

Chapter 4

**Genomic Characterisation of *UPKIB*,
Identification of the Putative
UPKIB Promoter and Detection
of a 500bp CpG Island**

4.1 Introduction

The human UPKIB locus (Locus ID7348) has previously been localised to chromosomal region 3q13.3-q21 by *in situ* hybridisation (Finch *et al.*, 1997). In 1998, Lobban *et al.* (1998), Yuasa *et al.* (1998) and Finch *et al.* (1999) all independently cloned cDNAs coding for the putative UPKIB open reading frame (ORF). The largest, of size 786bp, contained no 5'- or 3'- untranslated region (UTR) sequence, containing only the open reading frame, with the 5' initiation (ATG) and 3' termination (TAA) sequences (GenBank accession AF015234, BG401684 & AF042331). More recently, portions of the 5'UTR and 3'UTR mRNA have also been identified for UPKIB (Adachi *et al.*, 2000).

Finch *et al.* (1999) published a 783bp UPKIB genomic DNA segment spanning 42bp of the third transmembrane region, 78bp of the second extracellular domain and 663bp of intronic sequence (GenBank accession AF067147). At the time this study commenced, this was the only genomic DNA sequence available for *UPKIB*, and exon/intron structure and the promoter had not yet been identified. Characterisation of the promoter sequence would provide a basis to help determine the mechanisms of regulation of this gene by identifying a transcription initiation site, putative transcription factor binding motifs and transcription factors responsible for regulation of UPKIB mRNA expression in normal tissues and transitional cell carcinomas.

4.1.1 Genomic cloning and analysis techniques

To characterise the UPKIB genomic sequence and to identify regulatory factors controlling its expression in both normal and cancerous states, a number of techniques can be employed. Knowledge of the 5'UTR of the mRNA, the

Table 4.1
Comparison of Some Common Cloning Vectors

Vector	Host	Structure	Insert Size
P1 clones	<i>E. coli</i>	Circular plasmid	70-100 kb
BACs	<i>E. coli</i>	Circular plasmid	up to 300 kb
PACs	<i>E. coli</i>	Circular plasmid	100-300 kb
YACs	<i>S. cerevisiae</i>	Linear chromosome	100-2 000 kb

BAC- bacterial artificial chromosome; PAC- P1 artificial chromosome; YAC- Yeast artificial chromosome; *E. coli*- *Escherichia coli*; *S. cerevisiae*- *Saccharomyces cerevisiae*. Adapted from a review by Monaco & Larin (1994).

transcription start site and the promoter sequence of UPKIB are all required to provide information on some of the important regulatory mechanisms for this gene.

4.1.1.1 P1 artificial chromosomes: cloning vectors and inserts

YACs (Yeast artificial chromosomes), PACs (P1 artificial chromosomes), BACs (bacterial artificial chromosomes) and overlapping P1 clones are all large cloning vectors which, due to their size and properties, can carry large DNA segments inserted efficiently and stably. A summary of these properties (Table 4.1), illustrates that among these vectors, YACs can incorporate the largest inserts, with up to 2Mb of DNA.

In practical terms, PACs and BACs are considered more valuable tools than YACs. PACs and BACs can store large inserts of up to 300 kb, have high cloning efficiencies, and they are not subject to the problems associated with YACs, such as chimerism (two unlinked DNA segments in the same clone) or instability of clones. PACs have an advantage over BACs because they have a higher DNA yield when propagated in bacterial cultures (review by Monaco & Larin, 1994).

Once PACs have been identified that contain sequences which include the gene of interest, a series of techniques can be employed to identify and clone the genomic DNA, allowing further characterisation of the gene. Techniques that can be employed to define sequences include chromosome walking, primer extension and 5'RACE.

4.1.1.2 Chromosome walking

Chromosome (gene/primer) walking is a technique based on sequencing a stretch of DNA from one end to the other using a series of overlapping clones. It can be used to determine DNA sequences within cosmids, artificial chromosomes, or other vectors, using a series of either degenerate or vector-specific primers.

Initially, various restriction enzyme digestions of the vector constructs and restriction maps can be performed, followed by Southern hybridisations using probes of identified sequences to the gene (Southern, 1975). Unknown fragments that hybridise to the probe can then be isolated and sub-cloned into other vectors and sequenced. Subcloning also allows vector-specific primers to be used to gain sequence information. DNA segments can be sequentially cloned by chromosome walking from known points.

Hybridisation (Smith/Birnstiel) mapping, through contig assembly, can be performed with a series of overlapping clones. These clones are obtained through restriction enzyme digestion of DNA with various enzymes, Southern hybridisation with a radiolabelled probe of known sequence, and subsequent isolation of hybridised DNA fragments. These fragments are then sequenced and overlapping sequences are arranged on a contig map, allowing characterisation of the complete DNA sequence (Smith & Birnstiel, 1976).

Several groups, most recently including Weinel *et al.* (2001) working on the *Pseudomonas putida* KT2440 genome project, have modified this basic technique. Weinel and colleagues performed Southern hybridisation, after separation by pulsed-field gel electrophoresis, of DNA either completely digested with rare-cutter restriction enzymes or partially digested with frequent-cutter restriction enzymes. Probes were derived from DNA fragments located at the sequence ends originating from digestions with rare-cutter restriction enzymes. Hybridised fragments were isolated, sequenced, and assembled in high-resolution Smith/Birnstiel maps, providing gap closure for DNA sequences of interest.

4.1.1.3 Primer extension and 5'RACE

As performed by Hotta *et al.* (1992) for the tetraspanin family member CD63, primer extension is a useful tool to identify the 5'UTR of mRNAs, aiding in characterisation of promoter initiation and regulation. It involves a reverse transcription reaction with a radiolabelled primer specific for the mRNA of interest and subsequent direct sequencing, with the same primer, from genomic DNA containing exon 1 and the start of transcription. Sequencing from the genomic DNA will define the transcription start sequence, based on the size of the primer extension product. Further sequencing data from the genomic DNA will then define the upstream promoter sequence.

5'RACE (rapid amplification of cDNA ends/anchor PCR) is a technique similar to primer extension, which initially amplifies the cDNA with a reverse transcription step and then utilises the addition of a cytosine poly-linker to the 3' end of the newly synthesized cDNA. Using a primer annealing to the poly-linker tail and a second gene-specific primer, PCR will amplify the whole 5'UTR and the cDNA can subsequently be analysed through sequencing. The start of transcription of the

5'UTR is characteristic of exon 1. With these techniques, the UPKIB promoter can be compared with promoters of other tetraspanin members.

4.1.2 CpG islands, methylation, and acetylation of promoter sequences

In mammals and other vertebrates, approximately half of the gene promoters, in particular those of housekeeping genes, contain unmethylated CpG dinucleotides, with 60-90% of the remainder of the genome methylated at CpG sites (Ng & Bird, 1999). Cytosine residues adjacent to guanine residues at CpG dinucleotides are generally methylated to become 5'methyl-cytosine (5-MeC).

These CpG dinucleotides are generally found in clusters (islands) at the 5' ends of genes (Bird, 1986). CpG islands can range from 0.5kb to 5kb in size, have a GC content of approximately 60-70%, have a CpG:GpC ratio of at least 0.6 and the islands are usually unmethylated in housekeeping genes (Singal & Ginder, 1999). Due to the high GC content, a range of GC-rich transcription factor binding sites including that for Sp-1, can be found at these locations. Methylation can either reduce the binding affinities of activator transcription factors to the DNA or it can recruit sequence-specific transcriptional repressors (Zhang *et al.*, 1986; Prendergast & Ziff, 1991; Clark *et al.*, 1997).

Methylation and acetylation appear closely linked to transcriptional activity of CpG islands. Histone acetyltransferases and demethylases help to unravel chromatin and expose DNA at gene promoters to allow transcriptional activation through CpG sites (Nan *et al.*, 1997, 1998a, 1998b). Histone acetyltransferases bind various motifs on vertebrate histone amino-terminal tail domains to aid in transcriptional activation, with amino acid binding sites SGRGKQGGKARAKAK, PEPAKSAPAPKKGSKKAVTKT, ARTKQTARKSTGGKAPRKQLATKAA, and

SGRGKGGKGLGKGGAKRHRK on core histones H2A, H2B, H3 and H4 respectively (Roth *et al.*, 2001). When DNA becomes methylated, proteins including DNA methyltransferases (DNMTs), MeCP2 (methyl-CpG binding protein), and histone deacetylases act in co-operation to remodel and condense chromatin and inactivate genes associated with the CpG islands.

In humans and mice, DNA methyltransferases including DNMT1A (Bestor *et al.*, 1988; Yen *et al.*, 1992), DNMT1B (Bonfils *et al.*, 2000), DNMT2 (Yoder & Bestor, 1998), DNMT3a and DNMT3b (Okano *et al.*, 1998) co-ordinate the conversions of cytosines to 5-MeC, by the addition of methyl groups to the 5-positions of the cytosine rings (Bird, 1992; Robertson *et al.*, 1999). Studies in mice have shown the importance of DNMT1 for cellular viability, as defects in the DNMT1 gene show embryonic lethality (Li *et al.*, 1992, 1993a; Pradhan *et al.*, 1997). This indicates the importance of methyltransferases in transcriptional control in the embryo, which is likely to be associated with activation of normally imprinted genes (Section 5.1.1).

Germ line mutations in *DNMT3B*, observed in the immunodeficiency, centromeric instability and facial abnormalities (ICF) syndrome, are associated with abnormal methylation patterns in satellite DNA (Xu *et al.*, 1999). Deletion of *DNMT3B* in colon cancer-derived cell lines has exhibited a less than 3% reduction in cell methylation (Rhee *et al.*, 2002). In contrast, disruption of both *DNMT3B* and *DNMT1* reduced methylation by over 95% (Rhee *et al.*, 2002), showing that DNMT1 appears among the most important proteins in methylation, functioning to inactivate genes required for normal cellular control and viability.

4.2 Methods

4.2.1 Characterisation of a P1 Artificial Chromosome (PAC) containing the UPKIB gene in the region 3q13.3-q21

4.2.1.1 *Bacteria and plasmids*

A 778bp genomic portion of *UPKIB* was amplified by PCR using primers TM3 and ECD, derived from GenBank genomic segment AF067147 (Finch *et al.*, 1999). The genomic PCR product was sent to Research Genetics (Huntsville AL, USA), who carried out a hybridisation library screen to isolate a P1 artificial chromosome (PAC) containing an insert of the UPKIB sequence. RPCI (Roswell Park Cancer Institute) libraries containing more than 335 000 clones providing a 14x coverage of the human genome (<http://www.resgen.com/products/RPCIHum.php3>) were used for the screening. The library was established from DNA, obtained from a male normal fibroblast cell line, which was partially digested with the restriction enzyme *Mbo*I. The PAC vector (pNYPAC2N) (Appendix K) was derived by Research Genetics from the pAd10SacBII vector and contains a Kanamycin-resistance gene (Kan^R) for positive selection (Pierce *et al.*, 1992). PAC library clone 730-E-5 hybridised to the 778bp UPKIB genomic segment and was obtained as an agar stab in the host bacterium, *Escherichia coli* (*E. coli*).

4.2.1.2 *Maintenance of bacterial strains and selection of PACs with kanamycin*

E. coli containing the UPKIB-positive PACs were streaked from agar stabs onto nutrient agar plates (Section 2.2) containing kanamycin (75µg/ml) and incubated overnight in a 37°C incubator. Individual colonies were picked and placed into 200ml Luria Bertani broth (LB) with kanamycin (75µg/ml) and incubated overnight in Erlenmeyer flasks with aeration at 37°C. Fresh LB (50ml), containing

kanamycin (75µg/ml) and 1mM IPTG (isopropyl-β-D-galactosidase), was added to the overnight cultures and incubated at 37°C for a further 3 hrs. PAC plasmid DNA is normally maintained at a single copy per cell and the addition of IPTG induces a high copy number of PACs within a single bacterium (Dracopoli *et al.*, 1994). An 85µl aliquot of the culture was added to 15µl 100% glycerol, the mixture was vortexed and then stored at -70°C. From glycerol stocks, fresh cultures were prepared by seeding into LB with kanamycin (75µg/ml) and incubated with aeration at 37°C.

4.2.1.3 PAC isolation: Qiagen Maxi preparations

After propagation of bacteria for 19 hours (Section 4.2.1.2), 250ml cultures were centrifuged at 12 000 x g for 20 minutes at 4°C and cell pellets were processed according to instructions of the manufacturer of the Qiagen Maxi kit. The final PAC DNA product was resuspended in 500µl TE, divided into aliquots of 100µl, and stored at -20°C.

4.2.2 Southern hybridisation

4.2.2.1 Restriction enzyme digestions and gels for Southern

PAC DNA (approximately 2µg) was digested overnight at 37°C with *Hind*III, *Pst*I or *Xba*I (10U). Digests were then electrophoresed through a 0.8% 1x TAE low melting point agarose gel at 75V for 5 hours. An undigested PAC control and molecular weight size marker SPP1/*Eco*RI were included on all gels to allow identification of sizes of fragments and to check for adequate digestion of the PAC DNA.

4.2.2.2 Nitrocellulose membrane transfers

Gels were incubated in 0.25M HCl for 10 minutes at room temperature until the bromophenol blue bands from the loading dye turned from blue to green/yellow in colour. The gel was then placed face down on Whatman filter paper pre-soaked in 0.4M NaOH. A Hybond™ N+ nitrocellulose membrane cut to the length of the gel was placed on top, and two more Whatman filter papers were placed above the gel.

Paper towels were placed on top of the stack to absorb excess 0.4M NaOH and plastic sheeting was placed on exposed spaces around the NaOH container to prevent NaOH evaporation. DNA transfer was allowed to proceed overnight. The membrane was rinsed in 2x SSC, DNA was cross-linked with a UV crosslinker at 120 000 $\mu\text{J}/\text{cm}^2$ (Stratagene, UV Stratalinker 1 800), and stored dry at room temperature until hybridisation.

4.2.2.3 Prehybridising the membrane

The membrane was incubated for 3 hours at 65°C in 2.5ml 20x SSC and the blocking agents 0.5ml 100x Denhardt's reagent, 0.25ml 20% SDS and 200 μl sonicated salmon sperm DNA plus 6.75ml MilliQ-H₂O to give a final volume of 10.2ml. Prehybridising membranes with these blocking agents suppresses background hybridisation (Sambrook *et al.*, 1989).

4.2.2.4 Purifying a UPKIB cDNA probe for Southern blotting

A 796bp UPKIB cDNA sequence was obtained in pGEM-5zf(+) in *E. coli* from Dr Prue Cowled (Finch *et al.*, 1999) which coded for the UPKIB ORF (open reading frame) (GenBank accession AF042331). After a 100ml overnight culture in LB and ampicillin (final concentration 50 $\mu\text{g}/\text{ml}$), plasmid DNA was isolated using a BresaPure Midi kit (Geneworks) according to manufacturer's instructions and

resuspended in 100µl TE. Plasmid DNA (10µg) was digested with *SacI* and *SacII* restriction enzymes (10U), recognition sequences for which are located within the multiple cloning site of the vector flanking the cDNA. The 841bp cDNA product, with plasmid-derived ends, was separated on a 0.8% low melting point agarose gel, excised and purified using WIZARD technology according to the manufacturer's instructions (Promega). The DNA concentration was estimated by running products on a 1.0% agarose gel and manually comparing DNA fragments with known amounts of the size marker SPP1/*EcoRI*.

4.2.2.5 *Random labelling of probes*

A 7µl aliquot of UPKIB cDNA (approximately 150ng) was denatured at 100°C for 5 minutes. The denatured sample was then placed on ice for 2 minutes, followed by the addition of 6µl decanucleotide primer, 6µl nucleotide cocktail A, 1µl Klenow DNA polymerase I enzyme (1U/µl) (Gigaprime labelling kit, Section 2.1) and 5µl [$\alpha^{32}\text{P}$]-dATP (1.85MBq) to a final volume of 25µl. The reaction mixture was then incubated at 37°C for 30 minutes.

4.2.2.6 *Removal of unincorporated $\alpha^{32}\text{P}$*

Preparation of Sephadex G-50 columns

A 2ml syringe barrel was plugged with cotton wool and loaded with 1.5ml of Sephadex G50 gel beads (Section 2.2). To the column, 200µl 2mM β -mercapto-ethanol and 10µl sheared salmon sperm DNA (10mg/ml) were added and the column was centrifuged at 400 rpm for 10 minutes, in a Hettich Rotanta/RP bench-top centrifuge, to dry and settle the G-50 resin.

Removal of unincorporated [$\alpha^{32}\text{P}$]-dATP

The reaction mixture (Section 4.2.2.5) was combined with 2 μl of 400mM EDTA (pH 8.0), 180 μl 2mM β -mercapto-ethanol, 1 μl bromophenol blue (10%) and 200 μl phenol:chloroform:isoamyl-alcohol (25:24:1), centrifuged for 1 minute at 10 000 x g, and the blue (top) aqueous layer containing the probe mixture was loaded onto the G-50 column. The column was centrifuged at 800 rpm for 6 minutes to separate unincorporated nucleotides, which were retained in the G50 gel and the labelled probe was eluted from the column.

Hybridisation

The probe (~200 μl) was denatured at 100°C for 5 minutes, cooled on ice for 2 minutes, added to the pre-hybridisation solution and membrane and allowed to hybridise to the blot overnight at 65°C.

4.2.2.7 Membrane washes

The hybridisation mixture was removed and the membrane washed in 2x SSC at 65°C for 10 minutes, followed by washing in 100ml 1x SSC/0.1% SDS at 65°C for 30 minutes. Radioactivity on the membrane was monitored with a Geiger counter. Counts greater than 50 per minute in expected hybridising areas, or counts in non-hybridising areas such as in membrane corners, indicated the need for a higher stringency wash in 100ml 0.1x SSC/0.1% SDS (Section 2.2) for 20 minutes at 65°C. Counts of 10 to 20 per minute in DNA-containing areas were deemed adequate for autoradiography.

4.2.2.8 Autoradiography

The membrane (Section 4.2.2.7) was exposed to Amersham hyperfilm MP at

-70°C. The film was developed through an automated film processor (The Queen Elizabeth Hospital, Radiology Department).

4.2.2.9 Stripping the membrane for reuse

SDS (20%; 6.25ml) was added to boiling MilliQ-H₂O to a final volume of 250ml and final SDS concentration of 0.5%. The solution was poured over the membrane and agitated until the solution cooled to room temperature. The membrane was blotted on paper towels (DNA-side up), wrapped in cling wrap, and stored at room temperature.

4.2.3 Long-template PCR

Reaction mix for PCR

To 36.8µl sterile MilliQ-H₂O were added 5µl 10x Expand polymerase PCR buffer (Section 2.1), 1µl dNTPs (40mM), 2µl 5' sense primer (50ng/µl), 2µl 3' anti-sense primer (50ng/µl) and 0.1µl Expand polymerase (0.5U/µl) (Roche Diagnostics). PAC DNA (2µl, 20ng) was then added to a final volume of 50µl.

Cycling conditions for PCR

Cycling conditions for standard PCRs using Expand polymerase involved an initial 2 minutes at 94°C and then 45 cycles involving 1 minute at 94°C, 30 seconds at 58°C, and 1 minute at 68°C. Seven cycles were then performed involving 1 minute at 94°C, 30 seconds at 58°C, and an elongation step at 68°C for an initial 8 minutes in the first cycle, plus 20 seconds extra per additional cycle; followed by a final 7 minute elongation at 68°C before incubation at 4°C.

Analysis of PCR products and WIZARD purification

DNA samples (10µl of a 50µl PCR reaction) were run on a 1.0% agarose gel containing 1x TAE buffer and ethidium bromide and were subsequently visualised under a UV transilluminator (Ultra-lūm model MEB-20, Paramount, CA, USA) with Kodak image analysis software (v2.0.2) (Section 2.3.4). For extraction of specific DNA fragments, PCR products were run on a 0.8% low melting agarose gel containing 1x TAE buffer, appropriate bands were excised, and products were purified using a WIZARD column (Promega) according to the manufacturer's instructions, resulting in a final volume of 50µl. Five µl was run on a 1.0% agarose gel and visualised (Section 2.3.3.4) to check for DNA purity and allowing for manual quantitation by comparison with size marker SPP1/*EcoRI*.

4.2.4 Isolation of RNA from RT112 cells, RT-PCR and primer extension

4.2.4.1 Isolation of RNA from RT112 bladder carcinoma cell line

Cells from the bladder carcinoma-derived cell line RT112 were propagated as previously outlined (Section 2.3.1) and RNA isolated from these cells using TRI-Zol reagent (Invitrogen) according to protocols in Section 2.3.2.2.

4.2.4.2 RT-PCR of RT112 RNA

mRNA isolated from RT112 cells was reverse transcribed using Superscript II reverse transcriptase (Life Technologies) and an oligo-(dT)₁₅ primer (Promega), and the resulting cDNA was amplified by PCR with primers F1 and R1, as outlined in Sections 2.3.2.4 and 2.3.3.3.

4.2.4.3 Labelling 5' ends of 5'-UPKIB primer and molecular size marker SPP1/EcoRI with T4 polynucleotide kinase for subsequent primer extension

To 10µl 5'-UPKIB primer (50ng/µl) (sequence in Table 2.1) or SPP1/*EcoRI* DNA size marker (50ng/µl), were added 31µl sterile MilliQ-H₂O, 5µl 10x Kinase Buffer, 3µl [$\gamma^{32}\text{P}$]-dATP (1.11MBq) and 1µl T4 Polynucleotide Kinase (1U/µl) to a final volume of 50µl per reaction. The reaction mixtures were then incubated at 37°C for 30 minutes and the reactions stopped with the addition of 2µl 0.5M EDTA (pH 8.0). An equal volume (50µl) of phenol/chloroform/isoamyl alcohol (25:24:1) was added to each mixture, the solutions vortexed for 1 minute, centrifuged for 2 minutes at 12 000 x g and the upper aqueous phase transferred to fresh 1.5ml tubes. NaAc (3M; 0.1x volume) and absolute ethanol (2x volumes) were added and the DNA precipitated at -80°C for 30 minutes before being centrifuged at 12 000 x g for 5 minutes at 4°C. The supernatants were removed, and the pellets vacuum dried. ^{32}P -labelled SPP1/*EcoRI* was resuspended in 10µl TE and 2µl 6x loading buffer before being stored at 4°C. 5'-UPKIB primer was used dry in the subsequent primer extension (Section 4.2.4.4).

4.2.4.4 Primer extension using Superscript II reverse transcriptase **Annealing and extension**

RT112 RNA (6µl, 100ng) was added to [$\gamma^{32}\text{P}$]-dATP end-labelled 5'-UPKIB primer (dried) and 2.7µl DEPC-treated H₂O to a final volume of 8.7µl, which was incubated at 75°C for 15 minutes to allow annealing of the primer to the RNA template and then cooled on ice. To the annealed product was added 4µl 5x Superscript II first strand buffer, 1µl H₂O (RNase/DNase-free), 2µl MgCl₂ (25mM), 1µl dNTPs (10mM), 2µl DTT (100mM), 0.3µl RNasin (3U/µl), and 1µl Superscript

II (5U/ μ l) in a final volume of 20 μ l. The reaction mixture was incubated at 42°C for 90 minutes, 70°C for 15 minutes and then stored at 4°C before analysis on a polyacrylamide gel.

10% PAGE gel

PAGE (10%) gels were prepared with 20ml formamide (40% final), 21g urea (7M final), 12.5ml 40% acrylamide/bisacrylamide (19:1), and 5ml 10x TBE to a final volume of 50ml. TEMED (155 μ l) and 1ml of 0.1% ammonium persulfate were added immediately prior to pouring to induce polymerisation and the gels were allowed to set overnight.

The wells of the gel were washed out with 1x TBE buffer to remove excess unpolymerised acrylamide and the gels were pre-run in 1x TBE buffer at 15W for 15 minutes to warm up the buffer and gel. The samples were denatured at 100°C for 5 minutes, chilled on ice, and loaded immediately. The gels were run at 10W for 2 hours and soaked in PAGE-fixative (Section 2.2) for 30 minutes at room temperature to remove urea from the gel. Each gel was dried at 80°C for 2hrs in a gel-drying vacuum apparatus and subsequently exposed to autoradiography film for various times (Section 4.2.2.8).

4.2.5 5' RACE (Rapid Amplification of cDNA Ends)

Protocols were followed according to the manufacturer's instructions and Taq polymerase (Promega) was used for PCR. Two combinations of gene specific primers (GSPs) used to amplify RT112 mRNA sequences were (i) rTM3, rTM1c and 5'2-UPKIB; and (ii) 5'2-UPKIB, GSP1 and 5'1-UPKIB (see Section 2.3.3.3 for sequence information).

5'RACE steps involved (i) reverse transcription of RT112 RNA with GSP1 followed by poly-C tailing of the single-stranded cDNA product; (ii) a primary PCR amplification with GSP2 and AAP (abridged anchor primer) primers; and (iii) a secondary PCR with an internal GSP3 primer and AUAP (abridged universal anchor primer) primer. The AAP primer was specific for the poly-C tail and the AUAP primer annealed to the region specific for AAP.

Cycling conditions for 5'RACE PCR

After an initial 4 minutes incubation at 94°C; 35 cycles were performed for 1 minute at 94°C, 1 minute at 55°C, and 1 minute at 72°C; followed by a final elongation for 7 minutes at 72°C.

Analysis of 5'RACE PCR products

PCR products (10µl) were analysed on 2.0% low melting point agarose gels (Section 2.3.3.4). Two independent secondary PCR reactions using primers AUAP and 5'2-UPKIB were performed on the primary PCR product, which had used primers AAP and rTM1c. The two 5'RACE products were subsequently excised, purified using WIZARD methods and sequenced with primer 5'2-UPKIB to verify sequencing information (Sections 2.3.3.4 & 2.3.4).

4.2.6 Computer-based methodologies

4.2.6.1 *BLAST, GenBank, ClustalW and GeneDoc alignments*

UPKIB sequences were identified using GenBank and BLAST homology searches using human EST (expressed sequence tag), GSS (genome sequence survey), HTGS (high throughput genomic sequence) and NR (non-redundant) databases. Sequences used for the BLAST searches were sequences with GenBank accessions AF042331, AB002155 and long-template 2.5kb and 3.0kb sequences.

To identify the transcription start site, DNA sequences including the 5'RACE sequence were aligned using the program ClustalW and the alignments were viewed in GeneDoc (Section 3.2.2.2). Long-template sequences were aligned manually using Microsoft Word.

4.2.6.2 *Transcription factor binding motif prediction programs*

To analyse 1 000bp of DNA sequence including the putative UPKIB promoter, exon 1 and partial sequence of intron 1, Internet-based transcription factor binding motif programs were used. These programs comprise MatInspector (v2.2) (Quandt *et al.*, 1995; <http://transfac.gbf.de/cgi-bin/matSearch/matsearch.pl>), TESS (Schug & Overton, 1997; <http://www.cbil.upenn.edu/tess>) and TFSearch (v1.3) (Akiyama, 1998; <http://www.cbrc.jp/research/db/TFSEARCH.html>).

4.3 Results

4.3.1 Genomic assembly of *UPKIB* and identification of the *UPKIB* promoter

4.3.1.1 *PAC clones and hybridisation screening*

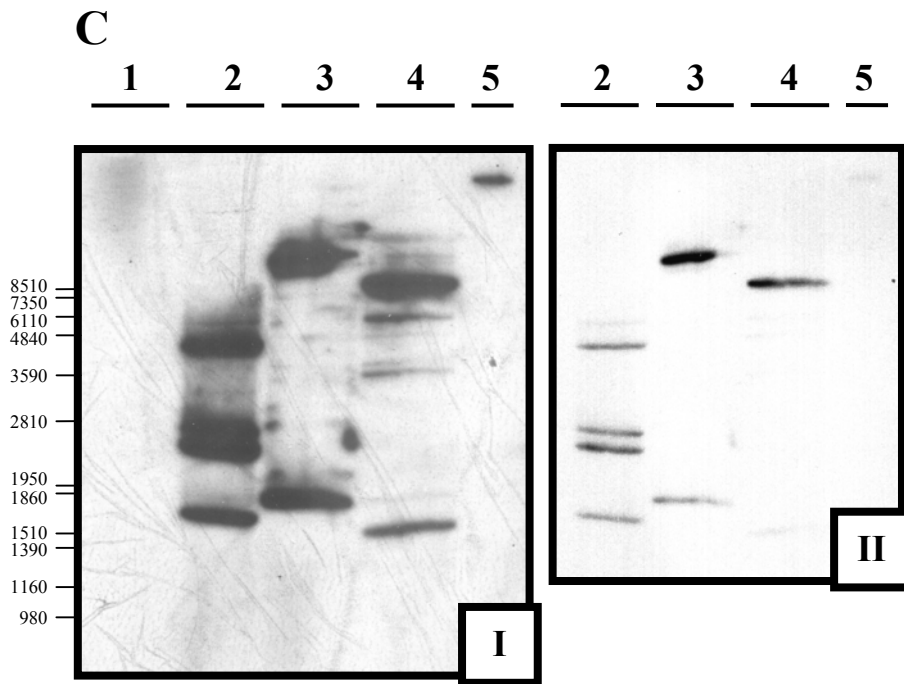
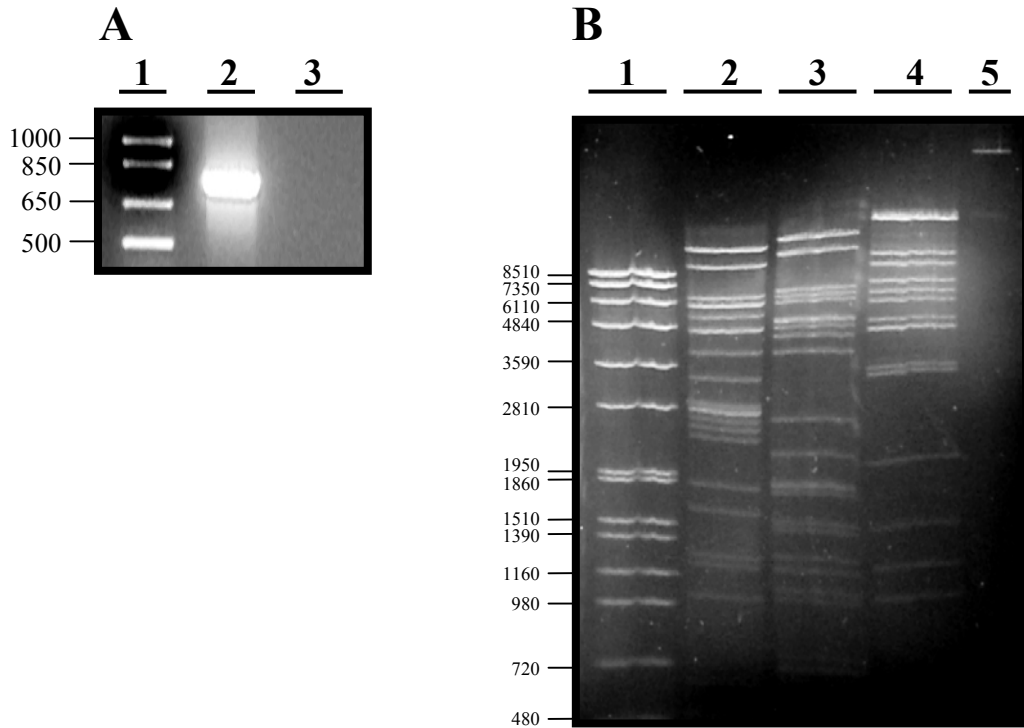
A 778bp UPKIB gDNA fragment, spanning two short cDNA segments and a small internal intron (Finch *et al.*, 1997) was sent to Research Genetics to identify and isolate PACs containing this UPKIB genomic sequence (Section 4.2.1.1). The PAC 730-E-5 construct obtained was propagated and DNA was extracted (Sections 4.2.1.2 & 4.2.1.3).

