Chapter 1

Literature Review

1.1 General Introduction

A balance between proliferation, growth-arrest and programmed cell death (apoptosis) regulates cellular homeostasis. Cancer encompasses a group of diseases characterised by unregulated cellular proliferation, which arise from defective regulators of the cell cycle. These regulators include tumour suppressor genes, proto-oncogenes, and growth factors (Reviews by Sandal, 2002; Ho & Dowdy, 2002). Transitional cell carcinomas (TCCs) are cancers arising from transitional epithelial cells, also known as urothelial cells, which lie adjacent to the lumen in the urinary system (Section 1.3.2). Four major structural proteins termed uroplakins (UPKs) are located in the asymmetric urothelial cell membranes of many mammalian organisms and have been shown to form hexameric complexes covering the luminal surface of the bladder (Wu *et al*., 1990, 1994; Yu *et al*., 1990, 1994). Originally isolated from cattle bladder, these four uroplakin proteins are termed UPKIa (27kDa), Ib (28kDa), II (15kDa) and III (47kDa). The expression of mRNA from the mink homologue of UPKIb, TI-1, has been shown to be induced by transforming growth factor β (TGFβ) during growth arrest (Kallin *et al*., 1991). The uroplakins have also been identified in humans and human *uroplakin IB* (*UPKIB*) will be the continuing focus of this thesis.

1.2 Growth Arrest, TGFβ, and the Cloning of TI-1 in Mink Lung Epithelial Cells

Cells propagated in tissue culture can be synchronised from asynchronous cell populations by serum starvation (Pardee, 1989; Rollins & Stiles, 1989). Following serum starvation, cells express regulatory proteins such as transforming growth factor β (TGFβ) (Moses *et al*., 1991), a cell-cell signalling protein capable of acting

as either a negative regulator of the cell cycle or being involved in the maintenance of cellular homeostasis. There are three TGFβ family members in mammalian cells; TGFβ1, TGFβ2, and TGFβ3, of which TGFβ1 is the best-characterised (Reviewed by Clark & Coker, 1998; Hoodless & Wrana, 1998; Roberts, 2002).

TGFβ1 can induce expression of various growth-arrest-specific genes including *TI-1* (Kallin *et al*., 1991). Kallin and colleagues (1991) aimed to identify cDNA clones upregulated by TGFβ1 in an epithelial cell line CCL64, derived from mink (*Mustela vison*) lung. A differential cDNA screen was performed to detect mRNA upregulated by TGFβ1 by analysis of three cell populations. Following two days of serum starvation to induce growth arrest, human TGFβ1 was added to one cell population and not to another as one control. A second control comprised cells with only serum added. Up-regulated poly A+ mRNA, extracted from the serum-starved TGFβ1-induced cells, was used to construct cDNA libraries, and increased expression of mink TI-1 mRNA was detected.

The mink TI-1 protein shares 93% amino acid identity with cattle (*Bos taurus*) uroplakin IB (UPKIB) and 93% amino acid identity with human uroplakin IB (UPKIB) (Kallin *et al*., 1991; Yu *et al*., 1994; Finch *et al*., 1999). Due to the high sequence conservation of the cDNA coding for the open reading frames between human, mink and cattle sequences, it could be predicted that the genes may similarly be regulated, with TGFβ binding motifs in the promoter region. The involvement of TGFβ1 in the direct activation of *TI-1* and human *uroplakin IB* currently remains uncertain because the analogous experiments analysing the promoters of these genes and the binding of the TGFβ protein to the promoters have not been performed.

1.3 Bladder Structure and Function

1.3.1 Diversification of waste excretion in a range of organisms, the urinary system, and the role and function of the bladder

Waste products are produced in all organisms and a variety of methods are employed by these organisms for the removal of them from their systems. Waste products, including urea and breakdown products of blood such as heme and urobilins, are compounds no longer required for cellular function. Generally, organisms have bladders because of a need to store urine: a solution of some of their waste products, particular nitrogenous waste as urea, which is usually dilute.

Invertebrates, including the well-studied *Drosophila melanogaster*, do not possess a bladder. In these organisms, there is no real need for liquid storage. As in all terrestrial insects, *Drosophila* excretion involves a system of malpighian tubules, which excrete nitrogen waste as uric acid. The structure of these tubules has been well studied by Eichelberg and Wessing (1971) and Wessing and Eichelberg (1969, 1972a, 1972b) using electron microscopy. Fish and reptiles vary in bladder usage. Primitive fish contain no urinary bladder; however some fish, including salmon, contain a urinary bladder, although their need for liquid storage is minimal. Amphibians vary in their requirements for bladders: in the case of the genus *Xenopus,* there is a urinary bladder emptying into the cloaca. Birds generally have a cloaca but no bladder; their nitrogenous waste is composed largely of solid uric acid. The exception is the ostrich which surprisingly does contain a bladder (Kardong, 1995).

In humans, urine is formed in the kidneys and passes from the renal pelvis to the ureters and then to the bladder. At the base of the human bladder is a ring-like sphincter muscle, which inhibits spontaneous urine release. The bladder has a

resilience to accommodate up to half a litre of urine and to accommodate contraction and distension with the release and storage of the urine. The urine is then spasmodically removed from the body through the urethra (Fig. 1.1). In females, the urethra functions solely as a urinary passage; and in males the urethra is located within the penis and also serves as a vessel for semen during sexual intercourse. This process is similar in all other mammals.

1.3.2 Urothelial tissue and the bladder

The bladder and ureters are composed of 3 tissue types (Fig. 1.2A). The outermost layer comprises *muscularis* (muscle) and *adventitia* (blood vessels), the latter providing vascularisation of the bladder (Douglas & Hossler, 1995). The next layer is the *lamina propria* (loose connective tissue) and then the urothelium, also known as transitional epithelium or urinary epithelium, which lies adjacent to the lumen and is covered with a mucosal layer. Urothelial tissue lines the renal pelvis, the ureters, lumen of the bladder and neck of the urethra, all regions where the epithelial cell layer encounters urine (Fig. 1.1) (Burkitt, 1996; Review by Lewis, 2000).

The urothelium is approximately 4-6 cells thick and is composed of 3 cell types, each with varying states of cellular differentiation. The most undifferentiated cells are the basal cells, which are located adjacent to the *lamina propria*. The intermediate cells are the central cells and show a degree of differentiation. The most differentiated cells are the superficial umbrella cells. The urothelium must maintain flexibility to accommodate for frequent bladder contraction and distension with the retention and passing of urine. It achieves this flexibility through the

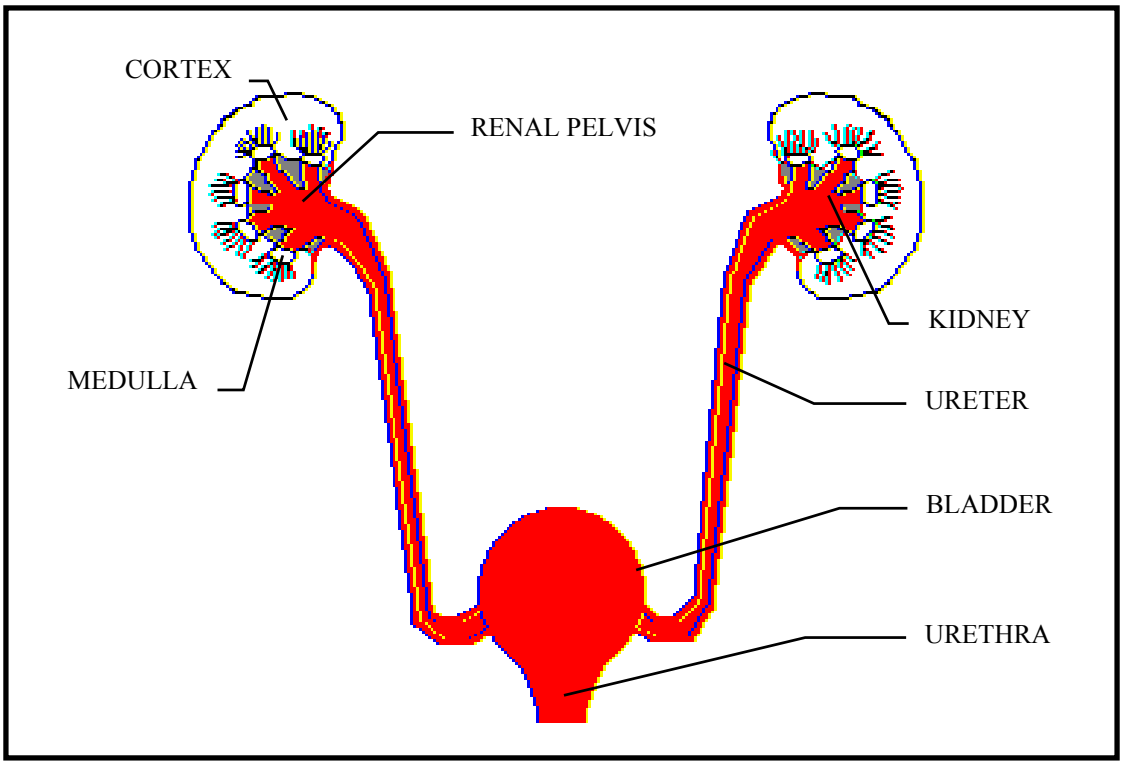


Figure 1.1

Location of Urothelial Tissues in the Body

The major regions containing urothelium are illustrated. These include the renal pelvis, bladder, ureters and the neck of the urethra.

The kidney comprises the outer cortex, containing glomeruli, the medulla (collecting ducts) and the renal pelvis. It is at the renal pelvis that urine is drained from the medulla before passing through the ureters into the bladder. From here, urine is removed from the body through the urethra. Urothelial tissue is highlighted in the diagram in red.

Figure adapted from Vander *et al*. (1986).

Figure 1.2

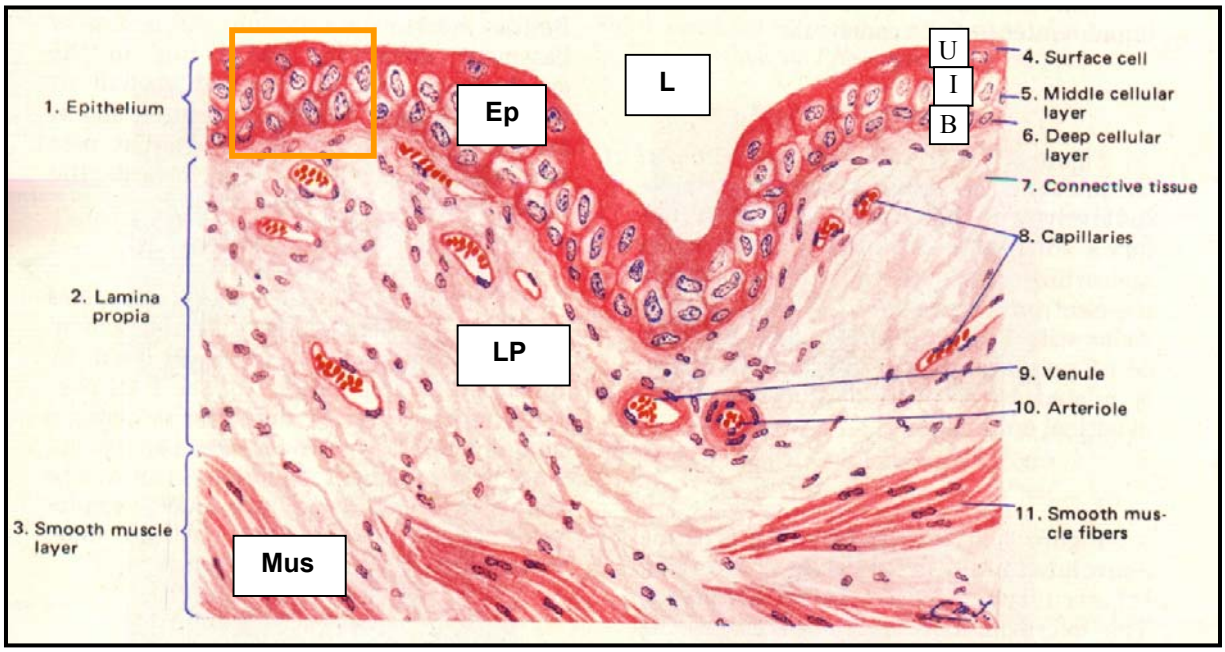
Histological Appearance of the Bladder

A. A haematoxylin and eosin-stained cross-section of the bladder wall, at 300x amplification. **B.** A haematoxylin and eosin-stained enlarged section of urothelial cells and *lamina propria* (LP) at 320x amplification, from a similar location to the boxed section in orange in part **A**. Some umbrella cells are binucleate.

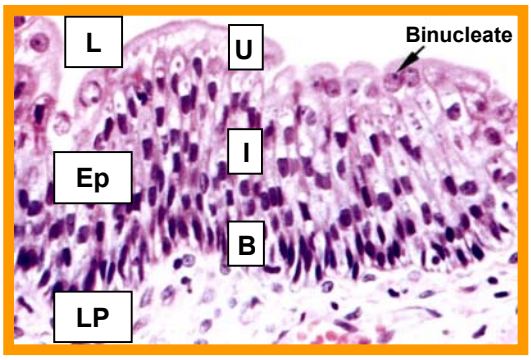
L	Lumen
Ep	Urothelium/transitional epithelium
LP	*Lamina propria*
Mus	*Muscularis*
U	Umbrella cells
I	Intermediate cells
B	Basal cells

Figures modified from di Fiore *et al*. (1978) (**A**) and Young & Heath (2000) (**B**).

A


1. Epithelium
2. Lamina propia
3. Smooth muscle layer
Ep
L
U
I
B
LP
Mus
4. Surface cell
5. Middle cellular layer
6. Deep cellular layer
7. Connective tissue
8. Capillaries
9. Venule
10. Arteriole
11. Smooth muscle fibers

B


L
U
Binucleate
I
Ep
B
LP

umbrella cells located closest to the lumen (Fig. 1.2B), which contain an asymmetric unit membrane (Section 1.3.3).

1.3.3 The Asymmetric Unit Membrane (AUM), plaques, and uroplakins

Umbrella cells are the most differentiated cells in the urothelium and must be flexible to deal with bladder distension and contraction; to enable this, they contain a highly specialised membrane known as the Asymmetric Unit Membrane (AUM). The AUM consists of an 8nm luminal leaflet and a 4nm cytoplasmic leaflet, giving the asymmetric appearance (Hicks, 1965; Nesci & Tessitore, 1969). The AUM contains tightly packed protein particles known as plaques, which are thought to strengthen and stabilise the urothelial luminal surface, along with other proteins such as urohingins, which are asymmetrically distributed along the luminal plasma membrane (Yu *et al*., 1990, 1992; Wu *et al*., 1990). The plaques are readily seen in electron microscopic images of the lumen of the bladder (Fig. 1.3) (Hicks & Ketterer, 1969; Wu *et al*., 1994). The plaques are ordered as hexagonal 12nm subunits with a centre-to-centre spacing of 16.5nm, organised into well-ordered two-dimensional crystalline arrays which cover over 70-90% of the urothelial apical surface (Chang *et al*., 1994).

The uroplakins are the main structural constituents of the urothelial plaques within the AUM, as demonstrated using antibodies directed against the uroplakin proteins (Yu *et al*., 1990, 1994; Wu *et al*., 1990, 1994). Wu and colleagues (1994) isolated AUMs from 9 mammalian species including human, cattle, monkey, sheep, pig, dog, rabbit, rat, and mouse to study the plaques and the role of the uroplakins in plaque formation, and electron microscopic analysis revealed all AUMs to be

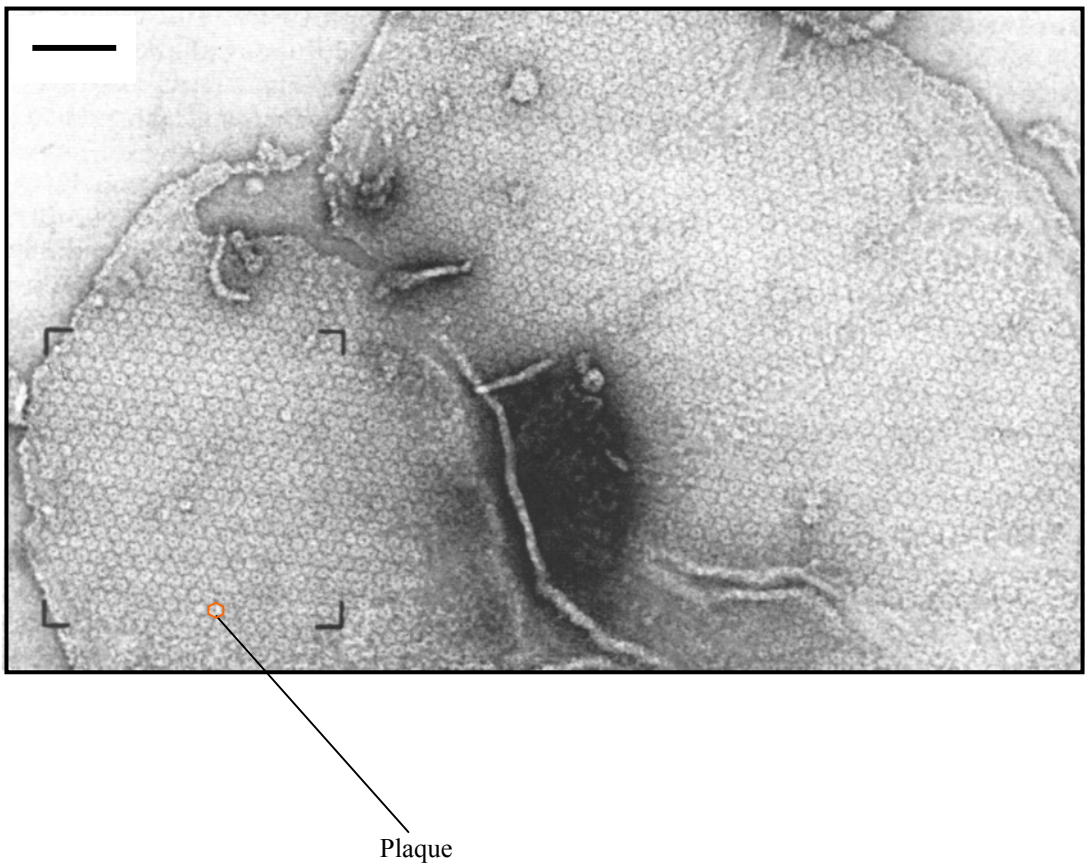


Figure 1.3

Electron Microscopic Image of Plaques on the Asymmetric Unit Membrane in Cattle

Uranyl-formate (UF)-stained asymmetric unit membrane showing tightly packed hexameric complexes on the luminal surface of urothelial tissue. Each 12nm complex is termed a 'plaque', and these are arranged in a two-dimensional crystalline lattice with centre-to-centre spacing of 16.5nm. These plaques are composed of four uroplakins. Scale: bar represents 100nm.

Figure adapted from Walz *et al*. (1995).

morphologically similar. The conservation of morphology between the species underlies the importance of the AUM in the bladder.

Uroplakins IA and IB are members of the tetraspanin family of proteins, so-called because they span the AUM four times and are characterised by four highly conserved transmembrane regions and a large second extracellular domain (discussed in more detail in Section 1.5). Uroplakins II and III are largely extracellular, single spanning membrane proteins with short cytoplasmic tails (Yu *et al*., 1990; Wu & Sun, 1993).

Wu and colleagues (1995) studied the interactions of the uroplakins using chemical cross-linking techniques and established that plaque assembly was not a random event. AUM proteins were isolated and incubated with cross-linking reagents including EGS (ethylene glycol bis-(succinimidylsuccinate)): EGS reacts with amino groups (NH_2), as located on the amino acid lysine. Products were then dissolved in SDS and the proteins analysed by 2-dimensional SDS-PAGE and Western blotting. Western blotting results of the cross-linked AUM proteins indicated specific binding between uroplakins IA and II, uroplakins IB and III and homodimerisation of uroplakin II. The complexes were assembled in domains in a model proposed by Wu and colleagues (1995) (Fig. 1.4). In these models, the uroplakins were ordered with either (i) UPKII proteins in the outer domains and UPKIA proteins in the inner domains (Fig. 1.4A) or (ii) UPKIII proteins in the outer domains and UPKIB in the inner domains (Fig. 1.4B), showing the specificity of plaque assembly.

Sun *et al*. (1996) proposed a model explaining the elasticity of the umbrella cells. In this model, the urothelial plaques are displayed on the surface of the lumen during bladder distension and are pulled into the cellular cytoskeleton during bladder

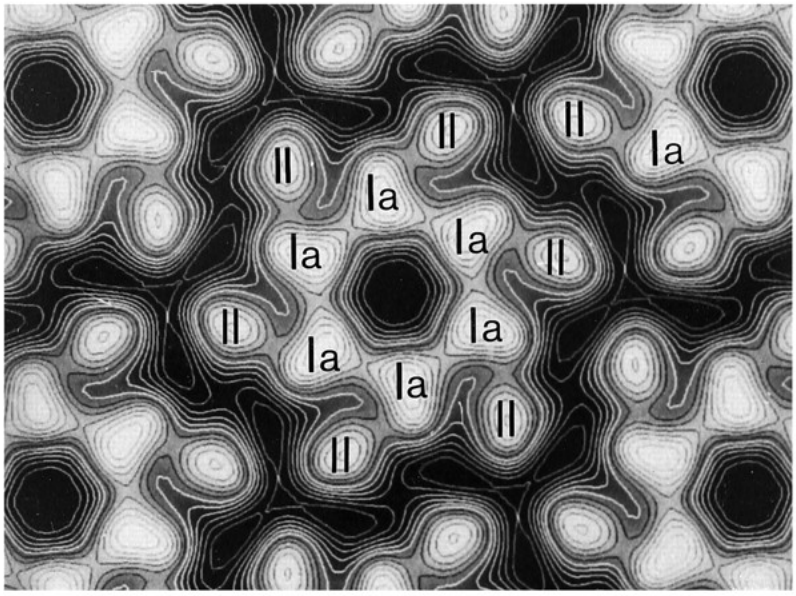
Figure 1.4

Proposed Model of Plaque Assembly in Cattle

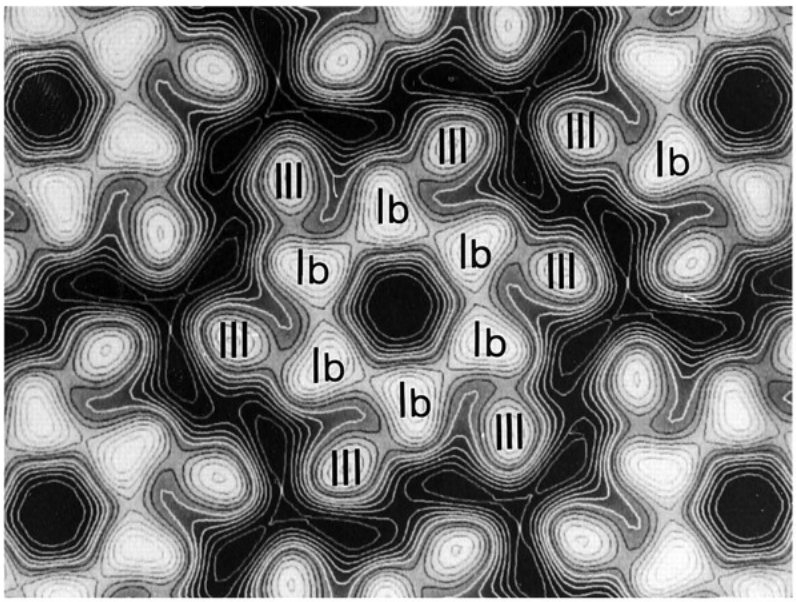
Model showing the possibility of two types of plaques along the asymmetric unit membrane of the cattle umbrella cells. The tetraspanins, uroplakins IA and IB, are thought to interact with the single transmembrane proteins uroplakins II and III respectively, in plaque complexes. **A.** Putative interactions between UPKIA and UPKII. **B.** Putative interactions between UPKIB and UPKIII. These plaque complexes are likely to be randomly distributed throughout the asymmetric unit membrane (AUM).

Figure taken from Wu *et al*. (1995).

A


II
II
II
Ia
Ia
Ia
Ia
II
II
Ia
Ia
Ia
II
II

B


III
III
III
Ib
Ib
Ib
Ib
III
III
Ib
Ib
Ib
III
III

contraction. It is proposed that the uroplakins, assembled in plaque complexes, interact with the underlying cytoskeleton, which contains keratin filaments. These filaments anchor the AUM to the cytoskeleton, stabilising the luminal surface to prevent cells from rupturing during bladder distension (Fig. 1.5) (Wu *et al*., 1994; Chlapowski *et al*., 1972; Staehelin *et al*., 1972; Minsky & Chlapowski, 1978; Sarikas & Chlapowski, 1986, 1989). The direct relationship of uroplakins with particular cytoskeletal proteins has not yet been conclusively identified, but may involve interactions with integrins (Sun *et al*., 1996).

Uroplakins IA and IB appear to be important in strengthening the plaque complexes, by spanning the AUM several times. Uroplakins IA and IB also have important immunological functions in the bladder, acting as receptors for the binding of type 1-fimbriated *Escherichia coli* that cause urinary tract infections. The interactions of type 1-fimbriated *Escherichia coli* with these uroplakins can be inhibited by the binding of urinary glycoproteins such as Tamm-Horsfall protein (THP) to the bacteria (Pak *et al*., 2001; Wu *et al*., 1996; Zhou *et al*., 2001). THP does not directly interact with urothelial cells, implying that THP also does not directly interact with uroplakins IA and IB (Fowler *et al*., 1987). Uroplakins IA and IB thus appear to facilitate urinary infections through their basic role as receptors.

1.3.4 Patterns of expression of uroplakins in non-urothelial tissues

The uroplakins have largely been studied in the bladder and using immunological approaches, uroplakins have been shown to be expressed in the bladder (Wu *et al*., 1990). In a study by Moll *et al*. (1995), UPKIII was detected in the human bladder but no UPKIII was detected in other normal human tissues including skin, mammary glands, oral mucosa, salivary glands, paranasal sinus

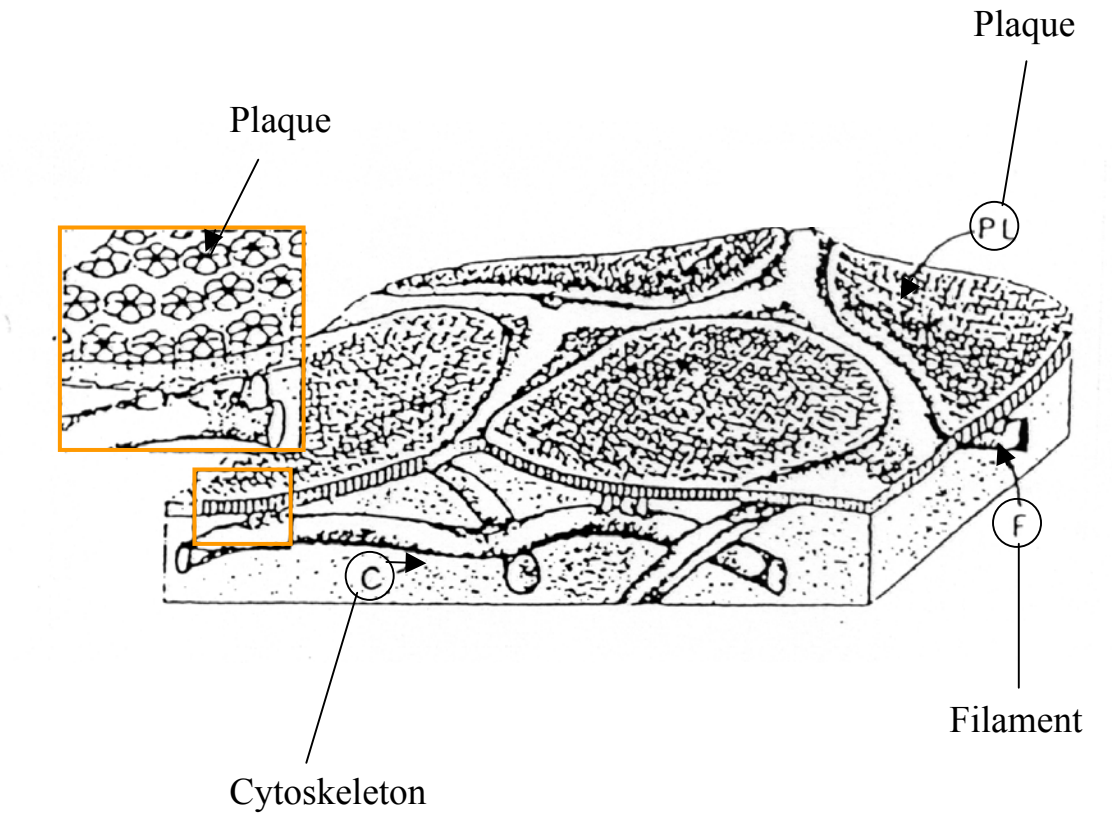


Figure 1.5

Plaque-AUM/Cytoskeletal Interactions in Superficial Umbrella Cells

Plaques (PL) interact with a supporting cytoskeleton (C), which also may contain keratin filaments (F). It has been proposed that AUM/cytoskeletal contact may help draw the plaques into the cytoskeleton during bladder contraction.

