single nucleotide polymorphisms (SNP) are present in the SNP database (Sherry et al., 2001). The other mismatches may be due to PCR induced errors or novel SNPs, the deletions are unlikely to be PCR induced errors, thus are likely to be caused by genetic diversity.

3.3 Creation of maspin promoter deletion series

3.3.1 Strategy

The full length construct was successfully cloned into the pGL3 Basic vector using *Sma* I and *Sac* I restriction sites in the MCS (see figure 3.5). The creation of a deletion series, combined with functional assays, allows the determination of potentially important regions of the maspin promoter. TESS predicted too many transcriptional effector sites within the maspin promoter, and so potential effectors were not obvious. Consequently deletions were created by digestion at conveniently located restriction endonuclease recognition sites in the maspin promoter (refer to 'map' in Appendix 6). An example of how these deletions were created is shown in figure 3.7.

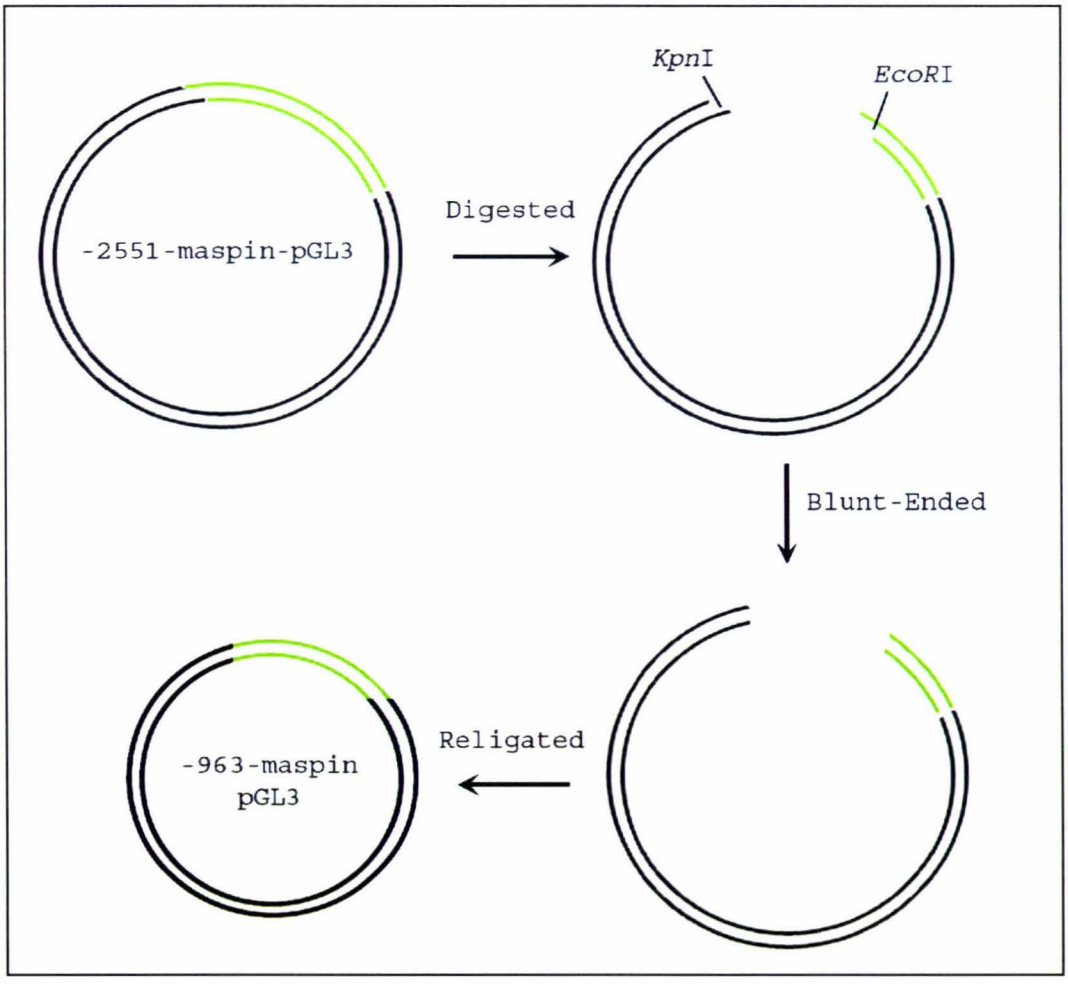


Figure 3.7. Schematic diagram of deletion series creation. The figure demonstrates the steps involved in the creation of -963 bp deletion. The full length construct was first cut with *Kpn* I, followed by *Eco* RI which releases a fragment of the promoter. The vector was then blunt-ended, gel purified, and religated to itself. This strategy was used in the creation of all deletions except for the -263-maspin-pGL3 construct. Green indicates cloned maspin promoter, black indicates pGL3 Basic vector.

As demonstrated by figure 3.7, the full length construct was digested with *Kpn* I restriction endonuclease, which cuts at 5' end of the MCS only. Another restriction enzyme was then used to truncate the 5' end of promoter (enzymes used are detailed in table 3.2); the digested vector was then rendered blunt-ended using *Klenow* fragment of *E.coli* DNA polymerase I (*Klenow*), gel purified, and religated to give a shorter promoter within the pGL3 Basic vector (refer to figure 3.8 for restriction map of all digests). This strategy was used for the creation of all deletions except the -263-maspin-pGL3 construct, where *Sca* I cut the vector too many times to efficiently gain a partial digestion product with complete vector (see figure 3.8G). In this case the maspin promoter fragment was removed by digestion with *Sca* I and *Nco* I, and sub-cloned (refer 3.3.2). The length of remaining insert and name of construct differ as all constructs contain approximately 180 bp of 5'UTR (table 3.2).

Some of the deletions were created by partial digestion (table 3.2). In these cases the second enzyme used had another restriction site elsewhere in the construct (see figure 3.8). Partial digests were stopped before reaching completion (see figure 3.9 for example), and the required fragment was gel extracted (example shown in figure 3.10). Uncut DNA was not run on the gel shown in figure 3.9 as the vector was cut to completion by *Kpn* I. This was confirmed by gel electrophoresis before partial digestion was carried out.

Name of construct	Length of remaining insert	Enzyme used	Type of digest	Purified Fragment
-2551	2.7 kb	-	-	-
-2234	2.4 kb	*Hind* III	Partial	7.2 kb
-1478	1.7 kb	*Bgl* II	Partial	6.5 kb
-1078	1.3 kb	*Hind* III	Partial	6.1 kb
-963	1.1 kb	*Eco* RI	Complete	5.9 kb
-660	0.8 kb	*Sph* I	Partial	5.6 kb
-295	0.48 kb	*Pst* I	Complete	5.3 kb
-263	0.45 kb	*Sca* I	Removed	~510 bp
-134	0.31 kb	*Sna* BI	Complete	5.1 kb

Table 3.2. Creation of deletion series. Table details the restriction enzymes used, type of digest, size of remaining insert DNA, and the name of the construct. Partial digest indicates that the vector was not cut to completion, complete indicates that the vector was completely cut, and removed indicates that the insert was cut out of the vector and sub-cloned into an empty pGL3 Basic vector. A schematic of all digests is shown in figure 3.8.

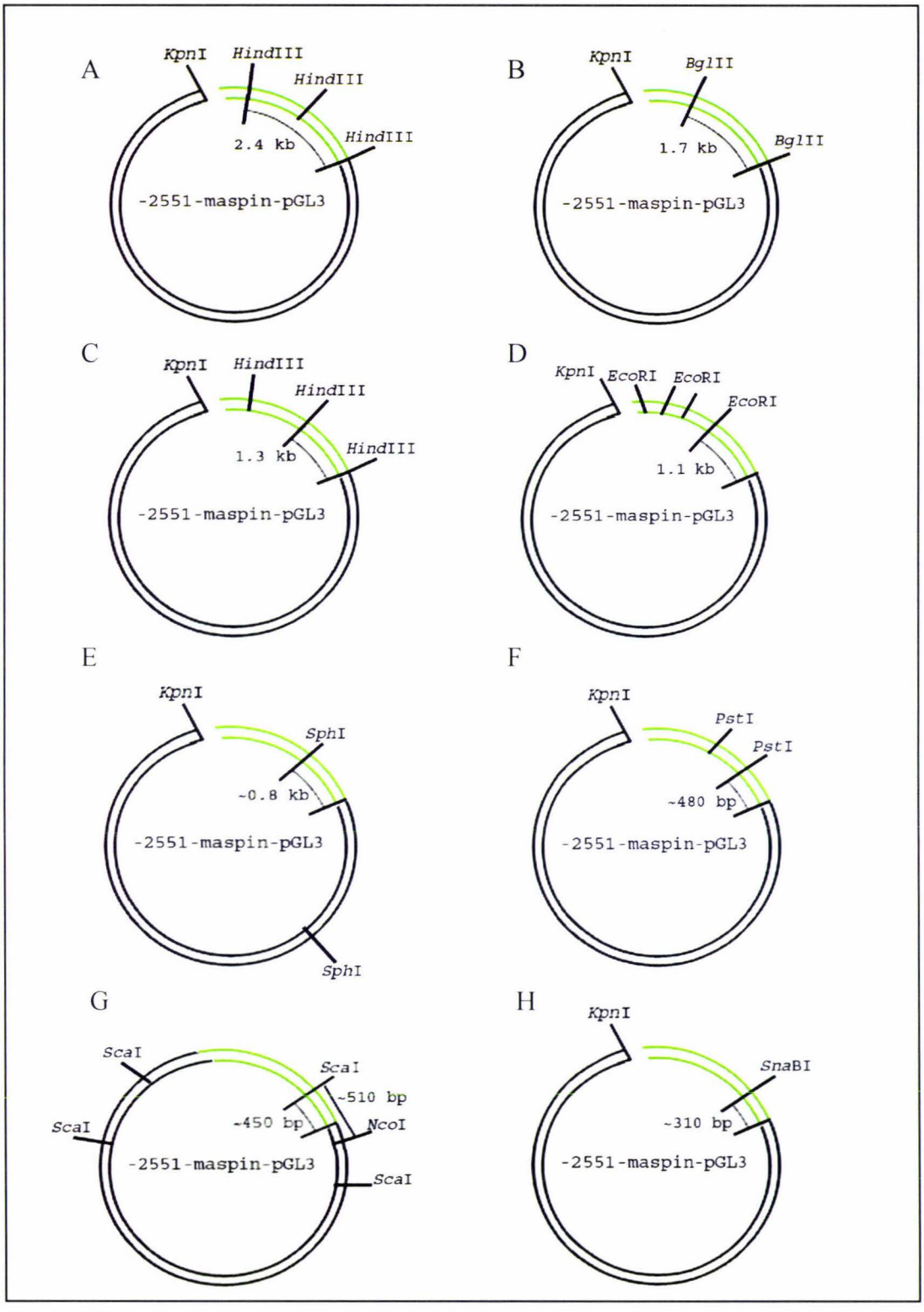


Figure 3.8. Digestion maps for the creation of deletion series. Full length construct was digested to completion with *Kpn* I before digestion with indicated enzyme (except G). Wanted fragment (indicated length plus intact vector (4.8 kb)) was gel purified and religated to create a shorter construct. A = -2234 bp deletion, B = -1478 bp deletion, C = -1078 bp deletion, D = -963 bp deletion, E = -660 bp deletion, F = -295 bp deletion, G = -263 bp deletion, H = -134 bp deletion. Note: -263 bp deletion was created from -295 bp deletion construct.

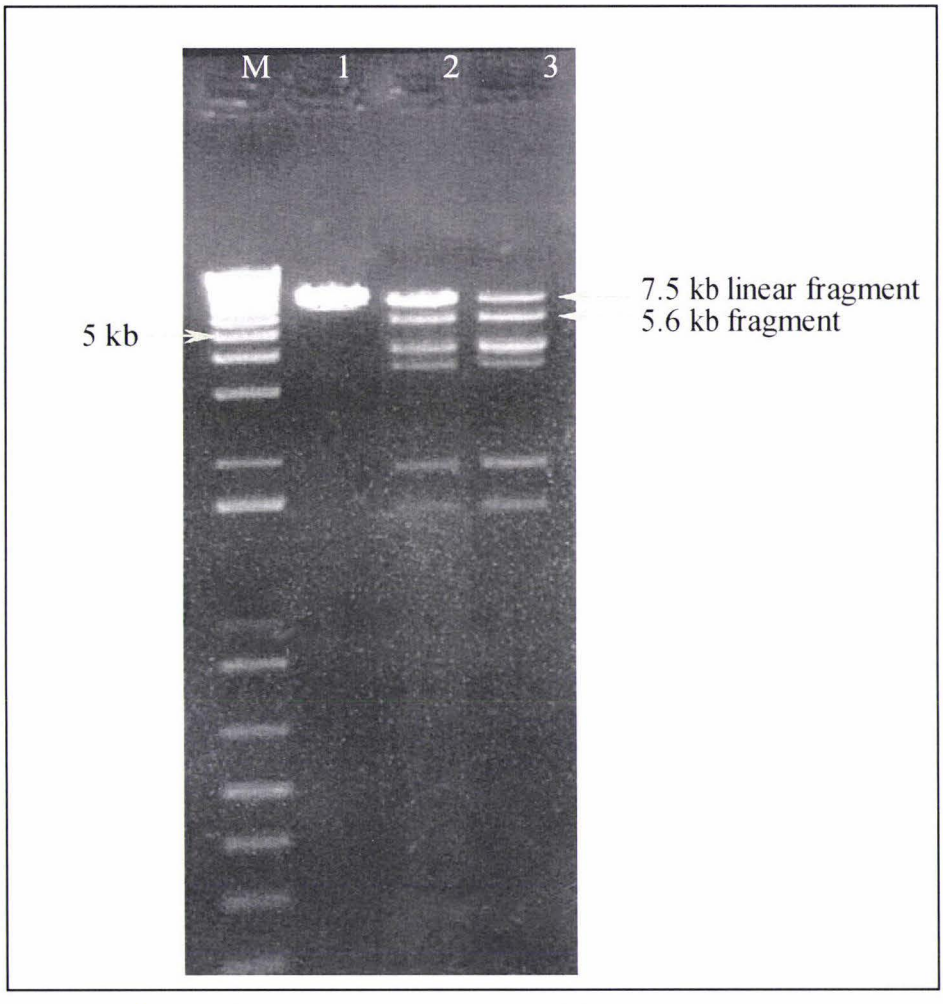


Figure 3.9. Example of a partial digest. The -2551-maspin-pGL3 construct was completely digested with *Kpn* I (lane 1), and then partially digested with increasing amounts of *Sph* I. Partially digested DNA was then electrophoresed on a 1% agarose 1 x TAE gel at 120 V for approximately 1 hour. A 5.6 kb fragment was required to create -660-maspin-pGL3 construct. The 5.6 kb fragment contains vector cut once at *Kpn* I site, and approximately 800 bp of maspin promoter (see figure 3.8E). 0.4 units of enzyme showed the best pattern (lane 3), as less of the 7.5 kb fragment is seen therefore making the 5.6kb fragment easier to gel purify. M = 1 kb plus DNA ladder. 1 = *Kpn* I digested linearised -2551-maspin-pGL3, 2 = linearised construct digested with 0.2 units *Sph* I, 3 = linearised construct digested with 0.4 units *Sph* I.

3.3.2 Sub cloning

The -263-maspin-pGL3 deletion construct was created by complete digestion of -295-maspin-pGL3 with *Sca* I, and rendered blunt-ended with the *Klenow*. The fragments were then digested with *Nco* I which cuts the vector only once (figure 3.8G, and pGL3 Basic vector map Appendix 5). This *Nco* I, *Sca* I fragment was then gel purified (figure 3.10) and ligated to *Sma* I, *Nco* I digested pGL3 Basic vector. This was the only deletion where the insert was sub-cloned.

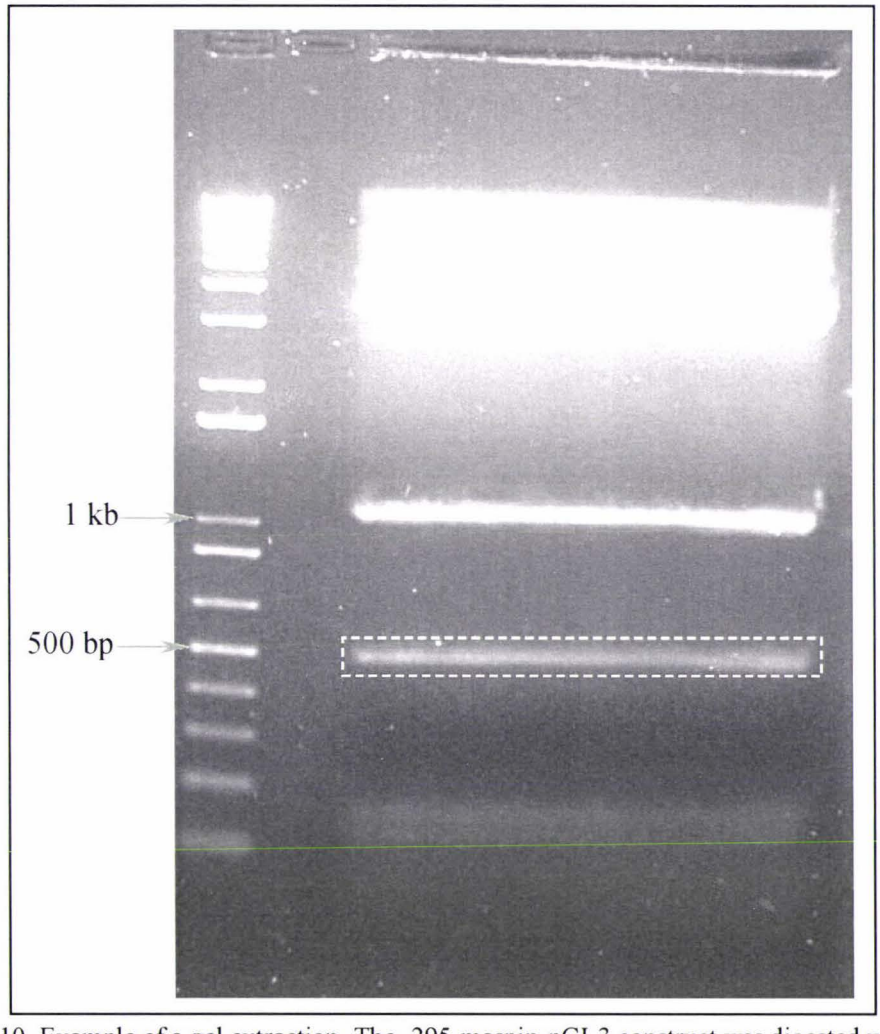


Figure 3.10. Example of a gel extraction. The -295-maspin-pGL3 construct was digested with *Sca* I, made blunt-ended with *Klenow* enzyme, and then digested with *Nco* I. Resulting samples were then combined and electrophoresed on a 1% agarose gel stained with ethidium bromide at 80 V for 1 hour in 1 x TAE. The band indicated by a dashed box was removed from the gel, and the DNA recovered using the freeze-squeeze method (refer to Chapter 2.2.5). The fragment was then ethanol precipitated, gel quantified, and ligated to *Sma* I, *Nco* I cut pGL3 Basic vector.

3.3.3 Confirmation of 5' ends

The 5' end of each deletion construct was sequenced to confirm the correct cloning and end of the truncated maspin promoter DNA. This was carried out with pGL3 Basic-specific primers (RV3 and GL2 refer Appendix 4) or internal primers, using PCR products amplified from the deletion vectors as sequencing templates. The sequences obtained were edited as for the full length construct, and aligned with the human genome derived maspin sequence using Bestfit (GCG, 1997). The exact ends were determined and are shown in table 3.2 as the construct name.

3.3.4 Construct Production

All except the -295 bp deletion were grown in *E.coli* XL-1 Blue strain. The -295-maspin-pGL3 construct was propagated in *E.coli* DH5α strain as the plasmid appeared to be either rearranged or lost in *E.coli* XL-1 Blue cells. The different *E.coli* strains should not affect the plasmid when used in reporter gene assays as both strains have similar genotypes (refer to Chapter 2.1.7) and both contain the endogenous *E.coli* methylation enzymes, indicating that plasmid DNA would be unlikely to have different methylation status (NEB catalogue).

All constructs were extracted from cell pellets using the Qiagen Maxi Prep Kit. Extracted plasmid DNA was quantified by spectrophotometer (refer Chapter 2.2.10). An example of spectrophotometric quantification is shown in figure 3.11. The absorbance of the sample was measured by the spectrophotometer at 260, 280 and 310 nm. An $A_{260} = 1.0$ indicates the sample has a concentration of 50 μg/mL, thus the results in figure 3.11 are in μg/mL. The A_{320} was a background reading for contaminating solvent, and was subtracted from the A_{260} reading before calculation of concentration. Ratio is a measure of purity of the DNA. A ratio of 1.8 indicates pure DNA, a higher ratio indicates protein contamination, while a lower ratio indicates RNA. The ratio is calculated by dividing the 260 nm reading by the 280 nm reading.

Sample	260nm	280nm	320nm	Ratio	Result
1	0.378	0.209	0.020	1.89	17.90
2	0.366	0.193	0.009	1.94	17.85

Figure 3.11. Example of a spectrophotometric DNA quantification. The absorbance of a 100 fold dilution of Maxi Prep DNA was measured at 260, 280 and 320 nm, using in 1 cm path-length quartz curvettes. Sample 1 = -295 -maspin-pGL3 Maxi Prep DNA, sample 2 = -134-maspin-pGL3 Maxi Prep DNA.

As only the 5' ends of deletions were sequenced, all constructs were digested to confirm that the constructs were as expected. *Nco* I and *Not* I restriction endonucleases each cut the pGL3 vector once, releasing the MCS in an approximately 250 bp fragment (figure 3.12). Thus any DNA cloned into the MCS would increase the size of this fragment therefore decreasing mobility on an agarose gel (figure 3.13). The linearised pGL3 Basic vector is present in all lanes as the ~4.6 kb band. The fragment representing cloned DNA has varied mobility, increasing as the cloned DNA shortens. The cloned DNA is approximately 250 bp shorter than the corresponding fragment. Uncut vector was not shown in figure 3.13, for an example of uncut vector refer to figure 3.5.

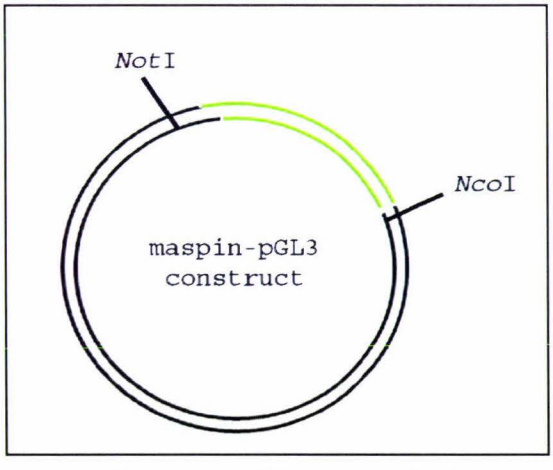


Figure 3.12. *Not* I and *Nco* I restriction map of the maspin-pGL3 constructs. Both enzymes cut the vector once only, on either side of the multiple cloning site. Thus these enzymes remove the MCS and cloned DNA. An ~250 bp fragment would be removed from empty pGL3 basic vector.

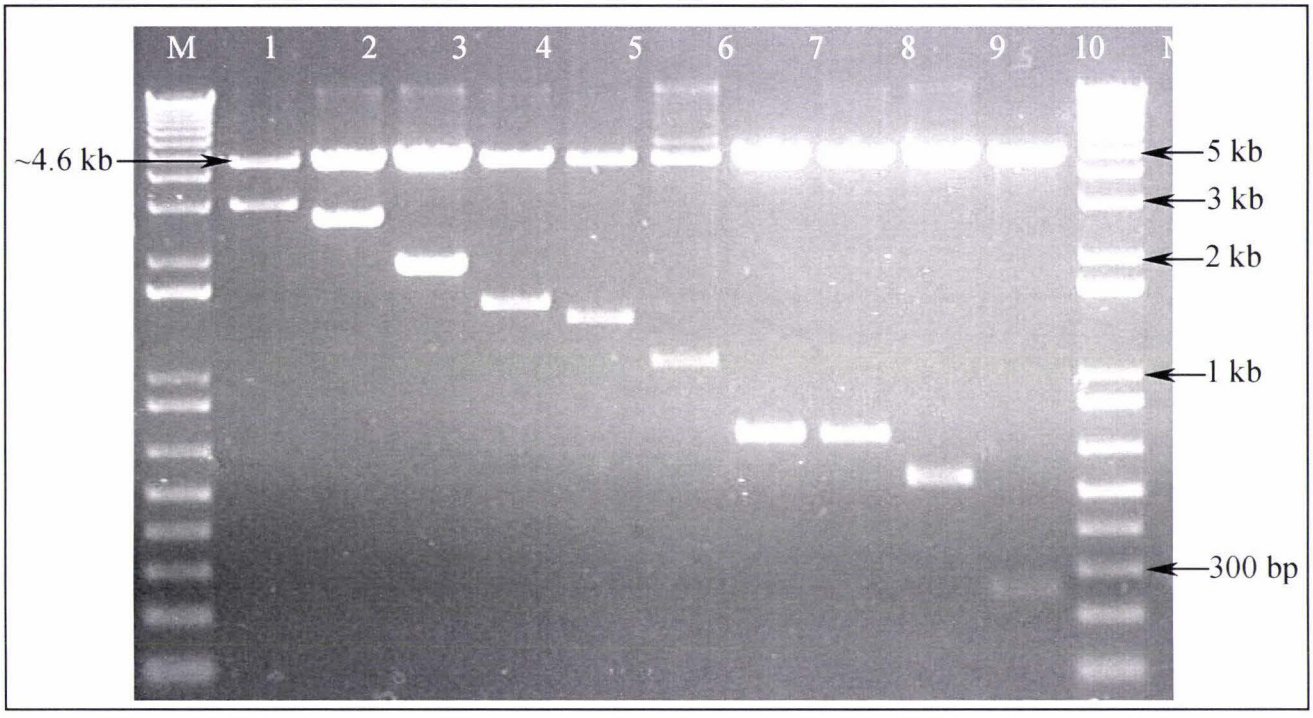


Figure 3.13 Maspin promoter deletion series digested with *Nco* I and *Not* I. Digestion was carried out with 1 µg of each construct and 1 unit of both *Nco* I and *Not* I restriction endonucleases in a 30 µL volume for 1 hour at 37°C. Digested DNA was then electrophoresed on a 1% agarose 1 x TAE gel, stained with ethidium bromide, at 120 V for approximately 1 hour. These enzymes release an approximately 250 bp fragment from pGL3 Basic which includes the MCS (lane 10). Therefore the inserted DNA is approximately 250 bp shorter than the fragment released. M = 1 kb plus ladder. 1 = -2551 bp. 2 = -2234 bp. 3 = -1478 bp. 4 = -1078 bp. 5 = -963 bp. 6 = -660 bp. 7 = -295 bp. 8 = -263 bp. 9 = -134 bp. 10 = pGL3 Basic

3.4 Chapter summary

The maspin promoter was successfully identified within the human genome sequence, using maspin mRNA sequences obtained from GenBank. The identified promoter was then analysed by TESS to predict transcriptional effectors whose sequence specificity could allow binding to the maspin promoter. Unfortunately TESS predicted too many potential effector sites, thus preventing identification of putative modes of regulation. Consequently, the TESS results were kept to aid later analysis of the results from transient transfection of the deletion series into cell lines.

To potentially minimise the number of transcriptional effector sites predicted by TESS, the mouse and rat maspin promoters were obtained and aligned with the human promoter. These alignments identified regions of the maspin promoter which were conserved in both sequence and position. Several regions were found to be conserved between all three species, including the previously identified Ap1, and p53 sites. Conservation of such regions between all three species indicates that these regions are likely to be important for the regulation of the maspin gene, and support the use of these alignments for identification of potentially important regions. The results from TESS analysis were then used to predict transcriptional effectors which could bind to these regions of the promoter. Unfortunately most of the regions found to be conserved between all three species were not predicted by TESS to bind transcription factors. Some regions conserved in human, and either mouse or rat, were predicted to bind transcription factors.

A conserved region approximately 4.1 kb upstream of the maspin promoter was also discovered, this region may act as an enhancer, and is predicted to bind NF-κB, which may explain the activation of maspin transcription by MnSOD.

Primers were designed to amplify 2.7 kb of the human maspin promoter including approximately 180 bp of 5' UTR. This PCR product was successfully cloned into the pGL3 Basic vector and sequenced. A total of 8 SNP's and 2 deletions were found in the sequence of the cloned human maspin promoter when compared to the human genome derived sequence obtained from GenBank. Four of the SNP's have been previously

identified but the remaining 4 may be novel SNP's, or PCR induced errors. The deletions are unlikely to be PCR induced errors and could represent genomic diversity. As the region being analysed was a promoter, these single differences may not be functionally significant depending on their positions, their importance can only be studied once a full investigation of the promoter region has been completed.

The identification and subsequent cloning of the maspin promoter allows functional assays to be carried out. The deletion series was created to identify potentially important regulatory regions within the maspin promoter. In total 9 maspin promoter deletion constructs were created. Due to the high number of transcriptional effector sites predicted by TESS, the deletion series was created by conveniently located restriction endonuclease sites. The resulting constructs each remove a region of DNA from the 5' end of the maspin promoter. The truncated constructs can then be transfected into cell lines, and the level of reporter gene activity measured. This allows relative differences in transcriptional efficiency of the truncated constructs to be observed.

CHAPTER 4 – Transient Transfections

4.1 Overview

Transient transfections followed by reporter gene assays were used to assess the transcriptional activity of the maspin promoter. Transient transfections are a convenient way of measuring the transcriptional activity of a promoter or enhancer in many different cell lines. The region of interest is cloned directly upstream of a reporter gene, and co-transfected into adherent cells with a transfection efficiency control vector (figure 4.1). The cells are then harvested and assayed for reporter and control gene expression. In this case the maspin promoter was cloned upstream of the luciferase reporter gene in the pGL3 Basic vector (Promega, refer to Chapter 3 and Appendix 5), and either pSV β–galactosidase (Promega, Appendix 5) or pCMV Sport-β gal (Invitrogen, Appendix 5) were used as transfection control vectors.

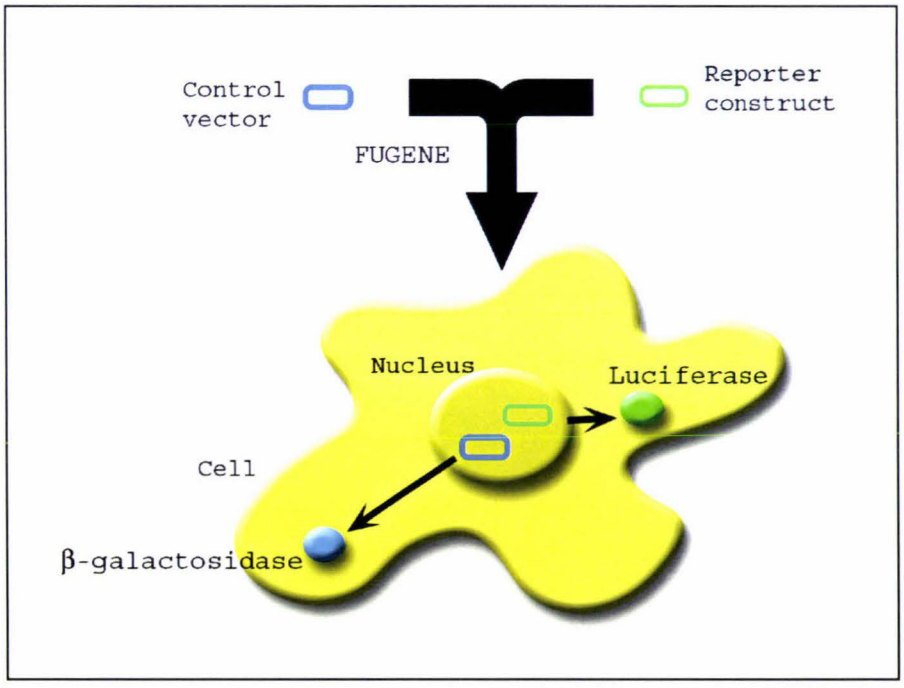


Figure 4.1. Schematic diagram of a transient transfection. The reporter construct and transfection efficiency control vector are co-transfected into cells using FUGENE™ transfection reagent. Once in the cell the plasmids locate to the nucleus where the reporter constructs are transcribed at rates relative to the efficiency of the cloned promoter. 40-48 hrs after transfection the cells are harvested and assayed for luciferase and β-galactosidase activity.

The utility of transient transfections as an experimental system is based on two assumptions. Firstly, a cloned promoter will activate the transcription of an easily measured reporter gene in the same way as its endogenously encoded gene. This means that the activity of the reporter gene is directly proportional to the efficiency of the promoter. The second assumption is that the transfection efficiency is the same for both reporter and control vectors. Cells were co-transfected with the maspin reporter constructs and a constitutively active β–galactosidase expression vector. The level of β–galactosidase activity in the resulting cell extracts was then measured by the cleavage of colourless ONPG (*o*-nitrophenyl-beta-D-galactopyranoside) forming the yellow ONP (*o*-nitrophenol), and resulting change in absorbance at 405 nm. Luciferase was the reporter gene that was used for analysis of the maspin promoter. Luciferase catalyses the oxidation of luciferin into oxyluciferin. Light is released as a by-product of this reaction and was easily measured by the FluoStar Galaxy (BMG Technologies). The level of luminescence was then normalised to β–galactosidase activity (figure 4.2). Statistical analyses were carried out on the results to determine significant differences. To this end the averages of each triplicate were analysed using a 1 tailed, 2 sample equal variance Students T-test (table of significant differences is shown in Appendix 10).

$$\text{Normalised Value} = \frac{\text{Luminescence}}{\beta\text{–galactosidase}}$$

Figure 4.2. Equation used to normalise transient transfection data. Luciferase activity in the cell lysates was divided by the level of β–galactosidase activity to gain a normalised value.

Deletions were made from the 5' end of the maspin promoter; these truncations removed potential transcription effector binding sites from the promoter (figure 4.3 shows shortest three deletions and reported transcription effectors). All deletion constructs were cloned into the pGL3 Basic vector (Chapter 3), used in transient transfections, and the effect of the truncations on reporter gene activity was measured. Significant changes in reporter gene activity implicate the removed region in transcriptional regulation of the gene. In this way the relative importance of specific regions of the maspin promoter was determined.

Removal of a region of DNA which results in decreased reporter gene activity indicates that the transcriptional activity of the promoter has been reduced. Thus the region that has been removed may contain a putative activation site(s). If the activity of the promoter is increased then a repressor site(s) may have been removed. Using the transcriptional activity of the truncated promoter constructs, a number of important regions in the promoter were identified. The activity of the full length construct (-2551-maspin-pGL3) represents relative activity of all the activation and repression sites in the maspin promoter. Maspin promoter activity was also measured in different cell lines, such that the relative expression in each cell line could be compared.

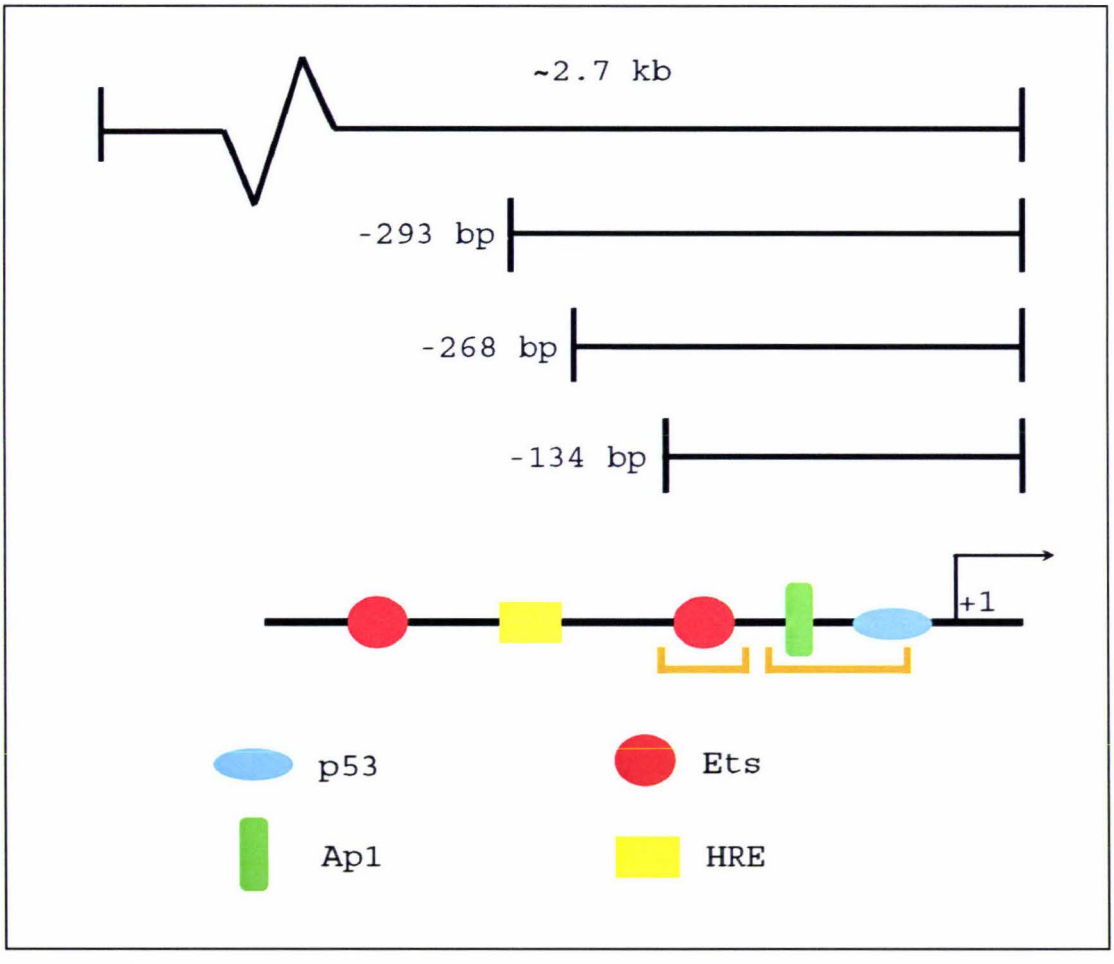


Figure 4.3. Reported elements within 500 bp of the human maspin TSS. Black line represents the promoter with bent arrow indicating transcription start site. Red, yellow, green, and blue shapes represent the location of the corresponding transcription effector sites reported to date (Zhang et al., 1997a; Zhang et al., 1997b; Zou et al., 2000). The shortest three deletions created for transfection deletion series are shown – indicating the elements present in each. Brown brackets indicate regions completely conserved between human mouse and rat (refer to Chapter 3). Note diagram is not to scale. For sequence information see figure 3.1.

4.2 HeLa cell transfections

4.2.1 Controls

Initial transfections were carried out in HeLa cells, varying the amount of full length construct, to determine the appropriate amount of DNA to be added to the cells (figure 4.4). Increasing amounts of full length reporter construct gave an almost linear response in HeLa cells. High levels of DNA can cause cell death, however this was not seen when using up to 2 μg of reporter construct and 1 μg of control vector. As a result of these transfections, 1 μg of both reporter construct and pSV β–galactosidase were used for all subsequent transfections in HeLa cells.

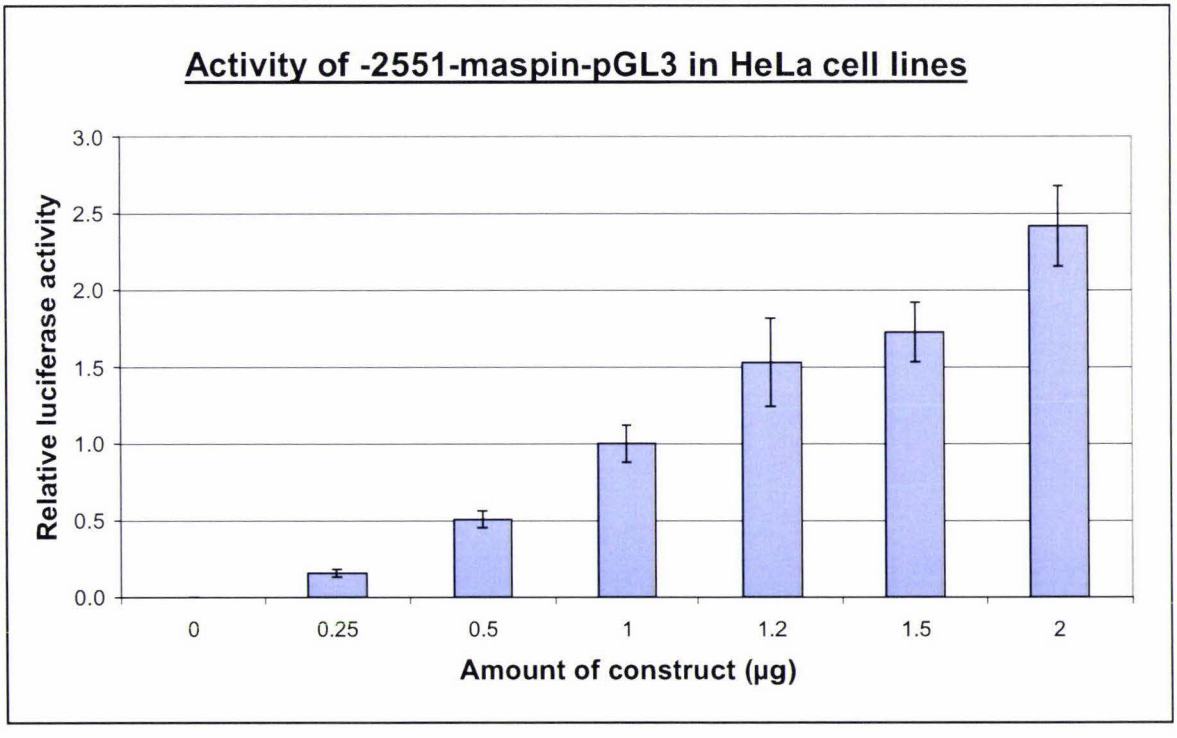


Figure 4.4. Effect of increasing transfected reporter construct in HeLa cells. Actively growing cells were transfected with -2551-maspin-pGL3 and pSV-β-galactosidase using a 3:2 ratio of FUGENE™:DNA. Transfections were carried out with increasing amount of reporter construct and 1 μg of pSV-β-galactosidase control vector. The cells were harvested 40-48 hours post-transfection and assayed for luciferase and β–galactosidase activity. Normalised results were graphed relative to 1 μg of reporter construct. Results are the average of at least 3 experiments. Error bars indicate average deviation of samples.