The PAC DNA was subsequently amplified using primers TM3 and ECD, to ensure that the UPKIB genomic segment was within the PAC supplied by Research Genetics (Fig. 4.1A). The 778bp genomic UPKIB fragment was sequenced in both directions using TM3 and ECD primers to confirm the identity of the sequence as

Figure 4.1

Southern Hybridisation of PAC 730-E-5 DNA Containing UPKIB Genomic DNA with a UPKIB cDNA Probe

A. 1.0% agarose gel stained with ethidium bromide, showing a 778bp PCR product of UPKIB gDNA from the PAC clone amplified with primers TM3 and ECD. Lanes: (1) 1kb+ molecular size marker, 500ng; (2) TM3/ECD primer PCR amplification of the PAC gDNA insert; (3) water control. **B.** 0.8% agarose gel stained with ethidium bromide, showing the digestion of the PAC clone 730-E-5 with a range of restriction enzymes. Lanes: (1) SPP1 molecular size marker, 500ng; (2-4) PAC clone (500ng) digested with *HindIII*, *PstI*, and *XbaI* respectively; (5) uncut PAC, 100ng. **C.** Southern hybridisation using the ORF of UPKIB cDNA as a probe, on a membrane prepared from the restriction digestion in Figure 4.1B. [I.] 16 hour autoradiographic exposure of hybridised membrane. [II.] 1 hour autoradiographic exposure of hybridised membrane. Lanes: as per Figure 4.1B. **D.** Fragment sizes obtained by Southern hybridisation from results in Figures 4.1B and 4.1C. Numbers represent manually estimated fragment sizes from each lane, and bracketed numbers represent actual sizes, as later determined from analysis of the chromosome 3q contig, GenBank accession AC083800.



D

PAC <i>Hind</i> III	PAC <i>Pst</i> I	PAC <i>Xba</i> I
5.9kb (5.963kb)	>10kb (15.313kb)	>10kb
4.6kb (4.731kb)	1.8kb (1.835kb)	>10kb (10.792kb)
2.6kb (2.672kb)	(0.700kb)	>8.5kb
2.4kb (2.406kb)	-	5.8kb (5.748kb)
1.7kb (1.654kb)	-	4.0kb (2.523kb)
(1.310kb)	-	3.6kb (2.412kb)
(1.054kb)	-	1.8kb
-	-	1.6kb (1.580kb)
-	-	(0.270kb)

UPKIB. This PAC was subsequently used to investigate the genetic organisation of *UPKIB*.

4.3.1.2 Southern hybridisation with the *UPKIB* cDNA probe

The *UPKIB*-positive PAC clone was digested with common 6-cutter restriction enzymes *Hind*III, *Pst*I and *Xba*I (Section 4.2.2.1). The digested DNA fragments were analysed together with uncut PAC DNA, to show multiple DNA bands from the digestions on an agarose gel stained with ethidium bromide (Fig. 4.1B). No undigested fragments were apparent in lanes 2-4 (Fig. 4.1B), thus demonstrating either partial or complete digestion of DNA samples. This gel was subsequently analysed by Southern blotting.

An 841bp *UPKIB* cDNA probe (Section 4.2.2.4) was used in Southern hybridisation of the PAC-digested DNA fragments (Fig. 4.1C, panels I & II). The panels represent different times of exposure, which were both shown to allow closer examination of the fragment sizes due to the strength or faintness of bands in panels I and II respectively. Molecular weights of the digested fragments were initially estimated manually by comparison with size marker SPP1/*Eco*RI (Fig. 4.1D). Other sizes in brackets (Fig. 4.1D) are measurements obtained later when contig AC083800 (Section 4.3.1.3) became available as a result of the Human Genome Project.

4.3.1.3 Comparison with contig AC083800 enzyme sites

Since the commencement of this study in 1998, the Human Genome Project has advanced swiftly and more extensive sequences of *UPKIB* have subsequently been obtained from the relevant databases. In this regard, the initial Southern hybridisation approach to genomic characterisation has progressively been replaced with updated human genome information. GenBank searches were initially used to

determine UPKIB sequences. Subsequently, BLAST searches were performed, initially with an UPKIB ORF sequence (GenBank accession AF042331) and later with a UPKIB mRNA sequence with a longer 5' UTR and complete 3'UTR (GenBank accession AB002155). A contig from region chromosome 3q was obtained from the HTGS (high throughput genomic sequence) database which initially contained 44 unordered sequences (circa March, 2000). This contig, accession AC083800, has slowly been assembled into a single sequence over the past 2 years (v13, March, 2002). By searching the AC083800 contig sequence in Microsoft Word for the restriction enzyme sites, the positions of the digestion sites from the enzymes used in Southern analysis were determined manually (Fig. 4.1D, numbers in brackets).

As seen from the data (Fig. 4.1C & 4.1D), *HindIII* and *PstI* appear to have digested the PAC clone completely. The smaller bands were not distinguishable on the gel (Fig. 4.1B) for these two enzyme digestions. *XbaI* appeared to only partially digest the PAC DNA, as not all bands produced by *XbaI* correlated with expected sizes in contig AC083800. Results from Southern blots will be further discussed in Section 4.4.

4.3.1.4 Long-template PCR

During the time that the Genome project revealed more sequence information, in addition to Southern hybridisation, long-template PCR was also used on the PAC clone to link together genomic sequences of unknown size. A BLAST search with the UPKIB cDNA ORF sequence (GenBank AF042331), using the GSS (genome sequence survey) database, identified a sequence (GenBank accession AQ318241) spanning one UPKIB exon (180bp) and flanking intronic sequences (311bp and 95bp for upstream and downstream sequences respectively). Primers TM1c and rTM3,

located in the cDNA ORF sequence, amplified a PCR product of approximate length 2.5kb, containing approximately 2.3kb of intronic sequence. Primers ex/int1 AQ318241 and rECD located further downstream, were designed within the intronic portion of the GSS entry and from the exonic portion of the second extracellular domain respectively and were used to amplify a PCR product of approximate size 3.0kb.

Amplification products of length 2.5kb and 3.0kb span exons 4-5 and 6-7 respectively (Fig. 4.2A) and can be plotted relative to the Finch *et al.* (1999) UPKIB ORF sequence (Fig. 4.2B). The 2.5kb and 3.0kb products were sequenced at both ends using TM1c and rTM3 primers, and ex/int1 AQ318241 and rECD primers respectively (Appendix L). Long-template PCR 2.5kb and 3.0kb sequences were used as targets in BLAST to identify overlaps with sequences from GSS accession AQ318241 and the latest version of AC083800 (v13 March, 2002). The long-template PCR sequences were aligned with Microsoft Word against AQ318241 and AC083800 sequences to confirm the identity of the long-template PCR products (Fig. 4.3 & 4.4).

4.3.2 Primer extension

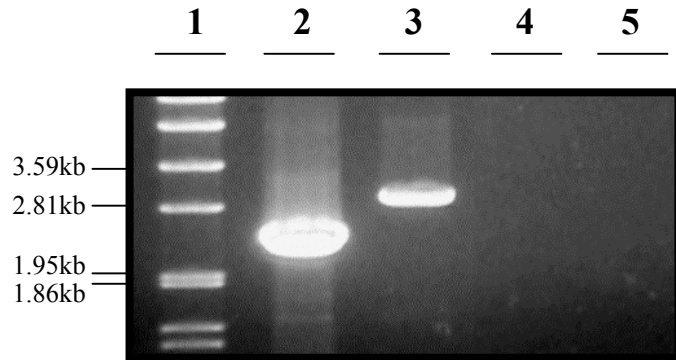
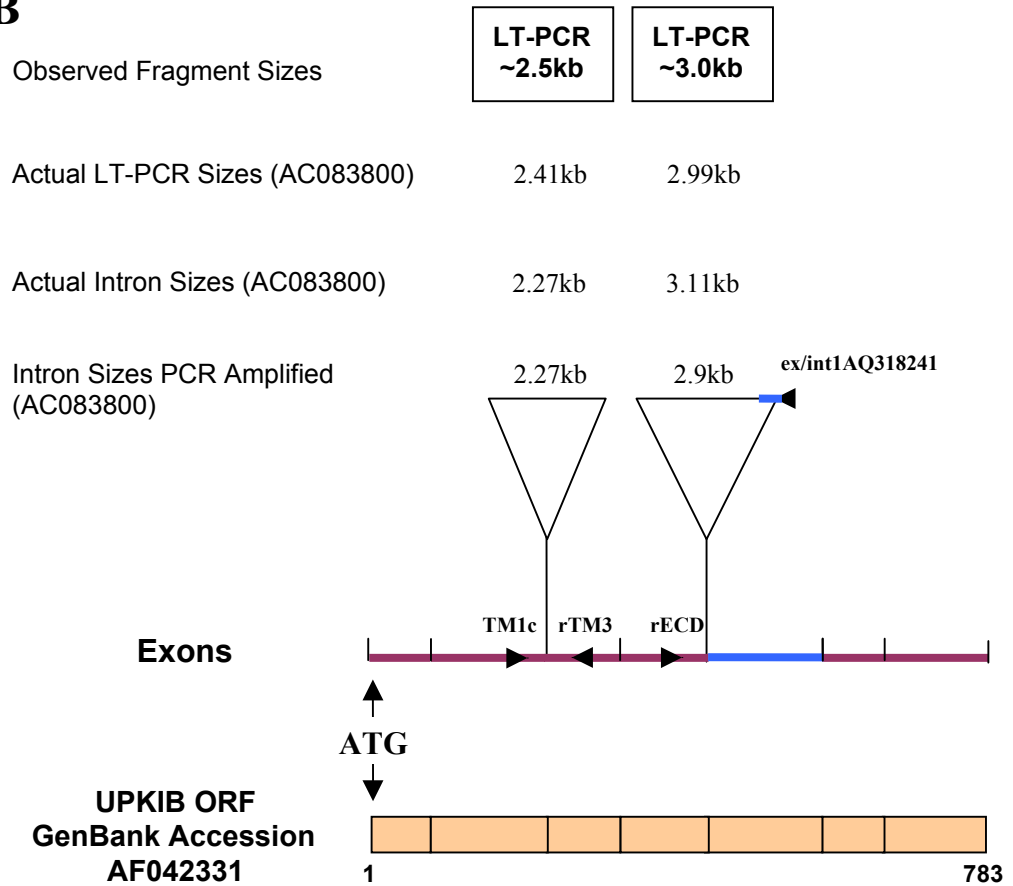
Primer extension is a method allowing extension of the mRNA, as single-stranded cDNA, in the 5' direction of the mRNA transcript (Sambrook *et al.*, 1989). For *UPKIB*, primer extension provides identification of the 5'UTR, allowing comparison with the 5'UTR proposed by Adachi *et al.* (2000) and determination of the transcription start site (Section 4.1.1.3).

The transitional cell carcinoma-derived cell line RT112 was used for primer extension, since it expresses UPKIB mRNA at high levels, as determined by RT-

Figure 4.2

Long-template PCR on PAC clone 730-E-5, Amplifying 2.5kb and 3.0kb Products

A. 1.0% TAE agarose gel stained with ethidium bromide, showing the long-template PCR (LT-PCR) products. Lanes: (1) 500ng SPP1/*Eco*RI molecular weight marker; (2) rTM1c/TM3 primers used to amplify PAC DNA giving a ~2.5kb fragment; (3) rECD/ ex/int1AQ318241 primers used to amplify PAC DNA giving a ~3.0kb fragment; (4) water control of (2); (5) water control of (3). **B.** Representation of long-template PCR (LT-PCR) products on the open reading frame of UPKIB. The ATG translation start site and relative positions of the long-template PCR primers are indicated. Exonic sequences from HTGS (high throughput genome survey) and GSS (genome survey sequence) databases, with GenBank accessions AC083800 and AQ318241, are shown in purple and blue respectively. Circa October 2000, the 3.0kb LT-PCR product spanned a portion of the sequence as yet undetermined by the HTGs and GSS sequences. The 2.9kb intronic sequence derived from the amplified 3.0kb LT-PCR product is shown as less than the actual intron 6 size (3.11kb), as the position of the ex/int1AQ318241 primer lies within intron 6. The actual LT-PCR size, the actual intron sizes and the intron sizes PCR amplified were determined by the analysis of the GenBank accession AC083800 sequence in Microsoft Word.

A**B**

— High Throughput Genomic Sequence GenBank Accession AC083800
— Genome Survey Sequence GenBank Accession AQ318241

A

```
2.5kbTM1c: 41          tctgttctaggcatttgtaggcatcatgaagtcagcaggaaaattcttctggcggtaac 100
AC083800 : 104293    tctgttctaggcatt-gtaggcatcatgaagtcagcaggaaaattcttctggcggtaac 104351

2.5kbTM1c: 101       acctcaaaattctctggagttgtctccctgagccaattgtaatggattctttcttcttca 160
AC083800 : 104352    acctcaaaattctctggagttgtctccctgagccaattgtaatggattctttcttcttca 104411

2.5kbTM1c: 161       ctgagcaatgctttggctttcagtgggggagcaggaagtaccagataaggggtgaggagg 220
AC083800 : 104412    ctgagcaatgctttggctttcagtgggggagcaggaagtaccagataaggggtgaggagg 104471

2.5kbTM1c: 221       ggaagactttccctgcagtgaaatgggtccacatggaggcacaagatgccctccactattgc 280
AC083800 : 104472    ggaagactttccctgcagtgaaatgggtccacatggaggcacaagatgccctccactattgc 104531

2.5kbTM1c: 281       aagcccagaaggctccagtacagagaagccaaggggagacacctgggggtgacctaatg 340
AC083800 : 104532    aagcccagaaggctccagtacagagaagccaaggggagacacctgggggtgacctaatg 104591

2.5kbTM1c: 341       agtattttccatattgtgcctctacagctgggtgtctgcatacacccagtaggtacagcta 400
AC083800 : 104592    agtattttccatattgtgcctctacagctgggtgtctgcatacacccagtaggtacagcta 104651

2.5kbTM1c: 401       ggagctatataggccaaggagagacacagtgagaattcattcattagattttttctct 460
AC083800 : 104652    ggagctatataggccaaggagagacacagtgagaattcattcattagattttttctct 104711

2.5kbTM1c: 461       tggcaataagttgtttaaaacattntgatattaattaggcattnaatatattaggatata 520
AC083800 : 104712    tggcaataagttgtttaaaacattntgatattaattaggcattnaatatattaggatata 104771

2.5kbTM1c: 521       aatataaaatagcatacatctntaaaagaaa 553
AC083800 : 104772    aatataaaatagcatacat-ttttaaaagaaa 104803
```

B

```
2.5kbrTM3: 26        atactataaacatcagaatgaaatactggaagggaaataaatgagataaaaagaattacat 85
AC083800 : 106640    atactataaacatcagaatgaaatactggaagggaaataaatgagataaaaagaattacat 106581

2.5kbrTM3: 86        ttaccogtcaaagagcgacttgcccaataattatgtttttcagcaaatttatcatata 145
AC083800 : 106580    ttaccogtcaaagagcgacttgcccaataattatgtttttcagcaaatttatcatata 106521

2.5kbrTM3: 146       atcatgtaatcactcatgtagttattcaacaaccctcattgaccagtggttggattga 205
AC083800 : 106520    atcatgtaatcactcatgtagttattcaacaaccctcattgaccagtggttggattga 106461

2.5kbrTM3: 206       ggcttgggtgctagtcagctcaattctctccttttctccagagaacaaatagccattgaa 265
AC083800 : 106460    ggcttgggtgctagtcagctcaattctctccttttctccagagaacaaatagccattgaa 106401

2.5kbrTM3: 266       acagaccttctcagtcagtttctcctgagtttctattttaactgggaaaaacaagtacaa 325
AC083800 : 106400    acagaccttctcagtcagtttctcctgagtttctattttaactgggaaaaacaagtacaa 106341

2.5kbrTM3: 326       gagaatgaactgggtgtgtatgtgtgcacacacacattgtgagcattttacatgctcattt 385
AC083800 : 106340    gagaatgaactgggtgtgtatgtgtgcacacacaca-tgtgagca-tttacatgctcattt 106283

2.5kbrTM3: 386       acatgggtgtgcattt 400
AC083800 : 106282    acat-gtgtgcattt 106269
```

Figure 4.3

Sequencing and Alignment of a 2.5kb Long-template PCR Product with High Throughput Genomic Sequence (HTGS) Database Contig AC083800

A. BLAST alignment of a portion of the sequence obtained from the 2.5kb LT-PCR product, derived from primer TM1c, aligned against part of the GenBank accession AC083800 sequence. **B.** BLAST alignment of a portion of the sequence obtained

Figure 4.4

Sequencing and Alignment of a 3.0kb Long-template PCR Product with Internet Database Entries

A. BLAST alignment of a portion of the sequence obtained from the 3.0kb LT-PCR product, derived from primer rECD, aligned against part of the GenBank accession AC083800 sequence. **B.** BLAST alignment of a portion of the sequence obtained from the 3.0kb LT-PCR product, derived from primer ex/int1AQ318241, aligned against portions of the sequences from GenBank accessions AC083800 and AQ318241. Red sequence shows the AQ318241 sequence, which was originally used for ex/int1AQ318241 primer design.

A

3.0kbrECD: 32 **caaaacctgggacaggctcatgctccaggtaagacctg**cggtcttggggagatgccattt 91
AC083800 : 107448 **caaaacctgggacaggctcatgctccaggtaagacctg**tggtcttggggagatgccattt 107507

3.0kbrECD: 92 **ttatttactgntgggagtggtgtcatcacatgactcag**tgggaccctatatattttattt 151
AC083800 : 107508 **ttacttactggtgggagtggtgtcatcacatgactcag**tgggaccctatatattttattt 107567

3.0kbrECD: 92 **ttatttactgntgggagtggtgtcatcacatgactcag**tgggaccctatatattttattt 151
AC083800 : 107508 **ttacttactggtgggagtggtgtcatcacatgactcag**tgggaccctatatattttattt 107567

3.0kbrECD: 152 **caagccatgatgattagcttacatttgttggaa**tgattttgtcaagaaacttaggttaat 211
AC083800 : 107568 **caagccatgatgattagcttacatttgttggaa**tgattttgtcaagaaacttaggttaat 107627

3.0kbrECD: 212 **catacactctccctcttagaactcaggtatcatt**tacattaaagaaaaatattttctct 271
AC083800 : 107628 **catacactctccctcttagaactcaggtatcatt**tacattaaagaaaaatattttctct 107687

3.0kbrECD: 272 **acccacagtgaaactcgagtggaataactgcct**gtttcgcttttgggtacctcttcac 331
AC083800 : 107688 **acccacagtgaaactcgagtggaataactgcct**gtttcgcttttgggtacctcttcac 107747

3.0kbrECD: 332 **aactcctgtgagctagtcttggccaaggaccag**gccctatatcttgatttgatcacccaat 391
AC083800 : 107748 **aactcctgtgagctagtcttggccaaggaccag**gccctatatcttgatttgatcacccaat 107807

3.0kbrECD: 392 **tttctccaactggcaatcatagtttatgtccat**gtttatctccaatgccatcaaaagtcc 451
AC083800 : 107808 **tttctccaactggcaatcatagtttatgtccat**gtttatctccaatgccatcaaaagtcc 107867

3.0kbrECD: 452 **aagtcttccagaatatccatcatccccacattg**ccctctgtatgcaaattg-tatcc 510
AC083800 : 107868 **aagtcttccagaatatccatcatccccacattg**ccctctgtatgcaaattgtttatcc 107927

3.0kbrECD: 511 **tctacataactaccaatctgacctagcttttct**tactgggcccggctcctactccacac 570
AC083800 : 107928 **tctacataactaccaatctgacctagc-tttct**tact-ggcccctgctcctactccacac 107985

3.0kbrECD: 571 **cccaccaccaanttttnacangtctggcatcag**caaaactcccttatagccacca 626
AC083800 : 107986 **cccaccaccaactttttaca-gtctgtcatcagc**-aaact-cccttatagccacca 108038

B

3.0kxbint: 44 **gtagccattgacntgttacttaaaactgtctaag**cctcagtttccttaactgtaaaattgg 103
AQ318241 : 504 **gtagccattgacatggttacttaaaactgtctaag**cctcagtttccttaactgtaaaattgg 563
AC083800 : 110361 **gtagccattgacatggttacttaaaactgtctaag**cctcagtttccttaactgtaaaattgg 110302

3.0kxbint: 104 **aatatagtagctatctcatagagtttttttaagg**gattaaatgagttaataacttaatgt 163
AQ318241 : 564 **aatatagtagctatctcatagag 586**
AC083800 : 110301 **aatatagtagctatctcatagagtttttttaagg**gattaaatgagttaataacttaatgt 110242

3.0kxbint: 164 **gctcaatacagtgccctggccttatggtaaaact**cttaataaaatgtaattatgattgatact 223
AC083800 : 110241 **gctcaatacagtgccctggccttatggtaaaact**cttaataaaatgtaattatgattgatact 110182

3.0kxbint: 224 **gtactaatatattgatataacaatgtgagagag**gaataaattattctgatatacagagca 283
AC083800 : 110181 **gctcaatacagtgccctggccttatggtaaaact**cttaataaaatgtaattatgattgatact 110122

3.0kxbint: 284 **aattgagaatcataaaataaagattatatcaaag**caaatgagaatcataaaatataaa 343
AC083800 : 110121 **aattgagaatcataaaataaagattatatcaaag**caaatgagaatcataaaatataaa 110062

3.0kxbint: 344 **atthttagtggtgaagttatgagtaattgttct**tgttttgcttttctatacatttaaacat 403
AC083800 : 110061 **atthttagtggtgaagttatgagtaattgttct**tgttttgcttttctatacatttaaacat 110002

3.0kxbint: 404 **tttttagagtaattaggtattatttttataatg**gggagaaaaataaagtgaaaaaataacc 463
AC083800 : 110001 **tttttagagtaattaggtattatttttataatg**gggagaaaaataaagtgaaaaaataacc 109942

3.0kxbint: 464 **tggtgacacagagctatgtctccacataggggag**ggagaccaattaatacatcccttaca 523
AC083800 : 109941 **tggtgacacagagctatgtctccacataggggag**ggagaccaattaatacatcccttaca 109882

3.0kxbint: 524 **ctctggatagaattgagaatagaggacaaaaaca**aattcatatagctttccttctttggt 583
AC083800 : 109881 **ctctggatagaattgagaatagaggacaaaaaca**aattcatatagctttccttctttggt 109822

3.0kxbint: 584 **atcaattctcctaaaaatccttcaaaa**ggggaggaagttaaaaccatgggttatcttggga 643
AC083800 : 109821 **atcaattctcctaaaaatccttcaaaa**-ggggaggaagttaaaaccatggg-tatctt-gga 109765

3.0kxbint: 644 **agttntagaatgaatgtattcaagt-aacttgat**gattaatcaaccatctttaaagggga 702
AC083800 : 109764 **acttttagaatgaatgtattcaagtaaaacttgat**gattaatcaaccatctttaa--gga 109707

3.0kxbint: 703 **tattcnaatttngtcactatggttatngggaaat**atctcaaccaagaaaa 753
AC083800 : 109706 **tatac-aattttgtcactat-gttat-gggaaat**atct-caaccaagaaaa 109660

PCR (Lobban *et al.*, 1998) (Sections 1.4.4 & 5.3.3). ³²P-labelled primer 5'1-UPKIB, 39bp in length and beginning in the open reading frame at position 32 was used to amplify the 5'UTR. The resulting primer extension gel, exposed for 20 minutes or 5 minutes (Fig. 4.5A & 4.5B respectively), showed a fragment of size ~120bp. Unincorporated ³²P was observed beneath the product, and a smear was detected above the 120bp product, indicating either longer 5'UTR variants or else non-specific binding of the radiolabelled primer to other mRNA species (Fig. 4.5A & 4.5B). From this data, it was estimated that the major 5'UTR 90-100bp in length. Sequence information could not be obtained, and the reasons for this are discussed in Section 4.4.

4.3.3 5'RACE (5'- Rapid Amplification of cDNA Ends)

5'RACE uses the same principle as primer extension, by producing cDNA using reverse transcriptase on an RNA template. An adaptor sequence is attached to the 5' end of the cDNA and primers provided in the 5'RACE kit amplify the 5' cDNA product. The product can then be sequenced or cloned (discussed in Section 4.1.1.3).

5'RACE using RT112 RNA, allowed amplification of specific 5' cDNA products. Two 5'RACE reactions were carried out, as shown in a diagrammatic representation of all experiments (Fig. 4.6A), together with sequences of 5'RACE primers AAP and AUAP, showing the AAP poly-C binding site required for the elongated 5'UTR transcript. The initial reaction illustrated in panel I used primer rTM3 for the reverse transcription (RT). A primary PCR was employed using an internal gene specific primer rTM1c and the supplied 5'RACE abridged anchor primer (AAP) (Fig. 4.6B, lane 3). Positive control primers for the RT reaction, TM1

Figure 4.5

Primer Extension of UPKIB from Total RT112 mRNA, with Primer 5'1-UPKIB

A. Autoradiography (20 minute exposure) of $\gamma^{32}\text{P}$ labelled primer extension product.

B. Autoradiography (5 minute exposure) of $\gamma^{32}\text{P}$ labelled primer extension product.

A. and **B.** lanes: (1) $\gamma^{32}\text{P}$ labelled pUC18 (HpaII) molecular weight marker, 500ng;

(2) $\gamma^{32}\text{P}$ labelled UPKIB primer extension product derived from primer 5'1-UPKIB.

C. Comparison of 5' regions from Finch *et al.* (1999) GenBank sequence AF042331, Adachi *et al.* (2000) GenBank sequence AB002155 and primer extension results with primer 5'1-UPKIB, situated 32bp inside the ORF.