Figure adapted from Sun *et al*. (1996).

mucosa, lung, stomach, small and large intestines, vagina, glans penis, pancreas, ovary and peripheral nerves. The urothelial-specific localisation of the uroplakins was further demonstrated by Yuasa *et al*. (1999), who also failed to detect mRNAs coding for uroplakins IA and II in brain, kidney, benign prostate tissue, prostate adenocarcinoma and testis. However, Olsburgh *et al*. (2003) recently hybridised a radiolabelled UPKIB cDNA probe to a masterblot containing poly (A)+ RNA from 43 adult human normal tissues and demonstrated some heterogeneity of expression patterns. UPKIA mRNA was expressed solely in the bladder, UPKII mRNA was expressed in the bladder and trachea, UPKIII mRNA was expressed in both the bladder and prostate and UPKIB mRNA was expressed in the bladder, trachea, placenta, pancreas and kidney. However, analysis of uroplakin expression by *in situ* hybridisation in this study did not detect the presence of UPKII mRNA in the trachea, indicating that both UPKIA and UPKII may be urothelium-specific. The *in situ* hybridisation analysis also did not detect the presence of UPKIB mRNA expression in the pancreas, indicating the likelihood of *in situ* hybridisation being a less sensitive technique. This study supported the relatively widespread expression of UPK1B mRNA in normal human tissues.

The Unigene database (NCBI) has reported expressed sequence tags (ESTs) with a high degree of sequence identity to UPKIA in human foetal heart and adult prostate tissues (GenBank accessions AI633692 and AA548647). ESTs homologous to UPKII were detected in human brain (neuroectoderm) (GenBank accession BF382615) and in cattle kidney tissues (Takasuga *et al*., 2001; GenBank accessions AV600502 and AV600503). ESTs homologous to UPKIII were detected in human prostate (GenBank accession BF447256), foetal eyes (Bonaldo *et al*., 1996; GenBank accessions BM727658 and BM681088), mouse head (Konno *et al*., 2000; GenBank

accession BB525631), lung (Aizawa *et al*., 2000; GenBank accession BB605052) and mammary glands (Marra *et al*., 1996; GenBank accession AA797509). However, the presence of uroplakin-like ESTs does not necessarily infer that these transcripts are translated into uroplakin protein in these tissues.

Adachi *et al*. (2000) identified a human uroplakin IB (UPKIB) transcript in human ocular surface epithelium (GenBank accession AB002155). This transcript extended the previous UPKIB cDNA sequences published by Finch *et al*. (1999) (GenBank accession AF042331), Yuasa *et al*. (1998) (GenBank accession AB015234) and Lobban *et al*. (1998) (GenBank accession AF082888) (discussed in Section 4.1), by providing the complete 3' untranslated region (UTR) and more of the 5'UTR coding for the mRNA. This thesis has found several UPKIB ESTs through Internet-based programs, and these are discussed in Chapter 4.

1.4 Transitional Cell Carcinomas and Their Association with Uroplakin IB Expression

Transitional cell carcinomas (TCCs) are cancers originating in the urothelial cells of the bladder, ureter and renal pelvis. Their development involves a multistage process of invasion, with sequential steps including initiation, promotion, and progression.

1.4.1 Population incidence and associated risk factors

The incidence of bladder cancer in South Australia in the year 2000 involved 136 males (3.3% of all cancers diagnosed) and 59 females (1.7% of all cancers diagnosed) (South Australian Cancer Registry, 2001). The male:female incidence of TCC is between 3:1 and 4:1 in Asia and the English-speaking world (North America,

British Isles, and Australasia) and is usually greater than 5:1 in Europe (Reviewed by McCredie, 1994).

Tobacco is the main risk factor associated with TCCs in western countries. The use of tobacco is related to approximately half of bladder cancers in men and a third of those occurring in women (Hartge *et al*., 1990; Hogan *et al*., 2001; Wada *et al*., 2001). Occupational exposure to exogenous bladder carcinogens including aromatic amines and aniline dye intermediates such as beta-naphthylamine and benzidine, also leads to TCC development, with 21-25% of bladder cancers in white males in the United States attributed to occupational exposure (Review by Mason *et al*., 1992). Other important risk factors are age, with people over 40 years of age being most at risk, or the development of chronic urinary tract infections (Hartge *et al*., 1990).

Squamous cell carcinomas (SCCs) of the bladder, arising from cornified cells (Fig. 1.6) also demonstrate the same risk factors. SCCs are less than 10% of all bladder cancers in Westernised countries, with much higher rates of occurrence in Africa and the Middle East. In these latter countries, SCCs are associated with infestation of the bladder by *Schistosoma haematobium* and subsequent schistosomiasis (Gelfand *et al*., 1967; Rosin & Anwar, 1992; Amonkar *et al*., 2001).

1.4.2 Primitive urothelium differentiation, formation of TCCs and TCC stages

Primitive urothelium can differentiate into three pathways: urothelium-, epidermis-, or glandular-type pathways (Fig. 1.6). These pathways are characterised by the production of UPKs, K1/K10 keratins, and secreted glycoproteins, including prostate specific antigen and prostate specific marker, respectively (Yi *et al*., 1995; Reviews by Baskin *et al*., 1996, 1997; Wu *et al*., 1998). Wu and colleagues (1998)

Figure 1.6

Pathways of Urothelial Differentiation

PSA	Prostate specific antigen
PSM	Prostate specific marker
SCC	Squamous cell carcinoma
TCC	Transitional cell carcinoma
UGSM	Urogenital sinus mesenchyme

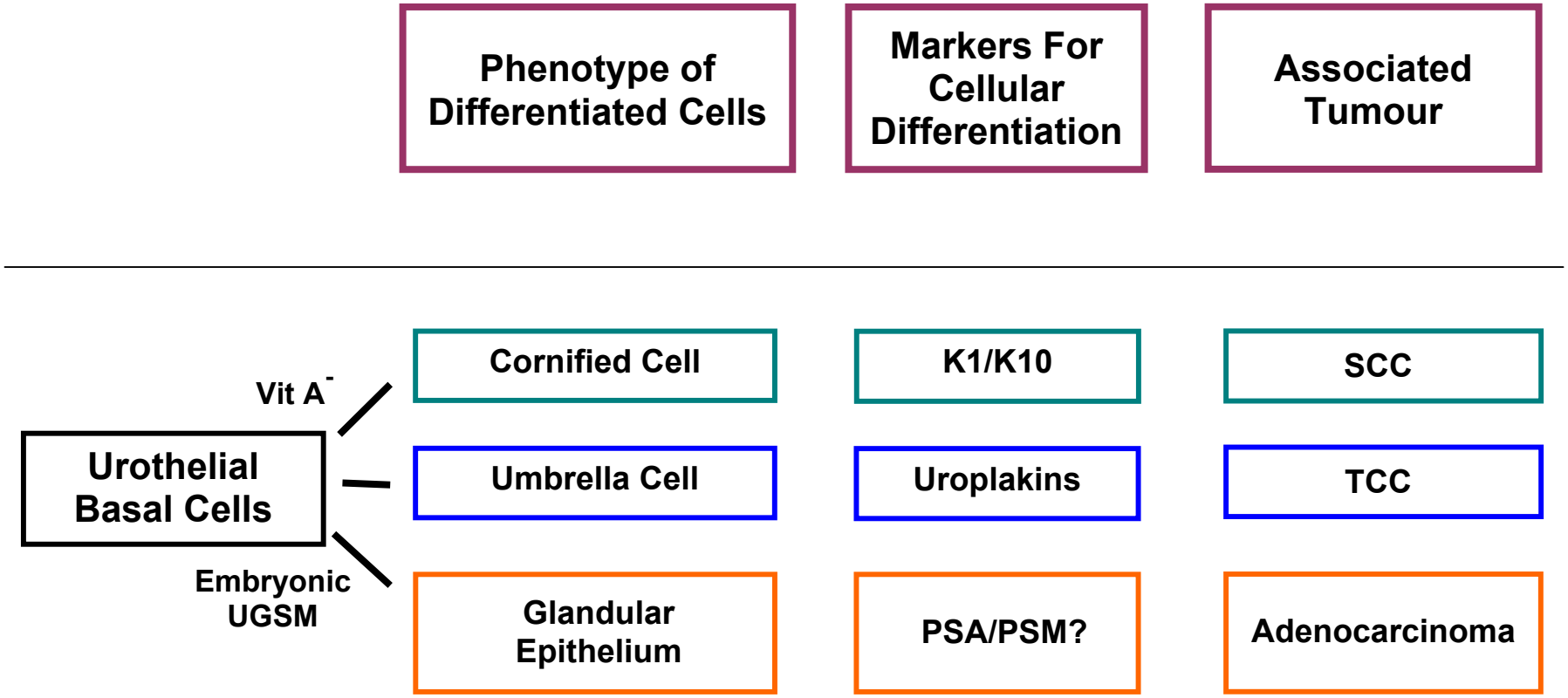
A common pool of pluripotent urothelial basal cells can differentiate through a variety of normal pathways or cancerous states. Deregulation of these cells can lead to TCC, SCC or adenocarcinoma respectively.

Vitamin A deficiency induces urothelial keratinisation, with the formation of cornified cell layers, which synthesise keratinisation markers K1 and K10. It has been proposed that vitamin A deficiency leads to the development of SCC (Wu *et al.,* 1998).

Uncontrolled urothelial growth can also lead to TCC. Normal urothelium expresses uroplakins as the main differentiation product and expression of these uroplakins has been detected in TCCs (Moll *et al*., 1995).

Embryonic urogenital sinus mesenchyme (UGSM) can induce *trans*-differentiation of normal postnatal urothelium into prostatic glandular epithelium (Cunha *et al*., 1983). Proteins expressed by this type of epithelium include prostate-specific epithelial markers PSM and PSA and mutations in these prostatic epithelial cells may form adenocarcinomas (Wu *et al.,* 1998).

Figure adapted from Wu *et al*. (1998).


Phenotype of Differentiated Cells
Markers For Cellular Differentiation
Associated Tumour
Urothelial Basal Cells
Vit A-
Embryonic UGSM
Cornified Cell
K1/K10
SCC
Umbrella Cell
Uroplakins
TCC
Glandular Epithelium
PSA/PSM?
Adenocarcinoma

suggested that urothelial malignancies including TCCs, SCCs and adenocarcinomas all arise from a common pool of undifferentiated basal cells of the urothelium (See comments on the legend to Fig. 1.6).

TCCs often present with multiple tumours, appearing at different times, and at different sites in the bladder (Sidransky *et al*., 1992). In the early development of TCCs, Tis (*in situ*) or Ta stages are either localised to the urothelium or have a slight outgrowth into the lumen respectively (Fig. 1.7). The differentiation between Tis and Ta stages is often difficult, due to similarities in chromosomal abnormalities (Fig. 1.8).

The T1 and T2 stages involve the invasion of the *lamina propria* (LP) and, in T2, can also involve early muscle invasion (Fig. 1.7). Advanced stages are designated T3 and T4, which comprise invasion of muscle by tumour cells, and N+/M+ that encompass distant lymph node and muscle metastases by the malignant cells. Less than 60% of patients with locally advanced bladder cancer (T2 onwards) will survive 3 years, despite aggressive, multimodal treatment with combinations of radiation, chemotherapy, and surgery (Logothetis *et al*., 1996).

1.4.3 Molecular genetics of TCCs

Normal cells can develop into malignant cells through a multistep process involving genetic and molecular alterations. Carcinogens, including chemicals within cigarettes (Section 1.4.1) can induce these changes. Initial chromosomal alterations in TCCs are associated with chromosome 9, suggesting the disruption of one or more candidate tumour suppressor genes in this region of the genome (Fig. 1.8). Alterations in chromosome 9 include homozygous deletions, microsatellite alterations and allelic loss (Nishiyama *et al*., 1999; Hartmann *et al*., 1999, 2002;

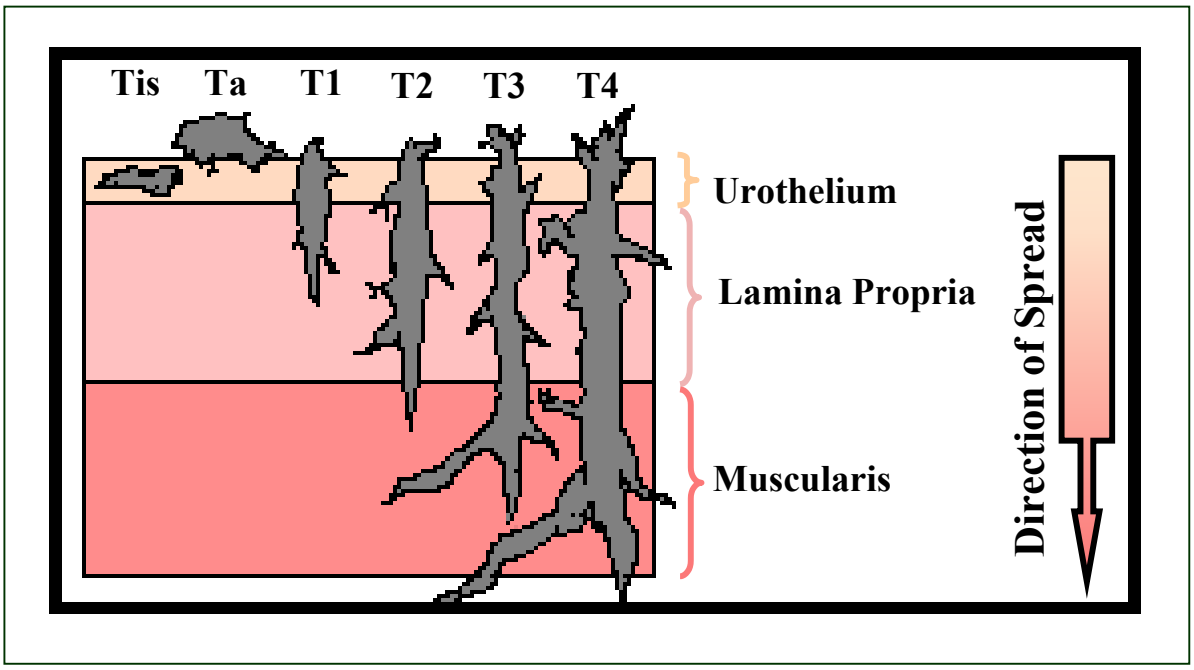


Figure 1.7

Representation of the Stages of Transitional Cell Carcinoma

Initially, cancers arise in the urothelium (Tis, carcinoma *in situ*) or bud into the lumen (Ta).

The cancer may subsequently spread into the *lamina propria* (T1) and muscle (T2). The cancer can spread further in stages T3 and T4 through local invasion before metastasis and the development of a secondary cancer in lymph nodes (N) and other muscle (M) throughout the body (N+/M+).

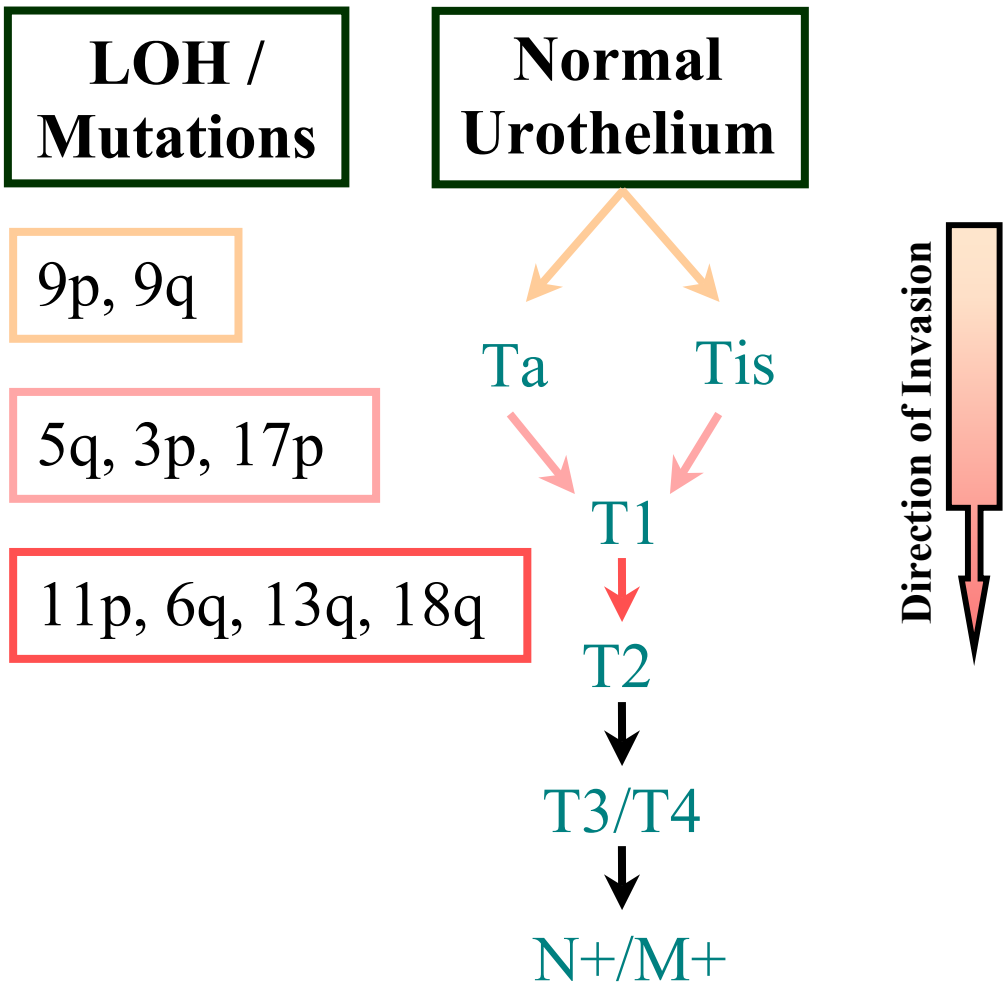
Figure adapted from Cordon-Cardo & Reuter (1997) and Dalbagni *et al*. (1993).

Figure 1.8

Genetic Indicators of Progression of Transitional Cell Carcinoma

Initially, transitional cell carcinomas arise in the urothelium (stages Tis/Ta) and in these early stages of progression, loss of heterozygosity (LOH) and mutations at loci on chromosome 9p and 9q are frequently observed. As the cancers become more advanced, a variety of other chromosomal regions are involved. Chromosomal regions 17p and 13q, containing tumour suppressor genes p53 and Rb respectively, are generally involved in latter stages of progression as LOH and mutations in these genes become more common.

Figure adapted from Dalbagni *et al*. (1993).


LOH / Mutations
Normal Urothelium
9p, 9q
5q, 3p, 17p
11p, 6q, 13q, 18q
Ta
Tis
T1
T2
T3/T4
N+/M+
Direction of Invasion

Orlow *et al*., 1994). Tumour suppressor genes *p16* and *p15* have been localised to chromosome band 9p21 (Section 1.4.3.2). Chromosomal region 17p13, containing tumour suppressor gene *p53*, has also been implicated in early stage TCCs (Section 1.4.3.1). This indicates that *p53*, *p16* and *p15* may all be involved in the early stages of bladder carcinoma progression.

UPKIB has been localised to human chromosomal region 3q13.3-q21 (Finch *et al*., 1997). This region has not been implicated in bladder cancer by either loss of heterozygosity (LOH) or deletion, indicating that 3q anomalies and, in particular, deletions in *UPKIB*, may not be significant in the progression of TCCs.

1.4.3.1 The tumour suppressor genes p53 and Rb

p53, located at chromosome band 17p13, and *Retinoblastoma* (*Rb*), located on chromosome band 13q14, are tumour suppressor genes and cell cycle regulators which, when deregulated, are frequently associated with cancer development. Alterations in *p53* and *Rb* occur in approximately 50% and 30% of bladder cancers respectively and may be used as indicators of disease progression and poor prognosis (Esrig *et al*., 1994; Ishikawa *et al*., 1991; Reznikoff *et al*., 1996; Spruck *et al*., 1994; Review by Brandau & Bohle, 2001). One theory suggesting how p53 functions involves its potential ability to downregulate expression of the metastasis suppressor genes such as *KAI-1* (Reviewed by Jackson & Puisieux, 2000) and metastasis-associated factors including the laminin receptor (Modugno *et al*., 2002). However, the specific function of p53 in the regulation of KAI-1 expression has been controversial and is discussed further in Section 1.5.1.2. A number of different *p53* alterations including loss of heterozygosity (LOH), mutations and polymorphisms have been linked to TCC development and spread, and these alterations vary between studies, with differences in patient epidemiology, risk factors (Section 1.4.1) and

cytogenetic variations within samples (Fig. 1.8). A difference in involvement between 17p (*p53*) and 13q (*Rb*) regions in the invasion of bladder carcinomas has been illustrated by Dalbagni *et al*. (1993a) (Fig. 1.8).

Several studies have used single strand conformation polymorphism (SSCP) techniques to analyse p53 gene mutations in bladder cancer. One region of *p53* frequently (30-40%) associated with point mutations in TCCs involves exons 4 through 9 (Miyamoto *et al*., 1993; Fujimoto *et al*., 1992; Oyasu *et al*., 1995). This was also demonstrated by Harano *et al*. (1999) who detected mutations in exons 4 to 9 in 36% of TCCs (38/105) and in these cases, the mutations were most frequently detected in Grade III tumours, indicating an association with tumour progression.

Spruck *et al*. (1993) examined *p53* exons 5 through 8 in a range of TCCs and did not demonstrate any significant differences between smokers and non-smokers. Inactivating mutations were detected in 40% (16/40) of tumours from smokers and 33% (13/40) of tumours from non-smokers. Berggren *et al*. (2001) also examined 189 TCC samples for mutations in exons 5 through 8 in *p53*. The mutations in this study were located at hotspots in codons 280 and 285 (exon 8), and were detected in 14% (26/189) of cases, which is lower than the 30-40% reported by Miyamoto *et al*. (1993), Fujimoto *et al*. (1992), Oyasu *et al*. (1995), Harano *et al*. (1999) and Spruck *et al*. (1993). The differences may be attributable to the different exons examined (exons 4-9 or exons 5-8) and to patient epidemiology, such as ethnicity or to differences in sample size. Bernardini *et al*. (2001) used PCR-SSCP and DNA sequencing to investigate *p53* mutations in exons 5-9 in 98 bladder carcinomas in current cigarette smokers, non-smokers or ex-smokers. Tumours were classified as Ta (40 cases), T1 (20 cases) and T2 to T4 (38 cases) and 28 mutations were detected in 24 of the 98 tumours (24.5%). Similarly to the study by Spruck *et al*. (1993), the

incidence of *p53* mutations was not significantly affected by smoking. However, this study showed that the nature of the mutations were different. Transversions G:C to T:A (1 case) and G:C to C:G (2 cases) were identified in tumours from current smokers. Multiple *p53* alterations, including point mutations and base pair deletions, were detected in the tumours from 4 current smokers, which were absent in ex-smokers or non-smokers. Bernardini *et al*. (2001) suggested that the G:C to A:T transversions in smokers were characteristic of carcinogen exposure, and in non-smokers the majority of mutations involving G:C base pairs at CpG sites, were likely representative of endogenous mutations.

Tsutsumi *et al*. (1997) and Dalbagni *et al*. (1993b) both reported LOH of the *p53* locus. Tsutsumi *et al*. (1997) detected LOH in approximately 30% (13/40) of samples, with higher incidence of LOH in Grade III tumours (57%) than Grade I tumours (10%). Dalbagni *et al*. (1993b) analysed 60 TCCs and detected LOH in 55% of T1-T4 tumours (21/38) and no LOH in Tis (0/1) or Ta (0/8) tumours.

Using analysis with allele-specific polymerase chain reactions (PCR), Soulitzis and colleagues (2002) recently studied an arg72pro (exon 4) polymorphism in *p53* from peripheral blood samples from 50 bladder cancer patients and 99 normal individuals. They linked an arg/arg genotype to an increased risk of bladder cancer development. This conflicts with previous findings by Toruner and colleagues (2001), who showed that any difference in frequency of the arg/arg genotype were not statistically significant between normal and tumour groups. This latter study was based on PCR and restriction digestion analysis on peripheral blood genomic DNA samples obtained from 121 bladder cancer patients and 114 normal individuals. The differences between the results in the two studies may be related to variations in methodology or to variations in patient epidemiology.

In addition to studies involving p53 in the bladder, the Retinoblastoma tumour suppressor gene (*Rb*) has also been analysed in transitional cell carcinomas. Wada *et al*. (2000) investigated loss of heterozygosity (LOH) at seven sites in the region 13q11 to 13q32 in 236 TCC cases by comparing DNA from tumour and matched peripheral blood samples. LOH was detected at 13q14.3, a region including *Rb*, in 32% of cases of TCC. This finding conflicts with earlier findings by Ishikawa *et al*. (1991), who did not demonstrate any correlation between LOH at the *Rb* locus and bladder cancer. Ishikawa *et al*. (1991) suggested that TCCs progressed either by *Rb* inactivation through independent mutations at each allele, or from the inactivation of a second tumour suppressor gene on chromosome 13 in a region not including *Rb*.

Miyamoto *et al*. (1996) examined LOH for both *p53* and *Rb* in 45 TCCs, using PCR and restriction fragment length polymorphism analysis. LOH was observed in 38% of cases at the *p53* locus and in 22% of cases at the *Rb* locus. Overall, these data suggest that LOH of *p53* and *Rb* occurs in TCCs, but the overall incidence of the LOH remains uncertain. The discrepancies between studies may be attributable to factors including varying techniques, variance of sample size and epidemiology of sampling. Both *p53*, and to a lesser extent *Rb*, appear to play important roles in the initiation and progression of TCCs, with mutations, LOH and polymorphisms of both of these loci being present in bladder tumours.

1.4.3.2 The tumour suppressor genes p15 and p16

Deletions of regions of chromosome 9p and 9q frequently occur in bladder cancers, with candidate genes identified by loss of heterozygosity (LOH) including *p16* (*CDKN2A/MTS1/TP16*) and *p15* (*CDKN2B/MTS2/TP15*) localised to 9p21; *DBCCR1* (*deleted in bladder cancer chromosome region 1*) at 9q32-33; and *PTCH* (*patched*) at 9q22.3. *p16* has been studied intensively in bladder carcinomas with

homozygous deletions and LOH detected at the *p16* locus, and LOH in the region being the region a strong indicator of poor prognosis in bladder cancer (Baud *et al*., 1999; Friedrich *et al*., 2001).

In quiescent cells, p16 is inactive. In mitogen-stimulated cells, cyclin D is activated and associates with cyclin-dependent kinases (CDKs) 4 and 6 to inactivate Rb protein, releasing E2F transcription factors (Review by Kolch *et al*., 2002). p16 can inhibit G1-phase cyclin-D/CDK complexes by binding to CDKs 4 and 6. Loss of *p16* may lead to uncontrolled cyclin/CDK activation and subsequent loss of G1/S checkpoint control and cancer proliferation.

Kamb *et al*. (1994) detected homozygous deletions of *p16* in 266 of 580 cell lines and primary tumours, derived from a range of tissues including lung, breast, brain, kidney, ovary and the bladder. This data suggests that *p16* homozygous deletions have widespread tissue distribution and may contribute to the development of many forms of malignancy. In addition to this finding, Cairns *et al*. (1995) and Williamson *et al*. (1995) also detected 9p21 homozygous deletions in bladder tumours. Cairns *et al*. (1995) used microsatellite analysis with markers surrounding the *p16* locus to show that approximately 60% of 545 primary tumours, including those of the bladder (177 of 285 cases), breast (13 of 20 cases) and colon (3 of 9 cases), had any 9p loss. With subsequent analysis of the 9p21 region, it was found that a region was deleted containing *p16* and excluding *p15*. Similarly, Williamson *et al*. (1995) detected *p16* homozygous deletions in 8/16 (50%) of bladder tumour cell lines and analysed 140 primary transitional cell carcinomas of the bladder. Among the TCCs, 13 tumours had small 9p21 LOH, 31 tumours had monosomy 9 (LOH at all informative loci), 5 tumours had LOH of 9q only, and 91 tumours had no LOH at all informative loci. Analysis revealed that all 13 tumours (100%) with

small defined 9p21 deletions had homozygous deletions of *p16*, 18 of 31 (58%) tumours with monosomy 9 had homozygous deletions of *p16*, 2 of 5 (40%) tumours with 9q LOH only had homozygous deletions of *p16*, and 9 of 91 (10%) tumours with no chromosome 9 LOH had homozygous deletions of *p16*.

To define further the involvement of 9p21 genes in bladder cancer, Stadler and Olopade (1996) examined the relevant chromosomal region in 16 bladder cancer cell lines. Deletions of 9p21 were identified in nine cell lines (56%) and, in these cell lines, the deletions involved one exon of *p16* and also involved *p15* and the closely linked gene encoding the enzyme methylthioadenoside phosphorylase (*MTAP*) in over 65% of cases. This showed the importance of *p16* deletions in the initiation or continuance of at least some bladder cancer cell lines. Studies have also shown LOH of *p16* to be a common occurrence in TCCs, with Baud *et al*. (1999) demonstrating LOH in 48% of 44 samples and Friedrich *et al*. (2001) in 35% of 37 cases.

Point mutations in *p16* are rare in clinical samples of bladder cancer and do not appear to be involved in the development of superficial disease or tumour progression. *p16* mutations have either not been detected, or detected in less than 15% of cases (Miyamoto *et al*., 1995; Orlow *et al*., 1995; Okajima *et al*., 1996; Sorlie *et al*., 1998). *p15* mutations are even less common than *p16* mutations, with no *p15* mutations detected by Miyamoto *et al*. (1995) in 50 bladder tumours.

