Evolutionary Comparison In The Vertebrate Lineage of *WT1*, a Wilms' Tumour Predisposition Gene

Gillian R.L. Kent

Ph.D.

University of Edinburgh

and

MRC Human Genetics Unit

1994

Declaration

I declare

a) This thesis was composed by myself alone

b) The work herein is my own unless clearly stated

In memory of my grandmother

Lavinia Paterson

Acknowledgements

There are many people without whom this project would never have got off the ground. Working on so many species would not have been possible without very many generous gifts from around the world. Hopefully I have managed to mention all of these within the text but there are several people I would particularly like to thank.

Anne Marie Coriat for the alligator *in situ* data.

Trevor Jowett for the zebrafish material and also allowing me to spend a week in a contamination-free lab.

Stephen Wilcox and Jenny Marshall Graves for screening the marsupial library.

David Housman, Daniel Haber and Alan Buckler for the mouse and human *WT1* cDNA clones.

P.A. Tsonis for sending me the axolotl *WT1* sequence.

During the last three and a half years at the MRC there have been an immense number of kind people who have been very generous with both their time and their advice. Among those I probably pestered the most, pinched their solutions, borrowed their books etc. are;

Duncan Davidson, Jonathan Bard, Bob Hill for trying to teach me development.

Isabel Hanson, Ian Jackson, Wendy Bickmore, Anne Seawright, Andreas Schedl, Cathy Abbott for excellent technical advice.

Lesley McInnes for teaching me the whole-mount technique.

John Inglis for advice on degenerate PCR and evolution analysis.

Also without the help of the following I probably would have lost a lot more sleep over this thesis.

Kathy Williamson and Duncan for initiating me into the mysteries of Papayrus

Andreas, Sarah Danes, James Williams for proofreading.

Douglas, Norman and Sandy in Photography.

Justin for always finding one more interesting article.

Also the whole lab. for putting up with me while I've been writing.

Finally I've been lucky enough to have some excellent supervisors.

Andy Leigh-Brown for being my University supervisor.

Veronica van Heyningen for encouragement, ideas, keeping my English at bay and generally trying to mother us all.

Nick Hastie for always wanting more data!

Abstract

The *WT1* gene was isolated by positional cloning as a candidate gene implicated in predisposition to the paediatric kidney cancer Wilms' tumour. Because of its early onset and histological resemblance to immature kidneys, Wilms' tumour was thought to arise from an aberration in the normal developmental pathway. From the amino acid sequence, *WT1* was predicted to be a transcription factor with an N-terminal proline/glutamine rich transregulatory domain and four C-terminal *TFIIIA*-like zinc fingers. Two alternative splices have been found in the transcript, resulting in the insertion of 17aa or 3aa (KTS). Functional analysis has shown that *WT1* is capable of sequence specific DNA binding and of regulating transcription. Analysis of *WT*1 mRNA in the developing embryo has shown a pattern of expression consistent with an important role in nephrogenesis, specifically an involvement in mesenchymal-epithelial cell transition. *WT1* analysis in Wilms' tumours revealed that both functional copies are lost in about 10% of sporadic Wilms' tumours. This confirmed the role of *WT1* as a tumour suppressor gene in at least some cases, (following Knudson's two hit hypothesis for tumorigenesis). Predisposition to Wilms' tumour can also be associated with developmental abnormalities, including genitourinary malformations. Expression of *WT1* is seen in the earliest stages of gonad development and constitutionally heterozygous mutations in *WT1* are frequently associated with genitourinary abnormalities. The most severe abnormalities, in Denys-Drash syndrome, have been linked to heterozygous constitutional *WT1* missense mutations in the zinc fingers. The cloning of the mouse *Wt1* gene revealed a very high level of similarity to the human gene, as well as a very similar pattern of expression during development. The major problem with the use of this system as a model is that Wilms' tumours have never been observed in mouse, even in the Sey$^{(Dey)}$ mouse which is hemizygous for *Wt1*.

In order to complement the current lines of investigation it was decided to attempt to isolate a number of *WT1* orthologs from a range of vertebrates to produce wider sequence comparison and eventually a functional comparison. Evolutionary analysis is a particularly appropriate line of investigation for kidney development because across all the vertebrates nephrogenesis seems to occur in a similar manner and kidney tumours which histologically resemble Wilms' tumours have been found in many vertebrates. Also mammalian kidney development passes through three stages. In the lower vertebrates only the first two stages occur so that the human foetal kidney resembles the adult kidney of the anamniotes. In all the vertebrates the close association of the urinary and gonadal systems is maintained, with variations.

Therefore the investigation of *WT1* expression in these different but similar systems may point to important roles for *WT1*. It will also produce more manipulable animal models in which the roles of many other developmental genes are now being investigated.

The conservation of sequences homologous to *WT1* in a range vertebrates was first demonstrated by zoo blot hybridisation with human and mouse *WT1* probes. Partial clones of *WT1* orthologs from rat, pig, chick, alligator, *Xenopus laevis* and zebrafish were then isolated by cross-species PCR, using degenerate primers designed on the basis of the mouse and human sequences. These were then used as probes to isolate larger regions of the gene from cDNA libraries. Clones for marsupial mouse, chick, alligator and zebrafish were obtained. Comparison of these has shown that there is a high level conservation of the predicted protein sequence due to selection pressure. The highest level of conservation is found in the DNA binding zinc fingers, which is consistent with that observed in comparisons of other transcription factors. Investigation of the conservation of the two alternative splice forms of the gene indicated that the KTS alternative splice which is involved in the modulation of DNA binding is very probably conserved in zebrafish, but the 17 amino acid alternative splice was not observed outside the eutherian mammals and was also adjacent to one of the most highly substituted regions of the protein indicating that this could be involved in class specific development. Much of the transregulatory domain was highly conserved, the exceptions being two homopolymer motifs indicating that these may have arisen due to instability of a trinucleotide repeat and may not be functionally important. The mutations so far described in *WT1* have all, bar one, been located at conserved positions.

The expression of *WT1* in chick, alligator and *Xenopus* has been investigated by Northern blotting and whole mount *in situ* hybridisation and the expected pattern during kidney development seen, consistent with a conserved function. Expression was also observed in lateral plate mesoderm and ectodermal derived structures which has also been observed in mammals.

Therefore *WT1* is a highly conserved gene at both the structural and expression level, indicating an evolutionary conserved role in development. This study paves the way for a detailed comparison of genitourinary development and Wilms' like tumours in the vertebrates.

Table of Contents

	Page
Abbreviations	16
Chapter 1 Introduction	
1.1.Introduction	20
1.2. Wilms' tumour and the identification of a putative predisposition gene-*WT1*	21
1.2.1 Incidence of Wilms' tumour	21
1.2.2 Knudson and Strong two hit model for tumorigenesis.	23
1.2.3 Mapping of the WAGR locus to 11p13	24
1.2.4 Isolation of *WT1*	24
1.3 Structure and function of *WT1*	26
1.3.1 Zinc fingers	26
1.3.2 Proline/glutamine rich region	26
1.3.3 Alternative splices	29
1.3.4 Identification of the WT1 protein	30
1.3.5 Transcription factor activity	30
1.3.5.1 DNA binding	30
1.3.5.2 Target promoters	30
1.3.5.3 Mapping of functional domains	31
1.3.5.4 Protein-protein interactions and modulation of *WT1* activity	32
1.3.5.5 Regulation of *WT1* expression	32
1.4 Evidence that *WT1* is a predisposition gene for Wilms' tumour	33
1.4.1 Mapping deletions in Wilms' tumours to the *WT1* locus	33
1.4.2 Intragenic deletions	33
1.5 Kidney development and *WT1*	36
1.5.1 The developing kidney	36
1.5.2 The molecular basis of kidney differentiation	40
1.5.3. Wilms' tumours	43
1.6 WT1 and intermediate mesoderm development	44
1.7 WT1 expression outside the intermediate mesoderm	47
1.8 Involvement of *WT1* in genitourinary abnormalities	48
1.8.1 Mapping of genitourinary abnormalities to the *WT1* region	48
1.8.2 Denys-Drash Syndrome	49
1.8.3 Dominant-negative hypothesis	49
1.9 Involvement in other developmental abnormalities	52
1.10 The *Wt1* knockout mouse	53
1.11 Is *WT1* the only gene responsible for predisposition to Wilms' tumour?	53
1.11.1 Lack of mutations in *WT1*	53
1.11.2 Dominant mutations and trans-splicing effects	54
1.11.3 A possible role for imprinting	54
1.11.4 A second Wilms' tumour locus on 11p	55
1.11.5 Not all Wilms' tumour predisposition loci map to chromosome 11	56

1.11.6 Progression of Wilms' tumours 56
1.12 Evolutionary comparison of the vertebrate genitourinary system 56
1.12.1 The vertebrate kidney 57
1.12.2 Amniote kidney development 57
1.12.3 Anamniote kidney development 60
1.12.4 The metanephros 60
1.12.5 The gonads and genital ducts 60
1.13 Evolutionary comparison of Wilms'-like tumours 62
1.13.1 Mammals 62
1.13.2 Birds 63
1.13.3 Anamniotes 64
1.14 Evolutionary comparison as a method for defining gene function 64
1.14.1 Structural analysis 64
1.14.2 Functional analysis 66
1.14.3 Animal models 67
1.15 Conclusion and aims of the project 67

Chapter 2 Materials and methods

2.1 Bacterial manipulation 70
2.1.1 Genotypes 70
2.1.1.1 Cell strains 70
2.1.1.2 Bacteriophage 70
2.1.2 Electro-transformation of *E.coli* 70
2.1.2.1 Preparation of electro-competent cells 70
2.1.2.2 Electro-transformation 71
2.1.3 Selection of cells harbouring recombinant plasmids 71
2.1.4 Bacterial colony lifts 71
2.1.5 Small scale preparation of plasmid DNA 72
2.1.5.1 Method 1 72
2.1.5.2 Method 2 72
2.1.6 Large scale preparation of plasmid DNA 73
2.1.7 Glycerol stocks of bacteria 74
2.1.8 Isolation of clones from bacteriophage libraries 74
2.1.8.1 Preparation of host cells for bacteriophage 74
2.1.8.2 Titering of libraries 74
2.1.8.3 Screening libraries 75
2.1.8.4 Amplification of phage by the plate lysate method 76
2.1.8.5 *In vivo* excision of λZAP phagemids 76
2.1.8.6 Purification of λgt11 DNA using LambdaSorb™ 77
2.2 DNA/RNA manipulations 77
2.2.1 DNA preparation from tissue 77
2.2.2 RNA preparation from tissue or cell lines 78
2.2.3 Agarose gel electrophoresis 78
2.2.3.1 DNA electrophoresis 78
2.2.3.2 RNA electrophoresis 79

2.2.4 Vacuum transfer of DNA and RNA 79
2.2.5 Hybridisation 80
2.2.6 Purification of DNA 81
2.2.6.1 Phenol:chloroform extraction 81
2.2.6.2 GENECLEAN® II (Bio 101) 81
2.2.6.3 Purification of a specific DNA fragment 81
2.2.6.3.1. GENECLEAN®II (Bio 101) 81
2.2.6.3.2. Spin-X 82
2.2.6.4 "Double GENECLEAN®" protocol for cloning PCR products 82
2.2.7 Quantification of DNA fragments 82
2.2.8 Enzyme modification of DNA and RNA 82
2.2.8.1 Restriction digest of plasmid DNA 82
2.2.8.2 Restriction digests of genomic DNA 83
2.2.8.3 Production of radiolabelled DNA probes 83
2.2.8.3.1 Random primed labelling 83
2.2.8.3.2 Specific primed labelling 84
2.2.8.4 Preparation of riboprobes 84
2.2.8.5 cDNA synthesis 85
2.2.8.6 Modifying ragged ends of PCR products 86
2.2.8.7 Ligation of DNA fragments into vectors 86
2.2.8.7.1 Cohesive end subcloning 86
2.2.8.7.2 Blunt ended ligation of PCR products 87
2.2.8.7.3 Turbo cloning of PCR products 87
2.3 Sequencing DNA 87
2.3.1 Double stranded sequencing of plasmid DNA 88
2.3.2 Direct sequencing of PCR products . 88
2.3.3 Polyacrylamide gel electrophoresis (PAGE) of sequencing reactions 89
2.4 Detection of radioactive signal 90
2.4.1 Autoradiography 90
2.4.2 PhosphorImager analysis 90
2.5 Incubation and analysis of embryos 90
2.5.1 Rearing of *Xenopus* embryos 90
2.5.2 Whole mount *in situ* hybridisation 91
2.5.2.1 Incubation and dissection of embryos 91
2.5.2.2 Preparation of embryos 91
2.5.2.3 Hybridisation 92
2.5.2.4 Antibody conjugate binding and staining 92
2.5.2.5 Photography of embryos 93
2.6 Polymerase chain reaction (PCR) 93
2.6.1 Synthesis and purification of oligonucleotides 93
2.6.2 Reaction conditions 93
2.6.3 PCR template 94
2.7 Solutions 94
2.8 Computer analysis of DNA sequences 98

Chapter 3 Identification of sequences cross hybridising to *WT1* in the vertebrates

3.1 Introduction 100
3.2 Identification of sequences cross hybridising to *WT1* in mammals 100
3.3 Detection of sequences cross hybridising to *WT1* in non mammalian vertebrates 102
3.4 Conclusions 102

Chapter 4 Isolation of *WT1* orthologs

4.1 Isolation by cross species PCR 106
4.1.1 Introduction 106
4.1.2 Experimental design 108
4.1.2.1 Template 108
4.1.2.2 Primer design 108
4.1.2.3 Reaction conditions 112
4.1.3 Testing PCR primers 112
4.1.4 Isolation of rat, pig and alligator *WT1* orthologs 112
4.1.4.1 Cross species PCR 112
4.1.4.2 Cloning PCR products 114
4.1.5 Redesign of PCR primers 121
4.1.6 Isolation of chick, *Xenopus* and axolotl *WT1* orthologs 121
4.1.7 Isolation of the zebrafish *WT1* ortholog 121
4.1.7.1 Degenerate PCR 121
4.1.7.2 Turbo cloning 123
4.2 Isolation by library screening 126
4.2.1 Introduction 126
4.2.2 Library screens 128
4.2.3 *WT1* cDNA clones isolated 128
4.2.4 Analysis of zinc finger 4 in chick 132
4.2.5 Isolation of zebrafish *WT1* genomic clone (Z1a) 136
4.3 Discussion 136

Chapter 5 Structural and evolutionary comparison of *WT1* orthologs

5.1 Introduction 141
5.2 Structural analysis 141
5.2.1 Sequence motifs 141
5.2.1.1 Structure prediction 141
5.2.1.2 Post-translational modification 144
5.2.1.3 Exon structure 144
5.2.2 5' Non coding region 144
5.2.3 The predicted initiation codon is conserved 146
5.2.4 Transregulatory domain 146

5.2.4.1 Homopolymer domains 146
5.2.4.2 The amino acid composition is similar to other transregulatory domains 149
5.2.4.3 Mutations 151
5.2.5 The conservation of the putative leucine zipper in exon 4 151
5.2.6 Exon 5 - the 17aa alternative splice is not conserved in the reptilian lineage 154
5.2.7 The divergence and conservation of exon 6 154
5.2.8 The zinc finger region is the most highly conserved region of *WT1* 156
5.2.9 The KTS alternative splice has been conserved through vertebrate evolution 161
5.2.10 The C-terminal region is not changed in mouse 161
5.2.11 3' Non coding region 161
5.3 Evolutionary analysis 164
5.3.1 Rate of sequence evolution of *WT1* 164
5.3.1.1 Introduction 164
5.3.1.2 Analysis of *WT1* 165
5.3.2 Phylogenetic analysis 170

Chapter 6 Investigation of the expression of *WT1*

6.1 Introduction 174
6.2 Northern analysis of *WT1* mRNA 174
6.3 *In situ* hybridisation 177
6.3.1 Comparison of the developmental stages of amniotes 177
6.3.2 *WT1* expression in chick embryos 179
6.3.3 *WT1* expression in alligator embryos 183
6.3.4 *Xenopus* whole-mount hybridisation 193
6.4 Comparison of WT1 expression 193

Chapter 7 Discussion

7.1 Evolution of transcription factors in vertebrates 196
7.2 Rate of evolution of *WT1* 198
7.3 The transregulatory domain 198
7.3.1 Evolution of homopolymer domains 200
7.3.2 Function of homopolymers 202
7.3.3 Conservation of the putative leucine zipper 203
7.4 A region of *WT1* with species specific function? 203
7.5 Selection of alternative splice sites 204
7.6 Evolution of the zinc fingers 206
7.7 Conservation of expression of the *WT1* gene in the intermediate mesoderm 208
7.7.1 Urinary system 209
7.7.2 Gonads 210
7.8 Additional sites of expression 212
7.9 *WT1* and limb development 212
7.10 Evolution of the vertebrate kidney and *WT1* 214

7.11 Origin of *WT1* 217
7.12 Future work 217

Appendix A Sequences of *WT1* orthologs 220

Appendix B Structure prediction plots for *WT1* orthologs 253

Bibliography 261

Figures

Figure 1.1 Schematic representation of the human *WT1* mRNA transcript 25
Figure 1.2 Secondary and tertiary structure of zinc fingers 27
Figure 1.3 Differentiation of the mammalian kidney 37
Figure 1.4 Differentiation of the metanephros 39
Figure 1.5 Nephron development 41
Figure 1.6 Secondary sexual determination 45
Figure 1.7 Phylogenetic tree of the vertebrates 58
Figure 1.8 The relationship between kidney development and evolution 59
Figure 1.9 The interrelationship between the gonad and urinary systems 61

Figure 3.1 Detection of sequences that hybridise to *WT1* in a range of mammalian genomes 101
Figure 3.2 Detection of sequences that hybridise to *WT1* in a range of vertebrate genomes 104

Figure 4.1 Predicted amino acid sequence of human *WT1* showing the exon structure and positions of degenerate PCR primers used in this study 109
Figure 4.2 Testing of degenerate primers 113
Figure 4.3 Isolation of *WT1* orthologs 115
Figure 4.4 Blunt ended cloning strategy used for cloning PCR products 116
Figure 4.5 Schematic representation of the human *WT1* cDNA and the *WT1* clones isolated by degenerate PCR 120
Figure 4.6 Isolation of chick and *Xenopus WT1* orthologs 122
Figure 4.7 Isolation of zebrafish *WT1* 124
Figure 4.8 Turbo cloning of PCR products 125
Figure 4.9 Comparison of human and zebrafish *WT1* genomic structures in the zinc finger region 127
Figure 4.10 Diagram of *WT1* clones isolated by library screening, in relation to the human *WT1* gene 130
Figure 4.11 Isolation of ZF4 from chick 133
Figure 4.12 Genomic structure of zebrafish *WT1* 137

Figure 5.1 Comparison of the predicted amino acid sequences of the *WT1* orthologs 142
Figure 5.2 Multiple sequence alignment the of non coding region of exon 1 using CLUSTAL V 147
Figure 5.3 Investigation into the conservation of homopolymer domains in exon 1 150
Figure 5.4 Helical representation of the putative leucine zipper 152
Figure 5.5 Investigation into the conservation of the 17amino acid alternative splice -exon 5 155
Figure 5.6 Mutation, editing and conservation of exon 6 157
Figure 5.7 Schematic representation of the four zinc fingers of *WT1* 159
Figure 5.8 Sequence of the mouse *Wt1* cDNA clone KS/2 162

Figure 5.9 Graphs plotting K_S against divergence time 169
Figure 5.10 Unrooted phylogenetic tree for exons 7-9 of *WT1* 171
Figure 5.11 Graph of divergence of *WT1* against divergence time 172

Figure 6.1 Detection of *WT1* mRNA in pig, chick, and alligator 175
Figure 6.2 *Xenopus WT1* mRNA expression 176
Figure 6.3 Expression of *WT1* in the developing chick 180
Figure 6.4 Whole mount *in situ* hybridisation of alligator embryos-onset *WT1* expression 184
Figure 6.5 *WT1* expression in stage 9-11 alligator embryos 186
Figure 6.6 *WT1* expression in stage 13-14 alligator embryos 188
Figure 6.7 *In situ* hybridisation of sections of the alligator mesonephros 196

Tables

Table 1.1 Congenital abnormalities associated with Wilms' tumour 22
Table 1.2 *WT1* mutations in WT with no associated congenital abnormalities 35
Table 1.3 *WT1* mutations in WT with associated congenital abnormalities 50
Table 1.4 *WT1* mutations associated with Denys-Drash syndrome 51

Table 3 Sizes of vertebrate genomes 103

Table 4.1 Sources of RNA from which isolation of *WT1* orthologs was attempted 107
Table 4.2 Sequence of degenerate PCR primers for cross species PCR 111
Table 4.3 Similarity of PCR clones isolated to human and mouse *WT1* at the nucleotide and amino acid level 118
Table 4.4 PCR primers designed using sequence from PCR clones 119
Table 4.5 Summary of libraries screened for *WT1* and the clones isolated 129
Table 4.6 Similarity of clones isolated from libraries to human and mouse *WT1* at the nucleotide and amino acid level 131

Table 5.1 Conservation of splice sites in *WT1* 145
Table 5.2 Percentage nucleotide identity between *WT1* orthologs in the non coding and coding regions of exon 1 148
Table 5.3 Comparison of the Kozak sequence with the putative *WT1* initiation codon 148
Table 5.4 Percentage amino acid composition of the transregulatory domain 148
Table 5.5 Percentage similarities between coding and non coding regions of exon10 163
Table 5.6 Figures obtained for K_S and K_A using Li's program 166
Table 5.7 Comparisons of the K_S and K_A values for *WT1* with other genes 167

Table 6 Development of the chick showing comparisons with human, mouse and alligator 178