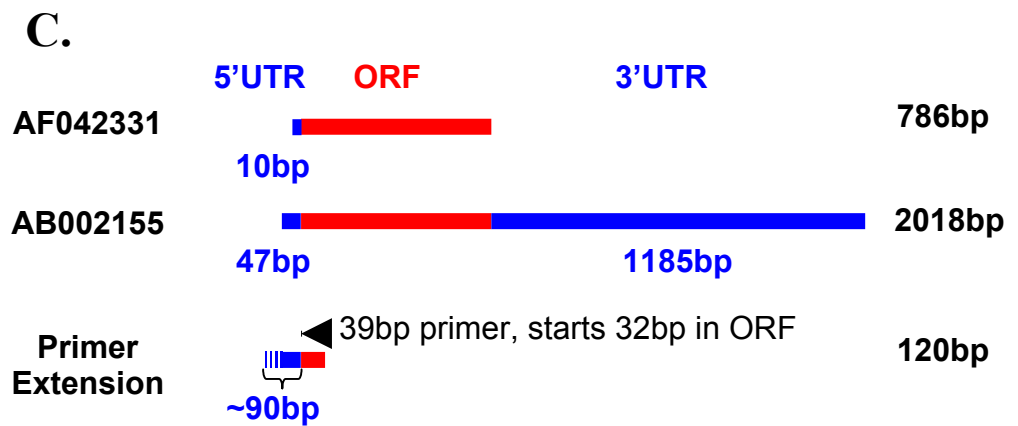
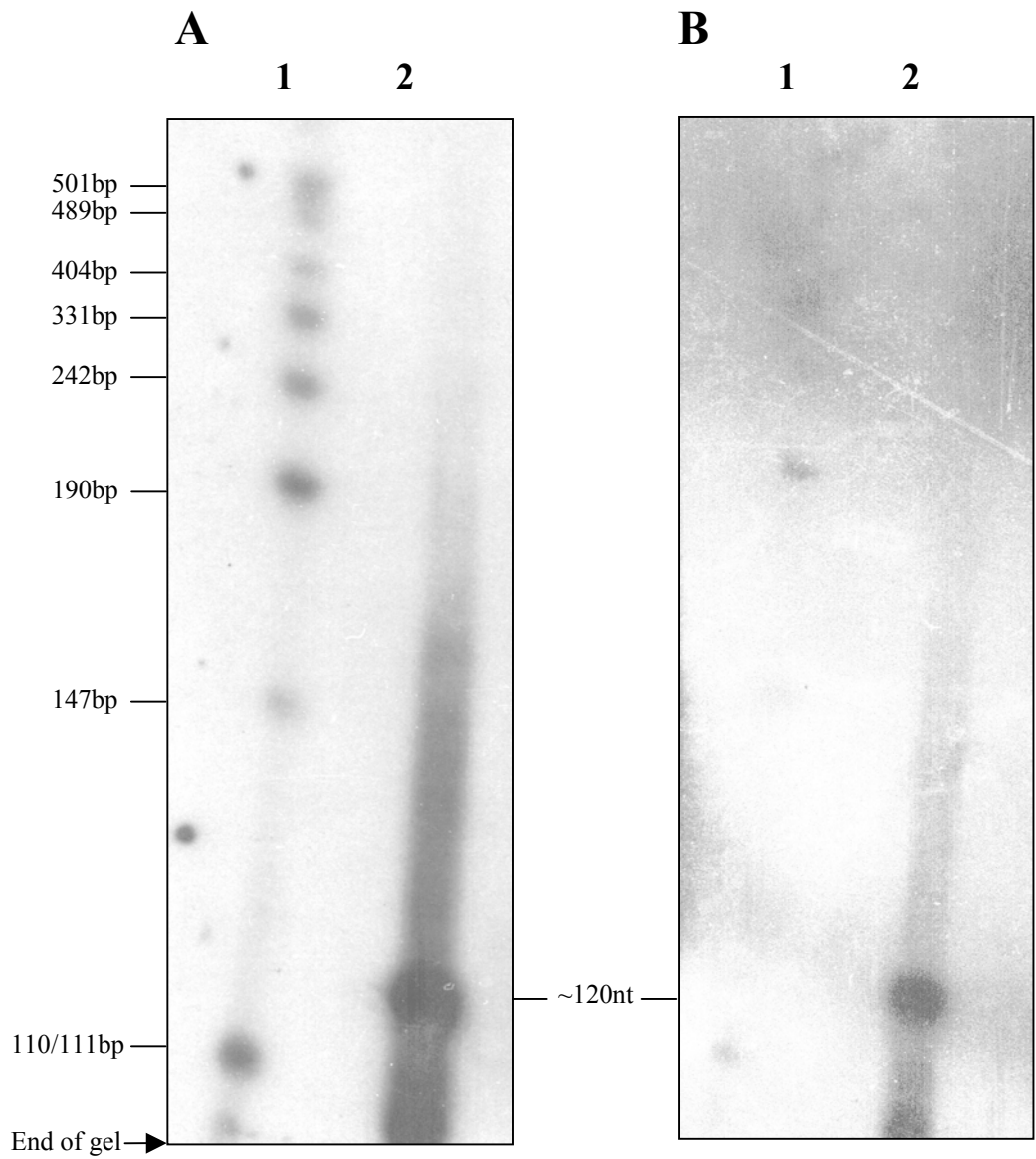
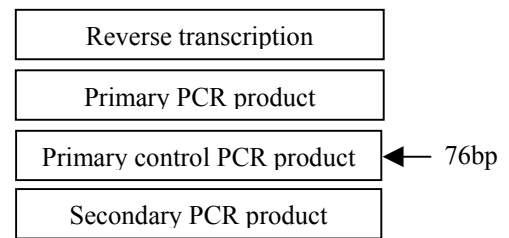
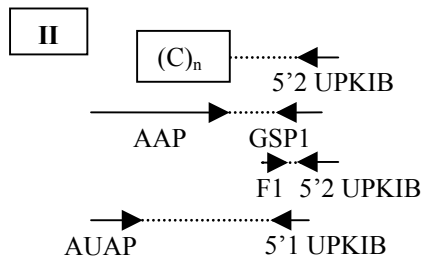
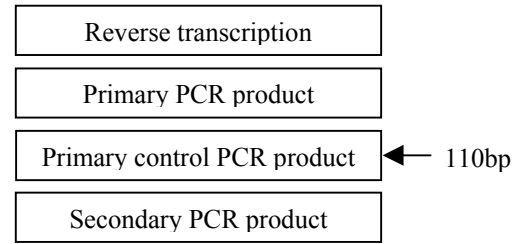
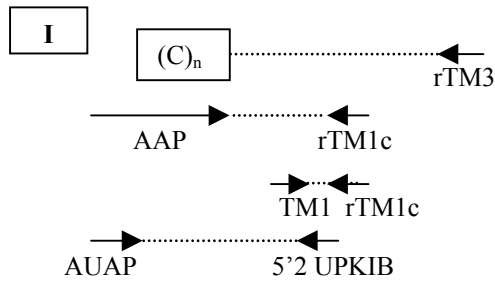


Figure 4.6

5'RACE Results after PCR of UPKIB cDNA from Cell Line RT112

A. UPKIB primers positions in the 5'RACE reactions. **[I.]** RT112 5'RACE 1 primary and secondary PCR products, as shown in B., lane 3 and C., lane 3 respectively, with product lengths of approximate 330bp and 220bp respectively. **[II.]** RT112 5'RACE 2 primary and secondary PCR products, as shown in B., lane 5, and C. lane 4 respectively, with products lengths of approximately 310bp and 200bp respectively. **B.** Primary PCR of 5'RACE with cell line RT112. Lanes: (1) 500ng 1kb+ molecular weight marker; (2) 1 μ g pUC18/HpaII molecular weight marker; (3) rTM1c/AAP primers with 5'RACE RT product derived from rTM3; (4) GSP1/AAP primers with 5'RACE RT product derived from 5'2-UPKIB, (5) positive control rTM1c/AAP+TM1 primers with RT 5'RACE product derived from rTM3; (6) positive control 5'2-UPKIB/AAP+F1 primers with RT 5'RACE product derived from 5'2-UPKIB; (7) water control rTM1c/AAP primers with RT 5'RACE product derived from rTM3; (8) water control GSP1/AAP primers with RT 5'RACE product derived from 5'2-UPKIB. **C.** Secondary PCR of 5'RACE with cell line RT112. Lanes: (1) 300ng pUC18/HpaII molecular weight marker; (2) 500ng 1kb+ molecular weight marker; (3) rTM1/AUAP primers with 5'RACE RT product derived from rTM3; (4) 5'2-UPKIB/AUAP primers with 5'RACE RT product derived from rTM3; (5) 5'1-UPKIB/AUAP primers with 5'RACE RT product derived from 5'2-UPKIB; (6) water control with primers rTM1/AUAP.

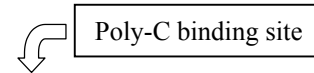
A



5' RACE Primers:

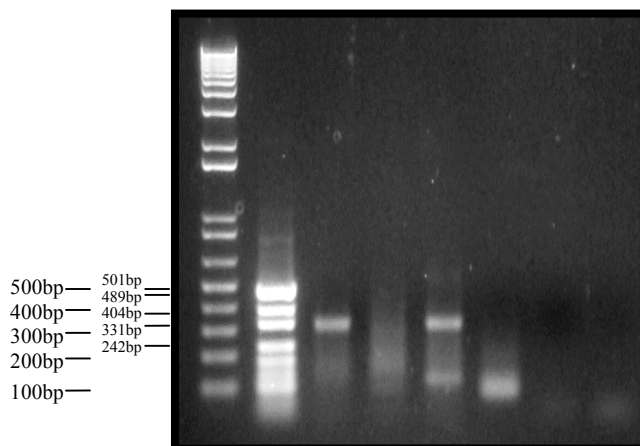
AAP 5' - GGC CAC GCG TCG ACT AGT AC GGI IGG GII GGG IIG -3'

AUAP 5' - GGC CAC GCG TCG ACT AGT AC -3'



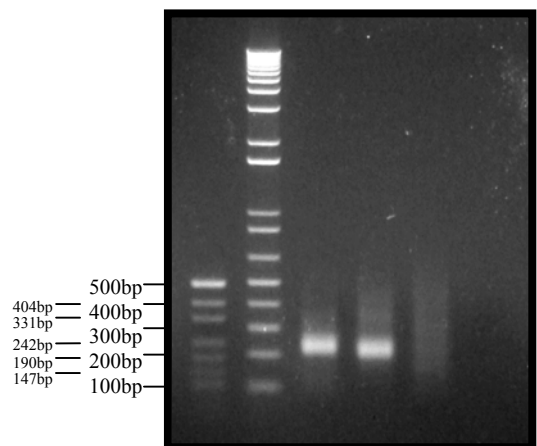
B

1 2 3 4 5 6 7 8



C

1 2 3 4 5 6



and rTM1c, were added to the AAP primer to show both amplification of an expected internal ~110bp fragment specific for *UPKIB*, and the primer extension product (~340bp) (Fig. 4.6B, lane 5). A secondary PCR with primer 5'2-UPKIB and supplied primer 5'RACE abridged universal anchor primer (AUAP) gave a product of approximate 220bp size (Fig. 4.6C, lane 3).

A second 5'RACE reaction (Fig. 4.6A, panel II), used primer 5'2-UPKIB for reverse transcription and the product was then amplified with primers GSP1 and supplied 5'RACE AAP (Fig. 4.6B, lane 4), which gave a non-specific smeared product. Internal positive control primers F1 and 5'2-UPKIB amplified an expected 76bp product (Fig. 4.6B, lane 6). A secondary PCR reaction, with the addition of the PCR product from primers GSP1 and AAP, was performed using primers 5'1-UPKIB and 5'RACE AUAP. A specific 5'RACE product of approximate size 200bp was observed in this reaction (Fig. 4.6C, lane 4).

Both 5'RACE results were indicative of a 5'UTR of approximate size 100bp. The 220bp 5'2-UPKIB/AAP sequence was sequenced twice in the direction of the transcription start site with primer 5'2-UPKIB (Appendix M). BLAST searches using the 5'RACE sequence were performed against EST and HTGS databases (NCBI). The results of these BLAST searches were aligned using ClustalW and displayed in GeneDoc (Fig. 4.7) to show the 5'RACE poly-5' tail region (highlighted in yellow), the transcription start site (highlighted in red) and the positions of the 5'2-UPKIB primer (highlighted in blue), exons and the ATG translation start site. An EST with GenBank accession BG401684 appears to express a non-coding exon 2 of size 112bp, but this exon has not been identified in any other cDNA or EST. The 5' end of this EST agreed with the guanine start site, identified in the 5'RACE sequencing results (Fig. 4.7). The transcription start site was also determined by

Figure 4.7

5'RACE Analysis: Identifying the Transcription Start Site and Exon 1

ClustalW multiple sequence comparison of the UPKIB 5'RACE sequence with other sequences from the National Centre for Biochemistry (NCBI) databases. The alignment was generated in the European Molecular Biology Lab ClustalW database and analysed with Genedoc, showing nucleotide conservation. Sequences were obtained through BLAST searches include: AC083800 genomic sequence contig, initially obtained through the high throughput genomic sequence (HTGS) database and then the non-redundant (NR) database; NR database entries BC012851 and cDNA sequence AB002155 (Adachi *et al.* 2000); sequence of 5'RACE product obtained from secondary PCR with primer 5'2-UPKIB (Figure 4.5); and EST database entries AU100584, BG401684, AL549174, BE019924, BE019779, BF034348 and BF033367. The start of translation at the ATG nucleotides is indicated by a star.

	Key
Black	100% conservation of sequence
Dark grey	80 to <100% conservation of sequence
Light grey	60 to <80% conservation of sequence
Green	Last nucleotide of primer 5'2UPKIB
Red	Expected <i>UPKIB</i> transcription start site and exon 1, at guanine (G) residue
Yellow	Confirmation of the 5'RACE poly-cytosine tail (Figure 4.5), which allowed attachment of 5'RACE primers AAP and AUAP

Transcription start
Exon1

Exon2

```

*           *           *           *
      20           40           60
AC083800 : CTGGGTGGGGTGCAGACTGCAGAGCGGGCCCTACCGTGTGCGCAGAAAGAGGAGGCGCTG : 60
BC012851 : -----ACCGTGTGCGCAGAAAGAGGAGGCGCT----- : 27
AB002155 : -----CGCAGAAAAGAGGAGGCGCT----- : 19
5' RACE  : --GGGGGGGGGGCGGGGTGGAGAGCGGGCCCTACCGTGTGCGCAGAAAGAGGAGGCGCT : 57
AU100584 : -----GCGGGCCCTACCGTGTGCGCAGAAAGAGGAGGCGCT----- : 36
BG401684 : -----GAGCGGGCCCTACCGTGTGCGCAGAAAGAGGAGGCGCTG : 41
AL549174 : -----AGNGCCCTACCGTGTGCGCAGAAAGAGGAGGCGCT----- : 35
BE019924 : -----TACGGGCCCTACCGTGTGCGCATAAAGAGGAGGCGCT----- : 37
BE019779 : -----GCGGGCCCTACCGTGTGCGCATAAAGAGGAGGCGCT----- : 36
BE747079 : -----GAGGCGCT----- : 8
BF034348 : -----GGCGCT----- : 6
BF033367 : ----- : -

```

gagggcgct

```

*           *           *           *
      80           100          120
AC083800 : GACTTATCCTACCTAAGTTGAAGCAGACCAGCAATTGTTGTGACCTACAATCTCCACAC : 120
BC012851 : ----- : -
AB002155 : ----- : -
5' RACE  : ----- : -
AU100584 : ----- : -
BG401684 : GACTTATCCTACCTAAGTTGAAGCAGACCAGCAATTGTTGTGACCTACAATCTCCACAC : 101
AL549174 : ----- : -
BE019924 : ----- : -
BE019779 : ----- : -
BE747079 : ----- : -
BF034348 : ----- : -
BF033367 : ----- : -

```

Exon3

```

*           *           *
      140          160          180
AC083800 : CCATCTTTACTCTGAGCCAAGGAAGTGTCTGTTCTTGTGCTGAGTTTCAGGTGCCTTCAG : 180
BC012851 : -----TGCTTCAG----- : 36
AB002155 : -----TGCTTCAG----- : 28
5' RACE  : -----TGCTTCAG----- : 66
AU100584 : -----TGCTTCAG----- : 45
BG401684 : CCATCTTTACTCTGAGCCAAGGAAGTGTCTGTTCTTGTGCTGAGTTTCAGGTGCCTTCAG : 161
AL549174 : -----TGCTTCAG----- : 44
BE019924 : -----TGCTTCAG----- : 46
BE019779 : -----TGCTTCAG----- : 45
BE747079 : -----TGCTTCAG----- : 17
BF034348 : -----TGCTTCAG----- : 15
BF033367 : ----- : -

```

Translation start

```

*           *           *           *
      200          220          240
AC083800 : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCC : 239
BC012851 : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCC : 95
AB002155 : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCC : 87
5' RACE  : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTNCAGGGCC : 126
AU100584 : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCC : 104
BG401684 : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCC : 220
AL549174 : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCC : 103
BE019924 : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCC : 105
BE019779 : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCC : 104
BE747079 : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGG--CC : 75
BF034348 : CTTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGTCC : 74
BF033367 : --TGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCC : 57
ctTGTGGGAAATCCCGAAGATGGCCAAAGACAACCTCAACTGTTTCGTTGCTTCCAGGGCC

```

```

*
AC083800 : TGCTGATTTTGGAAAT : 256
BC012851 : TGCTGATTTTGGAAAT : 112
AB002155 : TGCTGATTTTGGAAAT : 104
5' RACE  : TGCTGATTTTGGAAAT : 143
AU100584 : TGCTGATTTTGGAAAT : 121
BG401684 : TGCTGATTTTGGAAAT : 236
AL549174 : TGCTGATTTTGGAAAT : 120
BE019924 : TGCTGATTTTGGAAAT : 122
BE019779 : TGCTGATTTTGGAAAT : 121
BE747079 : TGCTGATTTTGGAAAT : 91
BF034348 : TGCTGATTTTGGAAAT : 91
BF033367 : TGCTGATTTTGGAAAT : 73
TGCTGATTTTGGAAAT

```

analysing a section of contig AC083800 into a promoter and transcription start site predictor program, found on the fruitfly Internet site (Section 2.3.4; Table 2.2) (Fig. 4.8). The transcription start site predicted was in agreement with the 5'RACE data.

A summary of the UPKIB exon and intron positions and proposed genomic organisation of *UPKIB* is shown in Figure 4.9. In support of this, a BLAST search with GenBank sequence AB002155, the largest UPKIB mRNA sequence available, revealed sequence matches with GenBank sequences with accessions AC083800 (version 13, circa March, 2002), AQ318241 and BG401684. Positions of exon/intron boundaries were determined with the aid of data from Mount (1982) (Fig. 4.9), which defines sequences characteristic of intron/exon junctions. In the junctions, the donor consensus sequences at the exon/intron boundaries are CAAG/GTAGAGT and the acceptor consensus sequences at the intron/exon boundaries are (TC)_nNCTAG/G. The UPKIB intron/exon and exon/intron junctions falls into these categories in all junctions, namely at the junctions at exon 4/intron 4, exon 6/intron 6, intron 6/exon 7, exon 7/intron 7 and intron 7/exon 8. The start of transcription was also identified, with the transcription start site located at base 89 971 in sequence from GenBank accession AC083800. The position of the translation start site (ATG) was located within exon 3, and the TAA translation stop sequence was located in exon 9, suggesting a 5'UTR of 63bp in length (or 176bp in the case of EST BG401684 which contains exon 2) and a 3'UTR of length 1 185bp. Thus, the first two exons and the start of exon 3 encode the 5'UTR sequence, and the remaining sequence of exon 9 encodes a long 3'UTR.

The two initial introns are of total length 12.96kb, with one large second intron (Fig. 4.9, highlighted in light blue). This explains difficulties in isolating the whole UPKIB gene by Southern hybridisation (discussed in Section 4.4). Approximate

Start	End	Score	Promoter Sequence
443	493	0.86	gcaggagggtagaaaagcacctgggtcgggtgcagactgc g gagcgggcc




Figure 4.8

Promoter and Transcription Start Site Prediction: Identifying the Transcription Start Site and Exon 1 of *UPKIB*

Further evidence of the transcription start through analysis of contig AC083800. A 1 000bp sequence from genomic fragment AC083800 was placed into a promoter and transcription start site predictor program at http://www.fruitfly.org/cgi-bin/seq_tools/promoter.pl (Table 2.2; section 2.3.4). Results are shown for the *UPKIB* gene, with score cutoff 0.80, confirming sequencing results obtained for 5'RACE (Figure 4.7). The predicted transcription start site is shown in larger font.

Figure 4.9

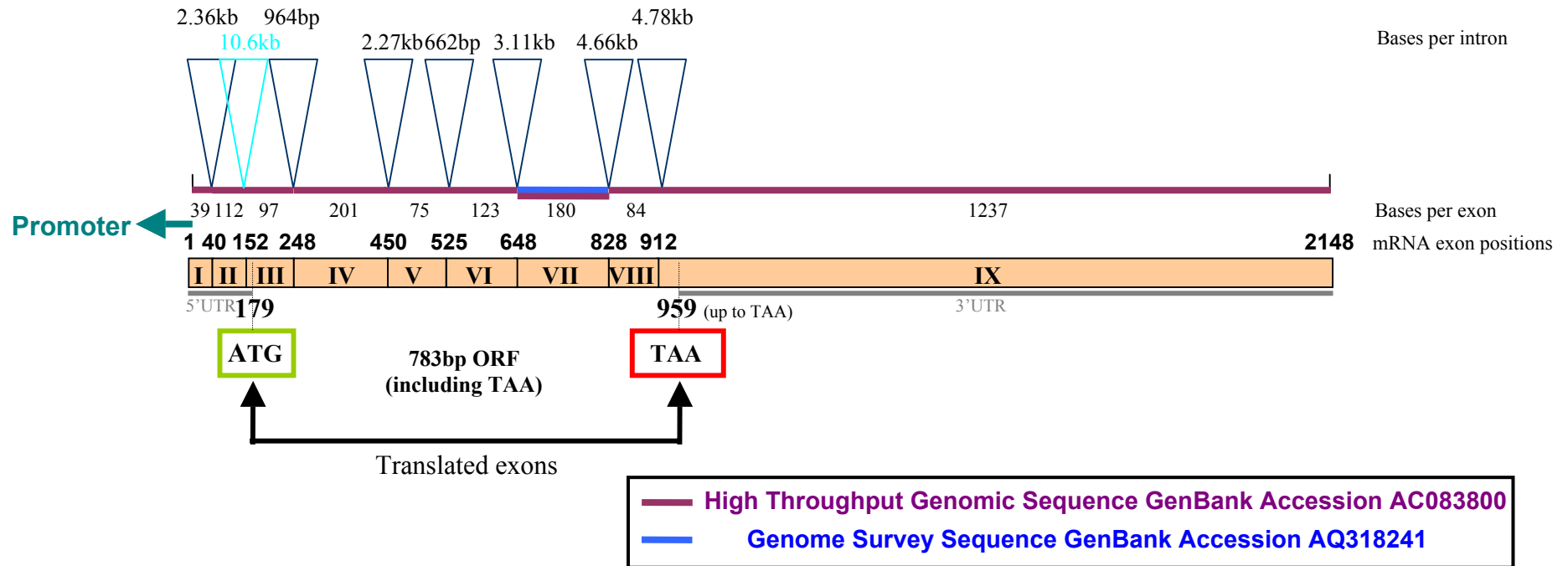
Summary of the Proposed UPKIB Genomic Structure and Intron/exon Boundary Positions

The UPKIB gene is comprised of 9 exons. Seven exons, 3 through 9, are translated into the UPKIB protein. Positions of translation start (green) and stop (red) regions are indicated.

The identity of exon 1 was determined from both 5'RACE results and through a promoter prediction program (section 4.3.3). Exon 2 is an uncommon splice variant detected in only one EST database entry, BG401684. Intron 2 is 10.6kb in size and is indicated in light blue. The genomic DNA for the UPKIB gene spans 31 533bp, as shown in the figure and table (GenBank accession AC083800 bases 89 971 to 121 524).

Key:

Top row of figure shows intron lengths derived from the LT-PCR products. Exon base sizes are indicated beneath the intronic values, with the respective mRNA positions of the exons. The UPKIB 5'UTR and 3'UTR are illustrated in grey beneath the exons. * in the table indicates a – DNA strand in GenBank sequence AQ318241, used in LT-PCR analysis (section 4.3.1.4).



Exon Position	AC083800 v13	AQ318241	Exon #; Size	Intron #; Size
1 to 39	89 971-90 009		1; 39bp	1; 2 359bp
40 to 151	92 369-92 480		2; 112bp	2; 10 604bp
152 to 248	103 085-103 181		3; 97bp	3; 964bp
248 to 449	104 146-104 346		4; 201bp	4; 2 269bp
450 to 524	106 616-106 690		5; 75bp	5; 662bp
525 to 647	107 353-107 475		6; 123bp	6; 3 114bp
648 to 827	110 590-110 769	275-96*	7; 180bp	7; 4 658bp
828 to 911	115 428-115 511		8; 84bp	8; 4 776bp
912 to 2 148	120 288-121 524		9; 1 237bp	

intron lengths deduced from LT-PCR analyses (Section 4.3.1.4) are shown in brackets (Fig. 4.9).

4.3.4 UPKIB putative transcription factor binding motifs

A 1 000bp UPKIB sequence from HTGS accession AC083800, containing both the transcription start site and upstream sequence, was analysed for the presence of putative transcription factor binding motifs using transcription factor prediction programs TESS, TFSearch and MatInspector (Section 2.3.4) (Schug & Overton, 1997; Akiyama, 1998; Quandt *et al.*, 1995). All transcription factor prediction programs gave differing motifs, although they were all based on TRANSFAC, a database of transcription factors and DNA-binding motifs (Heinemeyer *et al.*, 1998, 1999; Wingender *et al.*, 2000).

Transcription factor binding motif recognition is based on ‘core sequences’, which comprise highly conserved consecutive portions of sequence. Predicted transcription factor binding motifs are shown in Figure 4.10, but many binding sites are not shown, and several sites varied between the different prediction programs, based on different recognition principles. TATA (Goldberg/Hogness) boxes, AT-rich sequences located approximately 30bp from the nucleotide coding for the first base of the mature mRNA, are apparently absent from the UPKIB putative promoter. However, the UPKIB putative promoter contains two ‘CCAAT’ boxes with one in close proximity to the transcription start site.

CCAAT regions are known to recruit a number of transcription initiation factors including CCAAT-enhancer binding protein (C/EBP) -A, -B, and -C, and E2F (Bucher, 1990). A cap site (Bucher, 1990) is located only 11bp upstream of the CCAAT site closest to the predicted transcription start site (Fig. 4.10). Several

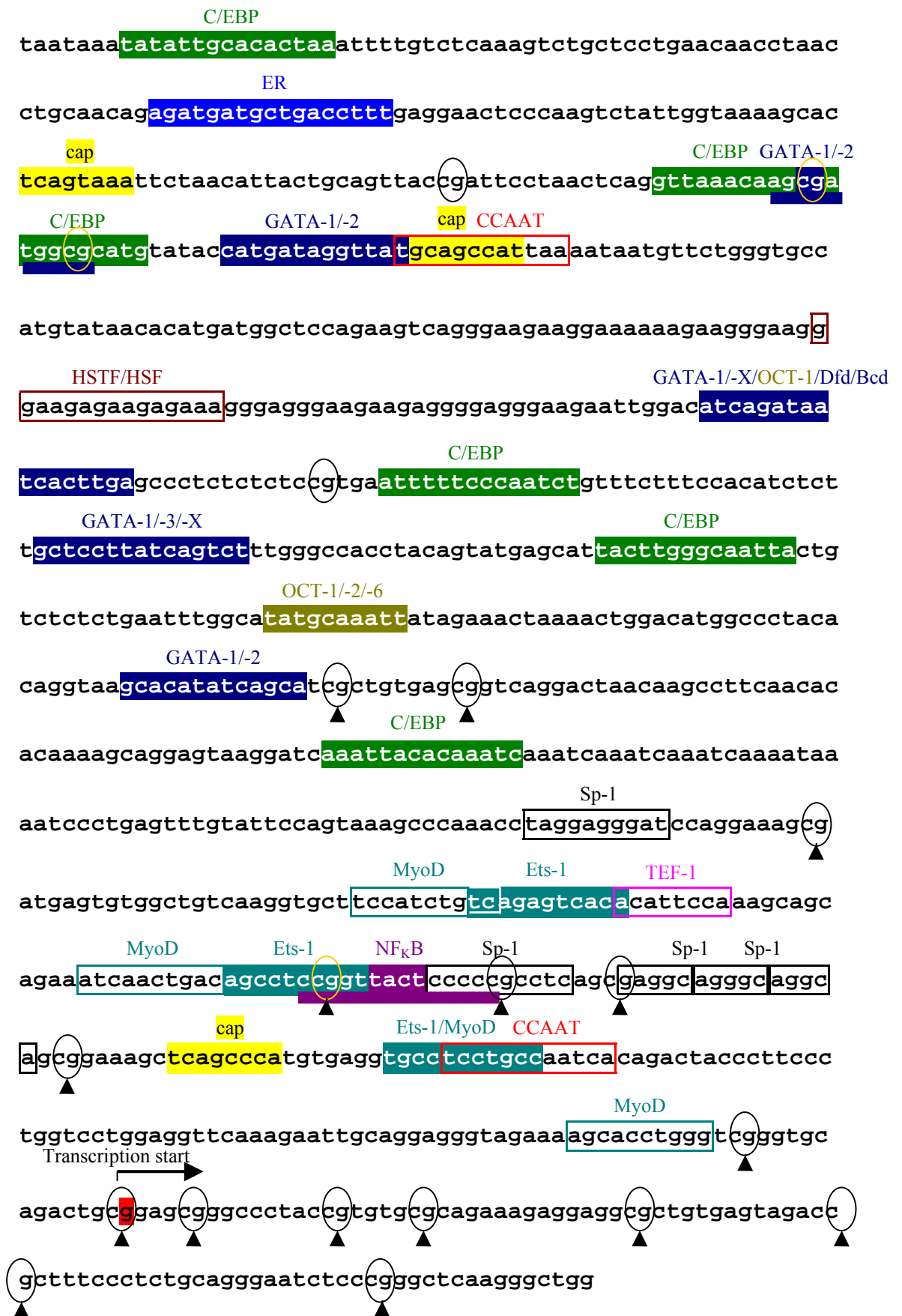
Figure 4.10

The *Uroplakin 1B* promoter

A 1000bp sequence of the *UPK1B* promoter region, from contig AC083800 (89061-90060), was placed into several transcription factor prediction programs TESS, MatInspector and TFSearch, which are all based on the TRANSFAC algorithm and matrix. The predicted transcription start site at exon 1 is highlighted in red and in bold font at guanine. Predicted transcription factor binding sites are boxed and highlighted.

Sites shown are those of C/EBP (green); GATA-1, -2, -3 and X (dark blue); ER (blue); CCAAT (red); the cap site (yellow); HSTF (brown); HSF (brown); Dfd (dark blue); Bcd (dark blue); OCT-1, -2, -6 (dark yellow); Sp-1 (black); MyoD (green); c-Ets-1 (light green); TEF-1 (pink); and NF κ B (purple).

Circled are CpG residues located in the promoter region. Those with arrows beneath circles indicate residues within a 470bp CpG island.