In bladder cancer, increased levels of p16 protein can reduce the levels of Rb protein and conversely, decreased levels of p16 protein increase levels of Rb protein. Benedict *et al*. (1999) used immunohistochemical techniques to demonstrate that p16 protein was reduced as a result of alterations in the 9p21 chromosomal region or in the p16 gene. In four cases with absent p16 staining and strong homogeneous Rb staining, LOH at 9p21 at one allele was observed: three of the cases also had a

mutation in the second *p16* allele and one case showed a homozygous deletion of the second *p16* allele. Presumably as a consequence of absent p16, strong nuclear staining of Rb protein was observed by immunohistochemistry and correlated with poor prognosis in the patient. Conversely, high levels of p16 have also been detected in bladder tumours with more advanced stage and grade, including lymph node metastases, and these elevated levels of p16 are regarded as denoting poor prognosis (Wu *et al*., 2000). The reasons for these discrepancies are not clear and these studies have shown the difficulties inherent in predicting disease stages, with both high and low levels of p16 expression leading to poor prognosis.

An interesting alternative mechanism of *p16* inactivation in bladder carcinomas may involve methylation (Chapter 5), an epigenetic mechanism regulating gene expression. A study by Akao *et al*. (1997) has demonstrated 50% (19/38) of primary urothelial tumours contained *p16* mutations, homozygous deletions or hypermethylation of the 5' CpG island. This suggests that inactivation of a second *p16* allele through LOH or epigenetic alterations including methylation may render *p16* completely inactive. Many tetraspanins (Section 1.5) contain a large number of CpG residues in their promoter regions, and the tetraspanin family member UPKIB may therefore be inactivated through methylation. This will be discussed in more detail below.

1.4.4 Uroplakin IB and other uroplakins in bladder cancer

Patterns of expression of uroplakins have been studied extensively in both transitional cell carcinoma-derived cell lines and in TCC samples obtained from patients. In TCC cell lines, uroplakins are not always expressed, as reported by Finch *et al*. (1999) in which 5637, T24, J82 and TCC-Sup cell lines showed no

detectable UPKIB mRNA transcripts. Similarly, Lobban *et al*. (1998) demonstrated that TCC cell lines RT112 and HT1376 expressed UPKIB mRNA in high abundance but they could not detect mRNA from any of the other three uroplakins in these cell lines. In the same study, it was also reported that the RT4 cell line expressed mRNA for all four uroplakins, VM-Cub3 cell lines expressed UPKIA, UPKIB and UPKII but not UPKIII transcripts and COLO232, KK47 and EJ cell lines all had no detectable uroplakin mRNA expression.

Moll *et al*. (1995) used immunohistochemical techniques to detect levels of UPKIII proteins in TCC patient samples. The study showed that 14/16 (88%) of papillary non-invasive TCCs, 29/55 (53%) of invasive (T1-T4) TCCs and 23/35 (66%) of metastases were positive for UPKIII. Lobban *et al*. (1998) used *in situ* hybridisation to detect expression of UPKIB mRNA in 9 non-invasive TCCs and in 8 primary invasive TCCs with matched lymph node metastases in 5 individuals. Among the non-invasive tumours, 7/9 expressed UPKIB mRNA (78%), and among the invasive samples, 50% of lymph node metastases retained expression of UPKIB mRNA. These results differs from Finch *et al*. (1999), who showed reduced or absent expression of UPKIB mRNA in approximately 70% of TCCs, a larger proportion of loss of expression than shown in studies by Yuasa *et al*. (1998) and Lobban *et al*. (1998). The reductions in expression described by Finch and colleagues (1999) were observed over a range of grade and stage, with more advanced tumours having more frequent loss of UPKIB mRNA expression. However, these latter findings were obtained using Northern hybridisation analysis, which is less sensitive than reverse transcription-polymerase chain reaction (RT-PCR) for analysis of mRNA expression. In addition to this, patient ethnicity may

have varied between populations in the three studies (Finch *et al*., 1999; Yuasa *et al*., 1998; Lobban *et al*., 1998).

Yuasa *et al*. (1998) analysed UPKIB and UPKIII mRNA by *in situ* hybridisation and detected expression of these mRNA in all samples of both normal urothelium (3/3) and bladder carcinoma (12/12). Peripheral blood lymphocytes (PBLs) from patients with metastatic disease and normal individuals were also analysed for UPKIB and UPKIII expression using nested RT-PCR to detect circulating tumour cells arising from the TCCs. This technique assumes that uroplakin expression is urothelium-specific in normal individuals and is maintained in TCCs. None of the PBL samples (3/3) from normal individuals expressed either uroplakin. However, 3/12 (25%) of PBL samples from patients with Grade II metastatic TCCs expressed these markers. This assay appears to be a useful technique in the detection of more advanced metastatic bladder cancers, by monitoring circulating tumour cells.

These studies were subsequently extended to analyse patterns of expression of the other uroplakins in bladder cancer. Using RT-PCR, Yuasa *et al*. (1999) detected UPKIA and UPKII mRNA in 12 TCC samples and in 3 samples of macroscopically normal urothelium. Similarly to the analysis of UPKIB and UPKIII expression (Yuasa *et al*., 1998), this new study used nested RT-PCR to show that UPKII transcripts were detected in PBL samples from 3 individuals with bladder metastases but not in PBL samples from healthy individuals nor from individuals with non-metastatic disease. In support of this finding, Li *et al*. (1999) and Lu *et al*. (2000) have also detected UPKII mRNA in the PBL of patients with metastatic bladder cancer. These studies suggest that the detection of uroplakin expression in the peripheral blood may be employed as a diagnostic marker for the presence of

circulating metastatic TCC cells. However in their study, Lu *et al*. (2000) also detected UPKII expression in the PBL from a patient with a papillary non-invasive (Ta) TCC, which is usually a non-metastatic carcinoma. This finding suggests that the strategies employed for detecting circulating metastatic tumour cells may not be as specific for metastatic cells as initially presented in the other reports, or else that the initial Ta diagnosis may have been inaccurate.

Olsburgh *et al*. (2003) recently analysed 10 advanced stage (T3-4) and grade (GIII) TCCs for the presence of uroplakin transcripts using *in situ* hybridisation. Primary tumours varied in their expression of the uroplakins, with a minimum of 20% of tumours demonstrating expression of UPKIII, 70% expression of UPKII, and 40% expression of both UPKIA and UPKIB. In lymph node metastases, expression of UPKIA, IB, II and III mRNA were detected in 50% of cases for UPKIA, 50% for UPKIB, 60% for UPKII and in 50% of cases for UPKIII. This finding is slightly different to findings by Seraj *et al*. (2001), who reported that all three samples of lymph node metastases expressed UPKIII. However, the latter sample size was extremely small. Olsburgh *et al*. (2003) proposed that the variations between the studies were due to sensitivities in technique. However, the differing sample sizes may also have affected any conclusions drawn of patterns of UPKII expression in the metastatic TCCs within the lymph nodes.

Deregulation of expression of the uroplakins has also been implicated in TCCs in cattle (Ambrosio *et al*., 2001). In this study, 20 bladder tumours from cattle that had been suffering from chronic enzootic hematuria, were stained with antibodies against all of the UPKs. Umbrella cells in urothelium from normal cattle showed strong immunoreactivity to the UPK antibodies on the cell surface, but not in the cytoplasm. In Grade I TCCs, UPK immunoreactivity was detected discontinuously

in umbrella cells and intracytoplasmic staining was also observed in intermediate cells. In both Grade II and III TCCs, UPK staining was distributed irregularly in clusters on the cell surface and diffuse intracytoplasmic staining was also detected in tumours. This data rather nicely indicates increasingly disordered distribution of UPKs within the urothelium, correlating with more advanced tumour grading.

The distribution patterns of UPK proteins in TCCs in cattle were similar to human studies performed by Kaufmann *et al*. (2000) and Moll *et al*. (1995). Kaufmann *et al*. (2000) and Moll *et al*. (1995) focused on human UPKIII, and using immunohistochemical staining, detected UPKIII positivity in between 53% and 88% of TCC cases, with detection of UPKIII protein reducing with increasing degree of invasiveness. Kageyama *et al*. (2002) recently investigated UPKIA protein in human TCCs using immunochemistry and showed positive staining in 96.8% of primary cancers from cystectomy patients, high expression in 94.4% of moderate to well-differentiated TCCs and in 80.0% of poorly differentiated TCC. These studies suggest a progressive reduction in UPK proteins with more advanced tumours, but these studies did not report on patterns of UPK expression within specific cells.

The importance of the uroplakins as urothelial cell markers in TCCs seems to be established; their normal distribution being within the umbrella cells of the urothelium. It is unclear how reduced levels of UPK mRNA seen in tumours (Yuasa *et al*., 1998, 1999; Finch *et al*., 1999) correlate with the progressive loss and disordered cellular distribution of the UPK proteins observed by Ambrosio *et al*. (2001), Kaufmann *et al*. (2000), Moll *et al*. (1995) and Kageyama *et al*. (2002) and the presence of UPK mRNA observed by Olsburgh *et al*. (2003) in up to 70% of advanced primary TCCs and in their paired lymph node metastases. Uroplakin proteins could be used as prognostic indicators, with lower uroplakin protein levels

indicative of advanced disease based on immunohistochemical analyses. Current mRNA expression data for the uroplakins appears to vary markedly between studies, making the prognostic significance of the mRNA data uncertain.

1.5 The Tetraspanins

Uroplakins IA and IB are tetraspanin family members. The tetraspanin family (transmembrane 4 superfamily (TM4SF)/tetraspans/4TM) was independently identified by two groups in 1990 with the recognition that amino acids from CO-029, Sm23 and CD81 (Section 1.5.1.1) were similar in nature to tumour antigens CD63 (Section 1.5.1.4) and CD37 (Szala *et al*., 1990; Oren *et al*., 1990). Members of the tetraspanin family are cell surface proteins, which may mediate cell development, activation, and motility through cellular signalling (Reviews by Berditchevski, 2001; Yanez-Mo *et al*., 2001). UPKIB is a tetraspanin as it conforms to criteria required for tetraspanins as described below.

Hemler *et al*. (1996) and Maecker *et al*. (1997) have both described the conserved protein structure of the tetraspanin family members (Fig. 1.9). As their name suggests, the tetraspanins span cell membranes four times, in addition they have highly conserved hydrophobic transmembrane domains, but have divergent extracellular and cytoplasmic domains (Wright & Tomlinson, 1994). A number of amino acid residues are highly conserved within the tetraspanins, which may influence how these membrane-associated proteins function. These include lysine (K) in the first cytoplasmic domain; asparagine (N) in the first transmembrane domain; glycine (G) in the second transmembrane domain; glutamic acid (E) in the second cytoplasmic domain; glutamic acid or glutamine (E/Q) in the third

Figure 1.9

Conserved Structure of the Tetraspanin Family Members

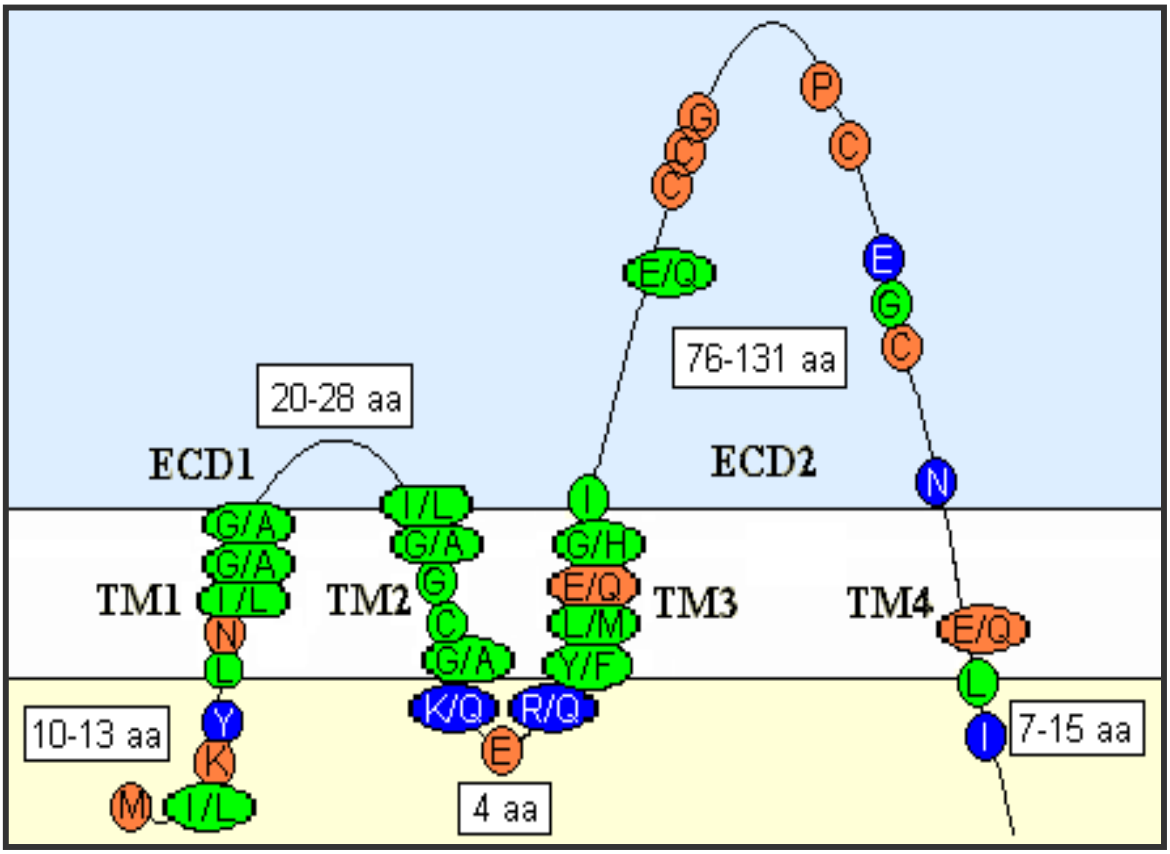
Tetraspanins contain one small and one large extracellular domain (ECD; light blue sector), four highly conserved transmembrane domains (TM, uncoloured sector) and three short cytoplasmic domains (yellow sector).

Approximate sizes of the various extracellular and cytoplasmic domains are shown. Colours at amino acid residues denote differences between studies by Hemler *et al*. (1996) and Maecker *et al*. (1997): orange circles denote conserved amino acid (aa) residues in both studies; conserved amino acid residues proposed by Hemler *et al*. (1996) alone are shaded in dark blue; and conserved amino acid residues proposed by Maecker *et al*. (1997) alone are shaded in green.

Amino Acid Residues

A	Alanine	K	Lysine
C	Cysteine	L	Leucine
E	Glutamic Acid	M	Methionine
F	Phenylalanine	N	Asparagine
G	Glycine	P	Proline
H	Histidine	R	Arginine
I	Isoleucine	Y	Tyrosine

Figure adapted from Hemler *et al*. (1996) and Maecker *et al*. (1997).


ECD1
ECD2
TM1
TM2
TM3
TM4
20-28 aa
76-131 aa
10-13 aa
4 aa
7-15 aa
G/A
G/A
I/L
N
L
Y
K
M
I/L
I/L
G/A
G
C
G/A
K/Q
R/Q
E
I
G/H
E/Q
L/M
Y/F
G
C
C
P
C
E/Q
E
G
C
N
E/Q
L
I

transmembrane domain; and the highly conserved series cysteine-cysteine-glycine (CCG) in the second extracellular domain.

There is some disagreement as to the nature of the conserved residues. A likely explanation may be the analysis of different tetraspanins in studies by Hemler *et al*. (1996) and Maecker *et al*. (1997). In their respective analyses, Hemler *et al*. (1996) based the conserved residues on 7 tetraspanin family members and Maecker *et al*. (1997) compared 18 tetraspanin members. In Figure 1.9, orange depicts residues in agreement with both Hemler *et al*. (1996) and Maecker *et al*. (1997), blue depicts conserved residues proposed by Hemler *et al*. (1996) and green depicts conserved residues presented by Maecker *et al*. (1997). The conservation of these amino acid residues and transmembrane domains identifies these proteins as tetraspanins and the considerable variation within the large second extracellular domains give the tetraspanins their individual characteristics.

Wright *et al*. (1993) suggested that the TM4SF members arose by gene duplication, given that *CD53* is very similar in gene structure to members *CD63* and *CD81* (*TAPA-1*) and hypothesised that the tetraspanins may be derived from an ancestral chromosome segment as a consequence of gene duplication. Although there is variation in exon lengths between family members, the positions of six CD53 introns are conserved in both CD63 and CD81 genes, indicating divergence from a primordial gene.

Tetraspanins are known to interact both with each other and with a variety of other proteins, including integrins. Integrins in turn can bind to extracellular matrix molecules such as collagen and are essential for processes including signal transduction, cell attachment and cell cycle progression (Review by Brakebusch *et al*., 2002). Berditchevski and colleagues (1996) suggest that integrin/tetraspanin

complexes are important in both cellular maintenance and structure, as these proteins appear co-localised in human primary cells including smooth muscle cells and foreskin fibroblasts and in cell lines including breast cancer-derived MDA-MB-231 and sarcoma-derived HT1080-C9 cells.

1.5.1 Members of the tetraspanin family

There are currently over 100 members of the tetraspanin superfamily identified, with 105 members listed in Appendix A. It can be seen that the members cover a broad spectrum of proteins expressed in a diverse range of organisms, including primitive schistosomes and nematodes. This evolutionary conservation suggests that these proteins play important biological roles. As discussed above, tetraspanin members are involved in cell motility and metastasis, as well as cell activation, signal transduction, and adhesion through interaction with integrins. Some key tetraspanin members involved in these processes include CD81 (TAPA-1), KAI-1 (CD82), CD9 (MRP-1), and CD63 (ME491). These tetraspanins have been studied in a range of carcinomas and have been implicated in suppression of metastasis. This property of tetraspanins will be discussed in detail below, as it is particularly relevant to studies described in this thesis.

1.5.1.1 Tetraspanin CD81

CD81 (TAPA-1/Target of Antiproliferative Antibody-1/M38/S5.7) is a 26kD cell-surface protein with broad tissue distribution including human thymocytes (Todd *et al*., 1996), human keratinocytes (Okochi *et al*., 1997) and glial cells in the developing rat brain (Sullivan & Geisert, 1998). CD81 associates with T-lymphocyte surface molecules CD4 and CD8 and with B-lymphocyte surface molecules CD19 and Leu-13 (Boismenu *et al*., 1996; Tseng *et al*., 2001; Horvath *et*

al., 1998; Takahashi *et al*., 1990). *CD81* has been localised to human chromosomal region 11p15.5 (Virtaneva *et al*., 1994) and loss of heterozygosity occurs in this region in Wilms tumours, lung, ovarian, and breast cancers, rhabdomyosarcomas and adrenocorticoid carcinomas (Review by Hu *et al*., 1997).

Pileri *et al*. (1998) demonstrated that the hepatitis C virus (HCV) envelope protein E2 could bind to CD81 on its second extracellular domain, and others have shown that this binding inhibits activation of natural killer (NK) cells in the immune response against HCV (Crotta *et al*., 2002; Tseng & Klimpel, 2002). CD81 also associates with a range of integrins including $\alpha_4\beta_1$ (CD49d/CD29) and $\alpha_{6A}\beta_{1,}$ and with factors HLA-DR and VLA (very late antigen) (Mannion *et al*., 1996; Domanico *et al*., 1997; Rubinstein *et al*., 1996). Taken together, these molecular interactions implicate CD81 in cell adhesion, integrin-mediated cell migration and the immune response.

1.5.1.2 Tetraspanin KAI-1/CD82

Metastasis suppression is associated with the KAI-1 (KANGAI-1/CD82/IA4 antigen/C33/R2 leucocyte antigen/4F9/ST6) gene. Initially, studies by Ichikawa *et al*. (1992) created rat/human hybrid prostatic cells by introducing human chromosomal region 11p13-p11.2 into metastatic rat prostatic cells, a region containing genes including *KAI-1*. Introduction of this region in the metastatic cells prevented metastasis but not local tumour growth in an animal model. This study implicated several genes in the 11p region, as being potentially important metastasis suppressors.

The KAI-1 gene was subsequently localised to chromosome 11, at band p11.2, by Virtaneva *et al*. (1993) and Dong *et al*. (1995). KAI-1 mRNA was shown to be

expressed in a range of cell types, including prostatic epithelial cells, early haematopoietic progenitor cells, normal granulocytes and monocytes (Gil *et al*., 1992; Burchert *et al*., 1999). The KAI-1 protein was also shown to be a co-stimulatory protein along with the accessory molecule CD3, activating T cells and leading to IL-2 production (Lebel-Binay *et al*., 1995).

Dong *et al*. (1995) cloned the KAI-1 cDNA and demonstrated metastasis suppression upon the restoration of expression of KAI-1 mRNA to metastatic prostate cancer cells. In these important experiments, the KAI-1 cDNA was subcloned into the constitutive vector pCMVneo and transfected into parental metastatic cells from rat prostate carcinoma cell line AT6.1. The resulting high levels of KAI-1 mRNA were shown to suppress the invasive ability of the parental AT6.1 cells through Boyden chamber invasion assays using matrigel. To further investigate the metastasis suppressive ability of KAI-1, the AT6.1-KAI-1 hybrid clones were inoculated into severe combined immunodeficiency (SCID) mice and inhibition of lung metastasis was observed when compared with the parental cells.

White *et al*. (1998) demonstrated reduced levels of KAI-1 mRNA in 31 of 42 cancer cell lines including those derived from prostate, ovary, bladder and lung carcinomas. White *et al*. also proposed that KAI-1 mRNA downregulation is an indicator of metastatic potential in cancers of urogenital, gynaecological, and pulmonary origin and in melanomas. KAI-1 mRNA was expressed in early stage colorectal carcinomas and lost in advanced tumours (Lombardi *et al*., 1999). Muneyuki *et al*. (2001) investigated 70 patients with advanced colorectal cancer and identified a statistically significant correlation between reduced KAI-1 mRNA expression and poor disease outcome. KAI-1 mRNA expression was also reduced in metastatic hepatocellular carcinomas (Guo *et al*., 1998) and there was an inverse

correlation between levels of KAI-1 mRNA and invasive behaviour in pancreatic cancers (Guo *et al*., 2000; Sho *et al*., 1998). KAI-1 cell surface expression levels in various human leukaemias, including chronic myeloid leukaemia (CML), acute myeloid leukaemia (AML) and chronic lymphocytic leukaemia (CLL) were analysed using two-colour flow cytometry by Burchert *et al*. (1999). It was found that KAI-1 was overexpressed compared with normal white blood cells, in which only one third of lymphocytes were KAI-1+ (CD82+). Analysis of KAI-1 transcription in normal blood and leukaemia by semi-quantitative PCR also revealed increased expression of KAI-1 in leukaemic cells. Burchert and colleagues (1999) also reported abundant expression of KAI-1 on CD34+ progenitor cells. These findings led this group to propose that overexpression of KAI-1 in leukaemia may be associated with the increased immature progenitor cell types associated with leukaemia. This finding showing upregulation of KAI-1 mRNA and expressing protein is in contrast to previous findings showing downregulation of KAI-1 expression in solid tumours.

The use of KAI-1 as an indicator for prognosis has also been investigated in lung and breast cancer. Using immunohistochemistry, Higashiyama *et al*. (1998) analysed 200 samples of non-small cell lung carcinoma: 104 samples (52%) were negative for KAI-1 protein and 31 (15%) showed reduced levels. The study did not correlate KAI-1 protein levels with evidence of cancer invasion, but did suggest the presence of KAI-1 protein in the tumour favoured a good prognosis and overall survival. In breast cancer, Huang *et al*. (1998) analysed the expression of KAI-1, CD63 and CD9 in 109 breast cancers using both quantitative RT-PCR and immunohistochemical assays. In this study, CD63 was expressed in all 109 samples and was CD63 positive in all tumours. In 73 samples (67%), CD9 positive staining was detected and 36 samples (33%) were CD9 negative, and 44 tumours (40.4%)

were KAI-1 positive and 65 samples (59.6%) were KAI-1 negative. Huang *et al*. (1998) also noted that disease-free survival rates were higher in patients expressing both CD9 and KAI-1 mRNA, indicating that the expression of both tetraspanin members CD9 and KAI-1 is a good prognostic indicator of favourable disease outcome in breast cancer patients. Similarly, studies by White *et al*. (1998) and Higashiyama *et al*. (1998) described above indicated that lower KAI-1 expression, in both mRNA and protein respectively, favoured poor disease prognosis. Analogous to CD81, KAI-1 interacts with integrin $\alpha_4\beta_1$ and with factors HLA-DR and VLA (Mannion *et al*., 1996; Rubinstein *et al*., 1996), indicating a similar role to CD81 in adhesion and integrin-associated cell migration in bladder cancers (Section 1.5.1.1).

The molecular mechanisms regulating *KAI-1* remain uncertain, with several studies analysing potential regulatory mechanisms including methylation (Section 1.5.2) and the interactions of upstream transcription factors with the KAI-1 promoter. A direct interaction of p53 protein with KAI-1 upstream enhancer elements in the activation of *KAI-1* has been proposed by Mashimo *et al*. (1998, 2000). In their studies, both *in vitro* and *in vivo* approaches have been undertaken. In these studies, gel mobility shift assays demonstrated that p53 bound to a DNA segment 860bp upstream from the *KAI-1* transcription initiation site. The transfection of a p53 expression plasmid LNp53B into a p53-negative prostate cancer cell line also induced a 3.8-fold elevation of KAI-1 mRNA levels. Using immunohistochemistry, 177 prostate carcinomas were analysed for the expression of p53 and KAI-1. p53-positive carcinomas were generally KAI-1-positive (93/110) and p53-negative tumours were generally KAI-1-negative (41/67). Taken together, the evidence suggests that p53 is a major component of the nuclear transcription factors which

bind to and activate the KAI-1 promoter, leading to increases in KAI-1 mRNA and protein levels.

Studies by Lombardi *et al*. (1999) and Jackson *et al*. (2002) appear to indicate that p53 and other p53-activated proteins are not solely responsible for interacting with and activating the KAI-1 promoter. Lombardi *et al*. (1999) analysed 20 colon carcinoma cell lines using immunohistochemical techniques and detected no correlation between p53 and KAI-1 protein levels. Jackson *et al*. (2002) analysed p53 and KAI-1 levels in 22 bladder and prostate cell lines and showed that exposure of cells to UV and other DNA damaging agents activated expression of *p53* and of the p53-activated gene *p21* (*WAF1*), but did not alter levels of KAI-1 mRNA.

Marreiros *et al*. (2003) recently analysed a region within the KAI-1 promoter important in the activity of *KAI-1*, including the p53 site initially analysed by Mashimo *et al*. (1998). The region was analysed using transfection experiments in 2 bladder cell lines, using CAT reporter vectors containing varying sequences from the KAI-1 promoter. Binding motifs for p53, AP-1 and AP-2 were all required for activity of the reporter system. Mutational analysis of the individual p53, AP-1 or AP-2 sites within the reporter constructs reduced expression of KAI-1 mRNA in the transfected cell lines. The only variation to this pattern of expression was obtained with a construct containing a *p53* mutation, which was transfected into the bladder carcinoma cell line HT1376 and showed similar reporter activity to the construct without a mutation. The transcriptional activity in this region of the KAI-1 promoter suggests the possible combined role of p53 and other factors such as c-jun and AP-2α in the regulation of *KAI-1*.

1.5.1.3 Tetraspanin CD9

Originally detected by the murine monoclonal antibody 602-29 (Andrews *et al*., 1981), and shown to be expressed by most human cells, CD9 (MRP-1/Motility-related protein-1/MIC3) has been shown to inhibit cell growth in mouse melanoma cells (Ikeyama *et al*., 1993) and be an important member of the tetraspanin family. CD9 also has the ability to suppress cell motility and metastasis in lung adenocarcinoma and breast cancer and demonstrates an inverse correlation with the presence of metastases in breast cancer (Ikeyama *et al*., 1993; Miyake *et al*., 1995). CD9 is expressed in activated T-cells, platelets, and neural cells, also suggesting a possible role for CD9 in intracellular signalling in the nervous system (Tai *et al*., 1996; Peng *et al*., 1997; Kaprielian *et al*., 1995).

CD9 has been localised to human chromosome 12, band p13, a region containing several genes whose losses of expression can lead to cancer. These include *p27*, involved in G1/S transition in the cell cycle, and *TEL*, a member of the ETS gene family of transcription factors (Van Cong *et al*., 1989; Benoit *et al*., 1991; Dahia *et al*., 1998; Stegmaier *et al*., 1995). *TEL* often shows loss of heterozygosity (LOH) and is also involved in translocations, deletions and complex rearrangements with *AML1* (*acute myeloid leukaemia 1)*, which can lead to leukaemias including acute lymphoblastic leukaemia (ALL) (Bernardin *et al*., 2002; Mitani, 2002; Kanerva *et al*., 2001). However, no direct association exists between CD9 and AML1 genes in the literature.

CD9 protein and mRNA levels have been studied in a number of tumours and a similar prognostic significance has been shown as has been identified for KAI-1. Reduced levels of CD9 protein have been correlated with poor prognosis in breast cancer (Miyake *et al*., 1996) and reduced expression of CD9 mRNA have correlated

with a worse prognosis, in terms of both overall survival and disease-free survival among 132 patients with lung adenocarcinomas (Higashiyama *et al*., 1997).