Table 7.1 Comparison of the conservation of transcription factors and their DNA binding domains 197
Table 7.2 Genes containing homopolymer domains 201

Abbreviations

A	adenosine
aa	amino acid
ALL	acute lymphoid leukemia
AML	acute myeloid leukemia
AMV	avian myeloblastosis virus
BCIP	5-bromo-4-chloro-3-indolyl-phosphate
bp	base pair
BSA	bovine serum albumin
BWS	Beckwith Wiedemann Syndrome
C	cytidine
cDNA	complementary DNA
CML	chronic myeloid leukemia
d	days
DDS	Denys Drash syndrome
DEPC	diethylpyrocarbonate
dH_2O	distilled water
DIG	digoxigenin
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dNTP	2'-deoxyribosylnucleoside 5'-triphosphoshates
dpc	days post coitum
DTE	dithioerythritol
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetate
EMBL	european molecular biology laboratory
G	guanosine
GCG	Genetics Computer Group
$(Glycine)_5$	GGGGG
GTF	general transcription factor
GU	genitourinary
HMG	high mobility group
HPhillic	hydrophilic
HPhobic	hydrophobic
I	inosine

ILNR	intralobar nephrogenic rest
IPTG	isopropylthiogalactose
kb	kilobase
K_A	number of nonsynonymous changes per nonsynonymous site
K_S	number of synonymous changes per synonymous site
KTS	lysine threonine serine
L Broth	Luria broth
LOH	loss of heterozygosity
MOPS	morpholinopropansulphonic acid
mRNA	messenger RNA
Myr	millions of years
Myra	millions of years ago
NBT	nitroblue tetrazolium
N/A	not applicable
N/D	not done
nt	nucleotide
p	short arm
PFA	paraformaldehyde
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
PLNR	perilobar nephrogenic rests
$(Proline)_4$	PPPP
$(Proline)_{13}$	PAPPPAPPPPPPPPP
OD	optical density
q	long arm
RACE	rapid amplification of PCR ends
RNA	ribonucleic acid
rpm	revolutions per minute
rATP	5'-(pyro) adenosine triphoshate
RT	room temperature
RT-PCR	reverse transcriptase PCR
SDS	sodium dodecyl sulphate
SRO	smallest region of overlap
SSC	standard sodium citrate
St	stage

T	thymidine
TCA	trichloroacetic acid
TEMED	N,N,N'N'-tetramethylethylenediamine
Tm	melting temperature
Tris	2-amino-2-(hydroxymethyl)-1,3-propandiol
tRNA	transfer RNA
U	units of enzyme as defined by the manufacturer
UTP	uridine 5'-triphosphate
UV	ultraviolet
w	weeks
WAGR	Wilms' tumour aniridia genitourinary abnormalities mental retardation
WT	Wilms' tumour
X-gal	5-bromo-4-chloro-3-indolyl-galactose
X-phosphate	5-bromo-4-chloro-3-indolyl-phosphate
ZF	zinc finger

Chapter 1
Introduction

1.1 Introduction

One of the major achievements of molecular genetics so far has been to increase the understanding of the basis of human genetic disease. In recent years there has been a flood of information pinpointing genes in which lesions can lead to a particular disease. In many cases the reason why a genetic lesion is giving rise to a disease is easily understood. The classic examples being haemophilias A and B, which result from Factor VIII and Factor IX mutations, and the thalassemias which are due to mutations of the globin genes. In the vast majority of cases the gene was isolated using some information about the gene product - functional cloning. In the last few years genes have been isolated by virtue of their physical or genetic linkage to a disease locus i.e. by reverse genetics, otherwise known as positional cloning. These include the *CFTR* gene in cystic fibrosis, *NF1* in neurofibromatosis type 1, *FMR* in Fragile X syndrome and *RB1* in retinoblastoma. It is then necessary to discover the function of the gene product and understand how malfunction can result in the particular disease phenotype.

Cancers and developmental abnormalities are two of the most complex disease types, the causes of which are just beginning to be unravelled through molecular genetics. Not infrequently these two superficially different classes of disease turn out to be intimately related. Cancers arise from a loosening of the strict growth controls on cells and development proceeds by the regulation of cell proliferation and differentiation. Wilms' tumour is an example of both a kidney developmental abnormality and a cancer.

A predisposition gene for Wilms' tumour, *WT1*, has been isolated by positional cloning on the basis of its genetic linkage to the disease. Since its isolation an important role for *WT1* in kidney and gonad development, as well as in tumorigenesis, has been demonstrated. That it is not the sole gene involved in formation of Wilms' tumour is also clear. *WT1* has been shown to have activities consistent with a role as a regulatory transcription factor but as yet its interactions *in vitro* and *in vivo* are not clearly understood. This particularly applies to the part it is playing in a mesenchyme to epithelial transition, and how *WT1* mutants give rise to genitourinary abnormalities. Investigations have so far been hampered by the lack of a relevant model system.

An evolutionary comparison of *WT1* will have many uses. A wider sequence comparison of the *WT1* gene will facilitate the targeting and interpretation of structural, functional and mutational analysis. The comparison between the mouse and human sequences reveals very little due to the almost complete similarity between

the two genes. At the same time the isolation of *WT1* orthologs will lay the foundations for a series of animal models in which *WT1* can be investigated and manipulated. In the case of *WT1* this is a particularly appropriate line of investigation because vertebrate kidney development is based on an initial plan which is then modified in the different classes. This seems to have occurred through modifications at the later stages of development so that the lower vertebrate kidneys have resemblances to foetal mammalian kidneys. By looking at lower vertebrate species it may be possible to elucidate the role of *WT1* in early human kidney development. The investigation of *WT1* may also increase the understanding of the evolution of the vertebrate kidney.

1.2 Wilms' Tumour and the identification of a putative predisposition gene-*WT1*

1.2.1 Incidence of Wilms' tumour

Wilms' tumour or nephroblastoma is an embryonic cancer of early childhood with 90% of cases presenting in the first 7 years of life, affecting 1 in 10,000 children (Young and Miller 1975). The early age of onset and the often undifferentiated histopathology of the tumour has suggested that Wilms' tumour can be a result of an error in the kidney developmental process, and it is thought to arise from the nephrogenic cells of the kidney (Nicholson 1931, Mierau *et al,* 1987). A comprehensive study of the incidence of Wilms' tumours has been produced by the US National Wilms' tumour Survey (Beckwith 1983, Breslow *et al.,* 1988, Olson *et al* 1993). Although the tumour is observed world-wide, the incidence varies between ethnic groups: it is higher in the American Black population than Caucasian (Kramer 1984) and lower in Asian than European populations (Tadokoro *et al,* 1992). Most cases are sporadic, but 1% of cases are familial with tumour predisposition inherited as an autosomal dominant trait with variable penetrance (Matsunga 1981). In addition incidence of associated congenital abnormalities in WT patients is greater than in the general population. Up to 15% of Wilms' tumour patients have associated congenital abnormalities. In particular, 1% have aniridia, 4% have genitourinary abnormalities, 1.5% have cardiopulmonary abnormalities and 3% have hemihypertrophy (Breslow and Beckwith 1982, Olson *et al.,* 1993). In males who have Wilms' tumour and aniridia the frequency of the genitourinary abnormalities rises to 40-50% (Shannon *et al.,* 1982, Turleau *et al.,* 1984). The abnormalities can be grouped into several syndromes; WAGR, DDS, Perlman and BWS (Table 1.1). In the case of WAGR patients large cytogenetic deletions have been observed.

Table 1.1 Congenital abnormality syndromes associated with Wilms' tumour

Syndrome	Abbreviation	Phenotype	Reference
WAGR	WAGR	Wilms' tumour, Aniridia, *Genitourinary Abnormalities, Mental Retardation	Miller *et al.*, 1964
Denys-Drash	DDS	Wilms' tumour, Gomerular Nephropathy, **Genitourinary Abnormalities	Denys *et al.*, 1967 Drash *et al.*, 1970
Beckwith Wiedemann	BWS	Wilms' Tumour, Adrenocortical Carcinoma, Hepatoblastoma, Foetal Gigantism, Exomphalos, Macroglossia, Organomegaly, Hemihypertrophy, Gonadal Abnormalities	Beckwith 1969 Wiedemann 1964
Perlman		Wilms' Tumour, Foetal Gigantism, Bilateral Nephromegaly, Nephroblastomatosis, Cryptorchidism	Perlman *et al.*, 1973

*Genitourinary Abnormalities — Cryptorchidism, Hypospadias, Fused Kidneys, Double Collecting System

**Genitourinary Abnormalities — A spectrum of abnormalities, from the abnormalities above all the way to Pseudohermaphroditism

1.2.2 Knudson and Strong two hit model for tumorigenesis.

Knudson (1971) originally proposed his two hit model of tumorigenesis for retinoblastoma (RB), a childhood eye tumour. Using statistical analysis he showed that the incidence of unilateral (in one eye), bilateral (in both eyes), familial and sporadic cases fitted a minimum of two mutations or 'hits' occurring to initiate tumorigenesis. He proposed that for familial cases the first mutation is present in the germline and so only one somatic mutation had to occur. In non germline cases both mutations had to be somatic and so the later onset than in familial cases was explained by the longer time for two mutations to occur. The onset of bilateral cases tended to be earlier, in the same range as familial tumours, also indicating the presence of a germline predisposition mutation. The tumour is limited to childhood because the target cell population are embryonic retinal precursor cells which disappear with time after birth. That the mutations resulted in the loss of a tumour suppressor was suggested by the association of retinoblastoma with deletions at chromosome 13q14 as well as loss of heterozygosity for this region in the tumours, the latter being consistent with a second hit (which would effectively expose the first hit in a recessive gene) (Cavenee *et al.*, 1983).

That the malignant phenotype can be due to the loss of gene function as opposed to dominantly acting oncogenes had been suggested by experiments which had demonstrated that a wild type cell could suppress the tumorigenic effect of malignant cells on cell fusion (Harris *et al.*, 1969, Stanbridge 1976). The tumour suppressor gene involved in retinoblastoma, *RB1*, has been cloned (Friend *et al.*, 1986, Lee *et al.*, 1987) and so far all cases have been found to be consistent with Knudson's model (Goodrich and Lee 1990).

Comparison of the incidence of Wilms' tumour with retinoblastoma showed many similarities, an earlier onset of the presumed germline and bilateral cases, and in the few familial cases autosomal dominant inheritance. On the basis of this it was proposed that Wilms' tumorigenesis could also arise from two 'hits' (Knudson and Strong 1972). The major difference between the two was that although the sporadic bilateral (i.e. germline) cases of both were frequent the number of familial cases was much lower in WT. These *WT* familial cases did not segregate with chromosome 11p to which the sporadic cases had been mapped. This indicated that although there were similarities in the molecular basis of the two childhood tumour types, WT maybe more complicated, with more than locus involved.

1.2.3 Mapping of the WAGR locus to 11p13

The first information concerning the location of a predisposition gene came from the karyotypic analysis of rare cases where Wilms' tumour was associated with congenital abnormalities and visible cytogenetic deletions. An interstitial deletion in the short arm of chromosome 11 was found in several cases of WAGR syndrome (Franke *et al.*, 1977). The localisation was refined in two cases where the only deleted region was 11p13 (Franke *et al.*, 1979, Riccardi *et al.*, 1978). This would be consistent with the first hit of Knudson theory. That 11p may also be involved in second hit was shown by loss of heterozygosity for this region in tumours compared to somatic tissue (Fearon *et al.*, 1984, Orkin *et al.*, 1984, Reeve *et al.*, 1984). Loss of heterozygosity on 11p was also found to occur in other types of tumours indicating a more general tumour suppressing activity (Koufos *et al.*, 1985). That there was a tumour suppressing activity present on chromosome 11 was demonstrated experimentally by suppressing tumorigenicity of a putative Wilms' tumour cell line in nude mice by introducing chromosome 11 (Weissman *et al.*, 1987).

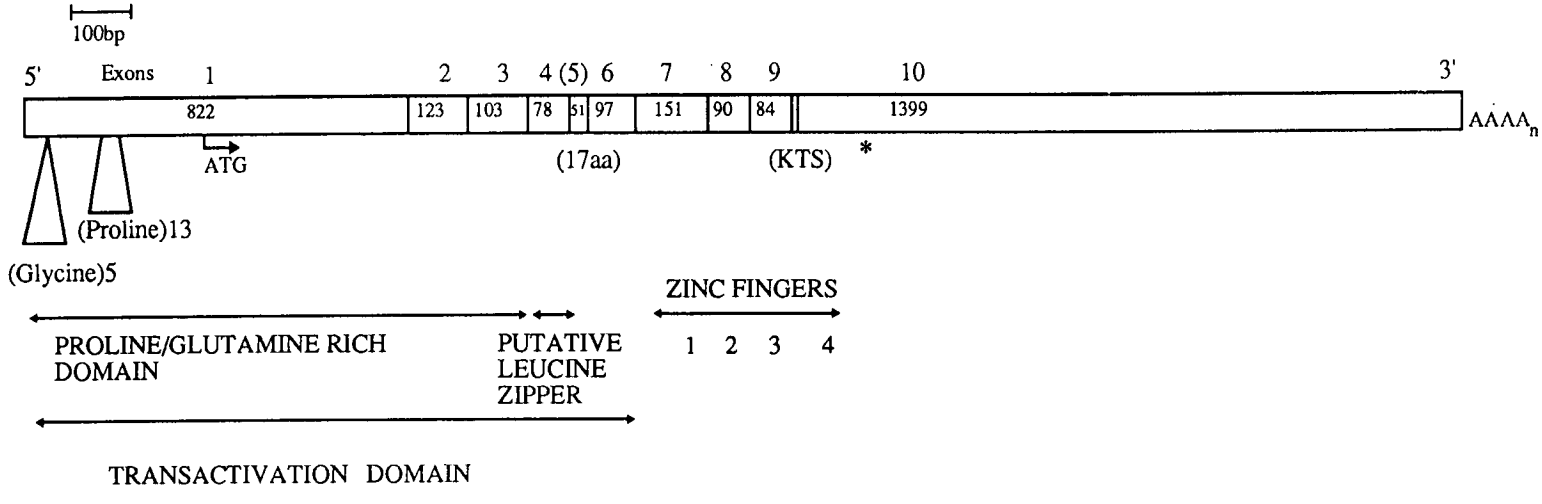
1.2.4 Isolation of *WT1*

More WAGR cases containing constitutional deletions of 11p13 were used to define the WAGR region to between the catalase and follicle stimulating hormone β (FSHβ) genes (van Heyningen *et al.*, 1985, Glaser *et al.*, 1986, Porteous *et al.*, 1987, Bickmore *et al.*, 1989). To define the exact extent of the deletions detailed physical and genetic maps were built up using newly isolated markers for this region (Porteous *et al.*, 1987, Gessler *et al.*, 1989, Compton *et al.*, 1990). Rare cutting enzymes and pulsed-field gel electrophoresis were used to identify CpG islands within the region (Bickmore *et al.*, 1989) as they are often located at the 5' end of genes (Bird 1986). A very detailed physical map of the WAGR region was produced using irradiation reduced hybrids in which the only human component was a small part of chromosome 11p (Rose *et al.*, 1990).

Lewis *et al.*, (1988) were the first to demonstrate the homozygous deletion of an 11p13 marker in a sporadic Wilms' tumour (WiT-3) in a patient with no associated abnormalities. The smallest region of overlap (SRO) between the deletions in WiT-3 was defined, allowing the Wilms' tumour predisposition region to be limited to 345kb (Rose *et al.*, 1990). Two single copy probes isolated from chromosome 11 cosmids were found to map to this region and were used to isolate a partial cDNA from a pre-B cell library that contained an open reading frame (WT33). This was found to hybridise to an RNA transcript in mouse and baboon kidney (Call *et al.*, 1990). This

Figure 1.1 Schematic representation of the human *WT1* mRNA transcript

The exon structure is shown including the two alternative splices (17aa-exon5 and KTS) (Haber *et al.*, 1991, Gessler *et al.*, 1992). The major features of the encoded protein, the four zinc fingers, the putative leucine zipper, and the proline/glutamine rich transcription regulatory domain including the glycine and proline rich domains within it, are shown.



same gene, called *WT1*, was also isolated by chromosome jumping between CpG islands (Gessler *et al.*, 1990).

1.3 Structure and function of *WT1*

The cDNA isolated by Gessler *et al.*, (1990) was 3kb long, corresponding to the size of transcript seen in Northern analysis, and contained a 1,725bp open reading frame (Figure 1.1). Subsequent analysis of the genomic sequence has shown that the transcript is made up of 10 exons, covering 50kb (Haber *et al.*, 1991, Gessler *et al.*, 1992).

1.3.1 Zinc fingers

The sequence of the C-terminal end of the predicted protein was found to correspond to four zinc finger motifs of the Cys_2 His_2 type, first identified in *Xenopus TFIIIA* (Miller *et al.*, 1985). Zinc fingers contain conserved cysteine and histidine residues which chelate a zinc ion holding together the tertiary structure of the finger (Miller *et al.*, 1985, Pavletich and Pabo 1991, Figure 1.2a). Over 1000 finger motifs have been identified in more than two hundred proteins, mostly transcription factors e.g. *GLI* family (Ruppert *et al.*, 1988), and *Sp1* (Kadonaga *et al.*, 1987). These motifs have been shown many times to have sequence specific DNA binding activities (reviewed in El-Baradi and Pieler 1991) and the orientation of three finger regions, *Zif 268* (*EGR1*) (Pavletich and Pabo 1991), *GLI1* (Pavletich and Pabo 1993) and *tamtrack* (Fairall *et al.*, 1993) when bound to DNA has been elucidated by X-ray crystallography of DNA-protein co-crystals (Figure 1.2b). The *WT1* fingers 2-4 most closely resemble the three fingers in *EGR2* (*Krox 20*) (Joseph *et al.*, 1988) with an identity at the amino acid level of 67% and a similarity of 75% (Figure 1.2c). No sequence similarity is seen outside the zinc fingers except for a similar richness in certain amino acids. *EGR2* belongs to the *EGR* family which contain virtually identical zinc fingers but with only a few regions of homology in the rest of the gene (Crosby *et al.*, 1992). They have a conserved function in the early growth response to serum stimulation (Sukhatme 1992) and development (Nieto *et al.*, 1991).

1.3.2 Proline/glutamine rich region

Upstream of the zinc fingers the *WT1* coding sequence is particularly rich in proline and glutamine residues including a run of fifteen amino acids, thirteen of which

Figure 1.2 Secondary and tertiary structure of zinc fingers

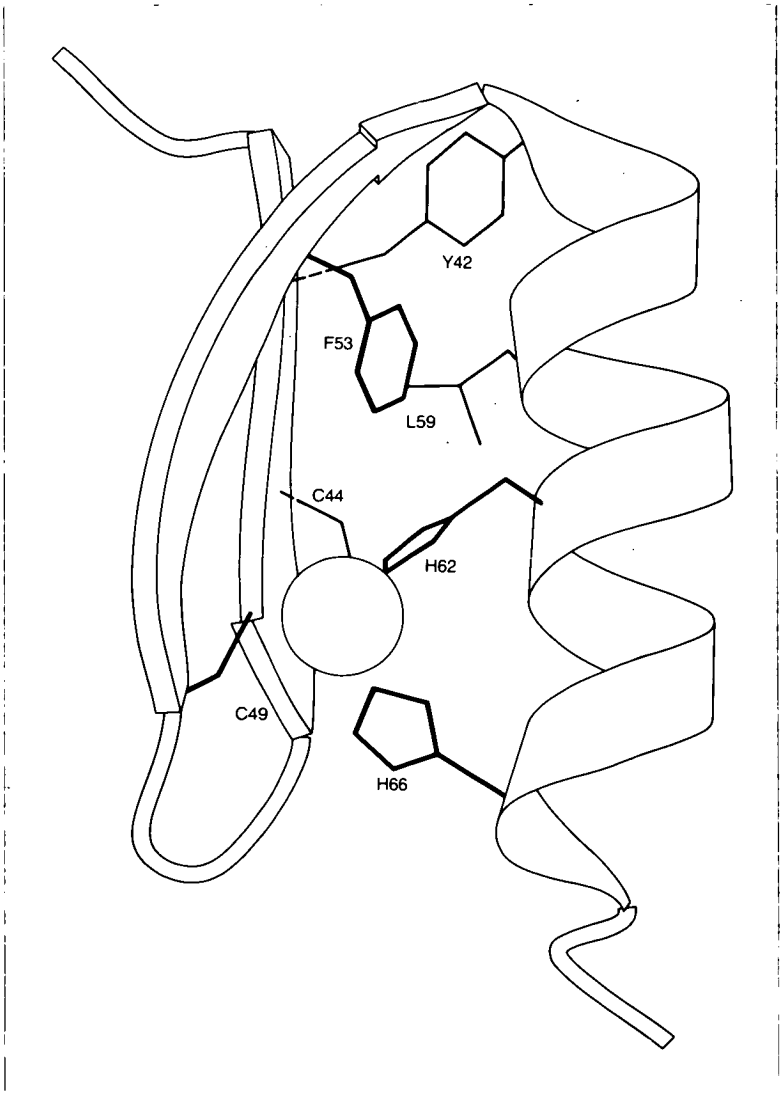
a NMR determined structure for the second zinc finger of *SWI5*. This shows the antiparallel β strands followed by the α helix which are held in the 'finger' conformation by the chelation of a $zinc^{2+}$ ion by two cysteine and two histidine residues and the hydrophobic core (Neuhaus *et al.*, 1990).

b The orientation of the three zinc fingers of *EGR1* in relationship to a DNA binding site as determined by the X-ray analysis of the protein-DNA co-crystal (Pavletich and Pabo 1991). The finger contacts the DNA in the major groove with the 'tip' of the finger closest to DNA, i.e. the C-terminal end of the β strand and N-terminal end of the α helix. The fingers are also shown to be wrapping around the DNA. (Illustrations taken from Travers 1993)