C/EBP sites are also present (Fig. 4.10), which may enhance activation of UPKIB mRNA through the CCAAT box (Akira *et al.*, 1990). Other putative transcription factor binding sites in the UPKIB promoter include c-Ets-1, MyoD and those of GATA family members GATA-1, -2, -3 and -X (Nye *et al.*, 1992; Merika & Orkin, 1993).

One of the most common regulatory elements is the GC box (5'-GGGCGG-3') and the related GT/CACC box, which is widely distributed in promoters, enhancers, and locus control regions (LCRs) of housekeeping as well as tissue-specific genes. Sp-1 (selective promoter factor 1) sites are found in GC rich regions and are often found at the sequence 5'-GGGCGG-3' (Thiesen & Bach, 1990). Four Sp-1 sites can be seen in the putative UPKIB promoter. CG-rich regions in promoters can also contain many CpG sites, often found in clusters termed CpG islands, which can regulate gene expression (Section 4.1.2).

4.3.5 The UPKIB CpG island

CpG islands can range in size from 0.5kb to 5kb, are usually unmethylated in housekeeping genes, have a GC content of approximately 60-70% and have a CpG:GpC ratio of at least 0.6 (Singal & Ginder, 1999). Analysis of the UPKIB putative promoter shows a CG-rich region containing 15 CpG dinucleotides within a CpG island of approximate size 0.5kb which spans the proximal promoter, exon 1 and the beginning of intron 1 (Fig. 4.10). The GC content of the UPKIB island is approximately 54% and the CpG:GpC ratio is 0.43, which falls below the expected range of a CpG island (Singal & Ginder, 1999); indicating the presence of a small and weak UPKIB CpG island. The role of methylation of this CpG island in the control of UPKIB gene expression will be examined in Chapter 5.

4.4 Discussion

Until recently, only the ORF of the UPKIB cDNA sequence had been identified (Lobban *et al.*, 1998; Yuasa *et al.*, 1998; Finch *et al.*, 1999). Adachi and colleagues (2000) subsequently published a full-length cDNA, which defined the 3'UTR and proposed a 5'UTR of length 47bp.

To analyse the transcription start site and exon 1, Southern hybridisation was initially employed, to hybridise cDNA sequence from the ORF to a PAC containing UPKIB genomic DNA. It was hoped that the cDNA would hybridise to genomic DNA containing the first exon of UPKIB, allowing delineation of the promoter through further cloning and sequencing analysis. However, subsequent data accrued from the Human Genome Project within the NCBI databases demonstrated that Southern analysis would not have been a successful strategy to determine the transcription start site (exon 1) nor the promoter, because the cDNA probe used for Southern blotting was found to begin in exon 3. A 12.96kb gap between the ATG translation start site in exon 3 and exon 1 would not have allowed hybridisation of the probe to exon 1. However, the HTGS contig AC083800 and a range of ESTs were gradually made available through the Human Genome Project and provided detailed sequence information of this region of *UPKIB*.

To verify the length of the 5'UTR sequence previously reported by Adachi *et al.* (2000), both primer extension and 5'RACE were employed. Primer extension suggested a 5'UTR of approximate length 90bp, showing that the 47bp 5'UTR reported by Adachi *et al.* (2000) was not the full length 5'UTR. Parallel sequencing with the available PAC-UPKIB DNA was also not possible, as for Southern analysis, because the gene-specific primer used for the extension was located within exon 3, a

considerable distance (12.96kb) from exon 1, and would have sequenced intronic DNA rather than the 5'UTR.

5'RACE analysis of RNA from the TCC cell line RT112 revealed the start of transcription of the UPKIB mRNA, with a 5'UTR of length 67bp. The identity of the transcription start site was determined by the 5'RACE sequence and with a promoter prediction program. An EST with GenBank accession BG401684 and a partial EST with GenBank accession AU100584 also had similar 5' ends, longer than that previously reported (Adachi *et al.*, 2000) (Fig. 4.7 & 4.8). The discrepancy of sizes between primer extension (90-100bp) and 5'RACE (67bp) results cannot clearly be explained; but could be attributed in part to the specificity of the reverse transcription primer toward all mRNA species. However, the primer sequences used in both studies were analysed using BLAST, which showed their specificity to their respective UPKIB sequences. 5'RACE employs three gene specific primers to amplify a specific 5'UTR, whereas primer extension utilises only a single primer and one round of amplification and may be less specific. 5'RACE gene-specific primers in the first round PCR amplification show several products (Fig. 4.6B), so a second round of PCR was subsequently performed to improve specificity, as outlined in the manufacturer's instructions.

In summary, the sequence data has allowed the characterisation of the UPKIB gene by defining a putative promoter, exon/intron boundaries and has shown a possible exon 2 splice variant from EST BG401684 (Fig. 4.7 & 4.9). The UPKIB promoter has no TATA box but it has a CCAAT box and 4 putative Sp-1 sites. The putative promoter region also comprises a weak 0.5kb CpG island containing 15 CpG dinucleotides in a GC-rich portion spanning the promoter, exon 1 and intron 1.

A recent article by Olsburgh *et al.* (2002) reported the identification of the UPKIB promoter by RNase-protection assays and CapSelect 5'RACE using the TCC-derived cell line VM-Cub3. However, in our laboratory, expression of UPKIB mRNA was not detected in VM-Cub3 cells (Section 5.3.3). Olsburgh and colleagues (2002) reported a 5'UTR of length 116bp which contrasts with results presented herein showing a 67bp UPKIB 5'UTR in RT112 cells. A possible explanation for the difference in the expression of UPKIB is that the VM-Cub3 cells have developed differing expression characteristics whilst in culture in the two laboratories. However, the differences in the two 5'UTR lengths would seem to be due to the use of differing techniques.

CapSelect, a technique derived from Schmidt & Mueller (1999), relies on reverse transcribing poly(A+) mRNA and utilising the mRNA cap (7-methylG-5'-ppp-5'-G). The cDNA is subsequently tailed by CRTC (controlled ribonucleotide tailing of cDNA) with rATP, ligated to an adaptor sequence, PCR amplified and sequenced. The 5'RACE described in this chapter is a similar technique to CapSelect, but used total RNA and relied on the assumption that UPKIB mRNA transcripts retain a cap. Homopolymeric tailing with dCTP was used instead of CRTC to extend the 3'end of the cDNA and ligation of an adaptor sequence was not required, due to the specificity of the AAP primer for the poly-C tail.

A major difference between the techniques in these studies was the use of the reverse transcription primer. UPKIB-specific primers TM3 and 5'2-UPKIB were used in the present investigation for the reverse transcription and Olsburgh *et al.* (2002) used an anchored oligo (dT) primer initially providing less specificity for the UPKIB product. In hindsight, CapSelect appears to be a more specific technique for selecting transcripts containing the complete 5' UTR sequence, as our 5'RACE

technique did not specifically select for 5' capped transcripts. This, the 5'RACE used in the present study may have produced a shorter 5'UTR sequence than that shown by Olsburgh *et al.* (2002).

Another finding from the Olsburgh *et al.* (2002) study involved the detection of multiple transcription start sites, using RNase-protection assays. This entailed the production of antisense ³²P-radiolabelled UPKIB probes, which were hybridised against total RNA from RT112 and VM-Cub3 cell lines and yeast. After RNase digestion, yeast showed no hybridisation to the UPKIB probe, suggesting the absence of UPKIB mRNA expression in yeast cells. Several bands were protected from RNase digestion in both RT112 and VM-Cub3 cell lines, ranging in size from 130bp to 185bp, suggesting multiple transcription initiation sites within their 88bp exon 1, including one with the sequence 5'-GCAGACT-3'. This sequence, suggested by Olsburgh *et al.* (2002) as the site of alternate transcription, is located only 9bp upstream from the expected transcription start site identified in this investigation.

In Olsburgh's study, 8 UPKIB exons were proposed, with exon 1 of size 88bp and an exon 2 of size 96bp, with translation commencing at the ATG sequence 28bp from the start of exon 2, and 116bp from the beginning of exon 1. There was no mention of the possibility of an alternately spliced exon 2, as occurs in EST BG401684 (derived from kidney full-length enriched cDNA library (NIH_MGC_75), constructed by Clontech Laboratories (Palo Alto, CA)). Other exons reported by Olsburgh *et al.* (2002) were of length 200bp, 75bp, 123bp, 180bp, 87bp and 1 236bp for each of the other 6 exons, based on accession AC083800. This is in contrast from findings presented here, where 9 exons were detected, of sizes 39bp, 112bp, 97bp, 201bp, 75bp, 123bp, 180bp, 84bp and 1 237bp for exons 1 to 9. The minor discrepancies in sizes may be based on the techniques of exon/intron

boundary analysis. The positions of these exon/intron boundaries were clarified through the analysis of consensus donor and acceptor sequences reported by Mount (1982) (Section 4.3.3).

The genomic structures of a number of other tetraspanins have been reported. Although there appears to be similarity in exon sizes between some members including CD53 and CD63, this does not imply that these genes are similarly regulated nor that they have similar functions (Wright *et al.*, 1993). The promoters vary between members: tetraspanins *CD81* and *CD63* do not have TATA boxes, similar to *UPKIB*, however *CD82* does have a TATA box at site -25 (review by Nagira *et al.*, 1994). These tetraspanin promoters are often GC rich and Sp-1 binding sites have been found in a number of tetraspanin promoters including CD63, CD81, CD82 and CD9 (Nagira *et al.*, 1994; Le Naour *et al.*, 1996). The differences between tetraspanin transcription factor binding sites, as detailed above, indicate the complexity of regulation of the various tetraspanin family members. This was confirmed with the extensive distribution of the tetraspanins in different tissues (Section 1.5.1).

It was initially hypothesised that TGF β 1 may induce transcription of human *UPKIB*, because expression of the mink homologue TI-1 mRNA was induced in lung epithelial cells by TGF β 1 (Section 1.2) (Kallin *et al.*, 1991). However, the programs used in this study to detect transcription factor motifs did not detect a TGF β 1 regulatory element within the human *UPKIB* putative promoter. However, based on sequence analyses by Kerr *et al.* (1990), Olsburgh *et al.* (2002) found close homology to a TGF β 1/fos binding element, TGF β 1 inhibitory element (TIE) with sequence 5'-GAATTGGACA-3', upstream of the transcription start site (-589 to -579). This suggests that either TGF β 1 or TGF β 1-related factors may be interacting

with TIE-like motifs in the UPKIB promoter. It is still uncertain whether TGF β 1 controls human UPKIB mRNA expression directly through enhancer elements such as TIE, or whether TGF β 1 interacts with other factors, which can bind to *cis*-acting elements along the UPKIB promoter, such as Sp-1. Other studies would need to be performed to examine the role of TGF β 1 in human UPKIB mRNA expression.

To study transcription factor binding motifs in the UPKIB promoter, and to show their functionality in controlling UPKIB mRNA expression, techniques including luciferase reporter assays could be employed. To investigate the UPKIB region with this technique, Olsburgh *et al.* (2002) analysed the putative promoter for transcriptional activity with a luciferase reporter construct containing up to 2.6kb of the UPKIB promoter, transfected into HT29 or HDF cell lines not normally expressing UPKIB mRNA. After transfection, Olsburgh *et al.* (2002) detected activation of transcription in a 235bp region including and upstream of their reported 88bp exon 1. In our study, this same region (-196 to +39) was shown to include putative transcription factor binding motifs for TEF-1, MyoD (x3), c-Ets-1 (x2), NF κ B, and Sp-1 (x3).

Olsburgh *et al.* (2002) also analysed a sequence of UPKIB including this region and 153bp further upstream (-349 to +39). This region, including putative transcription factor binding motifs C/EBP, Sp-1 and MyoD, showed reduced UPKIB expression, indicating the roles of possible of inhibitory factors binding in this 153bp sequence. Further analysis with band shift assays and supershift assays at these and other binding motifs may reveal transcription factors directly binding to the UPKIB promoter, allowing analysis of subtle variations in transcriptional activation, particularly in cultured cells.

Chapter 5

The Role of Promoter Methylation in Regulating Expression of the UPKIB Gene

5.1 Introduction

A variety of factors contribute to the regulation of gene expression, including the binding of upstream transcription factors such as activators and repressors, which often act in synergy to regulate expression. An important mechanism in the regulation of gene expression is DNA methylation, an epigenetic mechanism in which DNA methyl transferases act in concert with other proteins, including methylation-dependent binding proteins and histone deacetylases (Section 4.1.2).

Aberrant methylation patterns of CpG islands are hallmarks of many cancers and are characterised by the inactivation of regulatory genes such as DNA-repair and tumour suppressor genes. As presented and discussed in Sections 4.3.5 and 4.4, the UPKIB gene contains a 5' CpG island of approximate size 0.5kb spanning the proximal promoter, exon 1 and the beginning of intron 1. The island has a GC content of approximately 54%, a CpG:GpC ratio of 0.43 and contains 15 CpG dinucleotides. The aims of this chapter were to study methylation as a potential mechanism of regulation of UPKIB mRNA expression. Techniques to examine methylation status include bisulfite-modification of genomic DNA, SSCP, restriction enzyme analysis and direct sequencing. 5-Aza-2'-deoxycytidine (5-Aza-CdR) assays were performed to determine if a methylated UPKIB promoter could be induced to re-express the gene. This chapter investigates 11 CpG dinucleotides located in the UPKIB CpG island to determine if methylation of these sites plays a role in control of regulation of mRNA expression.

5.1.1 Imprinting in development and relevance to cancer

De novo methylation is involved in normal embryonic development. This form of genomic imprinting (gametic or parental imprinting) occurs in autosomal genes and also silences X-linked genes in female mammalian X-chromosome inactivation. Imprinting is a heritable epigenetic mechanism occurring in gametogenesis, where the cellular environment changes, allowing factors to bind to and to interact with the DNA, leading to methylation of cytosine residues at CpG sites. The imprinting can be parent-of-origin specific (maternal or paternal) or biallelic. Genomic imprinting has been identified in genes including p73 (Kaghad *et al.*, 1997; Mai *et al.*, 1998) on chromosome band 1p36 and the tetraspanin member CD81 (TAPA-1) (Gabriel *et al.*, 1998; Reid *et al.*, 1997; Morison & Reeve, 1998) on chromosome band 11p15.5. This imprinting is then maintained during somatic cell division (Heby, 1995).

The process by which these genes are marked or inactivated is complex and not well understood. Abnormal imprinting patterns in the child or adult generate genetic diseases affecting foetal, placental and adult development (Reik & Surani, 1997; Falls *et al.*, 1999). The tetraspanin family member *CD81* is imprinted (discussed above) and in adults, methylation of the tetraspanin *KAI-1* (*CD82*) has been studied in cancer. However, the role of methylation in the regulation of *KAI-1* appears controversial, with insufficient evidence to suggest methylation silences *KAI-1*, as discussed in the literature review (Section 1.5.2). Similar to *CD81*, *Uroplakin IB* may be methylated at promoter or other CpG sites, leading to silencing of the gene and cancer progression.

5.1.2 Methylation changes in cancer

Methylation of cell cycle regulatory and tumour suppressor genes has been well studied and has been found to facilitate cancer progression. Alterations in methylation patterns are associated with many cancers and diseases, and this section will focus on methylation as a possible mechanism for down-regulation of expression of UPKIB mRNA in TCCs (transitional cell carcinomas).

5.1.2.1 *An overview of CpG methylation in cancer*

Aberrant *de novo* methylation in cancer was first detected in the Rb (retinoblastoma) gene in sporadic retinoblastomas (Greger *et al.*, 1989). Tumour-specific changes in methylation status have also subsequently been reported in other genes including E-cadherin (Herman *et al.*, 1996) and the cell cycle regulator p16 (CDKN2/INK4a/MTS1) (Herman *et al.*, 1995, 1996; Merlo *et al.*, 1995; Gonzalez-Zulueta *et al.*, 1995) (Section 1.4.3.2). Some of these genes, including p16, also undergo loss of heterozygosity (LOH) or mutations in one allele in tumours, increasing risk of cancer development with the second allele possibly being inactivated through methylation. This is based on Knudson's 2-hit hypothesis in which both alleles of the tumour suppressor gene Retinoblastoma (*Rb*) are inactivated during the development of cancers (Knudson *et al.*, 1973).

Methylation appears to be an important component in the silencing of *p16*, and methylation of the p16 CpG island may relate to tumour grade and stage. p16 can inhibit G1-phase cyclin-D/CDK (cyclin-dependent kinase) complexes and loss of *p16* may lead to uncontrolled cyclin/CDK activation and subsequent loss of G1/S checkpoint control and cancer proliferation (discussed in Section 1.4.3.2). Herman and colleagues (1995) reported methylation of the p16 CpG island in a range of

primary tumours and cell lines including colon and breast. The p16 5' CpG island was methylated in 33% of breast cancer cell lines, 60% of prostate cancer cell lines, 92% of colon cancer cell lines, 31% of primary breast tumours and 40% of primary colon tumours. This finding supported an important role for methylation in the regulation of this tumour suppressor gene in cancer. This study also indicated that methylation of the p16 CpG island was more prevalent in cancer cell lines than in primary tumours. Merlo *et al.* (1995) also demonstrated methylation of the p16 CpG island in a number of other cancers including lung cancers, gliomas, and head-and-neck squamous cell carcinomas.

5.1.2.2 *Aberrant methylation in bladder cancers*

Methylation has been investigated extensively in cancers of the bladder, concentrating on tumour suppressor genes, including p16 and p53. A recent study by Bornman *et al.* (2001) has shown that in normal urothelial tissue, increased levels of methylation of *E-cadherin* may occur normally with aging, and may not be specific to bladder cancer. It was also suggested by Bornman and colleagues (2001) that increased DNA methylation in elderly populations may put them at greater risk of developing urothelial abnormalities due to the increased chance of deregulation of genes such as E-cadherin. This correlation between advancing age and increased methylation and deregulation of genes, including p16 and the mismatch repair gene hMLH1, has similarly been observed in colorectal carcinomas (Toyota *et al.*, 1999; Toyota & Issa, 1999). *p16* was also analysed in transitional cell carcinoma (TCC) samples by Gonzalez-Zulueta *et al.* (1995) and *de novo* methylation was observed in 67% of both primary bladder cancers and TCC cell lines.

Since tetraspanin family members frequently contain a high GC content in their promoter regions, methylation may silence expression of these genes, due to the high

5' region GC content. Nagira *et al.* (1994) analysed the promoter of the mouse tetraspanin family member *C33 (KAI-1/CD82)* and reported a 760bp island with a G+C content of 64% from -650 to +110 and a CpG:GpC ratio of 0.68. The promoter matched CpG/HTF (*HpaII* Tiny Fragment) island criteria (Bird, 1986), suggesting methylation as a possible regulator of mouse *C33* expression. The human *KAI-1* homologue has been studied in TCCs for hypermethylation of the promoter region. However, methylation does not appear to be an important factor responsible for deregulation of the gene or to correlate with progression of tumours and cancer cell lines (Jackson *et al.*, 2000b; Sekita *et al.*, 2001; Uzawa *et al.*, 2002) (Section 1.5.2). Thus, this chapter will explore promoter methylation as a possible mechanism for *UPKIB* regulation.

The member of the tetraspanin family examined in the current study, uroplakin IB, has been proposed as a marker for tumour progression, with reduced mRNA expression indicative of more invasive tumours (Finch *et al.*, 1999). However, no methylation studies for this gene have yet been reported in the literature. The 5' region of the uroplakin IB gene contains a weak CpG island of approximately 500bp in length, which could be methylated in TCCs, thus silencing the gene.

5.1.3 Techniques for analysis of methylation status

There are two main methods for analysing CpG island methylation, with the first using Southern hybridisation and restriction enzyme analysis, and the other modifying genomic DNA with sodium bisulfite. Studies using these methodologies will be discussed below.

5.1.3.1 Southern hybridisation and restriction enzyme analysis to detect CpG methylation

Singer and colleagues (1979) were the first to publish a technique incorporating both Southern hybridisation and methylation-sensitive restriction analysis, with enzymes *MspI* and *HpaII*, to detect methylated CpG sites in mouse promoter regions at the DNA recognition sequence CCGG. The methylation-sensitive restriction enzyme *HpaII* does not cleave this sequence when the internal cytosine residue is methylated. However, the restriction enzyme *MspI*, an isoschizomer of *HpaII*, cleaves the sequence irrespective of the methylation status at the recognition site. Disadvantages of this technique are that large quantities of high molecular weight DNA are required; subtle differences in methylation of heterogeneous samples cannot be detected nor differentiated; and not all CpG sites can be analysed, since analysis is dependent on the restriction enzymes employed. Due to these limitations, this technique has been superseded by bisulfite techniques.

5.1.3.2 Bisulfite modification and PCR-based strategies

Frommer *et al.* (1992) developed a technique to modify genomic DNA with sodium bisulfite, which converts all unmethylated cytosines throughout the genome to uracil, but methylated cytosines remain as cytosines (5-MeC). In a subsequent PCR amplification, uracil is amplified as thymidine. After this initial step, Frommer *et al.* (1992) used strand-specific primers for PCR that specifically amplified the bisulfite-modified promoter region of the human kininogen gene, which was subsequently cloned and sequenced.

The principles of Frommer *et al.* (1992) were adapted by Herman *et al.* (1996), in a technique known as MSP (methylation-sensitive PCR). MSP involves sodium bisulfite modification of DNA and subsequent PCR with primers spanning a

particular CpG island. MSP primers contain 1-3 CpG dinucleotides at their 3' ends and the numbers of cytosines in the primers determine the specificity for unmethylated or methylated alleles (Herman & Baylin, 1994). In this method, the primers do not amplify unmodified genomic DNA and the only CpG dinucleotides analysed are those located in the primer pairs. MSP is a sensitive technique in the analysis of methylation not requiring direct sequencing with amplified products typically 80-250bp in length. However, limitations are that only a few CpG dinucleotides can be assayed in each primer pair and any CpG sites located between the primer pairs are not analysed.

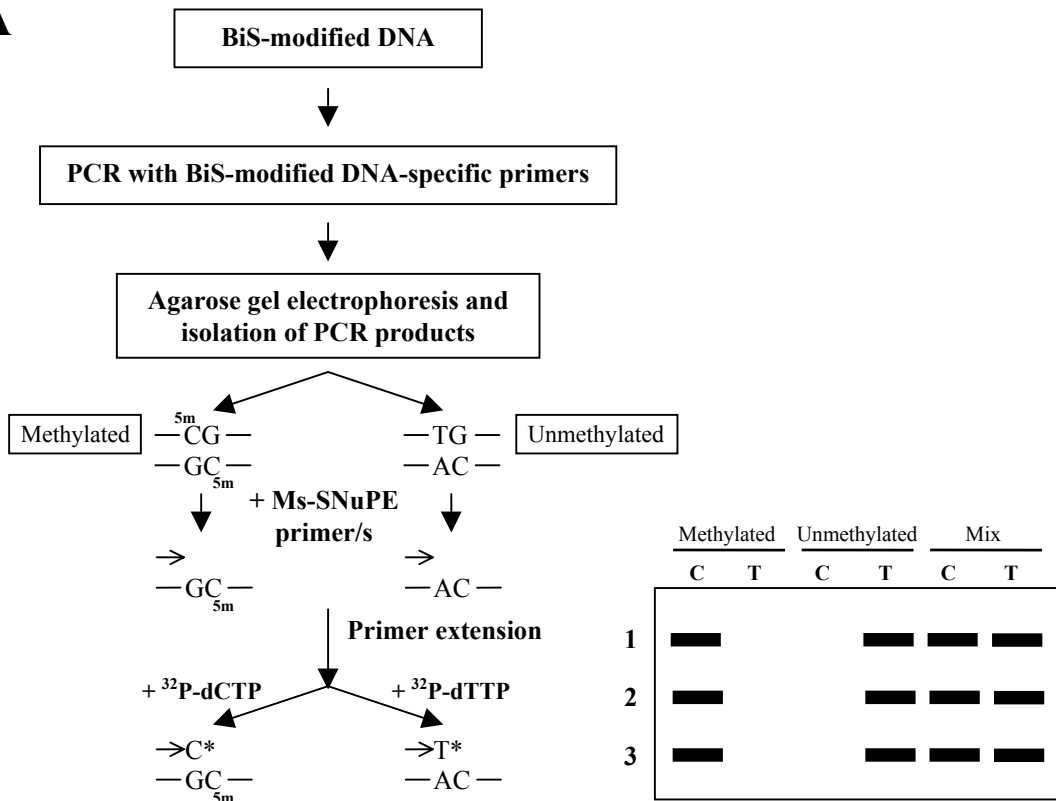
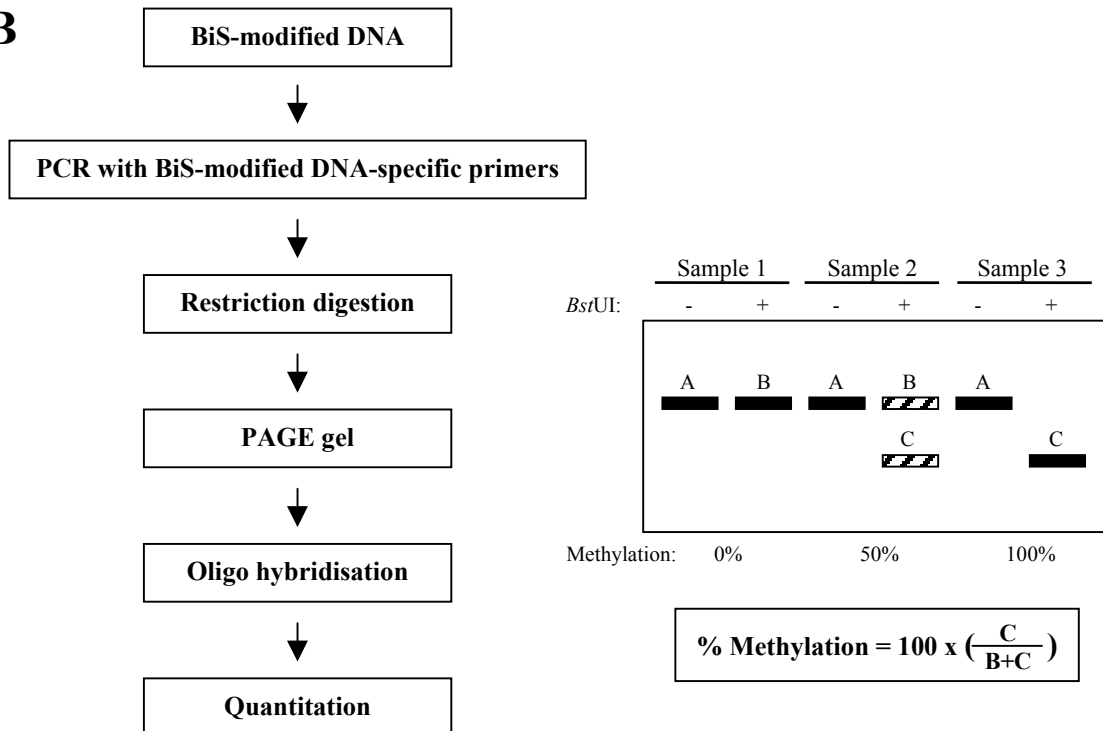
Gonzalzo and Jones (1997) and Xiong and Laird (1997) reported two quantitative techniques, termed Ms-SNuPE (methylation-sensitive single nucleotide primer extension) and COBRA (combined bisulfite restriction analysis) respectively. Ms-SNuPE (Fig. 5.1A) involves bisulfite treatment of genomic DNA, PCR with primers amplifying CpG regions and single-nucleotide primer extension with Ms-SNuPE primers located at cytosines in the top strand of amplified DNA. The primer extension incorporates radiolabelled dNTPs, allowing visualisation of products by exposure to autoradiographic film or quantitation using a phosphorimager. COBRA (Fig. 5.1B) relies on the use of restriction enzymes to digest methylated or unmethylated CpG residues in PCR-amplified bisulfite-treated DNA and involved probing the fragments on a membrane with 5'-end labelled probes to quantitate the degree of methylation. The COBRA technique uses restriction enzymes, so it can only be used to analyse certain CpG sites and there may be problems associated with partial restriction enzyme digestion. A further disadvantage of these two techniques is that they are very labour-intensive and thus not suitable for screening large numbers of samples for CpG methylation.

Figure 5.1

Ms-SNuPE and COBRA: Two Techniques to Determine Strand-specific Methylation Status

A. Ms-SNuPE: genomic DNA is modified with sodium bisulfite (BiS), amplified by PCR and the products isolated from agarose gels after electrophoresis. Products are then incubated with Ms-SNuPE primer/s, Taq polymerase, buffer, [³²P]dCTP or [³²P]dTTP, and other dNTPs for primer extension. Ms-SNuPE primer extension primer(s) anneal to bisulfite-treated PCR products and terminate immediately 5' of the single nucleotide to be assayed. Products are then run on a denaturing polyacrylamide gel and visualised with autoradiography. Quantitation of unmethylated cytosines (thymines) to methylated cytosines (cytosines) at the original CpG sites can subsequently be determined by phosphoimager analysis. Abbreviations: C, cytosine; T, thymine. **B. COBRA:** genomic DNA is modified with sodium bisulfite (BiS), amplified by PCR, restriction enzyme digested and quantitated. A restriction enzyme, such as *Bst*UI, is used to digest DNA containing a recognition sequence in the unmodified genomic DNA and does not cut methylated sites. The digested PCR products are finally analysed on a polyacrylamide gel, DNA is transferred to a charged membrane by electroblotting and the membrane is hybridised with 5' -end-labelled oligonucleotides followed by quantitation with a phosphoimager.

Figures adapted from Gonzalgo & Jones, 1997 and Xiong & Laird, 1997.