In common with many of the other tetraspanins, CD9 associates with a number of integrins, including $\alpha_3\beta_1$ and $\alpha_6\beta_1$. CD9 is also an accessory subunit of VLA-integrin complexes (Okochi *et al*., 1997; Schmidt *et al*., 1996; Rubinstein *et al*., 1994). The physical interaction of integrin complexes with the tetraspanins has been demonstrated in studies by Tachibana & Hemler (1999), which showed that expression of both CD9 and CD81 mRNA can increase fusion between muscle cell membranes through their integrin-mediated binding and can increase structural stability. CD9 is also involved in sperm-egg fusion, a process involving the integrin $\alpha_6\beta_1$ and is expressed on the mouse egg membrane (Kaji *et al*., 2000; Le Naour *et al*., 2000). The above investigations all mark CD9 as an important developmental-regulatory gene, a prognostic marker for disease progression, and an important factor in integrin-mediated cell adhesion and signalling.

1.5.1.4 Tetraspanin CD63

The gene encoding CD63 (ME491/MLA-1/granulophysin/murine monoclonal antibody 710F) was localised to human chromosome 12, at bands q12-14 (Hotta *et al*., 1988), and the expressed protein product induces cellular adhesion and spreading of monocytic cells on tissue culture dishes (Koyama *et al*., 1990; Koyama *et al*., 1998). Sho *et al*. (1998) investigated the potential of CD63 as a prognostic indicator in pancreatic cancers by assessing the levels of mRNA using RT-PCR. However, in this study, levels of expression of CD63 mRNA did not appear to correlate with tumour stage.

In common with other tetraspanins, the CD63 protein is associated with integrins including $\alpha_4\beta_1$ and $\alpha_5\beta_1$, with factors HLA-DR, VLA-3 and VLA-6 and also co-localises with tetraspanins CD151 (Peta-3) and CD9, showing the complexity of tetraspanin-integrin interactions (Radford *et al*. 1995; Mannion *et al*., 1996; Sincock *et al*., 1997; Rubinstein *et al*., 1996; Berditchevski, 2001; Berditchevski *et al*., 1995).

In studies by Hotta *et al*. (1991) and Radford *et al*. (1995), CD63 has been shown to regulate motility, adhesion and suppression of metastasis of melanoma cells. Hotta *et al*. (1991) transfected CD63 into H-*ras*–transformed mouse NIH-3T3 cells and observed reduced growth of the cells after injection into athymic nude mice. Similarly, Radford *et al*. (1995) transfected CD63 into human KM3 melanoma cells not normally expressing CD63, which were subsequently injected subdermally into nude mice. Reduced numbers of tumours were detected in mice injected with melanoma cells containing vector and *CD63* insert (KM3/26) compared with cells containing pREP9 vector alone (KM3/T), suggesting that CD63 expression suppressed tumour growth. However, the *in vitro* growth properties of KM3 cells were also tested and the relative growth rates between KM3/26 and KM3/T cells were not significantly different.

The studies described above for the tetraspanins CD81, KAI-1, CD9 and CD63 all suggest roles for tetraspanins in regulating cell development, proliferation, activation, motility and metastasis and may provide some insight into functions for uroplakins IA and IB in the bladder. KAI-1, CD9, CD63 and CD81 all appear to be important metastasis suppressor proteins that interact with integrins and are important structural constituents of plasma membranes. The loss of expression of tetraspanins, as discussed in the previous sections, can indicate the degree of cancer

progression. It is hypothesised that deregulation of tetraspanins *UPKIA* and *UPKIB* would similarly lead to a poor prognosis in TCCs.

1.5.2 Tetraspanins in bladder cancer

Tetraspanins have diverse roles in a range of tissues and interact with factors including other tetraspanins and integrins, suggesting that they may be important determinants of cancer progression. Levels of tetraspanin mRNA have been studied in a variety of cancers (Reviews by Yanez-Mo *et al*., 2001; Berditchevski, 2001; Boucheix & Rubinstein, 2001; Hemler, 2001). Of the tetraspanins, only KAI-1 and the uroplakin family members UPKIA and UPKIB have been studied in transitional cell carcinomas.

KAI-1 mRNA is highly expressed in normal bladder, inflammatory bladder and non-invasive papillary Grade I/II TCCs and its expression has been shown to be lost in high grade and invasive TCCs, suggesting a potential role in suppression of metastasis in TCCs (Yu *et al*., 1997). KAI-1 mRNA and protein levels were investigated by Ow *et al*. (2000) in 135 TCCs, by *in situ* hybridisation and immunohistochemistry respectively. A reduction of both KAI-1 mRNA and protein levels was observed in TCC tissues when compared to normal samples, and a further reduction in KAI-1 levels was also observed in invasive TCCs when compared to non-invasive TCCs. Recent studies by Jackson *et al*. (2000a, 2000b) have also supported these previous studies. In 18 bladder cancer cell lines, Jackson and colleagues (2000a) have shown that low levels of KAI-1 mRNA were associated with reduced adhesion to fibronectin *in vitro*, reduced cell-cell adhesion and increased invasiveness through matrigel. These findings support the participation of KAI-1 in the suppression of bladder tumour invasion.

Bladder cancers exhibit increased *de novo* methylation of CpG islands in the promoters of a range of genes including *p16* (Salem *et al*., 2000). Jackson *et al*. (2000b) extracted DNA from invasive TCCs and bladder cancer cell lines to determine the methylation status (Section 5.1) of the CpG island in the KAI-1 promoter. Evidence for hypermethylation was not detected, suggesting that mechanisms other than methylation were presumably responsible for reduced KAI-1 mRNA expression in these cancers. Sekita *et al*. (2001) examined methylation of *KAI-1* in 4 prostate cancer cell lines in a 331bp CpG island, which contained the KAI-1 transcription initiation site. Direct sequencing of DNA cloned from cell lines revealed that, of the 33 CpG dinucleotides in the region, cell lines LNCaP and PC-3 had 3 methylated CpG sites, DU-145 had 2 methylated sites, and TSU-Pr1 had no methylated CpG sites. With the use of RT-PCR, KAI-1 expression was reactivated with the demethylating agent 5-Aza-2' deoxycytidine (5-Aza-CdR) in 2 of the 4 prostate-carcinoma-derived cell lines, suggesting that demethylation of the entire genome could reactivate the KAI-1 gene. In contrast to these findings, Northern blot analysis and immunohistochemistry did not reveal activation of KAI-1 mRNA or protein in any cell line after treatment with 5-Aza-CdR, presumably due to the reduced sensitivity between these techniques for the detection of KAI-1 mRNA and in the detection of KAI-1 protein. This study suggested that methylation may not be the main mechanism of inactivation of *KAI-1* in the prostate cancer cell lines studied. Recently, Uzawa *et al*. (2002) also showed methylation of the KAI-1 CpG island to be infrequent in oral tumours, primary oral squamous cell carcinomas (OSCC), pre-cancerous lesions, metastatic OSCCs and within OSCC-derived cell lines.

Hence, from the studies described above in a range of cancers, it can be proposed that the metastasis suppressor protein KAI-1 may play a role in bladder

cancer metastasis. However, the nature of the molecular mechanisms regulating KAI-1 mRNA expression are largely unknown, with a combination of factors including methylation and the presence of enhancer binding motifs including AP-1, AP-2 and p53 (Section 1.4.3.1) important in the regulation of *KAI-1*.

1.6 Aims of This Study

The uroplakins are expressed in mammals, with UPKIB detected in the bladders of humans, monkeys, cattle, sheep, pigs, dogs, rabbits, mice and rats and in the mink lung. This data largely points to common patterns of expression of UPKIB mRNA in the bladders of higher vertebrates. No data has yet been reported on UPKIB mRNA expression in other vertebrates and invertebrates, and the tissue-specific expression of UPKIB in humans has not yet been fully elucidated.

A primary aim of this study was therefore to investigate Internet-based databases for UPKIB mRNA expression in human organs other than the bladder and eye, to divulge the full extent of expression of UPKIB transcripts in human tissues. A second aim was to detect other UPKIB homologues to characterise further patterns of UPKIB mRNA expression in a range of species other than mammals. Sequences of these UPKIB homologues can then be compared for similarity and divergence, using a range of evolutionary techniques including hydrophilicity plots and parsimony analysis. Human UPK tetraspanin members UPKIA and IB could also be compared against other human tetraspanins and tetraspanins from other organisms to verify the evolutionary relatedness of these tetraspanins to other members of the tetraspanin family, including those published in insects.

A third aim was to study how expression of UPKIB is regulated in bladder carcinoma cells. The relationship of UPKIB to metastasis suppressor genes KAI-1

and CD9 make it a candidate for metastasis suppression in TCC. Characterisation of the genomic sequence of *UPKIB* was initially required to define the gene structure, establish intron/exon boundaries, identify the promoter sequence and the transcription start site and to determine factors and sequences involved in promoter regulation. This could be performed with techniques including Internet-based databases and 5'rapid amplification of cDNA ends (5'RACE). Sequence information of the UPKIB promoter may reveal AT- or GC- rich areas, which would aid in the discovery of common DNA binding sequences such as Sp-1, AP-1 and the TATA initiation sequences. Sequence data would also reveal factors that may control UPKIB mRNA expression, such as TGFβ1 or p53.

Once transcription start site and promoter sequences have been determined, the UPKIB promoter can be analysed for transcription factor binding motifs and CpG islands, as reported in promoters of other tetraspanin family members. It is important to determine how *UPKIB* is regulated in normal tissues, how it is deregulated in TCCs, and how its regulatory mechanisms can be compared to other tetraspanins to determine transcription factors involved in regulation of both UPKIB and other tetraspanin family members.

An investigation of UPKIB promoter regulatory mechanisms, such as methylation, may give insight into the deregulation of this tetraspanin in TCCs. Methylation is a regulatory mechanism appearing to increase in cancer cells with advancing TCC stage and patient age and is one area of interest to investigate, as the promoters of all tetraspanins studied so far are GC rich and may therefore have CpG islands available for methylation. The methylation status of *UPKIB* can then be investigated in a number of TCC cell lines and tissues including normal urothelium,

normal colonic epithelium and bladder tumours to gauge the impact of methylation on UPKIB mRNA expression in cancer development.

Chapter 2

Materials and General Methods

2.1 Materials

Reagents for Tissue Culture	List of Suppliers
5-Aza-2'- deoxycytidine ($C_8H_{12}N_4O_4$)	Sigma, St. Louis, MO, USA
Dulbecco's Modified Eagle's Medium (DMEM)	Sigma, St. Louis, MO, USA
Foetal Calf Serum (FCS)	CSL, Melbourne, VIC, Australia
L-Glutamine (L-Glu)	Gibco BRL Life Technologies (Invitrogen), Gaithersburg, MD, USA
Trypsin-EDTA (0.5% Trypsin, 5.3mM EDTA.4Na) (10x)	Gibco BRL Life Technologies (Invitrogen), Gaithersburg, MD, USA

Molecular Biology Components Used	List of Suppliers
5'RACE Kit: 10x PCR Buffer [200mM Tris-HCl (pH 8.4), 500mM KCl] 25mM MgCl2 10mM dNTP mix 0.1M DTT Superscript II Reverse Transcriptase (200U/µl) RNase Mix 5x Tailing Buffer [50mM Tris-HCl (pH 8.4), 125mM KCl, 7.5mM $MgCl_2$] 2mM dCTP Terminal Deoxynucleotidyl Transferase 5'RACE Abridged Anchor Primer (AAP, 10µM) Abridged Universal Anchor Primer (AUAP, 10µM)	Gibco BRL Life Technologies (Invitrogen), Gaithersburg, MD, USA
Agarose; DNA grade; Molecular Biology grade	Progen Industries Ltd., Darra, QLD, Australia
Bresapure Midi Plasmid Isolation Kit	GeneWorks, Adelaide, SA, Australia
dNTP solutions (dATP, dCTP, dGTP, dTTP), 100mM each	Promega Corporation, Madison, WI, USA

Gigaprime Labelling Kit**:** Nucleotide / Buffer Cocktail C Enzyme Mix (Klenow) Decanucleotide Solution	GeneWorks, Adelaide, SA, Australia
Mussel Glycogen (5µg/µl)	Ambion, USA
Hybond N+ Nitrocellulose Membrane	Amersham International plc, Buckinghamshire, England
Molecular Size Markers: pUC19 (*Hpa*II) SPP1 (*Eco*RI) 1kb+	 GeneWorks, Adelaide, SA, Australia GeneWorks, Adelaide, SA, Australia Gibco BRL Life Technologies (Invitrogen), Gaithersburg, MD, USA
Oligo-(dT) (15-mer) 500µg/ml	Promega Corporation, Madison, WI, USA
Oligonucleotides (see sections 2.3.3.3 and Table 2.1)	GeneWorks, Adelaide, SA, Australia Gibco BRL Life Technologies (Invitrogen) Pty. Ltd., NSW, Australia Operon Pty. Ltd., Seattle, USA
Polymerases: Expand Long Template PCR System Expand Polymerase 10x PCR Buffer 1: 17.5mM $MgCl_2$ HotStar Taq DNA Polymerase 10x PCR Buffer $MgCl_2$ Taq Polymerase 10x Buffer, $MgCl_2$-free $MgCl_2$	 Roche Diagnostics Australia Pty. Ltd., Castle Hill, NSW, Australia Qiagen, Clifton Hill, VIC, Australia Promega Corporation, Madison, WI, USA
Low Melt Agarose; Molecular Biology grade	Progen Industries Ltd., Darra, QLD, Australia
Maxi Plasmid Isolation Kit	Qiagen, Clifton Hill, VIC, Australia
MDE™ (Mutation Detection Enhancement) Gel Solution	Cambrex / BioWhittaker Molecular Applications (BMA), Rockland, ME, USA
Proteinase K (PK)	Merck Pty. Limited, Darmstadt, Germany.
Restriction Enzymes (with 10x Buffers): *Hind*III (A/AGCTT) *Pst*I (CTGCA/G) *Pvu*II (CAG/CTG) *Xba*I (T/CTAGA)	 New England Biolabs Inc., Beverly, MA, USA

RNase A	Sigma, St. Louis, MO, USA
Salmon Sperm DNA	Sigma, St. Louis, MO, USA
Superscript II Reverse Transcriptase and 10x Buffer	Gibco BRL Life Technologies (Invitrogen), Gaithersburg, MD, USA
SYBR-Gold Nucleic Acid Gel Stain 10 000x Concentrate in DMSO	Molecular Probes, Eugene, OR, USA
T4 DNA ligase (3U/µl)	Promega Corporation, Madison, WI, USA
T4 Polynucleotide Kinase	Promega Corporation, Madison, WI, USA
Tissue-Tek^{R} OCT Compound	Ted Pella Inc., Redding, CA, USA
TRI-Zol Reagent	Gibco BRL Life Technologies (Invitrogen), Gaithersburg, MD, USA
UltraClean^{TM} Standard Mini Plasmid Prep Kit	Mo Bio laboratories, Solana Beach CA, USA
WIZARD^{TM} DNA Gel Purification Kit	Promega Corporation, Madison, WI, USA

Radiochemicals	**List of Suppliers**
Alpha-^{32}P-labelled dATP	GeneWorks, Adelaide, SA, Australia
Gamma-^{32}P-labelled dATP	GeneWorks, Adelaide, SA, Australia

Antibiotics	**List of Suppliers**
Ampicillin (Amp)	Gibco BRL Life Technologies (Invitrogen), Gaithersburg, MD, USA
Kanamycin (Kan)	Gibco BRL Life Technologies (Invitrogen), Gaithersburg, MD, USA

Film	**List of Suppliers**
Fujichrome DX400 Colour-Positive Slide Film	Fuji Photo Film Co., Ltd., Tokyo, Japan
Hyperfilm MP	Amersham International plc, Buckinghamshire, England

General Reagents	List of Suppliers
Ammonium Persulphate (APS)	Sigma, St. Louis, MO, USA
Bacto™-Agar	Difco Laboratories, Detroit, MI, USA
Bacto™-Tryptone	Difco Laboratories, Detroit, MI, USA
Bacto™-Yeast Extract	Difco Laboratories, Detroit, MI, USA
β-Mercaptoethanol	Sigma, St. Louis, MO, USA
Boric Acid	Sigma, St. Louis, MO, USA
Bovine Serum Albumin (BSA)	New England Biolabs, Beverley, MA, USA
Bromophenol Blue	Bio-Rad Laboratories, Hercules, CA, USA
Chloroform ($CHCl_3$)	Sigma, St. Louis, MO, USA
Diethyl pyrocarbonate (DEPC)	Sigma, St. Louis, MO, USA
Dimethyl Sulphoxide (DMSO)	Sigma, St. Louis, MO, USA
Eosin	Sigma, St. Louis, MO, USA
Ethylene Diamine Tetra Acetic Acid (EDTA)	Sigma, St. Louis, MO, USA
Ethanol	Merck Pty. Limited, Darmstadt, Germany.
Ethidium Bromide	Sigma, St. Louis, MO, USA
Ficoll-Paque	Pharmacia LKB, Uppsala, Sweden
Formamide	Merck Pty. Limited, Darmstadt, Germany
Glacial Acetic Acid	Merck Pty. Limited, Darmstadt, Germany
Glycerol / Glycerin	F.H. Faulding & Co. Ltd., Adelaide, Australia
Harris' Haematoxylin Solution	Australian Biostain Pty. Ltd., Traralgon, VIC, Australia
Hydrochloric Acid (HCl)	Ajax Chemicals Pty. Ltd., Auburn, NSW, Australia
Hydroquinone	Sigma, St. Louis, MO, USA
Isopropyl-β-D-Thiogalactopyranoside (IPTG)	Promega Corporation, Madison, WI, USA
Isoamyl Alcohol	Sigma, St. Louis, MO, USA
Isopropanol	Sigma, St. Louis, MO, USA
Methanol	Merck Pty. Limited, Darmstadt, Germany
Phenol	Sigma, St. Louis, MO, USA

Polyvinylpyrrolidone (PVP)	Sigma, St. Louis, MO, USA
Potassium Chloride (KCl)	Merck Pty. Limited, Darmstadt, Germany.
Potassium Di-hydrogen Orthophosphate (KH_2PO_4)	Ajax Chemicals Pty. Ltd., Auburn, NSW, Australia.
Sodium Acetate (NaAc)	Ajax Chemicals Pty. Ltd., Auburn, NSW, Australia.
Sodium Azide	Sigma, St. Louis, MO, USA
Sodium Chloride (NaCl)	Merck Pty. Limited, Darmstadt, Germany.
Tri-sodium Citrate	Merck Pty. Limited, Darmstadt, Germany.
Sodium Bisulfite ($Na_2S_2O_5$)	Sigma, St. Louis, MO, USA
Sodium Di-hydrogen Orthophosphate (NaH_2PO_4 (anhydrous))	Merck Pty. Limited, Darmstadt, Germany.
Di-sodium Hydrogen Orthophosphate (Na_2HPO_4 (anhydrous))	Merck Pty. Limited, Darmstadt, Germany.
Sodium Dodecyl Sulphate / Sodium Lauryl Sulphate (SDS)	Sigma, St. Louis, MO, USA
Sodium Hydroxide (NaOH) Pellets	Merck Pty. Limited, Darmstadt, Germany.
Sucrose	Merck Pty. Limited, Darmstadt, Germany.
Tris Base	Sigma, St. Louis, MO, USA
Urea	Sigma, St. Louis, MO, USA

Note:

All reagents are AR grade unless stated otherwise.

2.2 Solutions and Buffers

Agar plates	1L

Bacto™-agar (15g) was dissolved in 1L of LB (Luria Bertani) broth. The solution was autoclaved and cooled to below 55˚C before the addition of kanamycin (final concentration 75μg/ml) or ampicillin (final concentration 25μg/ml) for agar plates. Surplus antibiotic-free agar was stored at room temperature.

10% Ammonium Persulfate (APS)	1ml

Ammonium persulfate (0.1g) was dissolved in 1ml sterile MilliQ-H_2O and the solution was immediately used in the polymerisation of polyacrylamide gels.

Ampicillin (Amp) (stock 50mg/ml)	2ml

Amplicillin (100mg) was dissolved in 1.5ml sterile MilliQ-H_2O and the volume was adjusted to 2ml with MilliQ-H_2O. This mixture was sterilised by filtration through a 0.22μm filter and stored in 400μl aliquots at -20˚C.

Chloroform/Isoamyl Alcohol (24:1 v/v)	50ml

Chloroform (48ml) and isoamyl alcohol (2ml) were mixed and stored at 4˚C.

100x Denhardt's Reagent	100ml

Ficoll (2g), polyvinylpyrrolidone (2g) and BSA (2g) were made up to a final volume of 100ml with MilliQ-H_2O and 10ml aliquots were stored at -20˚C.

EDTA (0.5M; pH 8.0)	1L

Di-sodium ethylene diamine tetra-acetate.2H_2O (186.1g) was added to 800ml MilliQ-H_2O and the pH adjusted to 8.0 with NaOH (10M). The volume was adjusted to 1L with MilliQ-H_2O, autoclaved and stored at room temperature.

Eosin (1% in 25% ethanol)	100ml

Eosin (1g) was dissolved in 100ml 25% ethanol and the solution was stored at room temperature.

L-Glutamine (L-Glu) (200mM)	10ml

L-Glutamine (292.3mg) was dissolved in 10ml sterile MilliQ-H_2O, aseptically dispensed into 1ml aliquots and stored at -20˚C.

100mM Hydroquinone	10ml

Hydroquinone (0.11g) was dissolved in 10ml sterile MilliQ-H_2O and used fresh in bisulfite nucleotide conversions.

IPTG (stock 500mM)	5ml

IPTG (0.596g) was dissolved in 5ml sterile MilliQ-H_2O, 0.22µm filter sterilised in 1ml aliquots and stored at –20˚C.

Kanamycin (Kan) (stock 10mg/ml)	10ml

Kanamycin (100mg) was dissolved in 9ml sterile MilliQ-H_2O and the volume adjusted to 10ml with MilliQ-H_2O. The solution was sterilised by filtration through a 0.22µm filter and stored in 1ml aliquots at -20˚C.

6x Loading Buffer (DNA)	100ml

Bromophenol blue (0.25g) and sucrose (40g) were dissolved in 80ml sterile MilliQ-H_2O and the pH adjusted to 10.0 with NaOH (1M). The volume was adjusted to 100ml with sterile MilliQ-H_2O and the solution stored in aliquots at -20˚C.

2 x Loading Buffer (RNA)	1ml

Deionised formamide (500µl) was mixed with 100µl 10x MOPS buffer, 167µl 37% deionised formaldehyde, 133µl DEPC-treated H_2O, 100µl glycerol and a few grains of bromophenol blue and stored at –20˚C.

Luria-Bertani (LB) Broth (pH 7.5)	1L

Bacto™-tryptone (10g), Bacto™-yeast extract (5g), and NaCl (10g) were dissolved in 900ml sterile MilliQ-H_2O and the pH adjusted to 7.5 with 10M NaOH. The solution was made up to 1L with sterile MilliQ-H_2O, autoclaved and stored at room temperature.

10x MOPS Buffer	1L

MOPS free acid (4.8g) and sodium acetate (4.1g) were dissolved in 900ml DEPC-treated H_2O (Section 2.3.2.1). The pH was adjusted to 7.0 with 6M DEPC-treated NaOH, 20ml 0.5M EDTA (pH 8.0) was added and the final volume was adjusted to 1L with DEPC-treated H_2O. The solution was stored at 4˚C in a foil-wrapped bottle away from light.

PAGE Fixative (20% methanol / 5% acetic acid)	1L

Methanol (200ml) and acetic acid (50ml) were mixed with MilliQ-H_2O to a final volume of 1L and stored at room temperature.

10x Phosphate Buffered Saline (PBS) (Ca^{2+}/Mg^{2+}-free)	1L

NaCl (8g), KCl (0.2g), Na_2HPO_4 (1.15g) (anhydrous), and KH_2PO_4 (0.2g) were dissolved in 800ml MilliQ-H_2O and the pH adjusted to 7.4 with concentrated HCl before solution was made up to 1L with MilliQ-H_2O, autoclaved and stored at 4˚C.

Potassium Chloride (0.075M) (KCl)	1ml

Potassium chloride (27.96g) was dissolved in 500ml MilliQ-H_2O and stored at 4˚C.

Proteinase K (stock 10mg/ml)	10ml

Proteinase K (100mg) was dissolved in 10ml sterile MilliQ-H_2O and stored in 1ml aliquots at -20˚C.

Salmon Sperm DNA, Sheared (10mg/ml)	25ml

Salmon sperm DNA (100mg) was resuspended into a total of 10ml sterile MilliQ-H_2O and boiled for 5 minutes at 95˚C. The DNA was sheared by passing 10x through a 23-gauge needle, divided into 400µl aliquots and stored at 4˚C.

Sephadex G-50	25ml

Sephadex G-50 medium grade (1g) was suspended in 25ml MilliQ-H_2O and allowed to swell overnight at 65˚C. Sodium azide was added to a final concentration of 0.02% and the slurry was stored at 4˚C.

Sodium Acetate (NaAc) (3M; pH 5.2)	1L

Sodium acetate (246.1g) was dissolved in 900ml MilliQ-H_2O and the pH was adjusted to 5.2 with glacial acetic acid. The volume was adjusted to 1L with MilliQ-H_2O, autoclaved and stored at room temperature.

Sodium Bisulfite ($Na_2S_2O_5$) (4.8M; pH 5.0)	20ml

Sodium bisulfite (10g) was added to 10ml sterile MilliQ-H_2O and adjusted to pH 5.0 with 500µl NaOH (10M). The final volume was adjusted to 20ml with sterile MilliQ-H_2O, and used immediately.

Sodium Bisulfite (4M) / Urea (6.24M) (pH 5.0)	20ml

Sodium bisulfite (7.6g) and urea (7.5g) were added to 10ml sterile MilliQ-H_2O and

the pH adjusted to 5.0 with 500µl NaOH (10M). The final volume was adjusted to 20ml with MilliQ-H_2O and used immediately.

10% Sodium Dodecyl Sulfate (SDS)	200ml

SDS (20g) was dissolved in 180ml sterile MilliQ-H_2O with gentle heating and stirring. The volume was adjusted to 200ml with sterile MilliQ-H_2O and the solution stored at room temperature.

Sodium Hydro-Carbonate ($NaHCO_3$) (0.5M)	10ml

$NaHCO_3$ (0.42g) was dissolved in 9ml MilliQ-H_2O, the pH was adjusted to 9.0 with NaOH (10M), the volume made up to 10ml with MilliQ-H_2O, and the solution stored at room temperature.

20x SSC (pH 7.0)	1L

NaCl (175.32g) and sodium citrate (88.2g) were dissolved in 900ml MilliQ-H_2O, the pH adjusted to 7.0 with HCl (10M) and the volume was adjusted to 1L with MilliQ-H_2O. The solution was subsequently autoclaved and stored at room temperature.

50 x TAE	1L

Tris base (242g), EDTA (18.6g; pH 8.0) and 57.1ml glacial acetic acid were dissolved in 900ml MilliQ-H_2O and the volume was adjusted to 1L with MilliQ-H_2O. The solution was stored at room temperature and diluted 1 in 50 with MilliQ-H_2O for a 1x working solution.

5 x TBE	1L

Tris base (54g), boric acid (27.5g) and 20ml 0.5M EDTA (pH 8.0) were dissolved in

900ml MilliQ-H_2O. The volume was adjusted to 1L with MilliQ-H_2O and the solution stored at room temperature. The solution was diluted 1 in 5 with MilliQ-H_2O for a 1x working solution.

TE (pH 8.0)	50ml

A 0.25ml aliquot of 2M Tris-Cl (pH 8.0) was combined with 0.1ml 0.5M EDTA (pH 8.0) and MilliQ-H_2O to a volume of 40ml. The final volume was adjusted to 50ml with MilliQ-H_2O, the solution was autoclaved and stored at room temperature.

TES (STE) Buffer	100ml

A 0.5ml aliquot of 2M Tris-Cl (pH 8.0) was combined with 0.2ml 0.5M EDTA (pH 8.0), 0.58g NaCl and sterile MilliQ-H_2O to a final volume of 100ml. The solution was stored at room temperature.

2M Tris-Chloride (pH 8.0)	1L

Tris base (242.2g) was dissolved in 800ml MilliQ-H_2O. The pH was adjusted with concentrated HCl and the volume adjusted with MilliQ-H_2O to 1L. The solution was autoclaved and then stored at room temperature.

1x SSC; 0.1% SDS	1L

To 50ml 20x SSC were added 800ml MilliQ-H_2O and 5ml 20% SDS. The volume was adjusted to 1L with MilliQ-H_2O and stored at room temperature.

2.3 General Methods

2.3.1 Tissue culture: growth and maintenance of cell lines

The cell lines T24, J82, TCC-Sup and Sca-BER were obtained from the American Type Culture Centre (Rockville, MD). Dr David Leavesley (formerly Royal Adelaide Hospital) provided cell line 5637. Dr Paul Jackson and Professor Pamela Russell (Prince of Wales Hospital, NSW) provided cell lines VM-Cub3, HT1376 and RT112.

All cell lines except Sca-BER were derived from transitional cell carcinomas. Sca-BER was a derived from a squamous cell carcinoma. Information regarding the cell lines was obtained from the American Type Cell Culture database at http://www.atcc.org.