4.2.2 Transfections

Maspin is expressed in normal tissues of epithelial origin, and reportedly down-regulated in cancerous tissues (Zhang et al., 1997b; Futscher et al., 2002). HeLa cells are derived from a cervical carcinoma and so maspin expression in these cells was not expected. In fact high levels of endogenous maspin expression were seen (western blot in figure 1.1) as well as expression of the reporter gene (figure 4.5). For an example of data obtained from transient transfection experiments, and working refer to Appendix 9. Averages of transient transfection data, and statistical analysis is shown in Appendix 10.

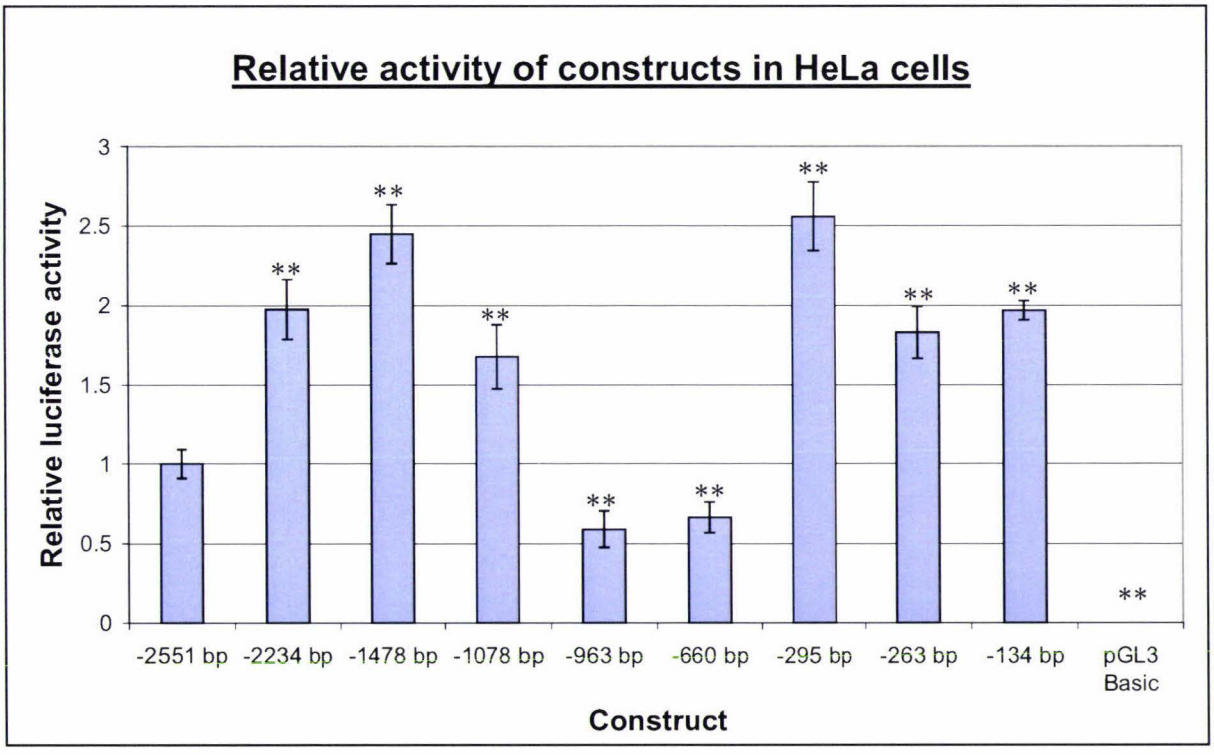


Figure 4.5. Relative activity of maspin promoter constructs in HeLa cells. Transfections were carried out using the FUGENE™ transfection reagent as described in Chapter 2. Cells were transfected with each construct in triplicate, and the level of reporter gene activity was expressed relative to the full length (-2551-maspin-pGL3) construct. Results shown are the average of at least 3 experiments. Error bars indicate average deviation of samples, and asterisks indicate a significant difference from the -2551 bp construct. ** <0.01.

As a general trend, constructs containing truncations from the 5' end of the maspin promoter showed an increase in reporter gene activity in HeLa cells, up to the -1478 bp deletion. At this point the promoter was approximately 2.5 fold more active than the full length construct. As this was a step-wise increase it indicates that there are several active repressor sites in the region of the promoter distal to the -1478 position. A highly significant ($P < 0.01$), approximately 5 fold, decrease in reporter gene expression was seen with the removal of the regions between -1478 bp and -963 bp, with no significant difference between -963 bp and -660 bp. At this point the reporter gene activity was approximately half that of the full length construct. This step-wise decrease between -1478 bp and -963 bp suggests the presence of positive effector sites. Removal of the region between -660 bp and -295 bp showed a 5 fold increase ($P < 0.01$). This indicates that a strong repressor site is present in this region. From this point there is a decrease to the shortest two constructs, which show just under a 2 fold increase when compared to the full length construct.

These results showed that the maspin promoter was active in HeLa cell lines. The shortest deletion (-134 bp) contains all activator sites reported to date (figure 4.3), and so the high levels of expression observed were expected. These results also indicated that there are regulatory sites present in the maspin promoter other than those reported to date.

4.3 Breast cell lines

Transfections were next carried out in MCF12a and MDA-MB-231 breast cell lines. MCF12a cells are derived from normal breast cells, while MDA-MB-231 cells are an invasive breast cancer cell line. These cells were chosen as they are both easy to grow, and MDA-MB-231 cells have previously been transfected with efficiency (Allen, 2003).

4.3.1 Transfections

The experiment shown in figure 4.4 was repeated in the MCF12a cell line to optimise the amount of DNA for use in transfections; 1µg of reporter construct and 1µg of

pSV-β-galactosidase was found to be suitable. The pSV-β-galactosidase vector did not produce measurable levels of β-galactosidase in MDA-MB-231 cells. The pCMV Sport-β gal vector was trialled and found to express β-galactosidase at easily measurable levels. In these cells 1 μg of reporter construct and 0.5 μg of pCMV Sport-β gal was used for all subsequent transfections. Using these ratios MDA-MB-231 and MCF12a cell lines were transfected with each of the deletion constructs (figure 4.6). To enable cell line comparison, MCF12a cells were transfected identically to, and at the same time as an MDA-MB-231 cell transfection. For tabular data refer to Appendix 10.

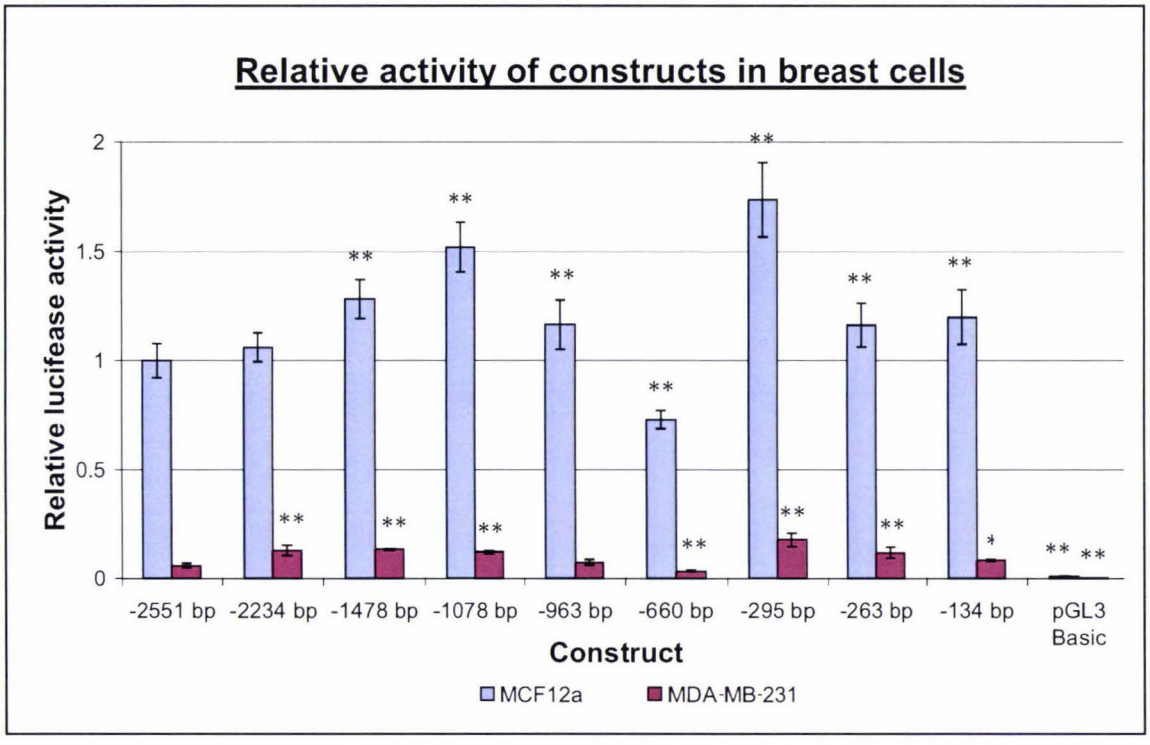


Figure 4.6. Relative activity of the deletion constructs in MCF12a and MDA-MB-231 cell lines. Cells were transfected using FUGENE™ transfection reagent as detailed in Chapter 2. Results are the average of at least 3 experiments and are expressed relative to the activity of the -2551 bp construct in MCF12a cells. Error bars represent the average deviation, and asterisks indicate a significant difference in relative luciferase activity from the -2551 bp construct in the corresponding cell line. * <0.05, ** <0.01.

As shown in figure 4.6 there is a large difference in reporter gene activity between MCF12a and MDA-MB-231 cell lines. MDA-MB-231 cells have been reported to express very low levels of maspin (Oshiro et al., 2003); this is supported by the western blot (figure 1.1) and the reporter gene assays (figure 4.6). Decrease/loss of maspin expression has been reported to be due to the loss of transactivation by p53, Ets or Ap1

elements, and/or methylation of the maspin promoter (Zhang et al., 1997a; Zhang et al., 1997b; Domann et al., 2000; Zou et al., 2000; Futscher et al., 2002; Maass et al., 2002; Oshiro et al., 2003).

MCF12a cells (figure 4.6) do not show the increase in expression with deletion up to -2234 bp as observed in HeLa cells (figure 4.5). This indicates that the negative effector site that is active in HeLa cell lines is not active in MCF12a cells. Interestingly this increase is present in MDA-MB-231 cells (for tabular results for each cell line chart refer to Appendix 10), indicating that this is not likely to be a factor specific to HeLa cells. HeLa cells showed a peak in expression at -1478 bp deletion with a significant ($P < 0.05$) decrease in expression to the -1078 bp deletion. In MCF12a cells the -1078 bp deletion shows peak activity, however in both MCF12a and MDA-MB-231 cell lines there is no significant difference between -1478 bp and -1078 bp deletions.

A large, step wise decrease in relative luciferase activity was observed between -1478 bp and -963 bp deletions in HeLa cell lines. This trend was similar for MDA-MB-231 cells, with the decrease continuing to the -660 bp deletion, however the decrease is less dramatic in MCF12a cells and was between -1078 bp and -660 bp deletions. The decrease in both breast cell lines between -1078 and -660 bp deletions is highly significant, as is the -963 bp to -660 bp comparison. This indicates that positive effector sites may be present in these regions, however the relative decrease is less in MCF12a cells, suggesting the sites may be less active in the normal breast cell line. A large increase was seen between -660 bp and -295 bp ($P < 0.01$) indicating that a repressor may be acting in this region of the maspin promoter. This deletion shows an increase in all cell lines that were tested suggesting that a ubiquitous repressor may bind within this region. The -263 bp, and -134 bp deletion constructs show a decrease in activity from the -295 bp construct in both breast, and HeLa cell lines, however both constructs show higher activity than the full length promoter. The activity observed in the shortest deletions indicates that the activators binding to this region of the promoter are sufficient to drive expression of maspin in these cell lines.

4.3.2 DNA methylation

As shown in the western blot (figure 1.1) and in reporter gene assays (figure 4.6) MCF12a cells express high levels of maspin, whilst MDA-MB-231 cells do not. The difference in maspin expression between normal and cancerous cells has been reported to be due in part to differential methylation of the proximal CpG island contained within the maspin promoter (Domann et al., 2000; Futscher et al., 2002; Maass et al., 2002; Oshiro et al., 2003). A distal CpG island has also been reported in the maspin promoter, this island is reported to be methylated in both maspin expressing and non-expressing cells (Domann et al., 2000). The proximal CpG island overlaps the transcription start site and the positive effector sites reported to date, but not the reported negative HRE (figure 4.7) (Domann et al., 2000). Methylation of the proximal CpG island would prevent the transactivation by the factors reported to bind to this region. This proximal island has been shown to be methylated in MDA-MB-231 cells, suggesting that this may be the cause for the down-regulation of endogenous maspin in these cells (Oshiro et al., 2003; Primeau et al., 2003). Other reports have suggested that the loss of activation by Ets, Ap1, or p53 may cause a down-regulation in maspin expression (Zhang et al., 1997a; Zhang et al., 1997b; Zou et al., 2000).

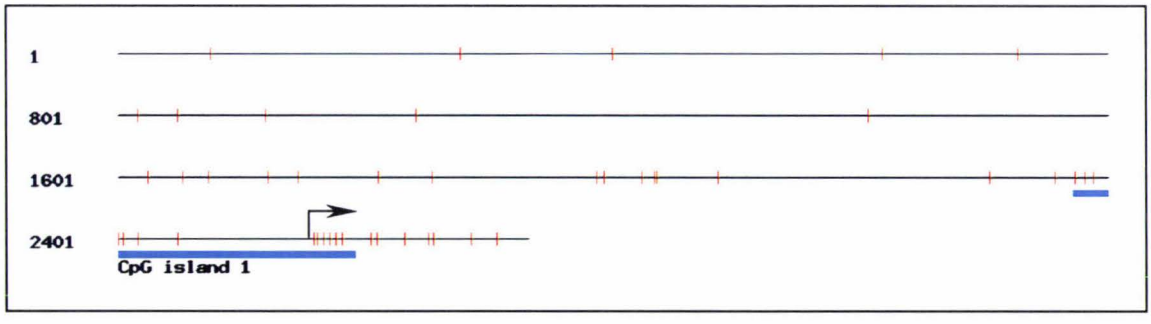


Figure 4.7. *In silico* prediction of a CpG island in the maspin promoter. Black lines indicate promoter sequence, red lines CpG di-nucleotides, and blue bar predicted CpG island. The transcription start site is located at base number 2551 (+1) (indicated by bent arrow); the CpG island is predicted to start at base 2372 (-178) and finish at base 2592 (+41). The reported distal CpG island was not predicted by this program. CpG island prediction from www.uscnorris.com/cpgislands/cpg.cgi.

The large difference seen in reporter gene assays between MDA-MB-231 and MCF12a cell lines (figure 4.6) could be due to loss of an activator/gain of a repressor; or due to methylation of the reporter construct. To differentiate between these possibilities, MDA-MB-231 cell lines were transfected with -2551-maspin-pGL3, and increasing

amounts of 5-azacytidine (methyltransferase inhibitor). As 5-azacytidine is a cytotoxic drug, it was added at 0 hrs and 24 hrs post-transfection to ensure cells were not being killed before expression of the reporter gene could take place. The cells were harvested at 40-48 hrs and the levels of reporter gene activity measured (figure 4.8).

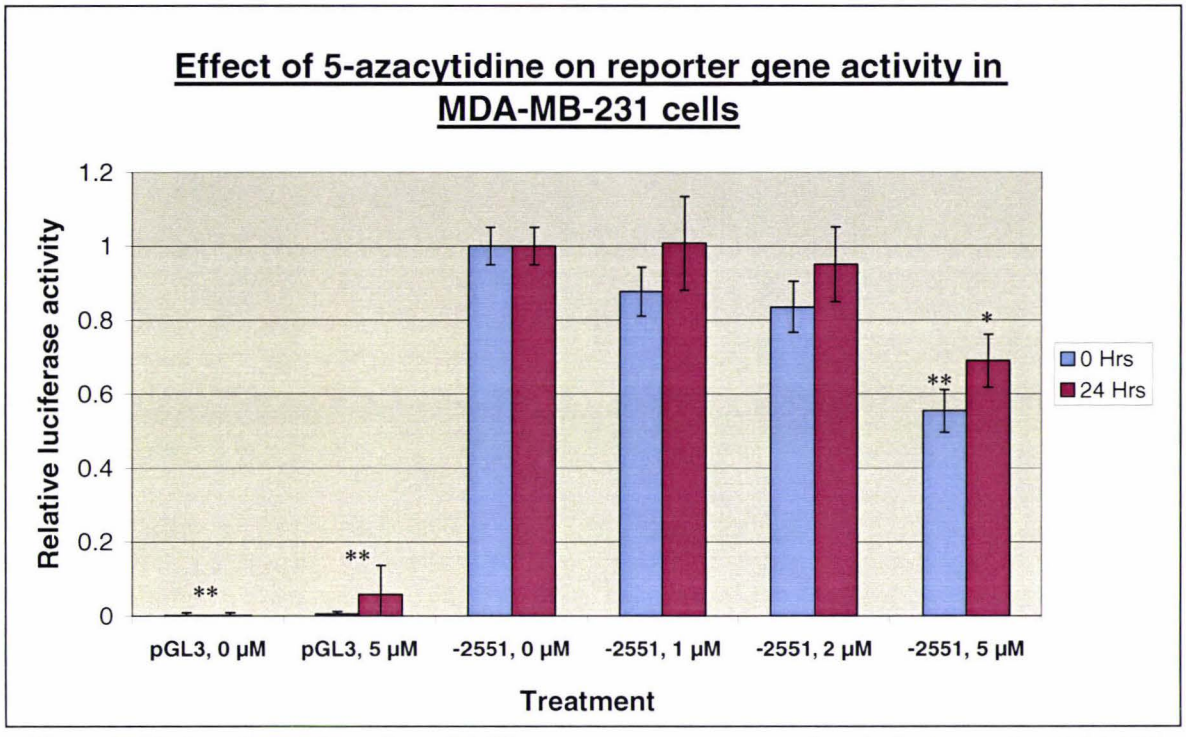


Figure 4.8. Effect of methyltransferase inhibitor on reporter gene expression. Actively growing MDA-MB-231 cells were transfected with -2551-maspin-pGL3 construct using FUGENE™ transfection reagent as detailed in Chapter 2. 5-azacytidine was added to the cells to a final concentration of 0 µM, 1 µM, 2 µM, or 5 µM at 0 hrs or 24 hours post-transfection. Cells were harvested at 40-48 hours and assayed for luciferase and β-galactosidase activity. Results are expressed relative to the activity of -2551 bp 0 µM 5-azacytidine control, and are average of at least 3 experiments. Error bars indicate average deviation, and asterisks indicate a significant difference from -2551 bp 0 µM 5-azacytidine. * <0.05, ** <0.01.

Addition of methyltransferase inhibitors to different non-expressing cell lines including MDA-MB-231 has shown an up-regulation of maspin expression, indicating that the endogenous maspin promoter is methylated, and that at least some of the required activating factors are present in these cell lines (Domann et al., 2000; Maass et al., 2002; Primeau et al., 2003). Therefore, if the down-regulation of reporter gene expression in transient transfections was due to the methylation of the promoter construct; then addition of a methyltransferase inhibitor should prevent methylation of the promoter construct, causing an increase in reporter gene expression.

As shown in figure 4.8, an increase in reporter gene activity was not observed upon addition of 5-azacytidine; instead a decrease in expression was seen with increasing amounts of the drug. This decrease was likely to be due to the drug being cytotoxic at high concentrations; this is supported by the samples with 5-azacytidine added at 24 hrs showing higher expression levels than those treated at 0 hrs. The lack of an increase in reporter gene activity indicates that the construct was not being methylated. Therefore, the difference observed between the cell lines in transient transfections was likely to be due to the loss of an activating, or gain of a repressive factor. Interestingly it has been reported that MDA-MB-231 cell lines are $p53^{-/-}$ (Oshiro et al., 2003). Thus the loss of p53 may be, in part, the cause of the decrease in reporter gene expression in MDA-MB-231 cell lines. However the presence or absence of the required Ets and Ap1 effectors is unknown in this cell line.

4.4 Prostate cell lines

Prostate cell lines CA-HPV-10 and PC3 were transfected with the maspin deletion series construct. Both cell lines are derived from prostate cancers, however in the nude mouse model, only PC3 formed tumours (www.atcc.org).

4.4.1 Transfections

It was determined that in both cell lines pCMV Sport-β gal was the best transfection control vector, and was used at 0.5 μg per well. The reporter construct was found to give the best results at 1 μg per well for CA-HPV-10 and 0.5 μg per well for PC3 cells. Using these ratios transient transfections of all deletion constructs were carried out in both prostate cancer cell lines (figure 4.9). To enable cell line comparison, PC3 cells were transfected identically to, and at the same time as a CA-HPV-10 cell transfection. For tabular data refer to Appendix 10.

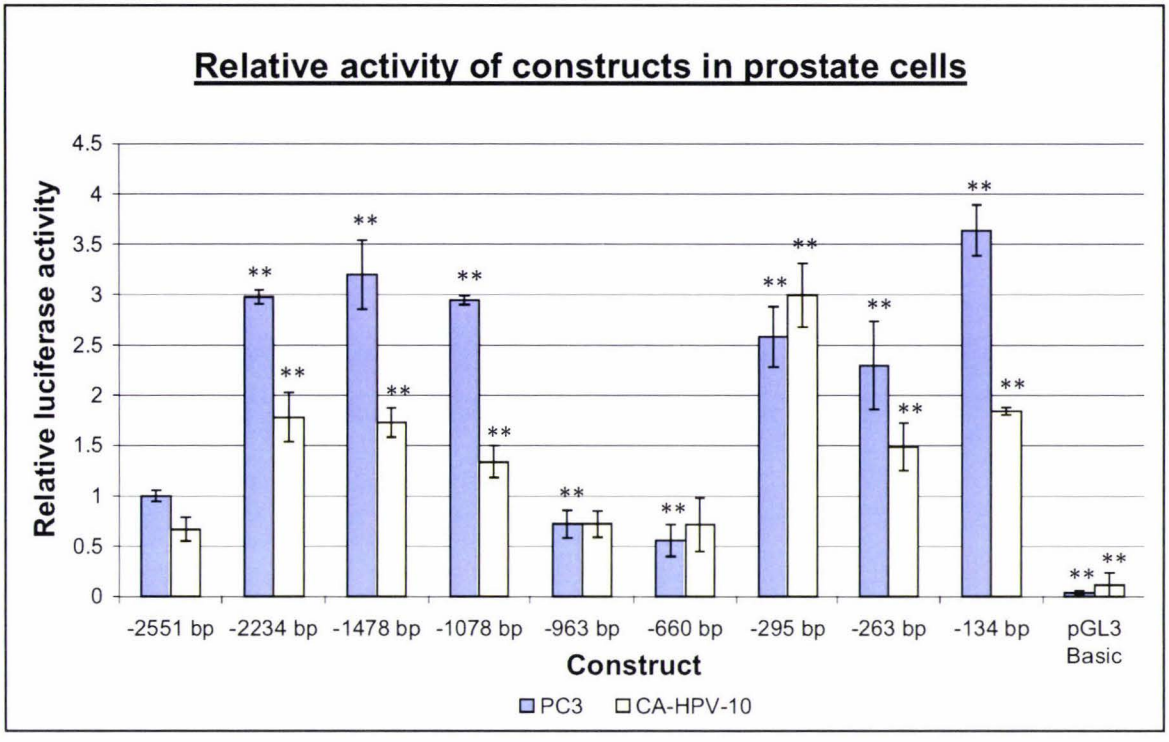


Figure 4.9. Relative activity of deletion constructs in CA-HPV-10 and PC3 cell lines. Actively growing cells were transfected using FUGENE™ transfection reagent as detailed in Chapter 2. Normalised values are expressed relative to the activity of the full length construct in PC3 cells. Results shown are the average of at least 3 experiments, and error bars indicate the average deviation of samples. Asterisks indicate significant difference from the full length construct in the relevant cell line. ** <0.01.

Whilst they are both cancerous cell lines, both CA-HPV-10 and PC3 cells express maspin (see fig1.1), and as shown in figure 4.9 the full length construct supports similar expression levels in both cell lines. Both cell lines also show similar trends, only differing in the relative increases/decreases between deletions. This suggests that regulatory mechanisms are similar in both cell lines. As with HeLa and MDA-MB-231 cells there is an increase with the removal of the most 5' region indicating a repressive factor binding site. The increase seen with the removal of this region is greatest in PC3 cells, suggesting that this factor may be more active or more highly expressed in these cells. The levels of expression stay relatively steady until the removal of the -1078 bp to -963 bp region, where levels of expression dropped significantly.

No significant difference was observed between the -963 bp and -660 bp deletion in both cell lines. This suggests that no transcriptional effectors bind in this region, or that equally balanced effector sites have been removed i.e. sites for effectors that activate

and repress transcription 2 fold have been removed, therefore no net difference in reporter gene activity would be seen. A significant ($P<0.01$) increase was again seen with the removal of the -660 bp to -295 bp region in both cell lines. The -263 bp construct shows a significant decrease ($P <0.01$) from the -295 bp deletion in CA-HPV-10 cells, followed by an increase in expression to the shortest deletion (-134 bp), which is significant in PC3 cells. Interestingly PC3 was the only cell line in which a significant difference between the shortest two deletions was observed.

4.5 Hormone response element

The region between -295 bp and -263 bp deletions contains a reported hormone response element (HRE) (figure 4.3) which is reportedly active in prostate cell lines, in response to the androgen receptor (AR) (Zhang et al., 1997b). This element repressed the transcription of a reporter gene in transfection assays independent of the addition of androgen hormone (Zhang et al., 1997b). Another report however, has indicated a role for androgen in maspin expression (Zou et al., 2002). This report indicated that in tumours, when androgen is present, maspin expression is decreased; suggesting androgen is required for the repression of maspin expression. *In vivo* regulation of maspin by androgen has also been demonstrated (Zou et al., 2002).

PC3 and CA-HPV-10 cells are both androgen insensitive (Bonaccorsi et al., 2000; Weijerman et al., 1998). Supporting this, an increase in reporter gene expression was not seen with the removal of the negative HRE, instead a decrease is observed (compare -295 bp and -263 bp constructs). These cells also do not show a decrease in maspin expression when grown in the presence of androgen (Bradbury, 2004). Attempts to duplicate electrophoretic mobility shift assay (EMSA) binding experiments performed by Zhang, *et. al* (1997b), showed weak binding to the HRE by androgen receptor indicating that there may be an interaction (van Dijk, 2003; Bradbury, 2004).

Addition of hydrocortisone, progesterone, oestrogen, or tamoxifen did not affect maspin expression in the normal breast cell line MCF12a (Bradbury, 2004). Interestingly however, tamoxifen has been shown to induce maspin expression in cancerous breast

cell lines, and increase maspin secretion in myoepithelial breast cell lines (Shao et al., 2000; Khalkhali-Ellis et al., 2004). The increase in maspin expression observed with tamoxifen treatment in breast cancer cell lines was abrogated by mutation of the maspin HRE (Khalkhali-Ellis et al., 2004). This suggests that tamoxifen activates maspin though the HRE. Tamoxifen is known to act though the oestrogen receptor β (ERβ), so it is likely that the ERβ also binds to the maspin HRE (Shao et al., 2000; Khalkhali-Ellis et al., 2004).

In electrophoretic mobility shift assays using the maspin HRE, glucocorticoid receptor (GR) showed weak binding, however oestrogen and progesterone receptors did not produce reproducible results using this technique (van Dijk, 2003; Bradbury, 2004). These reports indicate that the maspin HRE may be responsive to the androgen receptor; and tamoxifen through the oestrogen receptor β; however more experiments need to be carried out to confirm this. Co-transfections using the relevant maspin promoter constructs, receptors, and ligands may provide evidence of the functionality of the HRE to these treatments.

4.6 TESS interpretation of transient transfections

Transient transfections were used to identify important regions of the maspin promoter. The results from TESS (see Chapter 3 and Appendix 2) were then used to identify transcriptional effectors, which bind to these important regions, and could be responsible for the trends seen. TESS predicts transcriptional elements which could bind to a region of DNA by aligning the consensus sequence for a specific transcription factor to the DNA sequence to be analysed. Hence binding sites can used to predict potential transacting factors.

4.6.1 -2551 bp to -2234 bp region

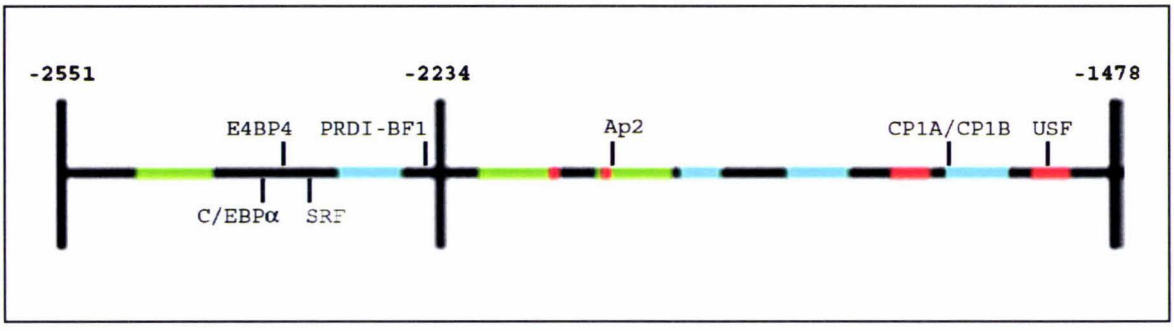


Figure 4.10. Schematic diagram of the -2251 bp to -1478 bp region. Solid horizontal line indicates DNA fragment, and solid vertical lines indicating end of deletion constructs. This region of the maspin promoter contains many regions conserved between human and mouse (blue), rat (red), or all three (green). Some of the putative effector sites predicted by TESS are indicated.

All cancerous cell lines showed a significant ($P < 0.01$) increase in reporter gene activity when the region between -2551 bp and -2234 bp was removed, suggesting the presence of a repressor site. This increase was not observed in the normal breast cell line MCF12a. This may indicate that a repressive protein which is up-regulated in cancerous cells, binds to a site in this region. This region contains a sequence highly conserved between human, mouse, and rat, and another region conserved only in rat and human (figure 4.10). Many factors were predicted to bind to this approximately 300 bp region; however TESS did not predict any transcriptional effectors to bind to either of these conserved regions. Potential factors predicted by TESS included PRDI-BF1, SRF, and E4BP4.

PRDI-BF1 also know as Blimp1 (B-lymphocyte-induced maturation protein 1) is a transcriptional repressor which is required for B-cell differentiation (Yu et al., 2000). Expression of this factor is limited to B-cells, and so this factor is unlikely to be responsible for the trend seen in transient transfections (Yu et al., 2000).

E4BP4 (E4 Binding Protein 4) is a bZIP negative regulator of transcription (Cowell, 2002). E4BP4 is believed to be part of the mammalian circadian clock, as well as playing a roll in prevention of apoptosis (Cowell, 2002). As maspin sensitises cells to apoptosis, and E4BP4 prevents apoptosis, it is possible that E4BP4 binds to this region and represses maspin transcription.

SRF (Serum Response Factor) is a transcriptional activator of many genes involved in cellular growth and differentiation (Chai and Tarnawski, 2002). As an activator of transcription this factor is unlikely to be responsible for the increased reporter gene expression seen with the removal of the -2551 bp to -2234 bp region. C/EBPα was also predicted to bind to this region of the maspin promoter, however this is also an activator of transcription and so is unlikely to be the factor responsible for the trend seen (Milde-Langosch et al., 2003).

4.6.2 -2234 bp to -1478 bp region

The -2234 bp to -1478 bp region of the maspin promoter contains many regions which show conservation between human, mouse, and or rat (figure 4.10). Interestingly there are two small regions closely spaced which are conserved in all three organisms. Even though there are many conserved regions contained within this fragment, there is little difference in reporter gene expression, however there is a significant difference between these deletions in MCF12a, and HeLa cell lines ($P < 0.01$ and $P < 0.05$ respectively). The lack of difference in reporter gene assays may indicate that nothing binds to this region of the maspin promoter, or that equally balanced activators and repressors bind to this region. The significant difference shown in MCF12a and HeLa cell lines may indicate that an activator binds to this region. To test this, further deletions of this approximately 800 bp region, including separation of the two closely spaced conserved regions, and transfection of these deletions into cell lines needs to be carried out.

4.6.3 -1078 bp to -963 bp region

When the -1078 bp to -963 bp region is removed a decrease in activity is observed, indicating the removal of an activator site. This difference is significant in CA-HPV-10, PC3, and MDA-MB-231 cell lines (Appendix 10). No conserved regions are contained within this region (figure 4.11). GR, PR and c-MYC are transcriptional activators that were predicted by TESS to bind to this region of the maspin promoter (figure 4.11). No difference was seen in maspin expression when MCF12a cells were grown in presence of progesterone or hydrocortisone (Bradbury, 2004), suggesting that PR and GR are not responsible for the trends shown in the transient transfections. c-MYC is an oncogene, which is known to regulate cell growth, differentiation, apoptosis, genomic instability,

and angiogenesis (Oster et al., 2002). C-MYC is known to induce apoptosis, and as an oncogene, it is often over-expressed in cancers. Little is known about the downstream targets of c-MYC in apoptosis, and so it is possible that c-MYC may activate maspin as part of its apoptotic induction (Oster et al., 2002).

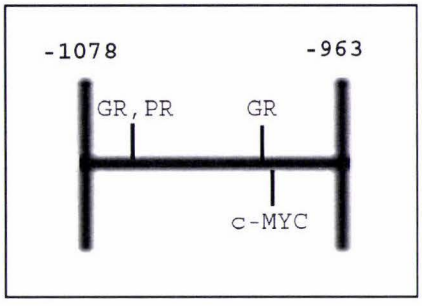


Figure 4.11. Schematic diagram of the -1078 bp to -963 bp region. No conservation between human and mouse or rat was observed within this region of the promoter. Putative transcriptional activators that are predicted by TESS to bind to this region of the maspin promoter are indicated.

4.6.4 -660 bp to -295 bp region

In all cell lines, removal of the region between -660 bp and -295 bp showed a significant ($P < 0.01$) increase in reporter gene activity, thus indicating the presence of a repressor element in this region. The increase in reporter gene activity from the -660 bp construct ranged from 238% in MCF12a cells, to 582% in MDA-MB-231 cell lines (see table 4.1).

Cell Line	Expression level		% increase
	-660 bp	-295 bp	
HeLa	0.66	2.56	385%
MDA-MB-231	0.53	3.07	582%
MCF12a	0.73	1.74	238%
Ca-HPV-10	1.08	4.49	418%
PC3	0.56	2.58	463%

Table 4.1. Change in reporter gene expression between -660 bp and -295 bp constructs. All cell lines were transfected with the -660 bp and -295 bp deletion constructs, and the level of reporter gene expression measured. The level of expression of both constructs was normalised to the -2551 bp maspin construct in the relative cell line. The increase in expression with the removal of the -660 bp to -295 bp region was significant ($P<0.01$) in all cells, and percent increase is shown.

TESS predicted many transcriptional elements within this ~360 bp region of the maspin promoter; however most of these were not repressor sites. Two transcriptional repressors were predicted to bind to this region (figure 4.12); F2F is a protein isolated from rat, and is 'most likely a transcriptional repressor' (www.cbil.upenn.edu/tess). Presence of the F2F binding site has been reported in other human promoters, however functionality of this site has not been tested (Minami et al., 1998; Schoonderwoert and Martens, 2002).

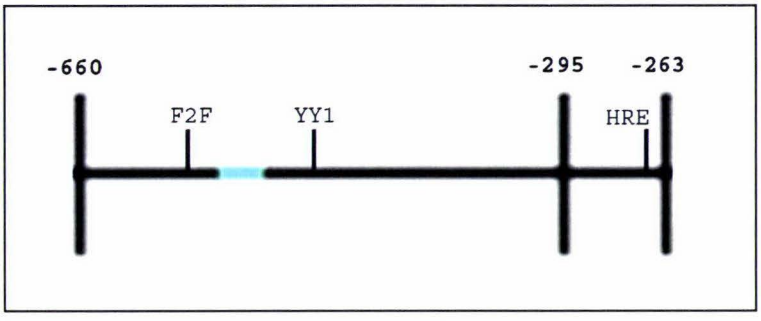


Figure 4.12. Schematic diagram of the -660 bp to -263 bp region. Conservation of one region of this fragment of the maspin promoter was observed between human and mouse (blue), and no conservation between human and rat was observed. Putative repressor sites predicted by TESS, and the previously reported HRE are indicated.

YY1 (Yin and Yang 1) also known as NMP-1 (Nuclear Matrix Protein 1) was predicted to bind to the -660 bp to -295 bp region of the maspin promoter. YY1 is a ubiquitous zinc finger transcription factor whose default function is as a repressor (Thomas and Seto, 1999). YY1 is known to regulate many genes including the oncogene c-MYC and tumour suppressor gene p53 (Thomas and Seto, 1999).

It is possible that YY1, or the human F2F bind to the maspin promoter within the -660 bp to -295 bp region and repress transcription. Functional assays with YY1 and human F2F would need to be carried out to test this hypothesis. A 12 bp region was found to be conserved between the mouse and human maspin promoters (figure 4.12). Although no transcriptional effector sites were predicted by TESS to bind to this region, it is possible that an unknown factor represses transcription at this conserved region.

4.6.5 -295 bp to -263 bp region

The region between -295 bp and -263 bp shows a significant decrease in reporter gene expression in HeLa, MDA-MB-231, and CA-HPV-10 cell lines (Appendix 10), suggesting the presence of an activating factor. TESS did not predict any transcription factors to bind to this region of the maspin promoter, and no conserved regions were identified (figure 4.12). This region contains the reported HRE (Zhang et al., 1997b) which was not predicted by TESS. The activity of this site could not be tested in either PC3 or CA-HPV-10 cells without co-transfections with the androgen receptor, as both cell lines lack this receptor. This is a small region of the maspin promoter (~30 bp), and may be small enough for use in EMSA to determine if a protein binds to this region, which may account for the significant decrease observed.

4.6.6 -134 bp to transcriptional start site

This region of the maspin promoter contains all the transcriptional activators reported to date. Supporting this, the -134 bp maspin promoter construct shows high levels of expression in all cell lines tested. These reported activation sites are conserved in all three species (figure 4.13), and so are likely to be important for the regulation of the maspin gene in all species. The conservation of the regulatory sites reported to date in all three species suggests that the other conserved regions of the maspin promoter may be relevant to the regulation of the gene.

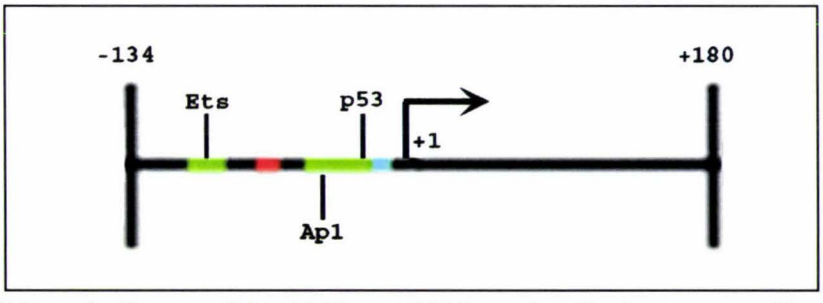


Figure 4.13. Schematic diagram of the -134 bp to +180 bp region. Regions conserved between human, mouse, and rat are shown in green, and regions conserved in human and rat are shown in red, conservation between human and mouse is shown in blue. Previously reported transcription activator sites that are present within this region are indicated.