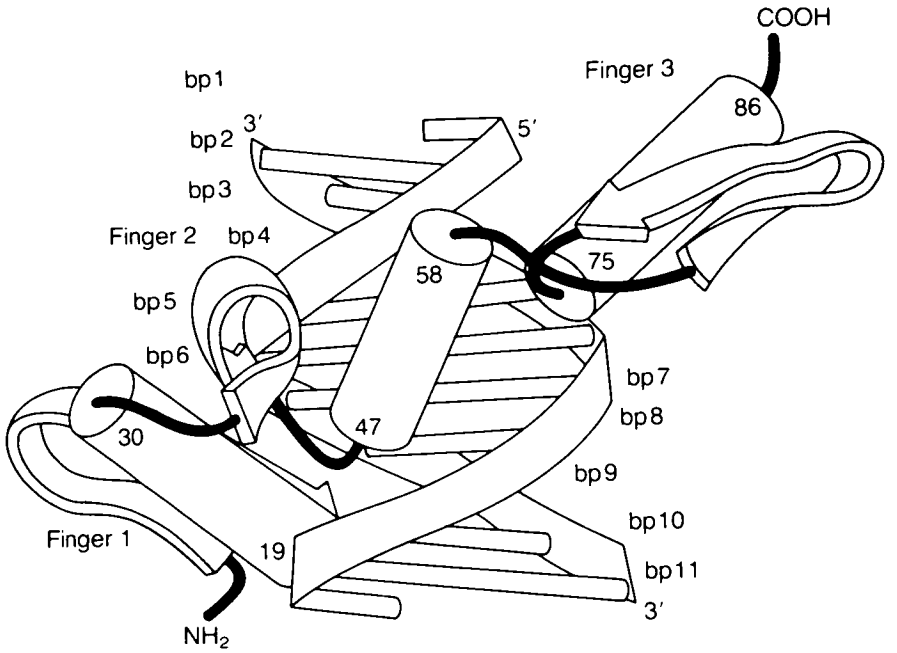
c Comparison of the amino acid sequence of the zinc fingers of *EGR1*&2 with fingers 2-4 of *WT1*. The conserved zinc chelating residues are shown in bold and the structural regions of the fingers indicated, including the regions involved in DNA binding. The residues directly involved in base recognition are underlined.

d Pattern of contacts of the three zinc fingers of *EGR1* (Zif 268) with their triplet base recognition sequence. The pattern is identical for fingers 1&3 but is different in ZF2 due to the substitution of an arginine for a threonine residue.
The positions within the zinc fingers involved in base contacts are circled and the α helix is shaded. *Indicates an aspartate residue which interacts with the first arginine to stabilise its interaction with the DNA. (Illustration from Fairall *et al.*, 1993)

a



b

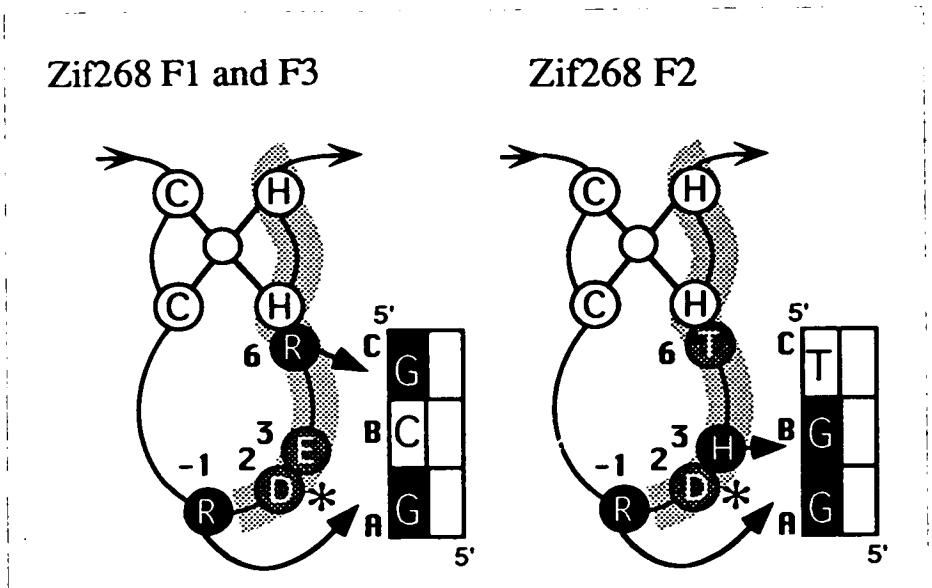


c

Zinc Finger Consensus XF/YXC $X_{2/4}$ CXXXFXXXXXLXXH$X_{3/4}$HTGEK

```
WT1 ZF2        PYQCDFKDCERRFSRSDQLKRHQRRHTGVK
               ||.|. .:|:||||||||:|.|| | ||| |
EGR2 ZF1       PYPCPAEGCDRRFSRSDELTRHIRIHTGHK
               ||:||.|:|||||||||||||||||||:|
Egr1 ZF1       PYACPVESCDRRFSRSDELTRHIRIHTGQK

WT1 ZF3        PFQC..KTCQRKFSRSDHLKTHTRTHTGKTSEK
               ||||  :.| |.|||||||.|||||||   ||
EGR2 ZF2       PFQC..RICMRNFSRSDHLTTHIRTHTG...EK
               ||||  |||||||||||||||||||||   ||
Egr1 ZF2       PFQC..RICMRNFSRSDHLTTHIRTHTG...EK
                ←    β sheet   →  ← α helix →←linker→
WT1 ZF4        PFSCRWPSCQKKFARSDELVRHHNMH
               ||.| :  |.:|||||||  || .:|
EGR2 ZF3       PFACDY..CGRKFARSDERKRHTKIH
               |||||.  ||||||||||||||||||
Egr1 ZF3       PFACDI..CGRKFARSDERKRHTKIH
                              ← DNA binding →
```

d



are proline. Although there is no actual sequence homology between them, regions rich in proline, glutamine or acidic amino acids have been found in an increasing number of transcription factors, *OTF2* (proline and glutamine rich), *Sp1* (glutamine rich), *EGR* family (proline rich) and *CTF/NF1* (proline rich) (Mitchell and Tjian 1989). Transcription regulatory activity has been localised to this region in *CTF/NF1* (Mermond *et al.*, 1988). The mechanism by which these domains act to alter transcription is just beginning to be elucidated. In higher eukaryotes transcription is carried out by RNA polymerase II in association with many general transcription factors (GTF) e.g. *TFIIA-I* (Zawel and Reinberg, 1993). Recently interactions between these GTFs and regulatory transcription factors have been described e.g. between *c-rel* and *TFIIτ*–the TATA box binding protein (Kerr *et al.*, 1993).

1.3.3 Alternative splices

WT1 is subject to alternative splicing of the RNA transcript at two positions (Figure 1.1). One is an insertion of 3 amino acids, KTS (lysine, threonine, serine) in the linker region between zinc fingers 3 and 4 (Haber *et al.*, 1990) and the other occurs between the proline/glutamine rich region and the fingers inserting an extra 17 amino acids (Haber *et al.*, 1991). From characterisation of the genomic structure of *WT1* it is known that the 3 amino acid insertion is achieved through the use of alternative splice donor sites in exon 9, and the 51 nucleotides which specify the 17 amino acid splice are contained within a separate exon, exon 5 (Haber *et al.*, 1991, Gessler *et al.*, 1992). The *in vivo* function of these splices is not known. The 17 amino acid insertion is at the C-terminal end of a putative leucine zipper (Landshultz *et al.*, 1988). Since zippers are dimerisation motifs (Kouzandes *et al.*, 1988) insertion at its end might change dimerisation characteristics of the protein.

The KTS insertion is in a finger linker region, which is a highly conserved sequence of amino acids in the *TFIIIA* type fingers. Insertions and changes in this sequence have been shown to change the DNA binding characteristics of *TFIIIA* (Choo and Klug 1993). The mRNA transcripts containing the KTS insertion have been observed to be the most plentiful RNA transcript (Brenner *et al.*, 1992).

In contrast to many genes a tissue specific pattern of expression of the four possible forms of the RNA has not been observed. Instead in the tissues examined a constant ratio is present (Haber *et al.*, 1991). On cloning the mouse homolog of *WT1* (Buckler *et al.*, 1991), which is 95% conserved at the amino acid level, both

alternative splices were also seen and similar ratios were observed in a wide range of tissues (Haber *et al.*, 1991).

1.3.4 Identification of the *WT1* protein

Using *WT1* specific antibodies, *WT1* has been shown to be a 49-54kDa protein, this correlates well with the size predicted from the primary sequence of 47-48.9kDa (Morris *et al.*, 1991, Telerman *et al.*, 1992). The range in size is thought to arise from alternative splicing or to phosphorylation. The latter could not be detected when transiently transfected COS-1 cells were ^{32}P labelled although there are potential phosphorylation sites (Morris *et al.*, 1991). Subcellular localisation to the nucleus consistent with *WT1* being a transcription factor has been shown by cellular fractionation, immunohistochemistry (Morris *et al.*, 1991, Telerman *et al.*, 1992) and epitope tagging (Pelletier *et al.*, 1991c).

1.3.5 Transcription factor activity

1.3.5.1 DNA binding

Numerous studies have tried to define *in vitro,* in various ways, the consensus DNA binding sites for the *WT1* zinc fingers. The -KTS but not the +KTS isoform was found to bind sequences extremely similar to and including the *EGR1* consensus (GCGGGGGCG) but not to other GC rich sequences, such as those bound by *Sp1* and *myb* (Rauscher *et al.*, 1990). It is not surprising that *WT1* is capable of binding the *EGR1* consensus considering the similarity of fingers 2-4 to those of *EGR1*, especially in those regions known to interact with the DNA (Figure 1.2c&d). The +KTS isoform has also been shown to bind DNA in a sequence specific manner including some of the -KTS binding sites (Bickmore *et al.*, 1992, Wang *et al.,* 1993b). As yet no true consensus sequence has been identified for either isoform.

1.3.5.2 Target promoters

WT1 has been shown in transfection systems to bind and down-regulate transcription from genes whose protein products are known to have mitogenic properties, *EGR1* (Madden *et al.*, 1991), *IGF2* (Drummond *et al.*, 1992), *PDGFA* chain (Gashler *et al.,* 1992, Wang *et al.*, 1992) and *IGF1R* (Werner *et al.*, 1993).

These genes could well have a physiological role in kidney development. *EGR1* is expressed in response to serum stimulation of cells, *WT1* may antagonise this mitogenic response to growth factors by repressing *EGR1* expression or repressing expression of *EGR1* activated genes, as it binds to *EGR1*-like binding sites. This may only occur in certain circumstances because unlike *EGR1*, *WT1* is not expressed in response to serum stimulation (Morris *et al.*, 1991). *IGF2* (insulin-like growth factor 2) is a growth factor which is thought to have an important role in development. The overexpression of *IGF2* mRNA in Wilms' tumours (Reeve *et al.*, 1985, Scott *et al.*, 1985) makes it a good candidate for regulation by *WT1*. Down-regulation of *IGF2* gene expression can be seen in the kidney as *WT1* is up-regulated (Drummond *et al.*, 1992, Yun *et al.*, 1993). *IGF1R* (insulin-like growth factor 1 receptor) is a transmembrane tyrosine kinase thought to mediate the effects of IGF1&2. In Wilms' tumours there is an inverse relationship between the levels of *WT1* and *IGF1R* expression which correlate with histological differences (Werner *et al.*, 1993). Blockage of *IGF1R* in nude mice using antibodies can prevent growth of Wilms' tumours after inoculation with transformed cells (Gansler *et al.*, 1989). *PDGFA* (platelet derived growth factor A chain) is another growth factor and is a powerful chemoattractant and mitogen for cells of mesenchymal origin (Deuel 1987). The importance of *PDGFA* in kidney differentiation has yet to be demonstrated but mesenchymal cells play a large role in kidney formation (see below). Therefore if functional *WT1* is lost in Wilms' tumours the loss of repression of these genes and therefore their expression may lead to a stimulation of growth as opposed to differentiation (Sukhatme 1992).

1.3.5.3 Mapping of functional domains

It has been possible, for some transcription factors, to define regions responsible for the transacting activity by making deletions in these regions. It has been shown that *EGR1* contains four regions which are responsible for transactivation and one that mediates transrepression (Russo *et al.*, 1993). Binding of *WT1* to DNA is necessary for transrepression and activation. The transrepression activity has been mapped to the proline/glutamine rich region as transferring this region to a heterologous DNA binding motif (*GAL4*) also transfers the transrepression activity (Madden *et al.*, 1993). Finer mapping of this region has failed to produce any convincing further localisation of function (Madden *et al.*, 1993, Wang *et al.*, 1993a). The deletion of three homopolymer motifs, $(Glycine)_5$ (GGGGG), $(Proline)_{13}$ (PAPPPAPPPPPPPPP), $(Proline)_4$ (PPPP) and mutations in the putative leucine

zipper reduce but do not ablate transrepression activity suggesting that the effect is due to a structural disturbance rather than function residing in this region (Figure 1.1).

1.3.5.4 Protein-protein interactions and modulation of *WT1* activity

From the initial simple model of transcriptional regulation in prokaryotes it has become clear that eukaryotic transcriptional regulation occurs through the interaction of multiple factors often bound to multiple promoter sequences (Zawel and Reinberg, 1993). There is some evidence for the modulation of *WT1* action by association with another tumour suppressor, p53 (Maheswaran *et al.*, 1993). Coimmunoprecipitations have shown the presence of p53 and *WT1* in the same complex. Wild type *WT1* (-KTS) could enhance the activation by wild type p53 of one of its targets; *MCK* (muscle creatine kinase) promoter. Also the suppression of transcription using EGR1 consensus sites by *WT1* (-KTS) was found only to occur in the presence of wild type p53, showing that transcriptional repression mediated by *WT1* is not an intrinsic property, but modulated by the cellular environment. The most often observed effect on transcription by *WT1* is transrepression but this is dependent on the number and positioning of binding sites and on cell type. The need for multiple binding sites for repression to occur may be due to *WT1* interacting with itself as has been shown for *Sp1* (Su *et al.*, 1991).

1.3.5.5 Regulation of *WT1* expression

As yet very little is known about what controls the expression of *WT1*. It may be that other WT predisposition genes are located upstream of *WT1* and are involved in its regulation. The region 5' to the transcriptional start position has been cloned in both mouse and human (Pelletier *et al.*, 1991c, Campbell *et al.*, 1993, Hofmann *et al.*, 1993). Similarities exist between these regions in that neither possess TATA or CCAAT elements but there are numerous putative *Sp1*, *EGR1*, *WT1*, *PAX8*, *PAX2*, *AP2*, *AP4* and GAGA binding sites. The presence of numerous *Sp1* sites is consistent with other TATA-less, GC-rich promoters. The binding of *Sp1* to the promoter region has been demonstrated *in vitro* (Hofmann *et al.*, 1993). From the sequence of the *WT1* promoter it was predicted that *WT1* may bind it own promoter (Hofmann *et al.*, 1993). This has now been shown to be the case *in vitro* (Campbell *et al.*, 1993). All four isoforms of *WT1* can down-regulate its expression, the isoform which most efficiently does this was the +17aa, +KTS isoform (Rupprecht *et al.*, 1994). A basal promoter region has also been defined and similar to other GC-rich TATA-less

promoters it is non-cell-specific. It has been proposed that in at least one cell line an enhancer of transcription lies 50kb downstream in exon 10 (Frazier *et al.*, 1994).

1.4 Evidence that *WT1* is a predisposition gene for Wilms' tumour

On the basis of the structure of the predicted protein it was proposed that *WT1* could be acting as a transcription factor involved in the control of kidney development and that loss of the gene function would perturb development giving cells the opportunity to become neoplastic. In order to demonstrate that *WT1* is involved in the development of Wilms' tumours the presence of intragenic mutations, the ultimate test for a candidate gene, in tumour DNA was investigated. In addition the expression pattern was investigated to see if it was consistent with a role in kidney development.

1.4.1 Mapping deletions in Wilms' tumours to the *WT1* locus

Very few Wilms' tumours have homozygous deletions in *WT1* detectable by Southern blotting (Cowell *et al.*, 1991, Royer-Pakora *et al.*, 1991, Ton *et al.*, 1991a). In fact most of the Wilms' tumours examined expressed *WT1*, indicating that in most cases more subtle lesions must be occurring than complete ablation of the gene product (Gessler *et al.*, 1990, Huang *et al.*, 1990, Pritchard-Jones *et al.*, 1990). Comparison of two tumours WiT-13 and PER which are both homozygously deleted for the WT region defined a region of over 150kb (Gessler *et al.*, 1990, Rose *et al.*, 1990). As the approximate genomic region covered by a gene is 50kb it was possible that *WT1* was just very closely linked to the real predisposition gene. A second mRNA transcript had already been isolated within the WT region which was also expressed in the developing kidney (Bonetta *et al.*, 1990, Huang *et al.*, 1990). The WT region was reduced by analysis of another sporadic Wilms' tumour homozygously deleted for the WT region (Ton *et al.*, 1991a). The deletions were found not to extend 3' of the first few exons of *WT1* and with the 5' end of the WT region defined by the 5' extent of the deletion in PER, the WT region was reduced to 16kb. The smallest region of overlap of the homozygous deletions in a sporadic Wilms' tumour was found to extend 3' of the *WT1* gene deleting only the last zinc finger and not extending as far as the next CpG island (Cowell *et al.*, 1991). These last two deletions do not overlap but both delete part of the *WT1* gene: this heavily implicates *WT1* as the predisposition gene.

1.4.2 Intragenic deletions

In a sporadic unilateral Wilms' tumour, a small intragenic deletion was found, the deletion of 25bp at an exon/intron junction resulted in the deletion of zinc finger 3, so altering or abolishing DNA binding activity (Haber *et al.*, 1990, Rauscher *et al.*, 1990, Wang *et al.*, 1993a). LOH on 11p was detected but the tumour was heterozygous for the *WT1* deletion, and so had one normal copy of the gene, suggesting that *WT1* mutation was the second hit and the predisposing mutation was due to LOH for 11p.

The first reported case where a tumour became homozygous for an intragenic deletion in *WT1* was in a patient with bilateral tumours, with no associated abnormalities, who had a constitutional deletion of <11kb within the *WT1* gene. This caused a 97bp deletion in the RNA transcript and truncation of the protein by an in-frame stop codon. Both tumours were shown to have become homozygous for this deletion of *WT1* but analysis of chromosome 11 markers identified these as separate events (Huff *et al.*, 1991). A second phenotypically normal case has been shown to have a constitutional deletion of two exons that became homozygous in the tumour (Tadokoro *et al.*, 1992) (Table 1.2).

Analysis of the remaining allele in four unilateral WAGR patients with constitutional deletions spanning the WT locus showed that in these cases the tumours contained mutations in the *WT1* gene that would produce a truncated protein product (Brown *et al.*, 1992, Baird *et al.*, 1992a, Gessler *et al.*, 1993) (Table 1.3). A second independent hit would be expected in these cases as becoming homozygous for these large deletions would probably lead to cell lethality. These cases therefore show that in most instances Knudson's two hit model does apply for Wilms' tumour and *WT1*.

So far *WT1* mutations have been detected in only about 10% of the Wilms' tumours analysed, and most have been limited to the zinc fingers (Tables 1.2-1.4). These will be expected to change or ablate the DNA binding activity. There have been a few reports of changes in the rest of the protein. A point mutation has been found in a Wilms' tumour that changes a glycine to aspartic acid in exon 3 (codon 201) (Park *et al.*, 1993c); this has been shown to change *WT1* from a transrepressor to transactivator. The removal of exon 2 by alternative splicing has also been shown to change the properties of *WT1* in the same way. This exclusion of exon 2 has been found only in cells derived from Wilms' tumours and in the absence of other *WT1* mutations. The resultant protein is found to be unable to repress growth of cells unlike wild type *WT1*, implicating it in tumorigenesis (Haber *et al.*, 1993). Another

Table 1.2 *WT1* mutations in WT with no associated congenital abnormalities

Summary of characterised *WT1* mutations in Wilms' tumour cases where there are no associated congenital abnormalities. The genetic events which are thought to be the first and second hits in tumorigenesis are shown.