VM-Cub3, HT1376 and RT112 were grown in 15% FCS in DMEM containing 1% L-Glu (200mM) (P. Jackson, personal communication), while 5637, T24, J82, TCC-Sup and Sca-BER were grown in 10% FCS in DMEM containing 1% L-Glu (200mM). After reaching confluence, adherent cells were washed briefly in 1x PBS, incubated in 5ml trypsin to detach cells, centrifuged at 350 x *g* at 4˚C for 5 minutes and diluted 1/10 before resuspension in fresh media.

2.3.2 RNA methods

2.3.2.1 DEPC-treatment of MilliQ-H_2O

DEPC (diethyl pyrocarbonate) (1ml) was added to 1L MilliQ-H_2O and was left in a fumehood overnight. The solution was then autoclaved and stored at 4˚C. DEPC treatment removes contaminating RNases and autoclaving hydrolyses DEPC into CO_2 and ethanol.

2.3.2.2 Isolation of RNA from bladder carcinoma cell lines and tissue samples

Cell lines were grown to confluence in 75cm^3 tissue culture flasks (Life Technologies). RNA was extracted for both cell lines and tissue using TRI-Zol reagent (Invitrogen) according to manufacturer's instructions. RNA was resuspended in a final volume of 12µl with DEPC-treated H_2O (Section 2.3.2.1) and stored at –80˚C.

2.3.2.3 Verification of RNA integrity

Preparation of a MOPS gel for RNA analysis

DEPC-treated H_2O (36.6ml) was added to 0.5g agarose and the solution was boiled. To this solution, 5ml 10x MOPS was added, the mixture cooled to 60˚C, 8.3ml formaldehyde was added and the mixture poured into gel moulds.

Preparation of RNA samples

RNA (3µl) (Section 2.3.2.2) was mixed with 3µl 2x RNA loading buffer. The mixture was heated for 5 minutes at 65˚C, put on ice, and loaded on a 1x MOPS gel for analysis of RNA integrity.

2.3.2.4 Reverse transcription

To 1µl RNA (~50ng), 5.7µl DEPC-treated H_2O and 2µl oligo-$(dT)_{15}$ primer (Promega) were added to a final volume of 8.7µl. The reaction mixture was incubated at 75˚C for 15 minutes and cooled on ice. To this mixture was added 2µl $MgCl_2$ (25mM), 4µl 5x Superscript II First Strand Buffer, 1µl dNTPs (10mM), 2µl DTT (100mM), 1µl DEPC-treated H_2O, 0.3µl RNasin (3U/µl) and 1µl Superscript II (5U/µl) to a final volume of 20µl. The reaction mixture was incubated at 42˚C for 90 minutes. The enzyme was then inactivated at 70˚C for 15 minutes and the mixture

stored at 4˚C. For RT-PCR, PCR was performed according to conditions outlined in section 2.3.3.3.

2.3.3 DNA methods

2.3.3.1 DNA extraction: the TES method

The TES (STE) method for isolation of DNA was adopted from Miller *et al*. (1988), thereby avoiding the use of phenol and chloroform. For cell lines, flasks ($25cm^3$) containing adherent cell lines were washed in 1x PBS until residual media was removed, and trypsinised to detach adherent cells. For PBL, cells were obtained after Ficoll extraction (Section 5.2.2.1). Cells were collected in a 10ml tube and centrifuged at 350 x *g* for 5 minutes. The supernatant was then removed and the cells were resuspended in 5ml 1x PBS and centrifuged at 350 x *g* for 5 minutes. The supernatant was removed and the pellet resuspended in a solution containing 840µl STE Buffer, 100µl 10% SDS, 50µl Proteinase K (10mg/ml) and 10µl RNase A (10mg/ml), and incubated overnight at 37˚C. An equal volume (1ml) of 3M NaAc (pH 5.2) was added and the solution left on ice for 15 minutes to allow precipitation of any residual protein. The solution was centrifuged for 15 minutes at 11 000 x *g* to pellet the protein.

The supernatant was placed into 500µl aliquots in four 1.5ml tubes. Two volumes (1ml) of 100% cold ethanol were added and DNA was precipitated at -70˚C for 2 hours. Precipitates were centrifuged at 11 000 x *g* for 15 minutes and washed in 70% ethanol. Pellets were vacuum dried, resuspended in 100µl TE and left at 4˚C until completely dissolved.

2.3.3.2 Agarose gel electrophoresis

The final concentration of ethidium bromide in both gels and running buffers was 0.25μg/ml. Low melting 0.8% (w/v) agarose gels were used for cutting out PCR products, 0.8% (w/v) gels were required for Southern blotting and both were run at 55V in 1x TAE buffer.

The agarose concentration of other gels was 1.0% (w/v) or 3.0% (w/v), depending on size of DNA fragments to be analysed. Samples were loaded with 6x loading buffer (section 2.2) and electrophoresis was performed at 70-80V in 1x TAE buffer. All gels were viewed on a transilluminator under UV light and photographs were taken with a digital camera, connected to Kodak digital science ID (v2.0.2) image analysis software. Exposure was typically taken for 1 to 2 seconds on f'stop 2.5.

2.3.3.3 Polymerase chain reactions

Oligonucleotide synthesis and design

Oligonucleotide (primer) sequences (Table 2.1) were designed manually, based on GenBank accession numbers AF042331, AQ318241, AF067147, AC083800 or AB002155, for the amplification of mRNA and genomic fragments from UPKIB. *Porphobilinogen deaminase* (*PBGD*; *hydroxymethylbilane*; *HMBS*) and *GAPdH* (*glyceraldehyde-3-phosphate dehydrogenase*) are two housekeeping genes, which were used as controls in RT-PCR, because the commonly-used β-actin has been reported not to satisfy all criteria for use as an internal control (Chretien *et al.*, 1988; Yamada *et al.*, 1997; Serazin-Leroy *et al.*, 1998; Selvey *et al.*, 2001).

Table 2.1

List of Primers

	Primer	Sequence (5' → 3')	Origin
UPKIB Sequence with GenBank Accession AF067147			
χ	TM3	GAAGTGGCATCTTGTATCACAGCA	cDNA
χ	ECD	TCCATTGGTCATCATTGTTTGGAG	cDNA
Primer Extension and 5'RACE			
χ	RTM3	TGCTGTGATACAAGATGCCACTTC	cDNA
β	RTM1c	AGGCAGATGCCCACAAATATG	cDNA
β	TM1	GCAGAGTGCATCTTCTTCGTA	cDNA
β	F1	TGTTCGTTGCTTCCAGGGCCTGC	cDNA
α	5'1 UPKIB	AAGCAACGAACAGTTGAGTTGTCTTTCGCCATTGTCGGG	cDNA
α	5'2 UPKIB	CTCCGCAGTCAGGGCAATGCCGCAACAACCAATAATCACA	cDNA
α	GSP1	AGGCCCTGGAAGCAACGAACAGTTGAGT	cDNA
δ	AAP	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG	Poly-C binding
δ	AUAP	GGCCACGCGTCGACTAGTAC	AAP
Long-Template PCR			
β	TM1c	CATATTTGTGGGCATCTGCCT	cDNA
χ	RTM3	TGCTGTGATACAAGATGCCACTTC	cDNA
β	RECD	CTCCAAACAATGATGACCAATGGA	cDNA
β	ex/int1_AQ318241	TACACTTCCTACTNTACTAGTC	gDNA
RT-PCR for UPKIB, PBGD and GAPdH			
β	F1	TGTTCGTTGCTTCCAGGGCCTGC	cDNA
β	R1	AGTAGAACATGGTACCCAGGAGAACC	cDNA
α	PBGD-5'	CTTTCCAAGCGGAGCCATGTCTGG	cDNA
α	PBGD-3'	CATGAGGGTTTTCCCGCTTGCAGA	cDNA
α	GAPdH-5'	CCACCCATGGCAAATTCCATGGCA	cDNA
α	GAPdH-3'	TCTAGACGCAGGTCAGGTCCACC	cDNA
Bisulfite PCR			
α	5' Methyl UPKIB	GAAAGCGATGAGTGTGGTTGTTAAGGTGT	gDNA
α	Outer A68	CCAACCCTTAAACCCGAAAAATTCCCTAC	gDNA
α	Inner S67	TTTTTCGTTTTAGCGAGGTAGGGTAGGT	gDNA

Sources:

α Primers supplied by GeneWorks (Sth Aust.) Pty. Ltd.

β Primers supplied by Gibco BRL (Aust.) Pty. Ltd.

χ Primers supplied by Operon (Aust.) Pty. Ltd.

δ 5'RACE Kit primers supplied by Clontech Pty. Ltd.

Primers for PBGD, labelled PBGD-5' and PBGD-3', were designed to amplify a 377bp fragment (GenBank accessions NM_000190 & X04808) and primers for GAPdH, labelled GAPdH-5' and GAPdH-3', were designed to amplify a 598bp fragment (GenBank accession BT006893). Suitability of all primers for amplification was checked with the programs BLAST and Amplify. BLAST was used to check the primer sequences were unique for the genes of interest, and Amplify tested primer specificity for sequences to be amplified, ensuring the lack of formation of primer-dimers. Amplification conditions for both UPKIB and GAPdH are outlined in standard PCR reactions and cycling conditions (below). Amplification conditions for PBGD are outlined in HotStar Taq PCR reactions and cycling conditions (below).

Primer stocks (500ng/µl) were diluted to 50ng/µl working solutions, except for AAP and AUAP, which were provided by the company at 10µM. The Tm (melting temperature) for short primers (≤20-mer) was calculated: 2°C (A + T) + 4°C (G + C) (Wallace *et al.*, 1979). For longer primers (>20-mer), the Tm was calculated using the Internet site http://www.basic.nwu.edu/biotools/oligocalc.html (Oligonucleotide Properties Calculator) and the Tm values were also supplied from the suppliers with the primers.

Standard PCR reactions

Reaction mix for PCR using Taq polymerase

To 34.84µl sterile MlliQ-H_2O were added 5µl 10x Taq buffer ($MgCl_2$-free), 3µl $MgCl_2$ (25mM), 1µl dNTPs (10mM), 2µl 5' sense primer (50ng/µl), 2µl 3' anti-

sense primer (50ng/µl) and 0.16µl Taq polymerase (0.5U/µl), to which 2µl DNA or cDNA product were added to a final volume of 50µl.

Reaction mix for PCR using HotStar Taq polymerase

To 36.9µl sterile MlliQ-H_2O were added 1µl $MgCl_2$ (25mM), 5µl 10x HotStarTaq buffer, 1µl dNTPs (10mM), 2µl 5' sense primer (50ng/µl), 2µl 3' antisense primer (50ng/µl) and 0.1µl HotStarTaq polymerase (5U/µl), to which 2µl DNA or cDNA product was added for a final volume of 50µl.

Cycling conditions for PCR

An Eppendorf Mastercycler Gradient PCR cycler was used which did not require mineral oil as the lid was heated during PCR cycles.

Conditions for PCR using Taq polymerase

Cycling conditions for standard PCRs using Taq Polymerase involved an initial 8 minutes at 94˚C, 45 cycles involving 1 minute at 94˚C, 1 minute at the average Tm of the primers (Table 2.1) -5˚C, and 1 minute at 72˚C, and a final 7 minutes at 72˚C before storage at 4˚C.

Conditions for PCR using HotStar Taq polymerase

Cycling conditions for PCRs (PBGD primers) using HotStar Taq Polymerase involved an initial 15 minutes at 95˚C, 35 cycles involving 45 seconds at 95˚C, 1 minute at 68˚C, and 1 minute at 72˚C, and a final 7 minutes at 72˚C before storage at 4˚C.

2.3.3.4 Analysis of DNA products and DNA purification

DNA samples and PCR reaction mixtures (10µl for 50µl PCR reactions) were run on a 1.0% TAE gel with ethidium bromide and visualised under a UV transilluminator (Ultra-lūm model MEB-20, Paramount, CA, USA) with Kodak

image analysis software for analysis (Section 2.3.4). For extraction of DNA of particular sizes from PCR and restriction digestions, products were run on a 0.8% low-melting agarose gel and bands were excised. Products were then purified using a WIZARD manifold according to the manufacturer's instructions, to a final volume of 50µl, from which 5µl was run on a 1.0% gel and visualised as above (Section 2.3.4).

2.3.4 Computer programs

Three categories of computer software tools were used during the course of this study; (i) general computer tools; (ii) sequence analysis tools, and (iii) evolutionary analysis tools (Table 2.2).

For the general computer tools, the program Amplify was used in verifying PCR primer suitability for amplifying products of choice. Kodak image analysis software (v2.0.2) was used for gel image capture with a UV transilluminator (Ultra-lūm model MEB-20, Paramount, CA, USA), as outlined in the software instructions, at f'stop 2.5 for 1 second. Kodak image analysis software (v3.5) was used for gel image capture when used with a SYBR photographic filter (S-7569) (Section 5.2.3.7). GeneDoc, a multiple sequence alignment editor and shading utility, was used to align multiple sequences derived from *.msf format files from ClustalW, and to shade consensus regions. Programs in Microsoft Office 2000 were used during the course of these investigations, including Word, Excel and Powerpoint.

Among the sequence analysis tools, TESS (transcription element search software) was used to predict promoter transcription binding motifs and BLAST (basic local alignment search tool) was used to search the National Centre for Biotechnology Information (NCBI) databases for homology between sequences.

Table 2.2

Computer Software Tools

General computer tools	
Amplify v1.2	Engels, 1992
GeneDoc v.2.6.002	Nicholas & Nicholas, 2000
Kodak ds (digital science) 1D v2.0.2 and v3.5	Kodak Scientific Imaging Systems, New Haven, CT, USA
MS Office 2000	Microsoft Corporation, USA
Sequence analysis tools	
ABI Prism v3.4.1	Applied Biosystems, Foster City, CA, USA
BLAST	Altschul *et al.*, 1990
BLAST2 (BLASTP v2.1.2)	Altschul *et al.*, 1990
ClustalW	Thompson *et al.*, 1994; Lopez & Lloyd, 1994 http://www.ebi.ac.uk/clustalw
GenBank®	Benson *et al.*, 2002
GeneJockey Sequence Processor	Taylor, 1990
MatInspector v2.2	Quandt *et al.*, 1995
Promoter and Transcription Start Site Predictor	Reese, 1994; Reese & Eeckman, 1995; Reese *et al.*, 1996 http://www.fruitfly.org/cgi-bin/seq_tools/promoter.pl
TESS	Schug & Overton, 1997
TFSearch v1.3	Akiyama, 1998
Evolutionary analysis tools	
AddGaps	Kortschak, 1998 The University of Adelaide, Adelaide, SA, Australia
BAMBE v2.03	Simon & Larget, 2001
BioEdit v5.0.9	Hall, 2001
MacVector v.6.5.3 Goldman/Engelman/Steitz (GES) Hydrophilicity Plots	Accelrys Inc. (Oxford Molecular Group PLC), Oxford, UK
Phylip SEQBOOT DNAPARS CONSENSE	http://evolution.genetics.washington.edu/phylip.html Felsenstein, 1993c Felsenstein, 1993b Felsenstein, 1993a
TreeView (win32) v.1.5.0	Page, 1998

ClustalW is an Internet-based alignment program in which known sequences are input in FASTA format and these are then saved in *.msf format from alignment (*.aln) files. GenBank is a database presented by NCBI, containing genomic sequences from a range of organisms, in a downloadable and publicly accessible format. GenBank is part of the International Nucleotide Sequence Database Collaboration, which is comprised of the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. In summary, GenBank is the genetic sequence database for the National Institute of Health of the United States of America (NIH), an annotated collection of all publicly available DNA sequences (Benson *et al.*, 2002). The program GeneJockey was used to analyse restriction enzyme digestion sites within sequences. Promoter and Transcription Start Site Predictor is a program obtained through the *Drosophila*-based Internet site http://www.fruitfly.org, which predicts eukaryotic transcription initiation sites. ABI Prism (Applied Biosystems) is a sequencing program used by Flinders Sequencing Centre (Flinders Medical Centre, South Australia) for application to all DNA sequencing with sequencing equipment and dye chemistry.

For the programs used as evolutionary analysis tools, AddGaps was used to distinguish differences between bases from cDNA sequences, possibly identifying codon changes, and allowing more representative evolutionary tree topologies based on DNA rather than amino acid sequences. BAMBE (Bayesian analysis in molecular biology and evolution) allowed the study of evolutionary trees relating to bootstrapping for 100 sets of 100 sets, which could not be easily accomplished with the traditional maximum likelihood method. It allowed a more relative analysis of evolutionary relatedness between larger groups, as in all tetraspanins described in Chapter 3. BioEdit, a sequence alignment editor, was used in the process of

AddGaps to convert sequences to an acceptable format for alignment. MacVector was used to analyse Goldman/Engelman/Steitz hydrophilicity plots (window length 19) for all tetraspanins to verify their nature before comparison. Phylip was used for bootstrapping and Maximum Parsimony analysis; programs included SEQBOOT, used for bootstrapping and random sampling; DNAPARS, used for DNA Maximum Parsimony analysis; and CONSENSE which was used for compiling a consensus tree from Maximum Parsimony data. TreeView was obtained from the ANGIS (Australian National Genomic Information Service) to visualise trees from text files containing Newick-format obtained through BAMBE, and from Maximum Parsimony analysis.

Equipment used during the course of these studies included a Microm HM505N cryostat maintained at -25˚C and a 373 DNA Sequencer (Applied Biosystems) used by Flinders Medical Centre. Sequencing chemistry used was either Amersham Dyenamic ET Terminators by the Flinders Medical Centre Sequencing Facility, South Australia, or Big Dye v3 by the Institute of Medical and Veterinary Science (IMVS), South Australia.

Chapter 3

Bioinformatics: Comparative Studies of UPKIB and Other Tetraspanin Family Members

3.1 Introduction

Uroplakin IB (UPKIB) mRNA is expressed in urothelial tissue of humans, cattle, rabbits and mice (Lobban *et al*., 1998; Yuasa *et al*., 1998; Finch *et al*., 1999; Webb *et al*., 1999; Wu *et al*., 1990), in human ocular epithelium (Adachi *et al*., 2000) and in mink lung epithelium (Kallin *et al*., 1991). Olsburgh *et al*. (2003) recently detected UPKIB transcripts in organs other than the bladder, including the trachea and placenta (discussed in Section 1.3.4). These data support the widespread distribution of UPKIB throughout the human body and suggests that UPKIB proteins play an important role in human physiology.

Analysis of UPKIB cDNA sequences has revealed 93% sequence identity between human and cattle sequences and 94% sequence identity between human and mink cDNA sequences (Finch *et al*., 1999). Through immunohistochemistry, Wu *et al*. (1994) showed the existence of the UPKIB protein in the bladder of monkey, sheep, pig, dog and rat, indicating the functional and structural importance of UPKIB in the bladder (Section 1.3.3).

The question arises as to whether UPKIB is conserved among non-mammalian organisms, and whether UPKIB is a conserved structural component of the bladder among non-mammalian organisms. Currently, UPKIB has been identified only within class mammalia. It is therefore of interest to search sequence databases for homologues of human UPKIB within other orders of vertebrates. Identification of UPKIB in several vertebrate species may point to this protein playing a conserved role in bladder development.

UPKIB is a member of the tetraspanins, a large family of proteins whose members each contain 4 transmembrane domains with several conserved amino acid

residues (Section 1.5). Members of this family have diverse cellular roles including sperm-egg fusion in mice, muscle-cell fusion and cancer metastasis suppression in various tissues including prostate (Section 1.5). Finch (1998) hybridised a radiolabelled UPKIB cDNA probe to human metaphase chromosome spreads and detected multiple sub-peaks. This study suggested that these chromosomal regions were either pseudogenes or tetraspanin sequences with similarity to UPKIB. Tetraspanin ESTs have been reported in a wide range of animals including species of the genera *Drosophila*, *Caenorhabditis*, *Schistosoma* and *Xenopus* (Todres *et al*., 2000; Fan *et al*., 1997; Kedzierski *et al*., 1996). The evolutionary relationships between these more primitive tetraspanins and UPKIB are currently unknown.

Maecker *et al*. (1997), Hemler (2001) and Boucheix & Rubinstein (2001) have carried out phylogenetic analysis based on sequence comparisons of some members of the tetraspanin family. However, the reliabilities of the phylogenies they derived were not tested using bootstrapping. In the study by Maecker *et al*. (1997), the amino acid sequences analysed extended from the amino terminal to a conserved CCG (cys-cys-gly) residue (Section 1.5); the complete second extracellular domains of the proteins, which diverge markedly between tetraspanin members, were not included. Hemler (2001) presented a maximum parsimony tree, but he did not test branching arrangements with bootstrapping. Boucheix & Rubinstein (2001) presented a distance tree, based on complete amino acid sequences for several human and mouse tetraspanins, but did not analyse tetraspanins from more primitive organisms.

In this Chapter, an analysis is presented of the evolutionary relationships between the tetraspanin family members, including members of the UPKIB subgroup, in several species. The aim is to derive a phylogeny, which might allow

better understanding of the evolution of the tetraspanins among a diverse population of organisms, and may help identify the origins of the uroplakins.

3.1.1 Evolutionary analysis

Methods including multiple sequence alignments, hydrophilicity plots, maximum parsimony and maximum likelihood may be employed to study similarities and evolutionary relationships between genes or proteins. Most methods of evolutionary analysis rely on an initial multiple sequence alignment, often carried out using the algorithm ClustalW. ClustalW and other such sequence alignment algorithms require a subjective decision to be made about the relative "penalty" to attach to sequence mismatches and sequence gaps. This fact must always be born in mind when interpreting derived phylogenetic trees. The greater the degree of sequence dissimilarity, and the greater the degree of differences in the sequence length, the less accurate the derived multiple sequence alignment, and therefore the less accurate the derived phylogeny.

Hydrophilicity (or hydrophobicity) plots are based on amino acid sequences, and allow characterisation of membrane-spanning hydrophobic (transmembrane) domains and exposed extracellular or cytoplasmic hydrophilic domains. These protein profiles may allow detection, by eye, of subtle variations in amino acid sequences and this may be a useful technique in the analysis of tetraspanin family members, which vary markedly between their second extracellular domains (Section 1.5).

The two most commonly used methods for constructing gene and protein phylogenies are maximum parsimony and maximum likelihood. Maximum parsimony (MP) assumes that the most probable evolutionary tree is that which

minimises the total number of nucleotide substitutions. The method assumes that nucleotide sites evolve independently of one another. It also assumes that all possible substitutions are equally likely, and that substitutions are rare (Felsenstein, 1993b). MP has been a popular technique allowing rapid analysis of large data sets, and it can be used in conjunction with bootstrapping to give an indication of the robustness of the resulting phylogenetic tree.

Maximum likelihood (ML) has some advantages over MP in predicting evolutionary trees, as the underlying model can account for biological relevant features of sequence evolution, for example, different rates of transitions and transversions. Bayesian Analysis in Molecular Biology and Evolution (BAMBE) is a ML program, based on Markov Chain Monte Carlo (MCMC) methods, allowing analysis of phylogenies with Bayesian inference. An advantage of the BAMBE method is that it allows the resulting tree branching order to be assigned a direct probability, obtained from the posterior distribution (Simon & Larget, 2001). It was decided to use the BAMBE program for ML analysis in this investigation due to the high efficiency with which it handles large data sets (Larget & Simon, 1999; Mau *et al*. 1997).

MP and BAMBE evolutionary trees can be represented as rooted or unrooted. Rooted trees aim to depict the sequence of evolutionary events in a time frame. The base (root) of such a tree can be identified using an outgroup. Outgroup sequences must be known, from independent evidence, to be more distantly related to all the other sequences. Unrooted trees, on the other hand, aim to depict evolutionary relationships without placing them in a time frame. This latter method of drawing a tree becomes necessary if an outgroup is impossible to identify. Rooted trees are the

most useful representation of evolutionary events. However, basing conclusions on such trees is critically dependent on the choice of a "true" outgroup.

When constructing evolutionary trees based on DNA sequences from a variety of species and homologous genes, the selection of an appropriate outgroup can be difficult, and it is tempting to use a sequence from the most distantly diverged species as the outgroup. However, the genes that form part of a gene family may be related to one another as orthologues (ie genes that last trace their most recent common ancestor to a speciation event) or paralogues (ie genes that trace their most recent common ancestor to a gene duplication event) and this relationship may not be known. In such cases, the choice of an outgroup based on known species relationships may produce an incorrect tree. Such an outgroup will only define the base of a gene tree accurately if the sequence involved is known to be ancestral to all the ingroup sequences. If the supposed outgroup sequence is orthologous to some of the sequences but paralogous to others, its use to root the tree will produce an incorrect tree. This phenomenon is depicted in Figure 3.1. In this figure, the *Xenopus* sequence is orthologous to the marsupial sequence but paralogous to the bird and placental sequences. If a rooted tree were to be constructed based on these sequences, using the *Xenopus* sequence as an outgroup, the tree would be incorrect in regards to the speciation events that had occurred. Thus, care must be taken in the use of outgroups in the analysis of both UPKIB and of the related tetraspanin family members.

In summary, the derivation of hydrophilicity plots and rooted or unrooted MP and BAMBE trees will be used to determine the evolutionary relationship of members of the UPKIB and tetraspanin families, drawing on macromolecular sequence data from a wide range of species. It will be of interest to see if UPKIB

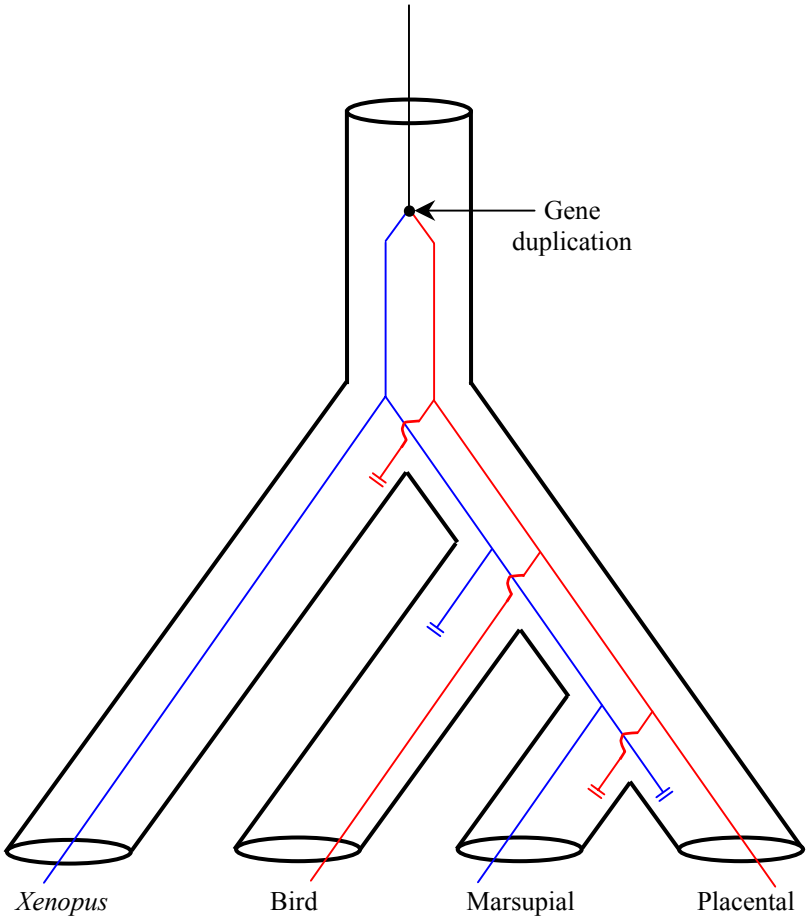


Figure 3.1

"Bronchial Tube" Diagram Showing the Evolution of a Hypothetical-gene Family

The tubes represent species, which have diverged at the joins. Duplicated gene lineages are shown as blue and red lines. *Xenopus* and marsupial genes (blue) are orthologous. Bird and placental genes (red) are also orthologous. Paralogous genes include *Xenopus* and bird sequences (red and blue) and *Xenopus* and placental genes (red and blue).

Paralogous genes originate from the gene duplication event depicted as  .

 and  depict gene loss or pseudogenes.

homologues can be identified in non-mammalian species, and if UPKIB is a conserved structural component of the bladder among non-mammalian vertebrates.

3.2 Methods

3.2.1 Data sets

DNA and protein sequences from a variety of tetraspanins, from a range of organisms, were obtained from the GenBank database and from PubMed literature searches. The tabulated data are shown in Appendix A. Maecker *et al*. (1997) initially suggested that peripherins are tetraspanin family members. However, several peripherins obtained through GenBank, including all human peripherins, did not have the characteristic 4 transmembrane domains upon study of hydrophilicity plots, so were excluded from this study.

For DNA sequences, only the open reading frames were analysed, with the stop codon triplets removed from the sequences for analysis. In multiple sequence alignments, all amino acids were included as it was deemed important to include amino acids from the second extracellular domains. The second extracellular domain has the greatest sequence variation in the tetraspanins, and may provide important phylogenetic information regarding the arrangement of terminal nodes in the resulting trees.

For BAMBE analyses it is necessary to define a notional outgroup. The outgroups were: *Xenopus laevis* 3' UPKIB sequence (GenBank accession BE507089) for UPKIB BAMBE analysis; *Homo sapien* CD9 sequence (GenBank accession L34068) for human tetraspanin BAMBE analysis, and *Caenorhabditis*

elegans B0563.2 sequence (GenBank accession U28740) for the analysis of tetraspanins from a range of organisms.