4.6.7 TESS summary

The results of TESS analysis of the maspin promoter, combined with mouse and rat alignments with the human promoter, have been used to analyse transient transfection results. Some putative transcriptional elements were identified by TESS in the regions defined by transient transfections as being important for the regulation of maspin expression.

YY1 and F2F were identified by TESS as repressor proteins which bind to the -660 bp to -295 bp region of the maspin promoter. These effectors may be responsible for the increase in reporter gene expression seen with the removal of this region. c-MYC was the only factor predicted by TESS to bind to the -1078 bp to -963 bp region that is likely to be responsible for the decrease in reporter gene expression observed when this region is removed.

E4BP4 is the most likely factor predicted by TESS to bind to the maspin promoter and repress transcription in the -2551 bp to -2234 bp region. E4BP4 plays a role in the prevention of apoptosis, and thus repression of maspin expression could help to explain this function of the protein. This region showed significant differences only in cancerous cell lines; so it is possible that the factor responsible for the reporter increase, with removal of the region, is more active or up-regulated in cancerous cells.

No conserved regions or potential effector sites were identified within the -295 bp to -263 bp region, yet 3/5 cell lines showed a significant decrease in reporter gene activity. This decrease suggests that there is an activation site present in this region. This region also contains the reported HRE.

4.6 Chapter summary

Overall the trends in transcriptional regulation of maspin are similar in all cell lines tested indicating that the major effectors are present in each of these cells. It has been suggested that the major cause of maspin down-regulation in cancerous cells, and in non-expressing tissues is due to methylation of the promoter (Domann et al., 2000; Futscher et al., 2002; Maass et al., 2002; Oshiro et al., 2003; Sato et al., 2003). Transfections in the presence of 5-azacytidine showed that this was not the cause of down-regulation of the reporter construct in MDA-MB-231 cell lines. The endogenous maspin promoter in these cells has been reported to be methylated (Oshiro et al., 2003), however since the reporter construct did not appear to be methylated, other causes are more likely to be responsible for the down-regulation of the reporter gene. This decrease in expression could be due to the reported mutation of p53, which has been reported to activate maspin transcription. The decrease could also be due to the over-expression of a repressor. Removal of this factor would be expected to raise expression levels to that of normal breast cell lines. This is not seen in transfection of the deletion series, so if this is the case the repressor must be active in the shortest deletion. The results shown do not distinguish between these possibilities, however further deletion of the -134 bp region, and transfection of these constructs may enable the differentiation of these hypotheses. *In vivo* footprinting (discussed in Chapter 5) may also be able to distinguish between these models.

Interestingly all cancerous cell lines showed an increase in reporter gene activity when the region between -2551 bp and -2234 bp was removed, indicating the presence of a repressor. This region contains a sequence highly conserved between mouse, rat, and human (figure 4.10). An unknown repressor protein may bind to this conserved region; alternatively E4BP4 may be responsible for the repression seen in this region. This putative repressor may be up-regulated in cancers, however this needs to be further investigated in more cancerous and normal cell lines. If this putative repressor is up-regulated, it may also be a factor in the down-regulation of maspin in many cancers.

A significant increase in reporter gene activity was observed between -660 bp and -295 bp in all cell lines transfected. This indicates the presence of an as yet unidentified repressor site. As this trend was seen in all transfected cell lines, the protein that recognises this site may be a ubiquitous repressor. YY1 is a ubiquitous protein which can act as a transcriptional repressor. This protein and F2F were both predicted to bind to this region of the maspin promoter.

Another ubiquitous factor may bind to the region between -1078 bp and -963 bp. This region shows a decrease in all cell lines tested, and so may contain a putative activator site. TESS predicted the activator c-MYC to bind to this region.

Transfection of a promoter deletion series is the first step in functional promoter analysis. These transfections have identified several important regulatory regions of the maspin promoter. The results of TESS predictions and mouse and rat alignments with the human maspin promoter were used to identify putative transacting factors; however the functionality of these putative effector sites needs to be tested. To address this *in vivo* DNaseI footprinting was attempted as the next step in the analysis. *In vivo* DNaseI footprinting can be used to predict exact protein binding sites on a specific region of DNA. Therefore *in vivo* DNaseI footprinting has the potential to define DNA consensus sequences which in turn may be used to predict DNA binding proteins with functional roles in regulation of the maspin gene. The role of these regulators can then be further investigated using co-transfections, mutation analysis, and electrophoretic mobility shift assays.

CHAPTER 5 - *In vivo* DNaseI Footprinting

5.1 Introduction

Transient transfections and reporter gene assays were used to identify regions of the maspin promoter of potential importance in transcriptional regulation. The next step was to refine these regions, and identify individual regulatory factors which may act on the maspin promoter. *In vivo* DNaseI footprinting was used in an attempt to identify important sequences at which regulatory proteins bind with functional significance. Footprinting is a technique in which DNA is randomly cleaved and amplified in a way that when separated on an acrylamide gel, a ladder of fragment sizes is seen (see figure 5.1). DNaseI cleaves only accessible DNA so that any DNA bound to protein will be protected from cleavage. Therefore gaps appear in the ladder in regions where proteins associate with DNA. By running a sequencing ladder beside the footprinting samples, the sequence which is protected from cleavage can be determined. It may then be possible to predict the identity of the bound proteins by analysis of the cognate DNA elements.

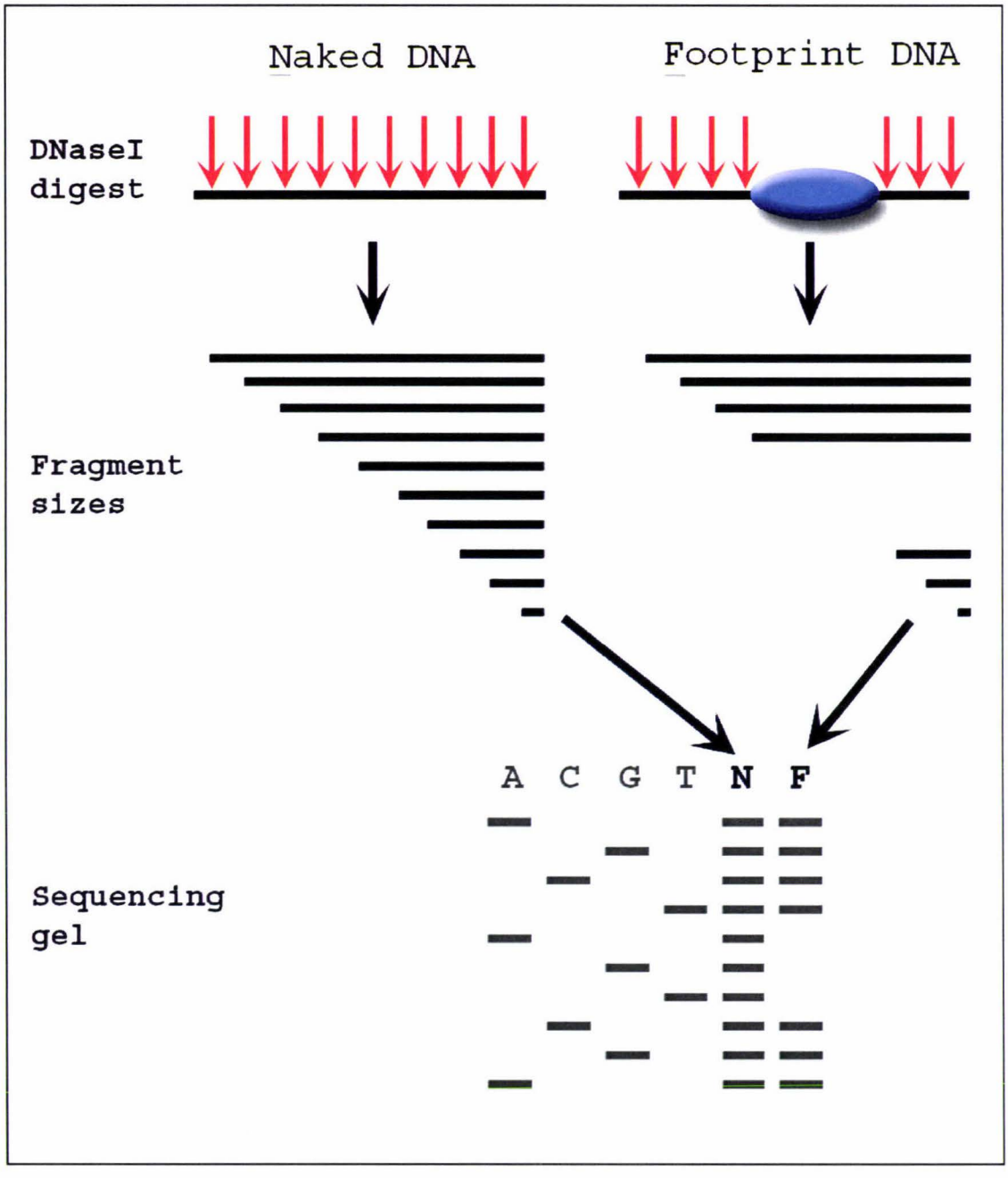


Figure 5.1. Schematic diagram of DNaseI footprinting. Naked DNA and DNA with bound proteins (Footprint DNA) is incubated with DNaseI, resulting in random cleavage of the DNA. Bound proteins protect DNA from DNaseI cleavage, resulting in gaps in the ladder when run on an acrylamide gel. A sequencing ladder is run along side the footprinting samples. This enables the identification of the sequence at which proteins are bound to the DNA. In the example shown the protein is bound to the sequence "AGT".

DNA is indicated by black lines, red arrows indicate DNaseI cleavage sites, blue ellipse indicates DNA binding protein, and the grey box shows acrylamide gel. A = adenine, C = cytosine, G = guanine, and T = Thymine. N stands for Naked DNA and F for Footprinting DNA.

5.1.2 *In vivo* DNaseI footprinting technique

The protocol used for *in vivo* DNaseI footprinting was modified from Carey and Smale, (2000); and Riggs et al, (1998) and was originally trialled by Dr Kirsty Allen (Allen, 2003). The technique involves the isolation and DNaseI digestion of intact nuclei from tissue culture cells. This results in DNA fragments which are 300 – 600 bases when made single-stranded. As shown by figure 5.2, primer extension is then carried out on this digested DNA using a biotinylated primer. This is followed by ligation of a double-stranded linker, and capture of the ligated product. The ligated product is melted and the 'old' strand is collected. PCR is then carried out with internal and linker primers. The resulting samples are then separated on an 8% acrylamide sequencing gel in the presence of urea. Bands are visualised by standard Southern blotting and hybridisation using a radioactively labelled specific probe.

5.1.3 *In vivo* versus *in vitro* footprinting

The technique used was designated *in vivo* as the DNaseI digestion was carried out on intact nuclei extracted from tissue culture cells, as compared to the *in vitro* technique, in which the digestion is carried out on a radio-labelled PCR product which is incubated with whole cell extracts. An *in vitro* footprint is unlikely to show differences between cell lines, unless there are major differences in protein complement e.g. total absence of a normally ubiquitous factor. For example, the same transcription factors may be present in two different cell lines, however in one cell line the sequence which the factor binds may be methylated and/or wrapped up in nucleosomes, and so inaccessible to the transcription factor. *In vitro* footprinting would show no difference between the cell lines in this situation as a PCR product is used; however *in vivo* footprinting would show the transcription factor absent in the methylated/nucleosome bound cell line. *In vitro* footprinting shows what can bind to a region of DNA, whilst *in vivo* footprinting shows what is actually bound, and therefore is more likely to have functional significance.

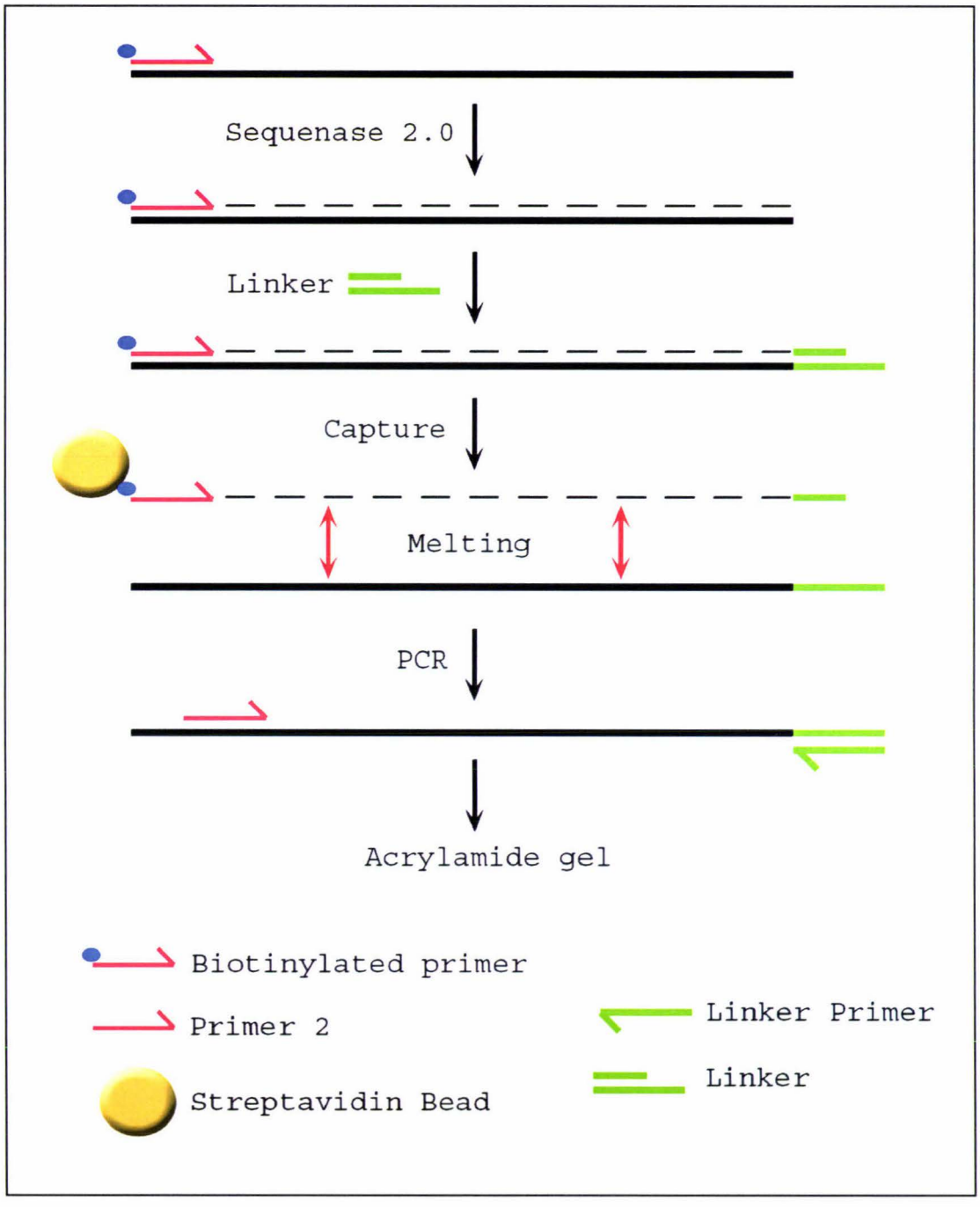


Figure 5.2. Schematic diagram of the *in vivo* footprinting protocol. A biotinylated primer is annealed to purified DNaseI digested DNA. This is then extended by Sequenase enzyme resulting in a blunt-ended product. A linker is then ligated to the product, followed by capture of the product using streptavidin beads, and denaturation of the product in alkaline conditions. The original strand is collected and used in PCR with primer 2, which overlaps the binding site for the biotinylated primer, and linker primer. Black lines and dashes indicate original and newly synthesised DNA respectively.

5.2 *In vivo* footprinting method

5.2.1 DNaseI digestion

Nuclei were extracted from actively growing cells using an NP-40 based lysis buffer, containing spermine and spermidine to stabilise the nuclei. The nuclei were then incubated with DNaseI in the presence of Ca^{2+} at 37°C. DNaseI introduces essentially random single-stranded nicks into the DNA, except where the DNA is protected by proteins. The DNA was then extracted, ethanol precipitated, and separated on an alkaline agarose gel. Alkaline gel electrophoresis causes double-stranded DNA to 'melt' into single-strands, and therefore the relative size of the single-strands can be determined. Single-strand sizes of approximately 600 - 300 bases are required for DNaseI footprinting. Figure 5.3 shows an example of DNaseI digested genomic DNA separated on an alkaline agarose gel.

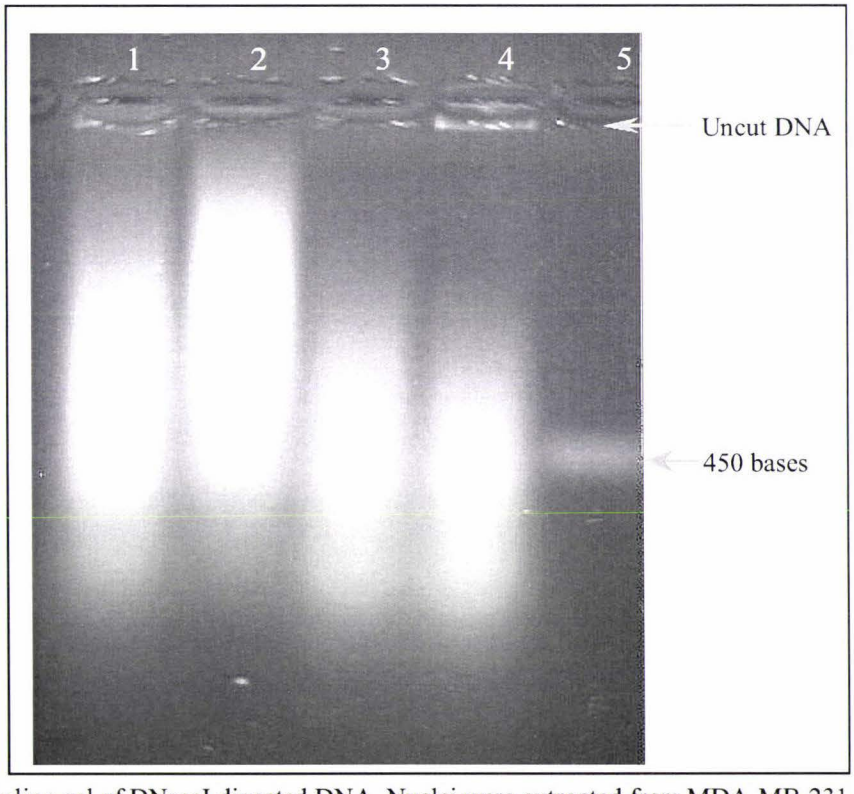


Figure 5.3. Alkaline gel of DNaseI digested DNA. Nuclei were extracted from MDA-MB-231 cells (Chapter 2.5.1), and incubated with 4.5 µg DNaseI for 5, 6, or 7 minutes (Chapter 2.5.2). Digests were stopped by addition of EDTA, SDS, and proteinase K and incubated overnight at 37°C. DNA was purified by phenol:chloroform extraction and RNase digestion, followed by ethanol precipitation. Samples were then electrophoresed on a 1.5% agarose gel under alkaline conditions at 10 V overnight. The gel was then neutralised and stained with ethidium bromide. 1 = 5 minute, 2-3 = 6 minute, 4 = 7 minute DNaseI digestion. 5 = 450 base PCR marker.

As demonstrated by figure 5.3 the DNaseI digestion was highly variable. Samples 2 and 3 were treated the same, yet showed large differences in amount of DNA digestion. This may be due to more cells being present in sample 2, therefore more DNA resulting in less overall digestion. Consequently all DNaseI digested DNA was analysed by alkaline agarose gel electrophoresis to ensure optimal digestion before subsequent footprinting analysis. Sample number 3 (figure 5.3) was considered to be adequately digested and used in footprinting reactions. Sample 4 contained too much uncut DNA to be used for footprinting reactions.

5.2.2 Amplification of footprint

In vivo DNaseI footprinting makes use of ligation mediated PCR (LMPCR). This allows for highly specific amplification of DNA fragments. The use of three nested primers in both directions (figure 5.4) enables footprinting of both DNA strands. Primers were designed so the 5' end overlapped the 3' end of the previous primer. This nested arrangement of primers increased the specificity for the region of interest, and overlapping meant that the primers were competing for the same binding site, therefore in the PCR reaction; the binding of primer 2 would prevent any remaining primer 1 binding, thus preventing the appearance of aberrant bands.

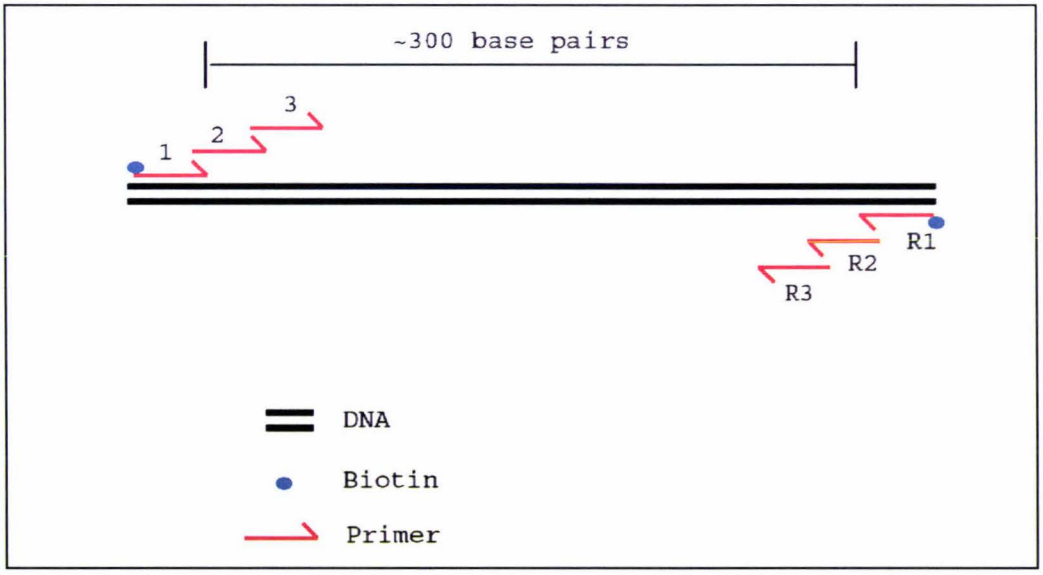


Figure 5.4. Schematic diagram of primers used in LMPCR. Primer 1 (1) and reverse 1 (R1) were used in primer extension. Primer 2 (2) and reverse 2 (R2) were used in PCR along with the linker primer. Primer 3 (3) and reverse 3 (R3) were used to create radio-labelled probe. Primers 1 and reverse 1 were biotinylated to enable capture and enrichment of extended/ligated product.

Using the arrangement of primers shown in figure 5.4, four sets of primers were designed to footprint four regions of the maspin promoter that were considered to be important for regulation, as determined by transient transfection (Chapter 4) (primer sequences are shown in Appendix 4). These regions were the -2551 bp to -2235 bp region which, when removed, showed an increase in reporter gene expression in 4/5 cell lines. The -1078 bp to -963 bp region which showed a decrease in 4/5 cell lines. An increase in all cell lines was seen in the -660 bp to -295 bp region, and so was considered important. The -134 bp to TSS was also considered important as this showed high levels of expression in all cell lines, and contains all positive effectors reported to date.

Primer extension using primer 1 of each primer set was carried out on both naked digested, and footprint DNA using Sequenase version 2.0 (USB). Extension was performed at 48°C to limit the addition of adenine to the 3' end of the product (Riggs et al., 1998). The blunt-ended products were then ligated to a double-stranded oligonucleotide linker. The ligation of the linker created a defined end from which PCR could be carried out using primer 2 (figure 5.2 and 5.4) together with a linker-specific primer.

DNaseI digested DNA can give a high background in footprinting reactions, which may be due to the 3'OH groups left by DNaseI acting as random primers in the PCR (Riggs et al., 1998). To avoid this complication, the extension product was captured. Primer 1 was biotinylated at the 5' end, allowing the extension product to be captured and purified from the remaining DNA using streptavidin coated magnetic beads (Dynal Biotech). Biotin binds to streptavidin in 1 M salt, and is stable at high pH. Once bound to the beads, the strands of the extended, ligated products were melted by the addition of sodium hydroxide. The newly synthesised strand (from primer extension) was left attached to the beads, and the original template strand, with ligated linker, was collected and precipitated. PCR was then carried out on the precipitated DNA (figure 5.2). This PCR amplifies all the different fragments, making use of the ligated linker as a fixed end. A *Taq* DNA polymerase booster mix was used to ensure that all products contain a 5' terminal adenine. This avoids the appearance of double bands on the final gel.

5.2.3 Detection of footprint

The amplified footprinting samples were separated on a denaturing acrylamide gel. The samples were then transferred to a nylon membrane using semi-dry electroblotting. The DNA was crosslinked to the membrane which was then hybridised overnight with a labelled probe. The probe was created by PCR amplification of the footprinted area using primer 3 and reverse 3 (figure 5.4). The use of both primer 3's in production of the probe allows detection of both forward and reverse footprinting reactions on the same membrane. Figure 5.5 shows an example of the products of this PCR reaction. The samples were pooled, column purified, and gel quantified. Approximately 50 ng of probe DNA was radio-labelled using Ready-to-go™ DNA labelling beads (Amersham). Labelled DNA was then purified using Probe Quant™ G-50 micro columns (Amersham) and added to the hybridisation buffer. Hybridisation was carried out at 60°C overnight. The hybridised membrane was washed, dried, and detected using either X-ray film (Kodak) or phosphoimaging.

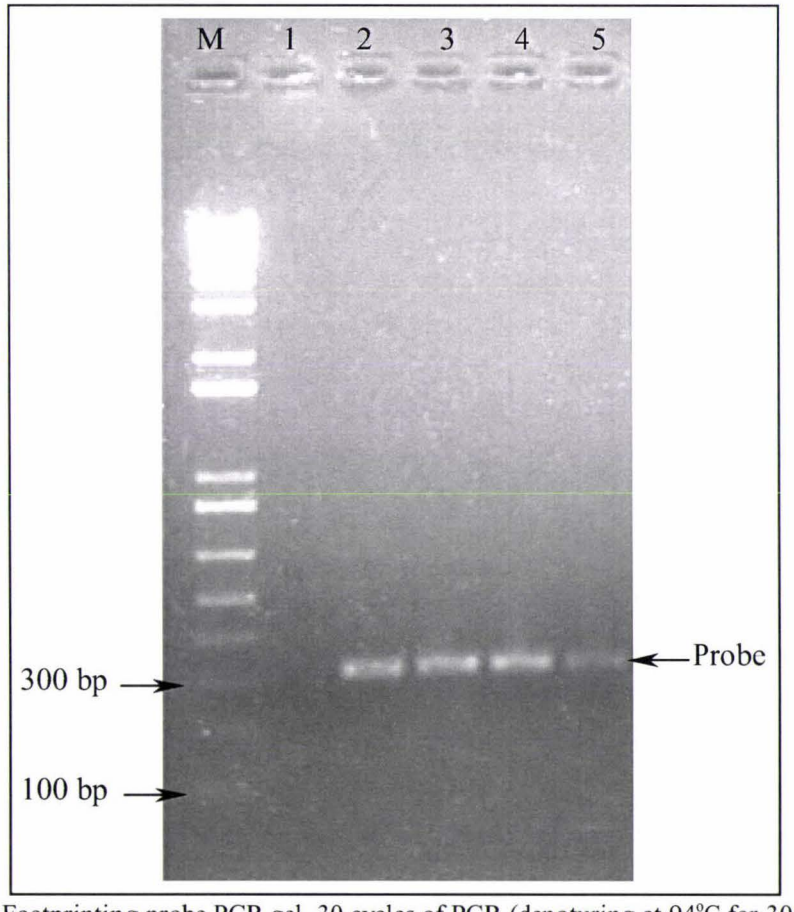


Figure 5.5. Footprinting probe PCR gel. 30 cycles of PCR (denaturing at 94°C for 30 seconds, annealing at 55°C for 30 seconds, extension at 74°C for 30 seconds) were carried out with primer 3 and reverse 3, in a 50 µL volume, using *Taq* DNA polymerase. Samples were then separated on a 1% agarose 1 x TAE gel stained with ethidium bromide. M = 1 kb plus DNA ladder, 1 = negative control, 2-5 = 5 µL of each PCR.

5.3 Controls

Several attempts using primer sets for four different regions of the maspin promoter, and DNA from four different cell lines (HeLa, MDA-MB-231, MCF12a, PC3), failed to result in a detectable footprint, and so controls were designed to test each step in the protocol.

5.3.1 Ligation

Blunt-end ligations can be inefficient, and so this was the first step of the protocol to be tested. The ligation was set up in identical conditions to those used in the footprinting protocol (refer chapter 2.5.6), except a PCR product partially digested with the blunt-end cutter *Hae* III was used instead of primer extended DNA. The gel shown in figure 5.6 shows a decrease in mobility of the linker ligated samples compared to the unligated sample. This indicates that the double-stranded linker has been ligated onto the blunt-end product. Interestingly neither a smear nor a double-band is observed; instead a definite band of lower mobility is seen in the ligated samples, indicating that all products were efficiently ligated in the reaction. The uncut PCR product has not shown a decrease in mobility, this is likely due to *Taq* DNA polymerase adding an adenine base overhang to the 3' ends of the PCR products, therefore preventing blunt-end ligation. A band of high mobility appears at the bottom of the gel in figure 5.6 in the ligated samples only; this band may be the 30 bp fragment which has been ligated to the linker, therefore decreasing its mobility.

This result indicated that the linker preparation was successful, and that the protocol for the ligation of this linker was functioning efficiently. Thus this was not the cause of the lack of bands on the final footprinting gel.

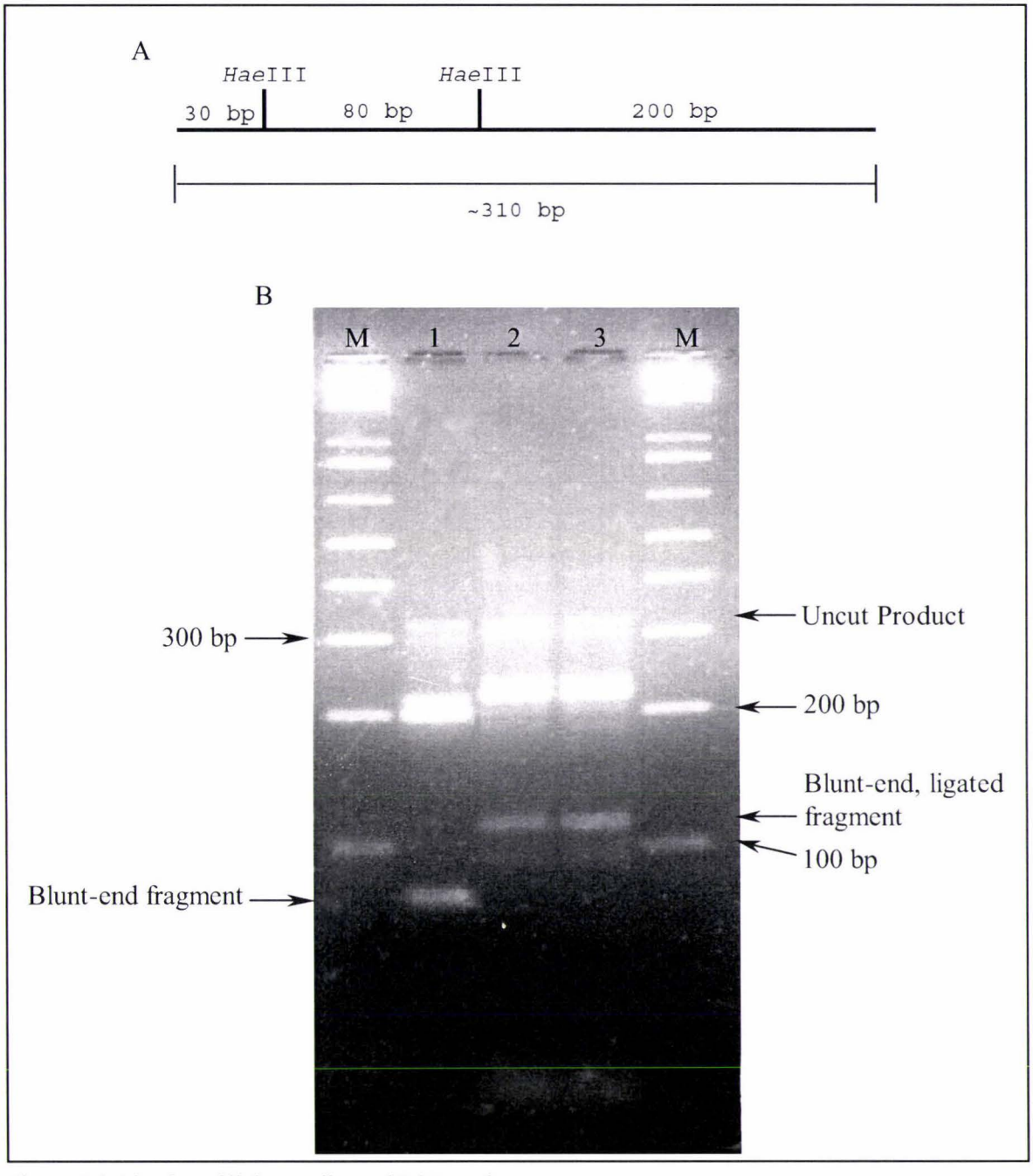


Figure 5.6. Ligation of linker to digested PCR product.

A. Schematic diagram showing where the blunt-end cutter *Hae* III cuts the PCR product.

B. The approximately 310 bp PCR product was partially digested with *Hae* III, leaving blunt-ended fragments. The digested fragments were then column purified (Qiagen PCR cleanup kit), ligated to linker overnight, and ethanol precipitated. Resulting DNA was electrophoresed on a 3.5% Nuseive agarose 1 x TAE gel stained with ethidium bromide, for ~3 hours at 60 V. M = 1 kb plus DNA ladder. 1 = cut unligated DNA, 2-3 = ligated DNA.

5.3.2 Primer extension

After confirming that the ligation step of the footprinting protocol was working the primer extension step was tested. Primer extension could have failed if the primer had not annealed to the DNA, or was not being extended to the end of the DNA, or if the enzyme did not leave blunt-ends. Primer extension must create a double-stranded, blunt-end extension product for ligation of the linker to occur. Several enzymes were trialled; Sequenase version 2.0 (USB) is the enzyme recommended in the protocol, but the *Klenow* fragment of *E.coli* DNA polymerase I (*Klenow*) (Roche), and T4 DNA polymerase (Roche) were also trialled. *Taq* DNA polymerase was not tested, as it adds a single adenine to the end of extended DNA, and so blunt-ends would not be produced by this enzyme.

DNaseI digested DNA was subjected to primer extension and ligation of linker as described in chapter 2.5.4 – 2.5.6. The resulting products were subjected to PCR to amplify the primer extended, blunt-ended products. As DNaseI digested DNA was used, a large range of primer-extended product sizes should have been present, thus when separated on a gel, a smear of many different fragment sizes was expected to be seen (figure 5.7).

The gel in figure 5.7 demonstrates that all enzymes used extended the primer, leaving blunt-ends, allowing the linker to be ligated. Sequenase enzyme has produced the largest range of product sizes. Not many fragments above 800 base pairs were produced by *Klenow*. T4 DNA polymerase was the least efficient at primer extending/blunt-end production, possibly because the enzyme was past its 'use by' date. This gel indicates that the primer extension protocol was successful, that Sequenase enzyme was the most efficient at primer extension, and that the double-stranded linker is efficiently ligated onto the end of the primer extended products.

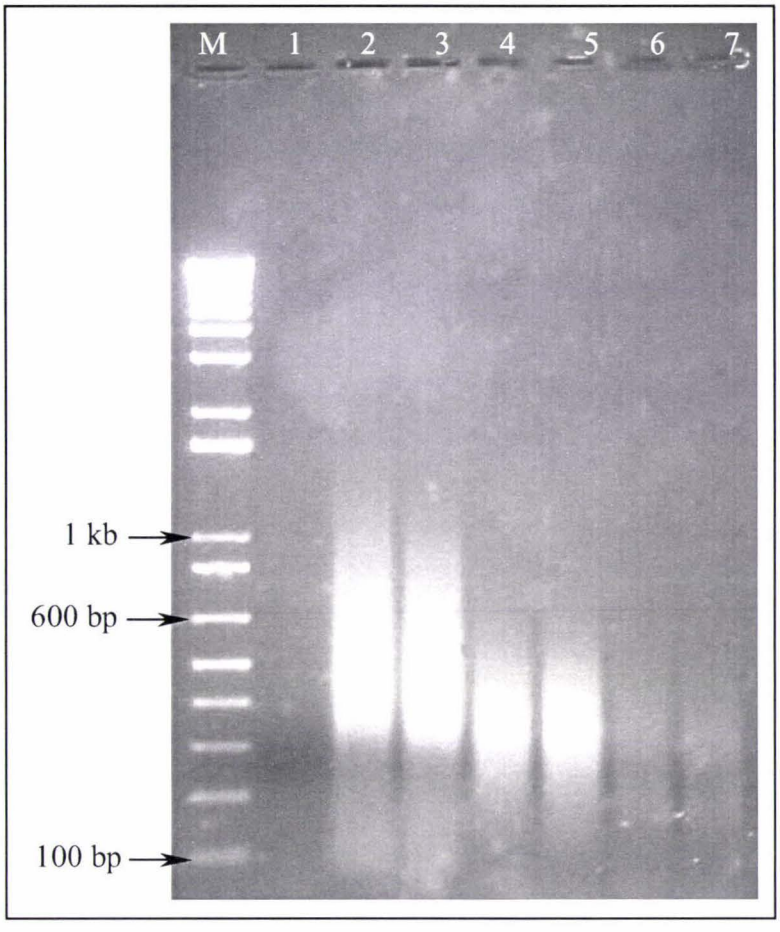


Figure 5.7. Gel photograph of the PCR of primer extension controls. Primer extension was carried out with 1 µg of MDA-MB-231 DNaseI digested DNA, using Sequenase version 2.0 (lanes 2 & 3), *Klenow* (lanes 4 & 5), or T4 DNA polymerase (lanes 6 & 7). Conditions for primer extension are detailed in chapter 2.5.4. Samples were then ligated to linker overnight, and 2 µL of the resulting sample was subjected to 30 cycles of PCR (denaturing = 94°C for 30 seconds, annealing = 55°C for 30 seconds, extension = 74°C for 30 seconds) using linker and Foot2 primers (Appendix 4) in a 20 µL volume. 5 µL of resulting samples were then electrophoresed on a 1% agarose gel in 1 x TAE stained with ethidium bromide, for 1 hour at 120 V. M = 1 kb plus DNA ladder, 1 = PCR negative control.

5.3.3 Product capture

The product capture step in the footprinting protocol is designed to minimise background, which may be due to 3'OH groups left by DNaseI digestion acting as random primers in the PCR step.

Different protocols were used for the capture of footprinting samples and for the samples used in the creation of the sequencing ladder (refer to Chapter 2.5.7 and 2.5.9). The footprinting protocol captures the products on streptavidin beads, melts the strands

in sodium hydroxide, and collects the eluted strand, whilst the sequencing protocol collects the bound strand. Single-stranded DNA in alkaline conditions can irreversibly bind to the sides of plastic tubes, and so the template may have been lost in this step. It was reasoned that the DNA does not need to be made single-stranded for the PCR step, and that the strand of DNA bound to the streptavidin beads would be as good a template as the non-bound strand.

A test was set up using the sequencing product capture protocol with the eluted strand collected. The footprinting product capture protocol was also modified to test captured, but not melted DNA; as well as captured, melted DNA – both bound to and melted from the beads. A new batch of streptavidin beads was also tested, to ensure the beads themselves were not faulty. Other samples were either not purified, or treated with terminal transferase to block the 3'OH end of DNA by the addition of di-deoxy nucleotides. This then prevents random primer extension in the PCR reaction (Pfeifer and Riggs, 1991). The resulting products were subjected to PCR as described in the footprinting protocol (Chapter 2.5.8), separated on an acrylamide gel, transferred, hybridised, and detected.

Unfortunately all of the above tests showed no bands on the sequencing gel. This suggested that the product capture step was not at fault, as samples with no purification, or treated with terminal transferase did not give rise to any visible bands. At the same time, the efficiency of the product capture was also not able to be evaluated.

5.3.4 PCR

The footprinting PCR protocol uses PCR buffer without Mg^{2+}, and adds $MgCl_2$ to a final concentration of 2 mM. Controls were set up to test the Mg^{2+}-free buffer, and the amount of Mg^{2+} in the PCR. This was done by using Mg^{2+} containing buffer, with and without additional Mg^{2+}. The PCR was carried out as per the protocol, using a PCR product as the template, and forward and reverse footprinting primers. Resulting products were analysed by agarose gel electrophoresis (figure 5.8).