A**B**

Recent techniques involving bisulfite conversion for screening patient samples for methylation include MS-SSCA (methylation-sensitive single-strand conformation analysis) and BiPS (bisulfite-PCR-SSCP (single-strand conformation polymorphism)) developed by Bianco *et al.* (1999) and Maekawa *et al.* (1999) respectively. These techniques are similar and both rely on PCR amplification of bisulfite-treated DNA and subsequent SSCP. Primers are designed to amplify bisulfite-modified DNA containing either methylated or unmethylated sequences over a 180-300bp CpG-rich region. The resulting PCR products can either be sequenced or analysed by SSCP, where products are denatured and run on a polyacrylamide gel. In MS-SSCA, aberrant bands can subsequently be stabbed in the gel, amplified using PCR and sequenced to determine the locations of the methylated cytosines. The main difference between MS-SSCA and BiPS techniques is the staining procedure of the polyacrylamide gel, with SYBR-Gold staining in MS-SSCA and silver staining in BiPS respectively. MS-SSCA and BiPS have been used to screen genes for methylation including BRCA1 (breast cancer 1) in breast cancer (Bianco *et al.*, 1999); and MLH1 (mutL homolog 1), HIC1 (hypermethylated in cancer 1), TERT (telomerase reverse transcriptase) and p16 in colorectal cancer (Maekawa *et al.*, 1999, Maekawa *et al.*, 2001; Nomoto *et al.*, 2002).

5.1.3.3 5-Aza-2'-deoxycytidine (5-Aza-CdR) experiments

Methylated genes showing no detectable mRNA expression can be induced to re-express their mRNA in carcinoma-derived cell lines by the addition of the drugs 5-Aza-2'-cytidine (5-Aza-CR) or 5-Aza-2'-deoxycytidine (5-Aza-CdR) (Jones & Taylor, 1980; Review by Karpf & Jones, 2002). These chemicals are demethylating and antileukaemic agents whose exact method of demethylation is not completely understood (Jones *et al.*, 1983; Jones, 1985a, 1985b; Pinto & Zagonel, 1993; Review

by Karpf & Jones, 2002). It has been reported that 5-Aza-CdR inhibits DNA methyltransferase DNMT1 (Section 4.1.2) in whole cells (Ferguson *et al.*, 1997). In this regard, 5-Aza-CdR-substituted DNA forms complexes (adducts) with DNMT1 through covalent bonding and these adducts thus prevent methylation and allow gene expression (Ferguson *et al.*, 1997; Review by Karpf & Jones, 2002). In this way, newly replicated DNA can not be remethylated, leading to genomic hypomethylation (Bender *et al.*, 1998).

This demethylation of genes in cancer cells by the inhibition of DNA methyltransferases has established these chemicals as possible therapeutic agents in advanced cancers. 5-Aza-CdR has already been investigated as a potential treatment of acute myeloid leukemias and myelodysplastic syndromes (Pinto & Zagonel, 1993; Pinto *et al.*, 1984), in sickle cell anaemia (Charache *et al.*, 1983) and in non-small cell lung cancer (NSCLC) (Momparler & Ayoub, 2001). However, the use of 5-Aza-CR and 5-Aza-CdR in clinical studies is of doubtful benefit to patients, as these drugs are also cytotoxic to normal cells (Review by Karpf & Jones, 2002). Cytotoxicity of 5-Aza-CR and 5-Aza-CdR is well recognised in numerous reports including Momparler and Goodman (1977), Murakami *et al.* (1995) and Glazer and Knode (1984). Momparler and Goodman (1977) demonstrated that 5-Aza-CdR was a very potent cytotoxic agent to A(T1)C1-3 hamster fibrosarcoma cells *in vitro*. Murakami *et al.* (1995) studied the effects of various concentrations of 5-Aza-CR on induction of apoptosis in promyelocytic leukaemic HL-60 cells. They showed that both the cell cycle position and the 5-Aza-CR dosage affected the degree of apoptosis, with increased cell death at G1 and S phases from increased dosage. Glazer and Knode (1984) treated the colon carcinoma cell line HT-29 with a range of cytidine analogues including 5-Aza-CR and 5-Aza-CdR and detected marked

cytotoxicity with a maximum of 30% of cells still viable after 2 hours of drug administration. These studies have shown that potential cytotoxicity that may affect the safety of these drugs in humans.

In this chapter, the methylation status of *UPK1B* in cell lines and patient samples will be investigated. The methods used are summarised in a flow diagram (Fig. 5.2).

5.2 Methods

5.2.1 Histological examination of tissue samples

5.2.1.1 Normal and tumour tissue samples

Fresh or frozen normal urothelium and TCC samples were obtained from The Queen Elizabeth Hospital through the assistance of Mr John Miller (Department of Urology) and Dr Prue A. Cowled (Department of Surgery). Tissues analysed in the current study were 2 normal urothelial samples from the ureter or renal pelvis and 8 TCC specimens of varying stages and grades (Table 5.1).

Other samples analysed consisted of 7 normal peripheral blood samples from healthy volunteers, of which 6 were obtained from Dr S. Stephenson (Department of Haematology/Oncology), and 2 samples of normal colonic epithelium, which were obtained from Dr J. Hardingham (Department of Haematology/Oncology). Fresh tissue samples were snap frozen in liquid nitrogen and stored at -80°C until use. All samples were collected with informed consent with the protocol approved by the Human Ethics Committee at The Queen Elizabeth Hospital.

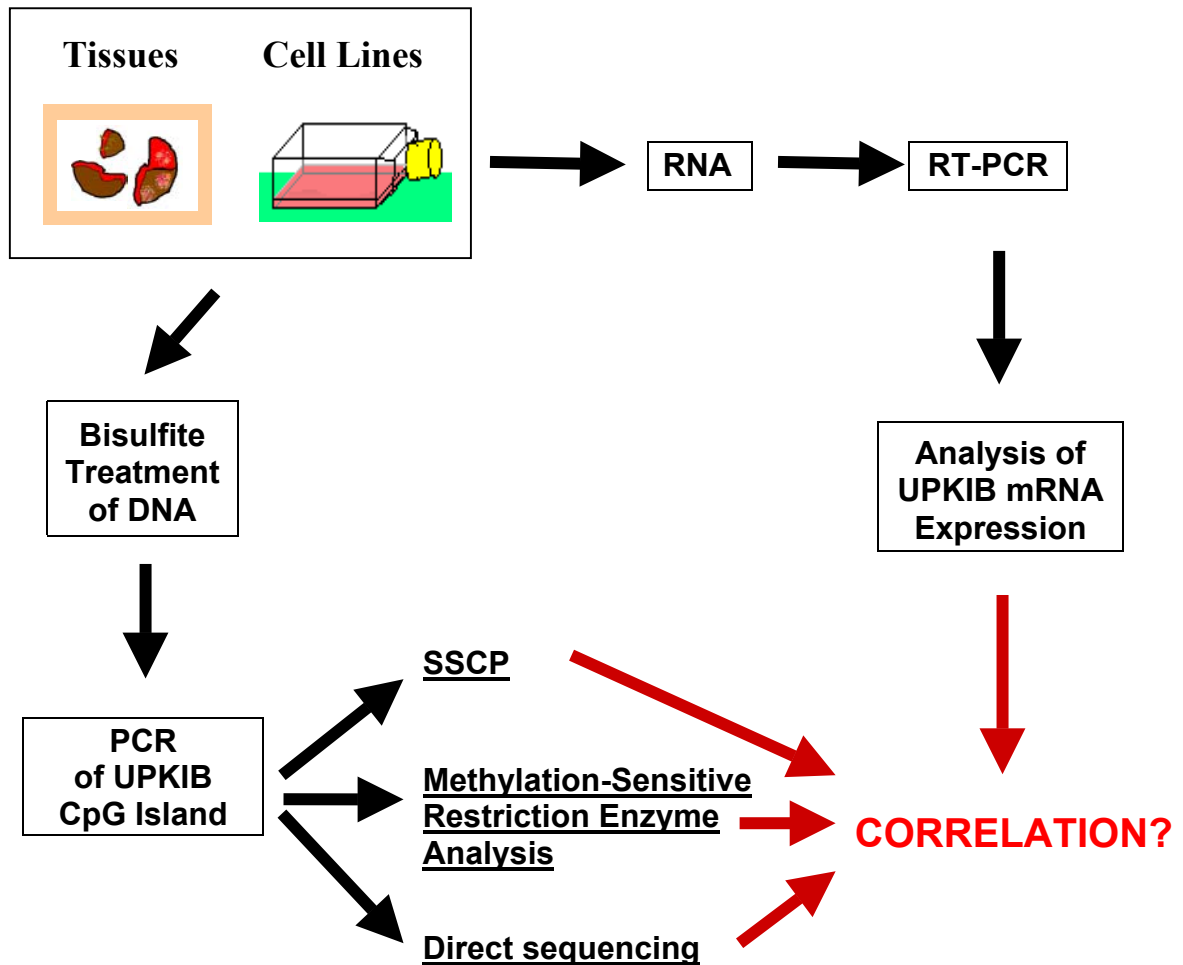


Figure 5.2

Flow Diagram of Methylation Studies

The strategy illustrated was employed in this study to determine if a correlation existed between promoter methylation and *UPK1B* expression in transitional cell carcinomas. Both cell lines and tissue samples were analysed for methylation by extracting DNA, treating the DNA with sodium bisulfite to convert unmethylated cytosines to uracils, using PCR of the Uroplakin IB 5' CpG island and analysing methylation status. RNA was also extracted and expression of UPK1B mRNA analysed by RT-PCR to correlate expression of UPK1B with DNA methylation status.

Table 5.1**Patient Samples Used in Methylation Analyses****A. Normal urothelial and colonic epithelial patient samples**

Patient	Gender	Date of Birth	Age (Years)	Tissue
SC-N	M	18/3/1936	74	Renal Pelvis
TE-N	F	27/2/1924	63	Ureter
COL-1	F	31/12/1919	81	Colonic epithelium
COL-2	M	21/12/1921	79	Colonic epithelium

B. TCC patient samples

Patient	Gender	Date of Birth	Age (Years)	Tissue	Grade	Stage	Type
BR-T1	M	3/8/1930	69	Bladder	GII	-	Papillary
HA-T1	M	1/12/1929	70	Bladder	GIII	-	Papillary
HA-T2	F	4/1/1921	77	Bladder	GI/II	-	Papillary
HA-T3	F	15/10/1924	76	Bladder	GII	T3	-
PE-T2	F	8/6/1932	67	Ureter	GI/II	Tis	Papillary
RA-T1	M	3/7/1921	78	Bladder	-	Tis	-
RO-T1	F	19/4/1927	72	Bladder	GIII	T3/4	-
WA-T1	F	8/6/1924	75	Bladder	GIII	T3	-

5.2.1.2 Poly-L-Lysine coating of slides for histological analyses

A 1 in 10 solution of poly-L-lysine was prepared in MilliQ-H₂O. Clean glass slides were placed in a rack, immersed in the poly-L-lysine solution for 5 minutes at room temperature, drained and left overnight to dry. This treatment was necessary for adequate adhesion of OCT sections to glass slides for histological examination.

5.2.1.3 OCT sectioning and Haematoxylin & Eosin (H&E) staining of tissue samples

Frozen tissue samples were embedded in OCT (optical cutting temperature)

compound and 10µm sections were cut using a cryostat (Microm, Section 2.3.4), set at -25°C. Sections were attached to poly-L-lysine coated slides (Section 5.2.1.2), fixed in acetone at -20°C for 10 minutes, and then dried at room temperature for 10 minutes.

Slides were stained using the following protocol: 2 minutes in haematoxylin, 2 minutes under running tap water, 2 minutes in eosin, a brief rinse in 95% ethanol, 1 minute in 100% ethanol, 1 minute in Sub-X solution and 5 minutes in fresh Sub-X solution. Sections were mounted with Sub-X mountant under 25 x 40mm coverslips, and examined under an Olympus microscope. Images were captured using Spot Advanced (v3.0) software under 40x or 200x total magnification.

5.2.2 RT-PCR of UPKIB, PBGD and GAPdH mRNA

5.2.2.1 Isolation of RNA from bladder carcinoma-derived cell lines and tissue samples

Cells were grown to confluence in 75cm³ tissue culture flasks. RNA was extracted from cell lines (Section 2.3.2.2) and from samples of normal colonic epithelium, normal urothelium and from TCC samples (Table 5.1). Peripheral blood lymphocytes (PBL) were isolated from whole blood using Ficoll-Paque (Pharmacia), according to the manufacturer's instructions and RNA was extracted as for other samples. All normal tissues and seven PBL samples were from individual with no reported incidence of TCC.

5.2.2.2 Reverse transcription and PCR

Reverse transcription was performed using oligo-(dT)₁₅ primer (Promega) and Superscript II reverse transcriptase (Life Technologies), as outlined in Section

2.3.2.4. PCR was subsequently employed to detect UPKIB mRNA and mRNA of housekeeping genes PBGD (Chretien *et al.*, 1988) and GAPdH, with primers F1/R1 which span the UPKIB open reading frame and housekeeping primers PBGD-5'/PBGD-3' or GAPdH-5'/GAPdH-3' respectively (Section 2.3.3.3).

5.2.3 Analysis of methylation status of the UPKIB CpG island

5.2.3.1 Maintenance and extraction of DNA from TCC and SCC cell lines

Cell lines were propagated in tissue culture as previously described (Section 2.3.1). DNA was extracted from the cell lines using the TES high salt DNA precipitation and extraction method (Section 2.3.3.1), the DNA was resuspended in 50µl sterile MilliQ-H₂O and then stored at 4°C. The concentration of DNA was estimated by running 2µl of the DNA the extracted DNA on a 1.0% agarose gel next to 500ng of SPP1/EcoRI molecular weight marker (GeneWorks).

5.2.3.2 Extraction of DNA from tissue samples

H&E stained sections containing papillary and/or solid TCC samples (Section 5.2.1.3) were examined to determine which regions of the tissue contained urothelial cells. Urothelial cells were then manually micro-dissected with a scalpel blade from the frozen tissue blocks and DNA extracted (Section 2.3.3.1). In the final stages of the DNA extraction, the pellet was vacuum dried, resuspended in 10µl sterile MilliQ-H₂O, and then stored at 4°C.

5.2.3.3 Bisulfite modification of DNA from cell lines

These techniques were derived from Clark & Frommer (1997), Bianco *et al.* (1999) and Bianco (personal communication). DNA (approximately 1µg) from each

cell line was suspended in sterile MilliQ-H₂O to a final volume of 18µl and 2µl of 3M NaOH was added, before the samples were incubated at 37°C for 20 minutes. Freshly prepared 4.8M sodium bisulfite (278µl) and 2µl of freshly prepared 100mM hydroquinone (Section 2.2) were added to the DNA and the tubes were mixed and centrifuged briefly. Samples were then incubated in an Eppendorf Master-Gradient PCR Cycler, on a continual cycle alternating between 55°C for 20 minutes and 95°C for 30 seconds, for a total of 4 hours. The modified DNA samples were subsequently purified (Section 5.2.3.5).

5.2.3.4 Bisulfite modification of small DNA amounts isolated from patient tissues

Salmon sperm DNA (1µg) was added to DNA samples of unknown concentrations (Section 5.2.3.2) and the final volume was adjusted to 18µl with sterile MilliQ-H₂O. To the mixture was added 2µl of 3M NaOH and the solution was then incubated at 37°C for 20 minutes. Freshly prepared 4M sodium bisulfite / 6.24M urea solution (278µl) and 2µl of freshly prepared 100mM hydroquinone (Section 2.2) were added to the DNA. The samples were mixed, centrifuged briefly, and incubated in an Eppendorf Master-Gradient PCR Cycler, on a continual cycle alternating between 55°C for 20 minutes and 95°C for 30 seconds, for a total of 4 hours. The bisulfite-modified DNA was subsequently purified (Section 5.2.3.5).

5.2.3.5 Purification and desulfonation of bisulfite modified DNA

Bisulfite-modified DNA samples (Sections 5.2.3.3 & 5.2.3.4) were purified according to standard WIZARD protocols supplied by the manufacturer (Promega). The samples were resuspended in a final volume of 50µl sterile MilliQ-H₂O at 80°C, 5.5µl of 3M NaOH was added and the DNA was incubated at 37°C for 15 minutes.

To samples containing very small amounts of DNA, 1µg glycogen was also added as a co-precipitant. To all samples, 5.6µl 3M NaAc and 150µl ethanol were added before allowing the DNA to precipitate at –80°C for 1 hour. The samples were then centrifuged at 16 600 x g at 4°C for 30 minutes and supernatants were removed. Pellets were washed in 70% ethanol, centrifuged as before, and left to dry at room temperature. The samples were resuspended in 50µl sterile MilliQ-H₂O, 5µl aliquots were dispensed into 0.5ml tubes, and stored at –20°C for subsequent PCR (Section 5.2.3.6).

5.2.3.6 *PCR of bisulfite-modified DNA*

Reaction mix for PCR

5'methyl-UPKIB or InnerS67 primers (50ng/µl, 2µl), 2µl OuterA68 primer (50ng/µl), 5µl 10x Taq Buffer, 1µl MgCl₂ (25mM), 1µl dNTPs (10mM), 0.1µl HotStar Taq (5U/µl) and 5µl bisulfite-modified DNA (variable amount) were combined to a final volume of 50µl with sterile MilliQ-H₂O. Sequence information and location of the primers in the UPKIB cDNA are provided in Sections 2.3.3.3 and 5.3.1.

Cycling conditions for PCR

Cycling conditions involved an initial 15 minutes at 95°C; 5 cycles for 1 minute at 96°C, “touchdown mode” for 45 seconds at 68-1°C per cycle (where the initial cycle was at 68°C and subsequent cycles were one degree less until the final cycle at 64°C), and 45 seconds at 72°C; 30 cycles for 1 minute at 96°C, 1 minute at 63°C, and 1 minute at 72°C; followed by a final elongation incubation for 5 minutes at 72°C.

Analysis of PCR products

PCR products were analysed on a 2.0% agarose gel (Section 2.3.3.4).

5.2.3.7 SSCP gels with PCR amplified bisulfite-modified DNA

Gel preparation

A 0.5x MDE gel was prepared, containing 6.25ml MDE Solution (2x), 3ml 5x TBE and 15.75ml MilliQ-H₂O to a final volume of 25ml. TEMED (10µl) and 100µl of freshly prepared 10% ammonium persulphate were added and the solution was poured between gel plates. The gel, with dimensions 150mm (height) x 160mm (width) x 0.15mm (depth), was allowed to polymerise for a minimum of 1 hour before combs were removed and wells washed with MilliQ-H₂O to remove residual unpolymerised gel.

Gel loading

DNA denaturing solution (DDS) was made by mixing together 950µl formamide, 1µl of 10M NaOH, 49µl MilliQ-H₂O plus a few grains each of bromophenol blue and xylene cyanol. DDS (15µl) was added to 6µl of bisulfite-DNA-amplified PCR product and DNA loading buffer (6x) (Section 2.2) was added to the SPP1/*Eco*RI molecular weight marker and non-denatured PCR control samples. PCR products containing DDS were heated at 95°C for 5 minutes, cooled briefly on ice and loaded immediately.

Gel running conditions

MDE 0.5x gels were electrophoresed at room temperature at 280V for 5 hours. Subsequently, gels were stained at room temperature in the dark for 45 minutes with 400ml SYBR-Gold (40µl aliquot (10 000x concentrate in DMSO) diluted in 0.6x TBE). Gels were visualised on a UV transilluminator (Ultra-lūm model MEB-20, Paramount, CA, USA) linked to a Kodak camera with a SYBR photographic filter (S-7569) and images were captured with Kodak image analysis software (v3.5).

5.2.3.8 Sequencing of PCR amplified bisulfite-modified DNA

PCR products amplified from bisulfite-modified DNA were purified using WIZARD columns (Promega) according to the manufacturer's instructions. Purified products were subsequently sequenced (Flinders Medical Centre Sequencing Facility) using primers 5'methyl-UPKIB and OuterA68, the same primers used to generate the PCR products, or alternatively, using plasmid-specific primers SP6 and T7 to sequence pGEM-Teasy-UPKIB PCR clones.

5.2.3.9 Cloning of PCR amplified bisulfite-modified DNA tissue samples

Competent JM109s

Escherichia coli JM109 bacteria were made competent, based on methodologies in Sambrook *et al.* (1989). Fresh JM109 cultures were prepared by seeding JM109 glycerol stocks into 20ml Luria Bertani media (LB) and incubating with aeration at 37°C overnight. The following morning, 1.5ml of the overnight JM109 culture was added to 20ml fresh LB and bacteria were incubated with aeration at 37°C for a further 2 hours until bacteria reached mid-log phase with OD₆₀₀ of 0.6. Cells (1.5ml) were placed on ice for 20 minutes and then centrifuged for 10 minutes at 13 000 x g at 4°C. Bacterial pellets were resuspended in 10ml cold 0.1M MgCl₂ and centrifuged for 10 minutes at 13 000 x g at 4°C. Cells were resuspended in 2ml of cold 0.1M CaCl₂ and left at 4°C for 1 hour before adding 500µl of 80% glycerol to 500µl of the cells and storing the vials at -80°C.

Ligation

PCR products (327bp; 5.45ng), amplified from bisulfite-modified DNA, were ligated into pGEM-Teasy vector (50ng) in the vector:insert ratio of 1:1. Reactions

were performed containing 1µl T4 DNA ligase (Promega, 3U/µl) and 2x ligase buffer (Promega) and were left overnight at 4°C to ligate.

Transformation of competent JM109 bacterial cells

Competent JM109s in glycerol (200µl) were added to each ligation mix (10µl) or to 50ng plasmid DNA as a control. A control was also prepared with JM109 competent cells in glycerol without ligation mix or plasmid. Samples were incubated on ice for 20 minutes and were then heat shocked at 42°C for 2 minutes. Samples were again placed on ice for 20 minutes, 1ml LB was added samples were incubated at 37°C for 30 minutes to allow the bacteria to recover and express the Ampicillin antibiotic resistance gene found in the pGEM-Teasy plasmid. Samples were then centrifuged for 10 seconds to remove excess LB and the cells were resuspended in the residual LB and were streaked onto nutrient agar plates (Section 2.2) containing ampicillin (50µg/ml), or lacking ampicillin for cells alone. Plates were incubated overnight in a 37°C incubator.

Isolation and identification of colonies containing PCR products

Individual colonies were picked and placed into 1.5ml Luria Bertani broth with ampicillin (50µg/ml) and incubated for 5 hours with aeration at 37°C. Plasmid DNA was purified using an UltraClean Mini Plasmid Prep Kit (Mo Bio laboratories). DNA (50ng) was digested with 5U of the restriction enzyme *PvuII* (10U/µl), 1µl of buffer (10x) and MilliQ-H₂O to a final volume of 10µl. Mixtures were incubated at 37°C for 2 hours and then products were run on 1.0% agarose gels in 1x TAE buffer to identify plasmids containing inserted PCR products. Plasmids containing the DNA inserts were sequenced in both directions with primers SP6 and T7 to analyse both strands for methylation of the CpG sites and for independent verification of the sequence at the CpG sites.

5.2.4 Re-activation of UPKIB mRNA expression with 5-Aza-CdR

This technique was derived from Bender *et al.* (1999) and Dr A. Dobrovic (personal communication).

5.2.4.1 Growth of T24 and J82 bladder cancer-derived cell lines

T24 or J82 TCC cell lines were seeded into 50mm Petri dishes at a density of 3×10^4 cells in 5ml DMEM supplemented with 1% L-Glu (200mM) and 10% FCS and incubated for 24 hours at 37°C with 5% CO₂.

5.2.4.2 Serum starvation

Serum starvation was necessary prior to 5-Aza-2'-deoxycytidine (5-Aza-CdR) treatment to synchronise cells (Pardee, 1989; Rollins & Stiles, 1989) (Section 1.2) and allow concurrent incorporation of 5-Aza-CdR into the newly synthesised DNA in the cells. Monolayers were rinsed in 1x PBS, followed by the addition of 5ml DMEM containing 1% L-Glutamine (L-Glu) (200mM) and 0.1% FCS to each Petri dish and cultures were incubated for 48 hours at 37°C with 5% CO₂.

5.2.4.3 5-Aza-CdR treatment of serum starved cells

T24 and J82 monolayers were rinsed with 1x PBS with the subsequent addition of 5ml DMEM containing 10% FCS and 1% L-Glu (200mM). 5-Aza-2'-deoxycytidine (2.5mM stock) was added to each cell line to give final concentrations of either 3µM (4.8µl per 4ml media) or 1µM (1.6µl per 4ml media). Sterile MilliQ-H₂O (3.2µl) was added as a 0µM control. All cultures were carried out in duplicate.

Due to the instability of 5-Aza-CdR in culture, fresh 5-Aza-CdR was added 24 hours and 48 hours after the first addition of 5-Aza-CdR without changing the culture medium. After 48 and 72 hours following the third addition of 5-Aza-CdR, the cells

were washed in 1x PBS and harvested. RNA was isolated (Section 5.2.2.1) and UPKIB mRNA expression was analysed by RT-PCR (Section 5.2.2.2).

5.3 Results

5.3.1 Primer design for bisulfite methylation studies

Primers were designed from the genomic contig, GenBank accession number AC083800, containing the UPKIB promoter region, exon 1 and the beginning of intron 1. The region analysed was chosen to maximise the number of CpG dinucleotides analysed. The primers 5'methyl-UPKIB, InnerS67 and OuterA68 were designed to span 11 CpG sites within the UPKIB CpG island, with specificity for bisulfite-modified DNA but not unmodified genomic DNA (Fig. 5.3A). 5' primers 5'methyl-UPKIB or InnerS67 and 3' primer OuterA68 were designed to amplify 327bp and 239bp fragments respectively. The program Amplify verified the specificity of these primers for the bisulfite-treated DNA (Fig. 5.3B).

5.3.2 Investigation of *UPKIB* methylation in TCC cell lines

5.3.2.1 *Methylation analysis of cell lines and PBL*

DNA was isolated from all cell lines and PBLs using the TES method (Section 2.3.3.1) and bisulfite-modified as described above (Sections 5.2.3.3).

Using bisulfite-treated DNA as the template, PCR amplified 327bp or 239bp products, but unmodified genomic DNA controls were not amplified (Fig. 5.4), demonstrating the specificity of the primers for bisulfite-treated DNA. The amplification step itself was not designed to show whether the PCR product amplified from the CpG island was methylated or unmethylated. SSCP and direct

Figure 5.3

The UPKIB CpG Island and Primer Design

A. Unmodified (unmod) and bisulfite-modified (BiS) DNA sequences, of GenBank AC083800 contig, used to determine the GC content, CpG:GpC ratio and primer sequences. The UPKIB CpG island is 470bp in length and contains 15 CpG dinucleotides. The GC content is calculated as $G+C/\text{total residues} = 252/470 = 54\%$. The CpG:GpC ratio is calculated as total CpG dinucleotides to total GpC dinucleotides = $15:35 = 0.43$. Primers shown are 5' methyl-UPKIB (green), InnerS67 (yellow) and OuterA68 (blue); and exon 1 is boxed in black. CpG sites are highlighted in red and bisulfite-modified cytosines, converted to uracil and amplified as thymidine, are shown in blue. **B.** The program Amplify was used to verify the suitability of the primers in the bisulfite-modified DNA PCR reactions. Primers shown: 5' methyl UPKIB (blue arrow), InnerS67 (blue arrow) and OuterA68 (red arrow). Products are generated only in reactions containing bisulfite-modified DNA. PCR products were not predicted using unmodified DNA, showing the suitability of the primer pairs for the methylation analyses.