Tumour Incidence	First mutation	Karyotype	Mutation in *WT1* gene	Change in *WT1* protein	Status of mutation in tumour	First hit	Second hit	Reference
unilateral	tumour	N/D	25bp deletion at exon 9/ intron 9 boundary	Exon skipping deletes ZF3	heterozygous	LOH 11	*WT1* mutation	Haber *et al.*, 1990
unilateral	tumour	N/D	missense mutation in exon 8	366R-C in ZF2	heterozygous	?	?	Little *et al.*, 1992a
unilateral	tumour	46XX	4bp insertion in exon 2	truncation in exon 2	homozygous	*WT1* truncation	LOH 11p	Park *et al.*, 1993b
bilateral	germline	46XX	<11kb deletion from intron 5 to 6	exon 6 deleted, truncation in exon 7	1) homozygous 2) homozygous	1) *WT1* truncation 2) *WT1* truncation	1) LOH 11p 2) LOH 11p	Huff *et al.*, 1991
bilateral	germline	N/D	chain termination mutation in exon 9	390R-STOP, truncation in ZF3	1) homozygous 2) heterozygous	1) *WT1* truncation 2) *WT1* truncation	1) LOH 11p 2) ?	Little *et al.*, 1992a
unilateral	germline	46XY	8kb deletion introns 5-7	truncation in ZF2	homozygous	*WT1* truncation	LOH 11p	Tadokoro *et al.*, 1992
unilateral	tumour	46XY	1bp insertion in exon 10	frameshift in ZF4, extension of 66aa	homozygous	*WT1* mutation	LOH	Coppes *et al.*, 1993a
unilateral	tumour	46XX	1bp deletion in exon 10	frameshift in ZF4, extension of 66aa	homozygous	*WT1* mutation	LOH	Coppes *et al.*, 1993a
unilateral	germline	46XX	chain termination mutation in exon 8	362R-STOP truncation in ZF2	homozygous	*WT1* truncation	LOH	Coppes *et al.*, 1993a
unilateral	germline	46XX	missense mutation in exon 9	394R-W	homozygous	*WT1* mutation	LOH	Akasaka *et al.*, 1993
unilateral	tumour	N/D	chain termination mutation in exon 9	390R-STOP truncation in ZF3	homozygous	*WT1* truncation	LOH	Varanasi *et al.*, 1994
unilateral	tumour	N/D	deletion & insertion in exon 9	truncation in ZF3	homozygous	*WT1* truncation	LOH	Varanasi *et al.*, 1994
unilateral	tumour	N/D	1) chain termination mutation in exon 8 2) 7bp insertion in exon 3	1) 362R-STOP truncation in ZF2 2) frameshift and truncation	heterozygous	*WT1* mutation	*WT1* mutation	Varanasi *et al.*, 1994
unilateral	tumour	N/D	chain termination mutation in exon 8	362R-STOP truncation in ZF2	homozygous	*WT1* truncation	LOH	Varanasi *et al.*, 1994
unilateral	tumour	N/D	deletion & insertion in exon 8	frameshift and truncation	heterozygous	?	?	Varanasi *et al.*, 1994
unilateral	tumour	N/D	missense mutation in exon 7	338S-Y	heterozygous	?	?	Varanasi *et al.*, 1994

point mutation in a Wilms' tumour associated with BWS has been reported. This is a phenylalanine to serine change in exon 2 (codon 154), as yet it has to be shown if this results in a change in the activity of *WT1* (Park *et al.,* 1993b).

1.5 Kidney development and *WT1*

In order to demonstrate how *WT1* could be involved in Wilms' tumour formation, the expression of *WT1* was investigated in the developing human embryo and Wilms' tumours (Pritchard-Jones *et al.*, 1990). Subsequently similar patterns of developmental expression have been seen in mouse and rat (Buckler *et al.*, 1991, Armstrong *et al.*, 1992, Sharma *et al.*, 1992).

1.5.1 The developing kidney

The development of the mammalian metanephric kidney has been investigated extensively as an example of how development occurs as a result of the interaction between the epithelial and mesenchymal components. The metanephric kidney is particularly amenable to investigation as it can be cultured and manipulated *in vitro*. It has been shown using transfilter culturing techniques that differentiation is dependent on a series of reciprocal inductions between these two components (Saxen 1970).

Kidneys arise from the nephrogenic mesenchyme of the intermediate mesoderm which lies between the lateral plate mesoderm and the somitic mesoderm and abuts onto the coelom (Gilbert 1991). A wave of differentiation extends caudally down the mesoderm from the region of the fourth somite forming the filtration units, the nephrons (Saxen 1987). Three kidneys are formed successively, the pronephros, mesonephros, and metanephros (Figure 1.3). A wave of degeneration follows the wave of differentiation removing the first two kidneys during foetal life leaving the metanephros as the functional adult kidney. Although the three kidneys are referred to as separate entities they are very closely associated and often it is hard to distinguish the very transitory pronephros from the beginning of mesonephric differentiation (Torrey 1965).

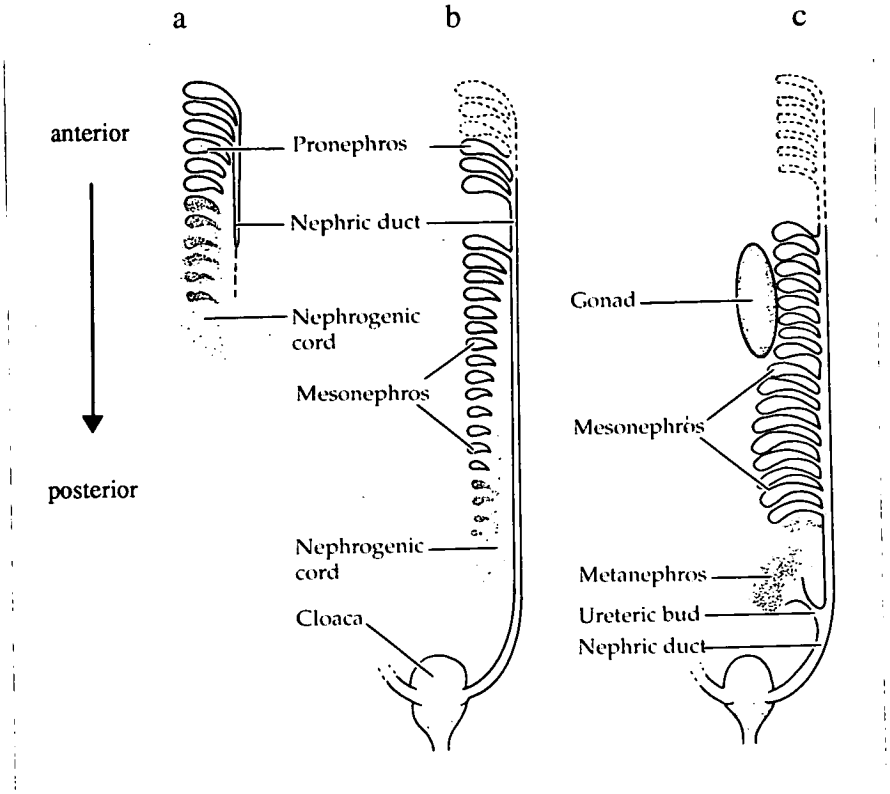
Kidney development is initiated when the pronephric analage form in the intermediate mesoderm, at about the level of the 4th somite, which then differentiate into the pronephric tubules. The pronephric duct extends caudally from this region until it joins the cloaca. As it extends it induces the differentiation of more pronephric

Figure 1.3 Differentiation of the mammalian kidney

a The first pronephric tubules are seen in human at about day 22 at the level of the 4th somite on either side of the midline. The nephric (pronephric) duct extends caudally inducing nephron differentiation.

b The first mesonephric differentiation occurs about day 24 in human. At this time the pronephric tubules are already regressing and the nephric (mesonephric) duct has reached the cloaca.

c The metanephros starts to differentiate about the 5th week of gestation, induced by the formation of the ureteric bud from the nephric duct near to the cloaca. The pronephros has completely regressed and the mesonephros is starting to do so. The gonad is differentiating on the medial side of the mesonephros. (Illustrations from Gilbert 1991 after Saxen 1987)



tubules and then the mesonephric tubules from the nephrogenic mesoderm. In the region of the mesonephros it is known as the mesonephric duct (Figure 1.3b).

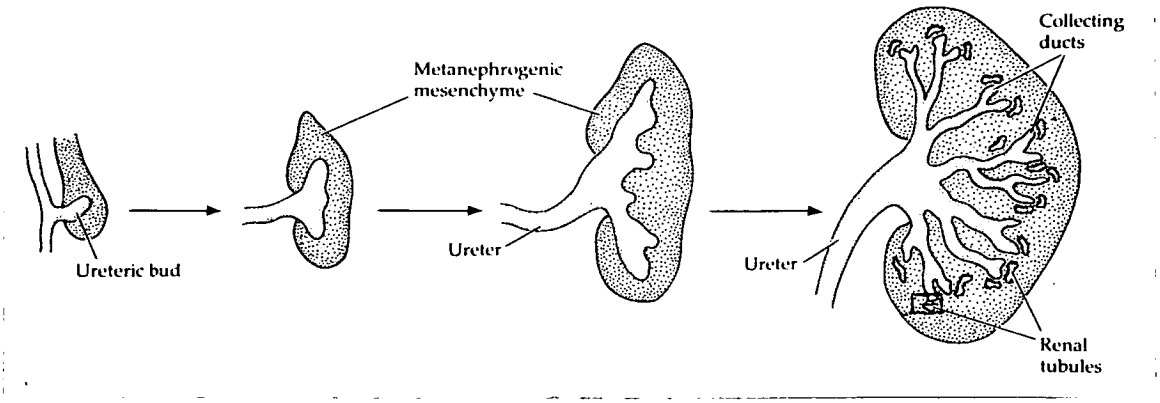
The initiation of the metanephros occurs when the ureteric bud branches off the mesonephric duct just above the cloaca (Figure 1.3c). This bud induces the condensation of the surrounding nephrogenic mesenchyme (Grobstein 1955, Sariola *et al.,* 1989, Bard and Woolf, 1992). The bud then divides dichotomously to form many branches (Figure 1.4a). As this is happening the mesenchyme proliferates and follows one of three pathways either leading to apoptosis (Coles *et al.,* 1993), formation of the supporting stroma or differentiation into epithelia of the nephrons.

Northern blotting and *in situ* hybridisation of foetal metanephric kidneys shows the up-regulation of *WT1* mRNA as the metanephric blastema is induced to condense by the ureteric bud. The induction leads to the formation of characteristic comma shaped and then S-shaped bodies as the nephric tubules differentiate (Figure 1.5a). The tubules further differentiate into Bowman's capsules, proximal, and distal tubules, including loops of Henle. The distal end joins up to the collecting duct system formed from the branched ureteric bud. As differentiation proceeds, *WT1* expression is limited to the developing nephrons becoming localised to the Bowman's capsule, and then to the podocytes (Pritchard-Jones *et al.*, 1990). No *WT1* expression has been seen in the ureteric bud or the collecting system. There is a second wave of differentiation of the nephrogenic blastema when the collecting ducts are no longer branching but are still competent to induce the metanephric blastema, forming arcades of nephrons which empty into the same collecting duct, thus greatly increasing the capacity of the kidney. In humans, mice, and rats the level of *WT1* expression declines after the last wave of nephron differentiation and loss of the renal blastema has occurred either before birth or neonatally (Pritchard-Jones *et al.*, 1990, Buckler *et al.*, 1991, Armstrong *et al.*, 1992, Sharma *et al.*, 1992). *WT1* expression is not completely down-regulated as *WT1* mRNA transcripts can be detected by PCR and the *WT1* protein is also present in the adult kidney (Mundlos *et al.,* 1993).

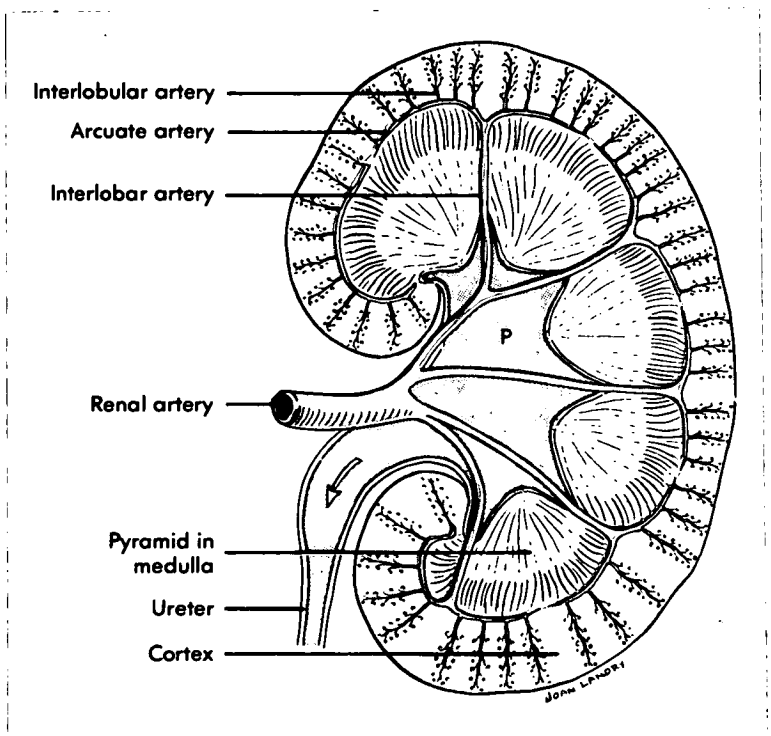
The differentiation of nephrons is thought to occur in a similar way in all three kidneys. A similar pattern of expression is seen during nephron differentiation in the mesonephros and the metanephros in both man and mouse. Up-regulation of *WT1* occurs during tubule differentiation, declining as regression of the mesonephros takes place (Pritchard-Jones *et al.*, 1990, Armstrong *et al.*, 1992). The earliest nephrons in the pronephros are different in that there is no Bowman's capsule but filtration is through the coelomic epithelia into the coelom. These structures are known as an external glomeruli (Figure 1.5b). In these the pattern of *WT1* expression is not known

Figure 1.4 Differentiation of the metanephros

a The ureteric bud within the metanephric mesenchyme divides to form the collecting ducts of the mature kidney and induces the differentiation of the mesenchyme into the nephrons from the centre of the kidney outwards. (Illustration from Gilbert 1991)



b Structure of the adult kidney, showing the division of the kidney into lobes and the cortex and the medulla. The glomeruli and proximal tubules are contained in the cortex and the collecting ducts and loops of Henle form the medulla. (Illustration from Kent 1987)



because of their very transient nature.

Vascularisation of the differentiating metanephric kidney is thought to occur mainly by invasion of cells from the surrounding capillaries (Saxen 1987) forming the glomeruli and the peritubular capillaries. But there is also by some angiogenesis within the kidney as demonstrated by capillary formation occurring in culture in the absence of the surrounding capillaries.

1.5.2 The molecular basis of kidney differentiation

The series of changes which occur within the cells as nephron development proceeds is only just becoming elucidated at the molecular level. The inductive signals from the nephric duct are thought to be permissive as opposed to instructive because other tissues can act as inducers of nephrogenesis demonstrating that the mesenchyme is predetermined (Sariola *et al.*, 1989). Following induction by the ureteric bud the mesenchymal cells divide and this is followed by a change in surface markers. These changes are associated with alterations in the adhesion properties of the cells which are very important for the differentiation of the epithelial nephrons from mesenchyme. Thirty six hours after induction typically epithelial markers such as laminin A chain and uvomorulin (*L-CAM*) are detected. These are involved in the formation of a basement membrane and the formation of cell to cell contacts and the polarisation of the cells. The mesenchymal marker protein vimentin is lost at this time (Ekblom 1981).

The control of these events is still unclear. The expression of several transcription factors is up-regulated as the cells are starting to differentiate; *N-myc* (24 hours after induction) (Mugrauer and Ekblom 1991) and then *PAX2*, *PAX8* and *WT1* (36 hours after induction) (Dressler *et al.*, 1990, Plachov *et al.*, 1990). Therefore the localisation and timing of expression of these genes points to a role in the control of the induction of nephrogenesis; specifically the mesenchymal to epithelial cell type transition taking place in the developing nephrons (Pritchard-Jones *et al.*, 1990). The importance of one of these genes, *PAX2*, has been experimentally demonstrated, as antisense oligonucleotides in culture can block the cell type transition (Rothenpieler and Dressler 1993) and overexpression produces deformed kidneys in mouse, which resemble the condition congenital nephrotic syndrome (Dressler *et al.*, 1993). It is thought that the *PAX2* gene may be a target for *WT1* regulation because in Wilms' tumours *PAX2* expression similarly to *WT1* expression is not down regulated as it is in normal kidney after tubular differentiation (Dressler and Douglas 1992). A fourth

Figure 1.5 Nephron development

a Mesenchyme is induced to condense by the nephric duct or the ureteric bud. The cells become polarised, form a basement membrane and take on the characteristics of epithelia. The condensate develops a lumen becoming comma and then S-shaped as the renal tubule forms. The end of the nephron nearest the collecting duct fuses to it while other end differentiates into the Bowman's capsule. Vascularisation occurs forming the glomeruli and the peritubular capillaries. (Illustration from Saxen 1987)

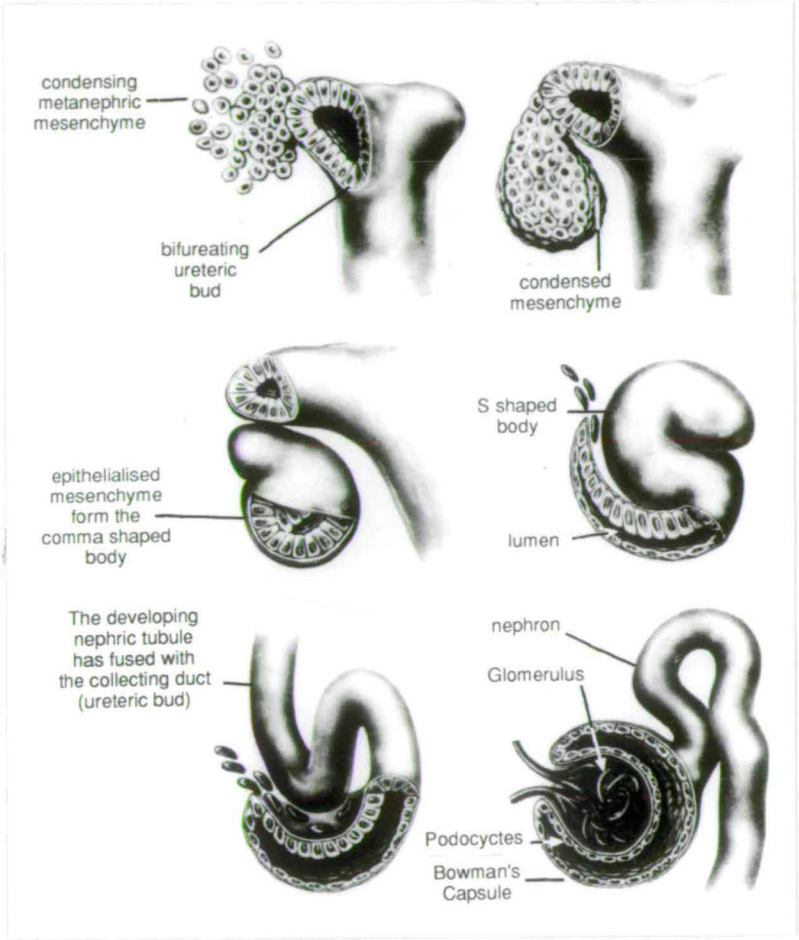
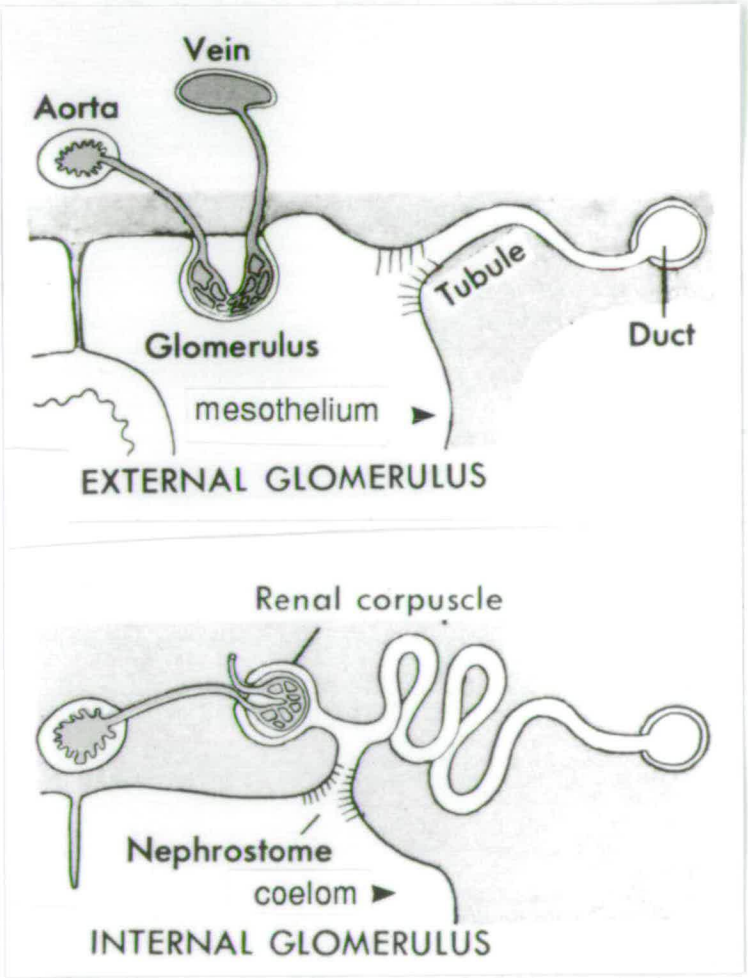


Figure 1.5 Nephron development (continued)

b In most mammalian nephrons the glomerulus is within the Bowman's capsule. Occasionally the nephron communicates to the coelom via a nephrostome. In the first pronephric tubules and lower vertebrates the glomerulus is in direct contact with the coelom (external glomerulus) with the coelomic epithelium equivalent to the podocytes of the Bowman's capsule. The nephric tubules also connect to the coelom. Therefore the formation of the Bowman's capsule is thought to have occurred by invagination of the glomerulus by the nephric tubule (internal glomerulus). (Illustration adapted from Kent 1987)



gene has also been implicated in this transition, *c-met* which is the ligand for hepatocyte scatter factor (Tsarfaly *et al.,* 1994).