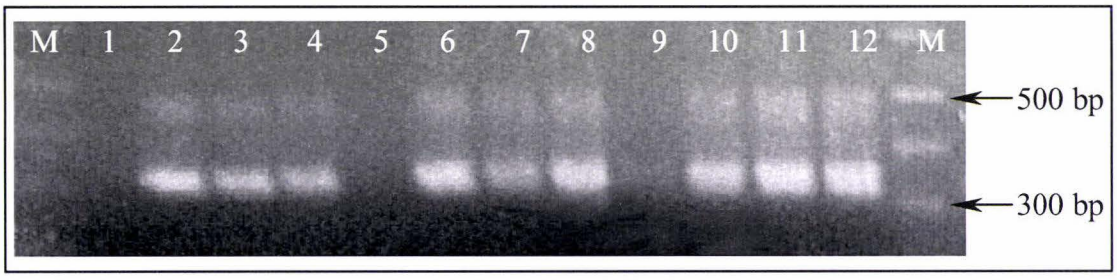


Figure 5.8. Photograph of footprinting PCR gel. PCR was setup as detailed in chapter 2.5.8 using PCR product as template, and -134 forward and reverse primer 2. 20 cycles of PCR were carried out with denaturing at 94°C for 1 minute, annealing at 55°C for 2 minutes, and extension at 74°C for 3 minutes. All of resulting PCR was then electrophoresed for 1 hour at 120 V, on a 1% agarose gel in 1 x TAE stained with ethidium bromide. M = 1 kb plus DNA ladder. Lanes 1, 5 and 9 are negative PCR controls. 1-4 = Mg^{2+} containing buffer supplemented with $MgCl_2$ to a final Mg^{2+} concentration of 2mM. 5-8 = footprinting PCR protocol unmodified, 9-12 = Mg^{2+} PCR buffer without additional Mg^{2+}.

As demonstrated in figure 5.8 all lanes, except negative controls show relatively similar intensity of the high mobility DNA fragment. The template used was an unpurified PCR product; this may explain the appearance of the lower mobility fragment. This result indicates that there is no significant difference between the different PCR conditions. Thus the PCR conditions used were not the cause of the poor footprints.

The footprinting protocol is designed to be quantitative i.e. a less intense band indicates that its position is protected from DNaseI cleavage more often than a more intense band. To ensure the technique is quantitative, only 20 cycles of PCR were carried out to 'minimise preferential amplification of short sequences'. If the template DNA concentration was low, then 20 cycles of PCR may not be enough for a banding pattern to be visible. In a final attempt to achieve a meaningful footprint banding pattern, the number of cycles of PCR was increased. An extra 10 cycles of PCR could potentially increase the number of DNA molecules per band by a factor of 1000. If the footprinting protocol was successful, but not enough DNA was present to give visible bands, then increasing the number of PCR cycles would increase the amount of DNA and therefore may produce visible bands (figure 5.9).

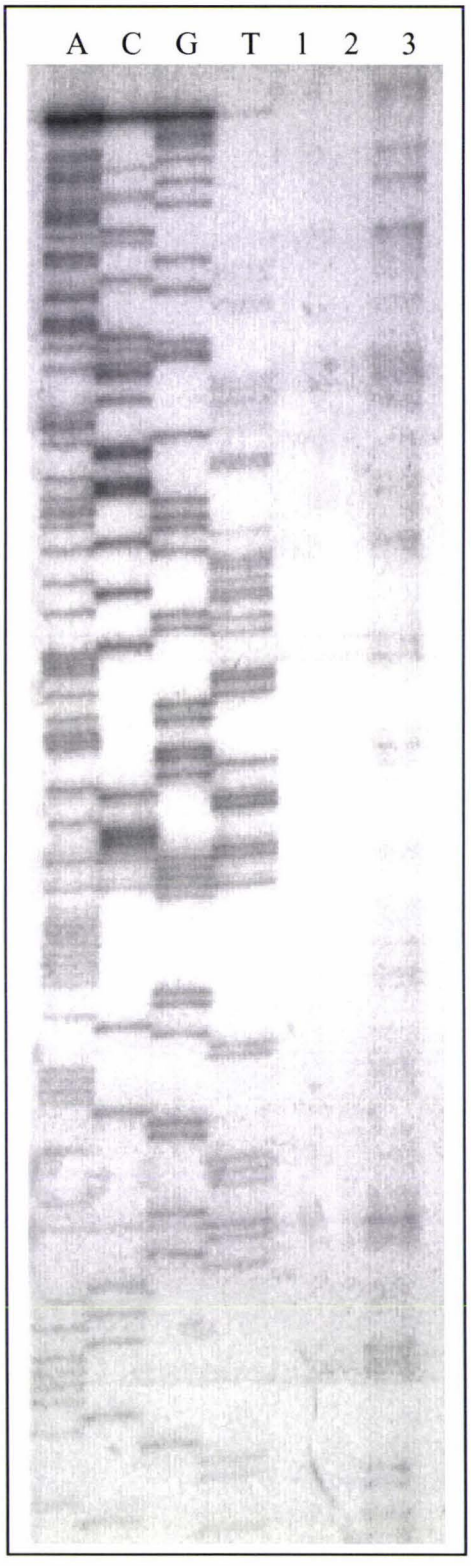


Figure 5.9. The effect of increased PCR cycles on footprinting banding pattern. Primer extension was carried out on 1.5 µg of DNaseI digested, naked DNA. Samples were then ligated, and captured as detailed in chapter 2.5. The captured product was then subjected to 20 cycles (lanes 1-2) (PCR conditions were denaturing = 94°C for 1 minute, annealing 55°C for 2 minutes, and extension 74°C for 3 minutes), or 30 cycles of PCR (lane 3) (cycling temperatures the same as lanes 1-2, but all steps were 30 seconds), with FootR2 and linker primers (Appendix 4). The samples were then run on a denaturing 8% poly-acrylamide gel for approximately 4 hours, and transferred to a nylon membrane. The membrane was then hybridised with labelled probe, washed and incubated with X-ray film. A = adenine, C = cytosine, G = guanine, T = thymine.

As shown in figure 5.9, 30 cycles of PCR has produced more intense bands when compared to 20 cycles. However the bands that are present are not well resolved. The appearance of bands in the 30 cycle lane indicates that the footprinting protocol is functional, but the amount of DNA per band may not have been enough to be visualised, even when using the sensitivity of radioactivity for detection. This could mean that the initial template DNA concentration was too low, or that DNA was being lost in one or more steps of the protocol, for example the product capture, or DNA precipitation steps.

5.5 Chapter Summary

DNaseI footprinting is designed to enable 'high resolution, nucleotide analysis of chromatin'(Riggs et al., 1998). The technique enables the identification of DNA sequences to which DNA binding proteins are bound. From the predicted sequence potential regulators can be identified, and then analysed further. Footprinting was used to follow up the results from transient transfection experiments using the maspin promoter deletion constructs. Transient transfections showed several important regions of potential importance in the regulation of transcription. Footprinting was attempted with the aim of identifying putative regulators that bind to these regions.

In vivo footprinting had been trialled once by Dr Kirsty Allen, with very limited success (Allen, 2003). The protocol has many steps which, if not performed optimally, could cause loss of bands on the final gel. Poor primer design can also reduce efficiency of primer extension and/or PCR; however four different sets of primers were used in an attempt to footprint four different regions of the maspin promoter. In each case only the sequencing ladder yielded bands on the analytical gel. Controls were then designed and tested to ensure each step of the protocol was successful. These controls showed that the ligation, primer extension, and PCR steps were all successful. The product capture step, while not definitively proven to be functioning efficiently, was unlikely to cause the lack of visible bands in the footprinting samples. This indicated that a more fundamental problem may have caused the absence of bands.

The number of cycles of PCR was increased to test if the protocol was successful but not enough DNA was being produced for bands to be visualised. 30 cycles of PCR gave obvious bands when compared to the PCR with 20 cycles. This meant that the protocol was functioning satisfactorily, but the amount of DNA in the PCR reaction may have been insufficient to produce visible bands.

To enable reproducible, quantitative footprinting results, the amount of DNA in the PCR needs to be increased and the number of cycles kept at 20. Increasing the number of cycles from 20 to 30 could cause a 1000 fold (2^{10} = 1024) increase in the amount of DNA in the final samples. In contrast, a 5 fold increase in the amount of DNA added to the PCR, results mathematically in a much more dramatic increase. For example if 2 single-stranded molecules are present in the initial PCR and 20 cycles were carried out then $2^{20} = 1 \times 10^6$ molecules could be present in the final sample. However if the number of template molecules was increased to 10 then $10^{20} = 1 \times 10^{20}$ molecules could be present. Thus increasing the template DNA 5 fold, far out-weighs the effect of increasing the number of cycles.

The protocol followed indicated that as little as 0.1 μg of DNA was enough to provide footprinting information, however to get reproducible results greater than 1 μg was recommended (Riggs et al., 1998). For all experiments 1.5 μg of DNaseI digested DNA was used. From the result of increasing the cycles of PCR and the calculations above, increasing the amount of DNA in the initial primer extension from 1.5 μg up to 7.5 μg may result in reliable footprinting results. The lack of bands may also be due to loss of DNA in the purification/precipitation steps of the protocol. Due to time restrictions the effect of increasing initial DNA concentration, and altering purification steps could not be tested.

CHAPTER 6 – Discussion and Future Work

6.1 Summary of the current knowledge of maspin

Maspin is a tumour suppressor gene which is expressed in tissues of epithelial origin, and reportedly lost in many cancers of these tissues. The tumour suppressive functions of maspin relate to its ability to inhibit cellular migration and invasion, sensitise cells to induced apoptosis, and inhibit angiogenesis. The molecular function of maspin is currently unknown; however reports are starting to elucidate the mechanisms of action. The inhibition of cellular mobility by maspin is dependent upon an intact reactive site loop (RSL); however this is unlikely to be due to a protease inhibitory function as maspin contains a short and highly divergent RSL. Maspin does not display the characteristic stressed-relaxed transition of inhibitory serpins; supporting the non-protease inhibitory mode of action. Maspin reportedly interacts with collagen, independent of its RSL; this may play a role in the inhibition of angiogenesis, which has been shown to be independent of the RSL. The role of maspin in apoptosis seems to require both C- and N-termini, as PAI-1 C- and N-terminals cannot substitute for maspin in this function. These results suggest that maspin contains at least two functional sites, both of which are required for its tumour suppressor activity.

Maspin was initially identified by subtractive hybridisation between normal and cancerous breast cell lines. A progressive down-regulation of maspin in tumours has been reported as prostate cancer progresses. The case for down-regulation of maspin in breast cancers is more controversial. Interestingly, maspin is not expressed in normal ovarian and pancreatic cancer, however it is expressed in many cancers of these tissues; this seems contradictory to the reported tumour suppressor functions of maspin.

To date maspin has been reported to be positively regulated by Ets, Ap1 and p53, whilst negatively regulated by the binding of androgen receptor to the hormone response element (HRE). These transcriptional effectors bind to their respective sites which are located within 500 bp of the transcription start site. The anti-oestrogen drug tamoxifen has been demonstrated to activate maspin expression. This activation is dependent on an intact HRE element within the maspin promoter and the presence of oestrogen receptor

β (ERβ), suggesting that the ER also binds to the maspin HRE, however no difference in maspin expression was seen when cells were treated with oestrogen. Nitric oxide, manganese containing super-oxide dismutase, and gamma linolenic acid, have also been shown to activate maspin expression, however the mechanisms of transcriptional activation caused by these effectors have not been demonstrated.

The aims of this study were to identify putative transcription factors which play a role in the regulation of maspin, in order to link these effectors with pathways or molecular mechanisms in which maspin acts.

6.2 Summary of results

6.2.1 *In silico* studies

The maspin gene was identified from the annotated human genome sequence, and the upstream regulatory region identified with the help of the reported transcription start site (Zhang et al., 1997a). This sequence was then analysed for transcription effector sites, and sequence conservation between human, mouse, and rat. Too many transcription effector sites were predicted, preventing identification of putative regulatory sites; and few sites mapped to regions which were conserved between the three species. Interestingly, the previously reported Ap1, Ets and p53 effector sites were conserved in all three organisms, indicating that these sites are likely to be important for the regulation of maspin expression in these organisms. The conservation of these sites also supported the use of this method for the identification of potentially important regions within the maspin promoter. A potential enhancer region was also identified by high conservation in all three species; this region was approximately 4.1 kb upstream of the human maspin transcription start site.

6.2.2 Cloning and functional analysis

To enable functional analysis, the maspin promoter was cloned into the pGL3 Basic vector. The cloned promoter was then subjected to truncation from the 5' end, thus removing putative transcription effector sites. The effect of these truncations on reporter gene activity was then measured by transient transfection of the deletion series into several cell lines. The results of this are summarised in figure 6.1.

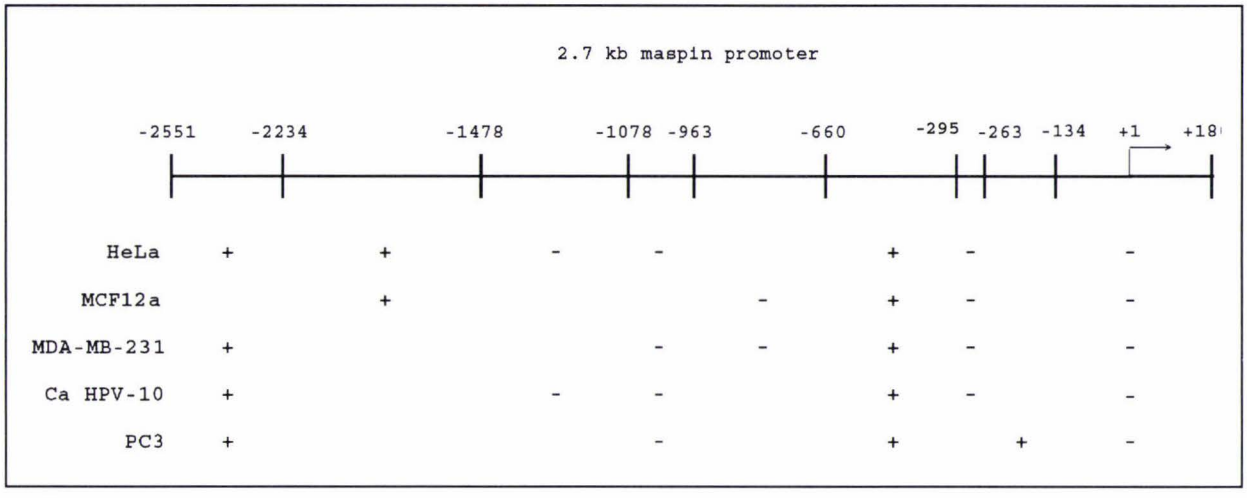


Figure 6.1. Summary of transient transfection experiments. Five different cell lines were transfected with the maspin deletion series, and the level of reporter gene expression measured. The effect of removing 5' regions of the maspin promoter on reporter gene activity was analysed, with significant increases and decreases observed in each cell line indicated by + or – respectively. This figure demonstrates the important regions of the maspin promoter.

From the results of transient transfection experiments, five regions of the maspin promoter were identified as being important in the regulation of gene expression. As shown in figure 6.1 these five regions showed a significant increase or decrease in reporter gene expression in 4/5 (-2551 bp to -2235 bp region, -295 bp to -263 bp region, and -1078 bp to -963 bp region) or 5/5 (-660 bp to -295 bp and -134 bp to +180) cell lines. This suggests that the factors that are responsible for the trends seen in these regions are likely to be ubiquitously expressed. Other regions showed a significant increase or decrease in reporter gene expression in only one or two cell lines. One such region, -963 bp to -660 bp, showed a significant decrease in both breast cell lines. These regions may also be important for the regulation of maspin; however the factors responsible for the regulation are not expressed, or are not active in cell lines that did

not show a significant difference. Thus the factors responsible for these trends may be tissue specific.

A large difference in reporter gene expression was found between the normal breast cell line MCF12a and cancerous breast cell line MDA-MB-231. As maspin is reportedly down-regulated in cancerous cells, this result was not unexpected. The maspin promoter has been reported to be methylated in some cancerous cell lines, including MDA-MB-231, resulting in chromatin compaction and loss of maspin transcription. Other reports suggest that the loss of maspin transcription in cancerous cell lines is due to the loss of transactivation by Ets, Ap1 and/or p53. Thus the loss of one or more of these transcription factors may be responsible for the loss of reporter gene expression. It is also possible that a transcriptional repressor which is capable of binding to the maspin promoter is up-regulated.

To distinguish between methylation, and gain or loss of transcriptional effectors, co-transfections with the methyl-transferase inhibitor 5-azacytidine were carried out. An increase in reporter gene expression was not observed when treated with 5-azacytidine, indicating that methylation of the maspin promoter is not the cause of the loss of reporter gene expression in MDA-MB-231 cell lines. Supporting this, p53 is reported to be mutated in MDA-MB-231 cells suggesting that this is in-part the reason for the loss of reporter gene expression in these cells.

6.2.3 DNaseI footprinting

Identification of sequences which putative transcriptional effectors bind to was attempted using *in vivo* DNaseI footprinting. Primers were designed to four regions of the maspin promoter identified as being important for the regulation of the gene. The -295 bp to -263 bp region of the promoter, which showed a significant decrease in 4/5 cell lines when removed was not used, however this region may be short enough for use in electrophoretic mobility shift assays. Despite using four sets of primers and DNA from four different cell lines, banding patterns were not observed on the final footprinting gel. After exhaustive control experiments, it was demonstrated that the amount of DNA in the final PCR was probably insufficient to produce visible bands. This may be caused by not enough DNA in initial primer extension, or due to loss of

DNA in one or more of the purification/precipitation steps of the protocol. Due to time restrictions however, the effect of adding more DNA, or altering purification protocols was not able to be tested.

6.3 Future work

To complete the aims of this project, *in vivo* DNaseI footprinting needs to be carried out successfully, with optimisation of DNA quantity, and DNA purification protocols. Once this is carried out meaningful data could be gained from the footprinting of the four regions of the maspin promoter. This technique allows the identification of DNA sequences at which DNA binding proteins are bound *in vivo*. From this sequence data, the protein which binds to this region may be able to be identified by comparison of bound sequence with the recognition sequence of the protein.

Once identified, confirmation of protein binding to the sequence could be demonstrated by electrophoretic mobility shift assays (EMSA) (see figure 6.2). The assay relies upon the ability of the transcription effector to specifically bind to a short piece of DNA representing the putative binding site and surrounding region; and the resulting decrease in mobility of the DNA though an acrylamide gel. The identity of this protein may then be confirmed by 'super-shifts', where a specific antibody to the protein is added to the assay, resulting in a further decrease in mobility, and thus a 'shift' toward the top of the gel. This technique could also be used to identify a protein binding specifically to the -295 bp to -263 bp region of the maspin promoter. In this case the ~30 bp region would be truncated from both 5' and 3' ends and the sequence mutated to identify the minimal sequence to which a protein binds. This sequence could then be used to identify the protein in the same way as footprinting. The identity of the protein could also be confirmed by 'super-shifts'.

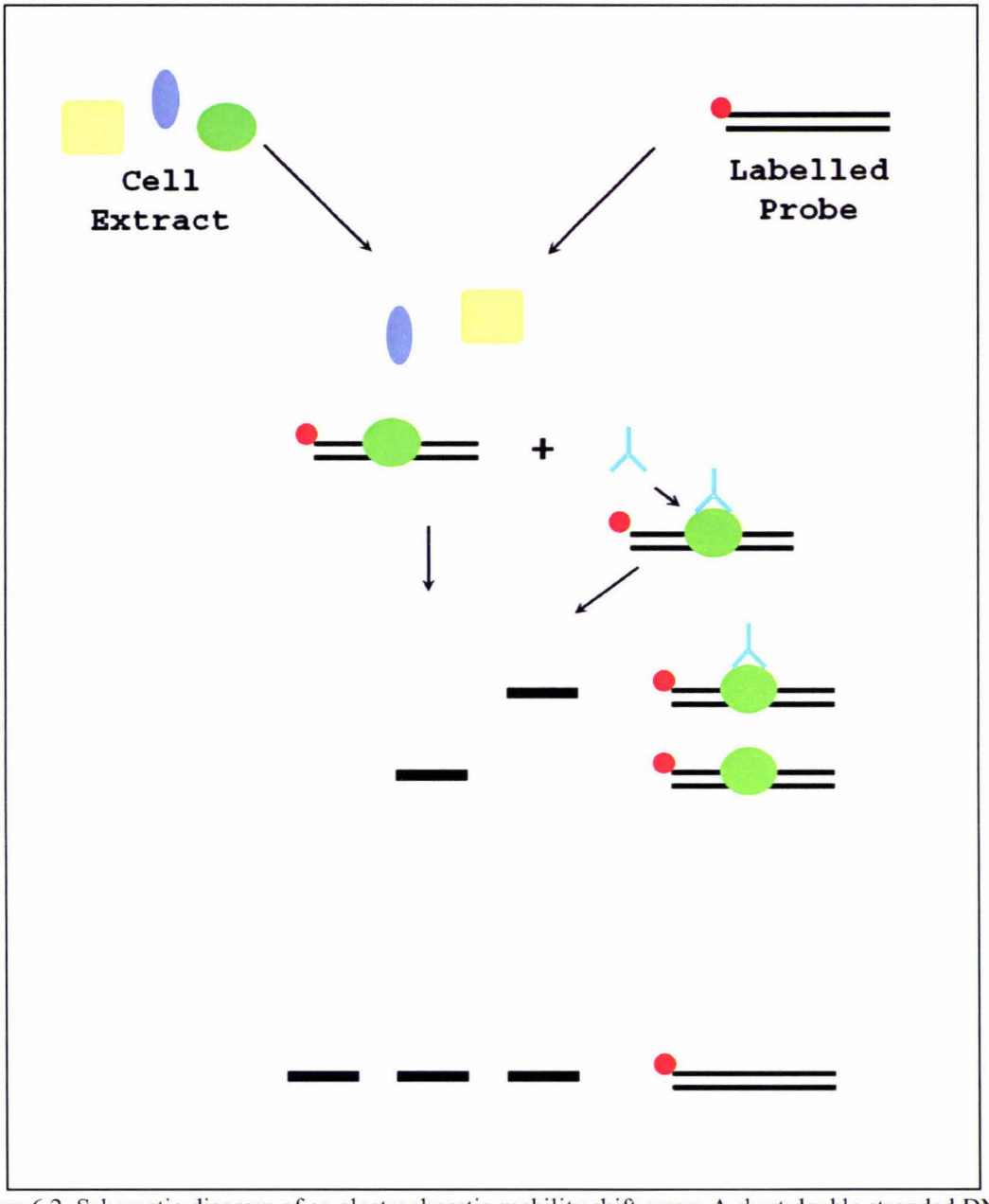


Figure 6.2. Schematic diagram of an electrophoretic mobility shift assay. A short double stranded DNA oligonucteotide (double black lines) is labelled with radioactive phosphate group (red dot), and added to whole cell extracts (proteins indicated by green, blue, and yellow shapes). DNA binding proteins recognise the oligonucleotide and bind, causing a decrease in mobility when electrophoresed on an acrylamide gel (grey). The identity of the binding protein can be demonstrated by addition of a specific antibody (light blue), which further decreases mobility. The complex which the individual bands represent is indicated to the right of the gel. Lanes on gel from left to right are no extract, probe plus extract, probe plus extract plus antibody (super-shift).

Once a protein is confirmed to bind to a region of the maspin promoter, then the effect of this protein on the transcriptional regulation of the gene can be measured. This could be done using transient transfections with maspin promoter constructs containing mutations in and/or deletion of the transcription factor binding site. Schematic diagrams of how such constructs can be created are shown in figure 6.3 and 6.4.

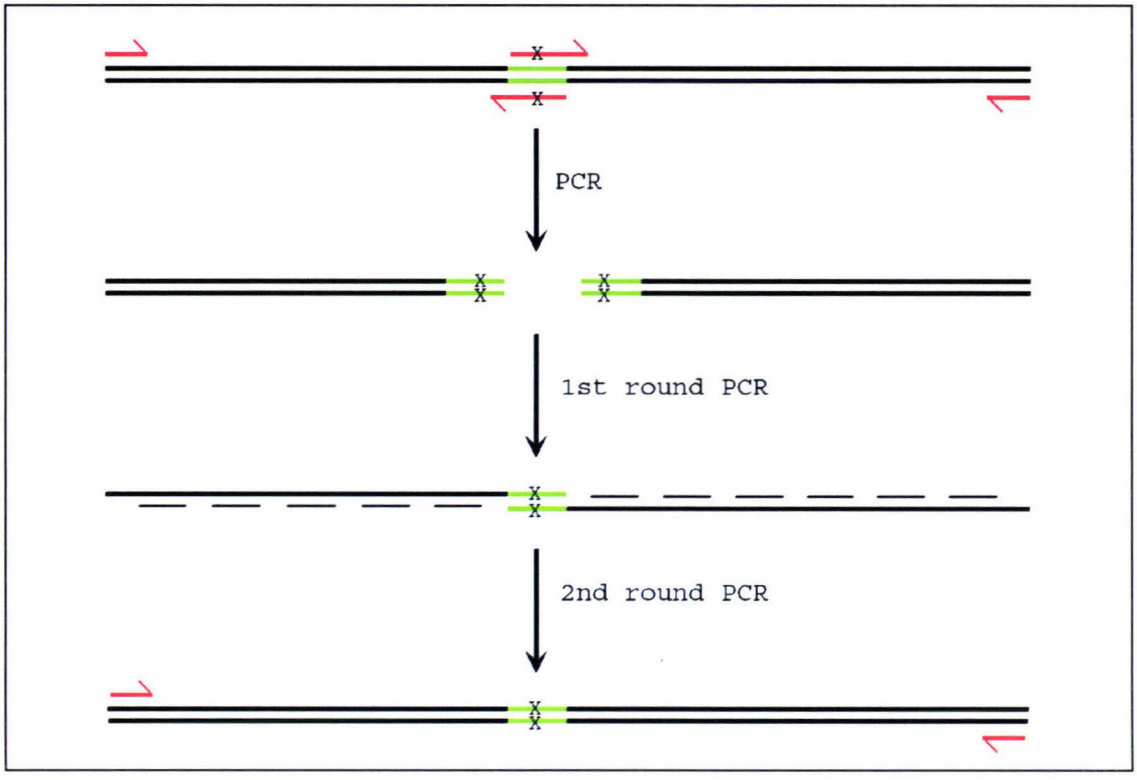


Figure 6.3. Schematic diagram of the creation of a mutated transcriptional effector binding site. The promoter fragment is amplified in two separate reactions using internal primers which anneal to the effector binding site. These primers contain mis-matches indicated by an X. When amplified the products will contain these mis-matches as mutations. The individual products are then added together, and PCR carried out; at first without primer to create a full length construct, then with primers to amplify the full length construct, containing the mutations. Green indicates transcriptional effector binding site. Red arrows indicate primers, black double bars indicate DNA, and dashed lines indicate new DNA synthesis.

A mutation in the recognition site for a transcriptional effector, creation of which is shown in figure 6.3, may not prevent binding of the effector, however is likely to at least decrease the affinity of the effector for the site. This decrease in affinity may cause a change in the transcriptional activity of the factor which may be seen in reporter gene assays, following transient transfection of the construct into cells. If the effector was an activator and the affinity for the binding site was reduced, then a decrease in reporter gene activity should be seen, if the effector was a repressor, then an increase in reporter gene activity should be seen.

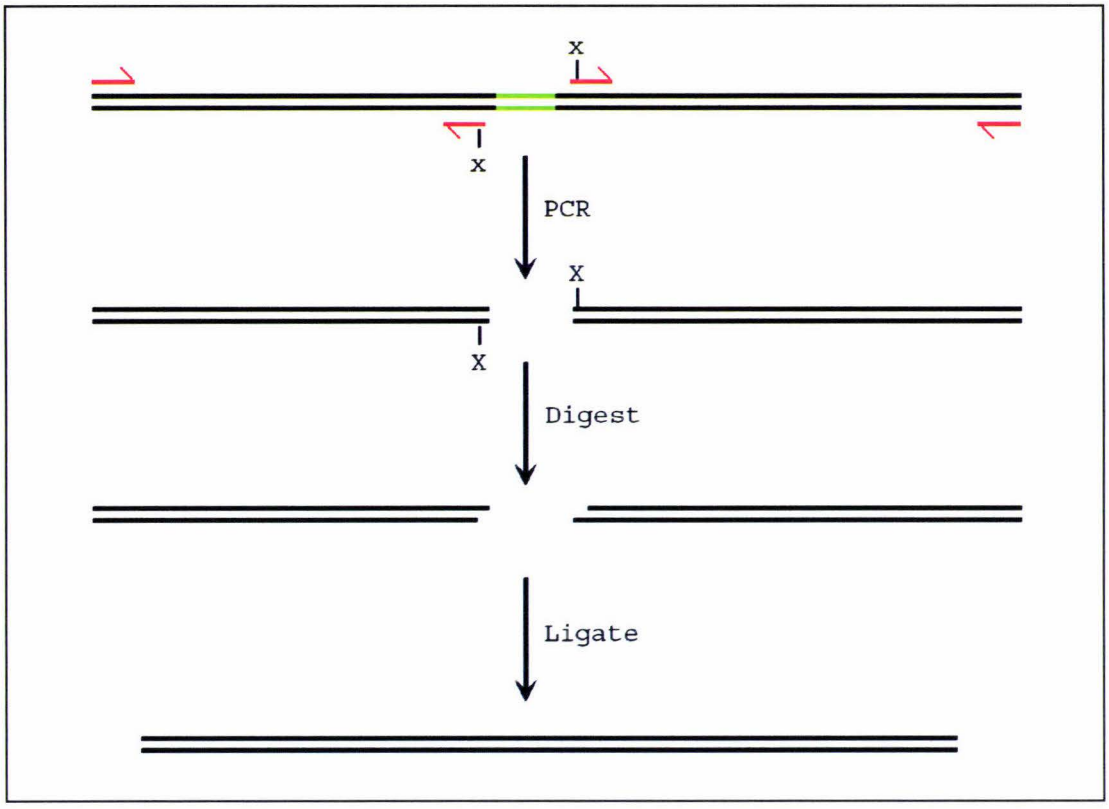


Figure 6.4. Schematic diagram of the creation of an internal deletion. The promoter fragment is amplified in two separate reactions using internal primers which anneal to either side of the effector binding site. These primers contain unique restriction sites indicated by an X. When amplified the products will contain these restriction sites at the ends of the products. The individual products are then digested and ligated together to create construct without the transcription effector binding site. Green indicates transcriptional effector binding site. Red arrows indicate primers, black double bars indicate DNA.

An internal deletion which removes the effector binding site (creation shown in figure 6.4) should prevent the factor from binding. This means that the transcriptional activation/repression activity of the factor is no longer present in the promoter, and thus a corresponding change in reporter gene activity should be seen in transient transfection of construct into cell lines.

In this way footprinting, followed by EMSA and functional assays with deleted and/or mutated effector binding sites would be used to identify, and assay the functional significance of transcription factors which play a role in the regulation of maspin. Other regions which may play regulatory roles on the transcription of maspin are the putative enhancer region identified in human, mouse and rat alignments, and the 7 kb

first intron. The cloning of these regions upstream of a minimal promoter, and the resulting effect on reporter gene expression after transfection of the constructs will test whether these regions regulate maspin transcription. If these regions are found to be important for maspin regulation then footprinting, EMSA, and functional assays can also be carried out on these regions.

With the identification and characterisation of transcriptional effectors important for the regulation of maspin, the clinical significance of these effectors can be tested. For statistically significant results, at least 100 samples from a tumour bank, clinically characterised with tumour markers, response to drugs, and maspin expression would be needed, such a bank exists at Palmerston North Hospital. Expression of the identified transcription effector could then be measured in these tumour specimens by immunohistochemistry, and the correlation between effector expression, maspin expression, cancer progression, and other clinical factors such as disease free survival could be assessed.

6.3 Conclusion

Several regions of the maspin promoter of potential importance to the transcriptional regulation of maspin have been identified. The factors which bind to these regions and are responsible for the trends seen in transient transfection and reporter gene assays are currently unknown. The use of *in vivo* DNaseI footprinting, followed by EMSA and transfection of mutation and internal deletion constructs will be useful in the identification of effectors, and analysis of their functional significance in maspin gene expression. The significance of these factors in clinical specimens can then be measured and correlated to clinical outcome. The identification of such factors may help in the search for a molecular function of maspin, and thus a greater understanding of this reported tumour suppressor.

References

Allen, K. (2003) An investigation into the regulation of the topoisomerase II alpha promoter in breast cancer cells exposed to doxorubicin., Massey University, Palmerston North.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Bass, R., Fernandez, A.-M. M., and Ellis, V. (2002). Maspin Inhibits Cell Migration in the Absence of Protease Inhibitory Activity. J Biol Chem *277*, 46845-46848.

Beecken, W. D., Kramer, W., and Jonas, D. (2000). New molecular mediators in tumor angiogenesis. J Cell Mol Med *4*, 262-269.

Biliran, H., Jr., and Sheng, S. (2001). Pleiotrophic inhibition of pericellular urokinase type plasminogen activator system by endogenous tumor suppressive maspin. Cancer Res *61*, 8676-8682.

Blacque, O. E., and Worrall, D. M. (2002). Evidence for a Direct Interaction between the Tumor Suppressor Serpin, Maspin, and Types I and III Collagen. J Biol Chem *277*, 10783-10788.

Bonaccorsi, L., Carloni, V., Muratori, M., Salvadori, A., Giannini, A., Carini, M., Serio, M., Forti, G., and Baldi, E. (2000). Androgen receptor expression in prostate carcinoma cells suppresses alpha6beta4 integrin-mediated invasive phenotype. Endocrinology *141*, 3172-3182.

Bradbury, P. (2004) An investigation into the hormonal control and prognostic significance of maspin in breast cancer., Massey University, Palmerston North.

Brown, R., and Strathdee, G. (2002). Epigenomics and epigenetic therapy of cancer. Trends Mol Med *8*, S43-48.

Carey, M., and Smale, S. T. (2000). Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Cato, A. C. B., and Peterziel, H. (1998). The Androgen Receptor as Mediator of Gene Expression and Signal Transduction Pathways. Trends in Endocrinology and Metabolism *9*, 150-154.

Chai, J., and Tarnawski, A. S. (2002). Serum response factor: discovery, biochemistry, biological roles and implications for tissue injury healing. J Physiol Pharmacol *53*, 147-157.

Chung, C. T., Niemela, S. L., and Miller, R. H. (1989). One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. Proc Natl Acad Sci U S A *86*, 2172-2175.

Costello, J. F., and Vertino, P. M. (2002). Methylation matters: a new spin on maspin. Nat Genet *31*, 123-124.

Cowell, I. G. (2002). E4BP4/NFIL3, a PAR-related bZIP factor with many roles. Bioessays *24*, 1023-1029.

Cullen, J. J., Weydert, C., Hinkhouse, M. M., Ritchie, J., Domann, F. E., Spitz, D., and Oberley, L. W. (2003). The role of manganese superoxide dismutase in the growth of pancreatic adenocarcinoma. Cancer Res *63*, 1297-1303.

DeVita, V. T. J., Hellman, S., and Rosenberg, S. A. (2001). Cancer: Principles and Practice of Oncology 6 - on CD-ROM, Lippincott Williams & Wilkins).

Domann, F. E., Rice, J. C., Hendrix, M. J., and Futscher, B. W. (2000). Epigenetic silencing of maspin gene expression in human breast cancers. Int J Cancer *85*, 805-810.

Feldman, R. J., Sementchenko, V. I., Gayed, M., Fraig, M. M., and Watson, D. K. (2003). Pdef Expression in Human Breast Cancer Is Correlated with Invasive Potential and Altered Gene Expression. Cancer Res *63*, 4626-4631.

Fitzpatrick, P. A., Wong, D. T., Barr, P. J., and Pemberton, P. A. (1996). Functional implications of the modeled structure of maspin. Protein Eng *9*, 585-589.

Futscher, B. W., Oshiro, M. M., Wozniak, R. J., Holtan, N., Hanigan, C. L., Duan, H., and Domann, F. E. (2002). Role for DNA methylation in the control of cell type specific maspin expression. Nat Genet *31*, 175-179.

Gasco, M., Shami, S., and Crook, T. (2002). The p53 pathway in breast cancer. Breast Cancer Res *4*, 70-76.

GCG (1997). Wisconsin Package Version 9.1 (Madison, Wisc, Genetics Computer Group).

Hofseth, L. J., Saito, S., Hussain, S. P., Espey, M. G., Miranda, K. M., Araki, Y., Jhappan, C., Higashimoto, Y., He, P., Linke, S. P., *et al.* (2003). Nitric oxide-induced cellular stress and p53 activation in chronic inflammation. Proc Natl Acad Sci U S A *100*, 143-148.

Hojo, T., Akiyama, Y., Nagasaki, K., Maruyama, K., Kikuchi, K., Ikeda, T., Kitajima, M., and Yamaguchi, K. (2001). Association of maspin expression with the malignancy grade and tumor vascularization in breast cancer tissues. Cancer Lett *171*, 103-110.

Holmes, D. S., and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal Biochem *114*, 193-197.

Hopkins, P. C., and Whisstock, J. (1994). Function of maspin. Science *265*, 1893-1894.

Jiang, N., Meng, Y., Zhang, S., Mensah-Osman, E., and Sheng, S. (2002). Maspin sensitizes breast carcinoma cells to induced apoptosis. Oncogene *21*, 4089-4098.

Jiang, W. G., Hiscox, S., Horrobin, D. F., Bryce, R. P., and Mansel, R. E. (1997). Gamma linolenic acid regulates expression of maspin and the motility of cancer cells. Biochem Biophys Res Commun *237*, 639-644.

Jiang, W. G., Redfern, A., Bryce, R. P., and Mansel, R. E. (2000). Peroxisome proliferator activated receptor gamma (PPAR gamma) mediates the action of gamma linolenic acid in breast cancer cells. Prostaglandins Leukot Essent Fatty Acids *62*, 119-127.

Khalkhali-Ellis, Z., Christian, A. L., Kirschmann, D. A., Edwards, E. M., Rezaie-Thompson, M., Vasef, M. A., Gruman, L. M., Seftor, R. E. B., Norwood, L. E., and Hendrix, M. J. C. (2004). Regulating the Tumor Suppressor Gene Maspin in Breast Cancer Cells: A Potential Mechanism for the Anticancer Properties of Tamoxifen. Clin Cancer Res *10*, 449-454.

Khalkhali-Ellis, Z., and Hendrix, M. J. C. (2003). Nitric Oxide Regulation of Maspin Expression in Normal Mammary Epithelial and Breast Cancer Cells. Am J Pathol *162*, 1411-1417.

Kim, D. H., Yoon, D. S., Dooley, W. C., Nam, E. S., Ryu, J. W., Jung, K. C., Park, H. R., Sohn, J. H., Shin, H. S., and Park, Y. E. (2003). Association of maspin expression with the high histological grade and lymphocyte-rich stroma in early-stage breast cancer. Histopathology *42*, 37-42.

Lee, S. W., Tomasetto, C., and Sager, R. (1991). Positive selection of candidate tumor-suppressor genes by subtractive hybridization. Proc Natl Acad Sci U S A *88*, 2825-2829.

Li, J. J., Colburn, N. H., and Oberley, L. W. (1998). Maspin gene expression in tumor suppression induced by overexpressing manganese containing superoxide dismutase cDNA in human breast cancer cells. Carcinogenesis *19*, 833-839.

Liotta, L. A., and Kohn, E. C. (2001). The microenvironment of the tumour-host interface. Nature *411*, 375-379.

Maass, N., Biallek, M., Rosel, F., Schem, C., Ohike, N., Zhang, M., Jonat, W., and Nagasaki, K. (2002a). Hypermethylation and histone deacetylation lead to silencing of the maspin gene in human breast cancer. Biochem Biophys Res Commun *297*, 125.

Maass, N., Hojo, T., Rosel, F., Ikeda, T., Jonat, W., and Nagasaki, K. (2001a). Down regulation of the tumor suppressor gene maspin in breast carcinoma is associated with a higher risk of distant metastasis. Clin Biochem *34*, 303-307.

Maass, N., Hojo, T., Ueding, M., Luttges, J., Kloppel, G., Jonat, W., and Nagasaki, K. (2001b). Expression of the tumor suppressor gene Maspin in human pancreatic cancers. Clin Cancer Res *7*, 812-817.

Maass, N., Hojo, T., Zhang, M., Sager, R., Jonat, W., and Nagasaki, K. (2000). Maspin a novel protease inhibitor with tumor suppressing activity in breast cancer. Acta Oncol *39*, 931-934.

Maass, N., Rolver, L., Ziebart, M., Nagasaki, K., and Rudolph, P. (2002b). Maspin locates to the nucleus in certain cell types. J Pathol *197*, 274-275.

Maass, N., Teffner, M., Rosel, F., Pawaresch, R., Jonat, W., Nagasaki, K., and Rudolph, P. (2001c). Decline in the expression of the serine proteinase inhibitor maspin is associated with tumour progression in ductal carcinomas of the breast. J Pathol *195*, 321-326.

Machtens, S., Serth, J., Bokemeyer, C., Bathke, W., Minssen, A., Kollmannsberger, C., Hartmann, J., Knuchel, R., Kondo, M., Jonas, U., and Kuczyk, M. (2001). Expression of the p53 and Maspin protein in primary prostate cancer: correlation with clinical features. Int J Cancer *95*, 337-342.