3.2.2 Sequence alignments

DNA and protein sequences were aligned using ClustalW and AddGaps. The latter is a computer program used to align multiple DNA sequences, allowing detection of nucleotide deletions and variance between DNA sequences (Kortschak, 1998, Section 2.3.4).

3.2.2.1 Pairwise sequence alignments

BLAST2 (v BLASTP 2.1.2, NCBI) was used for pairwise sequence comparisons of UPKIA and UPKIB sequences (Section 2.3.4, Section 3.3.3.4).

3.2.2.2 ClustalW multiple sequence alignments

BLAST and GenBank were employed to identify tetraspanins sequences through sequence similarity and key-word searches (Section 2.3.4). Groups of amino acid sequences were copied into http://www.ebi.ac.uk/clustalw/ in FASTA format and the gcg_msf output preference was selected. The alignment (*.aln) file was opened on the Internet from the output and saved as *.msf on the local hard drive with text file (*.txt) format selected on the drop-down bar. This *.msf file was then opened in GeneDoc and analysed for amino acid sequence comparisons.

3.2.2.3 AddGaps

To align cDNA sequences such that codons were not broken up in translation, AddGaps (Kortschak, 1998, Table 2.2) was used. cDNA sequences were placed into WordPad in FASTA format, saved as *.txt files, and placed into the AddGaps directory as infile.nuc. The file transdna.bat was opened, checkme was opened and

the output file was saved as pepin.pep, which was then opened in BioEdit. ClustalW alignment was chosen, saved as pepin.pep and these were replaced as the original output in AddGaps. The file was opened in WordPad, the amino acid numbers were deleted and the file was saved.

In AddGaps, addgaps.bat was opened. Checkme2 was opened in WordPad and then treeme.man was opened in WordPad. The file was opened in GeneDoc and saved in *.msf format for cDNA alignment. AddGaps cDNA alignments were subsequently used in MP and BAMBE analyses (Sections 3.2.4 & 3.2.5).

3.2.3 Hydrophilicity plots

Amino acid sequences were inserted into MacVector, saved in staden (*.SDN) format and analysed for hydrophilicity profiles with Goldman/Engelman/Steitz variables, with window length 19. Dr Hugh Campbell, Research School of Biological Sciences, Australian National University, assisted with these analyses. Plots were obtained, saved as *.pict files, and imported into PowerPoint and manually aligned for protein profile comparisons.

3.2.4 MP trees of cDNA sequences

The PHYLIP (v3.6 alpha2) computer package, containing programs DNAPARS, SEQBOOT and CONSENSE, was used for MP analysis using cDNA data from AddGaps (Section 3.2.2.3), and according to the manual accompanying the program (Felsenstein, 1993a, 1993b, 1993c). The reliability of branch order was assessed with 1 000 pseudo-replicates of the data set with bootstrapping program SEQBOOT and the input order of the sequences was randomised at the beginning of each search. A consensus tree was derived using CONSENSE, viewed in TreeView

(v 1.5.0). Internal edge labels were chosen and trees were rooted with outgroups where appropriate (Section 3.2.1).

3.2.5 BAMBE trees of cDNA sequences

BAMBE trees were generated from cDNA sequences (Section 3.2.2.3) by random sampling, as described in the manual accompanying the program and in the BAMBE publication (Simon & Larget, 2001). Mr Scott Spargo, Department of Molecular Biosciences, The University of Adelaide, assisted with these analyses. BAMBE (v 2.03) uses the TN93 model of sequence evolution with Markov Chain Monte Carlo (MCMC) algorithms to estimate model parameters, branch lengths and tree topologies. Seven runs were carried out each starting from a random tree, with a GLOBAL burn of 10 000 cycles followed by a main run of 2 000 000 cycles with the LOCAL algorithm, sub-sampling every tenth tree. Output data was obtained and Newick Tree format data and probabilities data were copied and pasted into Word for BAMBE tree analysis (Appendices B, C & D). The resulting trees were viewed with TreeView (v1.5.0), internal edge labels were chosen and trees were rooted with outgroups where appropriate (Section 3.2.1). Probability scores at the tree branches were assigned 100%, 80-<100%, 60-<80%, and 0-<60% for ease of visualisation.

3.3 Results

3.3.1 Human UPKIB mRNA expression and UPKIB family homologues

Analysis of Uroplakin IB homologues was undertaken to examine the divergence of UPKIB family members within both mammalian and non-mammalian

species. UPKIB homologues were identified using BLAST (National Centre for Biotechnology Information).

3.3.1.1 BLAST searches for sequences that show homology to the human UPKIB cDNA sequence

BLAST searches of the human EST database (NCBI) and the Unigene (NCBI) database, using the human cDNA open reading frame of UPKIB (GenBank accession AF042331), as the query sequence, revealed possible UPKIB expression in human tissues other than the adult eye and bladder. These tissues included the pancreas (GenBank accessions AL549174 and BG435965); ovary (GenBank accessions BE747079, BF033367, BF034348, BE019779, BE019924 and BE208200); genitourinary tract (GenBank accession AW630265); skeletal muscle (GenBank accessions AL709738, AL709739 and AL709753); kidney (GenBank accession BG401684, Chapter 4); foetal eyes (GenBank accessions BM711066 and BM666154); brain (GenBank accessions AI982899 and AW264021); uterus (GenBank accession AI811548); neuroblastoma-derived cell line SK-N-MC (GenBank accession AU100584, Chapter 4); and stomach (GenBank accession BM846565). This indicates that, far from being organ-specific in humans, UPKIB mRNA expression probably occurs in a wide variety of different tissues.

3.3.1.2 Determination of UPKIB homologues for analysis

An initial BLAST search for human UPKIB homologous using human UPKIB open reading frame cDNA (GenBank accession AF042331) as the query sequence revealed 5 additional organisms containing UPKIB-like sequences: *Xenopus,* mouse, cattle, rabbit and mink. The rabbit, mouse, and *Xenopus* sequences were shorter than the complete human UPKIB open reading frame sequence.

The *Xenopus laevis* UPKIB homologue

Nine *Xenopus* ESTs were identified, of varying length, but none covered the entire cDNA open reading frame. A multiple sequence alignment was performed on these sequences to determine which 5' and 3' sequences overlapped with each other and would be useful for evolutionary analyses (Fig. 3.2). Sequences with GenBank accessions BF024799 (assigned *Xenopus* UPKIB 5') and BE507089 (assigned *Xenopus* UPKIB 3') were chosen, as no sequences showed exact sequence identity at aligned amino acid sequences; BF024799 and BE507089 sequences were the longest ESTs with their complete 5' or 3' ends respectively.

These are the first reported ESTs with partial UPKIB open reading frame sequences from class amphibia. Information in GenBank states that these sequences were obtained from *Xenopus laevis* whole embryos and not from any specific *Xenopus* tissue.

3.3.1.3 Hydrophilicity plots of UPKIB homologues

Using MacVector Goldman/Engelman/Steitz hydrophilicity plots, the sequences of six UPKIB homologues, in addition to the human UPKIB sequence, were analysed to compare similarities between second extracellular domains (Fig. 3.3). Within the tetraspanin family, the second extracellular domain sequences vary between members, providing their individual protein-binding properties. It was of interest to view the *Xenopus* ESTs in relation to other UPKIB homologues. Most hydrophilic profiles of the UPKIB homologues (Fig. 3.3) indicate strong hydrophilic and amino acid conservation among the UPKIB family, with only slight variance observed in the second extracellular domains (ECD2) of both *Xenopus* and mouse. *Xenopus* 5' and 3' sequences have 85 amino acids overlap, with differences in 12

Figure 3.2

Multiple Sequence Alignment of *Xenopus laevis* UPKIB Sequences

Nine homologous *Xenopus* cDNA sequences were obtained from a BLAST search of a non-human database using human UPKIB cDNA (GenBank accession AF042331) as the query sequence. The *Xenopus* cDNA sequences were translated into the coding amino acids and subsequently aligned using ClustalW. The alignment is shown here using GeneDoc.

Sequences with GenBank accessions BG578239, BF615202, BE669299, AW200344, BF615799 and BF024799 were the most 5' sequences detected aligning with the human UPKIB sequence, with sequence BF024799 the longest transcript. Sequences with GenBank accessions BG555361, AW199299 and BE507089 aligned with the 3' terminus of the human UPKIB cDNA sequence, with sequence BE507089 as the longest transcript.

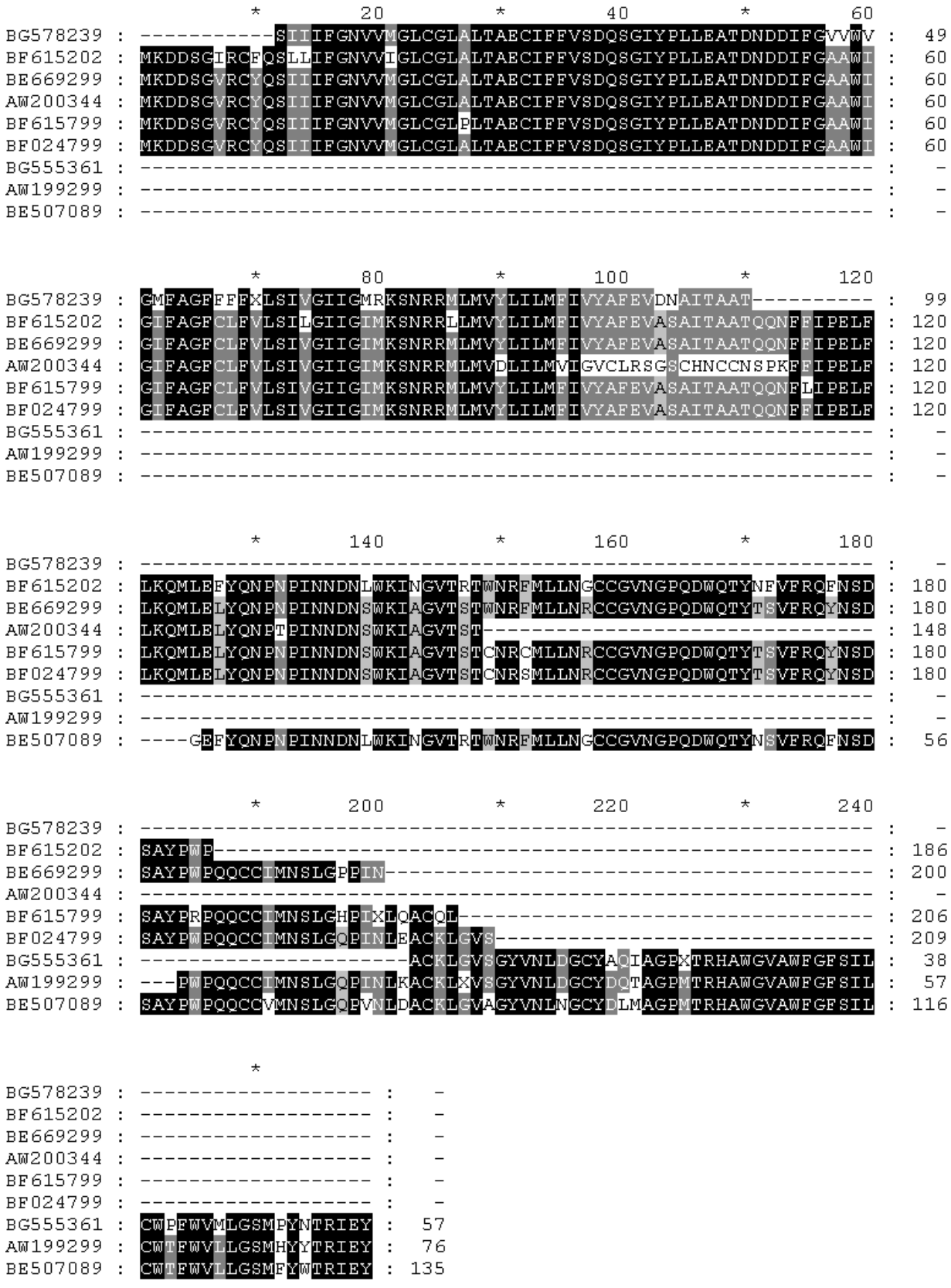
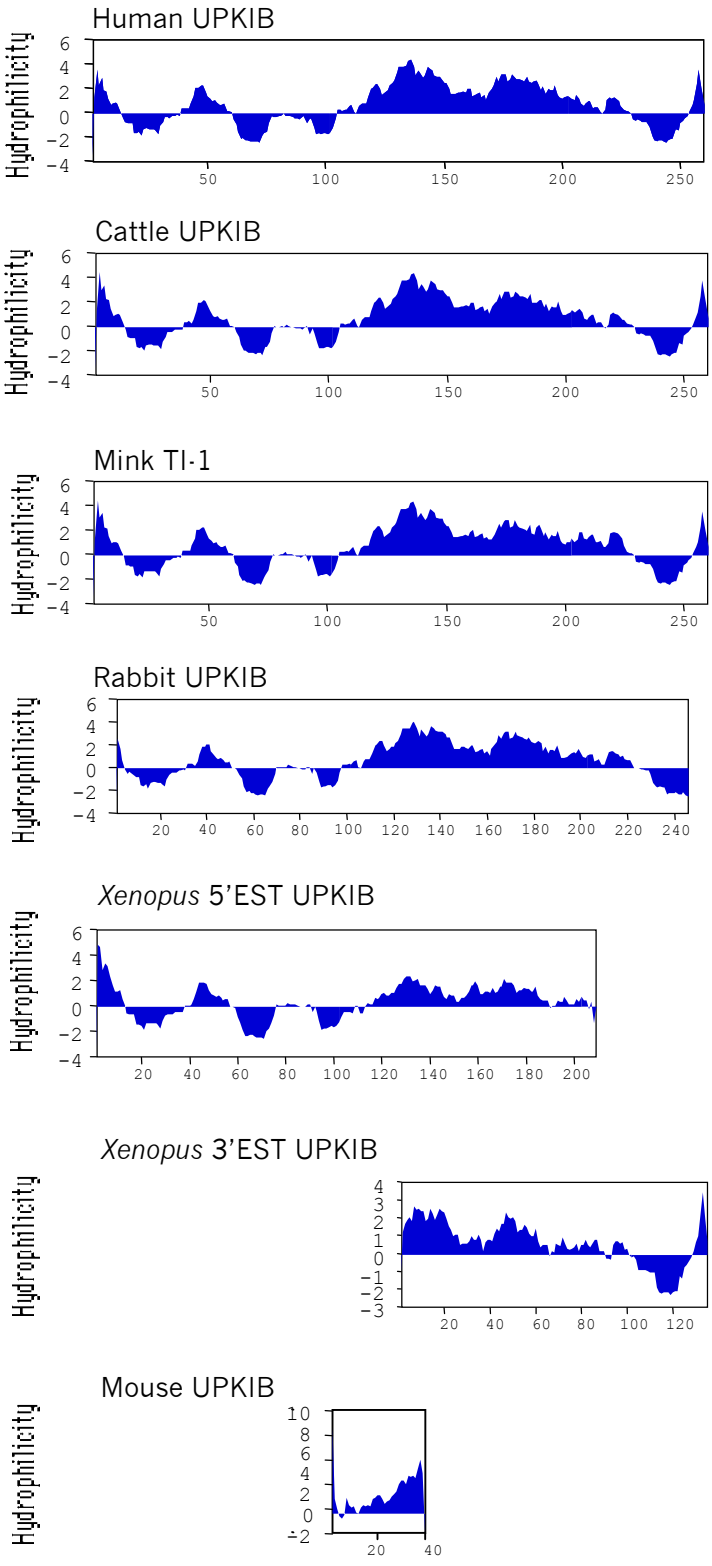


Figure 3.3

Goldman/Engelman/Steitz Hydrophilicity Plots (span 19) of UPKIB Members

Amino acid sequences from human, mink, cattle, rabbit, *Xenopus* and mouse were analysed using the MacVector hydrophilicity program. Goldman/Engelman/Steitz parameters were used with window length 19 to analyse amino acids within transmembrane domains. Negative hydrophilicity values indicate hydrophobic regions. All profiles were aligned manually to correspond to similar amino acid residues. Human, cattle and mink sequences were the only full-length open reading frames available.

Goldman/Engelman/Steitz hydrophilicity
Window = 19



amino acids (Fig. 3.2). The amino acid variations between these *Xenopus* sequences can be observed in Figure 3.3, with the hydrophilic profile in ECD2.

UPKIB is organised as: approximately 10-15 amino acids in CD1 (cytoplasmic domain), 20 amino acids in TM1 (transmembrane domain), 20 amino acids in ECD1 (extracellular domain), 10-15 amino acids in TM2, 20 amino acids in CD2, 20 amino acids in TM3, 125-130 amino acids in ECD2, 20 amino acids in TM4, and 5-10 amino acids in CD3.

3.3.1.4 Multiple sequence alignments

In the hydrophilicity data (Fig. 3.3), the *Xenopus* sequence appeared to have the most different plot compared to the other sequences. To study the variance of amino acid sequences, amino acid sequence alignments were initially performed with ClustalW (Appendix E). As can be seen in the alignment (Appendix E), the UPKIB sequences are highly conserved. However, only partial open reading frame sequences were available for rabbit, mouse, and *Xenopus*. To characterise further the amino acid variation between species, cDNA sequence alignments were carried out using AddGaps (Appendix F).

3.3.1.5 MP and BAMBE analyses of the UPKIB family

The UPKIB cDNA alignment was analysed using MP and BAMBE methods. The consensus MP trees are shown in rooted format in Figure 3.4, and in unrooted format in Figure 3.5. The BAMBE trees are shown in Figure 3.6 (rooted) and 3.7 (unrooted). For rooting trees, *Xenopus* UPKIB 5' and 3' sequences were chosen as outgroups because these amphibian sequences were most likely ancestral to the mammalian sequences. However, as referred to earlier (Section 3.1.1 & Fig. 3.1), the use of these sequences as outgroups assumes the sequences involved are

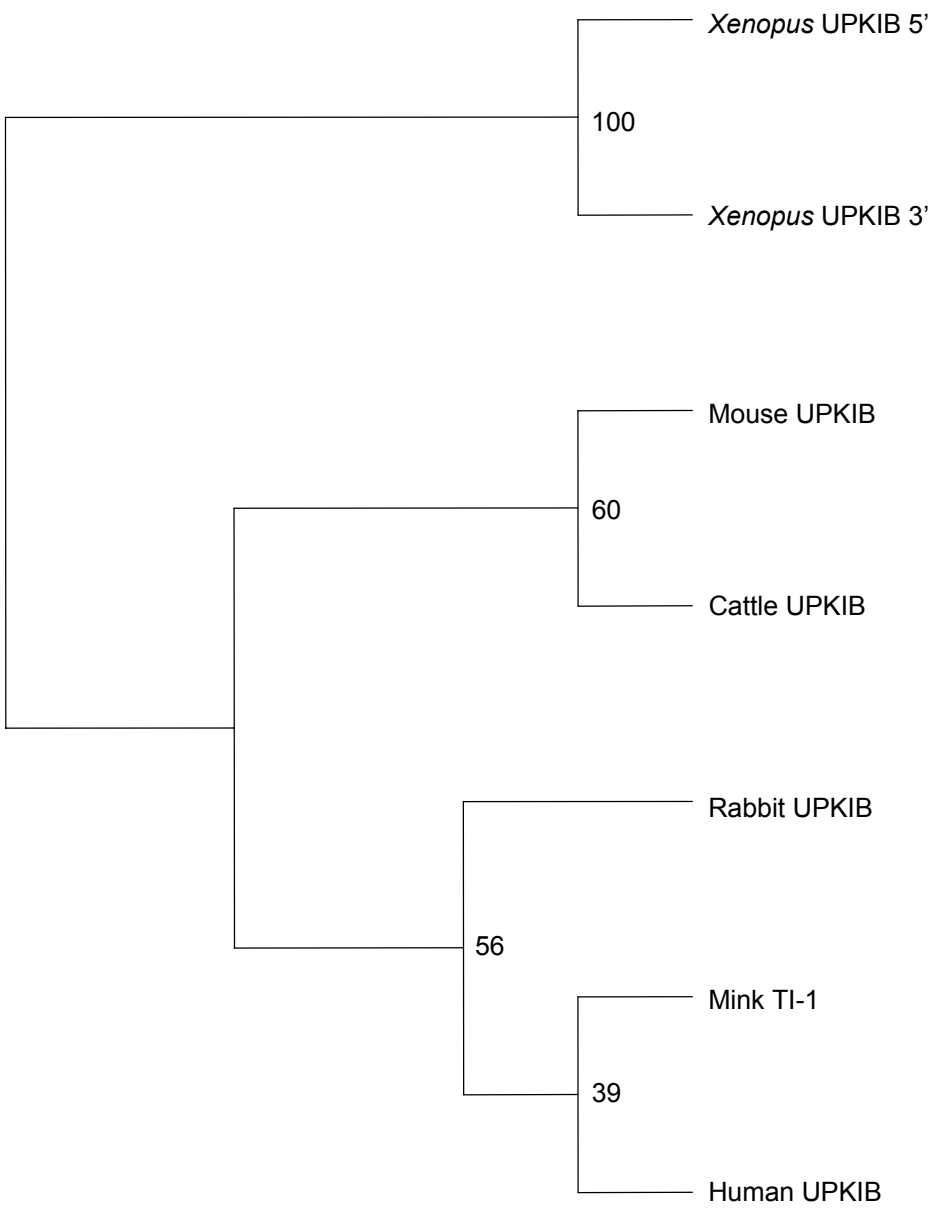


Figure 3.4

Consensus Maximum Parsimony Tree Based on Aligned cDNA Sequences from UPKIB Family Members

Bootstrap values (%) are based on 1 000 pseudoreplications of the data set. The tree has been rooted using *Xenopus* 5'EST and 3'EST sequences as outgroups. Appendix A contains the relevant UPKIB family members.

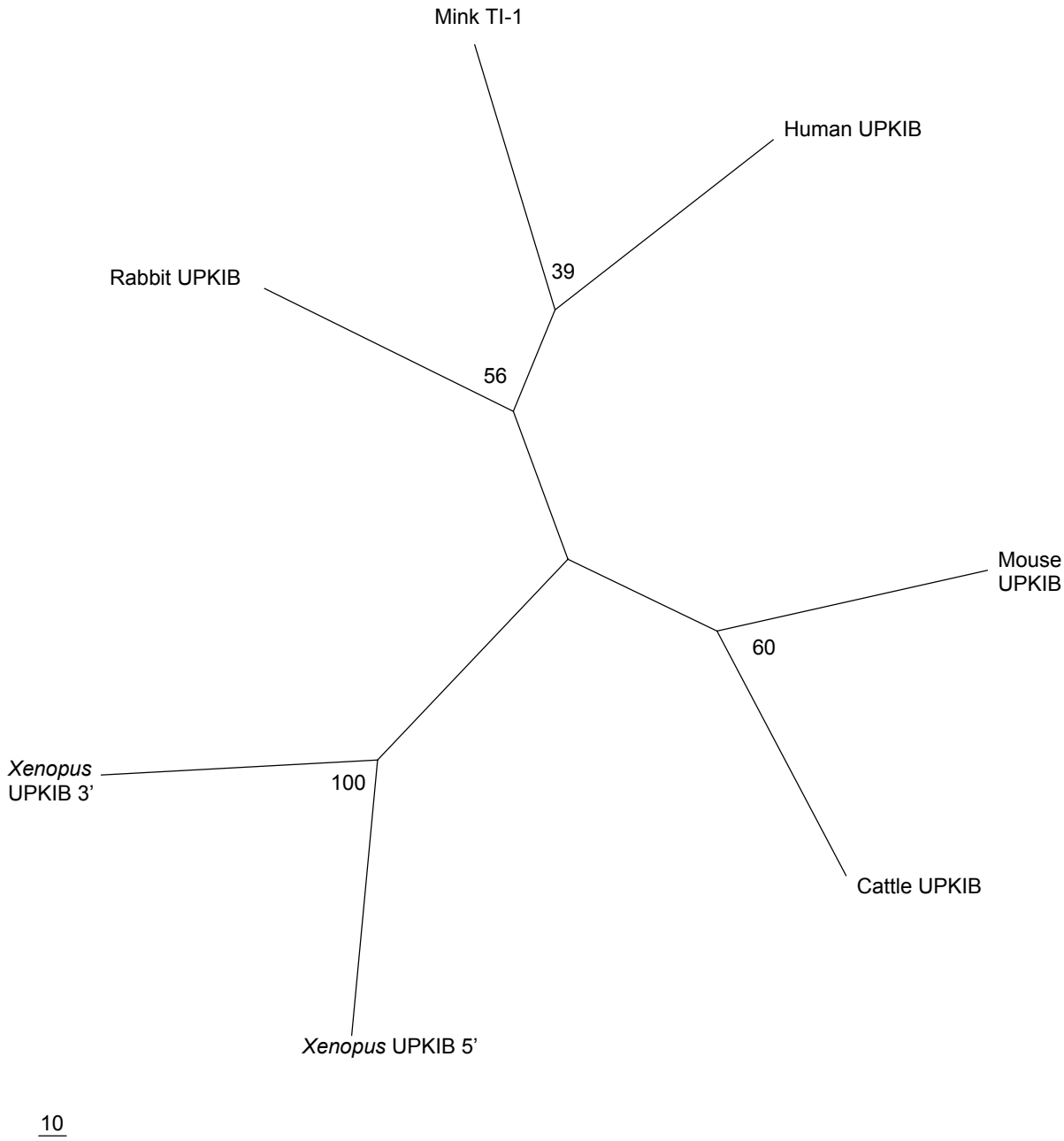


Figure 3.5

Unrooted, Consensus Maximum Parsimony Tree Based on Aligned cDNA Sequences from UPKIB Family Members

Bootstrap values (%) are based on 1 000 pseudoreplications of the data set. Appendix A contains the relevant UPKIB family members.

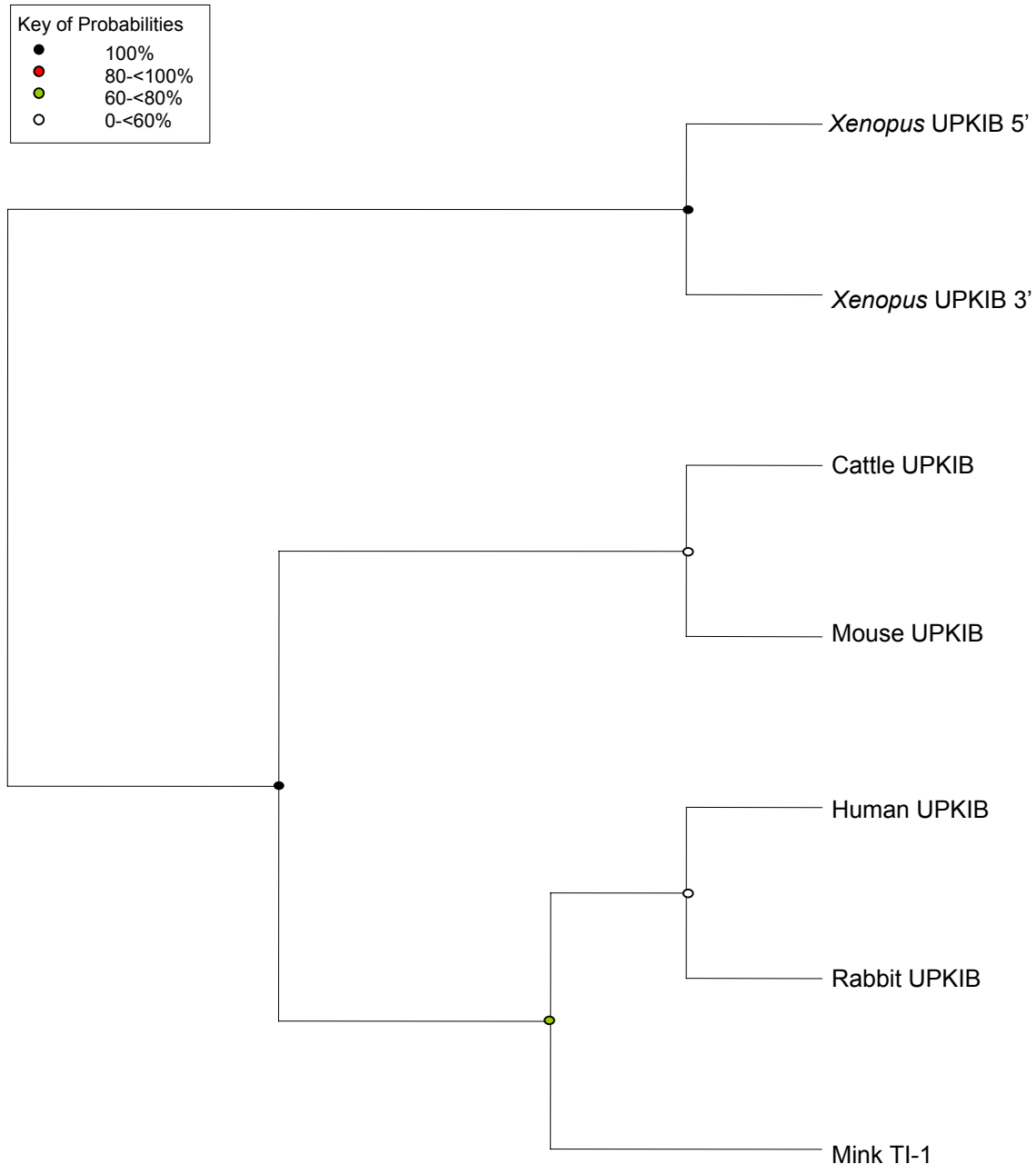


Figure 3.6

BAMBE Tree Based on Aligned cDNA Sequences from UPKIB Family Members

The tree has been rooted using *Xenopus* 5'EST and 3'EST sequences as outgroups. Branch lengths are drawn to scale (see key). Probabilities (%) are shown at each node. Appendix A contains the relevant UPKIB family members.