A

Unmod **TCGCTGTGAGCGGTCAGGACTAACAAGCCTTCAACACACAAAAGCAGGAG**
BiS **TCGTTGTGAGCGGTTAGGATTAATAAGTTTTTAATATATAAAAGTAGGAG**

Unmod **TAAGGATCAAATTACACAAATCAAATCAAATCAAATCAAATAAAATCCC**
BiS **TAAGGATTAAATTATATAAATTAATTAATTAATTAATTAATAAAATTTT**

Unmod **TGAGTTTGTATTCCAGTAAAGCCCAAACCTAGGAGGGATCCAGGAAAGCG**
BiS **TGAGTTTGTATTTTAGTAAAGTTTAAATTTAGGAGGGATTTAGGAAAGCG**
5'methyl-UPKIB →

Unmod **ATGAGTGTGGCTGTCAAGGTGCTTCCATCTGTCAGAGTCACACATTCCAA**
BiS **ATGAGTGTGGTTGTTAAGGTGTTTTATTTGTAGAGTTATATATTTTAA**

Unmod **AGCAGCAGAAATCAACTGACAGCCTCCGTTACTCCCCCGCCTCAGCGAG**
BiS **AGTAGTAGAAATTAATTGATAGTTTTCGTTATTTTTCGTTTAGCGAG**
InnerS67 →

Unmod **GCAGGGCAGGCAGCGGAAAGCTCAGCCCATGTGAGGTGCCTCCTGCCAAT**
BiS **GTAGGGTAGGTAGCGGAAAGTTAGTTTATGTGAGGTGTTTTTTGTTAAT**

Unmod **CACAGACTACCCTTCCCTGGTCTGGAGGTTCAAAGAATTGCAGGAGGGT**
BiS **TATAGATTATTTTTTTTTGGTTTTGGAGGTTTAAAGAATTGTAGGAGGGT**

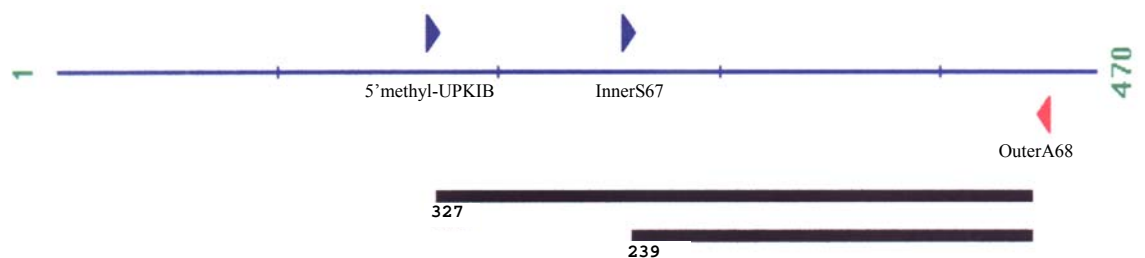
Unmod **AGAAAAGCACCTGGGTGGGTGCAGACTGCGGAGCGGGCCCTACCGTGTG**
BiS **AGAAAAGTATTTGGGTGGGTGTAGATTGCGGAGCGGGTTTTATCGTGTG**

Unmod **CGCAGAAAGAGGAGGCGCTGTGAGTAGACCGCTTTCCCTCTGCAGGGAAT**
BiS **CGTAGAAAGAGGAGGCGTGTGAGTAGATCGTTTTTTTTTTGTAGGGAAT**

Unmod **CTCCCGGGCTCAAGGGCTGG**
BiS **TTTTCGGGTTAAGGGTTGG**
← OuterA68

B

Bisulfite-modified sequence



Unmodified sequence



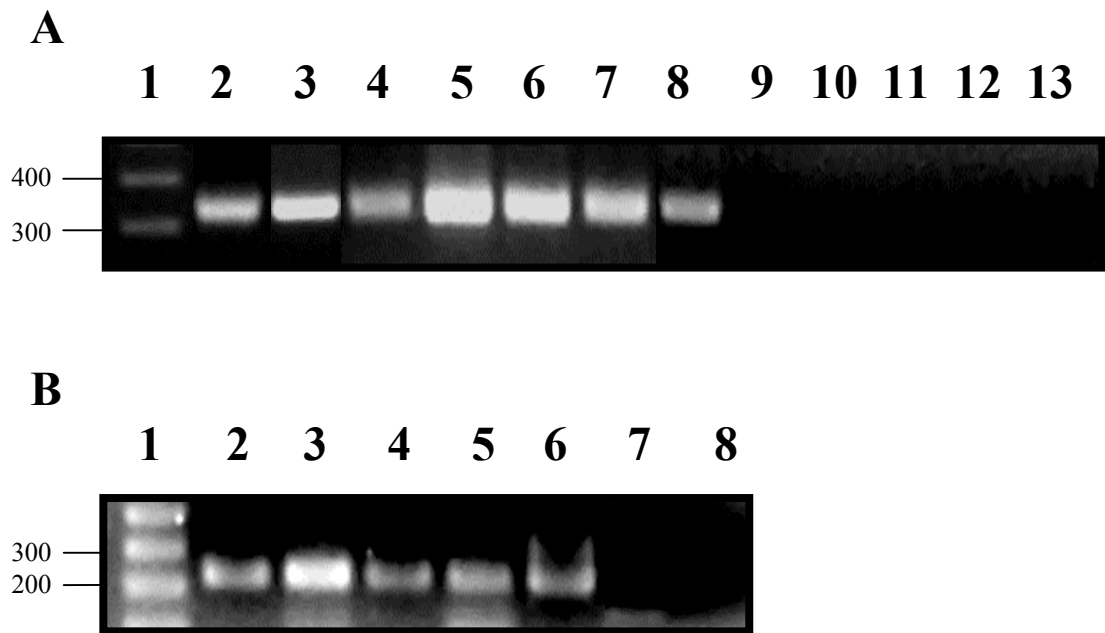


Figure 5.4

Bisulphite PCR Amplification of 327bp and 239bp Fragments

A. PCR products from amplification of 327bp bisulfite-treated DNA with primers 5'methyl UPKIB and OuterA68. Lanes: (1) 1kb+ molecular weight marker (500ng); (2) VM-Cub3; (3) HT1376; (4) RT112; (5) 5637; (6) T24; (7) J82; (8) TCC-Sup; (9) VM-Cub3; (10) 5637; (11) T24; (12) J82; (13) water control. DNA samples amplified in lanes (2-8) were bisulfite treated. DNA samples used in PCR reactions in lanes (9-12) were unmodified. **B.** PCR products from amplification of 239bp bisulfite-treated DNA with primers InnerS67 and OuterA68. Lanes: (1) 1kb+ molecular weight marker (500ng); (2) VM-Cub3; (3) HT1376; (4) RT112; (5) 5637; (6) T24; (7) VM-Cub3; (8) water control. DNA samples amplified in lanes (2-6) have been bisulfite treated. DNA used in PCR reactions in lane (7) was unmodified.

sequencing techniques were then used to analyse further the PCR fragments to determine the methylation status of the UPKIB CpG island in these samples.

5.3.2.2 *Direct sequencing*

Direct sequencing of the PCR products derived from bisulphite-modified DNA using primers 5'methyl-UPKIB and OuterA68 demonstrated a marked difference in methylation status of the UPKIB promoter between the different bladder cancer-derived cell lines. Figure 5.5 shows a multiple sequence alignment of the CpG regions for all cell lines examined (raw sequencing data is shown in Appendix N). This figure shows that bisulfite modification was effective in all of the cell lines, as all cytosines were converted to uracil and were amplified as thymines during PCR, except at methylated CpG sites. All cell lines have the same sequence except at the CpG dinucleotides, which show variations in methylation patterns. A summary of these results and of the positions of the CpG sites in relation to the UPKIB promoter, exon 1 and intron 1 is shown in Figure 5.6. In this figure, some of the sites are described as variably methylated. These are sites in which the C/T peaks in the sequencing traces at the CpG sites were similar in height (Appendix N). This may be due to the two alleles having a different methylation status at these CpG sites. All PCR products were sequenced at the Flinders Medical Centre Sequencing Facility (Section 2.3).

5.3.2.3 *SSCP analysis*

SSCP (single-strand conformation polymorphism) is a sensitive method used to detect changes of single or multiple nucleotides in a defined segment of DNA. In methylation analysis, it can be used in conjunction with DNA bisulfite-conversion to study alterations in banding patterns between samples with different methylation

Figure 5.5

Multiple Sequence Alignments between UPKIB PCR Fragments Amplified from Bisulfite-modified DNA from Cell Lines and PBL

Shown are alignments of sequencing results of bisulfite-treated DNA PCR products (Figure 5.4) for transitional cell carcinoma cell lines VM-Cub3, HT1376, RT112, 5637, T24, J82, TCC-Sup; squamous cell carcinoma cell line Sca-BER; and PBL (peripheral blood lymphocytes). DNA was sequenced with primer OuterA68.

	Key
Black	100% conservation of sequence
Grey	60 to <100% conservation of sequence
Green	Last nucleotide of primer OuterA68
Red	CpG dinucleotides

		*	20	*	40	*	60	
AC083800	:	CCTCCGGT	TA	CCCCGC	CTCAGCGAGG	CAGGGCAGGC	AGCGGAAAGCTCAGCC	ATGT : 60
VMCUB3	:	TTTTTAGT	TATTTTTT	CGTTTTAG	CCAGGTAGGG	TAGGTAGCC	GAAAGTTTAGTTT	ATGT : 60
HT1376	:	TTTTCGGT	TATTTTTT	CGTTTTAG	CCAGGTAGGG	TAGGTAGCC	GAAAGTTTAGTTT	ATGT : 60
RT112	:	TTTTTGGT	TATTTTTT	CGTTTTAG	CCAGGTAGGG	TAGGTAGCC	GAAAGTTTAGTTT	ATGT : 60
5637	:	TTTTCGGT	TATTTTTT	CGTTTTAG	CCAGGTAGGG	TAGGTAGCC	GAAAGTTTAGTTT	ATGT : 60
T24	:	TTTTNGGT	TATTTTTT	CGTTTTAG	CCAGGTAGGG	TAGGTAGCC	GAAAGTTTAGTTT	ATGT : 60
J82	:	TTTTNNGT	TATTTTTT	CGTTTTAG	CCAGGTAGGG	TAGGTAGCC	GAAAGTTTAGTTT	ATGT : 60
TCC-SUP	:	TTTTCGGT	TATTTTTT	CGTTTTAG	CCAGGTAGGG	TAGGTAGCC	GAAAGTTTAGTTT	ATGT : 60
PBL	:	TTTTCGGT	TATTTTTT	CGTTTTAG	CCAGGTAGGG	TAGGTAGCC	GAAAGTTTAGTTT	ATGT : 60
Sca-BER	:	TTTTCGGT	TATTTTTT	CGTTTTAG	CCAGGTAGGG	TAGGTAGCC	GAAAGTTTAGTTT	ATGT : 60

		*	80	*	100	*	120	
AC083800	:	GAGGTGCCT	CCTGCCAAT	CACAGACT	ACCCTTCCCT	GGTCTGGAGG	TCAAAGAATTGC	: 120
VMCUB3	:	GAGGTGTTTT	TGTTAATT	TATAGATT	ATTTTTTTTT	TGGTTTTGGAGG	TTAAAGAATTGT	: 120
HT1376	:	GAGGTGTTTT	TGTTAATT	TATAGATT	ATTTTTTTTT	TGGTTTTGGAGG	TTAAAGAATTGT	: 120
RT112	:	GAGGTGTTTT	TGTTAATT	TATAGATT	ATTTTTTTTT	TGGTTTTGGAGG	TTAAAGAATTGT	: 120
5637	:	GAGGTGTTTT	TGTTAATT	TATAGATT	ATTTTTTTTT	TGGTTTTGGAGG	TTAAAGAATTGT	: 120
T24	:	GAGGTGTTTT	TGTTAATT	TATAGATT	ATTTTTTTTT	TGGTTTTGGAGG	TTAAAGAATTGT	: 120
J82	:	GAGGTGTTTT	TGTTAATT	TATAGATT	ATTTTTTTTT	TGGTTTTGGAGG	TTAAAGAATTGT	: 120
TCC-SUP	:	GAGGTGTTTT	TGTTAATT	TATAGATT	ATTTTTTTTT	TGGTTTTGGAGG	TTAAAGAATTGT	: 120
PBL	:	GAGGTGTTTT	TGTTAATT	TATAGATT	ATTTTTTTTT	TGGTTTTGGAGG	TTAAAGAATTGT	: 120
Sca-BER	:	GAGGTGTTTT	TGTTAATT	TATAGATT	ATTTTTTTTT	TGGTTTTGGAGG	TTAAAGAATTGT	: 120

		*	140	*	160	*	180	
AC083800	:	AGGAGGGT	AGAAAAG	CACCTGGG	TCCGGTGC	AGACTGCCG	AGCGGGCC	TACCGTGTGCG : 180
VMCUB3	:	AGGAGGGT	AGAAAAG	TATTTGGG	TCCGGTGC	AGATTGCC	GAGCCGTTT	TATCGTGTGCG : 180
HT1376	:	AGGAGGGT	AGAAAAG	TATTTGGG	TCCGGTGC	AGATTGCC	GAGCCGTTT	TATCGTGTGCG : 180
RT112	:	AGGAGGGT	AGAAAAG	TATTTGGG	TCCGGTGC	AGATTGCC	GAGCCGTTT	TATCGTGTGCG : 180
5637	:	AGGAGGGT	AGAAAAG	TATTTGGG	TCCGGTGC	AGATTGCC	GAGCCGTTT	TATCGTGTGCG : 180
T24	:	AGGAGGGT	AGAAAAG	TATTTGGG	TCCGGTGC	AGATTGCC	GAGCCGTTT	TATCGTGTGCG : 180
J82	:	AGGAGGGT	AGAAAAG	TATTTGGG	TCCGGTGC	AGATTGCC	GAGCCGTTT	TATCGTGTGCG : 180
TCC-SUP	:	AGGAGGGT	AGAAAAG	TATTTGGG	TCCGGTGC	AGATTGCC	GAGCCGTTT	TATCGTGTGCG : 180
PBL	:	AGGAGGGT	AGAAAAG	TATTTGGG	TCCGGTGC	AGATTGCC	GAGCCGTTT	TATCGTGTGCG : 180
Sca-BER	:	AGGAGGGT	AGAAAAG	TATTTGGG	TCCGGTGC	AGATTGCC	GAGCCGTTT	TATCGTGTGCG : 180

		*	200	*	220	
AC083800	:	CAGAAAGAGG	AGGCGCTG	TGAGTAGA	CCGCTTTCC	CTG : 220
VMCUB3	:	TAGAAAGAGG	AGGCGCTG	TGAGTAGA	CGTTTTTTT	TG : 220
HT1376	:	TANAAAGAGG	AGGCGCTG	TGAGTAGA	TGTTTTTTT	TG : 220
RT112	:	TAAAAAGAGG	AGGCGCTG	TGAGTAGA	TGTTTTTTT	TG : 220
5637	:	TAGAAAGAGG	AGGCGCTG	TGAGTAGA	CGTTTTTTT	TG : 220
T24	:	TAGAAAGAGG	AGGCGCTG	TGAGTAGA	CGTTTTTTT	TG : 220
J82	:	TAGAAAGAGG	AGGCGCTG	TGAGTAGA	CGTTTTTTT	TG : 220
TCC-SUP	:	TAGAAAGAGG	AGGCGCTG	TGAGTAGA	CGTTTTTTT	TG : 220
PBL	:	TAAAAAGAGG	AGGCGCTG	TGAGTAGA	CGTTTTTTT	TG : 220
Sca-BER	:	TAGAAAGAGG	AGGCGCTG	TGAGTAGA	CGTTTTTTT	TG : 220

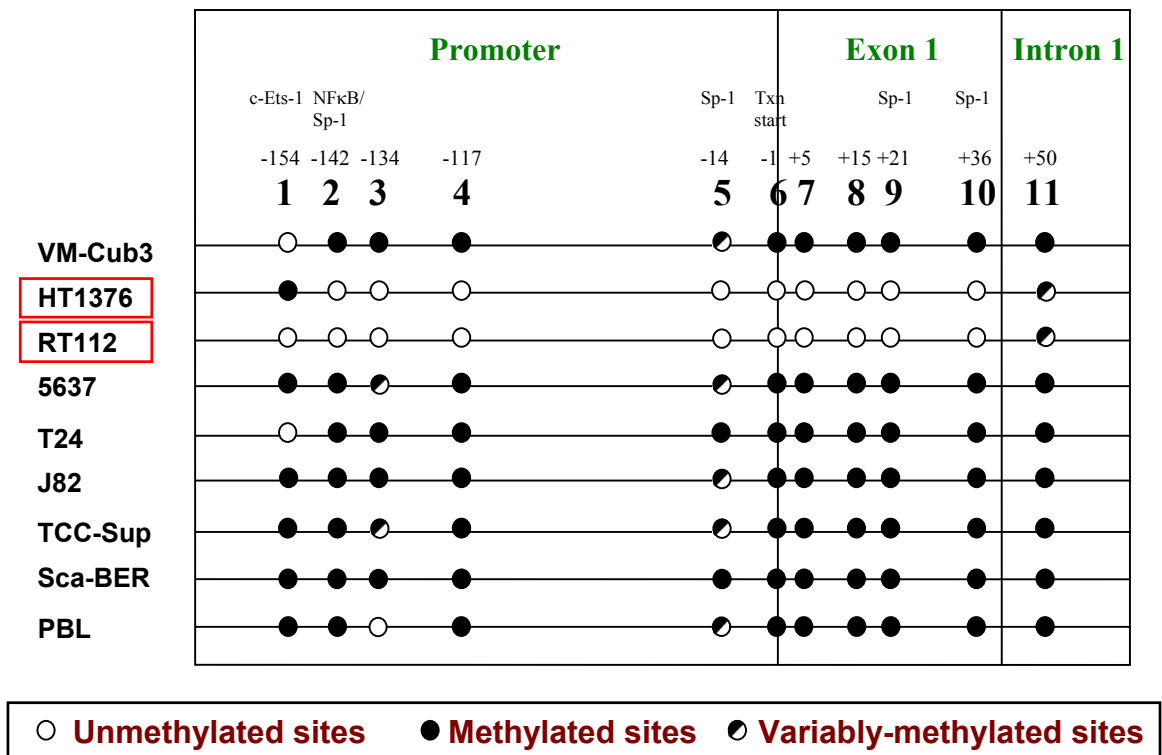


Figure 5.6

Direct Sequencing of PCR Products Amplified from Bisulphite-modified DNA from Cell Lines and PBL

CpG residues are numbered with the corresponding genomic location in the UPKIB putative promoter, exon 1 and intron 1 regions. DNA was isolated from cell lines VM-Cub3, HT1376, RT112, 5637, T24, J82, TCC-Sup, Sca-BER and peripheral blood lymphocytes (PBL) and was analysed for CpG methylation. The methylation at the CpG sites is based on the sequences in the multiple sequence alignment (Figure 5.5).

Open circles denote unmethylated sites, enclosed circles denote methylated sites and half circles denote variably methylated sites. Red boxes indicate cell lines expressing UPKIB mRNA, as shown by RT-PCR analysis (Figure 5.8).

states in regulatory CpG islands. In this method, PCR fragments from bisulfite-modified PCR amplification are denatured and run on a polyacrylamide gel. Changes in methylation status can be seen by subtle shifts in banding patterns in the gel. This approach is known as MS-SSCA (methylation-sensitive single-strand conformation analysis) (Bianco *et al.*, 1999).

MDE (0.5x) gel with 327bp PCR fragments

To determine the methylation status of *UPK1B* using MS-SSCA, bisulfite-treated DNA from cell lines shown to be unmethylated, HT1376 and RT112 was compared to cells lines which were methylated at CpG sites, including T24. However, SSCP analysis, using PCR products amplified from bisulfite-treated DNA from 7 cell lines and peripheral blood lymphocyte (Fig. 5.7A), showed no significant difference in banding patterns of PCR products generated from HT1376 and RT112 DNA samples and the other fully methylated samples. The HT1376 product showed no visible difference to the other samples including the fully methylated T24 and PBL-1 samples (Section 5.3.2.2), whereas the RT112 sample exhibited a slight conformer shift, however this difference was not sufficient to detect differences between the methylation profiles of these samples.

MDE (0.5x) gel with 239bp PCR fragments

As MS-SSCA is most effective in detecting conformer variance for PCR products between 200 and 300bp, the 327bp fragment may have been too large (A. Dobrovic, personal communication). Hence, primers InnerS67 and OuterA68 were subsequently used to amplify a shorter 239bp fragment for analysis. In this instance, 5 cell lines were analysed to determine if differences in mobility could be detected between PCR products generated from bisulfite-modified DNA from cell lines with unmethylated CpG sites (RT112 and HT1376) and those methylated at their CpG

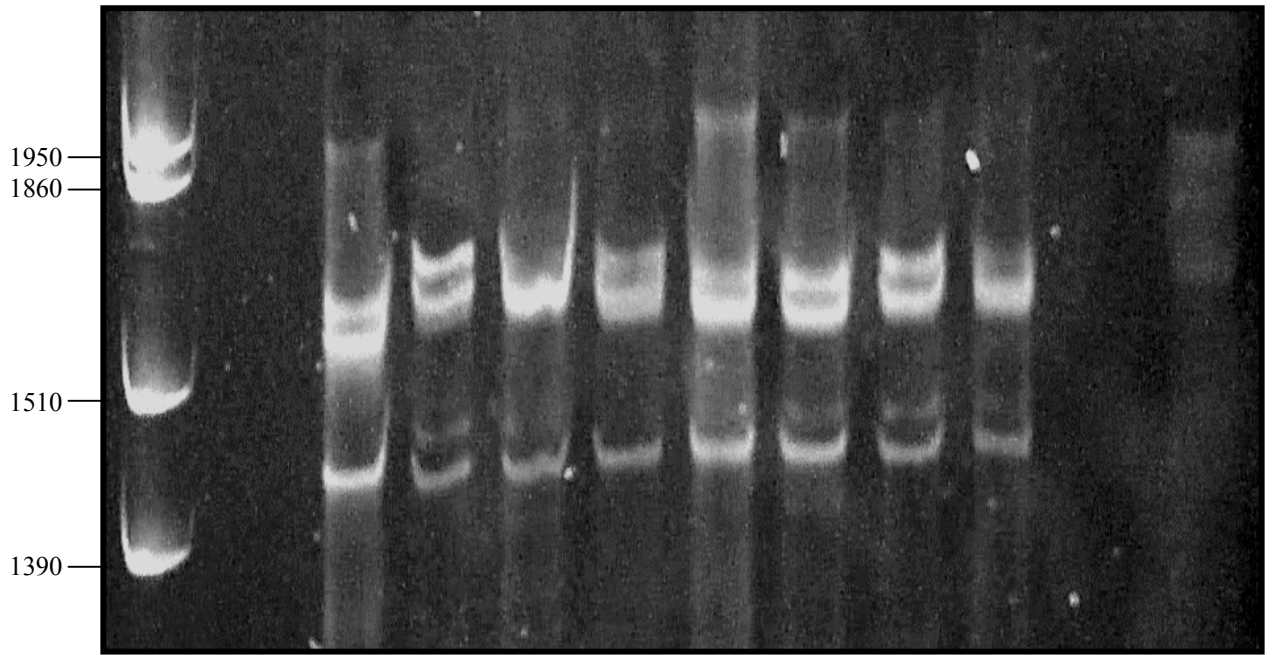
Figure 5.7

MS-SSCA Analysis of TCC Cell Lines and PBL

A. PCR product (327bp), containing 11 CpG sites was amplified using primers 5'methyl-UPKIB and OuterA68. Lanes: (1) 250ng SPP1/*Eco*RI molecular weight marker; (2) blank; (3) VM-Cub3; (4) HT1376; (5) RT112; (6) 5637; (7) T24; (8) J82; (9) TCC-Sup; (10) peripheral blood lymphocytes; (11) blank; (12) undenatured RT112 PCR product. **B.** PCR product (239bp), containing 8 CpG sites was amplified using primers InnerS67 and OuterA68. Lanes: (1) 250ng SPP1/*Eco*RI molecular weight marker; (2) VM-Cub3; (3) HT1376; (4) RT112; (5) 5637; (6) T24; (7) undenatured VM-Cub3 PCR product.

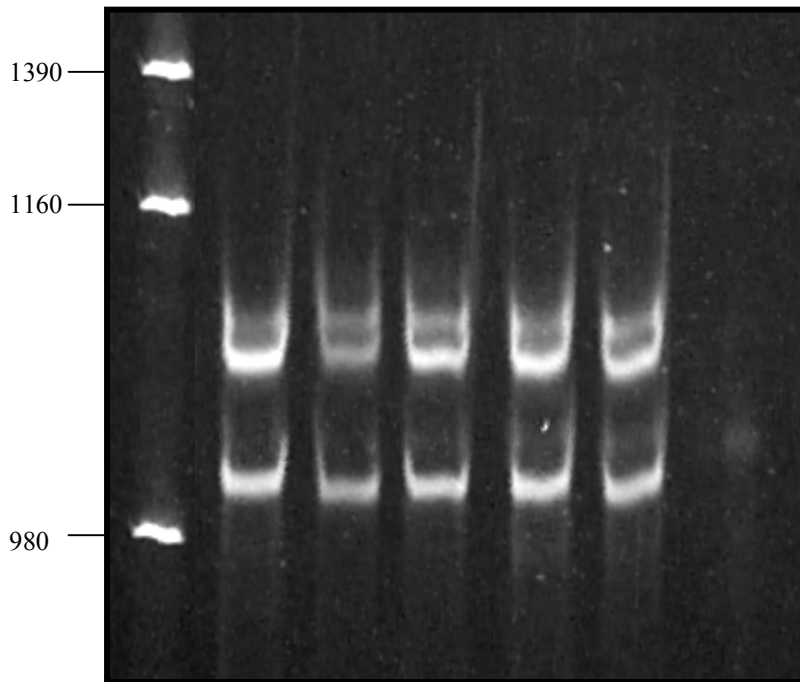
A

1 2 3 4 5 6 7 8 9 10 11 12



B

1 2 3 4 5 6 7



sites (VM-Cub3, T24 and J82). As with the 327bp fragments, no significant differences in banding patterns could be detected (Fig. 5.7B).

A range of different conditions was also tested in an attempt to optimise changes in band mobility. MDE (0.75x) gels, run for 20 hours at room temperature at 280V and MDE (0.5x) gels with 10% glycerol, run at 4°C for 5 hours were also attempted, but the PCR products showed no observable differences in band positions.

5.3.3 Investigation of UPKIB mRNA expression in TCC cell lines

TCC cell lines VM-Cub3, HT1376, RT112, 5637, T24, J82, and TCC-SUP and SCC (squamous cell carcinoma) cell line Sca-BER were analysed for expression of UPKIB mRNA to determine whether the expression of UPKIB mRNA is related to methylation of the UPKIB CpG island in these cell lines. Of the seven TCC cell lines, only two lines, HT1376 and RT112, expressed detectable levels of UPKIB mRNA, as shown by the amplification of a 743bp fragment (Fig. 5.8). Expression of mRNA of the housekeeping gene PBGD was uniform in all samples with the presence of a 377bp product.

5.3.4 DNA methylation in normal and TCC patient samples

5.3.4.1 Haematoxylin & Eosin staining and micro-dissection of 8 TCC tissue samples

Four samples of normal urothelium and 8 TCCs were analysed (clinical details are given in Table 5.1). OCT sections of each tissue were stained with Haematoxylin and Eosin and were either papillary or solid in nature (Section 5.3.4.3). These sections were used to identify which region of the tissue contained urothelium or tumour tissue, which was subsequently micro-dissected and analysed for expression of UPKIB mRNA and for methylation status of the UPKIB CpG island.

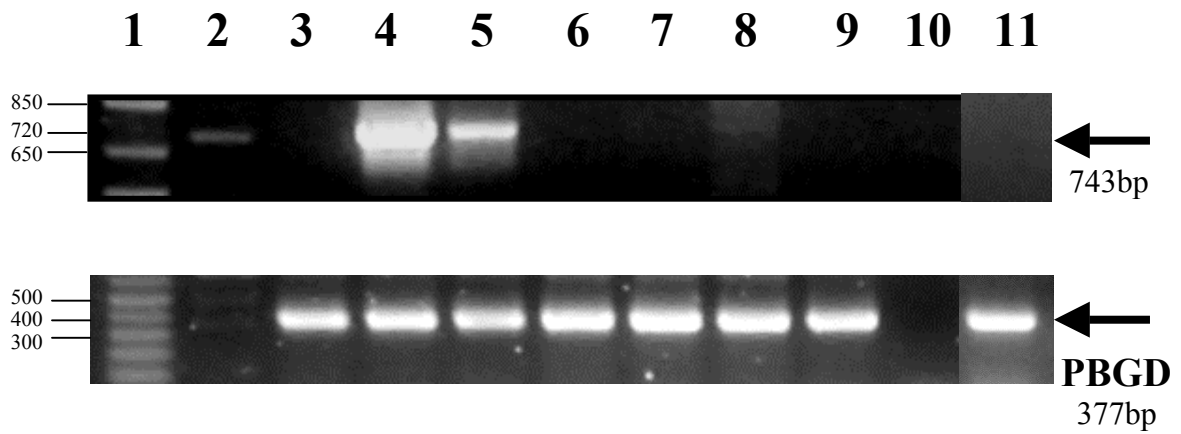


Figure 5.8

RT-PCR Analysis of Cell Lines for UPKIB mRNA Expression

Lanes: (1) 500ng 1kb+ molecular weight marker; (2) 500ng SPP1/*Eco*RI molecular weight marker; (3) VM-Cub3; (4) HT1376; (5) RT112; (6) 5637; (7) T24; (8) J82; (9) TCC-Sup; (10) water control; (11) Sca-BER.

Top panel: PCR with UPKIB primers F1/R1

Lower panel: PCR with PBGD primers PBGD-5'/PBGD-3'

5.3.4.2 RT-PCR analysis of normal urothelial and colonic samples

RT-PCR analysis of normal urothelial and colonic tissues demonstrated that UPKIB mRNA was expressed in the urothelium but not in colonic epithelium (Fig. 5.9A). Similarly, UPKIB mRNA was not detected in normal PBLs isolated from five different donors (PBL-2 to 6) (Fig. 5.9B).

5.3.4.3 Extraction, bisulfite modification, and amplification of genomic DNA from normal urothelium, colon and TCC samples

To gauge the methylation profiles of the CpG island in DNA from tumours (Fig. 5.10), genomic DNA (Section 2.3.3.1) and RNA (Sections 2.3.2.2 & 5.2.2.1) were extracted from the OCT-embedded samples. Bisulfite-modified genomic DNA was subjected to PCR, with primers OuterA68 and 5'methyl-UPKIB, and results are shown in Figure 5.11. Amplification of all modified DNA samples yielded products with expected size 327bp and no amplification was observed in unmodified DNA samples (Fig. 5.11A, lane 6; Fig. 5.11B, lane 14), demonstrating the specificity of the PCR for bisulfite-modified DNA.

5.3.4.4 Direct sequencing of patient samples

PCR products (327bp) from normal epithelial samples SC-N, TE-N, and COL-1 and tumour samples BR-T1, HA-T1, PE-T2 and RO-T1 were sequenced using primers 5'methyl-UPKIB and OuterA68 (raw sequencing data in Appendix O) and aligned with PCR products from cell lines RT112 and T24 (Figure 5.12). There was considerable variability at the CpG sites within individual samples as demonstrated by the presence of N bases at CpG sites in the alignment which could not be resolved into either cytosines or thymines. This heterogeneity in the sequencing data is represented in Figure 5.13 demonstrating ambiguities in the methylation analysis.