Magan, N. (2002) Regulation of human topoisomerase II gene expression, Massey university, Palmerston North.

Milde-Langosch, K., Loning, T., and Bamberger, A. M. (2003). Expression of the CCAAT/enhancer-binding proteins C/EBPalpha, C/EBPbeta and C/EBPdelta in breast cancer: correlations with clinicopathologic parameters and cell-cycle regulatory proteins. Breast Cancer Res Treat *79*, 175-185.

Minami, R., Kitazawa, R., Maeda, S., and Kitazawa, S. (1998). Analysis of 5'-flanking region of human Smad4 (DPC4) gene. Biochim Biophys Acta *1443*, 182-185.

Mohsin, S. K., Zhang, M., Clark, G. M., and Craig Allred, D. (2003). Maspin expression in invasive breast cancer: association with other prognostic factors. J Pathol *199*, 432-435.

NZHIS (2002). Cancer: New Registrations and Deaths 1999 (New Zealand Health Information Service, Ministry of Health).

Odero-Marah, V. A., Khalkhali-Ellis, Z., Schneider, G. B., Seftor, E. A., Seftor, R. E. B., Koland, J. G., and Hendrix, M. J. C. (2002). Tyrosine phosphorylation of maspin in normal mammary epithelia and breast cancer cells. Biochem Biophys Res Commun *295*, 800-805.

Ohike, N., Maass, N., Mundhenke, C., Biallek, M., Zhang, M., Jonat, W., Luttges, J., Morohoshi, T., Kloppel, G., and Nagasaki, K. (2003). Clinicopathological significance and molecular regulation of maspin expression in ductal adenocarcinoma of the pancreas. Cancer Lett *199*, 193-200.

Oikawa, T., and Yamada, T. (2003). Molecular biology of the Ets family of transcription factors. Gene *303*, 11-34.

Oshiro, M. M., Watts, G. S., Wozniak, R. J., Junk, D. J., L Munoz-Rodriguez, J., Domann, F. E., and Futscher, B. W. (2003). Mutant p53 and aberrant cytosine methylation cooperate to silence gene expression. Oncogene *22*, 3624-3634.

Oster, S. K., Ho, C. S., Soucie, E. L., and Penn, L. Z. (2002). The myc oncogene: MarvelouslY Complex. Adv Cancer Res *84*, 81-154.

Pemberton, P. A., Tipton, A. R., Pavloff, N., Smith, J., Erickson, J. R., Mouchabeck, Z. M., and Kiefer, M. C. (1997). Maspin is an intracellular serpin that partitions into secretory vesicles and is present at the cell surface. J Histochem Cytochem *45*, 1697-1706.

Pemberton, P. A., Wong, D. T., Gibson, H. L., Kiefer, M. C., Fitzpatrick, P. A., Sager, R., and Barr, P. J. (1995). The tumor suppressor maspin does not undergo the stressed to relaxed transition or inhibit trypsin like serine proteases. Evidence that maspin is not a protease inhibitory serpin. J Biol Chem *270*, 15832-15837.

Pfeifer, G. P., and Riggs, A. D. (1991). Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR. Genes Dev *5*, 1102-1113.

Pierson, C. R., McGowen, R., Grignon, D., Sakr, W., Dey, J., and Sheng, S. (2002). Maspin is up regulated in premalignant prostate epithelia. Prostate *53*, 255-262.

Primeau, M., Gagnon, J., and Momparler, R. L. (2003). Synergistic antineoplastic action of DNA methylation inhibitor 5 AZA 2' deoxycytidine and histone deacetylase inhibitor depsipeptide on human breast carcinoma cells. Int J Cancer *103*, 177-184.

Reis-Filho, J., Milanezi, F., and Schmitt, F. (2002). Maspin is expressed in the nuclei of breast myoepithelial cells. J Pathol *197*, 272-273.

Riggs, A. D., Singer-Sam, J., and Pfeifer, G. P. (1998). *In vivo* footprinting and chromatin analysis by LMPCR. In Chromatin- A Practical Approach., H. Gould, ed. (New York, Oxford University Press).

Sato, N., Fukushima, N., Matsubayashi, H., and Goggins, M. (2003). Identification of maspin and S100P as novel hypomethylation targets in pancreatic cancer using global gene expression profiling. Oncogene.

Schoonderwoert, V. T., and Martens, G. J. (2002). Structural gene organization and evolutionary aspects of the V-ATPase accessory subunit Ac45. Biochim Biophys Acta *1574*, 245-254.

Schug, J., and Overton, G. C. (1997). TESS: Transcription Element Search Software on the WWW (Pennsylvania, Computational Biology and Informatics Laboratory School of Medicine, University of Pennsylvania).

Seftor, R. E., Seftor, E. A., Sheng, S., Pemberton, P. A., Sager, R., and Hendrix, M. J. (1998). Maspin suppresses the invasive phenotype of human breast carcinoma. Cancer Res *58*, 5681-5685.

Shao, Z. M., Radziszewski, W. J., and Barsky, S. H. (2000). Tamoxifen enhances myoepithelial cell suppression of human breast carcinoma progression in vitro by two different effector mechanisms. Cancer Lett *157*, 133-144.

Shaulian, E., and Karin, M. (2001). AP-1 in cell proliferation and survival. Oncogene *20*, 2390-2400.

Sheng, S., Carey, J., Seftor, E. A., Dias, L., Hendrix, M. J., and Sager, R. (1996). Maspin acts at the cell membrane to inhibit invasion and motility of mammary and prostatic cancer cells. Proc Natl Acad Sci U S A *93*, 11669-11674.

Sheng, S., Pemberton, P. A., and Sager, R. (1994). Production, purification, and characterization of recombinant maspin proteins. J Biol Chem *269*, 30988-30993.

Sheng, S., Truong, B., Fredrickson, D., Wu, R., Pardee, A. B., and Sager, R. (1998). Tissue type plasminogen activator is a target of the tumor suppressor gene maspin. Proc Natl Acad Sci U S A *95*, 499-504.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res *29*, 308-311.

Shi, H. Y., Liang, R., Templeton, N. S., and Zhang, M. (2002). Inhibition of breast tumor progression by systemic delivery of the maspin gene in a syngeneic tumor model. Mol Ther *5*, 755-761.

Sood, A. K., Fletcher, M. S., Gruman, L. M., Coffin, J. E., Jabbari, S., Khalkhali-Ellis, Z., Arbour, N., Seftor, E. A., and Hendrix, M. J. C. (2002). The Paradoxical Expression of Maspin in Ovarian Carcinoma. Clin Cancer Res *8*, 2924-2932.

Sternlicht, M. D., Kedeshian, P., Shao, Z. M., Safarians, S., and Barsky, S. H. (1997). The human myoepithelial cell is a natural tumor suppressor. Clin Cancer Res *3*, 1949-1958.

Thomas, M. J., and Seto, E. (1999). Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? Gene *236*, 197-208.

Thuring, R. W., Sanders, J. P., Borst, P., Holmes, D. S., and Quigley, M. (1975). A freeze-squeeze method for recovering long DNA from agarose gels
A rapid boiling method for the preparation of bacterial plasmids. Anal Biochem *66*, 213-220.

Umekita, Y., Hiipakka, R. A., and Liao, S. (1997). Rat and human maspins: structures, metastatic suppressor activity and mutation in prostate cancer cells. Cancer Lett *113*, 87-93.

Umekita, Y., Ohi, Y., Sagara, Y., and Yoshida, H. (2002). Expression of maspin predicts poor prognosis in breast cancer patients. Int J Cancer *100*, 452-455.

Umekita, Y., and Yoshida, H. (2003). Expression of maspin is up-regulated during the progression of mammary ductal carcinoma. Histopathology *42*, 541-545.

van Dijk, A. (2003) The role of maspin in breast and prostate cancer., Massey University, Palmerston North.

Verger, A., and Duterque-Coquillaud, M. (2002). When Ets transcription factors meet their partners. Bioessays *24*, 362-370.

Vojtesek, B., and Lane, D. (1993). Regulation of p53 protein expression in human breast cancer cell lines. J Cell Sci *105*, 607-612.

Weaver, R. F. (1999). Molecular Biology, 1st edn (USA, WCG/McGraw-Hill).

Weijerman, P. C., Zhang, Y., Shen, J., Dubbink, H. J., Romijn, J. C., Peehl, D. M., and Schroder, F. H. (1998). Expression of prostatic factors measured by reverse transcription polymerase chain reaction in human papillomavirus type 18 deoxyribonucleic acid immortalized prostate cell lines. Urology *51*, 657-662.

Willis, A., Jung, E. J., Wakefield, T., and Chen, X. (2004). Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. Oncogene.

www.atcc.org (American Type Culture Collection).

www.cbil.upenn.edu/tess (TESS: Transcription Element Search System).

www.imaginis.com (Imaginis Corporation).

Xia, W., Lau, Y. K., Hu, M. C., Li, L., Johnston, D. A., Sheng, S. j., El-Naggar, A., and Hung, M. C. (2000). High tumoral maspin expression is associated with improved survival of patients with oral squamous cell carcinoma. Oncogene *19*, 2398-2403.

Yamada, N., Tamai, Y., Miyamoto, H., and Nozaki, M. (2000). Cloning and expression of the mouse Pse gene encoding a novel Ets family member. Gene *241*, 267-274.

Yu, J., Angelin-Duclos, C., Greenwood, J., Liao, J., and Calame, K. (2000). Transcriptional Repression by Blimp-1 (PRDI-BF1) Involves Recruitment of Histone Deacetylase. Mol Cell Biol *20*, 2592-2603.

Zhang, M., Maass, N., Magit, D., and Sager, R. (1997a). Transactivation through Ets and Ap1 transcription sites determines the expression of the tumor suppressing gene maspin. Cell Growth Differ *8*, 179-186.

Zhang, M., Magit, D., Botteri, F., Shi, H. Y., He, K., Li, M., Furth, P., and Sager, R. (1999). Maspin plays an important role in mammary gland development. Dev Biol *215*, 278-287.

Zhang, M., Magit, D., and Sager, R. (1997b). Expression of maspin in prostate cells is regulated by a positive ets element and a negative hormonal responsive element site recognized by androgen receptor. Proc Natl Acad Sci U S A *94*, 5673-5678.

Zhang, M., Sheng, S., Maass, N., and Sager, R. (1997c). mMaspin: the mouse homolog of a human tumor suppressor gene inhibits mammary tumor invasion and motility. Mol Med *3*, 49-59.

Zhang, M., Shi, Y., Magit, D., Furth, P. A., and Sager, R. (2000a). Reduced mammary tumor progression in WAP-TAg/WAP-maspin bitransgenic mice. Oncogene *19*, 6053-6058.

Zhang, M., Volpert, O., Shi, Y. H., and Bouck, N. (2000b). Maspin is an angiogenesis inhibitor. Nat Med *6*, 196-199.

Zou, Z., Anisowicz, A., Hendrix, M. J., Thor, A., Neveu, M., Sheng, S., Rafidi, K., Seftor, E., and Sager, R. (1994). Maspin, a serpin with tumor suppressing activity in human mammary epithelial cells. Science *263*, 526-529.

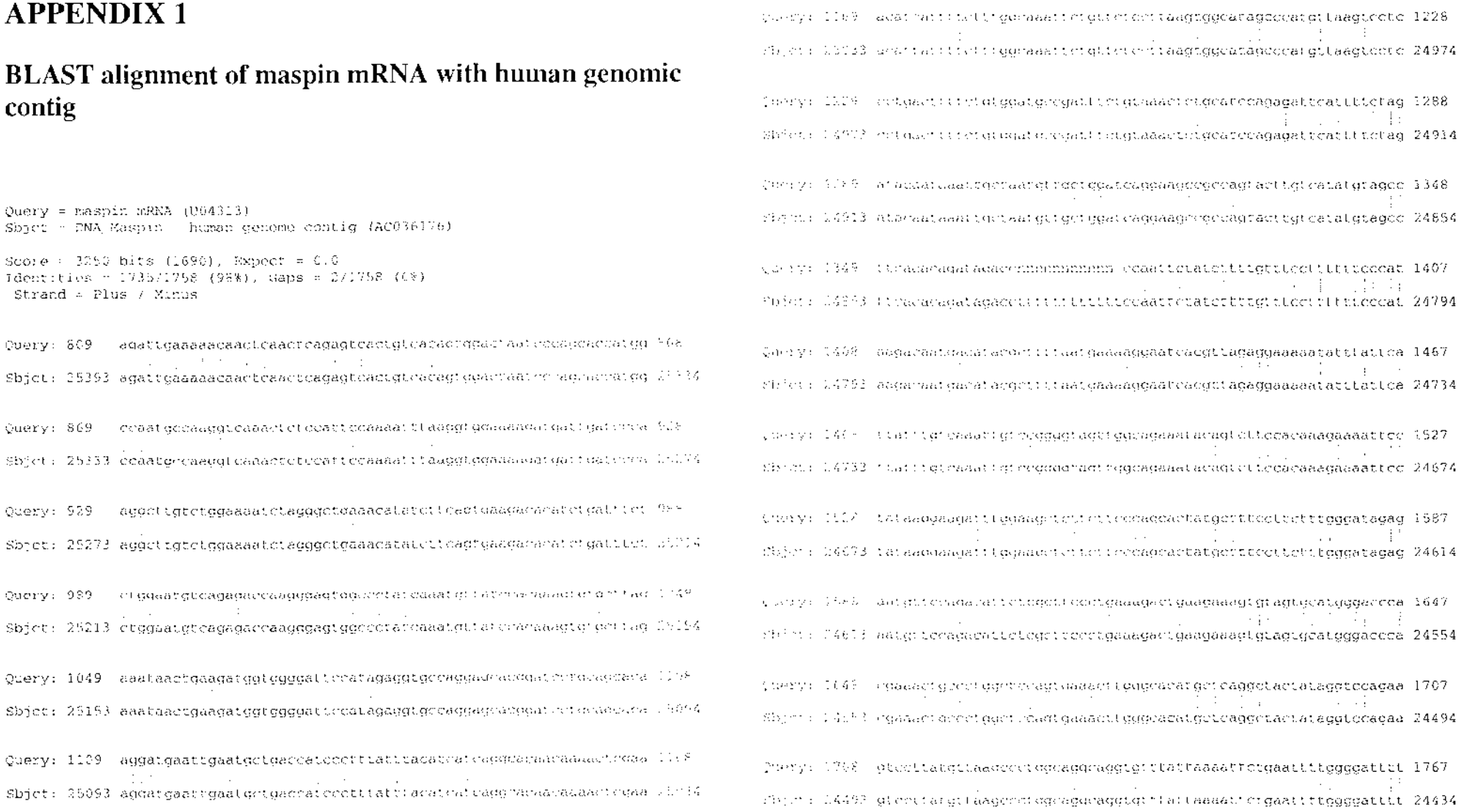
Zou, Z., Gao, C., Nagaich, A. K., Connell, T., Saito, S., Moul, J. W., Seth, P., Appella, E., and Srivastava, S. (2000). p53 regulates the expression of the tumor suppressor gene maspin. J Biol Chem *275*, 6051-6054.

Zou, Z., Zhang, W., Young, D., Gleave, M. G., Rennie, P., Connell, T., Connelly, R., Moul, J., Srivastava, S., and Sesterhenn, I. (2002). Maspin Expression Profile in Human Prostate Cancer (CaP) and in Vitro Induction of Maspin Expression by Androgen Ablation. Clin Cancer Res *8*, 1172-1177.

Appendix

APPENDIX 1

BLAST alignment of maspin mRNA with human genomic contig



```
Query: 1768  caaaagataatattttacatacactgtatgttatagaacttcatggatcagatctggggc 1827
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 24433 caaaagataatattttacatacactgtatgttatagaacttcatggatcagatctggggc 24374

Query: 1828  agcaacctataaatcaacaccttaatatgctgcaacaaaatgtagaatattcagacaaaa 1887
             |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 24373 agcaccctataaatcaacaccttaatatgctgcaacaaaatgtagaatattcagacaaaa 24314

Query: 1888  tggatacataaagactaagtagcccataaggggtcaaaatttgctgccaaatgcgtatgc 1947
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 24313 tggatacataaagactaagtagcccataaggggtcaaaatttgctgccaaatgcgtatgc 24254

Query: 1948  caccaacttacaaaaacacttcgttcgcagagcttttcagattgtggaatgttggataag 2007
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 24253 caccaacttacaaaaacacttcgttcgcagagcttttcagattgtggaatgttggataag 24194

Query: 2008  gaattatagacctctagtagctgaaatgcaagaccccaagaggaagttcagatcttaata 2067
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 24193 gaattatagacctctagtagctgaaatgcaagaccccaagaggaagttcagatcttaata 24134

Query: 2068  taaattcactttcatttttgatagctgtcccatctggtcatgtggttggcactagactgg 2127
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct: 24133 taaattcactttcatttttgatagctgtcccatctggtcatttggttggcactagactgg 24074

Query: 2128  tggcaggggcttctagctgactcgcacagggattctcacaatagccgatatcagaatttg 2187
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct: 24073 tggcaggggcttctagctgacttgcacagggattctcacaatagccgatatcagaatttg 24014

Query: 2188  tgttgaaggaacttgtctcttcatctaatatgatagcgggaaaaggagaggaaactactg 2247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 24013 tgttgaaggaacttgtctcttcatctaatatgatagcgggaaaaggagaggaaactactg 23954

Query: 2248  cctttagaaaatataagtaaagtgattaaagtgctcacgttaccttgacacatagttttt 2307
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 23953 cctttagaaaatataagtaaagtgattaaagtgctcacgttaccttgacacatagttttt 23894

Query: 2308  cagtctatgggtttagttactttagatggcaagcatgtaacttatattaatagtaatttg 2367
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 23893 cagtctatgggtttagttactttagatggcaagcatgtaacttatattaatagtaatttg 23834
```

```
Query: 2368  taaagttgggtggataagctatccctgttgccggttcatggattacttctctataaaaaa 2427
             ||||||||| |||||||||||||| |||||| ||||||||||||||||||||||||||||
Sbjct: 23833 taaagttggttggataagctatccgtgttgcaggttcatggattacttctctataaaaaa 23774

Query: 2428  tatatatttaccaaaaaattttgtgacattccttctcccatctcttccttgacatgcatt 2487
             ||| |||||||| |||||||||||||||||||||||||||||||||||||||| ||||||
Sbjct: 23773 tatgtatttacc-aaaaattttgtgacattccttctcccatctcttccttgacctgcatt 23715

Query: 2488  gtaaataggttcttcttgttctgagattcaatattgaatttctcctatgctattgacaat 2547
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct: 23714 gtaaataggttcttcttgttctgagattcaatattgaatttttcctatgctattgacaat 23655

Query: 2548  aaaatattattgaactac 2565
             |||||||||||||||||
Sbjct: 23654 aaaatattattgaactac 23637

Score =  350 bits (182), Expect = 5e-93
Identities = 182/182 (100%)
 Strand = Plus / Minus

Query: 64    tccaggcccgcaatggatgccctgcaactagcaaattcggcttttgccgttgatctgttc 123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 44305 tccaggcccgcaatggatgccctgcaactagcaaattcggcttttgccgttgatctgttc 44246

Query: 124   aaacaactatgtgaaaaggagccactgggcaatgtcctcttctctccaatctgtctctcc 183
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 44245 aaacaactatgtgaaaaggagccactgggcaatgtcctcttctctccaatctgtctctcc 44186

Query: 184   acctctctgtcacttgctcaagtgggtgctaaaggtgacactgcaaatgaaattggacag 243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 44185 acctctctgtcacttgctcaagtgggtgctaaaggtgacactgcaaatgaaattggacag 44126

Query: 244   gt 245
             ||
Sbjct: 44125 gt 44124
```

```
Score =  327 bits (170), Expect = 5e-86
Identities = 170/170 (100%)
 Strand = Plus / Minus


Query: 641   agacagacaccaaaccagtgcagatgatgaacatggaggccacgttctgtatgggaaaca 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 29603 agacagacaccaaaccagtgcagatgatgaacatggaggccacgttctgtatgggaaaca 29544


Query: 701   ttgacagtatcaattgtaagatcatagagcttccttttcaaaataagcatctcagcatgt 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 29543 ttgacagtatcaattgtaagatcatagagcttccttttcaaaataagcatctcagcatgt 29484


Query: 761   tcatcctactacccaaggatgtggaggatgagtccacaggcttggagaag 810
              ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 29483 tcatcctactacccaaggatgtggaggatgagtccacaggcttggagaag 29434


Score =  269 bits (140), Expect = 1e-68
Identities = 142/143 (99%)
 Strand = Plus / Minus


Query: 240   acaggttcttcattttgaaaatgtcaaagatataccctttggatttcaaacagtaacatc 299
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct: 41780 acaggttcttcattttgaaaatgtcaaagatgtaccctttggatttcaaacagtaacatc 41721


Query: 300   ggatgtaaacaaacttagttccttttactcactgaaactaatcaagcggctctacgtaga 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 41720 ggatgtaaacaaacttagttccttttactcactgaaactaatcaagcggctctacgtaga 41661


Query: 360   caaatctctgaatctttctacag 382
              |||||||||||||||||||||||
Sbjct: 41660 caaatctctgaatctttctacag 41638


Score =  266 bits (138), Expect = 2e-67
Identities = 142/144 (98%)
 Strand = Plus / Minus


Query: 499   ggccactttgagaacattttagctgacaacagtgtgaacgaccagaccaaaatccttgtg 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 35769 ggccactttgagaacattttagctgacaacagtgtgaacgaccagaccaaaatccttgtg 35710


Query: 559   gttaatgctgcctactttgttggcaagtggatgaagaaatttcctgaatcagaaacaaaa 618
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct: 35709 gttaatgctgcctactttgttggcaagtggatgaagaaattttctgaatcagaaacaaaa 35650


Query: 619   gaatgtcctttcagactcaacaag 642
              ||||||||||||||| ||||||||
Sbjct: 35649 gaatgtcctttcagagtcaacaag 35626


Score =  225 bits (117), Expect = 2e-55
Identities = 119/120 (99%)
 Strand = Plus / Minus


Query: 382   gagttcatcagctctacgaagagaccctatgcaaaggaattggaaactgttgacttcaaa 441
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct: 39374 gagttcatcagctctacgaagagaccgtatgcaaaggaattggaaactgttgacttcaaa 39315


Query: 442   gataaattggaagaaacgaaaggtcagatcaacaactcaattaaggatctcacagatggc 501
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 39314 gataaattggaagaaacgaaaggtcagatcaacaactcaattaaggatctcacagatggc 39255


Score =  117 bits (61), Expect = 6e-23
Identities = 61/61 (100%)
 Strand = Plus / Minus


Query: 9     ttgtgctcctcgcttgcctgttccttttccacgcattttccaggataactgtgactccag 68
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 51736 ttgtgctcctcgcttgcctgttccttttccacgcattttccaggataactgtgactccag 51677


Query: 69    g 69
              |
Sbjct: 51676 g 51676
```

```
CPU time:     1.42 user secs.       0.03 sys. secs          1.45 total secs.

Lambda     K      H
    1.33    0.621     1.12

Gapped
Lambda     K      H
    1.33    0.621     1.12


Matrix: blastn matrix:1 -2
Gap Penalties: Existence: 5, Extension: 2
Number of Hits to DB: 411
Number of Sequences: 0
Number of extensions: 411
Number of successful extensions: 55
Number of sequences better than 10.0: 1
length of query: 2566
length of database: 7,092,611,580
effective HSP length: 26
effective length of query: 2540
effective length of database: 7,092,611,554
effective search space: 18015233347160
effective search space used: 18015233347160
T: 0
A: 0
X1: 6 (11.5 bits)
X2: 26 (50.0 bits)
S1: 12 (23.8 bits)
S2: 21 (41.1 bits)
```

APPENDIX 2 - TESS analysis of maspin 8 kb promoter

8 kb of the putative maspin promoter was analysed by Transcriptional Element Search Software (Schug and Overton, 1997). Results are tabulated below.

Index is the base pair at which the protein binding site begins. M indicates the strand which the protein binds to, N normal, C complementary

LLH is the $\log_2$ of the likelihood that a protein would by chance recognise the site, mismatches are incorporated into the algorithm.

LLH Den is the average log likelihood over the whole site. 2 is the maximum value, -1 is the minimum.

LLH Q is the ratio of LLH to maximum possible LLH of that site. A scored of 1 indicates the best possible match, a lower score indicates a less than perfect match.

Sequence is the factor binding site sequence. Upper-case indicates matches, lower-case denotes mis-matches.

Factor is the name of the factor which binds to the sequence.

The transcription start site is located at base 8201, the cloned maspin promoter extends from base 5651 to 8381.

Maspin_all

With 10 percent mismatches allowed

Index	M	LLH	LLH Den	LLH Q	Sequence	Factor Name
2	N	12	1.09	1	CCWTNTTNNNW	YY1
4	C	12	2	1	ATTTTA	F2F
57	C	16	1.60	0.80	TGtTTTACAT	NF-IL-2A
57	C	16	1.60	0.80	TGtTTTACAT	Oct-2.1
57	N	16	1.60	0.80	TGCTTTgCAT	Oct-01
57	N	16	1.60	0.80	TGCTTTgCAT	octamer-binding factor
78	C	16	1.60	0.80	GaTGATTCAT	AP-1
94	C	16	1.60	0.80	gCCCCGCCCC	MAZ
94	C	16	1.60	0.80	gCCCCGCCCC	Sp1
94	C	16	1.60	0.80	gCCCCGCCCC	TFIID
97	C	12	2	1	CCGCCC	ETF
97	N	16	1.60	0.80	CCGCCCCcGC	EGR2
105	N	18	1.64	0.82	GCCAcATGACC	NF-muE3
105	N	18	1.64	0.82	GCCAcATGACC	TFE3
127	N	12	1.33	1	MATNNWAAT	N-Oct-3
140	C	16	1.60	0.80	TCCCTaCCCA	Sp1
143	C	12	2	1	CTTCCC	c-Ets-2
151	C	16	1.60	0.80	GGGGTGGGGc	Sp1
152	N	14	2	1	GGGTGGG	PuF
155	N	12	2	1	TGGGGA	AP-2
166	N	12	2	1	TGGGGA	AP-2
181	C	12	2	1	TGTACA	GR alpha
181	C	12	2	1	TGTACA	GR beta
181	C	12	2	1	TGTACA	PR
181	C	12	2	1	TGTACA	PR A
204	C	14	1.75	1	ATGRATAW	Pit-1
250	N	12	1.33	1	MATNNWAAT	N-Oct-3

257	C	16	1.60	0.80	ATTTAATTGa	Oct-2.1
261	N	12	1	0.75	AtTTNNNNATTT	HiNF-A
264	C	12	2	1	TGTACA	GR alpha
264	C	12	2	1	TGTACA	GR beta
264	C	12	2	1	TGTACA	PR
264	C	12	2	1	TGTACA	PR A
275	C	12	1	0.75	AAATNNNNAAAt	HiNF-A
278	N	14	2	1	TGACTAA	AP-1
294	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
317	N	12	1.71	1	MATWAAT	N-Oct-3
352	N	12	1.20	0.75	RTKAYGTaAY	E4BP4
363	C	12	1.20	1	ATTWNNNATK	N-Oct-3
371	C	16	1.60	0.80	TTGCATGCCc	Oct-01
385	N	12	2	1	TAAAAT	F2F
404	N	12	1.71	1	MATWAAT	N-Oct-3
405	N	14	2	1	ATAAATA	ETF
412	C	12	2	1	TGTACA	GR alpha
412	C	12	2	1	TGTACA	GR beta
412	C	12	2	1	TGTACA	PR
412	C	12	2	1	TGTACA	PR A
442	C	12	1	0.75	AAAtNNNNAAAT	HiNF-A
446	N	14	2	1	TAAAAAA	TBP
471	N	12	2	1	CTTTCC	NF-1
496	C	14	2	1	TGTGTTT	HiNF-A
550	N	16	2	1	TGACATCA	AP-1
550	N	16	2	1	TGACATCA	CREB
550	N	16	2	1	TGACATCA	c-Fos
577	N	16	1.60	0.80	aAGAGGAAAA	NFAT-1
577	N	16	1.60	0.80	aAGAGGAAAA	Pu box binding factor
592	C	16	1.60	0.80	TTCtTATCAA	GATA-1
606	C	16	2	1	TTTCTATA	TBP
618	N	12	1.20	1	MATNNNWAAT	N-Oct-3
621	N	12	2	1	TATAAA	TBP
621	N	14	2	1	TATAAAT	TMF
621	N	16	2	1	TATAAATA	TFIID
622	N	14	2	1	ATAAATA	ETF
630	C	12	1.20	0.75	RTTaCRTMAY	E4BP4
630	N	12	1	0.75	ATTTNNNNATTt	HiNF-A
644	C	16	2	1	ATTATATA	TBP
644	C	16	2	1	ATTATATA	Sp1
644	C	16	2	1	ATTATATA	TFIID
663	N	12	1	0.75	AtTTNNNNATTT	HiNF-A
671	C	12	2	1	ATTTTA	F2F
672	C	14	2	1	TTTTATA	TFIIA
672	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
672	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
672	C	14	2	1	TTTTATA	TFIIA-gamma
672	C	14	2	1	TTTTATA	TFIIB
672	C	14	2	1	TTTTATA	TFIIE
672	C	14	2	1	TTTTATA	TFIIF
672	C	14	2	1	TTTTATA	TFIIF-alpha
672	C	14	2	1	TTTTATA	TFIIF-beta
672	C	14	2	1	TTTTATA	TBP
672	C	14	2	1	TTTTATA	TFIID
672	C	14	2	1	TTTTATA	Dr1
675	N	16	2	1	TATATAAT	Sp1
680	N	12	1.33	1	MATNNWAAT	N-Oct-3
682	N	12	2	1	TATAAA	TBP
682	N	14	2	1	TATAAAT	TMF
682	N	16	2	1	TATAAATA	TFIID
683	N	14	2	1	ATAAATA	ETF
692	C	16	1.60	0.80	TGTTTTAcAT	NF-IL-2A
692	C	16	1.60	0.80	TGTTTTAcAT	Oct-01
692	C	16	1.60	0.80	TGTTTTAcAT	Oct-2.1
694	C	14	2	1	TTTTATA	TFIIA
694	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)

694	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)	735	C	14	2	1	TTTTATA	TFIIA-gamma
694	C	14	2	1	TTTTATA	TFIIA-gamma	735	C	14	2	1	TTTTATA	TFIIB
694	C	14	2	1	TTTTATA	TFIIB	735	C	14	2	1	TTTTATA	TFIIE
694	C	14	2	1	TTTTATA	TFIIE	735	C	14	2	1	TTTTATA	TFIIF
694	C	14	2	1	TTTTATA	TFIIF	735	C	14	2	1	TTTTATA	TFIIF-alpha
694	C	14	2	1	TTTTATA	TFIIF-alpha	735	C	14	2	1	TTTTATA	TFIIF-beta
694	C	14	2	1	TTTTATA	TFIIF-beta	735	C	14	2	1	TTTTATA	TBP
694	C	14	2	1	TTTTATA	TBP	735	C	14	2	1	TTTTATA	TFIID
694	C	14	2	1	TTTTATA	TFIID	735	C	14	2	1	TTTTATA	Dr1
694	C	14	2	1	TTTTATA	Dr1	743	C	12	2	1	ATTTTA	F2F
702	N	12	1	0.75	AtTTNNNNATTT	HiNF-A	744	C	14	2	1	TTTTATA	TFIIA
710	C	12	2	1	ATTTTA	F2F	744	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
711	C	14	2	1	TTTTATA	TFIIA	744	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
711	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)	744	C	14	2	1	TTTTATA	TFIIA-gamma
711	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)	744	C	14	2	1	TTTTATA	TFIIB
711	C	14	2	1	TTTTATA	TFIIA-gamma	744	C	14	2	1	TTTTATA	TFIIE
711	C	14	2	1	TTTTATA	TFIIB	744	C	14	2	1	TTTTATA	TFIIF
711	C	14	2	1	TTTTATA	TFIIE	744	C	14	2	1	TTTTATA	TFIIF-alpha
711	C	14	2	1	TTTTATA	TFIIF	744	C	14	2	1	TTTTATA	TFIIF-beta
711	C	14	2	1	TTTTATA	TFIIF-alpha	744	C	14	2	1	TTTTATA	TBP
711	C	14	2	1	TTTTATA	TFIIF-beta	744	C	14	2	1	TTTTATA	TFIID
711	C	14	2	1	TTTTATA	TBP	744	C	14	2	1	TTTTATA	Dr1
711	C	14	2	1	TTTTATA	TFIID	747	N	16	2	1	TATATAAT	Sp1
711	C	14	2	1	TTTTATA	Dr1	757	C	12	1	0.75	AaATNNNNAAAT	HiNF-A
714	N	16	2	1	TATATAAT	Sp1	762	N	12	2	1	TATAAA	TFIID
724	C	12	1	0.75	AaATNNNNAAAT	HiNF-A	762	N	12	2	1	TATAAA	TBP
729	N	12	2	1	TATAAA	TFIID	762	N	14	2	1	TATAAAT	TMF
729	N	12	2	1	TATAAA	TBP	767	C	12	2	1	ATTTTA	F2F
729	N	14	2	1	TATAAAT	TMF	768	C	14	2	1	TTTTATA	TFIIA
734	C	12	2	1	ATTTTA	F2F	768	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
735	C	14	2	1	TTTTATA	TFIIA	768	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
735	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)	768	C	14	2	1	TTTTATA	TFIIA-gamma
735	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)	768	C	14	2	1	TTTTATA	TFIIB

768	C	14	2	1	TTTTATA	TFIIE
768	C	14	2	1	TTTTATA	TFIIF
768	C	14	2	1	TTTTATA	TFIIF-alpha
768	C	14	2	1	TTTTATA	TFIIF-beta
768	C	14	2	1	TTTTATA	TBP
768	C	14	2	1	TTTTATA	TFIID
768	C	14	2	1	TTTTATA	Dr1
775	C	14	2	1	TATTTAT	ETF
775	C	16	2	1	TATTTATA	TFIID
776	C	14	2	1	ATTTATA	TMF
776	N	16	1.33	1	ATTTNNNNATTT	HiNF-A
777	C	12	2	1	TTTATA	TBP
784	C	12	2	1	ATTTTA	F2F
785	C	14	2	1	TTTTATA	TFIIA
785	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
785	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
785	C	14	2	1	TTTTATA	TFIIA-gamma
785	C	14	2	1	TTTTATA	TFIIB
785	C	14	2	1	TTTTATA	TFIIE
785	C	14	2	1	TTTTATA	TFIIF
785	C	14	2	1	TTTTATA	TFIIF-alpha
785	C	14	2	1	TTTTATA	TFIIF-beta
785	C	14	2	1	TTTTATA	TBP
785	C	14	2	1	TTTTATA	TFIID
785	C	14	2	1	TTTTATA	Dr1
788	N	16	2	1	TATATAAT	Sp1
803	N	12	2	1	TATAAA	TFIID
803	N	12	2	1	TATAAA	TBP
803	N	14	2	1	TATAAAT	TMF
808	C	12	2	1	ATTTTA	F2F
809	C	14	2	1	TTTTATA	TFIIA
809	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
809	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
809	C	14	2	1	TTTTATA	TFIIA-gamma
809	C	14	2	1	TTTTATA	TFIIB
809	C	14	2	1	TTTTATA	TFIIE
809	C	14	2	1	TTTTATA	TFIIF
809	C	14	2	1	TTTTATA	TFIIF-alpha
809	C	14	2	1	TTTTATA	TFIIF-beta
809	C	14	2	1	TTTTATA	TBP
809	C	14	2	1	TTTTATA	TFIID
809	C	14	2	1	TTTTATA	Dr1
809	N	12	1	0.75	aTTTNNNNATTT	HiNF-A
817	C	12	2	1	ATTTTA	F2F
818	C	14	2	1	TTTTATA	TFIIA
818	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
818	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
818	C	14	2	1	TTTTATA	TFIIA-gamma
818	C	14	2	1	TTTTATA	TFIIB
818	C	14	2	1	TTTTATA	TFIIE
818	C	14	2	1	TTTTATA	TFIIF
818	C	14	2	1	TTTTATA	TFIIF-alpha
818	C	14	2	1	TTTTATA	TFIIF-beta
818	C	14	2	1	TTTTATA	TBP
818	C	14	2	1	TTTTATA	TFIID
818	C	14	2	1	TTTTATA	Dr1
821	N	16	2	1	TATATAAT	Sp1
831	C	12	1	0.75	AaATNNNNAAAT	HiNF-A
836	N	12	2	1	TATAAA	TFIID
836	N	12	2	1	TATAAA	TBP
836	N	14	2	1	TATAAAT	TMF
841	C	12	2	1	ATTTTA	F2F
842	C	14	2	1	TTTTATA	TFIIA
842	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
842	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
842	C	14	2	1	TTTTATA	TFIIA-gamma

842	C	14	2	1	TTTTATA	TFIIB
842	C	14	2	1	TTTTATA	TFIIE
842	C	14	2	1	TTTTATA	TFIIF
842	C	14	2	1	TTTTATA	TFIIF-alpha
842	C	14	2	1	TTTTATA	TFIIF-beta
842	C	14	2	1	TTTTATA	TBP
842	C	14	2	1	TTTTATA	TFIID
842	C	14	2	1	TTTTATA	Dr1
845	N	16	2	1	TATATAAT	Sp1
855	C	12	1	0.75	AaATNNNNAAAT	HiNF-A
860	N	12	2	1	TATAAA	TFIID
860	N	12	2	1	TATAAA	TBP
860	N	14	2	1	TATAAAT	TMF
865	C	12	2	1	ATTTTA	F2F
866	C	14	2	1	TTTTATA	TFIIA
866	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
866	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
866	C	14	2	1	TTTTATA	TFIIA-gamma
866	C	14	2	1	TTTTATA	TFIIB
866	C	14	2	1	TTTTATA	TFIIE
866	C	14	2	1	TTTTATA	TFIIF
866	C	14	2	1	TTTTATA	TFIIF-alpha
866	C	14	2	1	TTTTATA	TFIIF-beta
866	C	14	2	1	TTTTATA	TBP
866	C	14	2	1	TTTTATA	TFIID
866	C	14	2	1	TTTTATA	Dr1
874	C	12	1.20	1	ATTWNNNATK	N-Oct-3
874	C	12	2	1	ATTTTA	F2F
875	C	14	2	1	TTTTATA	TFIIA
875	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
875	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
875	C	14	2	1	TTTTATA	TFIIA-gamma
875	C	14	2	1	TTTTATA	TFIIB
875	C	14	2	1	TTTTATA	TFIIE
875	C	14	2	1	TTTTATA	TFIIF
875	C	14	2	1	TTTTATA	TFIIF-alpha
875	C	14	2	1	TTTTATA	TFIIF-beta
875	C	14	2	1	TTTTATA	TBP
875	C	14	2	1	TTTTATA	TFIID
875	C	14	2	1	TTTTATA	Dr1
891	C	12	1	0.75	AaATNNNNAAAT	HiNF-A
896	N	12	2	1	TATAAA	TFIID
896	N	12	2	1	TATAAA	TBP
896	N	14	2	1	TATAAAT	TMF
901	C	12	2	1	ATTTTA	F2F
902	C	14	2	1	TTTTATA	TFIIA
902	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
902	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
902	C	14	2	1	TTTTATA	TFIIA-gamma
902	C	14	2	1	TTTTATA	TFIIB
902	C	14	2	1	TTTTATA	TFIIE
902	C	14	2	1	TTTTATA	TFIIF
902	C	14	2	1	TTTTATA	TFIIF-alpha
902	C	14	2	1	TTTTATA	TFIIF-beta
902	C	14	2	1	TTTTATA	TBP
902	C	14	2	1	TTTTATA	TFIID
902	C	14	2	1	TTTTATA	Dr1
910	C	12	2	1	ATTTTA	F2F
911	C	14	2	1	TTTTATA	TFIIA
911	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
911	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
911	C	14	2	1	TTTTATA	TFIIA-gamma
911	C	14	2	1	TTTTATA	TFIIB
911	C	14	2	1	TTTTATA	TFIIE
911	C	14	2	1	TTTTATA	TFIIF
911	C	14	2	1	TTTTATA	TFIIF-alpha