1.5.3 Wilms' tumours

Wilms' tumours can be classified on the basis of histology. The most common type is triphasic, containing undifferentiated blastema, but also stroma, and elements which resemble the tubules and glomeruli. Heterologous tissues such as skeletal muscle, cartilage, adipose tissue and bone are also observed and are believed to be derived from the stroma. It is thought that after not taking the correct developmental pathway these undifferentiated cells can respond to inductive signals to which they are subsequently exposed to try to form nephrons or other differentiated tissue types. Tumours are also classified according to their association with 'islands' of abnormally persistent blastema known as nephrogenic rests. It has been proposed that these arise because the system of kidney differentiation depends on a series of inductions and is not determined by an inalterable 'blueprint' and is therefore susceptible to error (Mierau *et al,* 1987). If the blastema either does not receive the inductive signal or fails to respond correctly to induction it can persist postnatally within the developed metanephric kidney. These are divided into either perilobar (PLNR) or intralobar (ILNR) types i.e. those found at the margins of the lobes of the metanephros or within the lobes (Beckwith *et al.*, 1990) (Figure 1.4b). ILNR associated tumours are often triphasic and contain heterologous tissues. PLNR associated tumours have a much more homogeneous histology, consisting mostly of blastema and epithelial elements. Since the metanephric blastema differentiates from the centre outwards, the peripheral location indicates that PLNR associated tumours have arisen later in nephrogenesis. At this time the cells maybe less pluripotent than in the initial wave of nephron formation resulting in the persistent blastema having fewer developmental options than in ILNR associated tumours. Nephrogenic rests have been reported in <1% of infant post-mortem (Bennington and Beckwith 1975). In contrast to the general population, rests are present in the kidneys of 40% of unilateral and 100% of bilateral Wilms' tumour patients (Bove and McAdams 1976, Beckwith *et al.*, 1990). The high incidence of WT and rest association and their similarities have suggested that tumours form from rests that have become neoplastic. In the case of ILNRs, these are often found at the edge of tumours indicating that the tumour may have arisen from a cell within the rest. Further evidence for this comes from two cases where the same *WT1* mutation was found in an associated rest as in the Wilms' tumour (Park *et al.,* 1993b). There is also a preferential association of rest and tumour type with different

associated congenital syndromes (Table 1.1). Both DDS and WAGR associate with the intralobar type (Mierau *et al.*, 1987, Heppe *et al.*, 1991) but perilobar types are more often to be found in BWS cases. This may indicate that there is genetically determined heterogeneity in the developmental abnormalities.

The expression of *WT1* has been investigated in tumours and *WT1* mRNA transcripts were found at high levels, but only in the analogous structures to those expressing *WT1* in normal development. Tumours with a high blastemal content in general express high levels of *WT1*. In more differentiated tumours *WT1* becomes localised to the immature tubules and glomeruloid bodies (Pritchard-Jones and Fleming, 1991, Miwa *et al.*, 1992a). Stromal rich tumours including stroma derived heterologous tissues (e.g. muscle and cartilage) were found to express very low or undetectable levels of *WT1* (Pritchard-Jones and Fleming, 1991, Huang *et al.*, 1990, Miwa *et al.*, 1992a, Yeger *et al.*, 1992). The origin of stromal tumours is not as clear as for other types. It is not known whether the precursor cells are derived from the uninduced or induced metanephric blastema. Pritchard-Jones and Fleming (1991) have suggested the second alternative on the basis that in order for a *WT1* lesion to affect cell growth, the cells should have passed though the stage of *WT1* expression. Mesenchymal cells which have not passed though the induced blastemal stage have not expressed *WT1*.

In very rare cases tumours histologically identical to nephroblastoma occur outside the metanephric kidney and in the adult (Aterman 1989). Extra-renal nephroblastoma is mainly found in the retroperitoneal space and in the gonads, or their associated ducts. Such tumours were originally thought to be due to displaced metanephric tissue, but the tumours could possibly arise from mesonephric kidney remnants that have failed to regress or from the gonads themselves as they are very closely related to the kidneys (Pritchard-Jones and Hastie 1990). The expression of *WT1* in these tumours has been investigated and it was found that in the endometrial derived tumours there is a persistence of high levels of *WT1* expression similar to some Wilms' tumours indicating that *WT1* may also be involved in their formation (Roberts *et al.*, 1993).

1.6 *WT1* and intermediate mesoderm development

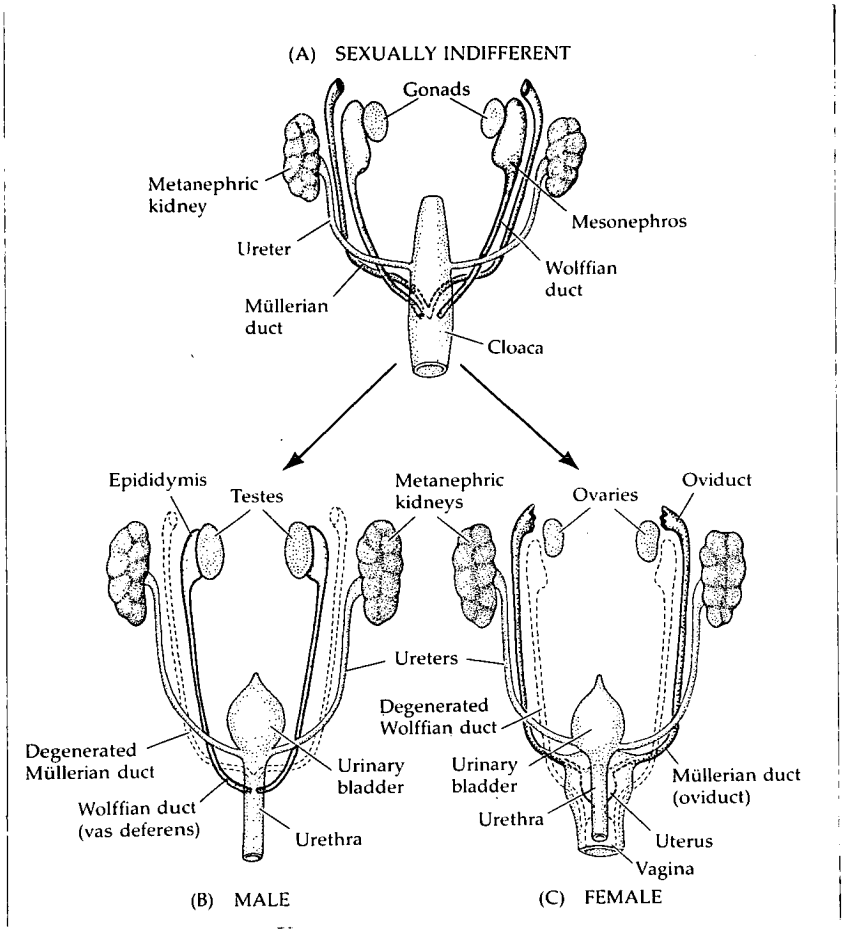
The gonads develop in close juxtaposition to the mesonephroi in the intermediate mesoderm (Figure 1.3a&1.6a) and in addition it has been shown that for complete development of the gonad some interstitial cells must be contributed by the

Figure 1.6 Secondary sexual determination

a The indifferent gonad differentiates on the medial side of the mesonephros. The Mullerian duct differentiates in both sexes in the intermediate mesoderm parallel with the Wolffian (mesonephric) duct and both ducts join the cloaca.

b In the male the testis produces testosterone and AMH which induces the retention and differentiation of the Wolffian duct into the vas deferens and the mesonephric tubules into the epididymis and the regression of the Mullerian duct.

c In females the mesonephric tubules and ducts degenerate and the Mullerian duct form the oviducts and the uterus under the influence of oestrogen. (Illustrations from Gilbert 1991)



overlying mesonephros (Byskov 1986, Buehr *et al.*, 1993). The non germ cell component of the gonad forms from a thickening of the coelomic epithelium in the intermediate mesoderm, differentiating into the interstitial component (including Leydig cells) and the epithelial component of the mature gonad involving a mesenchymal to epithelial cell type transition. From the first stages of development the gonad expresses high levels of *WT1*. On sexual differentiation this becomes localised to sex cords, in particular the Sertoli and granulosa cells of the adult gonad but not the germ cells (Pelletier *et al.*, 1991c, Armstrong *et al.*, 1992). *WT1* expression is not as down regulated in the adult gonad as in the kidney as it remains detectable by northern blotting (Pelletier *et al.*, 1991c). In the testis a smaller 2.5kb transcript was noted but not in other tissues (Pelletier *et al.*, 1991c). Since then, the presence of a smaller transcript has been observed in other tissues and species (Sharma *et al.*, 1992).

The mesothelium which covers all the internal organs also expresses *WT1* during development and is derived from the coelomic epithelium. The kidney, gonads and mesothelium are related by their derivation from the coelomic epithelium, arising from the intermediate mesoderm and they all undergo a cell type transition, mesenchyme to epithelia, during development. Therefore they are intimately related by virtue of position, cell origin, differentiation pathway and the expression of *WT1*, indicating that many similar developmental decisions are taking place in these cells (Pritchard-Jones *et al.*, 1990). The evolutionary origin of the glomerular podocytes is directly related to the mesothelium because in the primitive external glomeruli the filtration surface is not the podocytes but the coelomic epithelium. The Bowman's capsule seems to have been formed by the invagination of the glomeruli by the nephric tubules (Figure 1.5b). Interestingly the other organ to be derived from the intermediate mesoderm, the adrenal cortex, does not express *WT1*.

The genital ducts are also intimately associated with this region. In both sexes a second pair of ducts, the Mullerian ducts, forms from a longitudinal groove of the coelomic epithelium paralleling the mesonephric duct (Gilbert 1991). At this stage these ducts have not been observed to express *WT1*. Secondary sex determination which decides the fate of the Mullerian and Wolffian ducts (mesonephric ducts) is dependent on hormones produced by the developing testes and ovaries.

The developing testes produce testosterone from the Leydig cells and anti-Mullerian hormone (AMH) from the Sertoli cells. The former promotes the differentiation of the Wolffian ducts and the mesonephroi into the vas deferens and epididymis and the latter promotes the degeneration of the Mullerian ducts. (Figure

1.6b). Testosterone or its derivatives also affect the development of the external genitalia by acting on the urogenital sinus and swellings which then become the penis and scrotum (O'Rahilly 1977). After the initial development of the mesonephros none of these structures have been observed to express *WT1* (Pelletier *et al.*, 1991c).

In the female in the absence of these two hormones, and in the presence of oestrogen produced by the ovaries, the Wolffian ducts and mesonephroi regress and the Mullerian ducts differentiate into the internal genitalia (Figure 1.6c). The caudal ends of the ducts open into the cloaca and the anterior ends remain open to form the ostia, which collects the mature oocytes. In the middle their fusion produces the uterus and contributes to the vagina (Jiresek 1977). Of these structures the ostia, the myometrium and endometrial stroma of the uterus but not the epithelial component of the endometrium have been observed to express *WT1* in the mouse and rat (Pelletier *et al.*, 1991c, Zhou *et al.*, 1993). During normal hormonal cycling in the female the expression of *WT1* was found to remain constant, but 6 days after fertilisation as the myometrial stroma differentiated into the decidual cells, up-regulation of expression occurred, implicating *WT1* in differentiation of these cells.

1.7 *WT1* expression outside the intermediate mesoderm

Other cell populations with the same cell type transition have not been observed to express *WT1* by *in situ* hybridisation e.g. the somites, the endothelial lining of the capillaries, the corneal epithelium and the outer lining of the coelom derived from the lateral plate mesoderm (Armstrong *et al.*, 1992). By northern blotting heart, lungs, and thymus (Buckler *et al.*, 1991) express *WT1* but by *in situ* hybridisation this has been localised to their mesothelial covering (Armstrong *et al.*, 1992, Park *et al.*, 1993a).

At 11dpc the central region of the spinal cord in mouse expresses *WT1* and at day 15 this becomes specific to the ventral horn, this is also seen in human and rat. In the brain a small part of the medulla oblongata expresses, as does the area postrema in rats. Other non mesoderm derived structures, the tongue and the eye, express *WT1* when examined by RT-PCR (Armstrong *et al.*, 1992, Sharma *et al.*, 1992).

The first *WT1* clone was isolated from a preB-cell library (Call *et al,.* 1990), so a possible role in haematopoesis has been investigated. *WT1* expression has been shown to be limited to the splenic capsule and stroma, not the haematopoeitic cells (Park *et al.*, 1993a). Expression does occur in some immature leukaemic cells i.e. ALL, AML, CML but not in mature types (Call *et al,.* 1990, Miwa *et al.*, 1992b).

These regions of *WT1* expression outside the intermediate mesoderm and in structures not undergoing the mesenchyme epithelial transition indicate that *WT1* may have additional roles in other cell populations.

1.8 Involvement of *WT1* in genitourinary abnormalities

The association of Wilms' tumour with genitourinary abnormalities is particularly interesting because of the intimate association of kidney and genital systems and because both of them express high levels of *WT1* during development.

1.8.1 Mapping of genitourinary abnormalities to the *WT1* region

Mapping of constitutional WAGR deletions demonstrated that the locus involved in genitourinary abnormalities was in the same 345kb region that contained *WT1*. At the same time this clearly demonstrated the separate identity of the gene causing aniridia. It was proposed that genitourinary abnormalities could be due to a pleiotropic effect of *WT1* as opposed to the deletion of another gene, giving rise to this phenotype by haploinsufficiency (van Heyningen *et al.*, 1990). The lack of one functional copy of *WT1* in the developing genitourinary system could perturb developmental events such as the production of the hormones from the gonads (Pritchard-Jones *et al.*, 1990, Pelletier *et al.*, 1991a) which control the development of the external genitalia therefore leading to these malformations (Behringer *et al.*, 1990). This was further supported by two constitutional intragenic deletions in *WT1* being reported in two patients with Wilms' tumour, hypospadias and cryptorchidisim but no aniridia (Pelletier *et al.*, 1991a). As with WT, this cannot be a fully penetrant effect because one of these cases was familial but the father from whom the mutation was derived was unaffected. Several other cases of constitutional loss of one functional copy of *WT1* not resulting in genitourinary abnormalities have been reported (Table 1.2). The frequency of genitourinary malformations in XY WAGR is greater than for XX cases (Table 1.3). This could be consistent with the fact that the genitalia of females are not as dependent on hormone production from the developing gonad as males and so the perturbations caused by loss of one copy of *WT1* do not affect female genitalia development to such a great extent. This may be a reason why familial WT represents 1% of all cases (c.f. 35-45% for retinoblastoma (Knudson 1971)) because even where no obvious abnormalities are present fertility may be reduced.

1.8.2 Denys-Drash Syndrome

In addition to WAGR syndrome, Wilms' tumour is a component of Denys Drash syndrome which has much more severe genitourinary abnormalities in both XX and XY cases including ambiguous genitalia, pseudo-hermaphroditism, micropenis, cryptorchidism, hypospadias, and streak gonads (Denys *et al.*, 1967, Drash *et al.*, 1970). The kidneys also are affected and lose their function in childhood due to glomerular nephropathy (specifically mesangial sclerosis). Analysis of 10 DDS cases revealed constitutional point mutations in zinc fingers 2 and 3 of *WT1*, in these cases the mutations became homozygous in all the Wilms' tumours examined (Pelletier *et al.*, 1991b). The majority of DDS patients examined have a specific set of mutations. These are missense mutations in zinc fingers 2 and 3 (Table 1.4 and references therein). These mutations affect the residues chelating the zinc ion or those which had been shown to interact with DNA in the co-crystal structure of *EGR1* with its consensus sequence (Pavletich and Pabo 1991) (Figure 2). Approximately 60% of these are at R394 or D369 in ZF3. By analogy with the crystal structure of *EGR1*, the arginine residue probably binds a guanine base and the aspartate residue interacts with the arginine to stabilise the interaction with the DNA. Some of these mutations have been shown experimentally to destroy DNA binding to the *EGR1* consensus sequence (Pelletier *et al.*, 1991b). One interesting exception has been reported twice (Bruening *et al.*, 1992, Konig *et al.*, 1993) where a mutation in the splice donor site of exon 9 prevents the inclusion of the +KTS alternative splice. That this mutation in the heterozygous state can give rise to the DDS phenotype indicates that the ratio of the different isoforms of *WT1* must be of functional importance.

1.8.3 Dominant-negative hypothesis

The genitourinary abnormalities produced by these DDS mutations are more severe than those in WAGR cases and other constitutional deletion cases (Table 1.1). WAGR mutations normally lead to truncation of *WT1* or one *WT1* allele is completely deleted (Table 1.3) whereas DDS mutations are mostly missense mutations (Table 1.4). WAGR genitourinary abnormalities presumably are a result of *WT1* haploinsufficiency, while it is suggested that the DDS phenotype is due to a trans-dominant or dominant-negative effect (Herskowitz 1987, Su *et al.*, 1993). A dominant effect would be due to a new function of the *WT1* mutant causing the phenotype, and a dominant-negative effect where the remaining normal *WT1* function is interfered with by the mutant (Pelletier *et al.*, 1991b, Little *et al.*, 1993). The

Table 1.3 *WT1* mutations in WT with associated congenital abnormalities

Summary of *WT1* mutations reported in Wilms' tumours associated with congenital abnormalities (excluding Denys-Drash syndrome)

Phenotype	Tumour incidence	Karyotype	Mutation in *WT1* gene	Change in *WT1* protein	Status of mutation in tumour	First hit	Second hit	Reference
HP, C	bilateral sporadic	46XY	17bp deletion in exon 4	truncation in exon 4	1) homozygous 2) homozygous	*WT1* truncation	LOH	Pelletier *et al.*, 1991a
HP, C	unilateral Familial	46XY	1bp deletion in exon 6	truncation in exon 7	homozygous	*WT1* truncation	LOH	Pelletier *et al.*, 1991a
HP, C, DK, AN	unilateral sporadic	46XY del(11)p13p14.1	deletion of exons 2-3 in mRNA	truncation after exon 4	hemizygous	large constitutional deletion	*WT1* truncation	Brown *et al.*, 1992
GU, R, AN	unilateral sporadic	N/D del(11)p12p13	10bp insertion in exon 7	truncation in exon 7	hemizygous	large constitutional deletion	*WT1* truncation	Baird *et al.*, 1992a
GU, R, AN	unilateral sporadic	N/D del(11)p13	chain termination mutation in exon 8	362R-STOP in ZF2	hemizygous	large constitutional deletion	*WT1* truncation	Baird *et al.*, 1992a
HP, C, AN	unilateral sporadic	46XY del(11)p13p14.1	1bp deletion in exon 7	truncation in exon 7	hemizygous	large constitutional deletion	*WT1* truncation	Gessler *et al.*, 1993
C, AN, R	unilateral sporadic	46XY del(11)p13	missense mutation in exon 3	201G-D	hemizygous	large constitutional deletion	*WT1* mutation	Park *et al.*, 1993c
AGU, AN, R	unilateral sporadic	46XX del(11)p12p14.3	14bp duplication at exon/intron7 splice donor site	uncharacterised -may be exon skipping leading to truncation	hemizygous	large constitutional deletion	*WT1* mutation	Santos *et al.*, 1993
BWS	unilateral sporadic	46XX	missense mutation in exon 2	154F-S	heterozygous	genetic event at WT2	*WT1* mutation	Park *et al.*, 1993b

AGU, absence of gonads and uterine cavity: AN, aniridia: BWS, Beckwith Wiedemann Syndrome: C, cryptorchidism: DK, duplex kidney: GU, unspecified genitourinary abnormalities: HP, hypospadias: R, mental retardation.

Table 1.4 *WT1* mutations associated with Denys-Drash syndrome

Summary of constitutional *WT1* mutations found in patients with DDS.