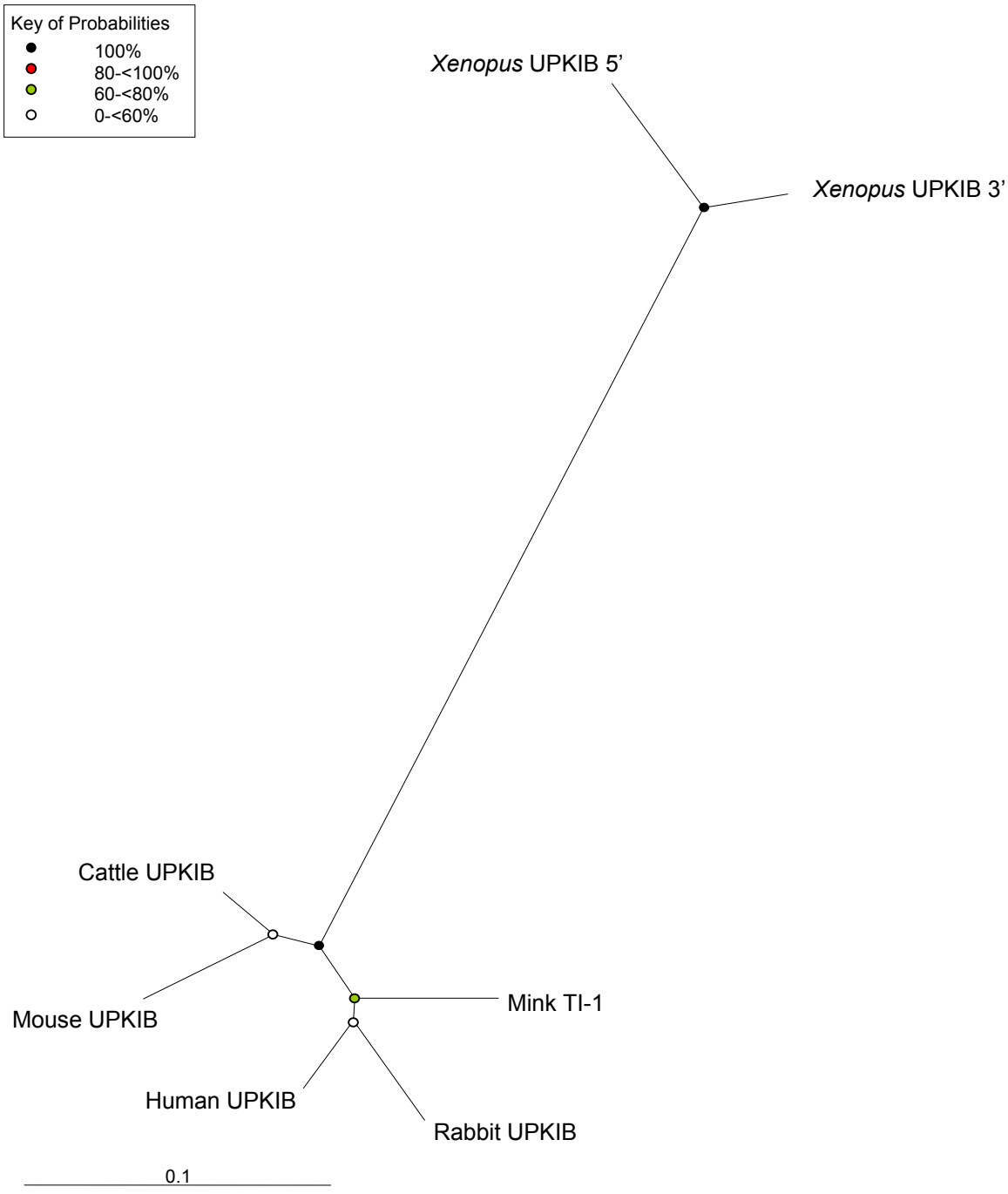


Figure 3.7

Unrooted BAMBE Tree Based on Aligned cDNA Sequences from UPKIB Family Members

Branch lengths are drawn to scale (see key). Probabilities (%) are shown at each node.

Appendix A contains the relevant UPKIB family members.

orthologous. It was because of this uncertainty that the MP and BAMBE trees were also presented in unrooted formats (Fig. 3.5 & 3.7)

3.3.2 Human tetraspanins

It was considered important to study the divergence of the tetraspanin family members within a single species, *Homo sapien*, to gauge the relationship of the tetraspanin paralogues. To this end, twenty-six human tetraspanins were compared.

3.3.2.1 Hydrophilicity plots of human tetraspanin family members

Using MacVector (v.6.5.3), Goldman/Engelman/Steitz hydrophilicity plots of human tetraspanins were analysed to compare the second extracellular domains (Fig. 3.8). Within the tetraspanin family, the second extracellular domains diverge between members, giving each protein its specificity (Section 1.5; review by Maecker *et al*., 1997).

As is characteristic of tetraspanins, there are four marked hydrophobic regions in all of the proteins, which represent the transmembrane domains (TMs) (Fig. 3.8). Other domains include the cytoplasmic- (CDs) and extracellular- domains (ECDs) (Section 1.5). The variation in size and hydrophilicity within the domains is distinct between the human tetraspanins (Fig. 3.8).

Human UPKIA is the tetraspanin most similar to human UPKIB in its hydrophilicity profile, as would be expected due to 39% sequence identity between cattle UPKIA and UPKIB amino acid sequences (Yu *et al*., 1994). The main differences between these two proteins were observed in the second extracellular domain (ECD2), with an apparent reduction in hydrophilicity at the centre of the domain.

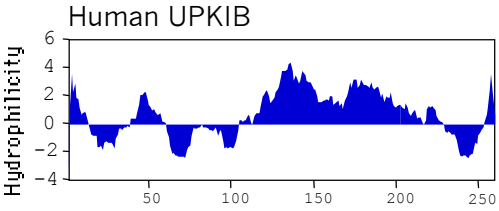
Figure 3.8

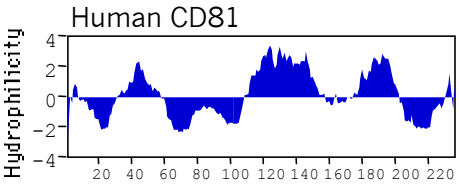
Goldman/Engelman/Steitz Hydrophilicity Plots (span 19) of Human Tetraspanins

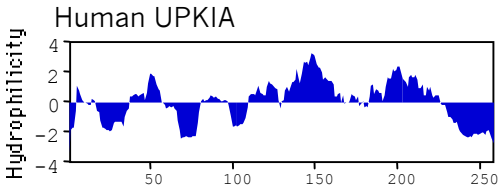
Amino acid sequences of twenty-six human tetraspanins were analysed using the MacVector hydrophilicity program. Goldman/Engelman/Steitz parameters were used with window length 19 to analyse amino acids within transmembrane domains. Negative hydrophilicity values indicate hydrophobic regions.

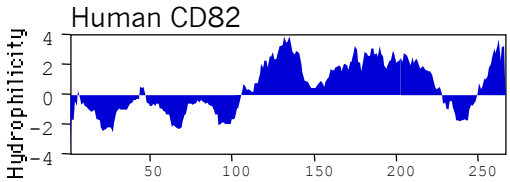
UPKIB is shown on the left above UPKIA. These two members have 34% sequence identity. Other human tetraspanin members are shown elsewhere in the figure. All profiles were manually adjusted and aligned to view hydrophilicity peaks and amino acid lengths proportionately for comparative analysis of members.

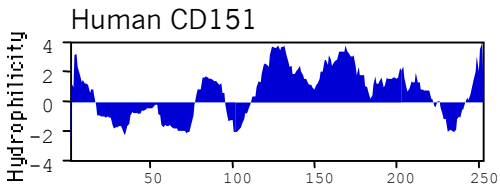
Goldman/Engelman/Steitz hydrophilicity
Window = 19

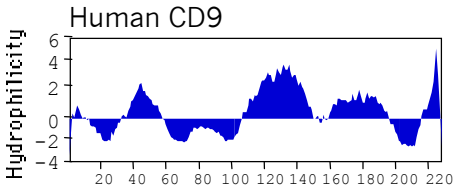


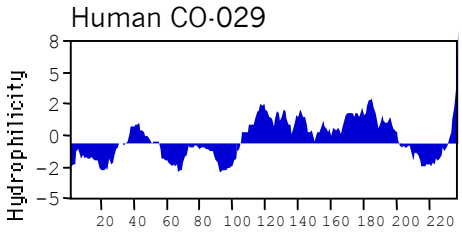


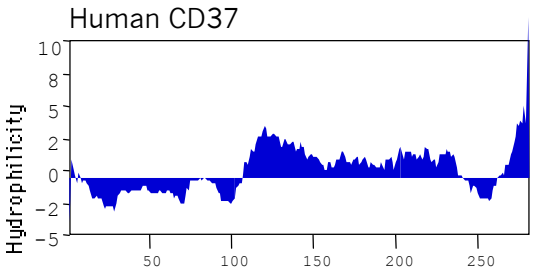


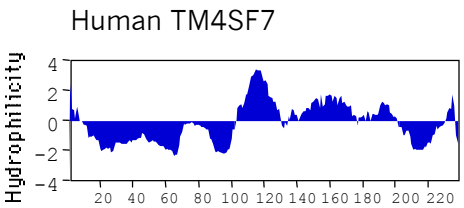


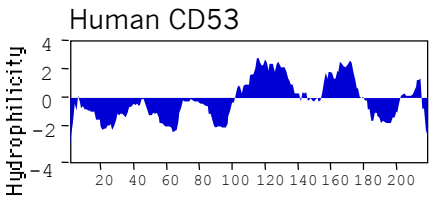


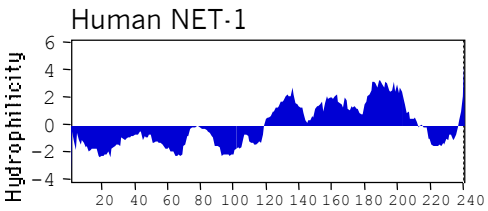


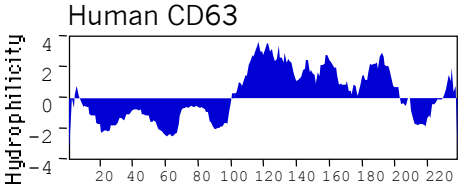












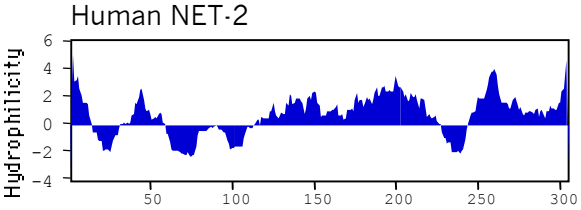
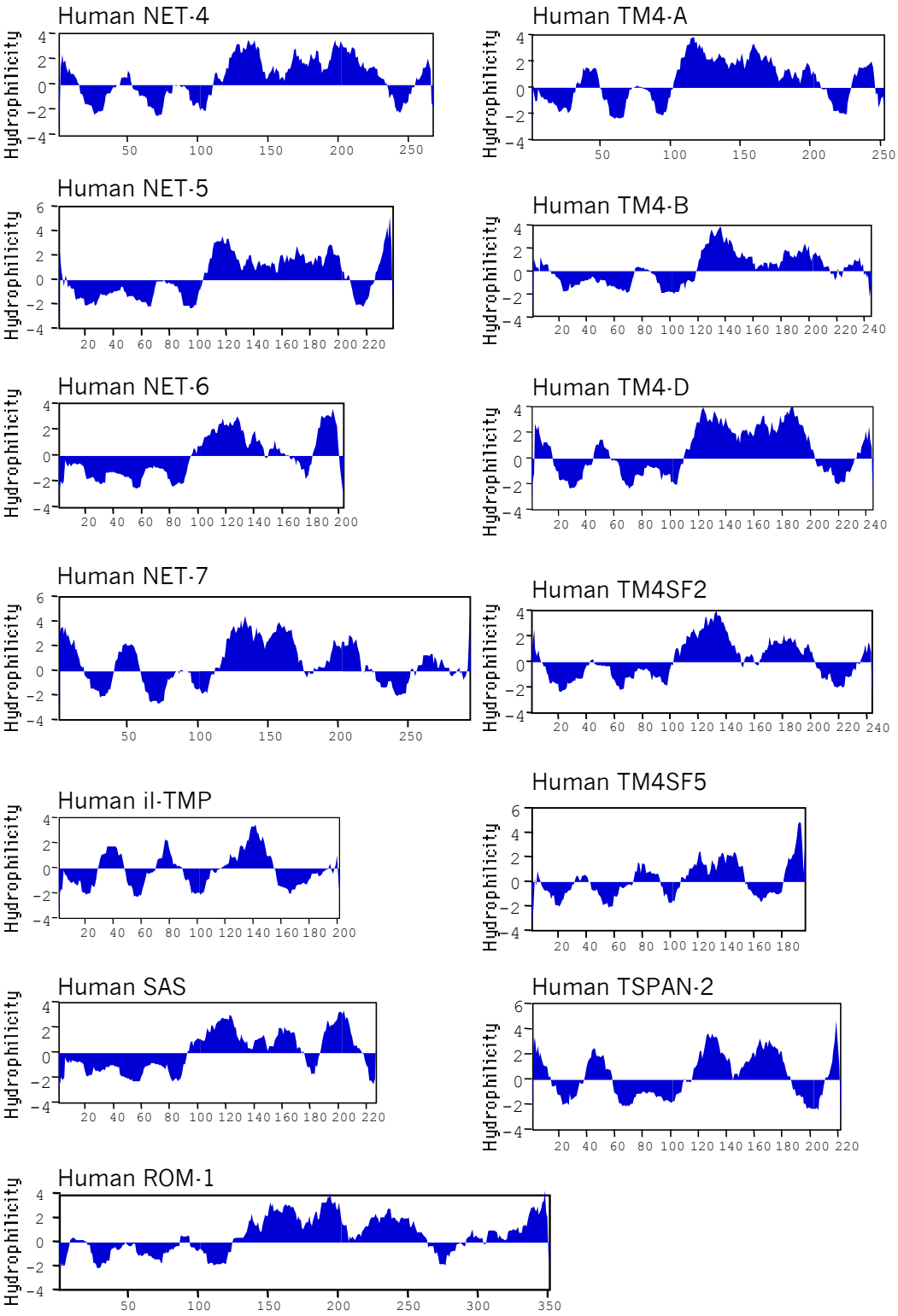


Figure 3.8 (continued)



Sizes of the domains ECD1 and ECD2 also appear to vary greatly between members. This data cannot show evolutionary trends between members, however the profiles contribute to an understanding of the relationship of the uroplakins to the other tetraspanins, as members of the tetraspanins. In this instance, the proteins are too varied for the plots to reveal any close association between most tetraspanin members.

3.3.2.2 Multiple sequence alignments

All human tetraspanin amino acid sequences were aligned with ClustalW (Appendix G), and the cDNA sequences for these proteins were aligned with AddGaps (Appendix H). In Appendix G, common motifs are highlighted in red and blue, with red highlights indicating previously reported conserved residues. The aligned cDNAs from AddGaps were subsequently used in MP and BAMBE analyses.

3.3.2.3 MP and BAMBE analyses of human tetraspanins

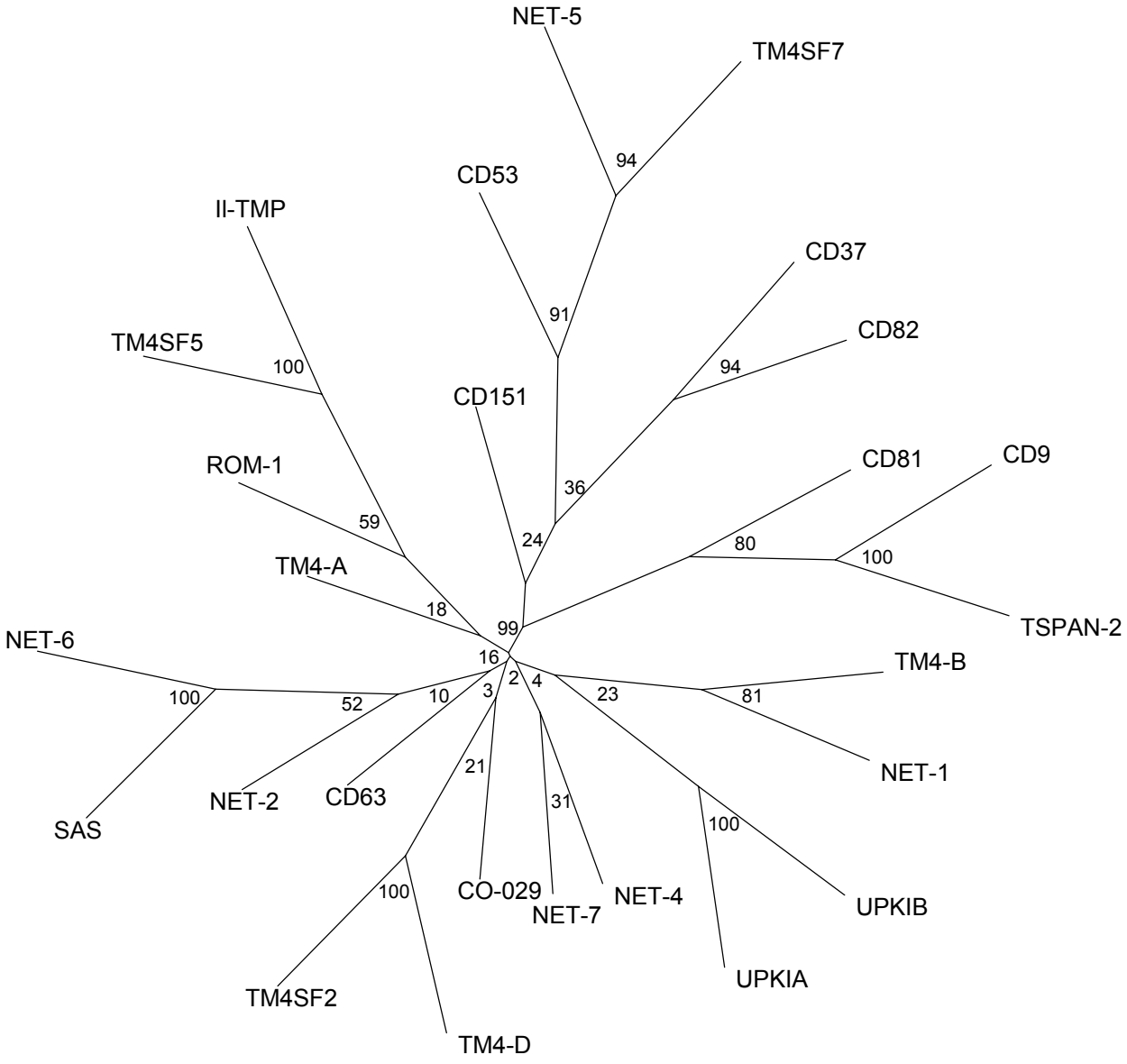
MP consensus and BAMBE trees, derived of human tetraspanin open reading frame cDNA sequences are shown in Figures 3.9 and 3.10 respectively.

3.3.3 Tetraspanin family members from a range of species

To study the tetraspanin sequences from a range of organisms and to examine the relationship of uroplakins IA and IB to more primitive tetraspanins, MP and BAMBE analyses were employed.

3.3.3.1 Multiple sequence alignments

Tetraspanin amino acid sequences from a range of organisms were aligned with ClustalW (Appendix I), and the cDNA sequences for these proteins were



10

Figure 3.9

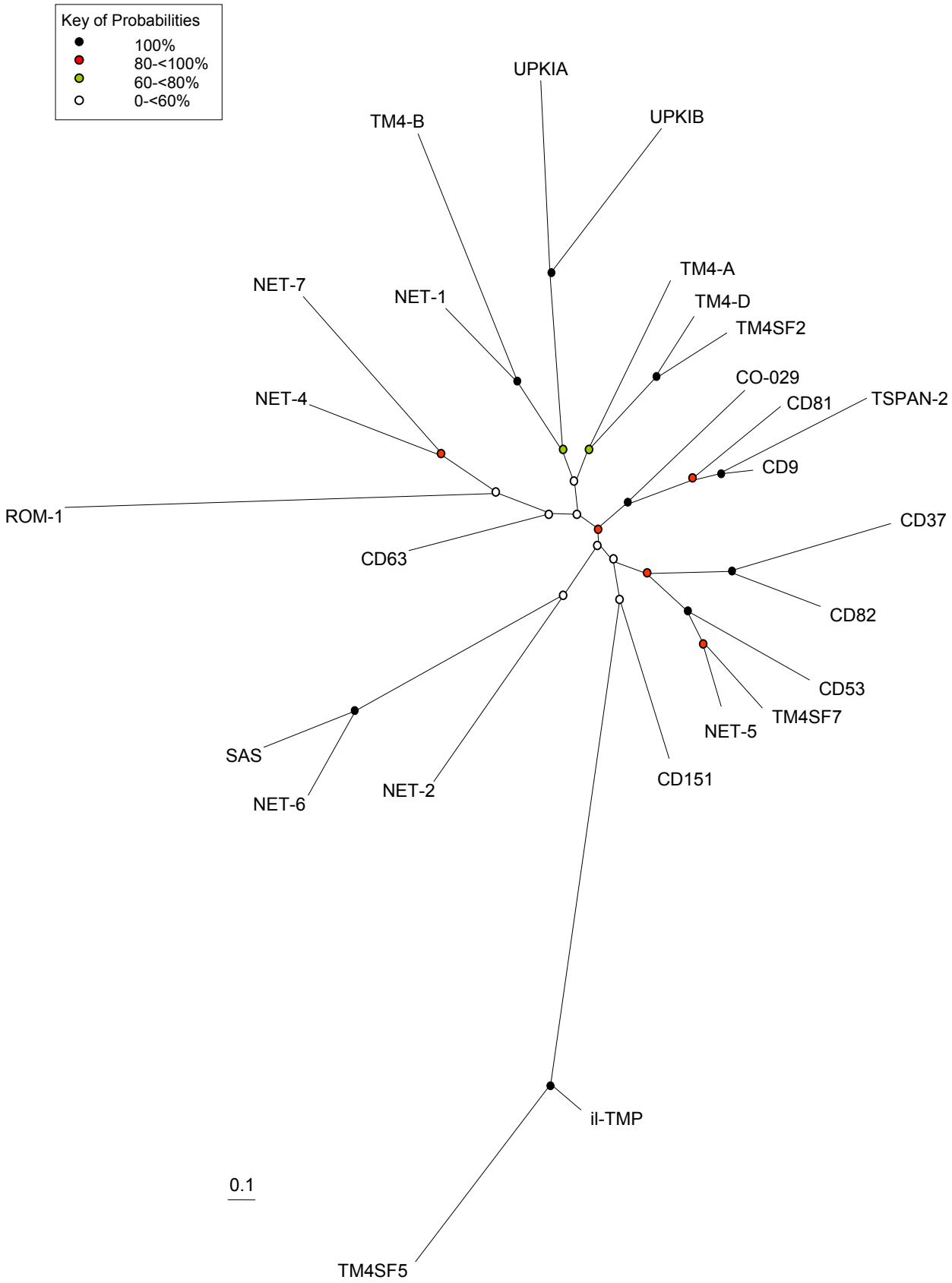
Consensus Maximum Parsimony Tree Based on 26 Aligned cDNA Sequences from Human Tetraspanin Family Members

Bootstrap values (%) are based on 1 000 pseudoreplications of the data set. The tree is unrooted. Appendix A contains the relevant human tetraspanin family members.

Figure 3.10

BAMBE Tree Based on Aligned cDNA Sequences from 26 Human Tetraspanin Family Members

The tree is unrooted. Branch lengths are drawn to scale (see key). Probabilities (%) are shown at each node. Appendix A contains the relevant human tetraspanin family members.


Key of Probabilities
100%
80-<100%
60-<80%
0-<60%
UPKIA
TM4-B
UPKIB
TM4-A
NET-7
NET-1
TM4-D
TM4SF2
CO-029
NET-4
CD81
TSPAN-2
CD9
ROM-1
CD37
CD63
CD82
CD53
TM4SF7
NET-5
SAS
CD151
NET-2
NET-6
il-TMP
0.1
TM4SF5

aligned with AddGaps (Appendix J). The aligned cDNAs were subsequently used in MP and BAMBE analyses.

3.3.3.2 MP analysis

MP analysis on the aligned tetraspanin cDNA sequences produced the phylogenetic tree shown in Figure 3.11.

3.3.3.3 BAMBE analysis

Analysis of the cDNA sequences using BAMBE produced the tree shown in Figure 3.12 (see also Appendix D). Similar to the MP tree where outer branches had higher bootstrap values, higher probabilities in BAMBE were observed in peripheral branches. In both MP and BAMBE data several groups were common, but as with the human tetraspanin (Section 3.3.2) and UPKIB family (Section 3.3.1) comparisons, there were slight differences in tetraspanins cDNA sequence placements in both MP and BAMBE methods.

3.3.3.4 Pairwise alignments of UPKIB and closely related tetraspanin amino acid sequences

As indicated in both MP and BAMBE data, UPKIA and UPKIB sequences are distantly related to a chicken tetraspanin. Pairwise sequence alignments of amino acid sequences were performed with BLAST2 (NCBI) (Section 2.3.4) both to provide further evidence of relatedness of these sequences, and to study the relationship of the *Xenopus* UPKIB homologue to both human UPKIB and UPKIA sequences.

Human UPKIA and UPKIB showed 40% sequence identity. There was 71% and 69% sequence identity between human UPKIB and with XenUPKIB5 and XenUPKIB3 sequences respectively, verifying the homologous nature of these

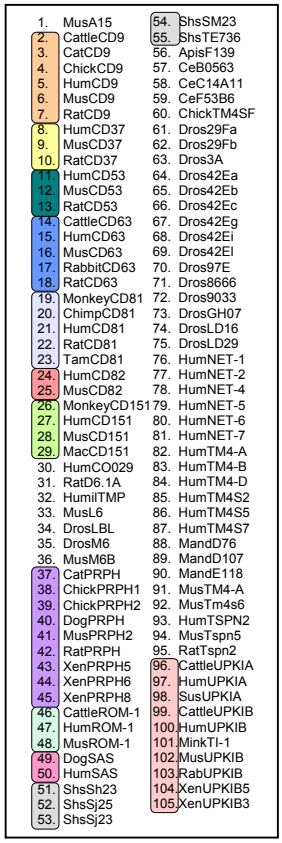
Figure 3.11

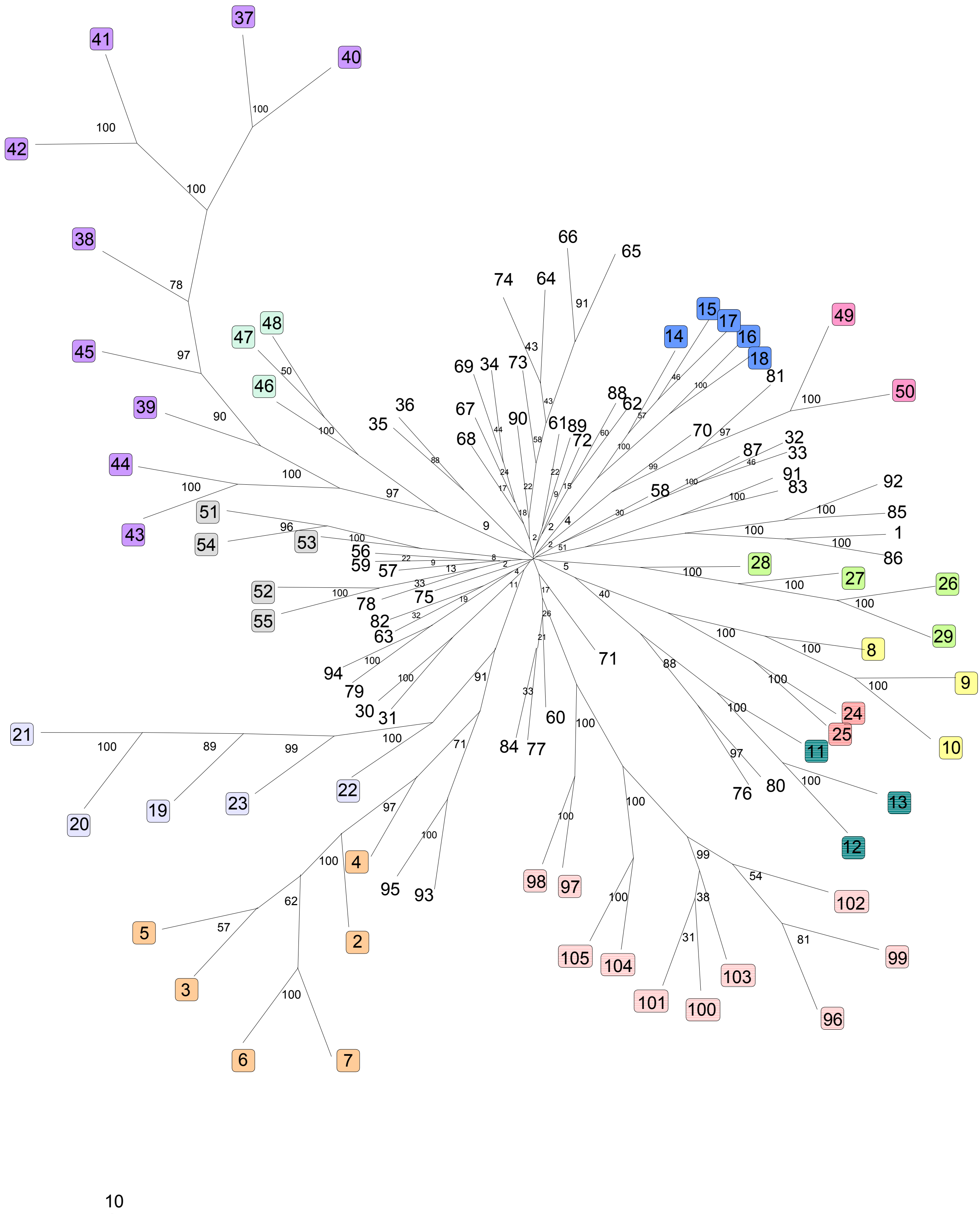
Consensus Maximum Parsimony Tree Based on Aligned cDNA Sequences from 105 Tetraspanin Family Members

Bootstrap values (%) are based on 1 000 pseudoreplicates of the data set. The tree is unrooted. Appendix A contains the relevant tetraspanin family members.