911	C	14	2	1	TTTTATA	TFIIF-beta
911	C	14	2	1	TTTTATA	TBP
911	C	14	2	1	TTTTATA	TFIID
911	C	14	2	1	TTTTATA	Dr1
914	N	16	2	1	TATATAAT	Sp1
924	C	12	1	0.75	AaATNNNNAAAT	HiNF-A
929	N	12	2	1	TATAAA	TFIID
929	N	12	2	1	TATAAA	TBP
929	N	14	2	1	TATAAAT	TMF
934	C	12	2	1	ATTTTA	F2F
935	C	14	2	1	TTTTATA	TFIIA
935	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
935	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
935	C	14	2	1	TTTTATA	TFIIA-gamma
935	C	14	2	1	TTTTATA	TFIIB
935	C	14	2	1	TTTTATA	TFIIE
935	C	14	2	1	TTTTATA	TFIIF
935	C	14	2	1	TTTTATA	TFIIF-alpha
935	C	14	2	1	TTTTATA	TFIIF-beta
935	C	14	2	1	TTTTATA	TBP
935	C	14	2	1	TTTTATA	TFIID
935	C	14	2	1	TTTTATA	Dr1
942	C	14	2	1	TATTTAT	ETF
942	C	16	2	1	TATTTATA	TFIID
943	C	14	2	1	ATTTATA	TMF
943	N	16	1.33	1	ATTTNNNNATTT	HiNF-A
944	C	12	2	1	TTTATA	TBP
951	C	12	2	1	ATTTTA	F2F
952	C	14	2	1	TTTTATA	TFIIA
952	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
952	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
952	C	14	2	1	TTTTATA	TFIIA-gamma
952	C	14	2	1	TTTTATA	TFIIB
952	C	14	2	1	TTTTATA	TFIIE
952	C	14	2	1	TTTTATA	TFIIF
952	C	14	2	1	TTTTATA	TFIIF-alpha
952	C	14	2	1	TTTTATA	TFIIF-beta
952	C	14	2	1	TTTTATA	TBP
952	C	14	2	1	TTTTATA	TFIID
952	C	14	2	1	TTTTATA	Dr1
955	N	16	2	1	TATATAAT	Sp1
977	C	14	2	1	TATTTAT	ETF
977	C	16	2	1	TATTTATA	TFIID
978	C	14	2	1	ATTTATA	TMF
978	N	16	1.33	1	ATTTNNNNATTT	HiNF-A
979	C	12	2	1	TTTATA	TBP
986	C	12	2	1	ATTTTA	F2F
987	C	14	2	1	TTTTATA	TFIIA
987	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
987	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
987	C	14	2	1	TTTTATA	TFIIA-gamma
987	C	14	2	1	TTTTATA	TFIIB
987	C	14	2	1	TTTTATA	TFIIE
987	C	14	2	1	TTTTATA	TFIIF
987	C	14	2	1	TTTTATA	TFIIF-alpha
987	C	14	2	1	TTTTATA	TFIIF-beta
987	C	14	2	1	TTTTATA	TBP
987	C	14	2	1	TTTTATA	TFIID
987	C	14	2	1	TTTTATA	Dr1
990	N	16	2	1	TATATAAT	Sp1
1011	N	14	1.75	1	WTATYCAT	Pit-1
1013	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
1021	C	12	2	1	ATTTTA	F2F
1022	C	14	2	1	TTTTATA	TFIIA
1022	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1022	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)

1022	C	14	2	1	TTTTATA	TFIIA-gamma
1022	C	14	2	1	TTTTATA	TFIIB
1022	C	14	2	1	TTTTATA	TFIIE
1022	C	14	2	1	TTTTATA	TFIIF
1022	C	14	2	1	TTTTATA	TFIIF-alpha
1022	C	14	2	1	TTTTATA	TFIIF-beta
1022	C	14	2	1	TTTTATA	TBP
1022	C	14	2	1	TTTTATA	TFIID
1022	C	14	2	1	TTTTATA	Dr1
1025	N	16	2	1	TATATAAT	Sp1
1046	N	14	1.75	1	WTATYCAT	Pit-1
1048	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
1056	C	12	2	1	ATTTTA	F2F
1057	C	14	2	1	TTTTATA	TFIIA
1057	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1057	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1057	C	14	2	1	TTTTATA	TFIIA-gamma
1057	C	14	2	1	TTTTATA	TFIIB
1057	C	14	2	1	TTTTATA	TFIIE
1057	C	14	2	1	TTTTATA	TFIIF
1057	C	14	2	1	TTTTATA	TFIIF-alpha
1057	C	14	2	1	TTTTATA	TFIIF-beta
1057	C	14	2	1	TTTTATA	TBP
1057	C	14	2	1	TTTTATA	TFIID
1057	C	14	2	1	TTTTATA	Dr1
1060	N	16	2	1	TATATAAT	Sp1
1081	N	14	1.75	1	WTATYCAT	Pit-1
1083	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
1091	C	12	2	1	ATTTTA	F2F
1092	C	14	2	1	TTTTATA	TFIIA
1092	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1092	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1092	C	14	2	1	TTTTATA	TFIIA-gamma
1092	C	14	2	1	TTTTATA	TFIIB
1092	C	14	2	1	TTTTATA	TFIIE
1092	C	14	2	1	TTTTATA	TFIIF
1092	C	14	2	1	TTTTATA	TFIIF-alpha
1092	C	14	2	1	TTTTATA	TFIIF-beta
1092	C	14	2	1	TTTTATA	TBP
1092	C	14	2	1	TTTTATA	TFIID
1092	C	14	2	1	TTTTATA	Dr1
1095	N	16	2	1	TATATAAT	Sp1
1116	N	14	1.75	1	WTATYCAT	Pit-1
1118	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
1126	C	12	2	1	ATTTTA	F2F
1127	C	14	2	1	TTTTATA	TFIIA
1127	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1127	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1127	C	14	2	1	TTTTATA	TFIIA-gamma
1127	C	14	2	1	TTTTATA	TFIIB
1127	C	14	2	1	TTTTATA	TFIIE
1127	C	14	2	1	TTTTATA	TFIIF
1127	C	14	2	1	TTTTATA	TFIIF-alpha
1127	C	14	2	1	TTTTATA	TFIIF-beta
1127	C	14	2	1	TTTTATA	TBP
1127	C	14	2	1	TTTTATA	TFIID
1127	C	14	2	1	TTTTATA	Dr1
1130	N	16	2	1	TATATAAT	Sp1
1151	N	14	1.75	1	WTATYCAT	Pit-1
1153	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
1161	C	12	2	1	ATTTTA	F2F
1162	C	14	2	1	TTTTATA	TFIIA
1162	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1162	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1162	C	14	2	1	TTTTATA	TFIIA-gamma
1162	C	14	2	1	TTTTATA	TFIIB

1162	C	14	2	1	TTTTATA	TFIIE
1162	C	14	2	1	TTTTATA	TFIIF
1162	C	14	2	1	TTTTATA	TFIIF-alpha
1162	C	14	2	1	TTTTATA	TFIIF-beta
1162	C	14	2	1	TTTTATA	TBP
1162	C	14	2	1	TTTTATA	TFIID
1162	C	14	2	1	TTTTATA	Dr1
1165	N	16	2	1	TATATAAT	Sp1
1186	N	14	1.75	1	WTATYCAT	Pit-1
1188	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
1196	C	12	2	1	ATTTTA	F2F
1197	C	14	2	1	TTTTATA	TFIIA
1197	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1197	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1197	C	14	2	1	TTTTATA	TFIIA-gamma
1197	C	14	2	1	TTTTATA	TFIIB
1197	C	14	2	1	TTTTATA	TFIIE
1197	C	14	2	1	TTTTATA	TFIIF
1197	C	14	2	1	TTTTATA	TFIIF-alpha
1197	C	14	2	1	TTTTATA	TFIIF-beta
1197	C	14	2	1	TTTTATA	TBP
1197	C	14	2	1	TTTTATA	TFIID
1197	C	14	2	1	TTTTATA	Dr1
1200	N	16	2	1	TATATAAT	Sp1
1221	N	14	1.75	1	WTATYCAT	Pit-1
1223	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
1231	C	12	2	1	ATTTTA	F2F
1232	C	14	2	1	TTTTATA	TFIIA
1232	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1232	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1232	C	14	2	1	TTTTATA	TFIIA-gamma
1232	C	14	2	1	TTTTATA	TFIIB
1232	C	14	2	1	TTTTATA	TFIIE
1232	C	14	2	1	TTTTATA	TFIIF
1232	C	14	2	1	TTTTATA	TFIIF-alpha
1232	C	14	2	1	TTTTATA	TFIIF-beta
1232	C	14	2	1	TTTTATA	TBP
1232	C	14	2	1	TTTTATA	TFIID
1232	C	14	2	1	TTTTATA	Dr1
1235	N	16	2	1	TATATAAT	Sp1
1256	N	14	1.75	1	WTATYCAT	Pit-1
1258	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
1266	C	12	2	1	ATTTTA	F2F
1267	C	14	2	1	TTTTATA	TFIIA
1267	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1267	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1267	C	14	2	1	TTTTATA	TFIIA-gamma
1267	C	14	2	1	TTTTATA	TFIIB
1267	C	14	2	1	TTTTATA	TFIIE
1267	C	14	2	1	TTTTATA	TFIIF
1267	C	14	2	1	TTTTATA	TFIIF-alpha
1267	C	14	2	1	TTTTATA	TFIIF-beta
1267	C	14	2	1	TTTTATA	TBP
1267	C	14	2	1	TTTTATA	TFIID
1267	C	14	2	1	TTTTATA	Dr1
1270	N	16	2	1	TATATAAT	Sp1
1291	N	14	1.75	1	WTATYCAT	Pit-1
1293	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
1301	C	12	2	1	ATTTTA	F2F
1302	C	14	2	1	TTTTATA	TFIIA
1302	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1302	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1302	C	14	2	1	TTTTATA	TFIIA-gamma
1302	C	14	2	1	TTTTATA	TFIIB
1302	C	14	2	1	TTTTATA	TFIIE
1302	C	14	2	1	TTTTATA	TFIIF

1302	C	14	2	1	TTTTATA	TFIIF-alpha
1302	C	14	2	1	TTTTATA	TFIIF-beta
1302	C	14	2	1	TTTTATA	TBP
1302	C	14	2	1	TTTTATA	TFIID
1302	C	14	2	1	TTTTATA	Dr1
1305	N	16	2	1	TATATAAT	Sp1
1326	N	14	1.75	1	WTATYCAT	Pit-1
1328	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
1336	C	12	2	1	ATTTTA	F2F
1337	C	14	2	1	TTTTATA	TFIIA
1337	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1337	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1337	C	14	2	1	TTTTATA	TFIIA-gamma
1337	C	14	2	1	TTTTATA	TFIIB
1337	C	14	2	1	TTTTATA	TFIIE
1337	C	14	2	1	TTTTATA	TFIIF
1337	C	14	2	1	TTTTATA	TFIIF-alpha
1337	C	14	2	1	TTTTATA	TFIIF-beta
1337	C	14	2	1	TTTTATA	TBP
1337	C	14	2	1	TTTTATA	TFIID
1337	C	14	2	1	TTTTATA	Dr1
1340	N	16	2	1	TATATAAT	Sp1
1358	N	12	2	1	CACATG	USF
1363	N	12	1	0.75	aTTTNNNNATTT	HiNF-A
1364	C	12	2	1	TTTATA	TFIID
1364	C	12	2	1	TTTATA	TBP
1371	C	12	2	1	ATTTTA	F2F
1372	C	14	2	1	TTTTATA	TFIIA
1372	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1372	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1372	C	14	2	1	TTTTATA	TFIIA-gamma
1372	C	14	2	1	TTTTATA	TFIIB
1372	C	14	2	1	TTTTATA	TFIIE
1372	C	14	2	1	TTTTATA	TFIIF
1372	C	14	2	1	TTTTATA	TFIIF-alpha
1372	C	14	2	1	TTTTATA	TFIIF-beta
1372	C	14	2	1	TTTTATA	TBP
1372	C	14	2	1	TTTTATA	TFIID
1372	C	14	2	1	TTTTATA	Dr1
1375	N	16	2	1	TATATAAT	Sp1
1393	N	12	2	1	CACATG	USF
1398	N	12	1	0.75	aTTTNNNNATTT	HiNF-A
1399	C	12	2	1	TTTATA	TFIID
1399	C	12	2	1	TTTATA	TBP
1406	C	12	2	1	ATTTTA	F2F
1407	C	14	2	1	TTTTATA	TFIIA
1407	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1407	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1407	C	14	2	1	TTTTATA	TFIIA-gamma
1407	C	14	2	1	TTTTATA	TFIIB
1407	C	14	2	1	TTTTATA	TFIIE
1407	C	14	2	1	TTTTATA	TFIIF
1407	C	14	2	1	TTTTATA	TFIIF-alpha
1407	C	14	2	1	TTTTATA	TFIIF-beta
1407	C	14	2	1	TTTTATA	TBP
1407	C	14	2	1	TTTTATA	TFIID
1407	C	14	2	1	TTTTATA	Dr1
1410	N	16	2	1	TATATAAT	Sp1
1433	N	18	1.64	0.82	GTTTATcTGTT	GATA-1
1434	C	12	2	1	TTTATA	TFIID
1434	C	12	2	1	TTTATA	TBP
1440	C	16	1.60	0.80	TGTTTTAcAT	NF-IL-2A
1440	C	16	1.60	0.80	TGTTTTAcAT	Oct-01
1440	C	16	1.60	0.80	TGTTTTAcAT	Oct-2.1
1442	C	14	2	1	TTTTATA	TFIIA
1442	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)

1442	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1442	C	14	2	1	TTTTATA	TFIIA-gamma
1442	C	14	2	1	TTTTATA	TFIIB
1442	C	14	2	1	TTTTATA	TFIIE
1442	C	14	2	1	TTTTATA	TFIIF
1442	C	14	2	1	TTTTATA	TFIIF-alpha
1442	C	14	2	1	TTTTATA	TFIIF-beta
1442	C	14	2	1	TTTTATA	TBP
1442	C	14	2	1	TTTTATA	TFIID
1442	C	14	2	1	TTTTATA	Dr1
1442	N	12	1	0.75	aTTTNNNNATTT	HiNF-A
1449	C	14	2	1	TATTTAT	ETF
1449	C	16	2	1	TATTTATA	TFIID
1450	C	14	2	1	ATTTATA	TMF
1451	C	12	2	1	TTTATA	TBP
1487	N	12	1	0.75	ATtTNNNNATTT	HiNF-A
1494	C	14	2	1	TATTTAT	ETF
1495	C	12	1.71	1	ATTWATK	N-Oct-3
1499	N	12	1	0.75	ATTTNNNNaTTT	HiNF-A
1500	C	14	2	1	TTTTTTA	TBP
1506	C	12	1.33	1	ATTWNNATK	N-Oct-3
1514	C	14	2	1	TTTTTTA	TBP
1565	N	12	2	1	CTATCT	GATA-1
1578	C	13	1.30	0.81	rTGMAAYCYC	NF-GMa
1592	C	12	2	1	CTTCCC	c-Ets-2
1619	C	12	1.33	1	MYYMGCCYM	Sp1
1651	C	12	2	1	CCCGCC	Sp1
1657	N	16	1.60	0.80	ACCcCGCCCA	Sp1
1671	N	12	1	0.75	AtTTNNNNATTT	HiNF-A
1694	N	16	1.60	1	GRGRTTKCAY	NF-GMa
1742	C	13	1.63	1	CWWWCCAC	C/EBPalpha
1794	C	12	2	1	TGTACA	GR alpha
1794	C	12	2	1	TGTACA	GR beta
1794	C	12	2	1	TGTACA	PR
1794	C	12	2	1	TGTACA	PR A
1808	C	14	2	1	TTTTTTA	TBP
1808	N	12	1	0.75	aTTTNNNNATTT	HiNF-A
1810	C	14	2	1	TTTTATA	TFIIA
1810	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (major)
1810	C	14	2	1	TTTTATA	TFIIA-alpha/beta precursor (minor)
1810	C	14	2	1	TTTTATA	TFIIA-gamma
1810	C	14	2	1	TTTTATA	TFIIB
1810	C	14	2	1	TTTTATA	TFIIE
1810	C	14	2	1	TTTTATA	TFIIF
1810	C	14	2	1	TTTTATA	TFIIF-alpha
1810	C	14	2	1	TTTTATA	TFIIF-beta
1810	C	14	2	1	TTTTATA	TFIID
1810	C	14	2	1	TTTTATA	Dr1
1875	C	12	2	1	CACGTG	c-Myc
1893	N	16	1.60	0.80	TTGaTAAGAA	GATA-1
1947	C	12	2	1	GACCTG	TGT3
1954	N	12	2	1	GGGAAG	c-Ets-2
1960	C	12	1.71	1	TKASTCA	AP-1
1989	C	16	1.60	0.80	TTTTCCTcTT	NFAT-1
1989	C	16	1.60	0.80	TTTTCCTcTT	Pu box binding factor
1991	N	12	2	1	TTCCTT	c-Ets-2
2000	C	14	2	1	ATTTATA	TMF
2001	C	12	2	1	TTTATA	TFIID
2001	C	12	2	1	TTTATA	TBP
2014	N	16	1.60	0.80	ATGcAAATAG	octamer-binding factor
2042	C	16	1.60	0.80	aTGCAAAGCA	Oct-01
2042	C	16	1.60	0.80	aTGCAAAGCA	octamer-binding factor
2054	N	14	2	1	TATAAAA	TFIIA
2054	N	14	2	1	TATAAAA	TFIIA-alpha/beta precursor (major)
2054	N	14	2	1	TATAAAA	TFIIA-alpha/beta precursor (minor)
2054	N	14	2	1	TATAAAA	TFIIA-gamma

2054	N	14	2	1	TATAAAA	TFIIB
2054	N	14	2	1	TATAAAA	TFIIE
2054	N	14	2	1	TATAAAA	TFIIF
2054	N	14	2	1	TATAAAA	TFIIF-alpha
2054	N	14	2	1	TATAAAA	TFIIF-beta
2054	N	14	2	1	TATAAAA	TBP
2054	N	14	2	1	TATAAAA	TFIID
2054	N	14	2	1	TATAAAA	Dr1
2080	N	14	2	1	TAAAAAA	TBP
2088	N	14	1.27	0.78	GNNCACtCAAG	TTF-1
2110	N	16	1.60	0.80	GTtTGTTCTT	HNF-1A
2110	N	16	1.60	0.80	GTtTGTTCTT	HNF-1B
2110	N	16	1.60	0.80	GTtTGTTCTT	HNF-1C
2113	C	12	2	1	TGTTCT	GR alpha
2113	C	12	2	1	TGTTCT	GR beta
2113	C	12	2	1	TGTTCT	PR
2113	C	12	2	1	TGTTCT	PR A
2127	C	12	1.09	1	WNNNAANAWGG	YY1
2188	C	12	0.86	0.75	TTGGCNNNNNKCcR	CTF
2188	C	12	0.86	0.75	TTGGCNNNNNKCcR	CTF-1
2188	C	12	0.86	0.75	TTGGCNNNNNKCcR	CTF-2
2188	C	12	0.86	0.75	TTGGCNNNNNKCcR	NF-1
2199	N	12	2	1	CAGTTG	c-Myc
2201	C	12	1.20	0.75	RTTaCRTMAY	E4BP4
2236	C	12	2	1	CAACTG	c-Myc
2239	C	16	2	1	CTGAGTCA	NF-E2
2240	C	14	2	1	TGAGTCA	AP-1
2240	C	14	2	1	TGAGTCA	c-Fos
2240	C	14	2	1	TGAGTCA	c-Jun
2282	N	24	1.50	1	GGNNGAGGGAGARRRR	Pur factor
2283	C	16	1.60	0.80	GAGGAGGGAt	Sp1
2283	N	16	2	1	GAGGAGGG	C/EBPalpha
2333	C	16	1.60	0.80	GCGGGGGCgG	EGR2
2339	C	13	1.86	1	GCTGACR	NF-S
2374	C	20	1.67	0.83	AGCCCtGGGCTT	AP-2
2412	N	16	2	1	GGGTGTGG	TEF-2
2427	N	16	1.60	0.80	TcGGGGTGGT	Sp1
2445	C	14	2	1	ATGACGT	E4F1
2446	C	12	2	1	TGACGT	CRE-BP1
2446	C	12	2	1	TGACGT	CREB
2446	C	12	2	1	TGACGT	EivF
2446	C	12	2	1	TGACGT	TREB-1
2446	C	12	2	1	TGACGT	deltaCREB
2446	N	12	2	1	TGACGT	ATF
2465	C	13	1.63	1	YKRCGTCA	ATF
2465	C	13	1.63	1	YKRCGTCA	ATF-1
2465	C	13	1.63	1	YKRCGTCA	ATF-a
2465	C	13	1.63	1	YKRCGTCA	ATF-adelta
2467	N	12	2	1	ACGTCA	CRE-BP1
2467	N	12	2	1	ACGTCA	CREB
2467	N	12	2	1	ACGTCA	EivF
2467	N	12	2	1	ACGTCA	TREB-1
2467	N	12	2	1	ACGTCA	deltaCREB
2468	N	13	1.86	1	YGTCAGC	NF-S
2519	C	16	1.60	0.80	TGGAATTTCt	CD28RC
2519	C	16	1.60	0.80	TGGAATTTCt	NF III-c
2519	C	16	1.60	0.80	TGGAATTTCt	NF-kappaB
2519	N	12	1.20	0.75	gGGAMTNYCC	NF-kappaB(-like)
2519	N	12	1.20	0.75	gGGAMTNYCC	NF-kappaB1
2519	N	12	1.20	0.75	gGGAMTNYCC	NF-kappaB1 precursor
2520	C	16	1.60	0.80	GGAATTTCCc	EBP-1
2520	C	16	1.60	0.80	GGAATTTCCc	NF-kappaB2
2520	C	16	1.60	0.80	GGAATTTCCc	NF-kappaB2 precursor
2537	N	14	2	1	AGAAATG	HiNF-A
2556	N	12	1.20	1	MATNNNWAAT	N-Oct-3
2560	N	14	2	1	AGAAATG	HiNF-A

2590	N	12	0.80	0.75	YGgMNNNNNNGCCAA	CTF
2590	N	12	0.80	0.75	YGgMNNNNNNGCCAA	CTF-1
2590	N	12	0.80	0.75	YGgMNNNNNNGCCAA	CTF-2
2590	N	12	0.80	0.75	YGgMNNNNNNGCCAA	NF-1
2607	C	12	1	0.75	AaATNNNNAAAT	HiNF-A
2619	N	12	1.20	1	MATNNNWAAT	N-Oct-3
2649	C	16	2	1	ATTATTAA	AFP1
2652	C	12	1.33	1	ATTWNNATK	N-Oct-3
2654	N	12	2	1	TAAAAT	F2F
2688	C	13	1.30	0.81	rTGMAAYCYC	NF-GMa
2719	N	12	2	1	GGGCGG	ETF
2719	N	12	2	1	GGGCGG	Sp1
2731	N	18	1.64	0.82	TGAGGTCAGGg	ATF
2731	N	18	1.64	0.82	TGAGGTCAGGg	CREB
2756	N	12	0.80	0.75	YGGMNNNNNNGcCAA	CTF
2756	N	12	0.80	0.75	YGGMNNNNNNGcCAA	CTF-1
2756	N	12	0.80	0.75	YGGMNNNNNNGcCAA	CTF-2
2756	N	12	0.80	0.75	YGGMNNNNNNGcCAA	NF-1
2788	C	16	1.33	1	AAATNNNNAAAT	HiNF-A
2804	N	16	1.60	0.80	TGGGCgTGGT	Sp1
2816	N	12	2	1	CACATG	USF
2855	C	13	1.30	0.81	GGRNAkTCCC	NF-kappaB
2855	C	13	1.30	0.81	GGRNAkTCCC	NF-kappaB(-like)
2855	C	13	1.30	0.81	GGRNAkTCCC	NF-kappaB1
2855	C	13	1.30	0.81	GGRNAkTCCC	NF-kappaB1 precursor
2881	N	13	1.30	0.81	GRGRTTKCAy	NF-GMa
2926	N	12	2	1	TATAAA	TFIID
2926	N	12	2	1	TATAAA	TBP
2950	C	12	1.20	1	ATTWNNNATK	N-Oct-3
2964	N	12	1	0.75	AtTTNNNNATTT	HiNF-A
2965	C	12	1.20	1	ATTWNNNATK	N-Oct-3
2990	C	12	1	0.75	AAATNNNNAAAt	HiNF-A
2995	N	14	2	1	TATAAAA	TFIIA
2995	N	14	2	1	TATAAAA	TFIIA-alpha/beta precursor (major)
2995	N	14	2	1	TATAAAA	TFIIA-alpha/beta precursor (minor)
2995	N	14	2	1	TATAAAA	TFIIA-gamma
2995	N	14	2	1	TATAAAA	TFIIB
2995	N	14	2	1	TATAAAA	TFIIE
2995	N	14	2	1	TATAAAA	TFIIF
2995	N	14	2	1	TATAAAA	TFIIF-alpha
2995	N	14	2	1	TATAAAA	TFIIF-beta
2995	N	14	2	1	TATAAAA	TBP
2995	N	14	2	1	TATAAAA	TFIID
2995	N	14	2	1	TATAAAA	Dr1
2997	N	12	2	1	TAAAAT	F2F
3019	C	16	1.60	0.80	GGAATTTCcC	EBP-1
3019	C	16	1.60	0.80	GGAATTTCcC	NF-kappaB2
3019	C	16	1.60	0.80	GGAATTTCcC	NF-kappaB2 precursor
3086	N	13	1.86	1	TCAGRTA	NF-GMb
3109	C	12	1.20	1	ATTWNNNATK	N-Oct-3
3116	C	12	2	1	ATTTTA	F2F
3120	C	14	2	1	TATTTAT	ETF
3121	C	12	1.71	1	ATTWATK	N-Oct-3
3121	N	12	1	0.75	ATTTNNNNATtT	HiNF-A
3123	N	14	1.75	1	WTATYCAT	Pit-1
3268	C	12	2	1	ATTTTA	F2F
3282	C	16	1.60	0.80	TTGcGCCACT	C/EBPalpha
3282	C	16	1.60	0.80	TTGcGCCACT	C/EBPbeta
3296	N	12	2	1	CTATCT	GATA-1
3362	C	16	1.60	0.80	TTTTcCTCTT	NFAT-1
3362	C	16	1.60	0.80	TTTTcCTCTT	Pu box binding factor
3366	N	14	2	1	TCTCTTA	c-Myc
3423	N	12	1	0.75	AtTTNNNNATTT	HiNF-A
3423	N	18	1.64	0.82	AaTTTATCATT	GATA-1
3435	N	12	1.20	1	MATNNNWAAT	N-Oct-3
3437	N	14	2	1	TATAAAA	TFIIA

3437	N	14	2	1	TATAAAA	TFIIA-alpha/beta precursor (major)
3437	N	14	2	1	TATAAAA	TFIIA-alpha/beta precursor (minor)
3437	N	14	2	1	TATAAAA	TFIIA-gamma
3437	N	14	2	1	TATAAAA	TFIIB
3437	N	14	2	1	TATAAAA	TFIIE
3437	N	14	2	1	TATAAAA	TFIIF
3437	N	14	2	1	TATAAAA	TFIIF-alpha
3437	N	14	2	1	TATAAAA	TFIIF-beta
3437	N	14	2	1	TATAAAA	TBP
3437	N	14	2	1	TATAAAA	TFIID
3437	N	14	2	1	TATAAAA	Dr1
3439	N	12	2	1	TAAAAT	F2F
3480	C	12	1	0.75	AAAtNNNNAAAT	HiNF-A
3489	N	12	1.71	1	MATWAAT	N-Oct-3
3492	C	12	1	0.75	AAATNNNNAAAt	HiNF-A
3494	C	19	1.19	0.83	RWYCTaWWWWTAGMGY	RSRFC4
3523	N	12	1.20	1	MATNNNWAAT	N-Oct-3
3526	N	12	2	1	TATAAA	TFIID
3526	N	12	2	1	TATAAA	TBP
3526	N	14	2	1	TATAAAT	TMF
3543	N	12	2	1	TAAAAT	F2F
3595	C	12	2	1	CTTCCC	c-Ets-2
3601	N	14	2	1	TGACTCA	AP-1
3601	N	14	2	1	TGACTCA	c-Fos
3601	N	14	2	1	TGACTCA	c-Jun
3601	N	16	2	1	TGACTCAG	NF-E2
3660	C	12	2	1	TTTATA	TFIID
3660	C	12	2	1	TTTATA	TBP
3668	C	12	2	1	CAACTG	c-Myc
3731	C	12	2	1	CATGTG	USF
3744	N	13	1.30	0.81	GGGAMTNyCC	NF-kappaB
3744	N	13	1.30	0.81	GGGAMTNyCC	NF-kappaB1
3744	N	13	1.30	0.81	GGGAMTNyCC	NF-kappaB1 precursor
3744	N	16	1.60	0.80	GGGAATCcCC	NF-kappaB(-like)
3776	N	12	1	0.75	AtTTNNNNATTT	HiNF-A
3830	N	14	2	1	TCTCTTA	c-Myc
3840	C	16	1.60	0.80	AAAGAGAaAG	IRF-1
3851	C	16	1.60	0.80	TTGcGTGAGA	Arnt
3851	C	16	1.60	0.80	TTGcGTGAGA	Arnt (774 AA form)
3888	C	16	1.60	0.80	AaAGAGAAAG	IRF-1
3905	N	12	2	1	CACATG	USF
3913	N	12	2	1	AGAACA	GR alpha
3913	N	12	2	1	AGAACA	GR beta
3913	N	12	2	1	AGAACA	PR
3913	N	12	2	1	AGAACA	PR A
3939	N	16	1.60	0.80	gGGGAGTGGC	NF-kappaB
3939	N	16	1.60	0.80	gGGGAGTGGC	Sp1
3940	N	14	2	1	GGGAGTG	p300
3953	N	14	2	1	TGTGTCA	AP-1
3953	N	14	2	1	TGTGTCA	c-Jun
3968	C	12	1.33	1	MYYMGCCYM	Sp1
3982	N	12	1.71	1	TGASTMA	AP-1
4059	C	12	1.20	1	ATTWNNNATK	N-Oct-3
4097	N	12	1.50	1	SMGGAWGY	PEA3
4097	N	12	1.50	1	SMGGAWGY	c-Ets-1
4097	N	14	1.75	1	SAGGAAGY	TCF-2alpha
4098	N	14	2	1	AGGAAGT	E1A-F
4134	N	12	1.20	1	MATNNNWAAT	N-Oct-3
4135	C	14	1.75	1	ATGRATAW	Pit-1
4166	N	14	2	1	CCCCTCC	Sp1
4184	N	16	1.60	0.80	TCaTTTAAAT	Oct-2.1
4200	N	16	1.60	0.80	TCAATtAAAT	Oct-2.1
4204	N	12	2	1	TAAAAT	F2F
4242	C	12	1.20	1	ATTWNNNATK	N-Oct-3
4252	N	12	1	0.75	ATTTNNNNAtTT	HiNF-A
4314	C	14	2	1	TTTTTTA	TBP

4321	N	12	1.71	1	MATWAAT	N-Oct-3
4329	C	12	1.71	1	ATTWATK	N-Oct-3
4387	N	16	1.60	0.80	GGCAGGTGGc	E12
4387	N	16	1.60	0.80	GGCAGGTGGc	E47
4393	N	14	2	1	TGGGTCA	ER
4393	N	14	2	1	TGGGTCA	c-Fos
4393	N	14	2	1	TGGGTCA	c-Jun
4402	N	24	1.71	0.86	AGGTCAGGAGgTCA	COUP
4402	N	24	1.71	0.86	AGGTCAGGAGgTCA	RXR-alpha
4436	C	16	1.60	1	RTGMAAYCYC	NF-GMa
4454	N	14	2	1	TAAAAAA	TBP
4455	C	12	1	0.75	AAAtNNNNAAAT	HiNF-A
4461	N	12	2	1	TAAAAT	F2F
4466	N	14	2	1	TAAAAAA	TBP
4482	N	16	1.60	0.80	TGGGCgTGGT	Sp1
4492	N	12	2	1	GGCGGG	Sp1
4521	N	12	1.33	1	KRGGCKRRK	Sp1
4553	C	16	1.60	0.80	GAGGcGGAGC	Sp1
4556	C	14	1.17	0.78	RTGRMcYTWGCM	HNF-4
4671	C	12	1.71	1	ATTWATK	N-Oct-3
4675	C	12	2	1	ATTTTA	F2F
4688	N	16	1.60	0.80	cAAGTGTTTG	TGT3
4700	N	16	1.60	0.80	ATGAgTCAGA	AP-1
4700	N	16	1.60	0.80	ATGAgTCAGA	c-Fos
4700	N	16	1.60	0.80	ATGAgTCAGA	c-Jun
4742	C	12	2	1	GGAAAG	NF-1
4761	C	12	2	1	AAGGAA	c-Ets-2
4765	C	12	1	0.75	AAATNNNNAAaT	HiNF-A
4849	C	16	1.60	0.80	TCcCTACCCA	Sp1
4948	N	12	2	1	GATTTC	H4TF-1
5008	N	16	2	1	TATAGAAA	TBP
5014	C	12	1	0.75	AAAtNNNNAAAT	HiNF-A
5018	N	14	2	1	TATAAAA	TFIIA
5018	N	14	2	1	TATAAAA	TFIIA-alpha/beta precursor (major)
5018	N	14	2	1	TATAAAA	TFIIA-alpha/beta precursor (minor)
5018	N	14	2	1	TATAAAA	TFIIA-gamma
5018	N	14	2	1	TATAAAA	TFIIB
5018	N	14	2	1	TATAAAA	TFIIE
5018	N	14	2	1	TATAAAA	TFIIF
5018	N	14	2	1	TATAAAA	TFIIF-alpha
5018	N	14	2	1	TATAAAA	TFIIF-beta
5018	N	14	2	1	TATAAAA	TBP
5018	N	14	2	1	TATAAAA	TFIID
5018	N	14	2	1	TATAAAA	Dr1
5020	N	12	2	1	TAAAAT	F2F
5050	N	22	1.69	0.85	GGTTCAgAGTTCA	RAR-gamma
5052	N	14	1.17	0.78	KgCWARGKYCAY	HNF-4
5052	N	14	2	1	TTCAAAG	SRY
5052	N	14	2	1	TTCAAAG	TCF-1A
5128	C	14	2	1	TCATGAG	Oct-01
5128	C	14	2	1	TCATGAG	Oct-02
5128	C	14	2	1	TCATGAG	Oct-2.1
5128	C	14	2	1	TCATGAG	Oct-2B
5128	C	14	2	1	TCATGAG	Oct-2C
5169	N	12	1.09	1	CCWTNTTNNNW	YY1
5171	N	12	1	0.75	ATtTNNNNATTT	HiNF-A
5179	C	12	1.33	1	ATTWNNATK	N-Oct-3
5192	C	16	1.60	0.80	TGCTaTCTGA	GATA-1
5199	C	14	2	1	TGAGTTA	AP-1
5205	C	12	1.33	1	ATTWNNATK	N-Oct-3
5221	N	12	1.09	1	CCWTNTTNNNW	YY1
5245	N	14	1.17	0.78	KgCWARGKYCAY	HNF-4
5281	N	16	1.60	0.80	gTTTGTTCTT	HNF-1A
5281	N	16	1.60	0.80	gTTTGTTCTT	HNF-1B
5281	N	16	1.60	0.80	gTTTGTTCTT	HNF-1C
5284	C	12	2	1	TGTTCT	GR alpha

5284	C	12	2	1	TGTTCT	GR beta
5284	C	12	2	1	TGTTCT	PR
5284	C	12	2	1	TGTTCT	PR A
5288	N	12	2	1	CTTTCC	NF-1
5295	N	16	1.60	0.80	GGaATTTCCT	NF-kappaB(-like)
5300	N	12	2	1	TTCCTT	c-Ets-2
5306	N	12	1.50	1	SMGGAWGY	c-Ets-1
5306	N	16	2	1	CAGGATGT	PEA3
5339	C	14	2	1	TTTTTTA	TBP
5367	C	13	1.86	1	TAYCTGA	NF-GMb
5369	C	16	1.60	0.80	TCTGACTcAT	AP-1
5369	C	16	1.60	0.80	TCTGACTcAT	c-Fos
5369	C	16	1.60	0.80	TCTGACTcAT	c-Jun
5376	N	16	2	1	TATAGAAA	TBP
5379	N	14	2	1	AGAAATG	HiNF-A
5383	C	16	1.60	0.80	ATGCAAAGCa	Oct-01
5383	C	16	1.60	0.80	ATGCAAAGCa	octamer-binding factor
5383	C	16	2	1	ATGCAAAG	oct-B3
5402	C	12	0.80	0.75	TtGGCNNNNNNKCCR	CTF
5402	C	12	0.80	0.75	TtGGCNNNNNNKCCR	CTF-1
5402	C	12	0.80	0.75	TtGGCNNNNNNKCCR	CTF-2
5402	C	12	0.80	0.75	TtGGCNNNNNNKCCR	NF-1
5451	C	14	2	1	TGACACA	AP-1
5451	C	14	2	1	TGACACA	c-Jun
5491	C	12	2	1	CCAATC	CCAAT-binding factor
5491	C	12	2	1	CCAATC	CP1
5491	C	12	2	1	CCAATC	CTF
5491	N	14	2	1	CCAATCA	AP-1
5493	N	12	1.33	1	MATNNWAAT	N-Oct-3
5516	N	12	2	1	GGGAAG	c-Ets-2
5536	N	12	2	1	ATTGAA	c-Myb
5567	N	12	2	1	CTTTCC	NF-1
5573	N	12	2	1	TAAAAT	F2F
5576	N	12	1.71	1	MATWAAT	N-Oct-3
5592	C	12	2	1	TGTTCT	GR alpha
5592	C	12	2	1	TGTTCT	GR beta
5592	C	12	2	1	TGTTCT	PR
5592	C	12	2	1	TGTTCT	PR A
5613	C	12	1	0.75	AAATNNNNaAAT	HiNF-A
5615	N	12	1.20	0.75	RTKaYGTAAY	E4BP4
5626	N	14	2	1	TAAAAAA	TBP
5711	N	12	2	1	TAAAAT	F2F
5790	C	13	1.63	1	CWWWCCAC	C/EBPalpha
5790	N	12	2	1	CTTTCC	NF-1
5847	C	14	2	1	TGTGTTT	HiNF-A
5859	C	16	1.60	0.80	ATGATGTcAC	AP-1
5859	N	12	1.20	0.75	RTKAYGTaAY	E4BP4
5872	N	16	2	1	TGAGACTT	Sp1
5872	N	16	2	1	TGAGACTT	AP-1
5891	C	16	1.60	0.80	CCTTAtTTGG	SRF
5902	C	12	2	1	TTCAAT	c-Myb
5917	N	12	2	1	AGAACA	GR alpha
5917	N	12	2	1	AGAACA	GR beta
5917	N	12	2	1	AGAACA	PR
5917	N	12	2	1	AGAACA	PR A
5960	N	16	1.60	0.80	AAGTGAAAGt	PRDI-BF1
5983	C	12	2	1	TGTTCT	GR alpha
5983	C	12	2	1	TGTTCT	GR beta
5983	C	12	2	1	TGTTCT	PR
5983	C	12	2	1	TGTTCT	PR A
6035	C	16	1.60	0.80	CCAGCTgTGG	AP-4
6103	C	12	2	1	CTTCCC	c-Ets-2
6105	C	12	2	1	TCCCCA	AP-2
6124	N	14	1.17	0.78	KGcWARGKYCAY	HNF-4
6152	N	12	1	0.75	aTTTNNNNATTT	HiNF-A
6169	N	14	2	1	TCTCTTA	c-Myc

6180	N	12	1.09	1	CCWTNTTNNNW	YY1
6237	N	12	2	1	TGGGGA	AP-2
6256	C	16	2	1	TTATAGCC	GATA-1
6299	N	16	1.60	0.80	GAAcCACCCA	Sp1
6311	N	12	1.20	0.75	GRgRTTKCAY	NF-GMa
6330	C	14	2	1	ACTTCCT	E1A-F
6332	N	12	2	1	TTCCTT	c-Ets-2
6358	N	12	1.20	1	MATNNNWAAT	N-Oct-3
6372	N	14	2	1	CTGCCCC	NF-E2
6451	N	16	1.60	0.80	tCATTTAAAT	Oct-2.1
6452	N	12	1.33	1	MATNNWAAT	N-Oct-3
6472	C	12	2	1	GAAATC	H4TF-1
6510	C	12	2	1	TTCAAT	c-Myb
6513	N	12	1.71	1	MATWAAT	N-Oct-3
6528	C	12	2	1	TTCAAT	c-Myb
6582	C	14	2	1	AACCAAT	CP1A
6582	C	14	2	1	AACCAAT	CP1B
6613	C	12	2	1	CATTAG	Isl-1
6623	N	12	2	1	CACATG	USF
6631	N	16	1.60	0.80	ATGaGTCAGA	AP-1
6631	N	16	1.60	0.80	ATGaGTCAGA	c-Fos
6631	N	16	1.60	0.80	ATGaGTCAGA	c-Jun
6654	C	12	0.80	0.75	TTGgCNNNNNNKCCR	CTF
6654	C	12	0.80	0.75	TTGgCNNNNNNKCCR	CTF-1
6654	C	12	0.80	0.75	TTGgCNNNNNNKCCR	CTF-2
6654	C	12	0.80	0.75	TTGgCNNNNNNKCCR	NF-1
6685	C	12	2	1	ATTTTA	F2F
6729	C	12	2	1	ATTTTA	F2F
6761	N	12	1	0.75	ATTtNNNNATTT	HiNF-A
6779	N	12	2	1	GATTTC	H4TF-1
6783	N	16	1.60	0.80	TCTCAcGCAA	Arnt
6783	N	16	1.60	0.80	TCTCAcGCAA	Arnt (774 AA form)
6796	N	13	1.63	1	GTGGWWWG	C/EBPalpha
6809	N	12	1	0.75	aTTTNNNNATTT	HiNF-A
6817	C	12	1.71	1	ATTWATK	N-Oct-3
6853	N	16	1.60	0.80	CcACAGCTGG	AP-4
6855	N	14	2	1	ACAGCTG	GT-IIBalpha
6855	N	14	2	1	ACAGCTG	GT-IIBbeta
6903	N	16	1.60	0.80	CCTtTGACCT	Tf-LF1
6911	C	12	2	1	CTTCCC	c-Ets-2
6922	C	16	1.60	0.80	TTGCgCCACT	C/EBPalpha
6922	C	16	1.60	0.80	TTGCgCCACT	C/EBPbeta
6938	N	12	2	1	TAAAAT	F2F
6989	N	12	1.33	1	MATNNWAAT	N-Oct-3
6992	N	12	2	1	TAAAAT	F2F
7009	N	18	2	1	GGGGGAGGG	H4TF-1
7009	N	20	2	1	GGGGGAGGGG	Sp1
7011	N	14	2	1	GGGAGGG	MAZ
7016	N	12	2	1	GGGAAG	c-Ets-2
7025	C	12	2	1	TTATCA	GATA-1
7039	C	16	2	1	TTAGTCAG	CRE-BP1
7039	C	16	2	1	TTAGTCAG	AP-1
7039	C	16	2	1	TTAGTCAG	c-Jun
7066	N	14	1.75	1	WTATYCAT	Pit-1
7073	C	12	1.71	1	TKASTCA	AP-1
7103	C	12	2	1	TGTACA	GR alpha
7103	C	12	2	1	TGTACA	GR beta
7103	C	12	2	1	TGTACA	PR
7103	C	12	2	1	TGTACA	PR A
7110	N	12	2	1	CAGGTC	TGT3
7134	N	14	2	1	TATAAAA	TFIIA
7134	N	14	2	1	TATAAAA	TFIIA-alpha/beta precursor (major)
7134	N	14	2	1	TATAAAA	TFIIA-alpha/beta precursor (minor)
7134	N	14	2	1	TATAAAA	TFIIA-gamma
7134	N	14	2	1	TATAAAA	TFIIB
7134	N	14	2	1	TATAAAA	TFIIE