Phenotype	Karyotype	*WT1* mutation in gene	Change in *WT1* protein	Incidence of Wilms' tumour	Status of mutation in tumour	Reference
F, SG	46XY	missense mutation in ZF2	366R-H	none gonadalblastoma	-	Pelletier *et al.*, 1991b
AM or F, SG	Two, 46XY Two, 46XX	missense mutation in ZF3	394R-W	unilateral	Homoygous in two cases examined	Pelletier *et al.*, 1991b
AM	N/D	missense mutation in ZF3	394R-W	none	-	Pelletier *et al.*, 1991b
AM	46XY	missense mutation in ZF3	394R-W	bilateral +JGCT	homozygous	Pelletier *et al.*, 1991b
F	46XX	missense mutation in ZF3	394R-W	bilateral	N/D	Pelletier *et al.*, 1991b
AM	46XY	missense mutation in ZF3	396D-G	none	-	Pelletier *et al.*, 1991b
F	46XY	missense mutation in ZF3	396D-N	none	-	Pelletier *et al.*, 1991b
F, SG	46XY	splice acceptor site mutation in intron 9	KTS alternative splice not inculuded	none	-	Bruening *et al.*, 1992
F	46XY	missense mutation in ZF3	394R-P	unilateral	homozygous	Bruening *et al.*, 1992
AM, SG	46XY	missense mutation in ZF3	394R-W	unilateral	N/D	Bruening *et al.*, 1992
CL	46XX	missense mutation in ZF1	330C-Y	none	-	Bruening *et al.*, 1992
F	46XY	missense mutation in ZF3	394R-W	none	-	Little *et al.*, 1993 Baird *et al.*, 1992b
F	46XX	missense mutation in ZF3	396D-N	bilateral	both homozygous	Little *et al.*, 1993
F	N/D	missense mutation in ZF2	360C-G	unilateral	heterozygous	Little *et al.*, 1993
MP, C, HK	46XY	chain termination mutation in ZF2	truncation in exon 8	bilateral	both homozygous	Little *et al.*, 1993
HP	46XY	missense mutation in ZF2	373H-Q	none	-	Little *et al.*, 1993
GU	46XY	missense mutation in ZF3	394R-W	unilateral	N/D	Baird *et al.*, 1992b
GU	46XY	missense mutation in ZF3	394R-W	unilateral	homozygous	Baird *et al.*, 1992b
NM	46XY	missense mutation in ZF3	396D-N	unilateral	homozygous	Baird *et al.*, 1992b
AM	46XY	missense mutation in ZF3	366R-H	unilateral	N/D	Baird *et al.*, 1992b
AM	46XY	insertion 1bp in exon 6	truncation exon 6	bilateral	homozygous	Baird *et al.*, 1992b
AM	46XY	missense mutation in ZF3	394R-W	none	-	Coppes *et al.*, 1992b
F	46XX	missense mutation in ZF3	394R-W	bilateral	N/D	Coppes *et al.*, 1992b
F, SG	46XY	missense mutation in ZF2	377H-Y	none	-	Coppes *et al.*, 1992b
F, SG	46XY	splice acceptor site mutation in intron 9	KTS alternative splice not included	none	-	Konig *et al.*, 1993
F, SG	46XY	missense mutation in ZF3	394R-W	unilateral	N/D	Sakai *et al.*, 1993
F	46XX	missense mutation in ZF3	394R-W	unilateral	N/D	Sakai *et al.*, 1993
F	46XX	missense mutation in ZF2	355C-Y	none	-	Sakai *et al.*, 1993
F	46XX	missense mutation in ZF3	401H-Y	unilateral	N/D	Baird and Cowell 1993
HP, C	46XY	insertion 1bp in exon 9	truncation in ZF3	none	-	Ogawa *et al.*, 1993b
AM, HP, C	46XY	splice donor site mutation exon 6	Exon skipping deletes exon 6 truncation in exon 7	unilateral	homozygous	Schneider *et al.*, 1993
F, CL	46XY	missense in ZF2	360C-Y	none	-	Clarkson *et al.*, 1993
HP, C	46XY	missense in ZF2	377H-R	unilateral	N/D	Nordenskjold *et al.*, 1994
HP, C	46XY	missense in ZF3	396D-N	unilateral	homozygous	Nordenskjold *et al.*, 1994
HP, C	46XY	missense in ZF3	394R-W	none	-	Nordenskjold *et al.*, 1994
F, GD	46XX	missense in ZF3	396D-N	unilateral	homozygous	Nordenskjold *et al.*, 1994

AM, ambiguous genitalia: C, cryptorchidism: GD gonadal dysgenesis: GU, unspecified genitourinary malformations: HK, horseshoe kidney: CL, clitromegaly: HP, hypospadias: MP: micropenis: NF, female external genitalia: NM, normal male genitalia: SG, streak gonads.
JGCT, juvenile granulosa cell tumour

evidence for the DDS phenotype being caused by a dominant-negative type of mutation, as opposed to a purely dominant one, comes from the observation that truncation of the *WT1* zinc fingers (which leaves the transregulatory and dimerisation domains intact) as well as the zinc finger missense mutations can lead to more severe genitourinary defects (Baird *et al.*, 1992b, Little *et al.*, 1993, Schneider *et al.*, 1993). If *WT1* functions only as an intact dimer, then such mutations might be expected to reduce the functional dimer concentration to 25%. Initially it was thought that DDS and WAGR phenotypes were due to different types of *WT1* mutants. This is not exclusively the case as intermediate phenotypes are now being observed e.g. an XX WAGR case with an 11p12-13 deletion but with nephropathy which is normally characteristic of DDS (Baird *et al.*, 1992b), and other cases lacking the intersex disorders but still showing the diagnostically characteristic glomerular nephropathy (Habib *et al.*, 1985). The same deletion of 11p13 was inherited as an unbalanced translocation in two cousins, the XX case had only aniridia but the XY case had DDS like genitourinary abnormalities (Henry *et al.*, 1993b). Therefore it may be that WAGR and the DDS represent different parts of a spectrum of genitourinary abnormalities dependent on the type of mutation, genetic background and a higher susceptibility in XY cases. Like WAGR, DDS may be incompletely penetrant. Two phenotypically normal cases where Drash-type *WT1* mutations are present constitutionally have been observed, a girl with Wilms' tumour (Akasaka *et al.*, 1993) and the father of a DDS son (Coppes *et al.*, 1992b). In all the other cases where inheritance was looked for parental mutations were not found; phenotypically normal carriers of Drash-type mutations are probably very rare. It is now becoming clear that different mutations in a single gene can give a range of effects on the developmental pathways. This has been found for *PAX6* mutations where a range of eye phenotypes are obtained (Hanson *et al.*, 1994) and for the *ret* oncogene (van Heyningen 1994).

1.9 Involvement in other developmental abnormalities

Since *WT1* is expressed at sites other than the kidney, the possible role of *WT1* in tumorigenesis of gonadal tumours and mesotheliomas has been investigated. One case has been reported of a cystic peritoneal mesothelioma in which a homozygous point mutation was found. It is debatable whether this is a tumour or a developmental abnormality and as yet no other mesotheliomas have been detected with mutations in *WT1*. The mutation found in exon 6 (codon 273) resulted in a serine to glycine substitution and has been shown to change the transrepressor activity of *WT1* into

transactivation activity in *in vitro* systems (Park *et al.*, 1993a). Mutations, such as this, that do not delete or change the sequence of the zinc fingers are so far very rare.

Gonadoblastoma has been observed in a few cases of DDS in which *WT1* point mutations have been found (Pelletier *et al.*, 1991b, Table 1.4). *WT1* mutations are probably not a major cause of these tumours as no mutations have been seen in Frasier's syndrome (nephrotic syndrome and gonadoblastoma) (Poulat *et al.*, 1993), ovarian tumours (Bruening *et al.*, 1993), sex cord-stroma tumours (Coppes *et al.*, 1993b) or other urogenital tumours (Quek *et al.*, 1993). In the ovarian tumours an up-regulation of expression was observed, similar to some Wilms' tumours which have no detectable *WT1* mutations (Bruening *et al.*, 1993).

1.10 The *Wt1* knockout mouse

The ultimate demonstration of the importance of a gene in a developmental event is the failure of this event to occur when the gene is homozygously deleted. Homozygous deletion of the 5' end of the *Wt1* gene leads to the failure of kidney and gonad development and embryonic lethality in mice. Mesonephric differentiation occurs, but 2-3 fold fewer tubules form. Formation of the uninduced metanephric blastema is also observed but no ureteric bud develops so that induction does not happen and the blastema undergoes apoptosis. The initial thickening of the coelomic epithelium at the onset of gonad differentiation appears but again development does not proceed beyond this point. The embryonic lethality is probably not due to the above, but to heart, lung, and diaphragm abnormalities which are also observed, and may be ascribed to a failure of the mesothelium which covers these organs to develop fully (Kreidberg *et al.*, 1993). This indicates that *Wt1* is absolutely required for murine kidney and gonad development and also suggests that the expression in the coelomic epithelium and later in the mesothelium also plays a key role in the normal development of heart, lungs, and diaphragm. The heterozygously deleted mice did not develop WT (Kreidberg *et al.*, 1993), this indicates that there are some distinct differences in the development of mouse and human kidneys.

1.11 Is *WT1* the only gene responsible for predisposition to Wilms' tumour?

1.11.1 Lack of mutations in *WT1*

Although it is clear that the homozygous loss of *WT1* function is the mechanism of tumorigenesis in some Wilms' tumours (summarised in Tables 1.2-1.4), it cannot be

the only mechanism. Only 1-2% of WT patients have visible cytogenetic deletions in 11p13 and only 30% show LOH (Little *et al.,* 1991c). The *WT1* gene was found to be unaltered in the genome as observed by Southern blotting (Cowell *et al.,* 1991, Royer-Pokora *et al.,* 1991), and *WT1* expression was observed in many Wilms' tumours (Pritchard-Jones *et al.*, 1990). Therefore it was expected that the gene would contain small changes resulting in a loss of function of the protein. Many Wilms' tumours have now been examined. Analysis of the literature shows that in only about 10% is there a loss of the functional gene, corresponding to Knudson's two hit model, with many tumours having no changes in *WT1* in either allele (Baird *et al.,* 1992a&b). This may be due to mutations outside of the transcribed or coding regions which may not have been analysed, affecting *WT1* transcription or translation, but this is unlikely for so many cases. The one group that does seem to follow Knudson's model are the DDS cases, except in two cases (Little *et al.,* 1993, K. Williamson personal communication) where the constitutional mutation in *WT1* becomes homozygous in the Wilms' tumour (Table 1.4).

1.11.2 Dominant *WT1* mutations and trans-splicing affects

The heterozygous loss of zinc finger 3 in the tumour AR has been further investigated to determine if this may have a dominant tumorigenic effect over the remaining wild type *WT1* allele (Haber *et al.*, 1990). The deleted *WT1,* but not wild-type *WT1,* can co-operate with adenovirus E1A protein to produce a transformed phenotype (Haber *et al.*, 1992). This demonstrates that mutant *WT1* can act in a dominant way, but not that it is responsible for the development of the Wilms' tumour. Another mechanism is now becoming apparent with the isolation of some *WT1* transcripts lacking exon 2 from Wilms' tumours but not from normal kidney cells. No mutations have been detected in the genomic DNA indicating *WT1* is affected by a trans-splicing mechanism (Haber *et al.,* 1993).

1.11.3 A possible role for imprinting

It has been observed that the paternal allele is preferentially retained in Wilms' tumours, suggesting that tumorigenesis is favoured by the retention of this allele. It was proposed that this could be due to imprinting of *WT1* (Schroeder *et al.*, 1987, Williams *et al.*, 1989, Pal *et al.,*1990). The expression of maternal and paternal alleles in kidney and tumour tissue has been shown, disproving this theory (Haber *et al.*, 1990, Little *et al.*, 1992b). Paternal alleles may be more frequently selected in

tumours because they are the ones which contain the mutation and therefore give the growth advantage. This is because mutations happen at a higher frequency during male meiosis and so will be transmitted through the paternal germline (Charlesworth 1993). Retention of the paternal allele could also be due the presence of a linked imprinted transforming gene which can synergise with *WT1* to produce a transformed phenotype (Wilkins 1988).

1.11.4 A second Wilms' tumour locus on 11p

From loss of heterozygosity studies a second locus on 11p, 11p15.5, has been implicated in the predisposition to Wilms' tumour (Mannens *et al.*, 1988, Henry *et al.*, 1989, Reeve *et al.*, 1989, Wadey *et al.*, 1990). The locus for Beckwith-Wiedemann syndrome has now been mapped to 11p15.5 (Waziri *et al.*, 1983, Koufos *et al.*, 1989). BWS patients are susceptible to several tumours (Table 1.1), although more than 50% are Wilms' tumours, indicating that there maybe a second tumour predisposing locus, *WT2*.

IGF2 has been proposed as a candidate gene for BWS (Little *et al.*, 1991b). It is located at 11p15.5 and has been shown to be imprinted in most tissues where it is expressed including the foetal kidney, lung, brain, muscle, liver, adult kidney and placenta (Kalscheuer *et al.*, 1993, Ohlsson *et al.*, 1993) with only the paternal allele being expressed. There is evidence for the overgrowth observed in BWS arising from an 'overdose' of *IGF2*, which is a growth hormone, either owing to deregulation of imprinting (Weksberg *et al.*, 1993), or paternal disomy (Henry *et al.*, 1991, Henry *et al.*, 1993a). This is consistent with the phenotype observed in mouse as overexpression of *Igf2* gives normally proportioned but differently sized mice (DeChiara *et al.*, 1991). That an 'overdose' of *IGF2* is also involved in Wilms' tumour formation is suggested by the observation of deregulation of *IGF2* imprinting (Ogawa *et al.*, 1993a&c) and LOH (Chao *et al.*, 1993) in Wilms' tumours. There may have to be further genetic changes than simple LOH to obtain a phenotype, as 11p homozygous tissues have also been found to develop with a normal phenotype (Chao *et al.*, 1993).

The loss of heterozygosity at 11p15.5 in Wilms' tumours as opposed to 11p13 suggests that changes at 11p15.5 (*WT2*) can co-operate with loss of one copy of *WT1* in tumorigenesis. Recently a heterozygous *WT1* mutation was detected in a perilobar Wilms' tumour from a patient with BWS stigmata (Park *et al.*, 1993b). This suggests that the first hit occurred at the *WT2* locus and the second hit was in *WT1*. This gives

weight to the suggestion that the *WT1* and *WT2* loci can co-operate in tumour formation.

1.11.5 Not all Wilms' tumour predisposition loci map to chromosome 11

Non random allele loss and frequent cytogenetic rearrangements have been observed on the long arm of chromosome 16 as well as on 11p, and so 16q has been proposed as the location of *WT3* (Coppes *et al.,* 1992a, Maw *et al.*, 1992). So far the most likely candidate gene is the epithelial protein, uvomorulin.

Familial cases of Wilms' tumour are rare, although there are now cases of hereditary *WT1* mutations (Pelletier *et al.,* 1991a), WAGR cases (Fantes *et al.,* 1992, Henry *et al.,* 1993b) and even one DDS mutation (Coppes *et al.,* 1992b). In several families predisposition to WT is found not to segregate with 11p13 or 11p15 markers (Grundy *et al.*, 1988, Huff *et al.*, 1988, Schwartz *et al.*, 1991) and in one study not with chromosome 16 either (Huff *et al.,* 1992).

1.11.6 Progression of Wilms' tumours

Secondary mutational events are probably involved in the progression of Wilms' tumour once *WT1* or other Wilms' tumour predisposition loci have been affected. An obvious candidate is p53 which is affected in many human tumours and may interact with *WT1* (Maheswaran *et al.*, 1993). p53 has been found to be highly overexpressed in Wilms' tumour, particularly in the epithelial elements (Lemoine *et al.,* 1992). In a tumour which was without a *WT1* mutation, a mutation in intron 1 of p53 was found (Velasco *et al.,* 1993). p53 mutations have now been specifically linked to the rare anaplastic type of Wilms' tumours, which are associated with poor prognosis (Bardeesy *et al.,* 1994). Chromosomal abnormalities and allele loss have been found in tumours on 1, 4p, 6, 7, 8q, 12, 14q, 16, 17, and 18 (Wang-Wuu *et al.,* 1990, Maw *et al.,* 1992).

1.12 Evolutionary comparison of the vertebrate genitourinary system

Comparison of the development of the genitourinary systems within the vertebrates can be used to explore the relationships between different organs by examining the way in which they have evolved. The theme that runs through kidney and gonad development is their intimate association with each other and with the

coelomic epithelium. This close evolutionary relationship may explain the expression of *WT1* in all three structures.

1.12.1 The vertebrate kidney

The kidney or holonephros is a characteristic feature of vertebrates. Holonephros is a term which includes all the nephrogenic tissue (Saxen 1987), and also refers to the archetypal vertebrate kidney (Kent 1987). The nephron in its various forms is found in all vertebrates, and nephron-like imprints have been found in fossils of the ostracoderms, the oldest known vertebrates (Torrey 1965), but not in any other sub-phyla of Chordata. The nephrons seem to have developed from specialisation of mesodermal ducts, known as coelomoducts, which often lead from the coelom. Coelomoducts of mesodermal origin do exist outside of Chordata, but do not resemble vertebrate kidneys (Barnes *et al.,* 1993). In all vertebrates the differentiation of the intermediate mesoderm occurs as a wave of differentiation followed by degradation. The final structures formed depend on the extent of these two events (Figure 1.8).

1.12.2 Amniote kidney development

The amniotes (mammals, birds and reptiles) all develop three successive kidneys, the first two being lost during foetal life or early in adulthood (Figure 1.7&1.8). The pronephros is not functional, and whether it actually exists in mammals has been debated (Torrey 1965). This partly arises from the difficulty in defining boundaries between the kidneys, in what is a continuous wave of differentiation (Saxen 1987). The origin of the pronephric duct is not well understood in mammals. In the chick a distinct duct rudiment differentiates from the lateral plate mesoderm before any tubules are seen (P. Lear personal communication), as opposed to being derived from the fusion of the first pronephric tubules as has been proposed for amphibians (Torrey 1965). The pronephros is less transitory in reptiles although it is thought not to be functional (Fox 1977). The mesonephros has been shown to be the functional foetal kidney in the reptilian lineages (Romanoff 1960). Its degeneration occurs later in the reptilian lineages, so that in some species of snakes it is present until after the first moult. This is in contrast to mammals where the mesonephros is vestigial and has disappeared well before birth.

Figure 1.7 Phylogenetic tree of the vertebrates

The tree shows the separation of the fish, then amphibian, then reptilian lineages from the mammalian lineage. The separation of the bird lineage from the reptilian lineage happened after the separation from the mammalian lineage so that birds and some reptiles belong to the same clade. (Tree adapted from Powers 1991)

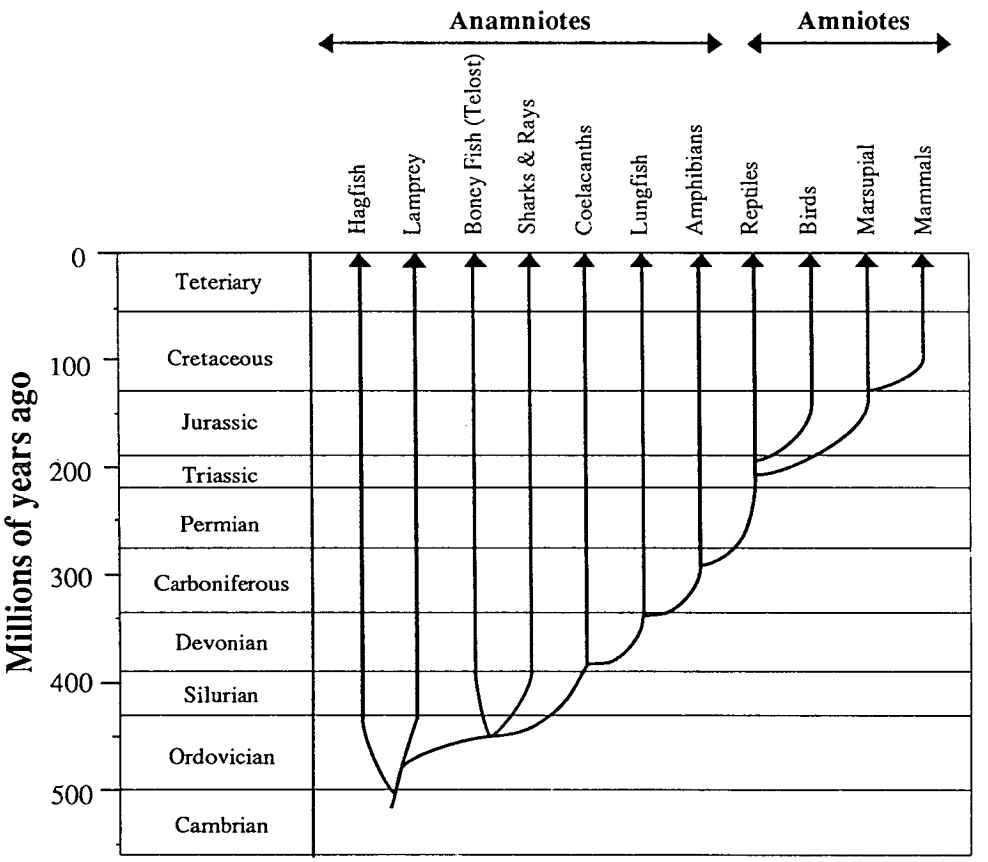
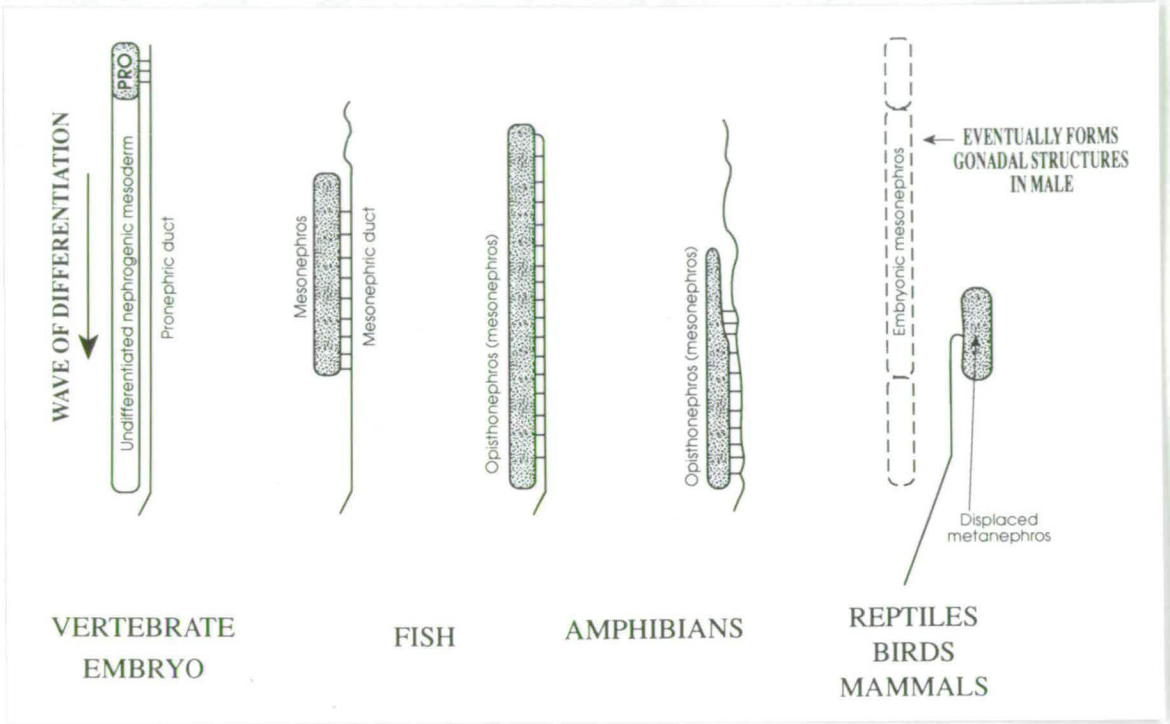


Figure 1.8 The relationship between kidney development and evolution

Vertebrate kidneys are all formed as a result of a wave of differentiation of the urogenital ridge of the intermediate mesoderm followed by a wave of degeneration. In fish and amphibians this can result in the formation of mesonephroi of different extents. In the reptiles, birds and mammals the mesonephros is replaced by the metanephros. (Illustration adapted from Kent 1987)



1.12.3 Anamniote kidney development

In anamniotes (amphibians and fish) the mesonephros is the adult kidney (sometimes called the opisthonephros) as no metanephros differentiates (Figure 1.8). The pronephros is the functional foetal kidney and in some teleosts it continues to function in the adult. Many more of the nephrons, especially the anterior ones, are found to have the more ancient external glomeruli (Figure 1.5b). The capacity for nephron development may not be limited to embryonic life in lower vertebrates (Fox 1977).