Sequence abbreviations:

Apis	*Apis mellifera* (honeybee)
Cat	*Felis catus*
Cattle	*Bos taurus*
Ce	*Caenorhabditis elegans*
Chick	*Gallus gallus*
Chimp	*Pan troglodytes*
Dog	*Canis familiaris*
Dros	*Drosophila melanogaster*
Hum	*Homo sapien*
Mac	*Macaca mulatta* (Rhesus monkey)
Mand	*Manduca sexta* (tobacco hornworm)
Mink	*Mustela vison*
Monkey	*Chlorocebus aethiops*
Mus	*Mus musculus*
Pig	*Sus scrofa*
Rabbit	*Oryctolagus cuniculus*
Rat	*Rattus norvegicus*
Shs	*Schistosoma haematobium* *Shistosoma japonicum* *Schistosoma mansoni*
Sus	*Sus scrofa*
Tam	*Saguinus oedipus* (Tamarind)
Xen	*Xenopus laevis*





41 37 40 42 38 45 39 44 43
48 47 46
51 54 53 52 55
21 20 19 23 22
5 3 6 7 4 2
15 17 16 18 14
49 50
28 27 26 29
8 9 10 24 25 11 13 12
98 97 105 104 101 100 103 102 99 96
10

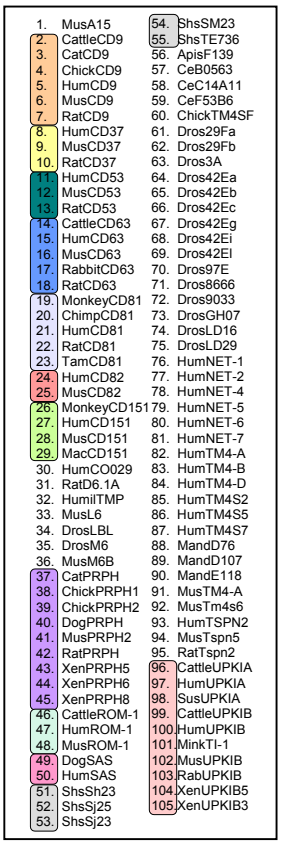
Figure 3.12

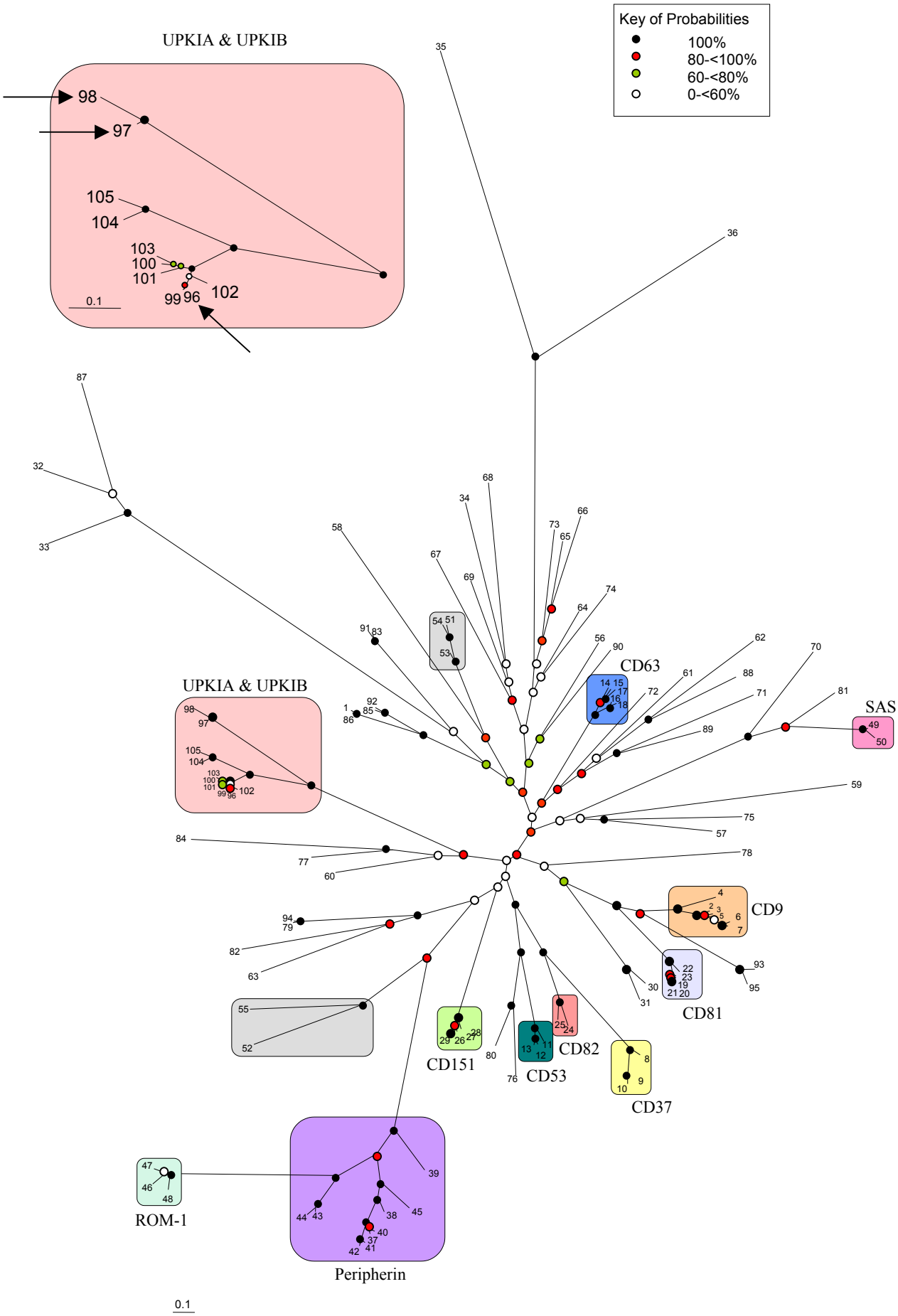
BAMBE Tree Based on Aligned cDNA Sequences from 105 Tetraspanin Family Members

Branch lengths are drawn to scale and probabilities (%) are shown at each node (see keys). Arrows indicate UPKIA sequences among the UPKIB sequences in the enlarged region of the UPKIB tetraspanins (top left corner). Appendix A contains the relevant tetraspanin family members.

Sequence abbreviations:

Apis	*Apis mellifera* (honeybee)
Cat	*Felis catus*
Cattle	*Bos taurus*
Ce	*Caenorhabditis elegans*
Chick	*Gallus gallus*
Chimp	*Pan troglodytes*
Dog	*Canis familiaris*
Hum	*Homo sapien*
Dros	*Drosophila melanogaster*
Mac	*Macaca mulatta* (Rhesus monkey)
Mand	*Manduca sexta* (tobacco hornworm)
Mink	*Mustela vison*
Monkey	*Chlorocebus aethiops*
Mus	*Mus musculus*
Pig	*Sus scrofa*
Rabbit	*Oryctolagus cuniculus*
Rat	*Rattus norvegicus*
Shs	*Schistosoma haematobium* *Shistosoma japonicum* *Schistosoma mansoni*
Sus	*Sus scrofa*
Tam	*Saguinus oedipus* (Tamarind)
Xen	*Xenopus laevis*




Key of Probabilities
100%
80-<100%
60-<80%
0-<60%
UPKIA & UPKIB
98
97
105
104
103
100
101
99
96
102
0.1
UPKIA & UPKIB
CD63
SAS
CD9
CD81
CD82
CD53
CD151
CD37
ROM-1
Peripherin
0.1

proteins. BLAST2 analyses revealed 38% and 37% sequence identity between human UPKIA and with XenUPKIB5 and XenUPKIB3 sequences respectively.

Human UPKIA and UPKIB both share 23% sequence identity with the chicken TM4SF sequence (GenBank accession AF206661). The chicken sequence was derived from the embryonic nervous system, a region where UPKIB expression has been detected in humans, in the neuroblastoma-derived cell line SK-N-MC (GenBank accession AU100584, Section 3.3.1.1).

3.4 Discussion

The uroplakins are a highly conserved group of proteins expressed in tissues including the mammalian bladder, the mink lung and the human ocular epithelium. They appear to be important as structural membrane proteins in the mammalian bladder (discussed in Sections 1.1, 1.3.3 & 1.3.4) and the present study has suggested UPKIB expression occurs in various other human tissues, with several ESTs detected (Section 3.3.1.1). This supports data from Olsburgh *et al*. (2003), indicating that this gene is expressed in a wide range of tissues (discussed in Sections 1.3.4 & 3.1). However, there are uncertainties surrounding the use of ESTs, as the presence of an EST does not necessarily imply expression. Several cDNA libraries that are used for EST identification are contaminated with DNA from other sources (Miller *et al*., 1999; White *et al*., 1993). Also, given the fragmentary nature of the data on the uroplakin expression patterns (Section 1.3.4), a thorough study should be carried out, perhaps using *in situ* hybridisation on tissue sections, to analyse the spatial and temporal expression patterns of the uroplakin IB gene.

The observation that UPKIB was expressed in the amphibian *Xenopus* (but not in *Drosophila* or other invertebrates) made it clear that this protein is not a strictly

"mammalian protein". In addition, no UPKIB ESTs, mRNA or genomic DNA sequences were detected in the Internet sequence-databases for birds. The *Xenopus* UPKIB 3' amino acid sequence (GenBank accession BF024799) retains up to 71% sequence identity with human UPKIB and the *Xenopus* UPKIB 5' amino acid sequence (BG507089) retains 69% sequence identity with human UPKIB. Sequence identity between UPKIA and UPKIB, two closely functionally and structurally related proteins is 39% in cattle (Yu *et al*., 1994) and 40% in humans, as shown through BLAST2 searches (Section 3.3.3.4).

In interpreting the various evolutionary trees (Fig. 3.4, 3.5, 3.6, 3.7, 3.9, 3.10, 3.11 & 3.12) presented in this Chapter, it was important to be aware of the uncertainty associated with the selection of outgroups, discussed earlier (Section 3.1.1). In constructing the UPKIB MP and BAMBE rooted trees (Fig. 3.4 & 3.6) it was assumed that the genes involved were orthologous and that *Xenopus* 5' and 3' sequences therefore provided appropriate outgroups, *Xenopus* being the most distantly related species amongst the group. The sequence similarities revealed by ClustalW (Appendix E) and the similarities in hydrophilicity plots (Fig. 3.3) provide some evidence to support the orthology of this group of genes. However, the MP and BAMBE trees were also shown in an unrooted format (Fig. 3.5 & 3.7). All other trees in this Chapter were shown in an unrooted format.

Analysis of UPKIB family members

The MP and BAMBE trees for UPKIB (Fig. 3.4 – 3.7) were poorly resolved, with either low bootstrap values (MP) or low probabilities (BAMBE), making conclusions problematical. Both forms of analysis show the mouse and cattle sequences forming a clade, to the exclusion of a second clade containing mink, rabbit

and human sequences. If all these genes were orthologues, the indicated species arrangement would be incorrect. The most likely explanation is that the genes are indeed orthologues and that there is insufficient sequence data to correctly resolve the branching arrangements. However, the possibility that the sequences contain a mix of orthologous and paralogous genes cannot be discounted.

The analysis of uroplakins in this study was to redefine the role of UPKIB in evolution of the bladder in mammals and to find other organisms expressing UPKIB mRNA. Several partial transcripts were obtained with high sequence identity to human UPKIB. This finding, together with the fact that *Xenopus* has a bladder, implies a functional role of UPKIB in organs secreting nitrogenous waste from the body and in the development of a bladder among certain organisms. The presence of a UPKIB homologue in *Xenopus* suggests that the UPKIB mRNA is required for creation of UPKIB proteins expressed in the *Xenopus* urinary bladder and that to have a structurally functional bladder, UPKIB is required. It would be very interesting to perform functional and expression studies of UPKIB in *Xenopus*, as the UPKIB protein may have a quite different role in this species, which may give some clues as to its evolution of function.

With the finding that UPKIB ESTs and mRNA were detected in a range of human tissue, it would suggest that UPKIB is an important strengthening and stabilising protein within a range of epithelial cells, as it is in the urothelium in bladder distension and contraction in mammals. This strength and flexibility would explain the existence of UPKIB in both the mink lung (Kallin *et al*., 1991), providing strength and flexibility in gas absorption and emission, and in the human eye (Adachi *et al*., 2000), containing a strong outer membrane required for the maintenance of eye shape and stability.

Analysis of human tetraspanin family members

The evolutionary relationships between UPKIB and other human tetraspanin paralogues are set out in Figures 3.9 (MP tree) and 3.10 (BAMBE tree). It can be seen from the MP tree (Fig. 3.9) that many of the nodes, particularly those near the hub of the tree, have low bootstrap values, and therefore little significance can be attached to them. However, some of the groupings were associated with high bootstrap values (arbitrarily defined here as 80% or over) and will be treated as "significant". On this basis, there were seven groups of closely related genes:

(i) UPKIA and UPKIB (as expected – see Yu *et al*., 1994);
(ii) CD37 and CD82;
(iii) CD53, NET-5 and TM4SF7;
(iv) CD81, CD9 and TSPAN-2;
(v) TM4SF2 and TM4-D;
(vi) NET-6 and SAS; and
(vii) TM4SF5 and il-TMP.

In relation to the BAMBE Tree (Fig. 3.10), if we arbitrarily define a "high" ("significant") probability as being one of 80% or greater, then all seven of the groups identified by MP analysis (see above) were also identified in the BAMBE analysis. Two of these groupings are enlarged as follows:

CO-029 is added to group (iv), which now contains 4 genes: CD81, CD9, TSPAN-2 and CO-029.
Groups (ii) and (iii) are combined into a single clade of 5 genes: CD37, CD82, CD53, NET-5 and TM4SF7.

Two additional groups were apparent from the BAMBE analysis: NET-7 and NET-4 form a clade (viii); and TM4-B and NET-1 form a clade (ix). Another notable feature of the BAMBE tree was the presence of a long branch, connecting the

TM4SF5 and il-TMP group to the remainder of the genes. This suggests that there has been rapid evolution along this branch, assuming that the root of the tree does not lie here. As scarce information is known about these particular proteins, there is nothing unusual or common about these proteins that fits in with their rapid rate of sequence divergence compared with the others tetraspanin family members.

The overall evolutionary pattern to emerge from these analyses is that a number of ancient gene duplications gave rise to at least 8 clades. More recent duplication events within each clade gave rise to more closely related sets of genes.

The clade containing CD37, CD82 and CD53 (Fig. 3.9 & 3.10) supports previous suggestions by Wright *et al*. (1993) that a duplication event occurred between members of the tetraspanin family to evolve the tetraspanins, as seen with the group containing CD53 and CD37, which are closely related (Angelisova *et al*., 1990). The data in this Chapter supports the possibility of such gene duplications and it is possible to refine the evolutionary model by suggesting that the evolutionary relationships are:

(((CD37)(CD82))(CD53 (NET-5, TM4SF7)))

Analysis of tetraspanins from a range of organisms

It can be seen that branches near the hub of the MP tree (Fig. 3.11) were poorly resolved with low bootstrap values. If we take a bootstrap value around 80% or more as indicating significant phylogenetic grouping, then the following groups of related genes are apparent:

(i) ((*Bos taurus* ROM-1 (*Homo sapien* ROM-1, *Mus musculus* ROM-1))((*Xenopus laevis* PRPH6, *Xenopus laevis* PRPH6)(*Gallus gallus* PRPH2 (*Xenopus laevis* PRPH8 (*Gallus gallus* PRPH1 ((*Mus musculus* PRPH2, *Rattus norvegicus* PRPH)(*Felis catus* PRPH, *Canis familiaris* PRPH)))))));

(ii) ((*Rattus norvegicus* CD81 (*Saguinus oedipus* CD81 (*Chlorocebus aethiops* CD81 (*Homo sapien* CD81, *Pan troglodytes* CD81))))(*Gallus gallus* CD9 (*Bos taurus* CD9 (*Mus musculus* CD9, *Rattus norvegicus* CD9)(*Felis catus* CD9, *Homo sapien* CD9))));

(iii) (*Drosophila melanogaster* M6, *Mus musculus* M6B);

(iv) (*Mus musculus* CD63, *Rattus norvegicus* CD63);

(v) (*Drosophila melanogaster* 97E (*Homo sapien* NET-7 (*Canis familiaris* SAS, *Homo sapien* SAS)));

(vi) (*Homo sapien* TM4-B, *Mus musculus* TM4-A);

(vii) ((*Mus musculus* Tm4sf6, *Homo sapien* TM4SF2)(*Mus musculus* A15, *Homo sapien* TM4SF5));

(viii) (*Mus musculus* CD151 (*Homo sapien* CD151 (*Macaca mulatta* CD151, *Chlorocebus aethiops* CD151)));

(ix) ((*Homo sapien* CD82, *Mus musculus* CD82)(*Homo sapien* CD37 (*Mus musculus* CD37, *Rattus norvegicus* CD37)));

(x) ((*Homo sapien* NET-1, *Homo sapien* NET-6)(*Homo sapien* CD53 (*Mus musculus* CD53, *Rattus norvegicus* CD53)));

(xi) (*Homo sapien* UPKIA, *Sus scrofa* UPKIA);

(xii) (*Xenopus laevis* UPKIB 5', *Xenopus laevis* UPKIB 3');

(xiii) (*Bos taurus* UPKIA, *Bos taurus* UPKIB);

(xiv) (*Homo sapien* TSPAN2, *Rattus norvegicus* Tspan2);

(xv) (*Homo sapien* CO-029, *Rattus norvegicus* D6.1A);

(xvi) (*Homo sapien* NET-5, *Mus musculus* Tspan5);

(xvii) (*Shistosoma japonicum* Sj25, *Shistosoma japonicum* TE736);

(xviii) (*Shistosoma japonicum* Sj23 (*Shistosoma haematobium* Sh23, *Shistosoma mansoni* SM23)); and

(xix) (*Drosophila melanogaster* 42Eb, *Drosophila melanogaster* 42Ec)

The BAMBE tree (Fig. 3.12) reveals a similar overall evolutionary topology to the MP tree (Fig. 3.11). If we arbitrarily define a "high" ("significant") probability as being one of 80% or greater, then all nineteen of the groups identified by MP analysis (see above) were also identified in the BAMBE analysis. Seven of these groupings are enlarged as follows:

Bos taurus CD63, *Homo sapien* CD63, *Oryctolagus cuniculus* CD63 and *Drosophila melanogaster* 9033 are added to group (iv), which now contains 6 genes: *Mus musculus* CD63, *Rattus norvegicus* CD63, *Bos taurus* CD63, *Homo sapien* CD63, *Oryctolagus cuniculus* CD63 and *Drosophila melanogaster* 9033.

Drosophila melanogaster 3A and *Homo sapien* TM4-A are added to group (xvi), which now contains 4 genes: *Homo sapien* NET-5, *Mus musculus* Tspan5, *Drosophila melanogaster* 3A and *Homo sapien* TM4-A.

Caenorhabditis elegans C14A11 is added to group (xviii), which now contains 4 genes: *Shistosoma japonicum* Sj23, *Shistosoma haematobium* Sh23, *Shistosoma mansoni* SM23 and *Caenorhabditis elegans* C14A11.

Drosophila melanogaster GH07 is added to group (xix), which now contains 3 genes: *Drosophila melanogaster* 42Eb, *Drosophila melanogaster* 42Ec and *Drosophila melanogaster* GH07.

Groups (i) and (xvii) are combined into a single clade of 14 genes: *Bos taurus* ROM-1, *Homo sapien* ROM-1, *Mus musculus* ROM-1, *Xenopus laevis* PRPH6, *Xenopus laevis* PRPH6, *Gallus gallus* PRPH2, *Xenopus laevis* PRPH8, *Gallus gallus* PRPH1, *Mus musculus* PRPH2, *Rattus norvegicus* PRPH, *Felis catus* PRPH, *Canis familiaris* PRPH, *Shistosoma japonicum* Sj25, *Shistosoma japonicum* TE736.

Groups (ii) and (xiv) are combined into a single clade of 13 genes: *Rattus norvegicus* CD81, *Saguinus oedipus* CD81, *Chlorocebus aethiops* CD81, *Homo sapien* CD81, *Pan troglodytes* CD81, *Gallus gallus* CD9, *Bos taurus* CD9, *Mus musculus* CD9, *Rattus norvegicus* CD9, *Felis catus* CD9, *Homo sapien* CD9, *Homo sapien* TSPAN2, *Rattus norvegicus* Tspan2.

Groups (ix) and (x) are combined into a single clade of 10 genes: *Homo sapien* CD82, *Mus musculus* CD82, *Homo sapien* CD37, *Mus musculus* CD37, *Rattus norvegicus* CD37, *Homo sapien* NET-1, *Homo sapien* NET-6, *Homo sapien* CD53, *Mus musculus* CD53, *Rattus norvegicus* CD53.

Four additional groups were apparent from the BAMBE analysis:

Drosophila melanogaster 29Fb and *Manduca sexta* D76 form a clade (xvii);

Drosophila melanogaster 8666 and *Manduca sexta* D107 form a clade (xviii);

Caenorhabditis elegans B0563 and *Drosophila melanogaster* LD29 form a clade (xix); and

Homo sapien NET-2 and *Homo sapien* TM4-D form a clade (xx).

Another notable feature of the BAMBE tree was the presence of two long branches, (i) connecting the *Drosophila melanogaster* M6 and *Mus musculus* M6B group and (ii) connecting the less significant *Homo sapien* il-TMP, *Mus musculus* L6

and *Homo sapien* TM4SF7 group to the remainder of the genes. This suggests that there has been rapid evolution along these branches, assuming that the root of the tree does not lie within these regions.

The overall evolutionary pattern to emerge from these analyses is that a number of ancient gene duplications gave rise to at least 19 clades. More recent duplication events within each clade gave rise to more closely related sets of genes.

It is interesting to note that the PRPH (peripherin) and ROM-1 (rod outer segment protein 1) sequences appear closely related (Fig. 5.11 & 5.12). The respective proteins have both been localised to eye rod disks using antibodies (Bascom *et al*., 1990; Travis *et al*., 1991) and these proteins form heterodimers with each other *in vivo* (Kedzierski *et al*., 1999). Within this group several *Shistosoma* sequences were also displayed in the BAMBE analysis (Fig. 5.12). Also of interest are the sequences with cluster of differentiation (CD) nomenclature, which have roles in the immunological field (see below). Some of these sequences appear comparatively distantly related, as seen in the BAMBE analysis (Fig. 5.12) with groups containing sequences for (i) CD53, CD37 and CD82; and (ii) CD81 and CD9.

Maecker *et al*. (1997), Hemler (2001) and Boucheix & Rubinstein (2001) compared the various tetraspanin sequences (Section 3.1). In these studies, uroplakins IA and IB did not appear to have clear relationships with other tetraspanins. Maecker *et al*. (1997) implied from their dendrogram that UPKIA and UPKIB were related to ocular proteins PRPH and ROM-1. This present study has shown that when comparing tetraspanins in a range of species, uroplakins IA and IB appear to be a unique group. In BAMBE analysis (Fig. 3.12), the uroplakins formed a clade and showed distant relationships with sequences from *Schistosoma*,

peripherins, ROM-1, CD151, CD53, CD82, CD37, CO-029, rat D6.1A, CD81 and CD9. There were no human peripherins for analysis, so comparison with previous findings was somewhat skewed. It was interesting to note that most *Drosophila* and *C. elegans* sequences were closely related, implying a functional group of early invertebrate tetraspanins. These sequences, interestingly, were closely related to both CD63 and SAS, as observed in both MP and BAMBE results (Fig. 3.11 & 3.12).

The analysis of larger data sets (Fig. 3.11 & 3.12) did not alter the original clades shown in UPKIB analyses for MP or BAMBE analyses (Fig. 3.4 - 3.7), with MP and BAMBE trees varying slightly in their UPKIB sequence organisation. MP trees (Fig. 3.4 & 3.5) for UPKIB members and for a range of tetraspanins (Fig. 3.11) revealed a clade with UPKIB arrangement:

((*Xenopus laevis* UPKIB 5', *Xenopus laevis* UPKIB 3')(*Mus musculus* UPKIB, *Bos taurus* UPKIB)((*Oryctolagus cuniculus* UPKIB)(*Mustela vison* TI-1, *Homo sapien* UPKIB)))

and BAMBE trees (Fig. 3.6 & 3.7) for UPKIB members and for a range of tetraspanins (Fig. 3.12) revealed a clade with UPKIB arrangement:

((*Xenopus laevis* UPKIB 5', *Xenopus laevis* UPKIB 3')(*Mus musculus* UPKIB, *Bos taurus* UPKIB)((*Mustela vison* TI-1)(*Oryctolagus cuniculus* UPKIB, *Homo sapien* UPKIB)))

As shown by the underlined sequences in these arrangements, the positions of rabbit and mink sequences varied between MP and BAMBE analyses.

UPKIB appeared paralogous with tetraspanin UPKIA, and was more distantly related to a chicken TM4SF sequence (GenBank accession AF206661) (Section 3.3.3.4). The association of the chicken sequence may show a re-adaptation of the

UPKIB protein within the avian system, however the sequence identity at 23% does not appear significant.

Among human studies (Section 3.3.2), UPKIA and UPKIB appeared most closely related to human tetraspanins NET-1 and TM4-B, however in studies from a range of organisms (Section 3.3.3), they appeared related to human NET-2 and TM4-D. Differences may be attributed to the number of sequences compared.

Tetraspanin nomenclature

There is no common gene/protein nomenclature among the tetraspanins, with names including UPK, CO-029, il-TMP, NET, ROM-1, PRPH and SAS. Several members of the family: CD9, CD53, CD63, CD81, CD82 (KAI-1) and CD151 also do not reflect the 4-transmembrane structures of the proteins, but reflect their roles in the immunological field as cluster of differentiation (CD) antigens.

Ideally, the system of nomenclature applied to the genes should reflect their evolutionary relationships, and it is apparent from Figures 3.9 and 3.10 that a number of nomenclature anomalies exist. If for no other reasons than one of improved communication between workers in the field, an attempt should be made to re-name a number of the genes in the tetraspanin family.

Members including human TM4SF2, 5 and 7 and TM4-A, B and D reflect their association with the tetraspanin family. However, these sequences appear to have been named in order of discovery and not in respect to any paralogues within the tetraspanin data set. For instance, in both MP (Fig. 3.11) and BAMBE (Fig. 3.12) analyses there is a strong association of members MusTm4sf6 (92), HumTM4SF2 (85), MusA15 (1) and HumTM4SF5 (86) in the arrangement:

((*Mus musculus* Tm4sf6, *Homo sapien* TM4SF2)(*Mus musculus* A15, *Homo sapien* TM4SF5))

In this instance, the clade may reflect a gene duplication event. Mouse Tm4sf6 may be a homologue of human TM4SF2, and similarly, mouse A15 may be a homologue of human TM4SF5. However, the current nomenclature does not reflect their evolutionary nature. An analysis of the tetraspanin data set indicates that the system of nomenclature that has been adopted by the scientific community should be revised to make more evolutionary sense. The tetraspanin superfamily is in itself a large family of proteins with similar amino acid composition, gene structure and possible function (Section 1.5). The members should reflect these characteristics with suitable nomenclature.

The evolutionary analysis that was reported in this Chapter was designed as a preliminary analysis aimed at detecting major features. It is appreciated that additional analyses, such as relative rates tests and tests for "saturation" of nucleotide sequence divergences could be undertaken in the longer term. The main features that the analyses have uncovered were that: (i) a non-mammalian UPKIB homologue was found in the amphibian *Xenopus laevis*, (ii) UPKIA and UPKIB formed a clade to the exclusion of all other tetraspanins, and (iii) several groups of closely-related tetraspanins were apparent. This study has also noted that the nomenclature currently in use for tetraspanins is inadequate for both their evolutionary relationships and conserved protein structure.