7134	N	14	2	1	TATAAAA	TFIIF
7134	N	14	2	1	TATAAAA	TFIIF-alpha
7134	N	14	2	1	TATAAAA	TFIIF-beta
7134	N	14	2	1	TATAAAA	TBP
7134	N	14	2	1	TATAAAA	TFIID
7134	N	14	2	1	TATAAAA	Dr1
7184	N	16	2	1	ATATAGGA	GR alpha
7184	N	16	2	1	ATATAGGA	GR beta
7197	C	14	2	1	TAAGAGA	c-Myc
7203	C	12	1.33	1	ATTWNNATK	N-Oct-3
7206	N	12	1.20	0.75	RTKAYGtAAY	E4BP4
7227	C	12	2	1	TGTACA	GR alpha
7227	C	12	2	1	TGTACA	GR beta
7227	C	12	2	1	TGTACA	PR
7227	C	12	2	1	TGTACA	PR A
7277	C	14	0.93	0.78	NTTCNNtTCNNTTCN	HSF
7277	C	14	0.93	0.78	NTTCNNtTCNNTTCN	HSF1 (long)
7277	C	14	0.93	0.78	NTTCNNtTCNNTTCN	HSF1 (short)
7280	N	12	2	1	CAGGTC	TGT3
7286	N	12	2	1	CTTTCC	NF-1
7320	C	16	1.60	0.80	GTCCCGCCcA	NF-kappaB
7320	C	16	1.60	0.80	GTCCCGCCcA	Sp1
7343	N	12	2	1	GGGAAG	c-Ets-2
7360	N	12	2	1	AGAACA	GR alpha
7360	N	12	2	1	AGAACA	GR beta
7360	N	12	2	1	AGAACA	PR
7360	N	12	2	1	AGAACA	PR A
7377	N	13	1.30	0.81	GRGRTTKCAy	NF-GMa
7379	N	12	2	1	GATTTC	H4TF-1
7403	N	12	1.09	1	CCWTNTTNNNW	YY1
7404	N	16	1.60	0.80	CTTTcTCTTT	IRF-1
7407	C	14	0.93	0.78	NTTCNNTTcNNTTCN	HSF
7407	C	14	0.93	0.78	NTTCNNTTcNNTTCN	HSF1 (long)
7407	C	14	0.93	0.78	NTTCNNTTcNNTTCN	HSF1 (short)
7450	N	13	1.63	1	GTGGWWWG	C/EBPalpha
7474	N	16	1.60	0.80	cCTTTGACCT	Tf-LF1
7503	C	18	2	1	CTCGTCTCA	AP-1
7503	C	18	2	1	CTCGTCTCA	Sp1
7513	N	12	2	1	CTTTCC	NF-1
7561	C	20	2	1	TGTTTTACAT	NF-IL-2A
7561	C	20	2	1	TGTTTTACAT	Oct-01
7561	C	20	2	1	TGTTTTACAT	Oct-2.1
7584	N	12	1.20	0.75	GRGRTtKCAY	NF-GMa
7642	C	12	2	1	ACGTCA	ATF
7642	N	12	2	1	ACGTCA	CRE-BP1
7642	N	12	2	1	ACGTCA	CREB
7642	N	12	2	1	ACGTCA	EivF
7642	N	12	2	1	ACGTCA	TREB-1
7642	N	12	2	1	ACGTCA	deltaCREB
7651	C	12	2	1	TCCCCA	AP-2
7695	N	12	1.71	1	TGASTMA	AP-1
7700	N	12	1.33	1	MATNNWAAT	N-Oct-3
7715	C	12	2	1	ATTTTA	F2F
7727	C	16	1.60	0.80	cACTTCCTGC	Elk-1
7727	C	16	1.60	0.80	cACTTCCTGC	Ets-1 deltaVII
7727	C	16	1.60	0.80	cACTTCCTGC	PEA3
7727	C	16	1.60	0.80	cACTTCCTGC	c-Ets-1
7727	C	16	1.60	0.80	cACTTCCTGC	c-Ets-2
7727	C	16	1.60	0.80	cACTTCCTGC	p38erg
7727	C	16	1.60	0.80	cACTTCCTGC	p49erg
7727	C	16	1.60	0.80	cACTTCCTGC	p55erg
7728	C	14	1.75	1	RCTTCCTS	TCF-2alpha
7728	C	14	2	1	ACTTCCT	E1A-F
7735	C	16	1.60	0.80	GCGCCgCCCA	Sp1
7814	C	14	2	1	TGTGTTT	HiNF-A
7818	C	12	2	1	TTTATA	TFIID

7818	C	12	2	1	TTTATA	TBP
7827	N	14	1.75	1	WTATYCAT	Pit-1
7834	N	12	2	1	TTCCTT	c-Ets-2
7836	N	12	1.09	1	CCWTNTTNNNW	YY1
7853	N	16	1.60	0.80	CTTTCaGTTT	ISGF-1
7853	N	16	1.60	0.80	CTTTCTcTTT	IRF-1
7881	C	12	0.86	0.75	TTGGCNNNNNKCcR	CTF
7881	C	12	0.86	0.75	TTGGCNNNNNKCcR	CTF-1
7881	C	12	0.86	0.75	TTGGCNNNNNKCcR	CTF-2
7881	C	12	0.86	0.75	TTGGCNNNNNKCcR	NF-1
7903	N	16	1.60	0.80	GCTtCAGTTT	ISGF-3
8031	N	16	1.60	0.80	GCGcGGGCCG	GCF
8077	N	14	2	1	TAACTCA	AP-1
8081	C	16	1.60	0.80	aCACAGCCCC	Sp1
8088	C	16	1.60	0.80	CaCTTCCTGC	Elk-1
8088	C	16	1.60	0.80	CaCTTCCTGC	Ets-1 deltaVII
8088	C	16	1.60	0.80	CaCTTCCTGC	PEA3
8088	C	16	1.60	0.80	CaCTTCCTGC	c-Ets-1
8088	C	16	1.60	0.80	CaCTTCCTGC	c-Ets-2
8088	C	16	1.60	0.80	CaCTTCCTGC	p38erg
8088	C	16	1.60	0.80	CaCTTCCTGC	p49erg
8088	C	16	1.60	0.80	CaCTTCCTGC	p55erg
8151	N	14	2	1	TGAATCA	AP-1
8156	C	14	2	1	CATTTCT	HiNF-A
8163	C	12	2	1	TTCAAT	c-Myb
8171	N	16	1.07	1	YGGMNNNNNNGCCAA	CTF
8171	N	16	1.07	1	YGGMNNNNNNGCCAA	CTF-1
8171	N	16	1.07	1	YGGMNNNNNNGCCAA	CTF-2
8171	N	16	1.07	1	YGGMNNNNNNGCCAA	NF-1
8193	C	14	1.75	1	GAGTANGA	C/EBPalpha
8203	C	16	1.60	0.80	GGAGgGCCGC	Sp1
8213	N	30	1.76	0.88	CGgGGCGGGGCGGGGCG	ETF
8214	N	16	1.60	0.80	GgGGCGGGGC	MAZ
8214	N	16	1.60	0.80	GgGGCGGGGC	Sp1
8214	N	16	1.60	0.80	GgGGCGGGGC	TFIID
8214	N	20	2	1	GAGGCGGGGC	octamer-binding factor
8215	N	14	1.40	1	NNGCGGGGCN	GCF
8215	N	20	1.25	0.83	gGGCGGNNNNGGGCGG	LSF
8224	N	16	1.60	0.80	GaGGCGGGGC	octamer-binding factor
8224	N	20	2	1	GGGGCGGGGC	MAZ
8224	N	20	2	1	GGGGCGGGGC	Sp1
8224	N	20	2	1	GGGGCGGGGC	TFIID
8225	N	14	1.40	1	NNGCGGGGCN	GCF
8252	N	16	1.60	0.80	TGGGCGTGGt	Sp1
8253	N	16	1.60	0.80	GGGCGgGGCG	GCF
8254	N	16	2	1	GGCGTGGC	NF-kappaB
8258	N	12	0.80	0.75	YGGMNNNNNNGCCaA	CTF
8258	N	12	0.80	0.75	YGGMNNNNNNGCCaA	CTF-1
8258	N	12	0.80	0.75	YGGMNNNNNNGCCaA	CTF-2
8258	N	12	0.80	0.75	YGGMNNNNNNGCCaA	NF-1
8318	N	12	2	1	CAGGTC	TGT3
8345	N	12	2	1	TTCCTT	c-Ets-2
8390	C	16	1.60	0.80	CAgGGTGGGG	CACCC-binding factor
8414	C	13	1.30	1	RYWTCCKKYK	Elk-1
8414	C	14	2	1	ACTTCCT	E1A-F
8416	N	12	2	1	TTCCTT	c-Ets-2
8428	N	16	1.60	0.80	GGGAAATTCc	EBP-1
8428	N	16	1.60	0.80	GGGAAATTCc	NF-kappaB2
8428	N	16	1.60	0.80	GGGAAATTCc	NF-kappaB2 precursor
8431	C	16	1.33	1	AAATNNNNAAAT	HiNF-A
8434	N	16	2	1	TTCATAAA	TFIID
8434	N	16	2	1	TTCATAAA	TBP
8436	N	12	1.71	1	MATWAAT	N-Oct-3
8437	N	14	2	1	ATAAATA	ETF
8459	C	12	1.20	1	ATTWNNNATK	N-Oct-3
8497	N	16	1.60	0.80	GGGGCTGTgT	Sp1

8506	N	14	2	1	TTCAAAG	SRY
8506	N	14	2	1	TTCAAAG	TCF-1A
8550	N	12	1.09	1	CCWTNTTNNNW	YY1
8555	C	16	1.07	1	TTGGCNNNNNNKCCR	CTF
8555	C	16	1.07	1	TTGGCNNNNNNKCCR	CTF-1
8555	C	16	1.07	1	TTGGCNNNNNNKCCR	CTF-2
8555	C	16	1.07	1	TTGGCNNNNNNKCCR	NF-1
8610	N	12	2	1	TAAAAT	F2F
8623	C	16	1.60	0.80	CCATAaAAGG	SRF
8648	C	16	1.45	0.80	GTCAYGTGSCc	USF
8650	C	12	2	1	CACGTG	c-Myc
8690	C	14	2	1	TGATTCA	AP-1
8749	C	14	1.17	0.78	RTGRMcYTWGCM	HNF-4
8777	N	16	2	1	GGGTGTGG	TEF-2
8796	N	12	2	1	TGGGGA	AP-2
8799	C	12	2	1	GGAAAG	NF-1
8844	C	13	1.86	1	GCTGACR	NF-S
8846	C	12	2	1	TGACGT	CRE-BP1
8846	C	12	2	1	TGACGT	CREB
8846	C	12	2	1	TGACGT	EivF
8846	C	12	2	1	TGACGT	TREB-1
8846	C	12	2	1	TGACGT	deltaCREB
8846	N	12	2	1	TGACGT	ATF
8864	N	12	1	0.75	ATTTNNNNATtT	HiNF-A

APPENDIX 3 - BLAST alignments

3.1 Rat and human maspin promoters

Only the matches which were considered to be conserved are shown (positioned within 500 bp). Rat maspin promoter from Accession number AC094626. Human maspin promoter from Accession number AC036176. Base number 1 is the predicted transcription start sites.

```
Human maspin promoter (User Entered)
- and -
Rat maspin promoter (User Entered)


Query:      >Human maspin promoter
            Length = 8171
Reference: Query= Human maspin promoter
          (8171 letters)

>Rat maspin promoter
           Length = 4771

 Score = 73.8 bits (37), Expect = 2e-15
 Identities = 172/214 (80%), Gaps = 6/214 (2%)
 Strand = Plus / Plus

Query: 4133 aaatatttaataacctaggagactgttttgttggtgctttttcttacctataggttgaag 4192
            |||||||||||||||||| |||||    ||||||| |||||||| || || |||| ||
Sbjct: 3334 aaatatttaataacctaggtgactg----gttggtggtttttcttgcccatgggtttaaa 3389

Query: 4193 ctttgcaaatattat-acttgattactcagcatcgaggctgagggatgatttgacacacc 4251
            | || |||  ||||| ||| |||||| | ||||    |||||||||||| | | ||||||
Sbjct: 3390 cattacaagcattatgactggattacccggcattctagctgagggatgactgggcacacc 3449

Query: 4252 cagccactcccaggaacaaatgaaaggagctttgttctagcatgtgcctctcgctttctc 4311
            |||||||||||||| | ||| | ||||||||||||||| || |||||  || |    |||||
Sbjct: 3450 cagccactcccagcagcaactcaaaggagctttgttataacatgt-tctttgatattctc 3508

Query: 4312 tctctctctctctctctgtctcgctctcgctctc 4345
            |||||||||||||||| |||| ||||| |||||
Sbjct: 3509 tctctctctctctctctctctctctctctctctc 3542


 Score = 34.2 bits (17), Expect = 0.002
 Identities = 35/41 (85%)
 Strand = Plus / Plus

Query: 14  ctcttggcagcttgtccacaattgaaagaaatgattcaggc 54
            |||||||||||| ||| ||  |||| |||| ||||||||||
Sbjct: 107 ctcttggcagcctgtgcaagattggaagacatgattcaggc 147


 Score = 30.2 bits (15), Expect = 0.031
 Identities = 15/15 (100%)
 Strand = Plus / Plus

Query: 2426 tgattctatttgact 2440
            |||||||||||||||
Sbjct: 1934 tgattctatttgact 1948


 Score = 30.2 bits (15), Expect = 0.031
 Identities = 15/15 (100%)
 Strand = Plus / Plus

Query: 67  tgtctgcacagcttc 81
            |||||||||||||||
Sbjct: 167 tgtctgcacagcttc 181


 Score = 30.2 bits (15), Expect = 0.031
 Identities = 21/23 (91%)
 Strand = Plus / Plus

Query: 1259 attttaggcatcagtggagcaag 1281
            |||| |||||||||||| |||||
Sbjct: 930  atttcaggcatcagtggggcaag 952


 Score = 26.3 bits (13), Expect = 0.48
 Identities = 13/13 (100%)
 Strand = Plus / Plus

Query: 2769 agaataagaccta 2781
            |||||||||||||
Sbjct: 2274 agaataagaccta 2286
```

```
 Score = 24.3 bits (12), Expect = 1.9
 Identities = 12/12 (100%)
 Strand = Plus / Plus

Query: 1265 ggcatcagtgga 1276
            ||||||||||||
Sbjct: 1667 ggcatcagtgga 1678


 Score = 24.3 bits (12), Expect = 1.9
 Identities = 12/12 (100%)
 Strand = Plus / Plus

Query: 102 cgggcaggaagg 113
           ||||||||||||
Sbjct: 201 cgggcaggaagg 212


 Score = 24.3 bits (12), Expect = 1.9
 Identities = 12/12 (100%)
 Strand = Plus / Plus

Query: 2118 attggacatcag 2129
            ||||||||||||
Sbjct: 2798 attggacatcag 2809


 Score = 22.3 bits (11), Expect = 7.5
 Identities = 11/11 (100%)
 Strand = Plus / Plus

Query: 1953 ttccctctccc 1963
            ||||||||||
Sbjct: 1508 ttccctctccc 1518


 Score = 22.3 bits (11), Expect = 7.5
 Identities = 11/11 (100%)
 Strand = Plus / Minus

Query: 1572 agcatgtgctt 1582
            ||||||||||
Sbjct: 1106 agcatgtgctt 1096


Score = 22.3 bits (11), Expect = 7.5
 Identities = 11/11 (100%)
 Strand = Plus / Plus

Query: 1673 aattcaccaca 1683
            ||||||||||
Sbjct: 1248 aattcaccaca 1258
```

```
Lambda     K      H
    1.37    0.711     1.31

Gapped
Lambda     K      H
    1.37    0.711     1.31


Matrix: blastn matrix:1 -3
Gap Penalties: Existence: 5, Extension: 2
Number of Hits to DB: 73
Number of Sequences: 0
Number of extensions: 73
Number of successful extensions: 73
Number of sequences better than 10.0: 1
Number of HSP's better than 10.0 without gapping: 1
Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 0
Number of HSP's gapped (non-prelim): 53
length of query: 4771
length of database: 8171
effective HSP length: 13
effective length of query: 4758
effective length of database: 8158
effective search space: 38815764
effective search space used: 38815764
T: 0
A: 0
X1: 6 (11.9 bits)
X2: 15 (29.7 bits)
S1: 12 (24.3 bits)
S2: 11 (22.3 bits)
```

3.2 Mouse and human maspin promoters

Only the matches which were considered to be conserved are shown (positioned within 500 bp). Mouse maspin promoter from Accession number NW_000170. Human maspin promoter from Accession number AC036176. Base number 1 is the predicted transcription start sites.

Human maspin promoter (User Entered)
- and -
Mouse maspin pro reverse (User Entered)

```
Query:      >Human maspin promoter
            Length = 8171
Reference: Query= Human maspin promoter
          (8171 letters)

>Mouse maspin pro
           Length = 3718

 Score = 48.1 bits (24), Expect = 1e-07
 Identities = 107/133 (80%), Gaps = 5/133 (3%)
 Strand = Plus / Plus

Query: 4133 aaatatttaataacctaggagactgttttgttggtgctttttcttacctataggttgaag 4192
            |||||||||||||||||| |||||    ||||||| |||||||| ||||| |||| ||
Sbjct: 3528 aaatatttaataacctaggtgactg----gttggtggtttttcttgcctatgggtttaaa 3583

Query: 4193 ctttgcaaatattat-acttgattactcagcatcgaggctgagggatgatttgacacacc 4251
            | || |||  ||||| ||| |||||| | ||||   || | |||||||| | ||||||||
Sbjct: 3584 cattacaagcattatgactggattacccggcattctggataagggatgactagacacacc 3643

Query: 4252 cagccactcccag 4264
            ||||||||||||
Sbjct: 3644 tagccactcccag 3656
```

```
 Score = 36.2 bits (18), Expect = 4e-04
 Identities = 42/50 (84%)
 Strand = Plus / Plus

Query: 5    tactcaagcctcttggcagcttgtccacaattgaaagaaatgattcaggc 54
             |||||| | |||||||||||| ||| ||  |||| |||| ||||||||||
Sbjct: 122 tactcatgactcttggcagcctgtgcaagattggaagacatgattcaggc 171


 Score = 30.2 bits (15), Expect = 0.024
 Identities = 15/15 (100%)
 Strand = Plus / Plus

Query: 2426 tgattctatttgact 2440
            |||||||||||||||
Sbjct: 2081 tgattctatttgact 2095


 Score = 24.3 bits (12), Expect = 1.5
 Identities = 12/12 (100%)
 Strand = Plus / Plus

Query: 428 actaccaaattg 439
           ||||||||||||
Sbjct: 473 actaccaaattg 484


 Score = 24.3 bits (12), Expect = 1.5
 Identities = 12/12 (100%)
 Strand = Plus / Plus

Query: 1918 ttagaggcacca 1929
           ||||||||||||
Sbjct: 1569 ttagaggcacca 1580


 Score = 24.3 bits (12), Expect = 1.5
 Identities = 12/12 (100%)
 Strand = Plus / Plus

Query: 1949 catgttccctct 1960
           ||||||||||||
Sbjct: 1598 catgttccctct 1609


 Score = 24.3 bits (12), Expect = 1.5
 Identities = 12/12 (100%)
 Strand = Plus / Plus

Query: 102 cgggcaggaagg 113
           ||||||||||||
Sbjct: 221 cgggcaggaagg 232
```

```
 Score = 22.3 bits (11), Expect = 5.9
 Identities = 11/11 (100%)
 Strand = Plus / Plus

Query: 2245 tgaagtatctc 2255
           ||||||||||| 
Sbjct: 1944 tgaagtatctc 1954


 Score = 22.3 bits (11), Expect = 5.9
 Identities = 11/11 (100%)
 Strand = Plus / Plus

Query: 2119 ttggacatcag 2129
           ||||||||||| 
Sbjct: 1818 ttggacatcag 1828


Score = 22.3 bits (11), Expect = 5.9
 Identities = 11/11 (100%)
 Strand = Plus / Plus

Query: 1606 ctttgcccatt 1616
           ||||||||||| 
Sbjct: 1488 ctttgcccatt 1498
```

```
Lambda     K      H
    1.37    0.711     1.31

Gapped
Lambda     K      H
    1.37    0.711     1.31


Matrix: blastn matrix:1 -3
Gap Penalties: Existence: 5, Extension: 2
Number of Hits to DB: 51
Number of Sequences: 0
Number of extensions: 51
Number of successful extensions: 51
Number of sequences better than 10.0: 1
Number of HSP's better than 10.0 without gapping: 1
Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 0
Number of HSP's gapped (non-prelim): 51
length of query: 3718
length of database: 8171
effective HSP length: 12
effective length of query: 3706
effective length of database: 8159
effective search space: 30237254
effective search space used: 30237254
T: 0
A: 0
X1: 6 (11.9 bits)
X2: 15 (29.7 bits)
S1: 12 (24.3 bits)
S2: 11 (22.3 bits)
```

APPENDIX 4 - Primers

All primer sequences are written 5' – 3'.

4.1 Cloning primers

Maspin Pro F	GCGCGAGCTCCATCCTCTCCCACTTTTCTG
Maspin Pro R	GCGCCTCGAGTCACAGTTATCCTGGAAAATGC

4.2 Sequencing primers

-2172_seq	ATCATCCCAGCTCTGGAC
Maspin 401_prime	GAGTACTTTTTGTGCCACCAACGTG
-1823_seq	CGGTTGAGCAAGCACTG
-1769_seq_rev	AGAGCCTCTCTATTCTCC
-1357_seq	TTTCCCTCCTACAGCTGG
-1320_seq_rev	GAGGTGTGGAAGGATGC
-950_seq	GCACAGCAGAGAAGCAAC
-903_seq_rev	TCCCTTCCAGCCGAATC
-583_seq	TCCTGGGCTCAAGCAATC
-551_seq_rev	TAATCCCAGCATTTGGGG
Maspin 2284_prime	GGAGACCATTTAAGCCTGAGAAC
-187_seq_rev	CCGGAAGGTGAAATATTCG
RVprimer3 (Promega)	CTAGCAAAATAGGCTGTCCC
GLprimer2 (Promega)	CTTTATGTTTTTGGCGTCTTCCA

4.3 Footprinting primers

11mer	GAATTCAGATC
25mer	GCGGTGACCCGGGAGATCTGAATTC

4.3.1 -2551 bp to -2251 bp region

Foot1	CATCCTCTCCCACTTTTC (5′ biotin)
Foot2	TCCTCTCCCACTTTTCTG
Foot3	CTCCCACTTTTCTGTTTTC
FootR1	ACATTTCCTGTCAAAAGC (5′ biotin)
FootR2	CCTGTCAAAAGCTTTCAC
FootR3	CAAAAGCTTTCACTTGTTG

4.3.2 -1078 bp to -963 bp region

-1078_fwd1	GCACTTATTGACTGCATACTG (5′ biotin)
-1078_fwd2	GCATACTGTACATCAGGTCTG
-1078_fwd3	GGTCTGTACCAAGCTTTGG
-1078_rev1	AGCCTAGTGTGGTGATGC (5′ biotin)
-1078_rev2	TGGTGATGCATGCCTG
-1078_rev3	GCCTGTGGTCCCAGTTAC

4.3.3 -660 bp to -295 bp region

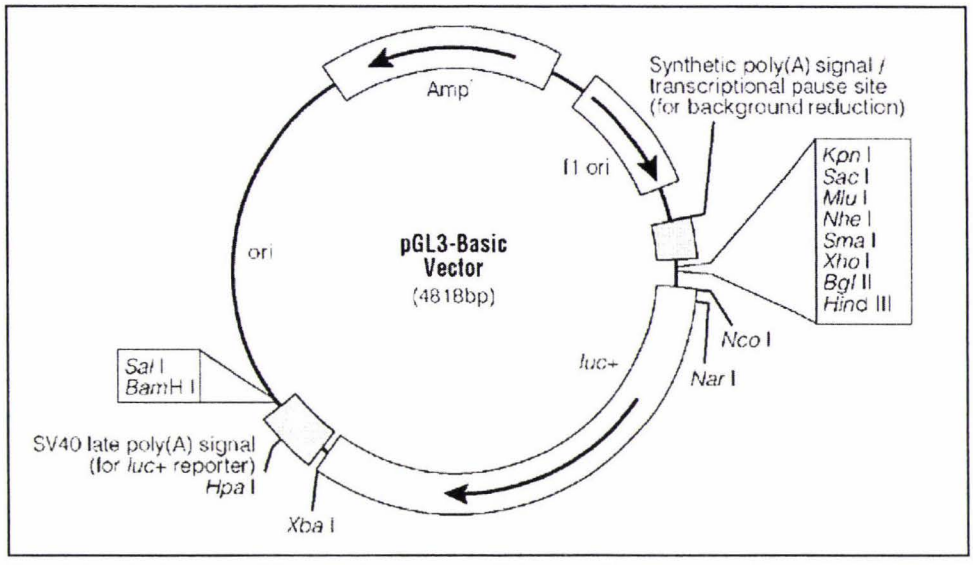
-660_fwd1	GTAACTGGGACCACAGGC (5' biotin)
-660_fwd2	ACAGGCATGCATCACCAC
-660_fwd3	ATGCATCACCACACTAGGC
-660_rev1	CGTTGGTGGCACAAAAAG (5' biotin)
-660_rev2	GGTGGCACAAAAAGTACTCTG
-660_rev3	GCACAAAAAGTACTCTGATCTCC

4.3.4 -134 bp to TSS region

-134_fwd1	CATCGAATATTTCACCTTCC (5' biotin)
-134_fwd2	ATTTCACCTTCCGGTCC
-134_fwd3	AGAGGATTGCCGTACGC
-134_rev1	GAGGAGCACAAAGACCTG (5' biotin)
-134_rev2	GCACAAAGACCTGGATGTG
-134_rev3	ATGTGGAGGCGACCG

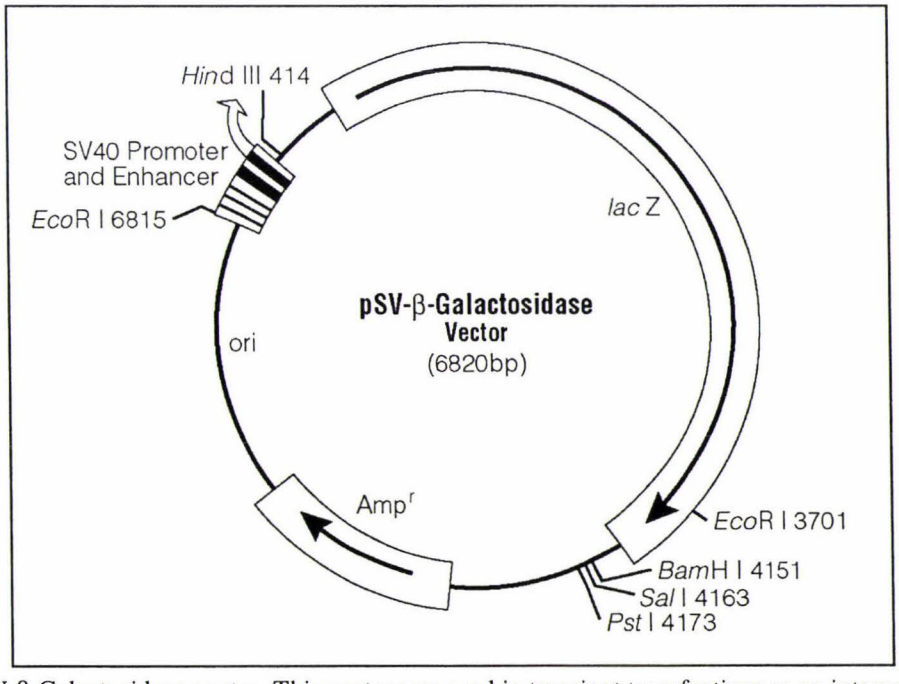
APPENDIX 5 - Vector maps

5.1 pGL3 Basic



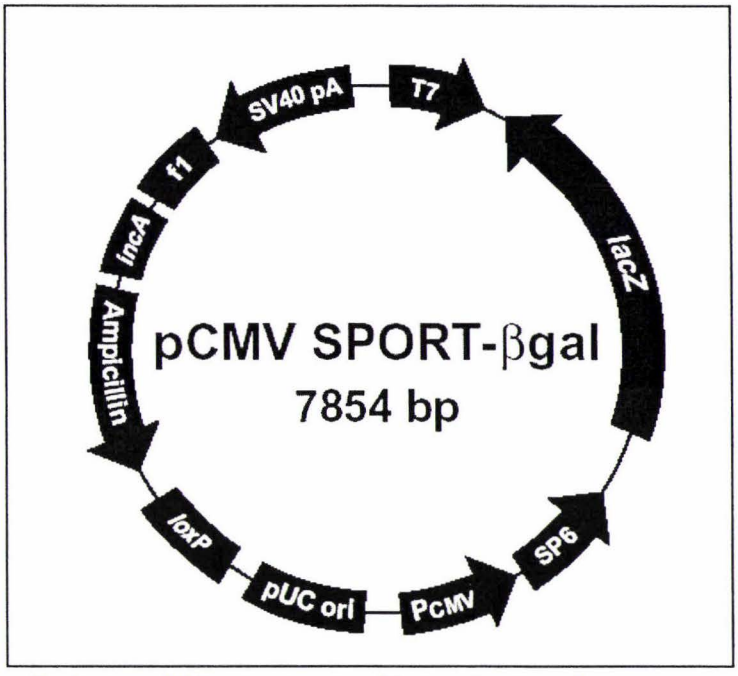
The pGL3 Basic vector (Promega). Vector used for maspin promoter functional assays. From Promega catalogue

5.2 pSV-β-Galactosidase



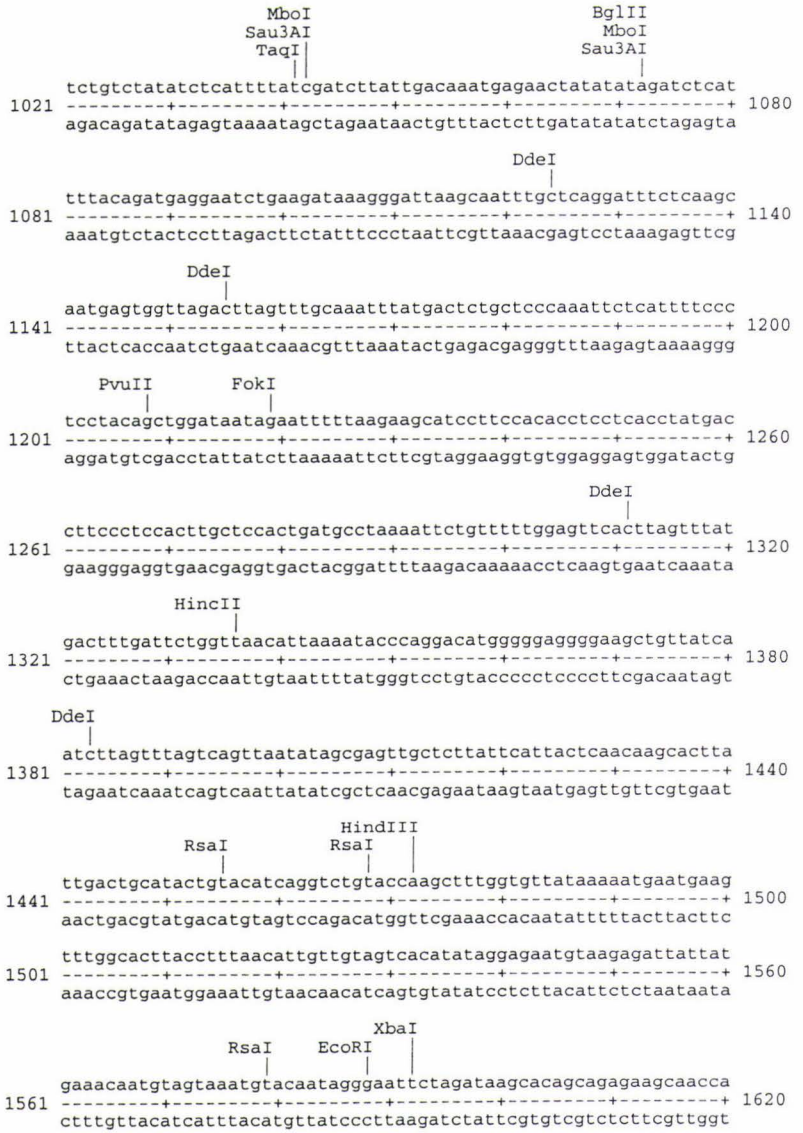
pSV-β-Galactosidase vector. This vector was used in transient transfections as an internal control for transfection efficiency. From Promega catalogue

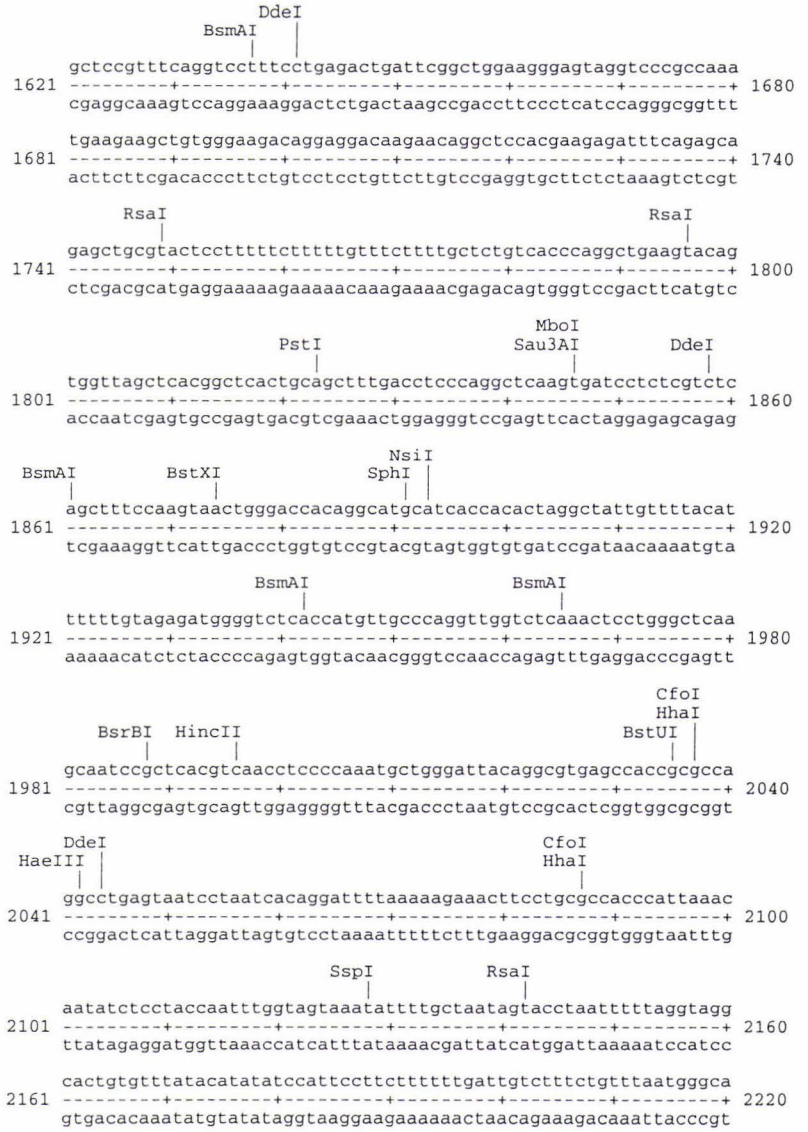
5.3 pCMV-βgal

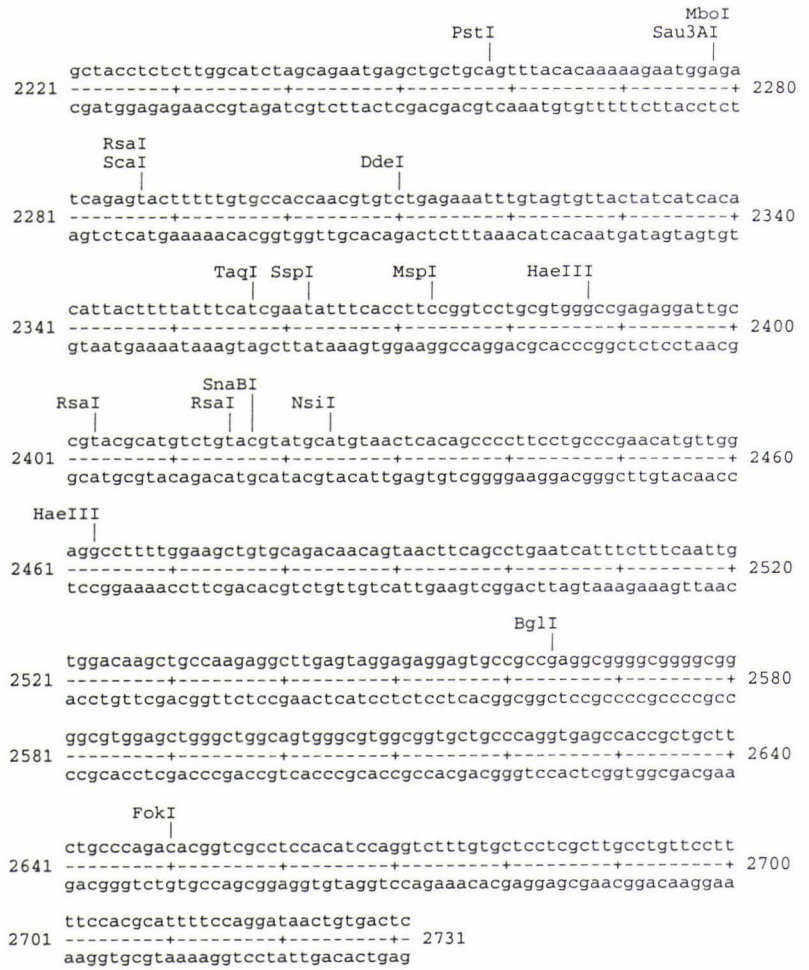


pCMV-Sport-βgal vector. This vector was used in transient transfections as an internal control for transfection efficiency. From Invitrogen catalogue.

APPENDIX 6 - Restriction map of 2.7 kb maspin promoter fragment


MboI
Sau3AI
TaqI
BglII
MboI
Sau3AI
DdeI
DdeI
PvuII
FokI
DdeI
HincII
DdeI
HindIII
RsaI
RsaI
XbaI
RsaI
EcoRI


DdeI
BsmAI
RsaI
RsaI
MboI
PstI
Sau3AI
DdeI
NsiI
BsmAI
BstXI
SphI
BsmAI
BsmAI
CfoI
HhaI
BsrBI
HincII
BstUI
DdeI
HaeIII
CfoI
HhaI
SspI
RsaI



Enzymes that do cut:

AccI	AlwNI	AvaI	BglI	BglII	BsmAI	BsrBI	BstUI
BstXI	CfoI	DdeI	EcoRI	FokI	HaeIII	HhaI	HincII
HindIII	MboI	MspI	NsiI	PstI	PvuII	RsaI	Sau3AI
ScaI	SnaBI	SphI	SspI	StyI	TaqI	XbaI	

Enzymes that do not cut:

ApaI	BalI	BamHI	BlnI	BssHII	BstBI	Bsu36I	EagI
EciI	EclXI	EcoRV	HgaI	KpnI	MluI	MscI	NarI
NcoI	NgoMI	NheI	NotI	PvuI	SacI	SalI	SmaI
SstI	XhoI						

APPENDIX 7 - Sequencing data

Sequencing was carried out with sequencing primers listed in Appendix 4 and maspin constructs, or PCR products representing the cloned DNA. An example chromatogram is shown over page. The edited sequence obtained from this sequence is shown below. Edited bases are in lower case, and bases whose identity could not be determined from chromatogram were termed 'n'.