1.12.4 The metanephros

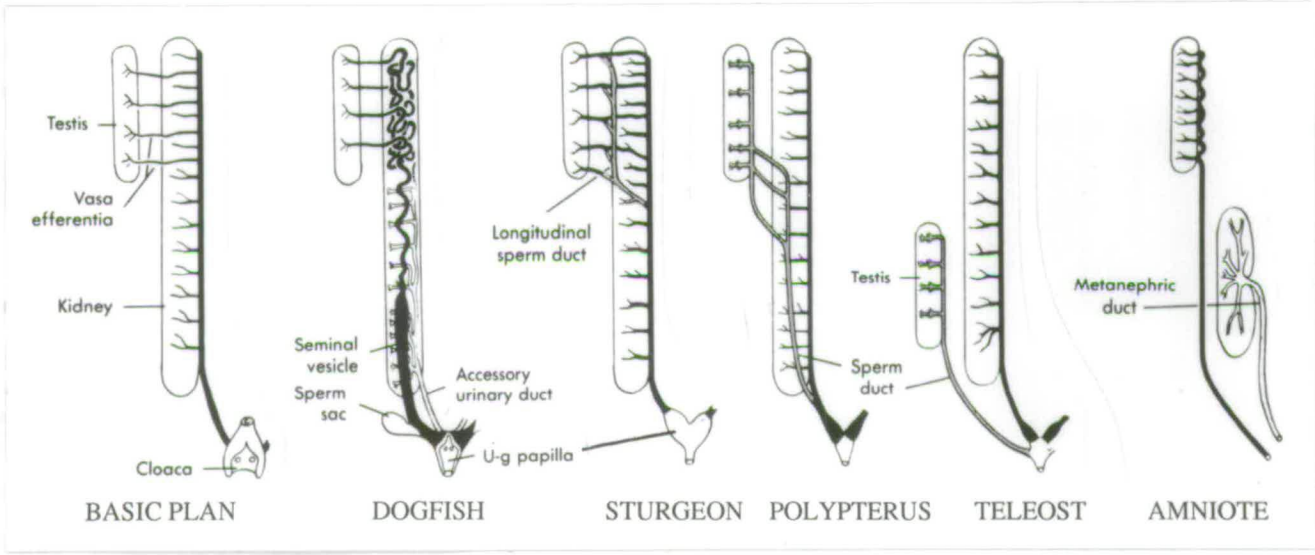
The feature that appears to be the evolutionary novelty that causes the metanephros to form is the ureteric bud. The region of nephrogenic tissue that forms the metanephros in amniotes, is capable of forming opisthonephric tubules when induced by the mesonephric duct in the lamprey, shark and urodele. The mesonephros in mammals contains 30-50 nephrons while the metanephric tubules are present in their millions. This may be a consequence of the mesonephric function becoming of lesser importance, but even in the opisthonephros only a few hundred nephrons are present (Kent 1987). The greater capacity of the metanephros could have been selected for by a greater requirement for fluid conservation with the colonisation of terrestrial habitats. Selection for greater fluid conservation within mammals by the metanephros is exemplified by the almost complete water conservation achieved by very long loops of Henle in desert species such as the kangaroo rat. A greater capacity for excretion may also have been involved in the evolution of homeothermy and the associated increase in metabolic rate and production of toxic wastes.

1.12.5 The gonads and genital ducts

In all vertebrates the gonads develop in very close proximity to the mesonephroi and to the coelom. The mechanism by which gonadal sex is determined is very variable. In fish and amphibia there is the capacity for sex reversal in adult life. In reptiles the mechanism of sex determination can be temperature dependent but can also be influenced genetically. In birds and mammals it is determined solely genetically (Gilbert 1991).

Figure 1.9 The interrelationship between the gonadal and urinary systems

In the earliest gnathostomes the kidney ducts became utilised by the gonads for transport of the sperm to give the basic genitourinary system. In the most primitive of fish this organisation has remained. In other groups of vertebrates sperm and urine transport have become separated by the production of a second duct, either to carry sperm (fish) or urine (amniotes). (Illustration adapted from Kent 1987)



In the most ancient vertebrates, the agnatha, there are no genital ducts and the gametes are discharged into the coelom and exit via abdominal genital pores. In the most primitive gnathostomes the bridge between the testes and kidneys has been made and the urinary duct function has been usurped by the gonads. This situation is still found in some fish e.g. the sturgeon. In most other systems there has been a separation of sperm and urine transport by the formation of a second set of ducts. In the fish and urodeles these carry sperm and are formed as outgrowths from the testes and in amniotes the urine is transported by the ureter, formed from the ureteric bud (Romer and Parsons 1977) (Figure 1.8). In females a second set of ducts is formed in all species from the gnathostomes onwards. In amphibians and elasmobranchii the relationship of the female genitalia to the urinary ducts is closer than in higher vertebrates, as the ostia are formed from the pronephric tubules and the Mullerian ducts from the longitudinal splitting of the pronephric duct, as opposed to a separate origin in the intermediate mesoderm. These ducts become specialised, forming shell glands, ovisacs, and uteri.

1.13 Evolutionary comparison of Wilms'-like tumours

1.13.1 Mammals

Although there is a high degree of similarity of the mouse *Wt1* gene to human *WT1* at both the predicted protein sequence and expression levels, very few nephroblastomas have been reported in the mouse (Liebelt *et al.*, 1989). This is surprising considering the number of mice in laboratories. This observation is supported by the fact that the heterozygous Sey^{Dey} (small eye) mice, which are deleted for the region corresponding to the human WAGR region (Glaser *et al.*, 1990), have an eye phenotype similar to aniridia (Jordan *et al.*, 1992), but there is no evidence of the genitourinary abnormalities or Wilms' tumours. In addition mice heterozygous for the *Wt1* deletion do not develop nephroblastomas (Kreidberg *et al.*, 1993). This is also the case in retinoblastoma, where the mechanism of tumorigenesis differs between mouse and human, as retinoblastomas are not observed in mice heterozygous for a non-functional retinoblastoma gene (Jacks *et al.*, 1992).

Reports of nephroblastomas in other animals tend to be rare and there is often the problem of ascertaining whether it is a true nephroblastoma and so comparable with Wilms' tumour or is derived from a different type of precursor cell (Hard 1984). The rat, has an almost identical *Wt1* gene to the mouse but has been shown to develop nephroblastomas with the characteristic blastemal cells and tubule and glomerular like

structures and certain strains have a higher level of predisposition (Hard and Noble 1981). Stromal and heterologous elements are also observed in mesenchymal tumours. These are not true nephroblastomas but have a histology similar to some human Wilms' tumours (Hard 1984). Both these tumours can be induced by exposure to nitroso compounds (Jasmin and Riopelle 1970). Wilms' like tumours have also been found in a number of eutherian mammals; rabbits, hamsters, pigs, cows, dogs, horses, sea-lion, elk, cats, sheep, goats, macaque and red monkey, and in a metatherian mammal; the opossum. In some species no age dependency for tumour development is observed (dogs), in others it is found at approximately the same time as humans (pigs), in others prenatally (cows) (Hard 1984). *WT1* expression has been observed in pigs and shows a similar down regulation in kidney expression in early life. In addition contra-lateral nephrectomy induced expression in adult pigs may be as a result of compensatory renal growth (Kushner *et al.*, 1992). Nitroso compounds have been found to induce tumours in many of these species, the window for exposure corresponding to the presence of metanephric blastema (Jurgelski *et al.*, 1976).

1.13.2 Birds

In fowl spontaneous nephroblastoma is one of the most common neoplasms, occurring mostly in young birds. In the laboratory nephroblastomas can be induced by infection with viruses, mainly avian myeloblastosis virus (AMV) (Heine *et al.*, 1962). Although virus induced tumours and spontaneous ones are histologically identical no virus has been shown as yet to be associated with spontaneous ones (Heine *et al.*, 1962, Ishiguro *et al.*, 1962). The sites of viral integration have been examined and any changes occurring in RNA expression investigated. The gene that has been implicated is not *WT1* but *novC* (Joliot *et al.*, 1992). In one case a truncated transcript was produced as a result of viral integration, in others expression was seen to be up-regulated in the resulting nephroblastomas. The predicted protein belongs to a cysteine rich secreted family of immediate growth response genes which show similarity to insulin-like growth factor 1 binding protein (*IGF1BP*) (Joliot *et al.*, 1992). The truncated form of the gene has been shown to transform cells in culture but the full length transcript repressed growth. The human homolog *NOVH* has been isolated and maps to 8q24.1 proximal to *c-myc*. This is a region found to be affected in Wilms' tumours with above random frequency (Martinerie *et al.*, 1992). *NOVH* may well have a role in human Wilms' tumour progression but this has yet to be demonstrated.

The nearest relatives to the aves are the crocodilians but there are no reports of nephroblastoma in the reptilian class. While this may be a true reflection of the situation, it is more likely that it is a rare tumour and has not yet been observed (Elkan 1963).

1.13.3 Anamniotes

Nephroblastomas do occur in classes only possessing the pronephros and mesonephros. In *Xenopus laevis* tumours consisting of undifferentiated blastema and tubules have been observed (Elkan 1963), and in the fire bellied newt (*Cynops pyrrhogaster*) primitive glomerular differentiation was seen (Zwart *et al.,* 1970). Most of the nephroblastomas in fish have been observed in the *Salmonidae*. In the rainbow trout (*Salmo gairdneri*) tumour induction is possible by exposure of embryos to nitroso compounds (Hendricks *et al.,* 1980). Recently there has been a report of characteristic Wilms'-type tumours in the Japanese eating eel (*Anguilla japonica*) containing blastemal cells, renal tubules and striated muscle (Masahito *et al.*, 1992).

1.14 Evolutionary comparison as a method for defining gene function

1.14.1 Structural analysis

The rate of change of amino acid sequence can vary by as much as a thousand fold; ranging from the fast evolving fibrinopeptides to the virtually static histone H4 (Nei 1987). Conserved regions between the same gene in different species are thought to arise from functional and structural constraints on the gene and gene product. Originally it was thought that mutations would become fixed or removed from the population owing to their advantageous or disadvantageous phenotype. This would be consistent with Darwin's theory of natural selection. Kimura's neutral theory of molecular evolution (Kimura 1968, Kimura 1983) was developed in order to explain the way in which evolution occurs at the molecular level. This states that most mutations are neutral or nearly neutral and these become fixed in the population due to factors such as genetic drift and that Darwinian selection will only act on a subset of mutations. Some of the best evidence in favour of this is the evolution of third bases of codons in protein coding sequences (Kimura 1977). Because of the degeneracy of the amino acid code many nucleotide substitutions will be neutral or almost so (RNA structure or tRNA abundance can affect this), but these have been shown to be some of the most prevalent sites of nucleotide substitution. Even more rapidly evolving are regions of DNA with no known function e.g. pseudogenes (Li *et*

al., 1981). Therefore it is postulated that, assuming a gene retains a similar function, those regions that have been conserved through evolution are important for its structure and function.

The number of known gene sequences is increasing very rapidly, and so the opportunities for comparison of orthologs (genes related by speciation) and paralogs (genes related by duplication) also increases. In particular the number of transcription factors studied has expanded. Comparative sequence analysis has identified several important amino acid sequence motifs e.g. the heptad repeat of leucines in the leucine zipper (Landshultz *et al.*, 1988), the POU domain (Strum and Herr 1988) and the REL homology domain (Kerr *et al.*, 1993). Also comparison of newly isolated genes containing known motifs may give clues to the characteristics and function of the gene product. In this way many genes are now being grouped into families and superfamilies e.g. *AP1* family (containing the *fos* and *jun* transcription factors) (reviewed by Lamb and McKnight 1991) and the immunoglobulin superfamily. In addition many related sequences are being isolated on the basis of the conserved motifs e.g. zinc fingers and tyrosine kinases.

In the absence of obvious similarities with known genes inter-species comparison can yield important information. In the case of the tumour suppressor gene p53, the cloning of a range of orthologs allowed comparison with the human and mouse genes revealing a distinctive pattern of conserved and non-conserved regions (Soussi *et al.*, 1990). Subsequent analysis of mutant p53 in tumours has shown clustering of nucleotide mutations resulting in amino acid changes in the evolutionarily conserved regions (Levine 1992).

With the isolation of more orthologs and paralogs for comparative analysis the study of chromosome evolution is becoming more feasible. The chromosomes of vertebrates are surprisingly variable, considering how well the actual genes themselves have been conserved. The way in which some of these changes have occurred can be recognised by defining homologous syntenic regions conserved between species. In vertebrate evolution several tetraploidisations have been known to occur and these have been followed by reorganisation to regain diploidy. The way in which this occurred may tell us about chromosome structure and behaviour (Lundin 1993).

1.14.2 Functional analysis

The genes present in organisms can to some extent be grouped by their function into classes such as metabolic, cell cycle control, and developmental specification. The conservation of a gene and its maintenance through evolution will depend on its function. In the case of the metabolic enzymes which fulfil very basic and necessary functions, many are conserved in both prokaryotes and eukaryotes. Glutamate synthetase is one such gene which duplicated 3,500 Myra (Kumada *et al.*, 1993) and *GSI* or *GSII* is found in every kingdom. The regulation of the eukaryotic cell cycle is another highly conserved function, the human homolog of yeast *Schizosaccharomyces pombe cdc2* has been found to be capable of complementing *Schizosaccharomyces pombe* lacking functional *cdc2* (Lee and Nurse 1987). Within the vertebrates cell regulatory and tumorigenesis mechanisms can be conserved, as was recently demonstrated in that zebrafish *c-myc* was capable of co-operating with mammalian *H-ras* to transform mammalian cells (Schreiber-Agus *et al.*, 1993).

The regulation of development has evolved with the advent of Metazoa. In recent years much has been learnt about the genetic control of development of the fruit fly *Drosophila.* One of the major findings is that many of the *Drosophila* developmental genes have homologs in the vertebrates which also have roles in development. This is the case for *Notch* (Bierkamp and Campos-Ortega 1993), the *Wnt* family (Sidow 1992), the *hedgehog* family (Ingham 1994), and the *HOM-C* and *Hox* gene clusters (Holland 1990). Most of the genes seem to be involved in position specification in the embryo as opposed to specifying specific structures. The resemblances between phyla in *Hox* gene expression has lead to the proposal of the 'zootype' stage where the animal body plan is being set up and can be recognised by characteristic gene expression patterns (Slack *et al.*, 1993). In many cases this coincides with the phylotypic stage which was based on morphological similarities.

WT1 is a gene that does not seem to be involved in pattern specification but rather in subsequent differentiation of a predetermined region. The evolution of the vertebrate genitourinary system, which is effectively the modification of a basic plan, is quite well understood. The similarities in nephrogenesis and nephroblastoma throughout the vertebrates would imply that the investigation of the anamniote kidney could be applied to the understanding of the development of the amniote kidney and Wilms' tumour.

1.14.3 Animal models

The mouse is currently the favourite animal model for understanding human development. This has become a much more powerful system since the advent of transgenic techniques. For developmental genes, the disadvantage of viviparity, making embryonic stages not particularly accessible has yet to be overcome. In the study of Wilms' tumour, the mouse has the added disadvantage of a very low level of incidence of this tumour. The more user friendly free-living embryos of anamniotes have been exploited. In particular much information has been derived from *Xenopus laevis* on control of early development, and now new contributions to knowledge are coming from zebrafish (*Brachydanio rerio*) systems which are genetically more manipulable (reviewed in Strahle and Ingham 1992) especially with the advent of transgenic fish. All of these different systems have advantages and disadvantages, their combined use can allow different questions to be asked and the answers to be compared. The most recent example of this are the studies of chick, mouse and zebrafish *hedgehog* paralogs (Echelard *et al.*, 1993, Krauss *et al.*, 1993, Riddle *et al.*, 1993).

1.15 Conclusion and aims of the project

Since its isolation in 1990 much has been learnt about *WT1* but it is far from being fully understood. The specific DNA binding activity and nuclear localisation of *WT1* is consistent with a role as a transcription factor, but true binding sites for both +KTS and -KTS isoforms remain to be defined. The most often observed effect on transcription by *WT1* was transrepression but this is dependent on the number and positioning of binding sites and on cell type. The need for multiple binding sites for repression to occur may be due to *WT1* interacting with itself and it maybe this interaction with which mutant *WT1* interferes. The cell type specificity and the interaction with p53 indicates that transrepression is not an intrinsic property of *WT1* but depends on the proteins with which it interacts. Therefore in different cell types *in vivo* and as cells differentiate the action of *WT1* may change from activator to repressor of genes or *vice versa*. Most of the analyses have been performed by transient transfection assays which result in very high non-physiological expression levels and it has been shown that the results obtained for *WT1* were very sensitive to the systems used so that its *in vivo* actions may be very different from those observed *in vitro*.

The histological and mutational observations described above suggest that loss of normal *WT1* function would affect normal differentiation, specifically disturbing the mesenchymal-epithelial transition of nephrogenesis. Disturbances of this type are thought to cause the nephrogenic rests, which would then be open to neoplastic transformation. From *in vitro* studies of *WT1* function, loss of regulation of growth factor expression and loss of *WT1* autoregulation are candidate mechanisms at the protein level. If the cells cannot pass though the *WT1* expressing stage, i.e. the induced blastema and differentiating nephrons, this would explain the persistent expression of the gene as seen in some tumours. In other cases the cells fail to form nephrons and take other developmental pathways where *WT1* is not expressed.

It seems that some cases of Wilms' tumour probably do involve the loss of both functional copies of *WT1*, but it is also clear that there is at least a second predisposing locus at 11p15.5 which can either act alone or in concert with *WT1* malfunction leading to tumorigenesis. Also there is some evidence for *WT1* acting in a dominant manner; as an oncogene. It is not yet possible to correlate tumour type, *WT1* expression and presence and type of *WT1* mutations. The only correlation that has been found so far is between BWS (and therefore *WT2*) and perilobar rest associated tumours, and between WAGR and DDS (and therefore *WT1*) and intralobar rest associated tumours.

In order to add to the understanding of how *WT1* is acting in development and tumorigenesis it was proposed to isolate *WT1* from as many relevant developmental systems as possible. First, this would produce a structural comparison which can be a powerful tool in the dissection of protein function. Second, the comparison of *WT1* expression patterns would demonstrate if *WT1* has a conserved function in genitourinary development and therefore whether the study of *WT1* in different classes of vertebrates would be relevant to the study of the function of *WT1* in humans.

Chapter 2

Materials and Methods

2.1 Bacterial manipulation

2.1.1 Genotypes

2.1.1.1 Cell strains

JM101 *supE, thi-1, Δ(lac-proAB), [F' traD36, proAB, lacI^{q}ZΔM15]* (Yanish-Perron *et al.,* 1985).

XL1-Blue *recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac,* [*F' proAB, lacI^{q}ZΔM15, Tn10(tet^r)]* (Stratagene).

Y1090 *Δ(lac)U169, Δ(lon)?, araD139, strA, supF, mcrA, trpC22::Tn10 (tet^r), (pMC9;tet^{r}amp^r)* (Young and Davis 1983a).

SURE *e14-(mcrA), Δ(mcrCB-hsdSMR-mrr)171, sbcC, recB, recJ, umuC::Tn5 (kan^r), UVrC, supE44, lac, gyrA96, relA1, thi-1, endA1 [F' proAB, lacI^{q}ZΔM15, Tn10 (tet^r)]* (Stratagene).

2.1.1.2 Bacteriophage

λZAP λ*sbh*Iλ1^o *chi*A131 (T *amp* ColE1 *ori lacZ'* T3 promoter-polycloning site-T7promoter I) *sr*Iλ3^o c*Its*857 *sr*Iλ4^o *nin*5o *sr*Iλ5^o *S*am100.

λZAPII as for λZAP but no *S*am100 mutation (Short *et al.,* 1988).

λgt11 λ*lac*5 Δ*shnd*IIIλ2-3 *sr*Iλ3^o c*Its*857 *sr*Iλ4^o *nin*5 *sr*Iλ5^o *S*am100 (Young and Davis 1983b).

2.1.2 Electro-transformation of *E.coli*

2.1.2.1 Preparation of electro-competent cells

Bacterial cells were streaked out onto an L-agar[1] plate from a -20^{o}C glycerol stock. 10ml of L-broth were inoculated with a single colony and grown overnight in a shaking incubator at 37^{o}C. This was used to inoculate 1 litre of L-broth. Cells were grown until they were in log phase, this was estimated for XL1-Blues or JM101s to be at an OD_{600nm} of approximately 0.8.

Cells were harvested by chilling the culture on ice as rapidly as possible and leaving on ice for 30 minutes. This was spun at 7,000rpm for 15 minutes and the cell pellet resuspended in 1 litre cold sterile dH_2O. This was centrifuged three times more and the cells concentrated down by resuspension in 500ml dH_2O, 20ml 10% glycerol

[1]Recipes for media and solutions are at the end of the methods section.

and 2-3ml 10% glycerol respectively to give an approximate final cell density of $3x10^{10}$cells/ml. Aliquots were frozen in liquid N_2 and stored at -70°C.

2.1.2.2 Electro-transformation

Salt was removed from the DNA to be electroporated by dialysis for 15 minutes on a 0.025μm type VM Millipore filter floating on dH_2O. Competent cells were thawed, and 40μl added to a cold 0.4 cm electroporation cuvette containing 1-5μl ligation mixture or 1ng plasmid DNA and this was allowed to stand on ice for 1 minute. The cells were pulsed on a BRL Gene Pulser™ set at 25μF, 2.5kV and 200Ω. 1ml of cold L-broth was added immediately afterwards and the cells were allowed to recover at 37°C for 1 hour with shaking before 100-150μl was plated on agar containing the appropriate selection.