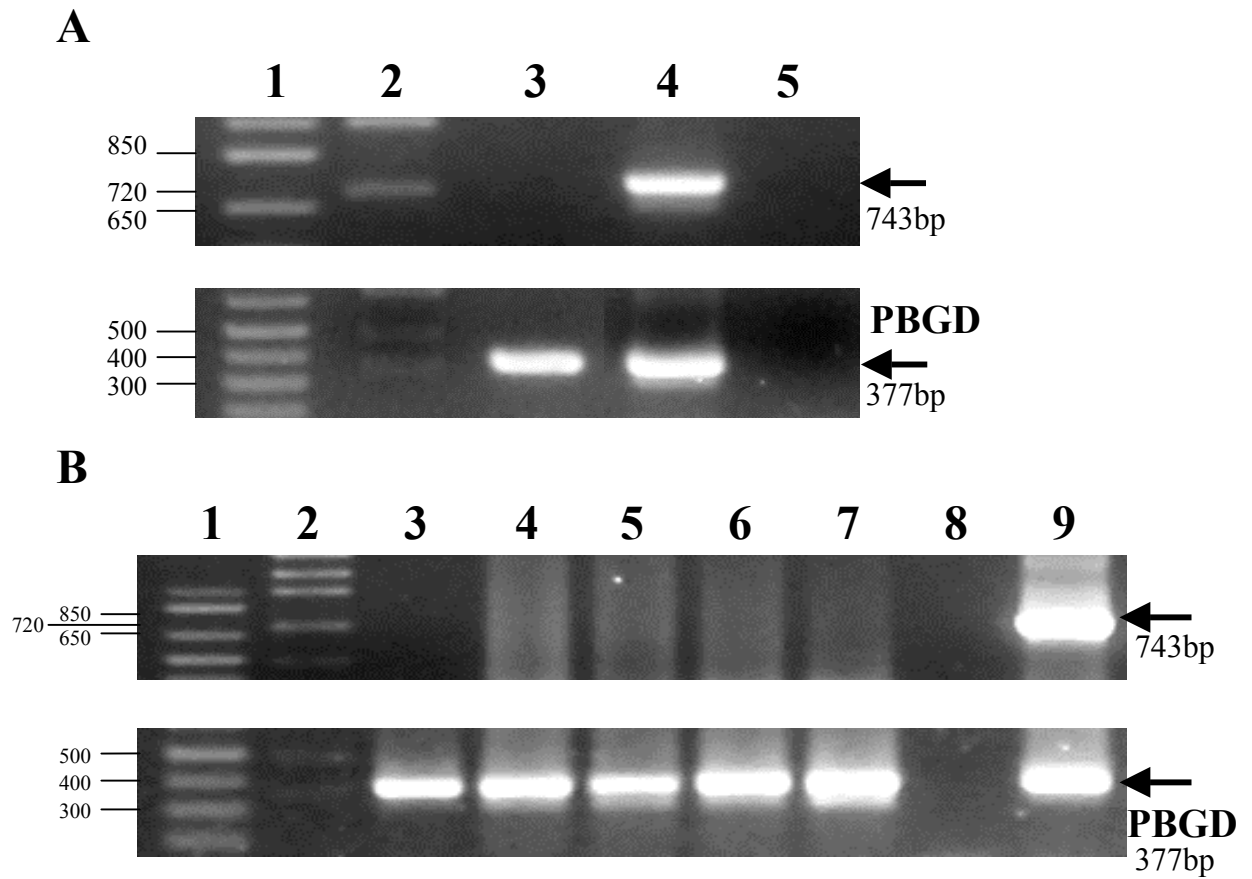


Figure 5.9

RT-PCR Analysis of UPK1B mRNA Expression in Normal Colonic Epithelium, Normal Urothelium and PBL

A. RT-PCR analysis of colonic epithelium and normal urothelial samples was performed using F1 and R1 primers. Lanes: (1) 1kb+ MWM (500ng); (2) SPP1/*Eco*RI MWM (500ng); (3) colonic sample COL-1; (4) urothelial sample TE-N; (5) water (negative control). **B.** RT-PCR analysis of PBLs from 5 normal volunteers was performed using F1 and R1 primers. Lanes: (1) 1kb+ MWM (500ng); (2) SPP1/*Eco*RI MWM (500ng); (3) PBL-1; (4) PBL-2; (5) PBL-3; (6) PBL-4; (7) PBL-5; (8) water (negative control); (9) RT112 (positive control).

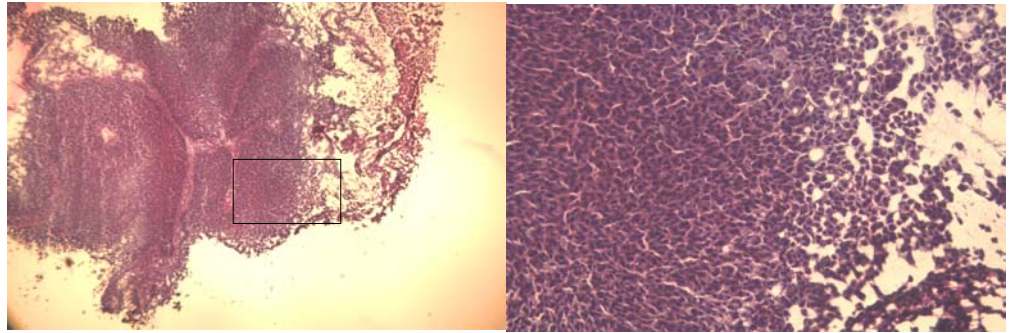
Figure 5.10

Haematoxylin and Eosin Staining of TCC Samples of Bladders and Ureter

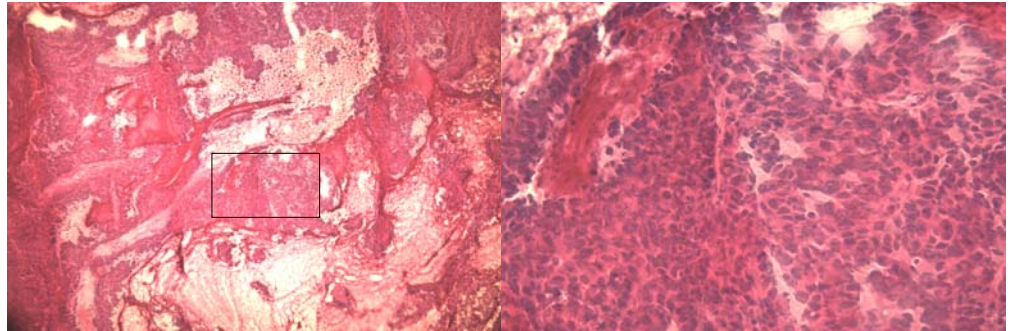
Patient TCC samples were H&E stained to determine regions containing urothelial cells for dissection. Shown are TCC patient samples BR-T1, HA-T1, HA-T2, HA-T3, PE-T2, RA-T1, RO-T1 and WA-T1. Regions containing TCC from patient samples RA-T1 and RO-T1 are indicated with blue hashed lines and arrows. The regions containing TCC were manually micro-dissected with a scalpel for analysis of UPK1B promoter methylation and mRNA expression (Figures 5.11-5.14).

Images in the left-hand columns, at 40x total magnification, are magnified in the second column at total 200x magnification with a light microscope. Scale bars represent 250 μ m and 50 μ m in 40x and 200x total magnifications respectively.

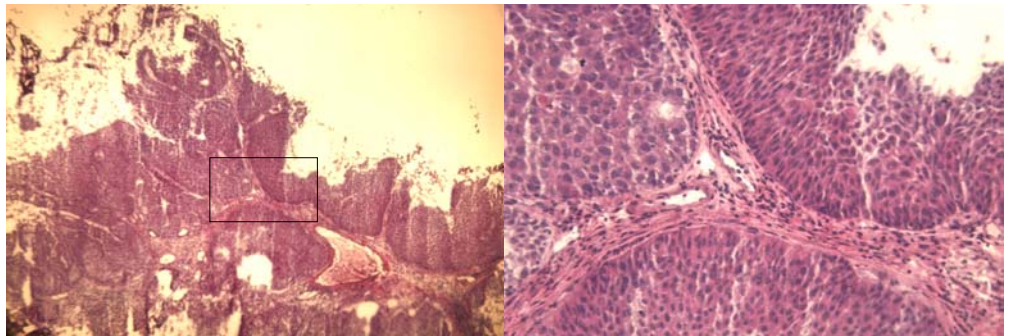
BR-T1



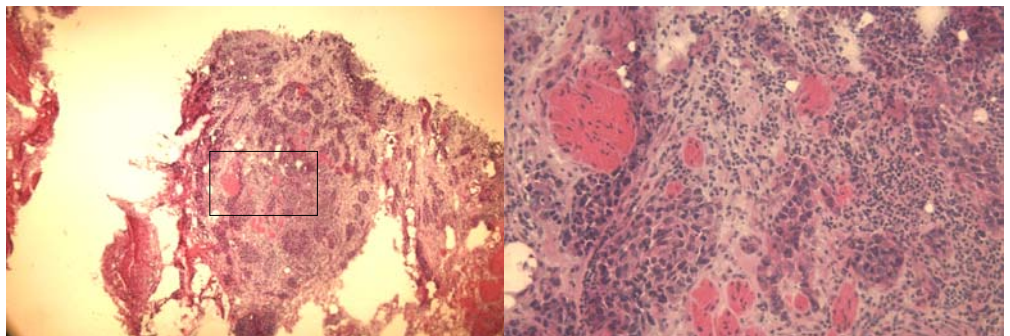
HA-T1



HA-T2



HA-T3



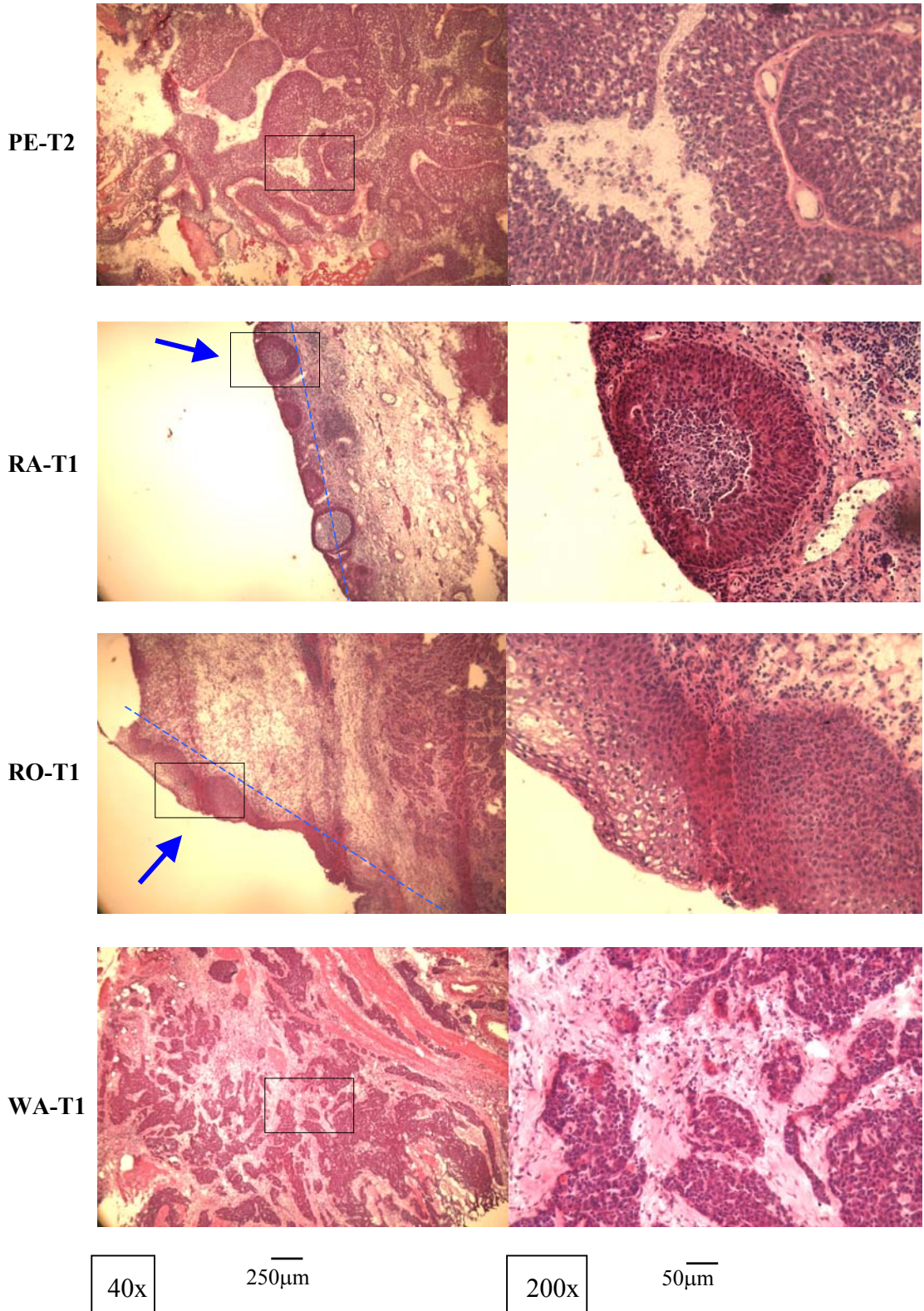
40x

250μm

200x

50μm

Figure 5.10 (continued)



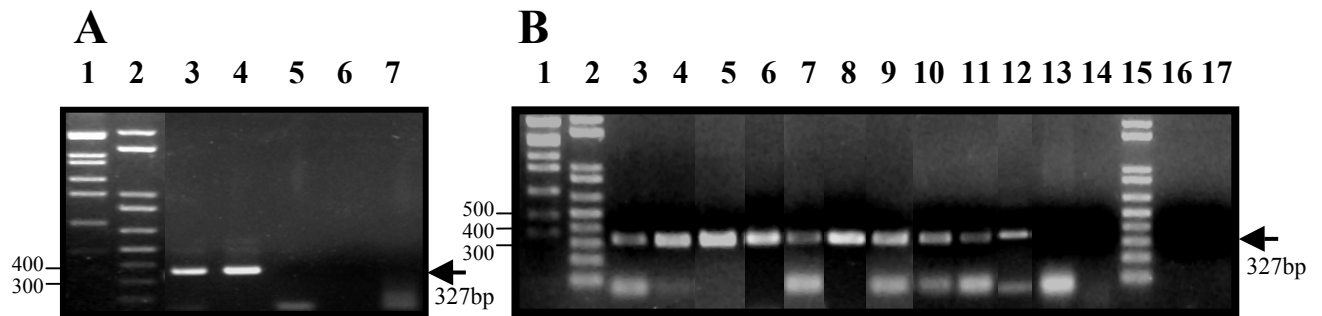


Figure 5.11

PCR Amplification of 327bp *UPKIB*-derived Products from DNA Treated with Bisulfite, and Derived from Normal Tissues of Colonic Epithelium, Urothelium and Ureteric Muscle and from TCCs

A. Normal urothelial DNA, bisulfite-modified and amplified by PCR using primers 5'methyl-*UPKIB* and OuterA68. Lanes: (1) 500ng SPP1/*Eco*RI molecular weight marker; (2) 500ng 1kb+ molecular weight marker; (3) SC-N; (4) TE-N; (5) VM-Cub3 unmodified DNA control; (6-7) water controls. **B.** Bisulfite-modified DNA from TCC and normal colonic epithelial tissues amplified by PCR with primers 5'methyl-*UPKIB* and OuterA68. Lanes: (1) 500ng SPP1/*Eco*RI molecular weight marker; (2) 500ng 1kb+ molecular weight marker; (3) COL-1; (4) COL-2; (5) BR-T1; (6) HA-T1; (7) HA-T2; (8) HA-T3; (9) PE-T2; (10) RA-T1; (11) RO-T1; (12) WA-T1; (13) VM-Cub3 unmodified; (14) RT112 unmodified; (15) 500ng 1kb+ molecular weight marker; (16-17) water controls.

Figure 5.12

Multiple Sequence Alignments between UPKIB PCR Fragments from Patient Sample Bisulphite-modified DNA and RT112 and T24 Cell Lines

Alignments of sequencing results of bisulfite-treated DNA PCR products (Figure 5.13) are shown for normal colonic epithelium COL-1, normal urothelium SC-N and TE-N, transitional cell carcinoma patient samples BR-T1, HA-T1, PE-T2 and RO-T1, and transitional cell carcinoma cell lines RT112 and T24. DNA was sequenced with primer OuterA68. There are N sequences at the cytosine residues at a number of CpG sites. These may be accounted for by variable methylation patterns at these CpG sites with similar cytosine and thymine peaks.

	Key
Black	100% conservation of sequence
Grey	50 to <100% conservation of sequence
Green	Last nucleotide of primer OuterA68
Red	CpG dinucleotides

```

          *           20           *           40           *           60
AC083800 : CCTCCGGTTACTCCCCCGCCTCAGCGAGGCAGGGCAGGCAGCGGAAAGCTCAGCCCATGT : 60
COL-1    : TTTTCGGTTATTTTTTNGTTTTAGTGAGGTACGGTAGGTAGTGGAAAGTTAGTTTNTGT : 60
SC-N     : TTTTCGGTTATTTTTTCGTTTTAGTGAGGTACGGTAGGTAGCGGAAAGTTAGTTTATGT : 60
TE-N     : TTTTCGGTTATTTTTTCGTTTTAGNGAGGTACGGTAGGTAGCGGAAAGTTAGTTTATGT : 60
BR-T1    : TTTTCGGTTATTTTTTCGTTTTAGNGAGGTACGGTAGGTAGNGGAAAGTTAGTTTATGT : 60
HA-T1    : TTTTCGGTTATTTTTTCGTTTTAGNGAGGTACGGTAGGTAGCGGAAAGTTAGTTTATGT : 60
PE-T2    : TTNTCGGTTATTTTTTCGTTTTAGNGAGGTACGGTAGGTAGCGGAAAGTTAGTTTATGT : 60
RO-T1    : TTTTCGGTTATTTCTTCGTTTTAGTGAGGTATGGTAGGTAGCGGAAAGTTAGTTTATGT : 60
RT112   : TTTTTGGTTATTTTTTTGTTTTAGTGAGGTACGGTAGGTAGTGGANAGTTAGTTTATGT : 60
T24     : TTTTNGGTTATTTTTTCGTTTTAGCGAGGTACGGTAGGTAGCGGAAAGTTAGTTTATGT : 60

```

```

          *           80           *           100          *           120
AC083800 : GAGGTGCCTCCTGCCAATCACAGACTACCTTCCCTGGTCTGGAGGTTCAAAGAATTGC : 120
COL-1    : GAGGTGTTTTTTGTTAATTATAGATTATTTTTTTTTTGGTTTTGGAGGTTAAAGAATTGT : 120
SC-N     : GAGGTGTTTTTTGTTAATTATAGATTATTTTTTTTTTGGTTTTGGAGGTTAAAGAATTGT : 120
TE-N     : GAGGTGTTTTTTGTTAATTATAGATTATTTTTTTTTTGGTTTTGGAGGTTAAAGAATTGT : 120
BR-T1    : GAGGTGTTTTTTGTTAATTATAGATTATTTTTTTTTTGGTTTTGGAGGTTAAAAAATNGT : 120
HA-T1    : GAGGTGTTTTTTGTTAATTATAGATTATTTTTTTTTTGGTTTTGNNGGTTAAAGAATTGT : 120
PE-T2    : GAGGTGTTTTTTGTTAATTATAGATTATTTTTTTTTTGGTTTTGGAGGTTAAAGAATTGT : 120
RO-T1    : GAGGTGTTTTTTGTTAATTATAGATTATTTTTTTTTTGGTTTTGGAGGTTAAANAATTGT : 120
RT112   : GAGGTGTTTTTTGTTAATTATAGATTATTTTTTTTTTGGTTTTGGAGGTTAAAGAATTGT : 120
T24     : GAGGTGTTTTTTGTTAATTATAGATTATTTTTTTTTTGGTTTTGGAGGTTAAAGAATTGT : 120

```

```

          *           140          *           160          *           180
AC083800 : AGGAGGGTAGAAAAGCACCTGGGTGGGTGCAGACTGCGGAGCGGGCCCTACCGTGTGCC : 180
COL-1    : AGGAGGGTAAAAAGTATTTGGGTTGGGTGTAAATTGCGGAGNGGGTTTTATCGTGTCC : 180
SC-N     : AGGAGGGTAAAAAGTATTTGGNTNGGTTGTNAAATTGCGGAGNGGNTTTTTATCGTGTCC : 180
TE-N     : AGGAGGGTAAAAAGTATTTGGGTNGGGTGTAAATTGCGGAGCGGGTTTTATCGTGTCC : 180
BR-T1    : AGGAGGGTAAAAAGTATTTGGGTNGGGTGTAAATTGCGGAGNGGGTTTTATCGTGTCC : 180
HA-T1    : AGGAGGGTAAAAACAGTATTTGGNTCGGGTGTAAATNGCGGAGCGGGTTTTATCGTGTGN : 180
PE-T2    : AGGAGGGTNAAANAGNATTTGGGTNGGGTGTAAATTGCGGAGNGGGTTTTATCGTGTGN : 180
RO-T1    : NGGNGGNTNAAAANGTATTTGGGTCCGGTGTANATTGCNGNGTGGGNTTTATCGTGTNC : 180
RT112   : AGGAGGGTAGAAAAGTATTTGGGTTGGGTGTAAATTGTGGAGTGGGTTTTATCGTGTGT : 180
T24     : AGGAGGGTAGAAAAGTATTTGGGTCGGGTGTAGATTGCGGAGCGGGTTTTATCGTGTCC : 180

```

```

          *           200          *           220
AC083800 : CAGAAAGAGGAGGCGCTGTGAGTAGACCGCTTCCCTCTG : 220
COL-1    : TAAAAAGAGGAGGCGTTGTGAGTAAATCGTTTTTTTTTTNG : 220
SC-N     : TAAAAAGAGGAGGCGTTGTGACTAGNTCGTTTTTTTTTTGG : 220
TE-N     : TAGAAAGAGGAGGCGTTGTGAGTAGATCGTTTTTTTTTTGG : 220
BR-T1    : TAGAAAGAGGAGGCGTTGTGAGTANATCGTTTTTTTTTTNGN : 220
HA-T1    : TAGAAAGAGGAGGCNTTGTGAGTAGATCGTTTTTTTTTTNG : 220
PE-T2    : TAGAAAGAGGAGNCGTTGTGAGTAGATCGTTTTTTTTTTTG : 220
RO-T1    : TAAAAAGAGGAGGCGTTGTGAGTAAATCGTTTTTTTTTTTG : 220
RT112   : TAANANGAGGAGGTCTTGTGAGTAGATNGTTTTTTTTTTNG : 220
T24     : TAGAAAGAGGAGGCGTTGTGAGTAGATCGTTTTTTTTTTNG : 220

```

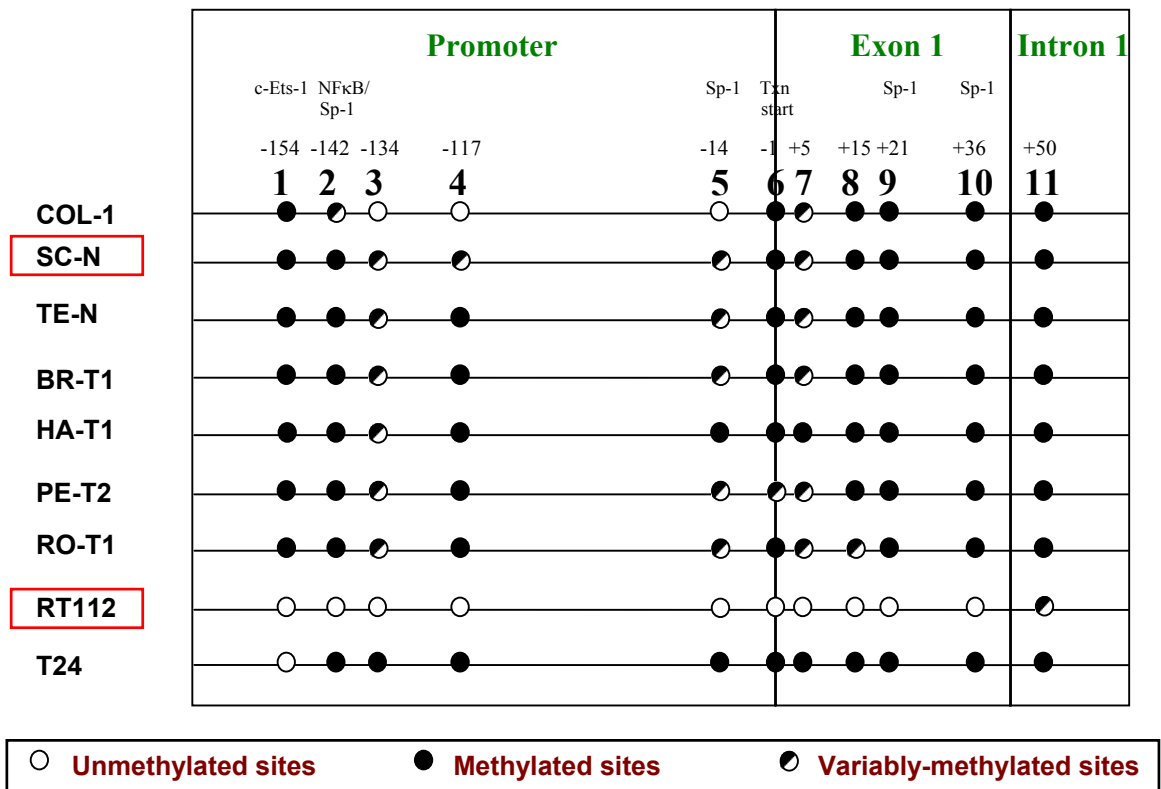


Figure 5.13

Direct Sequencing of PCR Amplified Bisulphite-modified DNA in Patient Samples

CpG residues are numbered above, with corresponding promoter, exon 1 and intron 1 sites for *UPKIB*. These results reflect data from the multiple sequence alignments (Figure 5.12). Shown are patient samples for normal colon COL-1, normal urothelium SC-N and TE-N, TCC patient samples BR-T1, HA-T1, PE-T2, RO-T1, and transitional cell carcinoma cell lines RT112 and T24.

Open circles denote unmethylated sites, enclosed circles denote methylated sites and half circles denote variably methylated sites. Red boxes indicate tissues or cell lines expressing *UPKIB* transcripts, as shown by RT-PCR analysis (Fig. 5.8 & 5.9A).

This ambiguity was likely due to the heterogeneous nature of the tumours, which contained both urothelium and surrounding lamina propria or muscle, where each tissue type would be expected to have different methylation patterns. To clarify this issue, a selection of PCR products were cloned and sequenced.

5.3.4.5 Cloning, direct sequencing and expression of UPKIB in tissues

PCR products (327bp) from four of the eight samples of TCC with a range of stages and grades were cloned into pGEM-Teasy (Promega) (Section 5.2.3.9). At least ten individual clones from each patient were sequenced in both directions with vector-specific primers SP6 and T7, to reveal methylation patterns in various cells (Mrs Lefta Leonardos) (Fig. 5.14).

Normal urothelial samples TE-N and SC-N, normal colonic epithelial sample COL-2 and peripheral blood lymphocyte sample PBL-7, from a normal individual (Fig. 5.14) showed quite variable methylation patterns. In all cases, CpG site 5, at a putative Sp-1 binding motif, appeared largely unmethylated. Both normal urothelial samples were methylated at CpG dinucleotides 7, 9, 10 and 11 at the 3' end of the CpG island, and at CpG site 1 at the 5' end. For sample SC-N, CpG dinucleotides 2-5 were unmethylated, and between 3-8 for sample TE-N. COL-2 and PBL-7 samples were fully methylated at most 3' CpG dinucleotides and at most 5' CpG sites. The PBL-7 sample was variably methylated at CpG sites 1, 3, 4 and 5. The main variation of methylation between all of these 4 samples is the methylation status of the putative NFκB and Sp-1 binding site, at CpG site 2. RT-PCR analysis of RNA obtained from micro-dissected frozen samples revealed expression of UPKIB mRNA only in sample SC-N among these 4 samples (Fig. 5.15). COL-2 and PBL-7 did not

Figure 5.14

Diagrammatic Representation of Sequencing Results of PCR Amplified and Cloned Bisulfite-modified DNA from Patient Tissue Samples

Shown are bisulfite-modified DNA obtained from samples of 2 normal urothelium (SC-N and TE-N), 1 colonic epithelium (COL-2), 1 peripheral blood lymphocyte (PBL-7) and 4 transitional cell carcinomas (RA-T1, HA-T2, HA-T3 and WA-T1). CpG residues are numbered above with putative transcription factor binding motifs and the presence of UPK1B mRNA expression is indicated by either (-) or (+) in the samples. Ten or twelve individual clones were sequenced for each patient.

Open circles denote unmethylated sites, enclosed circles denote methylated sites and half circles denote variably methylated sites in which there were equal cytosine and thymine peak heights. Cloned bisulfite-modified DNA products from tumour samples were sequenced in replicate sampling runs in both directions using primers T7 and SP6 from the vector pGEM-Teasy.

Detailed patient data is presented in Table 5.1.