-1478 bp construct sequenced with RV3 primer.

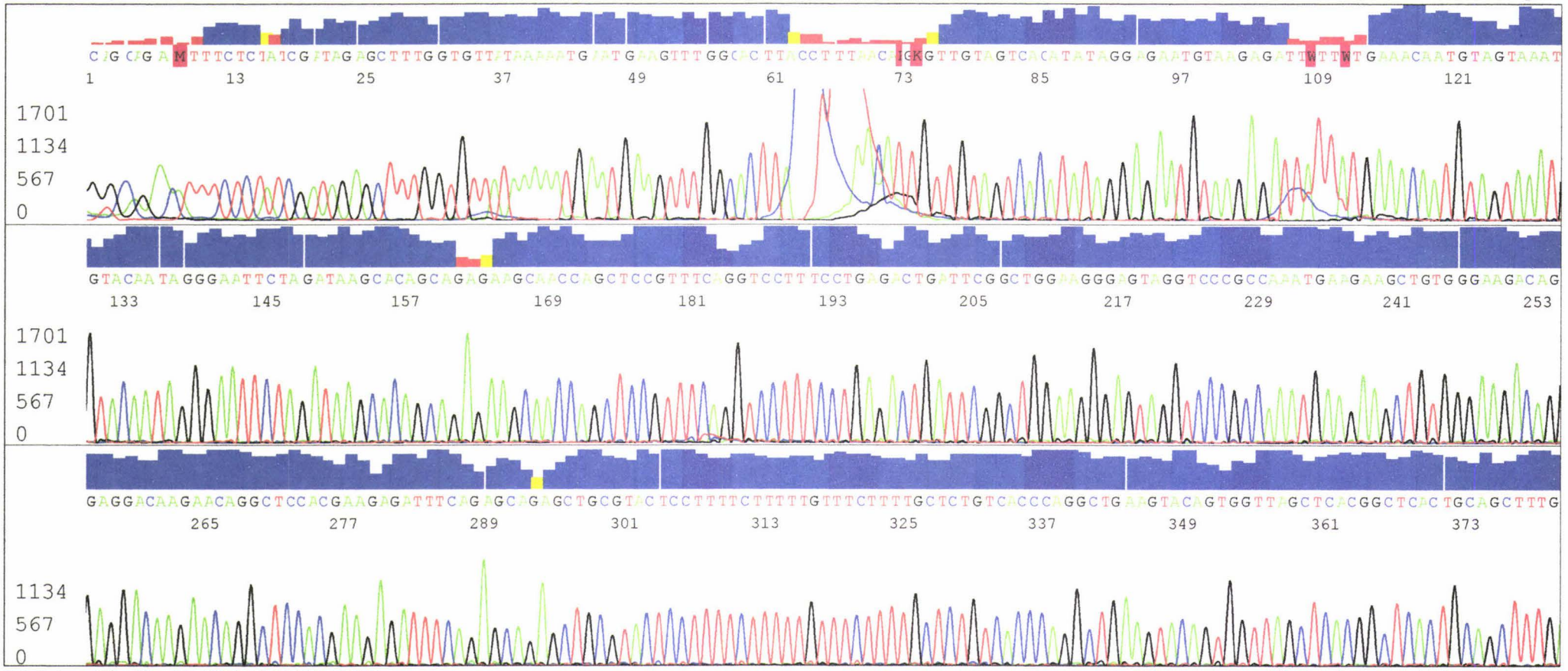
CAGCAGAnTTTCTCTATCGATAGAGCTTTGGTGTTATAAAAATGAATGAAGTTTGGCACT
TAnnnnnAACAttGTTGTAGTCACATATAGGAGAATGTAAGAGATTaTTaTGAAACAATG
TAGTAAATGTACAATAGGGAATTCTAGATAAGCACAGCAGAGAAGCAACCAGCTCCGTTT
CAGGTCCTTTCCTGAGACTGATTCGGCTGGAAGGGAGTAGGTCCCGCCAAATGAAGAAGC
TGTGGGAAGACAGGAGGACAAGAACAGGCTCCACGAAGAGATTTCAGAGCAGAGCTGCGT
ACTCCTTTTCTTTTTGTTTCTTTTGCTCTGTCACCCAGGCTGAAGTACAGTGGTTAGCTC
ACGGCTCACTGCAGCTTTGACCTCCCAGGCTCAAGTGATCCTCTCGTCTCAGCTTTCCAA
GTAACTGGGACCACAGGCATGCATCACCACGCTAGGCTATTGTTTTACATTTTTTGTAGA
GATGGGGTCTCACCATGTTGCCCAGGTTGGTCTCAAACTCCTGGGCTCAAGCAATCCGCT
CACGTCAACCTCCCCAAATGCTGGGATTACAGGCGTGAGCCACCACGCCAGGCCTGAGTA
ATCCTAATCACAGGATTTTAAAAAGAAACTTCCTGCGCCACCCATTAAACAATATCTCCT
ACCAATTTGGTAGTAAATATTTTGCTAATAGTACCTAATTTTTAGGTAGGCACTGTGTTT
ATACATATATCCATTCCTTCTTTTTTGATTGTCTTTCTGTTTAATGGGCAGCTACCTCTC
TTGGCATCTAGCAGAATGAGCTGCTGCAGTTTACACAAAAAGAATGGAGATCAGAGTACT
TTTTGTGCCACCAACGTGTCTGAGAAATTTGTAGTGTTACTATCATCACACATTACTTTT
ATTTCATCGAATATTTCACCTTCCGGTCCTGCGTGGGCCGAGAGGATTGCCGTACGCATG
TCTGTACGTATGCATGTAACTCACAGCCCCTTCCTGCCCGAACATGTTGGAGGCCTTTTG
GAAGCTGTGCAGACACAGCAACTTCAGCTGAATCATCTCTTTCAATTGTGGACAAGCTGC
CAAGAGCTGAGTAGAAGAGTGCGCGAGCGGGCGAGACGGGGCGTGAAGCTGGCTGGCAAT
TGT

Applied Biosystems

S/N G:91 A:82 T:71 C:88
KB.bcp
KB 1.0 Cap:3

1.7kb.RV3
andyhollings.6.1.7kb.RV3
KB_3730_POP7_BDTv3.mob
Pts 2200 to 18760 Pk1 Loc:2007
Version 5.1 HiSQV Bases: 960

Inst Model/Name 3730/MickeyMouse-1412-023
Aug 19,2003 10:56AM, NZST
Aug 19,2003 11:30AM, NZST
Spacing:16.36 Pts/Panel1500
Plate Name: Unassigned

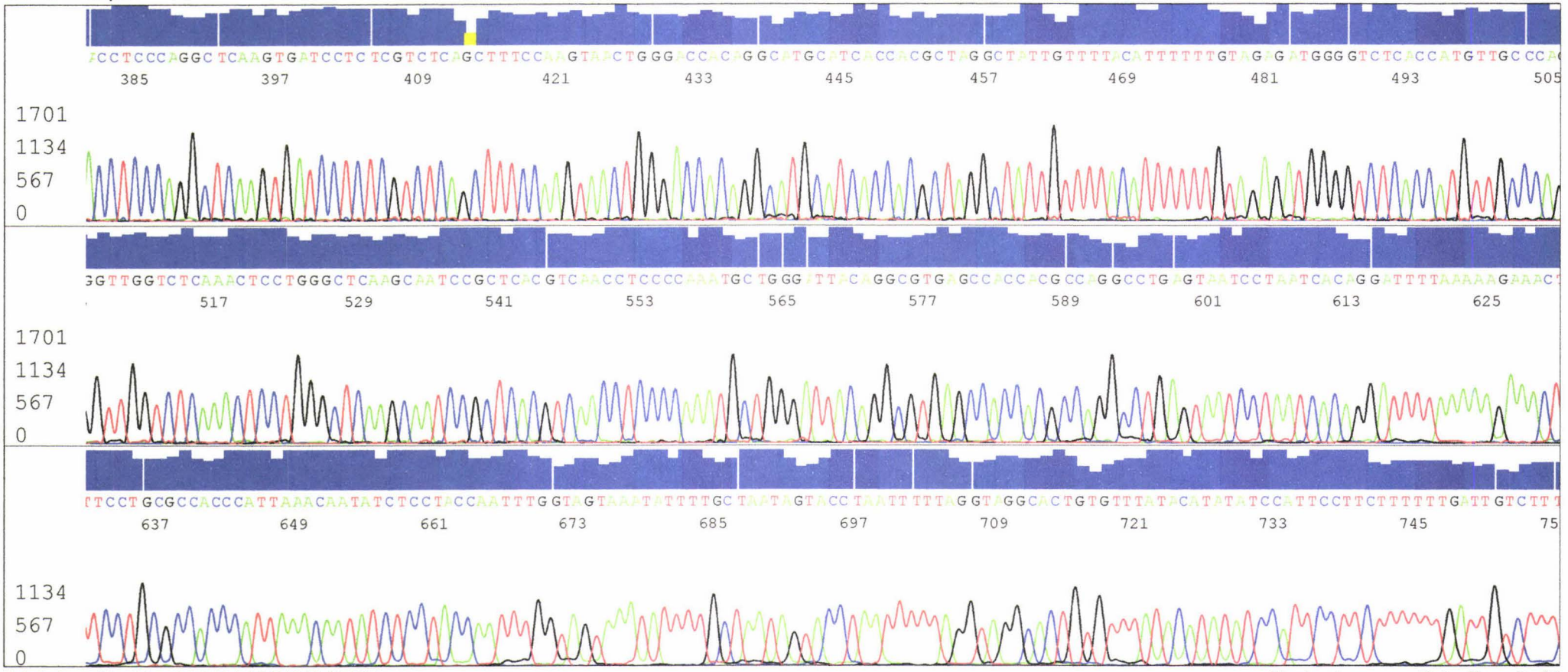


Applied Biosystems

S/N G:91 A:82 T:71 C:88
KB.bcp
KB 1.0 Cap:3

1.7kb.RV3
andyhollings.6.1.7kb.RV3
KB_3730_POP7_BDTv3.mob
Pts 2200 to 18760 Pk1 Loc:2007
Version 5.1 HiSQV Bases: 960

Inst Model/Name 3730/MickeyMouse-1412-023
Aug 19,2003 10:56AM, NZST
Aug 19,2003 11:30AM, NZST
Spacing:16.36 Pts/Panel1500
Plate Name: Unassigned



APPENDIX 8 - Alignment of cloned and human genome derived maspin promoters

GCG sequence alignment of cloned maspin (capitals) against the human genome maspin promoter sequence (lower case). The cloned promoter was sequenced in both directions and edited sequences were merged together (using Gelmerge function of GCG) to gain a contig. The contig was then aligned with the human genome derived maspin promoter sequence (from accession number AC036176) using the Bestfit function of GCG (GCG, 1997).

Mismatches are present at base numbers (human genome sequence) 2, 569, 655, 1103, 1901, 2036, 2491, and 2510. Deleted bases are at base numbers 162 and 1755. Transcription start site is located at base number 2551, thus the promoter region is from base number 1 to 2551, while bases 2552 to 2731 are the first exon.

```
BESTFIT of: 2.7kb_maspin_sequenced  check: 9500  from: 1  to: 3414
to: Maspin_2.7kb_promoter (from accession number AC036176)  check:
9322  from: 1  to: 2731

        Gap Weight:     50      Average Match: 10.000
     Length Weight:      3   Average Mismatch: -9.000

           Quality:  27032             Length:   2731
             Ratio:  9.905               Gaps:      2
 Percent Similarity: 99.707   Percent Identity: 99.707

       Match display thresholds for the alignment(s):
                  | = IDENTITY
                  : =    5
                  . =    1

 2.7kb_maspin_sequenced x Maspin_2.7kb_promoter (from accession
number AC036176) November 3, 19103 16:21  ..


   384 CCCCATCCTCTCCCACTTTTCTGTTTTCTAGTGTCTTTTTAGGTAGAATT 433
       | ||||||||||||||||||||||||||||||||||||||||||||||||
     1 ctccatcctctcccacttttctgttttctagtgtcttttaggtagaatt 50

   434 CTAAAGAACTTAAAATTTTGGCTGCCGTGGAAAAGAATATTACTGGGAAA 483
       ||||||:|||||||||||||||||||||||||||||||||||||||||||
    51 ctaaagaacttaaaatttggctgccgtggaaaagaatattactgggaaa 100

   484 TGGGTCTTCTCAGTCAAATAGAATCACAGTTTTAATACTCTTTCCACTGA 533
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   101 tgggtcttctcagtcaaatagaatcacagttttaatactctttccactga 150

   534 ATGGGAACCAC.TTTTTTTTTCCTGCAATTTTGCCTAAATCAACAGTGTG 582
       |||||||||||| ||||||||||||||||||||||||||||||||||||
   151 atgggaaccacttttttttttcctgcaatttgcctaaatcaacagtgtg 200

   583 TTTTGTCAATGATGTTACACCTGAGACTTGTGAGTCCTTGCCTTACTTGG 632
       |||||||||||||||||||||||||||||||.|||||||||||||||||
   201 ttttgtcaatgatgttacacctgagacttgtgagtccttgccttacttgg 250

   633 GTTCAATGAGAGGGTGAGAACATTCACGTACAGGGTCTGACCTCCAGAGA 682
       ||||||||||||||||||||||||||||||||||:|||||||||||||||
   251 gttcaatgagagggtgagaacattcacgtacagggtctgacctccagaga 300
```

```
 683 TACTTCAACAAGTGAAAGCTTTTGACAGGAAATGTTCTGTTTAATCTTTT 732
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 tacttcaacaagtgaaagcttttgacaggaaatgttctgtttaatctttt 350

 733 GTTAACATATCTGTCCAATATACCTTGTATCATCCCAGCTCTGGACCTCC 782
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 351 gttaacatatctgtccaatataccttgtatcatcccagctctggacctcc 400

 783 GTTCTCAGGCTTAAATGGTCTCCTGATGTCCAATGATGCTGCCTCCATGA 832
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 gttctcaggcttaaatggtctcctgatgtccaatgatgctgcctccatga 450

 833 TCCTTCCCCAGTCATAGCCAGACTGAAAAGGTCACACCCATCTCTGAACT 882
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 451 tccttccccagtcatagccagactgaaaaggtcacacccatctctgaact 500

 883 ACTTTTGTCATTTGCACCTCTCTTAAGTCCCTTATTACATGCCTCCAGAT 932
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 501 acttttgtcatttgcacctctcttaagtcccttattacatgcctccagat 550

 933 ACTTTCTTTTGTTTCTTATCTTCCTACTGTACCCTTTGGGGAGAGGGAAC 982
     |||||||||||||||| ||||||||||||||||||||||||||||||||
 551 actttcttttgtttcttgtcttcctactgtacccttggggagagggaac 600

 983 ATGTCTTATAGCCCCAGTCGAATGGTGCCTCTAATTTTCACCCAGTTAGA 1032
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 601 atgtcttatagccccagtcgaatggtgcctctaattttcacccagttaga 650

1033 ATCATCCAAAGGCATTTCACTTTGCTAGAACTTCCTTGGGTATTGTTGAA 1082
     |||| ||||||||||||||||||||||||||||||||||||||||||||
 651 atcacccaaaggcatttcactttgctagaacttccttgggtattgttgaa 700

1083 TTCCTGTCATTGGAAATGTTTCTGCCCCGGTTGAGCAAGCACTGAACAGG 1132
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 701 ttcctgtcattggaaatgtttctgccccggttgagcaagcactgaacagg 750

1133 AGTGTTTTAGAGGGAATTAAAACTTCAAAAAGGAGAATAGAGAGGCTCTA 1182
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 751 agtgttttagagggaattaaaacttcaaaaaggagaatagagaggctcta 800

1183 CCATTTAAATGCAAGTCGCTTGAAATCTGTAGACCACAGTAGCAAGCACG 1232
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 801 ccatttaaatgcaagtcgcttgaaatctgtagaccacagtagcaagcacg 850

1233 AAGCAGGTATTCAATAAATGTGGTGAATTCAATGTCAATAACATGCTCCA 1282
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 851 aagcaggtattcaataaatgtggtgaattcaatgtcaataacatgctcca 900

1283 GTCTTTTGTCCACTATTCCCCGAGCCTACCTAACCAATGGGCAAAGTGGT 1332
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 901 gtcttttgtccactattccccgagcctacctaaccaatgggcaaagtggt 950

1333 ACTAATTGCCATCATTAGGAAGCACATGCTATGTGTCAGAAAGTATGTAA 1382
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 951 actaattgccatcattaggaagcacatgctatgtgtcagaaagtatgtaa 1000

1383 GAATTGTCTCTTCATCCATCTCTGTCTATATCTCATTTTATCGATCTTAT 1432
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 gaattgtctcttcatccatctctgtctatatctcattttatcgatcttat 1050

1433 TGACAAATGAGAACTATATATAGATCTCATTTTACAGATGAGGAATCTGA 1482
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 tgacaaatgagaactatatatagatctcattttacagatgaggaatctga 1100

1483 AGCTAAAGGGATTAAGCAATTTGCTCAGGATTTCTCAAGCAATGAGTGGT 1532
     || ||||||||||||||||||||||||||||||||||||||||||||||||
1101 agataaagggattaagcaatttgctcaggatttctcaagcaatgagtggt 1150

1533 TAGACTTAGTTTGCAAATTTATGACTCTGCTCCCAAATTCTCATTTTCCC 1582
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1151 tagacttagtttgcaaatttatgactctgctcccaaattctcattttccc 1200

1583 TCCTACAGCTGGATAATAGAATTTTTAAGAAGCATCCTTCCACACCTCCT 1632
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 tcctacagctggataatagaatttttaagaagcatccttccacacctcct 1250

1633 CACCTATGACCTTCCCTCCACTTGCTCCACTGATGCCTAAAATTCTGTTT 1682
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1251 cacctatgaccttccctccacttgctccactgatgcctaaaattctgttt 1300

1683 TTGGAGTTCACTTAGTTTATGACTTTGATTCTGGTTAACATTAAAATACC 1732
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1301 ttggagttcacttagtttatgactttgattctggttaacattaaaatacc 1350

1733 CAGGACATGGGGGAGGGGAAGCTGTTATCAATCTTAGTTTAGTCAGTTAA 1782
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1351 caggacatgggggaggggaagctgttatcaatcttagtttagtcagttaa 1400
```

```
1783 TATAGCGAGTTGCTCTTATTCATTACTCAACAAGCACTTATTGACTGCAT 1832
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1401 tatagcgagttgctcttattcattactcaacaagcacttattgactgcat 1450

1833 ACTGTACATCAGGTCTGTACCAAGCTTTGGTGTTATAAAAATGAATGAAG 1882
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1451 actgtacatcaggtctgtaccaagctttggtgttataaaaatgaatgaag 1500

1883 TTTGGCACTTACCTTTAACATTGTTGTAGTCACATATAGGAGAATGTAAG 1932
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1501 tttggcacttacctttaacattgttgtagtcacatataggagaatgtaag 1550

1933 AGATTATTATGAAACAATGTAGTAAATGTACAATAGGGAATTCTAGATAA 1982
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1551 agattattatgaaacaatgtagtaaatgtacaatagggaattctagataa 1600

1983 GCACAGCAGAGAAGCAACCAGCTCCGTTTCAGGTCCTTTCCTGAGACTGA 2032
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1601 gcacagcagagaagcaaccagctccgtttcaggtcctttcctgagactga 1650

2033 TTCGGCTGGAAGGGAGTAGGTCCCGCCAAATGAAGAAGCTGTGGGAAGAC 2082
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1651 ttcggctggaagggagtaggtcccgccaaatgaagaagctgtgggaagac 1700

2083 AGGAGGACAAGAACAGGCTCCACGAAGAGATTTCAGAGCAGAGCTGCGTA 2132
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1701 aggaggacaagaacaggctccacgaagagatttcagagcagagctgcgta 1750

2133 CTCC.TTTTCTTTTTGTTTCTTTTGCTCTGTCACCCAGGCTGAAGTACAG 2181
     |||| ||||||||||||||||||||||||||||||||||||||||||||
1751 ctcctttttcttttttgtttctttttgctctgtcacccaggctgaagtacag 1800

2182 TGGTTAGCTCACGGCTCACTGCAGCTTTGACCTCCCAGGCTCAAGTGATC 2231
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1801 tggttagctcacggctcactgcagctttgacctcccaggctcaagtgatc 1850

2232 CTCTCGTCTCAGCTTTCCAAGTAACTGGGACCACAGGCATGCATCACCAC 2281
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1851 ctctcgtctcagctttccaagtaactgggaccacaggcatgcatcaccac 1900

2282 GCTAGGCTATTGTTTTACATTTTTTGTAGAGATGGGGTCTCACCATGTTG 2331
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1901 actaggctattgttttacatttttgtagagatggggtctcaccatgttg 1950
```

```
2332 CCCAGGTTGGTCTCAAACTCCTGGGCTCAAGCAATCCGCTCACGTCAACC 2381
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1951 cccaggttggtctcaaactcctgggctcaagcaatccgctcacgtcaacc 2000

2382 TCCCCAAATGCTGGGATTACAGGCGTGAGCCACCACGCCAGGCCTGAGTA 2431
     |||||||||||||||||||||||||||||||| ||||||||||||||||
2001 tccccaaatgctgggattacaggcgtgagccaccgcgccaggcctgagta 2050

2432 ATCCTAATCACAGGATTTTAAAAAGAAACTTCCTGCGCCACCCATTAAAC 2481
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2051 atcctaatcacaggattttaaaaagaaacttcctgcgccacccattaaac 2100

2482 AATATCTCCTACCAATTTGGTAGTAAATATTTTGCTAATAGTACCTAATT 2531
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2101 aatatctcctaccaatttggtagtaaatattttgctaatagtacctaatt 2150

2532 TTTAGGTAGGCACTGTGTTTATACATATATCCATTCCTTCTTTTTTGATT 2581
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2151 tttaggtaggcactgtgtttatacatatatccattccttctttttgatt 2200

2582 GTCTTTCTGTTTAATGGGCAGCTACCTCTCTTGGCATCTAGCAGAATGAG 2631
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2201 gtctttctgtttaatgggcagctacctctcttggcatctagcagaatgag 2250

2632 CTGCTGCAGTTTACACAAAAAGAATGGAGATCAGAGTACTTTTTGTGCCA 2631
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2251 ctgctgcagtttacacaaaaagaatggagatcagagtacttttgtgcca 2300

2632 CCAACGTGTCTGAGAAATTTGTAGTGTTACTATCATCACACATTACTTTT 2731
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2301 ccaacgtgtctgagaaatttgtagtgttactatcatcacacattacttttt 2350

2732 ATTTCATCGAATATTTCACCTTCCGGTCCTGCGTGGGCCGAGAGGATTGC 2781
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2351 atttcatcgaatatttcaccttccggtcctgcgtgggccgagaggattgc 2400

2782 CGTACGCATGTCTGTACGTATGCATGTAACTCACAGCCCCTTCCTGCCCG 2831
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2401 cgtacgcatgtctgtacgtatgcatgtaactcacagcccccttcctgcccg 2450

2832 AACATGTTGGAGGCCTTTTGGAAGCTGTGCAGACAACAGCAACTTCAGCC 2881
     |||||||||||||||||||||||||||||||||||| ||||||||||
2451 aacatgttggaggccttttggaagctgtgcagacaacagtaacttcagcc 2500
```

```
                .          .          .          .          .
2882 TGAATCATCTCTTTCAATTGTGGACAAGCTGCCAAGAGGCTTGAGTAGGA 2951
     |||||||| ||||||||||||||||||||||||||||||||||||||||||
2501 tgaatcatttctttcaattgtggacaagctgccaagaggcttgagtagga 2550
                .          .          .          .          .
2952 GAGGAGTGCCGCCGAGGCGGGGCGGGGCGGGGCGTGGAGCTGGGCTGGCA 2981
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2551 gaggagtgccgccgaggcggggcggggcggggcgtggagctgggctggca 2600
                .          .          .          .          .
2982 GTGGGCGTGGCGGTGCTGCCCAGGTGAGCCACCGCTGCTTCTGCCCAGAC 3031
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2601 gtgggcgtggcggtgctgcccaggtgagccaccgctgcttctgcccagac 2650
                .          .          .          .          .
3032 ACGGTCGCCTCCACATCCAGGTCTTTGTGCTCCTCGCTTGCCTGTTCCTT 3081
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2651 acggtcgcctccacatccaggtctttgtgctcctcgcttgcctgttcctt 2700
                .          .          .
3082 TTCCACGCATTTTCCAGGATAACTGTGACTC 3112
     |||||||||||||||||||||||||||||||
2701 ttccacgcattttccaggataactgtgactc 2731
```

APPENDIX 9 - Transfection example data

To obtain normalised values, the maximum level of luciferase activity, and β-gal activity were measured, and the blank readings subtracted from the raw values. The level of luminescence was then divided by the β-gal to gain normalised values for each well. Triplicates were then averaged and the average deviation between the well calculated. The values were then divided by the -2551 bp expression level, thus the normalised readings are expressed relative to the expression of the full length construct. Example data shows results from transfection of five constructs in PC3 cells. Blank indicates reagent blank wells.

Construct	Max luminescence	Luminescence - blank	Beta-gal reading	Beta-gal raw - blank	Normalised luciferase	Average activity	Average deviation	Relative activity	Rel. avg. deviation
Blank	112	0	0.053	0.000	#DIV/0!				
-2551 bp a	3827	3715	0.707	0.654	5680.4	4983.7	464.5	1.00	0.09
-2551 bp b	2509	2397	0.574	0.521	4600.8				
-2551 bp c	1980	1868	0.453	0.400	4670.0				
-963 bp a	2981	2869	0.625	0.572	5015.7	4789.3	189.4	0.96	0.04
-963 bp b	2306	2194	0.540	0.487	4505.1				
-963 bp c	2613	2501	0.569	0.516	4846.9				
-660 bp a	2552	2440	0.638	0.585	4170.9	3850.6	213.6	0.77	0.04
-660 bp b	2145	2033	0.602	0.549	3703.1				
-660 bp c	2212	2100	0.624	0.571	3677.8				
-295 bp a	7827	7715	0.689	0.636	12130.5	13311.5	787.3	2.67	0.16
-295 bp b	9346	9234	0.691	0.638	14473.4				
-295 bp c	7937	7825	0.640	0.587	13330.5				
-263 bp b	8382	8270	0.752	0.699	11831.2	10777.8	885.9	2.16	0.18
-263 bp c	6131	6019	0.690	0.637	9449.0				
-263 bp a	7573	7461	0.728	0.675	11053.3				
pGL3 B a	379	267	0.820	0.767	348.1	386.5	41.2	0.08	0.01
pGL3 B b	468	356	0.847	0.794	448.4				
pGL3 B c	392	280	0.824	0.771	363.2				

APPENDIX 10 - Transient transfection results

The average and average deviation of all transient transfection experiments have been tabulated and are shown below. For all cell lines, the full length construct (-2551-maspin-pGL3) was designated as 100% activity, and thus all other constructs are relative to the activity of this construct.

Statistical analysis was carried out on the average of the triplicate wells. A one tailed, two sample equal variance students T-test was carried out using Microsoft Excel 98. The test was used to identify significant differences, at 99% or 95% level, between a construct and every other construct transfected into the same cell line. Statistical analysis of constructs between cell lines was not carried out.

10.6 and 10.7 below show the tabular results of the cell line comparison results shown graphically in chapter 4. For these results, the activity of the full length construct in MCF12a for breast cells, and PC3 for prostate cells was set as 100% activity.

10.1 HeLa

Construct	Activity	Avg Dev
-2551 bp	1.00	0.09
-2234 bp	1.97	0.19
-1478 bp	2.45	0.19
-1078 bp	1.68	0.20
-963 bp	0.59	0.11
-660 bp	0.66	0.10
-295 bp	2.56	0.21
-263 bp	1.83	0.16
-134 bp	1.97	0.06
pGL3 Basic	-0.09	0.07

Above. Averaged activity and average deviation of constructs in HeLa cells.

Right. Results of T-test for significant difference between constructs. Significance at 95% and 99% confidence level is indicated.

Confidence level	95%	99%
-2551 to -2234	significant	significant
-2551 to -1478	significant	significant
-2551 to -1078	significant	significant
-2551 to -963	significant	significant
-2551 to -660	significant	significant
-2551 to -295	significant	significant
-2551 to -263	significant	significant
-2551 to -134	significant	significant
-2234 to -1478	significant	not
-2234 to -1078	not	not
-2234 to -963	significant	significant
-2234 to -660	significant	significant
-2234 to -295	significant	significant
-2234 to -263	not	not
-2234 to -134	not	not
-1478 to -1078	significant	not
-1478 to -963	significant	significant
-1478 to -660	significant	significant
-1478 to -295	not	not
-1478 to -263	significant	not
-1478 to -134	significant	not
-1078 to -963	not	not
-1078 to -660	not	not
-1078 to -295	significant	significant
-1078 to -263	not	not
-1078 to -134	not	not
-963 to -660	not	not
-963 to -295	significant	significant
-963 to -263	significant	significant
-963 to -134	significant	significant
-660 to -295	significant	significant
-660 to -263	significant	significant
-660 to -134	significant	significant
-295 to -263	significant	significant
-295 to -134	significant	significant
-263 to -134	not	not
pGL3b to -2551	significant	significant
pGL3b to -2234	significant	significant
pGL3b to -1478	significant	significant
pGL3b to -1078	significant	significant
pGL3b to -963	significant	significant
pGL3b to -660	significant	significant
pGL3b to -295	significant	significant
pGL3b to -263	significant	significant
pGL3b to -134	significant	significant

10.2 MCF12a

Construct	Activity	Avg Dev
-2551 bp	1.00	0.08
-2234 bp	1.06	0.07
-1478 bp	1.28	0.09
-1078 bp	1.52	0.11
-963 bp	1.16	0.11
-660 bp	0.73	0.04
-295 bp	1.74	0.17
-263 bp	1.16	0.10
-134 bp	1.20	0.12
pGL3 Basic	0.01	0.00

Above. Averaged activity and average deviation of constructs in MCF12a cells.

Right. Results of T-test for significant difference between constructs. Significance at 95% and 99% confidence level is indicated.

Confiidence level	95%	99%
-2551 to -2234	not	not
-2551 to -1478	significant	significant
-2551 to -1078	significant	significant
-2551 to -963	significant	significant
-2551 to -660	significant	significant
-2551 to -295	significant	significant
-2551 to -263	significant	significant
-2551 to -134	significant	significant
-2234 to -1478	significant	significant
-2234 to -1078	significant	significant
-2234 to -963	not	not
-2234 to -660	significant	significant
-2234 to -295	significant	significant
-2234 to -263	not	not
-2234 to -134	not	not
-1478 to -1078	not	not
-1478 to -963	significant	not
-1478 to -660	significant	significant
-1478 to -295	not	not
-1478 to -263	significant	not
-1478 to -134	significant	significant
-1078 to -963	not	not
-1078 to -660	significant	significant
-1078 to -295	not	not
-1078 to -263	significant	not
-1078 to -134	significant	significant
-963 to -660	significant	significant
-963 to -295	not	not
-963 to -263	not	not
-963 to -134	significant	significant
-660 to -295	significant	significant
-660 to -263	significant	significant
-660 to -134	significant	significant
-295 to -263	not	not
-295 to -134	significant	not
-263 to -134	not	not
pGL3b to -2551	significant	significant
pGL3b to -2234	significant	significant
pGL3b to -1478	significant	significant
pGL3b to -1078	significant	significant
pGL3b to -963	significant	significant
pGL3b to -660	significant	significant
pGL3b to -295	significant	significant
pGL3b to -263	significant	significant
pGL3b to -134	significant	significant

10.3 MDA-MB-231

Construct	Activity	Avg Dev
-2551bp	1.00	0.20
-2235bp	2.20	0.46
-1479bp	2.21	0.08
-1079bp	2.02	0.11
-963bp	1.18	0.33
-660bp	0.53	0.06
-293bp	3.07	0.61
-268bp	1.95	0.42
-134bp	1.38	0.07
pGL3 Basic	0.01	0.02

Above. Averaged activity and average deviation of constructs in MDA-MB-231 cells.

Right. Results of T-test for significant difference between constructs. Significance at 95% and 99% confidence level is indicated.

Confidence level	95%	99%
-2551 to -2234	significant	significant
-2551 to -1478	significant	significant
-2551 to -1078	significant	significant
-2551 to -963	not	not
-2551 to -660	significant	significant
-2551 to -295	significant	significant
-2551 to -263	significant	significant
-2551 to -134	significant	not
-2234 to -1478	not	not
-2234 to -1078	not	not
-2234 to -963	significant	significant
-2234 to -660	significant	significant
-2234 to -295	significant	not
-2234 to -263	not	not
-2234 to -134	significant	not
-1478 to -1078	not	not
-1478 to -963	significant	significant
-1478 to -660	significant	significant
-1478 to -295	not	not
-1478 to -263	not	not
-1478 to -134	significant	significant
-1078 to -963	significant	not
-1078 to -660	significant	significant
-1078 to -295	significant	not
-1078 to -263	not	not
-1078 to -134	significant	significant
-963 to -660	significant	not
-963 to -295	significant	significant
-963 to -263	significant	not
-963 to -134	not	not
-660 to -295	significant	significant
-660 to -263	significant	significant
-660 to -134	significant	significant
-295 to -263	significant	significant
-295 to -134	significant	significant
-263 to -134	not	not
pGL3b to -2551	significant	significant
pGL3b to -2234	significant	significant
pGL3b to -1478	significant	significant
pGL3b to -1078	significant	significant
pGL3b to -963	significant	significant
pGL3b to -660	significant	significant
pGL3b to -295	significant	significant
pGL3b to -263	significant	significant
pGL3b to -134	significant	significant

10.4 PC3

Construct	Activity	Avg Dev
-2551 bp	1.00	0.06
-2234 bp	2.98	0.07
-1478 bp	3.20	0.34
-1078 bp	2.95	0.04
-963 bp	0.72	0.14
-660 bp	0.56	0.16
-295 bp	2.58	0.30
-263 bp	2.30	0.44
-134 bp	3.64	0.25
pGL3 Basic	0.04	0.02

Above. Averaged activity and average deviation of constructs in PC3 cells.

Right. Results of T-test for significant difference between constructs. Significance at 95% and 99% confidence level is indicated.

Confidence level	95%	99%
-2551 to -2234	significant	significant
-2551 to -1478	significant	significant
-2551 to -1078	significant	significant
-2551 to -963	significant	significant
-2551 to -660	significant	significant
-2551 to -295	significant	significant
-2551 to -263	significant	significant
-2551 to -134	significant	significant
-2234 to -1478	not	not
-2234 to -1078	not	not
-2234 to -963	significant	significant
-2234 to -660	significant	significant
-2234 to -295	not	not
-2234 to -263	significant	not
-2234 to -134	significant	not
-1478 to -1078	not	not
-1478 to -963	significant	significant
-1478 to -660	significant	significant
-1478 to -295	significant	not
-1478 to -263	significant	not
-1478 to -134	not	not
-1078 to -963	significant	significant
-1078 to -660	significant	significant
-1078 to -295	not	not
-1078 to -263	significant	not
-1078 to -134	significant	not
-963 to -660	not	not
-963 to -295	significant	significant
-963 to -263	significant	significant
-963 to -134	significant	significant
-660 to -295	significant	significant
-660 to -263	significant	significant
-660 to -134	significant	significant
-295 to -263	not	not
-295 to -134	significant	significant
-263 to -134	significant	significant
pGL3b to -2551	significant	significant
pGL3b to -2234	significant	significant
pGL3b to -1478	significant	significant
pGL3b to -1078	significant	significant
pGL3b to -963	significant	significant
pGL3b to -660	significant	significant
pGL3b to -295	significant	significant
pGL3b to -263	significant	significant
pGL3b to -134	significant	significant

10.5 CA-HPV-10

Construct	Activity	Avg Dev
-2551 bp	1.00	0.17
-2234 bp	2.67	0.37
-1478 bp	2.59	0.22
-1078 bp	2.01	0.24
-963 bp	1.08	0.19
-660 bp	1.08	0.40
-295 bp	4.49	0.48
-263 bp	2.23	0.35
-134bp	2.77	0.05
pGL3 Basic	0.16	0.19

Above. Averaged activity and average deviation of constructs in CA-HPV-10 cells.

Right. Results of T-test for significant difference between constructs. Significance at 95% and 99% confidence level is indicated.

Confidence level	95%	99%
-2551 to -2234	significant	significant
-2551 to -1478	significant	significant
-2551 to -1078	significant	significant
-2551 to -963	not	not
-2551 to -660	not	not
-2551 to -295	significant	significant
-2551 to -263	significant	significant
-2551 to -134	significant	significant
-2234 to -1478	not	not
-2234 to -1078	significant	not
-2234 to -963	significant	significant
-2234 to -660	significant	significant
-2234 to -295	significant	significant
-2234 to -263	not	not
-2234 to -134	not	not
-1478 to -1078	significant	not
-1478 to -963	significant	significant
-1478 to -660	significant	significant
-1478 to -295	significant	significant
-1478 to -263	not	not
-1478 to -134	not	not
-1078 to -963	significant	significant
-1078 to -660	significant	not
-1078 to -295	significant	significant
-1078 to -263	not	not
-1078 to -134	significant	significant
-963 to -660	not	not
-963 to -295	significant	significant
-963 to -263	significant	significant
-963 to -134	significant	significant
-660 to -295	significant	significant
-660 to -263	significant	significant
-660 to -134	significant	significant
-295 to -263	significant	significant
-295 to -134	significant	significant
-263 to -134	not	not
pGL3b to -2551	significant	significant
pGL3b to -2234	significant	significant
pGL3b to -1478	significant	significant
pGL3b to -1078	significant	significant
pGL3b to -963	significant	significant
pGL3b to -660	significant	significant
pGL3b to -295	significant	significant
pGL3b to -263	significant	significant
pGL3b to -134	significant	significant

10.6 Breast cell line comparison

Cell Line	MCF12a		MDA-MB-231	
Construct	Activity	Avg Dev	Activity	Avg Dev
-2551 bp	1.00	0.08	0.06	0.01
-2234 bp	1.06	0.07	0.13	0.02
-1478 bp	1.28	0.09	0.13	0.01
-1078 bp	1.52	0.11	0.12	0.01
-963 bp	1.16	0.11	0.08	0.01
-660 bp	0.73	0.04	0.03	0.00
-295bp	1.74	0.17	0.18	0.03
-263 bp	1.16	0.10	0.12	0.03
-134 bp	1.20	0.12	0.08	0.00
pGL3 Basic	0.01	0.00	0.00	0.00

Average of transfection data for breast cell lines. For comparison, the -2551 bp construct in MCF12a cells was assumed to have 100% activity, all other constructs are expressed relative to this.

10.7 Prostate cell line comparison

Cell Line	PC3		CA-HPV-10	
Construct	Activity	Avg Dev	Activity	Avg Dev
-2551 bp	1.00	0.06	0.67	0.12
-2234 bp	2.98	0.07	1.78	0.24
-1478 bp	3.20	0.34	1.73	0.15
-1078 bp	2.95	0.04	1.34	0.16
-963 bp	0.72	0.14	0.72	0.13
-660 bp	0.56	0.16	0.72	0.26
-295bp	2.58	0.30	3.00	0.32
-263 bp	2.30	0.44	1.49	0.24
-134 bp	3.64	0.25	1.84	0.04
pGL3 Basic	0.04	0.02	0.11	0.12

Average of transfection data for prostate cell lines. For comparison, the -2551 bp construct PC3 cells was assumed to have 100% activity, all other constructs are expressed relative to this.

10.8 MDA-MB-231 5-azacytidine co-transfections

Treatment	0 Hrs		24 Hrs	
	Avg	Avg dev	Avg	Avg dev
pGL3 no	0.00	0.01	0.01	0.01
pGL3 10	0.01	0.01	0.06	0.08
-2551 no	1.00	0.05	1.00	0.04
-2551 1μM	0.88	0.07	1.01	0.13
-2551 2μM	0.84	0.07	0.95	0.10
-2551 5μM	0.55	0.06	0.69	0.07
-2551 10μM	0.22	0.09	0.67	0.09

Averaged data from methyltransferase inhibitor co-transfections above. Below results of one tailed two sample equal variance Students T-test statistical analysis.

Treatment		95%	99%
Full Length construct	1 μM 0hrs	not	not
	2 μM 0hrs	not	not
	5 μM 0hrs	significant	significant
	1 μM 24hrs	not	not
	2 μm 24hrs	not	not
	5 μM 24hrs	significant	not
pGL3 Basic	no	significant	significant
	10 μM 24hrs	significant	significant