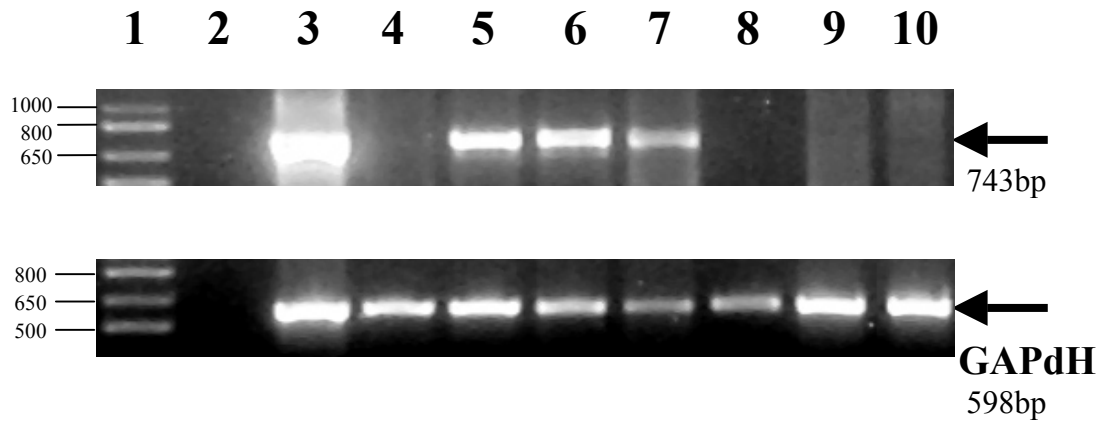


Figure 5.15

RT-PCR Analysis for UPKIB mRNA Expression in Patient Tissues and PBL

Lanes: (1) 500ng 1kb+ molecular weight marker; (2) water control; (3) HT1376; (4) TE-N; (5) SC-N; (6) RA-T1; (7) HA-T2; (8) HA-T3; (9) WA-T1; (10) PBL-7.

Top panel: PCR with UPKIB primers F1/R1

Lower panel: PCR with GAPdH primers GAPdH-5'/GAPdH-3'

express UPKIB mRNA (Figure 15, data not shown for COL-2), similarly to normal colon sample COL-1 and PBL samples 2-6 (Fig. 5.9).

Samples of a carcinoma *in situ* (Tis), a Grade I/II tumour, a T3 Grade II tumour and a T3 Grade III tumour were analysed for methylation of CpG sites (Fig. 5.14). Cloning and sequencing revealed a range of methylation profiles. As with normal samples and PBL-7, the 3' CpG sites were heavily methylated in all clones, as was CpG site 1, containing a putative c-Ets-1 binding motif. Interestingly, CpG site 5, containing a putative Sp-1 binding motif was unmethylated in the majority of clones, and did not appear to reflect differences in expression of UPKIB mRNA. RT-PCR on RNA obtained from these samples (Fig. 5.15) revealed expression of UPKIB mRNA in samples RA-T1 and HA-T2 but not in HA-T3 and WA-T1. This finding indicates that expression of UPKIB mRNA is maintained in less invasive tumours of carcinoma *in situ* (Tis) and a Grade I/II tumour, with loss of expression of UPKIB mRNA in T3 Grade II and T3 Grade III tumours. Overall, examination of methylation status in normal tissues, PBL and TCC samples revealed that, for UPKIB mRNA expression, the putative NF κ B/Sp-1 binding site at CpG site 2 appeared to be involved. Lack of methylation of this CpG site correlates closely with expression of UPKIB mRNA. The exception was the TCC HA-T3, which did not conform to this principle. The methylation of certain critical CpGs in the UPKIB CpG island may represent an important event in the transcriptional repression of this gene in the progression of TCCs and possibly also in silencing the gene in non-expressing normal tissues. Methylation could therefore play a role in controlling tissue-specific expression of *UPKIB*.

5.3.5 5-Aza-2'deoxyctidine treatment for UPKIB mRNA re-expression in T24 and J82 cell lines

The previous data (Sections 5.3.2, 5.3.3 & 5.3.4) demonstrate that in TCC cell lines and in clinical samples of TCC, there appears to be a direct correlation between methylation status at a number of CpG sites in the proximal promoter and UPKIB mRNA expression. The findings support the initial hypothesis outlined in Figure 5.2. The question thus arises whether demethylation of the UPKIB CpG island in cell lines not expressing UPKIB mRNA will allow re-activation of expression of UPKIB mRNA. Cell lines T24 and J82 were previously shown by RT-PCR analysis to not express UPKIB mRNA (Fig. 5.8) and to be methylated at most CpG residues (Fig. 5.6). It was therefore important to determine whether UPKIB mRNA expression could be activated with treatment of these cell lines with the demethylating agent 5-Aza-2'deoxyctidine.

T24 and J82 cell lines were incubated for 48 hours in 5-Aza-2'deoxyctidine (0, 1, or 3 μ M) (Section 5.2.4.3) and at 48 and 72 hour time points after the second addition of fresh 5-AzaCdR, RNA was extracted from the cells and was analysed for UPKIB mRNA expression. The results of RT-PCR analysis (Fig. 5.16) show re-expression of UPKIB mRNA only in cell line T24 at the initial 48-hour time post-induction. However, there were no detectable levels of expression at 72 hours in either cell line.

5.4 Discussion

In this study it was hypothesised that methylation may be one mechanism of regulation of transcription of *UPKIB*, since the UPKIB proximal promoter contains a CpG island (Chapter 4). In the initial investigation in transitional cell carcinoma cell

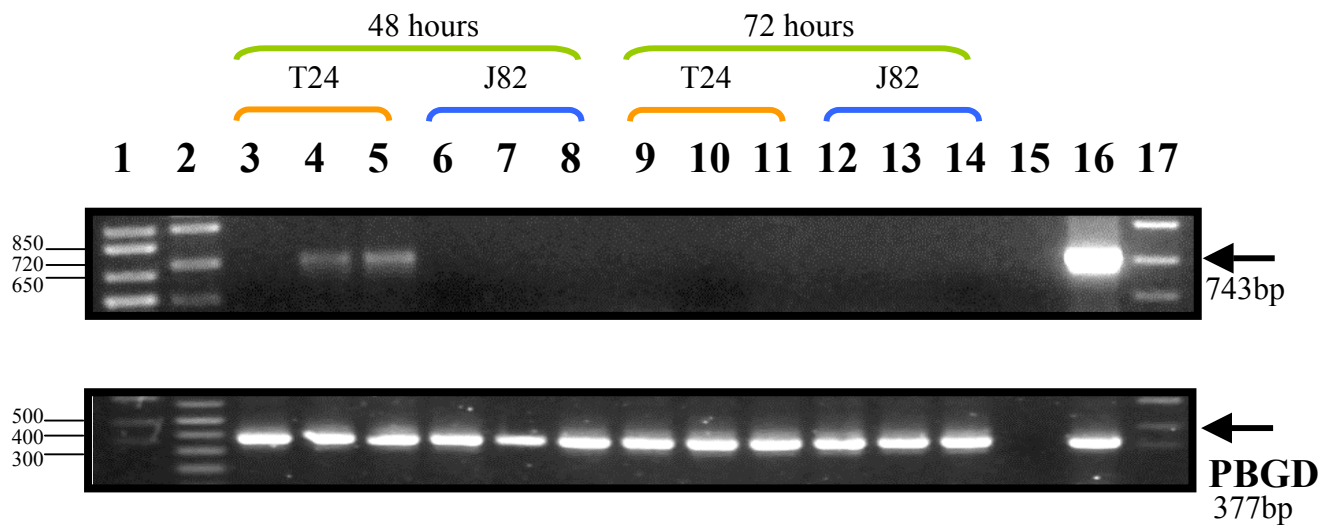


Figure 5.16

5-Aza-2'deoxyctidine Analysis in TCC Cell Lines T24 and J82: Re-induction of UPKIB mRNA Expression in T24 Cells

Top panel: PCR with F1/R1 primers

Lower panel: PCR with PBGD-5'/PBGD-3' primers

Lanes are labelled with cell lines and times of treatment

Lanes: (1) 500ng 1kb+ molecular weight marker; (2) 500ng SPP1/*EcoRI* molecular weight marker; (lanes 3-5 (48hrs) and 9-11 (472hrs)) T24 cell line; (lanes 6-8 (48hrs) and 12-14 (72hrs)) J82 cell line; (15) water control; (16) RT112 cell line; (17) 500ng SPP1/*EcoRI* molecular weight marker.

Treatments:

Lanes 3, 6, 9 and 12: 0 μ M 5-Aza-2'deoxyctidine

Lanes 4, 7, 10 and 13: 1 μ M 5-Aza-2'deoxyctidine

Lanes 5, 8, 11, and 14: 3 μ M 5-Aza-2'deoxyctidine

lines, it was found that RT112 and HT1376 cell lines expressed UPKIB mRNA transcripts. Direct sequencing of a 327bp PCR product containing the UPKIB CpG island (Chapter 4), amplified from bisulfite-modified DNA, revealed that DNA derived from RT112 and HT1376 cell lines was unmethylated at CpG sites 2 through 10 (Fig. 5.5). This 327bp region contained a putative NFκB/Sp-1 site, several Sp-1 sites and the transcription start site and exon 1 at CpG site 6. CpG sites in other TCC cell lines including T24, in SCC cell line Sca-BER and in peripheral blood sample PBL-1 were all heavily methylated throughout the region. However, cloning of PCR products generated from a second PBL sample, PBL-7 from a different volunteer (section 5.3.4.5) showed variable methylation at CpG sites 1, 2, 4 and 5, possibly showing that cloning revealed a more accurate picture of the methylation status of various cells within the PBL cell population. Analysis of cell lines and PBL revealed an inverse correlation between UPKIB mRNA expression and methylation of the CpG island in TCC and SCC cell lines and in PBLs. The PCR products from the cell lines were not cloned, as the sequencing chromatographs gave unequivocal data and the analysis of methylation status of different cell types was not an issue, as the cell lines were a relatively homogeneous source of DNA.

In this study, 8 TCC patient samples were analysed together with normal urothelium, normal colonic epithelium and PBL. RT-PCR showed expression of UPKIB mRNA in one normal urothelial sample, with no expression of UPKIB mRNA in 2 colonic epithelial samples and in 6 PBL samples. Yuasa and colleagues (1998) have previously reported a lack of expression of UPKIB mRNA in PBL from 3 normal individuals. Following cloning and direct sequencing, it was observed that the two normal urothelial samples had different methylation profiles and different mRNA expression. Both samples did however show CpG site 5 and CpG site 3 as

unmethylated. COL-2, shown not to express UPKIB mRNA, was heavily methylated except at the putative Sp-1 motif located directly before the transcription start site. The PBL-7 sample, also not expressing UPKIB mRNA, was also methylated or variably methylated at some sites, with the transcription start site and putative NF κ B/Sp-1 site both being fully methylated. The most methylated sites in the majority of cell line and tumour samples were clustered in the 3' region of the CpG islands, at CpG sites 8-11. These CpG sites incorporate sequence from exon 1 and intron 1 of the UPKIB island, implying that this region does not appear to play a role in the regulation of UPKIB mRNA expression.

It has been reported (Finch *et al.*, 1999) that up to 70% of TCCs have reduced expression of UPKIB mRNA and Lobban *et al.* (1998) reported reduced UPKIB mRNA expression which varied between 12% and 50% in non-invasive and invasive samples respectively (Section 1.4.4). Among the 4 tumour samples, Tis and Grade I/II tumours expressed UPKIB mRNA, and Grade II and Grade III tumours showed no UPKIB mRNA expression and methylation patterns appeared to reflect these differences. The more differentiated tumours expressing UPKIB mRNA were unmethylated at CpGs at both NF κ B/Sp-1 (site 2) and at the transcription start site (site 6) and less differentiated tumours with absent UPKIB mRNA expression were generally methylated at CpG site 2 and at the transcription start site (site 6). As mentioned previously (Section 5.3.4.5), HA-T3 was the exception to this rule: the HA-T3 sample was unmethylated at CpG site 2, yet showed no detectable levels of UPKIB transcript by RT-PCR, suggesting that other regions of the putative promoter might also be important in regulating expression of *UPKIB*.

CpG site 2 is contained within overlapping NF κ B and Sp-1 putative binding motifs. NF κ B is a nuclear transcription factor involved in the regulation of cell

proliferation, immune and inflammatory response and apoptosis. This transcription factor can activate genes including cytokines, growth factors and adhesion molecules (Review by Muller-Ladner *et al.*, 2002). Sp-1 (selective promoter factor 1) was the first identified transcription factor binding to and acting through these GC-rich elements (Dyran & Tjian, 1983). Sp-1 is a ubiquitous transcription factor which binds upstream of certain viral and cellular promoters and activates initiation of RNA synthesis by RNA polymerase II. It recognises GC boxes containing multiple copies of the sequence 5'-GGGGCGG-3' (section 4.3.4). Sp-1 expression varies between organs and during development (Saffer *et al.*, 1991). As discussed previously (Section 4.4), Sp-1 binding sites are seen in the promoters of tetraspanin members CD63, CD81, CD82 and CD9 (Nagira *et al.*, 1994; Le Naour *et al.*, 1996), suggesting that Sp-1 is an important transcription factor for the regulation of expression of these genes.

In the current study, SSCP analysis of samples, based on MS-SSCA, was not useful in the analysis of the methylation status in cell lines. PCR-amplified bisulfite-modified DNA from all of the cell lines, including RT112 and T24 appeared as similar bands in all SSCP gels. These band similarities were seen despite the various gel running procedures and PCR fragment sizes used to optimise the conditions and the marked differences in methylation status and sequences of the PCR products (Section 5.3.2.3, Fig. 5.7). The positions of the various bands did not support the direct sequencing data, which demonstrated that these sequences were fully unmethylated in RT112 and fully methylated in T24 at the CpG sites.

Restriction enzyme analysis was also considered in the initial design of these studies. The methylation-sensitive restriction enzyme *AciI* was the most suitable enzyme to digest methylated CpG sites in the 327bp bisulfite-treated PCR product.

DNA from cell lines, normal samples of urothelium, colonic epithelium and PBL and 18 samples of TCC was analysed by restriction digestion. DNA was bisulfite-treated and digested overnight with *AciI* (data not shown). Cell lines RT112 and HT1376 were not cut, showing their lack of methylation at the CpG sites. However, only partial digests were obtained for all other samples, even for the T24 cell line, which was methylated at all CpG sites, showing that this technique was neither informative nor effective in the analysis of methylation of this region. This result would suggest a possible heterogeneity in methylation patterns in the samples that may have been detected with cloning. Thus, it would appear that the technique originating from Frommer *et al.* (1992), based on bisulfite conversion and direct sequencing, is the most useful technique for analysis of UPKIB CpG residues, though more expensive than the other techniques due to sequencing costs.

In the 5-Aza-CdR experiments, TCC-derived cell lines T24 and J82 were analysed for re-expression of UPKIB transcript with demethylation of the genomic DNA by the agent 5-Aza-CdR. T24 cells re-expressed UPKIB mRNA 48 hours after the last 5-Aza-CdR treatment. However, expression of UPKIB mRNA had disappeared after 72 hours. Bender *et al.* (1999) demonstrated that p16 CpG islands could be re-methylated in T24 cells, even after treatment with 5-Aza-CdR, suggesting that this also occurs with the UPKIB CpG island in T24 cells. Loss of expression of UPKIB mRNA at 72 hours in T24 cells following detectable levels of expression at the 48 hour time point may also be due to the instability and breakdown of 5-Aza-CdR in culture (Covey & Zaharko, 1984).

The J82 cells did not express UPKIB mRNA at either time, or at any dosage of 5-Aza-CdR. Two studies have shown that the sensitivity of cells to 5-Aza-CdR is related to the endogenous levels of DNMT. It was reported by Jüttermann *et al.*

(1994) that mutant DNMT embryonic mouse stem cells and embryos with reduced DNMT levels were more resistant to 5-Aza-C. Ferguson *et al.* (1997) also showed that oestrogen receptor (ER)-negative cell lines Hs578t and MDA-MB-435 and ER-positive cell line MCF-7 contained only low levels of DNA methyltransferase and were more resistant to 5-Aza-C treatment than other breast cancer-derived cells lines. J82 may similarly have low levels of DNMT, which may have inhibited the activity of 5-Aza-CdR.

Thus, methylation appears to play a role in inhibiting expression of the UPKIB transcript, as the presence of the demethylating agent 5-Aza-CdR lead to the re-expression of UPKIB mRNA in T24 cells. Interestingly, there was no observable cytotoxic effect of 5-Aza-CdR on the cells. As detected by eye, cells grew viably and increasing 5-Aza-CdR concentrations only slowed the rate of cell growth.

In the T24 cells, UPKIB transcription in response to the addition of 5-Aza-2'deoxyctidine may also have been indirectly activated in several ways. These include the re-activation of an upstream transcription factor; the regulation of genes involved in a signal transduction pathway which may activate *UPKIB*, and non-specific changes in gene expression due to biological effects that occur as a result of an upstream gene re-activation event (Review by Karpf & Jones, 2002). In support of this, Liang *et al.* (2002) showed that approximately 60% of genes induced by 5-Aza-CdR did not contain CpG islands, so were activated indirectly.

In summary, UPKIB mRNA expression is reduced or absent in less differentiated tumours and appears to have a direct correlation with UPKIB CpG island methylation in bladder cancer-derived cell lines. Methylation of the NF κ B and transcription initiation CpG sites may inhibit binding of essential enhancers, leading to the subsequent down-regulation of UPKIB mRNA. From this finding, it is

surprising that several studies show increased levels of *UPK1B* mRNA with more advanced, less differentiated TCC. As methylation has been detected in older populations in a range of tissues (Section 5.1.2.2), one would expect that methylation of *UPK1B* may be one of several important determinants involved in bladder tumour progression.

Chapter 6

General Discussion and Future Directions

6.1 General Discussion

The preceding chapters report the characterisation of the uroplakin IB cDNA, genomic structure and putative promoter. These chapters report the analysis of the cDNA among various organisms, the genomic characterisation of the gene and the identification of the UPKIB promoter region and transcription start site and the regulation of gene expression through the mechanism of methylation of a 5' UPKIB CpG island.

A single partial human cDNA sequence containing the open reading frame was compared with the cDNA sequences of UPKIB in organisms of human, mink, cattle, rabbit, mouse and *Xenopus*, expressed in tissues including the bladder and lung. The human cDNA sequence was used to design primers to amplify genomic segments by long-template PCR to better define the genomic size of *UPKIB* by Southern hybridisation. However, this work was superseded by the advent of the human genome project, which led to the assembly of a contig containing the full *UPKIB* sequence. To identify the start of transcription the techniques of 5'RACE, primer extension and an Internet promoter prediction program were used. Potential mechanisms of regulation of *UPKIB* were studied by mining Internet databases, identifying the promoter sequence and the subsequent analysis of this region for potential transcription factor binding sites and for a possible UPKIB CpG island. Since a CpG island was identified, methylation analysis was employed to determine if expression of the UPKIB gene was regulated by CpG methylation.

6.1.1 Evolutionary studies of Uroplakin IB among the tetraspanins

A comparison of tetraspanin cDNA and protein sequences from a range of organisms including *Drosophila*, *Xenopus* and humans for UPKIB and for the other related tetraspanin family members revealed the closeness of the UPKIB subgroup within a single branch in the evolutionary trees. The *Xenopus* UPKIB 5' and 3' amino acid sequences had high sequence identity with the human homologue, at 71% and 69% respectively, and these sequences were also closely related to other mammalian UPKIB protein sequences. This suggests that, in the amphibian *Xenopus*, the mRNA required for the creation of the uroplakin proteins is presumably expressed in the urinary bladder and tracts as is the case in mammalian species.

From EST databases, the human uroplakins appear to be expressed in several tissue types including the brain, the eye, in skeletal muscle and in the stomach. However, a specific study has so far not been performed to screen for expression of uroplakin members in multiple tissue cDNA libraries. Microarrays may be a useful technique in the future to identify elevated or reduced levels of UPKIB transcripts in a range of human tissues.

There was no direct evolutionary relationship observed between UPKIB, UPKIA and other well-known members of the tetraspanins such as KAI-1 and CD9. Sequences have diverged to an extent that Maximum Parsimony and BAMBE analyses could not relate the uroplakin sequences to a primitive tetraspanin ancestor. This data is in contrast with Maecker *et al.* (1997) who inferred from their tree that the uroplakins were closely related to ROM-1 and to peripherins. This discrepancy appears to be related to methodological differences.

This study provides an example of how a specifically directed investigation of a group of proteins, the tetraspanins, can assemble largely random and poorly connected sequence data. It is predicted that this type of analysis will occur more frequently in the future. The data from the genomes of several organisms are now available, but will only be utilised when the needs arise.

6.1.2 Analysis of the UPKIB promoter

In the initial chapter, it was postulated that an investigation of UPKIB genomic and promoter sequences would reveal possible mechanisms of regulation of expression of UPKIB in normal cells and altered regulation in cancer. Characterisation of a 1 000bp region including the UPKIB proximal promoter, exon 1 and a portion of intron 1 revealed a range of putative transcription factor binding sites. These included sites for MyoD, NF κ B, c-Ets-1 and Sp-1 (Figure 4.10), with transcription factors binding to these sites possibly involved in the *in vivo* function of the UPKIB gene. A CCAAT box was also identified, suggesting this box may function as the recruitment centre for transcription initiation factors in the UPKIB promoter.

Mink *TI-1*, a homologue of human *UPKIB*, was regulated by TGF β 1 (Kallin *et al.*, 1991) and it was also predicted that human *UPKIB* would similarly be regulated by TGF β 1. However the *TI-1* promoter has not been characterised, so there is no direct evidence of a TGF β 1 binding motif regulating *TI-1* mRNA expression. Studies presented here do not show any evidence for a putative TGF β 1 transcription factor binding motif within the human putative proximal UPKIB promoter. However, as reported by Olsburgh *et al.* (2002), there was a region with close sequence identity to a TGF β 1 inhibitory element (TIE) in this region (Section 4.4).

Another likelihood of TGF β 1 regulation is that there is a TGF β 1 binding motif upstream of the region investigated. The role of TGF β 1 remains ambiguous and it may not be a key factor in the expression of human UPKIB mRNA. Another difference may be that human and mink genes have alternate transcription factor binding domains in their respective promoters, as in the case of the UPKII promoter in humans and mice, which has 80% sequence identity (Zhang *et al.*, 2002).

6.1.3 Analysis of bladder cancer cell lines and patient samples for methylation of the UPKIB CpG island

A weak CpG island of approximate size 500bp, associated with the putative UPKIB promoter, was identified in the current study that covered the UPKIB putative promoter, exon 1 and intron 1. An inverse correlation was observed between UPKIB expression and methylation status of the UPKIB CpG island. Of 7 TCC cell lines examined, methylation was observed at the majority of CpG sites in DNA prepared from 5 cell lines which did not express UPKIB mRNA. In the 2 cell lines that did express UPKIB mRNA, HT1376 and RT112, there was markedly less methylation throughout the CpG island. DNA from normal peripheral blood, colon and SCC cell line Sca-BER also showed no UPKIB mRNA expression and these samples were heavily methylated at the CpG sites. The precise degree of methylation in primary transitional cell carcinoma samples was less clear, due to the heterogeneity of the tissue collected from the patients. Sequencing of individual cloned PCR products revealed heterogeneity of methylation at several key CpG sites in the TCC tissues, including at a putative NF κ B/Sp-1 binding site.

Methylation is observed in several cancer types including colorectal carcinomas and is also generally increased in normal cells of older persons (Section

5.1.2.2). Increases in methylation in normal urothelial tissues (Bornman *et al.*, 2001) may trigger a reduction in expression of a range of genes including *UPKIB*. The question arises whether the methylation profiles for various genes are specific for certain tumour types or whether the process of methylation changes the expression of an array of genes similarly in any particular tissue type, by silencing their expression in a general fashion. The *UPKIB* promoter is methylated in other non-urothelial cells where there is no mRNA expression, as in PBL cells and in colonic epithelial cells. Invasive TCCs contain *UPKIB*-expressing urothelial cells and other non-urothelial cells including those of the *muscularis* and *lamina propria*, producing a heterogeneous cluster of cells with a mixed methylation profile and variable patterns of expression of *UPKIB*. This variety of *UPKIB* methylation profiles in the TCCs may not necessarily reflect changes in the bladder cancer cells themselves, but may mirror the combined methylation profiles of all cell types in the heterogeneous samples. The combination of both increased tissue heterogeneity and increases in methylation in cancers in elderly populations would alter the overall methylation patterns throughout more advanced cancers, and *UPKIB* expression in the TCCs is likely to be reduced as a result. However, this decrease in *UPKIB* expression and increase in methylation of the CpG island may not directly correlate with the development of bladder cancers.

6.2 Future Directions

Much of the work presented in this thesis was derived from mining Internet based programs and takes advantage of the rapidly expanding data derived from the human genome project. However, as seen in this thesis, analysis of the genomic sequences is still necessary to divulge the genomic structure of these genes and their

regulatory regions. It is only with a specific interest in particular genes that genes will be characterised and further studied.

In this thesis, analysis of the non-human EST databases has shown UPKIB to be expressed in amphibians, but not in insects. An interesting study would be to see which tetraspanin members are found in insect waste secretion. This may include the analysis of *Drosophila melanogaster* malpighian tubules, to examine the expression of tetraspanins other than UPKIB in *Drosophila*, and to see which tetraspanins are similar to UPKIB in the amphibian bladder. This may reveal a tetraspanin that has evolved into UPKIB in mammals, with similar structural and functional roles in waste excretion.

To further study the bladder and the involvement of other tetraspanins in the development and progression of TCCs, an initial approach may also include an analysis of the tetraspanin UPKIA, which could theoretically, like UPKIB, be regulated by methylation. This would involve searching the UPKIA promoter for a possible CpG island and subsequent bisulfite analysis in cell lines and patient samples. It is of interest to explore methylation further in bladder cancer, to determine the role of tetraspanins uroplakins IA and IB in TCC disease progression, and to explore the possibility that these uroplakins interact with proteins including integrins and other integral membrane proteins. There is also the possibility that uroplakins IA and IB act as metastasis suppressors, similarly to the tetraspanins CD9 and KAI-1.

It may be some time before all the genes within the human genome are characterised and the methods of their regulation identified, in the field of functional genomics or "transcriptomics". Assays will be required to determine which

regulatory motifs are important in expression of the gene. One important technique important in the understanding of the function of genes in both normal cells and within tumours is microarray technology. In mice, gene knockouts of UPK1B would lead to loss of expression of *UPK1B*. Phenotypic changes in mice lacking expression of UPK1B would be displayed, possibly involving the bladder and other tissues. With the use of microarray technology, the mRNA expression profiles of other genes within the mutant cells would be revealed; up- or down-regulated transcripts may show the involvement of UPK1B in signal transduction pathways. Using TCC cell lines such as RT112, expressing UPK1B transcript and T24, not expressing UPK1B transcript, microarrays could reveal the expression profiles of various genes in different grades and stages of TCCs, including analysis of UPK1B transcripts. Microarray technology can classify tumours based on gene expression profiles and could suggest diagnostic or prognostic markers within tumours (Alizadeh *et al.*, 2001; DeRisi *et al.*, 1996; Rew, 2001).

An alternative UPK1B promoter region was published by Olsburgh *et al.* (2002), where a luciferase reporter gene showed increased activity when expression was induced by a 235bp UPK1B promoter sequence. In this study, the 235bp region (-196 to +39) covered the CpG sites 1-10 analysed in the current study, including the putative proximal promoter and transcription start site. There are several strategies that could be employed to further characterise this region, with a focus on analysis of the putative NF κ B/Sp-1 motif at CpG site 2. One strategy involves site-directed mutagenesis of the putative NF κ B/Sp-1 binding region, to inhibit the binding of transcription factors Sp-1 or NF κ B. A second study may involve electrophoretic mobility shift assays (EMSA) or supershift assays using nuclear extracts of TCC cell lines such as RT112 and T24, which have different UPK1B expression profiles. The

nuclear extracts would be probed with radiolabelled DNA from the UPKIB NFκB/Sp-1 region, or with radiolabelled Sp-1 or NFκB consensus oligonucleotides, and analysed on polyacrylamide gels. Supershifts would involve using antibodies directed against specific nuclear extracts of Sp-1 or NFκB, which would be shifted further on polyacrylamide gels if the proteins bound to the radiolabelled DNA. It could be predicted that the RT112 cell line, which is unmethylated at CpG sites 1-10 and which expresses UPKIB transcripts, would bind NFκB or Sp-1 protein at the relevant promoter sequence. However, the T24 cell line which is methylated at CpG site 2 would be predicted to have no transcription factors bound to the promoter at this site. To study the possible binding of these two transcription factors and their possible co-operative roles in the activation of the UPKIB gene, the UPKIB promoter could be cloned into a luciferase reporter construct and co-transfected into TCC cell lines with Sp-1 and NFκB expressing plasmids.

Uroplakin mRNA is expressed in urothelial cells of the bladder (Sections 1.3.4) and this property may allow gene therapy approaches using the promoters from their respective genes to treat TCCs. The uroplakin II promoter has been used to induce expression of the lacZ reporter gene in urothelial cells of transgenic mice (Lin *et al.*, 1995). A recent study by Zhang *et al.* (2002) demonstrated that the use of the uroplakin II promoter attached to E1A and E1B (Δ 19kb) genes in the adenovirus variant CG8840 could target urothelial cells and hinder the spread of TCCs. However, as described in the Olsburgh *et al.* (2003) paper (Section 1.3.4), UPKII may also be expressed in the trachea, as shown in their masterblot. It was not reported by Zhang *et al.* (2002) if the UPKII-adenoviral vector was expressed in normal trachea, and no adverse effects were reported in the mice.

The UPK1B promoter could also be used to target gene therapy to urothelial tumours, however UPK1B mRNA has been detected in multiple tissues, including the trachea, placenta and kidney (Section 1.3.4), so this form of gene therapy may target areas not specifically associated with urothelium and TCCs. Further studies would be required to compare the promoters and expression patterns of the uroplakins members and to determine uroplakin 1B promoter sequences essential for urothelial-specific expression.