2.1.3 Selection of cells harbouring recombinant plasmids

All the vectors used in this study carry ampicillin resistance. Therefore cells harbouring plasmids were selected for after transformation by the presence of 100μg/ml ampicillin in the agar. The pBluescript vectors used contain a partial *lacZ* gene and the host cells used contain an F' episome harbouring the rest of the gene. *LacZ* expression can be induced by addition of 40μg/ml IPTG (isopropylthiogalactose) to the medium, the two parts of the gene are then capable of metabolising X-gal (5-bromo-4-chloro-3-indolyl-galactose) which is also added to the medium at 40μg/ml. The product formed is a blue precipitate. If an insert is present in the vector the *lacZ* gene is disrupted and the colonies are white. White colonies were streaked out onto a fresh agar plate.

2.1.4 Bacterial colony lifts

After streaking out, colonies were allowed to grow overnight, the plates were then chilled for at least two hours at 4°C. Gridded nitrocellulose circles (Schleicher & Schuell) were gently placed over the plates and left until the whole circle had absorbed moisture, approximately two minutes. The filters were treated by placing them DNA side up on to Whatmann 3MM paper soaked with denaturing solution for 5 minutes, and then on to paper soaked in neutralising solution for 5 minutes. These were washed briefly (< 1 minute) in 2xSSC. (The cell debris and agar were removed in the 2xSSC wash by rubbing with gloved fingers.) Excess moisture was then

removed with blotting paper. DNA was fixed to the nitrocellulose by baking for 2 hours at 80°C under vacuum. The agar plates were placed at 37°C overnight to allow the colonies to recover.

2.1.5 Small scale preparation of plasmid DNA

2.1.5.1 Method 1

A bacterial culture was prepared by the inoculation of 1-1.5ml of L-Broth or Terrific Broth containing selection with cells picked from a single colony. This was incubated overnight at 37°C in a shaker. The cells were spun down in a microfuge for 1 minute and as much growth medium as possible removed. The pellet was resuspended in 100µl GTE solution, then 200µl of 1%SDS, 0.2mM NaOH were added and the tube inverted several times to mix the solution. This was incubated on ice for 5 minutes, then 150µl 3M5MKAc solution was added, mixed by inversion, and replaced on ice for 5 minutes. The suspension was spun for 5 minutes and the supernatant retained. To this 1 volume of EtOH was added, vortexed and the DNA precipitate spun down for 5 minutes in a microfuge. The pellet was washed in 1ml 70% EtOH and freeze dried. This was resuspended in 20µl TE buffer or sterile dH_2O. This method produced high quality DNA which was suitable for DNA sequencing (Jones and Schofield 1990).

2.1.5.2 Method 2

The above procedure was used for most of the experiments but in the final year the Magic™ Miniprep purification system was produced by Promega. The main advantage was the increased yield, purity, and speed of the system.

Cells were harvested from 1-3ml of overnight culture as above. The cell pellet was resuspended in 200µl of cell resuspension solution and inverted until the solution cleared. To this was added 200µl cell lysis solution, the solutions were mixed by inverting the tube and then 200µl 2.55M KAc was added and mixed as before. This was spun in a microfuge for 5 minutes and the supernatant retained. 1ml of warmed (25°C) resin (containing 7M guanidine HCl) was added to the supernatant. This was pipetted into a 2ml syringe barrel attached to a Magic™ Column. The solution was pushed though the column and the column washed with 2ml column wash solution. The column was dried by spinning for 20s and leaving at 50-60°C for 5 minutes. 50µl of dH_2O at 65-70°C was applied to the column and the DNA was eluted into an

eppendorf by spinning for 20s. Elution was found to be more successful if the dH_2O was allowed to stand on the column for 5-10 minutes before spinning.

2.1.6 Large scale preparation of plasmid DNA

500ml of Terrific broth containing the appropriate selection were inoculated with bacteria from a single colony. This was incubated overnight at 37°C with shaking. The cells were harvested by centrifugation at 6,000rpm for 5 minutes and the cell pellets drained on ice to remove as much of the growth medium as possible. The cells were resuspended in 18ml of GTE and then 2ml of freshly prepared lysozyme solution (10mg/ml in 10mM Tris-HCl pH8.0) was added. To lyse the cells 40ml 1% SDS, 0.2M NaOH was added and mixed by inversion of the centrifuge bottle. After 10 minutes on ice 30ml 3M5MKOAc was added, mixed thoroughly, and replaced on ice for 10 minutes. This was centrifuged at 10,000rpm for 30 minutes and the supernatant filtered through muslin. To this 0.6 volume isopropanol was added and incubated at RT for 10 minutes. The DNA precipitate was then spun down at 8,000rpm for 10 minutes at RT. The pellet was drained, washed with 70% EtOH and freeze dried. The pellet was resuspended in 6ml TE buffer, care being taken to avoid shearing of the high molecular weight chromosomal DNA. The volume of the DNA solution was measured and 1g/ml CsCl was added and the volume re measured (the solution was warmed to 30°C to facilitate dissolution of the salt). To this 0.08 volume 10mg/ml ethidium bromide solution was added. This solution was put into Beckman Quick-Seal^R Bell Top centrifuge tubes and sealed. Balanced tubes were spun at 100,000rpm in a Beckman TL-100 ultracentrifuge for at least 4 hours. Of the two bands visible in the CsCl gradient the lower plasmid band was removed by piercing the tube with a needle and drawing out the band into a syringe while avoiding the upper chromosomal band.

The ethidium bromide was removed by butan-1-ol extraction until the solution was colourless. The volume of DNA solution was kept constant by adding H_2O. To the plasmid DNA, 3 volumes dH_2O and 2 volumes EtOH were added and precipitation of the DNA allowed to occur over 30 minutes at -70°C. The DNA precipitate was spun down at 12,000rpm for 10 minutes, washed with 70% EtOH and freeze dried. The DNA was resuspended in 1mlTE buffer and the concentration estimated by measuring OD_{260nm} in a spectrophotometer. This assumes that an OD_{260nm} of 1 is equivalent to 50ng/ml DNA. The ratio of the optical densities at 260nm and 280nm was used to assess the purity of the DNA, as at 280nm any protein

contamination is detected. For an acceptably pure DNA solution a ratio of approximately 2 was expected.

2.1.7 Glycerol stocks of bacteria

Stocks of bacterial cells were produced by setting up an overnight culture as for a small scale plasmid preparation. The cells were grown overnight and then chilled and a 1:1 ratio of 50% filter sterilised glycerol to culture added. Stocks were then stored at -70°C. When bacteria were required a sterile inoculation loop was dipped into the briefly defrosted stock and used to streak an agar plate containing selection, if appropriate. The plate was incubated at 37°C overnight and then stored at 4°C.

2.1.8 Isolation of clones from bacteriophage libraries

2.1.8.1. Preparation of host cells for bacteriophage

25 ml of L-Broth containing 0.2% maltose (filter sterilised) and 10mM $MgSO_4$ was inoculated with a single colony of bacterial host cells from a fresh agar plate. These were grown overnight with shaking at 30°C. The cells were harvested by centrifuging at 4,000rpm for 10 minutes. The cell pellet was resuspended in 10mM $MgSO_4$ to give an OD_{600nm} of 0.5 for screening or amplification, or 1.0 for *in vivo* excision. Cells were made fresh whenever possible but it was found that infection was not impaired for cells stored at 4°C for up to 2 weeks.

2.1.8.2 Titering of libraries

Serial dilutions of the lambda phage libraries were made in SM buffer. These were added to 200µl of host cells and incubated at 37°C for 15 minutes to allow infection to occur. 3ml of CY top agarose at 43°C was added, mixed quickly, and poured onto prewarmed, predried 9cm plates containing CY select agar. The top agar was allowed to harden for 10 minutes at RT and then incubated at 37°C overnight. The number of plaques produced were then counted and the number of plaque forming units (pfu) calculated for the undiluted library.

2.1.8.3 Screening libraries

24cm x 24cm plates (Nunc) were sterilised by wiping with ethanol and then exposing to ultraviolet light for a minimum of 1hour. These were poured with 150ml CY select agar and allowed to dry. 0.8ml cells were inoculated with an estimated $1x10^5$ phage and incubated at 37°C for 15 minutes. This was then plated as above using 30ml of CY top agarose and incubated overnight. Typically four plates were used so approximately $4x10^5$ phage were screened per experiment.

The plates were chilled at 4°C for at least 2 hours before lifts were taken. Lifts were taken by lowering a 20cm x 20cm Hybond N+ filter (Amersham) onto each plate and leaving for two minutes. The position of the filter on the plate was marked by pricking with a needle attached to a syringe containing ink through the filter into the agar to make an asymmetric pattern. The filter was removed with care in order not to disturb the top agar. Duplicate lifts were made, the second filter was left on the agar twice as long as for the first lift. The filters were then processed by immersion, phage side up in;

denaturing solution for 2 minutes

neutralising solution for 5 minutes

washing solution for < 1 minute (0.2M Tris-HCl pH 7.5, 2xSSC)

After removal from the last solution excess moisture was removed on blotting paper. Any top agar attached to the filters was removed by rubbing with gloved fingers at the denaturation and washing stages. The DNA was fixed to the filters by laying them DNA side up on blotting paper soaked in 0.4M NaOH and then washing them in 5xSSC. The filters were then hybridised in PVC sandwich boxes to an appropriate probe. Positive signals were considered to be those that were present on both duplicates. The filters were lined up with the plates using the needle holes and the positions of positives on the plates identified. These were removed by coring the agar with an inverted yellow tip and put into 1ml SM buffer. These were left overnight at 4°C or at least 2 hours at RT to allow the phage to diffuse into the buffer. The number of pfu was calculated by estimating $1x10^6$ phage per plaque. Serial dilutions were prepared to give 450 and 50 plaques per 9cm plate. These were plated as for titering and screened as before except gridded nitrocellulose circles (Schleicher & Schuell) were used. This meant that excess top agar was removed in the 2xSSC wash and DNA was fixed to the nitrocellulose by baking for 2 hours at 80°C under vacuum. Screening was repeated a third time so that an isolated and therefore pure

plaque could be cored. This plaque was put into 0.5ml SM buffer containing a drop of $CHCl_3$, and the phage allowed to diffuse out as before.

2.1.8.4 Amplification of phage by the plate lysate method

The isolated phage was amplified by infecting 0.3ml of host cells with approximately 50,000 phage and plating in 3ml CY top agarose onto 9cm CY agar plates as above. These were grown for 8 hours so that the plaques were 1-2mm across. The agar was overlaid with 4-5ml SM buffer and the phage allowed to diffuse into the buffer overnight at 4°C. The SM buffer was collected and spun at 8,000rpm for 10 minutes to remove cell debris. To the supernatant 0.3% $CHCl_3$ was added and in addition 7% DMSO was added to 1ml for a -70°C stock.

λZAP and λZAPII were plated out on SURE or XL1-Blue cells as above, and λgt11 on Y1090 cells as above except 50μg/ml ampicillin was included in the medium. The quality of the agar was found to be variable, and L-agar was sometimes substituted for CY(select) agar.

2.1.8.5 *In vivo* excision of λZAP phagemid

λZAP phage have been designed so that a phagemid (plasmid) can be excised from the phage and recirularised *in vivo*. This is due to the inclusion at separate positions in the phage of the sites of initiation and termination of the +strand origin of replication of the f1 bacteriophage which are adjacent to each other in the f1 phage. F1 (helper) phage are co-infected with λZAP, at the site of initiation the phage DNA is nicked and the +strand replicated as far as the terminator sequence and then the new strand is circularised. This is then packaged and secreted. When this is infected into new host cells the second strand is synthesised. The plasmid thus obtained is pBluescriptSK-.

1μl amplified phage isolate and 1μl R408 helper phage (>$1x10^6$pfu/ml Russel *et al.*, 1986) were added to 200μl host cells and 200μl SM buffer. The mixture was incubated for 15 minutes at 37°C and then 5ml Terrific broth was added and this shaken at 37°C for 3 hours. The cells were killed by incubation at 70°C for 15 minutes, then the cell debris were spun down at 6,000rpm for 5 minutes. 10μl or 10μl of a 100x dilution of the phagemid stock was added to 0.2ml host cells and incubated at 37°C for 15 minutes. 1-100μl was then spread on to L-agar plates containing 100μg/ml ampicillin and grown overnight at 37°C. Phagemids were then analysed.

To avoid contamination with the f1 phage, DNA was prepared by small scale plasmid preparation from colonies and transformed into electro-competent cells.

2.1.8.6 Purification of λgt11 DNA using LambdaSorb^{TM} (Promega)

Phage DNA is purified using antibodies to the phage particles which are added to the crude plate lysate. This initial purification eliminates purification steps needed at a later stage in other protocols.

1/100th volume of phage adsorbent was added to the plate lysate and incubated on ice for 30 minutes and then spun at 7,500rpm for 10 minutes. The pellet was resuspended in SM buffer (1/10th the original lysate volume), spun in a microfuge for 1.5 minutes and resuspended in 10mM Tris-HCl pH 7.8, 10mM EDTA (1/20th original lysate volume). This was heated for 15 minutes at 70^{o}C. The EDTA chelates Mg^{2+} ions which are necessary for the integrity of the phage coat, so the phage DNA is released. The phage solution was then spun for 2 minutes down to remove the absorbent and two phenol/chloroform extractions and one chloroform extraction were performed to remove the phage proteins. The DNA was ethanol precipitated with 0.5 volume 5M NH_4Ac and two volumes EtOH and incubation at -20^{o}C for 30 minutes. The DNA was washed with 70% EtOH and resuspended in 20-50μl TE buffer.

2.2 DNA/RNA manipulations

2.2.1 DNA preparation from tissue

The tissue (approximately 1cm^2) was chopped as finely as possible and homogenised. To this 10ml 10mM Tris-HCl pH7.8, 150mM NaCl, 10mM EDTA and then 50μg RNase was added. This was incubated at 37^{o}C for 30-60 minutes. 100μl 20% SDS and 5-10mg proteinase K was added and incubated overnight at 37^{o}C. This was purified by a phenol, a phenol:chloroform, and a chloroform extraction (chloroform contained 1 volume octan-2-ol for every 24 volumes chloroform). The DNA was precipitated by adding 0.5 volume 7.5M NH_4Ac and 2 volumes of EtOH. The DNA was spooled out onto a sealed Pasteur pipette and air dried and resuspended in 8ml TE. The DNA concentration was measured as for plasmid DNA.

2.2.2 RNA preparation from tissue or cell lines

Tissue was dissected as quickly as possible in sterile conditions in PBS (phosphate buffered saline). Cell pellets from cell lines were produced by trypsinisation of the cell cultures (to release attached cells if necessary) and spinning down the resulting cell suspension and then washing the pellet.

The sample was added to 0.6ml 6M Urea, 3M LiCl in an eppendorf tube (approx. $5x10^6$ cells) and homogenised on ice using a syringe and progressively smaller gauges of needle as far as 25. The homogenate was sonicated using an amplitude of 22μ for 3x20s on ice and then incubated overnight at 4°C. The precipitated RNA was spun down for 15 minutes in a microfuge at 4°C. The pellet was resuspended in 0.5ml LiCl/Urea and incubated 2-3hours at 4°C and spun as before. The pellet was resuspended in 0.4ml 10mM Tris-HCl pH8.0, 0.5% SDS, 20mg proteinase K added and incubated at 37°C for 30 minutes. The protein was removed by a phenol, a chloroform (containing 1/25 octan-2-ol) and then two ether extractions. The RNA was precipitated with 2 volumes EtOH and 1/10th volume 3M NaAc pH 5.2. After overnight incubation at -20°C the RNA was spun down and resuspended in DEPC dH_2O. The OD_{260nm} was measured and the concentration was then calculated assuming OD_{260nm} of 1 is equivalent to an RNA concentration of 40μg/ml. The OD_{280nm} was also measured to assess the purity of the sample. A ratio of OD_{260nm}/OD_{280nm} of approximately 2 is expected for a acceptably pure sample.

All solutions to come into contact with the RNA were either treated by adding 1/1000 volume DEPC (diethylpyrocarbonate), leaving for at least 2 hours and then autoclaving, or were made up using DEPC treated dH_2O (Tris and urea containing solutions).

2.2.3 Agarose gel electrophoresis

2.2.3.1 DNA electrophoresis

DNA fragments were resolved on agarose gels, the agarose concentration being optimised for maximum resolution at the expected sizes of fragments. For concentrations 0.7-1.2g/l agarose was used but for higher concentrations a mixture of agarose and NuSieve GTG low melting point agarose.

For analytical gels the running buffer used was 1xTBE. Routine analysis was performed using a BRL Horizon 58^{TM} minigel system at 100V, or 60V for low melting point agarose.

For preparative gels and genomic gels, the running buffer was 1xTAE. Preparative gels also used the BRL Horizon 58^{TM}minigel system and genomic gels on maxi (in house made) gel tanks. Preparative gels were run at 60V and genomic gels at 35V overnight.

6x glycerol loading buffer was added to samples before loading. DNA standards were loaded either, λDNA/HindIII fragments, ΦX174/HaeIII fragments, or 1kb DNA ladder (Promega). The DNA was visualised by adding one drop of ethidium bromide to agarose solution before the gel was cast. The DNA-ethidium complex then fluoresced under UV light.

2.2.3.2 RNA electrophoresis

Total RNA was resolved by electrophoresis in 1% agarose gels containing 2.2M formaldehyde. Either midi or maxi gel tanks were used. Before loading the RNA was mixed in a 1:1 ratio with formaldehyde sample buffer (1 volume 10xgel running buffer, 2 volumes formamide, 1.2 volumes formaldehyde) and heated to 65^{o}C for 5-10 minutes and the chilled on ice to denature any secondary structure. To this 1 volume of formamide loading dye was added and 1μl 10% ethidium bromide. 0.24-9.5kb RNA markers (Gibco BRL) were treated in the same way and loaded as size standards. Gels were run in 1xgel running buffer initially for 10 minutes at 100V and then overnight at 45V.

2.2.4 Vacuum transfer of DNA and RNA

Vacuum transfer of DNA and RNA on to nylon filters was performed using the LKB vacuum blotting apparatus. The apparatus was set up according to the manufacture's instruction. A plastic mask was placed over the porous support with a window cut out approx. 5mm smaller than the gel. The filter was cut slightly smaller than this window and carefully placed within it. The gel was placed over the filter so that no gaps between the gel and mask are present and no air bubbles were trapped between the filter and gel. The top frame was screwed down and vacuum pump started and the tightness of the gel-mask seal checked.

DNA was depurinated by the addition of 0.25M HCl to just cover the surface of the gel. After 4 minutes this was replaced with denaturing solution. After a further 3 minutes the denaturing solution was removed and replaced with alkaline transfer buffer. RNA was transferred using 10xSSC. Depending on the size of the gel, transfer took from 35-60 minutes. That the transfer was successful was verified by restaining the gel with ethidium bromide and inspecting the filter under UV light. The filters used were Hybond N or N+ (Amersham). In the case of the positively charged nylon the DNA was already fixed to the filter by alkali, otherwise the RNA or DNA was fixed using a Stratalinker (Stratagene) which gives a fixed UV exposure.

2.2.5 Hybridisation

Filters onto which DNA or RNA had been fixed were wet with 2xSSC. Filters, separated by a nylon mesh, were rolled up and uncurled in Hybaid hybridisation bottles containing 2xSSC, leaving as few air bubbles as possible. The DNA was on the side not contacting the glass. 2xSSC was replaced with Church and Gilbert's buffer, prewarmed to the hybridisation temperature. Bottles were put into a Hybaid rotisserie oven and filters were prehybridised for at least 1 hour. The probe was incubated at 100°C for 10 minutes and immediately added to the bottles. Approximately $5x10^5$ counts/ml hybridisation mix were used when hybridising to cloned DNA or PCR products, and $1x10^6$ counts/ml hybridisation mix for genomic DNA[2]. Library primary screen filters were hybridised in sandwich boxes in Perspex containers. These were treated in the same way as the bottles except that they were incubated in a shaking oven. After overnight hybridisation the solution was removed and the filters washed with the appropriate concentrations of SSC solution containing 0.1%SDS. Following washing, filters were briefly dried on blotting paper, then wrapped in Saran wrap.

Hybridisation conditions were determined by whether high or low stringency was required, depending on the level of identity of the probe to the target DNA. High stringency hybridisation and washes were at 68°C, with 0.5xSSC and 0.1xSSC washes. Reduced stringency was variable, in general hybridisation and washing took place at 58°C with washes of 3xSSC and 1xSSC, unless indicated in the text.

---

[2]Counts are in dpm (degenerations per minute) and were calculated using the formula:
Counts in dpm = μCi ^{32}P added to labelling reaction x % incorporation of label x $2.2x10^4$

2.2.6 Purification of DNA

Purification of DNA away from modification enzymes (and unused primers and nucleotides in the case of PCR) was performed either by phenol:chloroform extraction or GENECLEAN® II (Bio101).

2.2.6.1 Phenol:chloroform extraction

The reaction was made up to at least 100µl by addition of dH_2O. To this an equal volume of water saturated phenol was added. The phases were mixed thoroughly by vortexing and then separated by spinning in a microfuge for 1 minute. The top phase was then removed, avoiding any protein at the interface, and retained. This was repeated and then again with a chloroform:octan-2-ol (24:1) mixture. The DNA was then ethanol precipitated.

2.2.6.2 GENECLEAN® II (Bio 101)

To the reaction mixture 3 volumes 6M NaI were added and then 5µl glassmilk (suspension of silica matrix in H_2O). If more than 5µg of DNA was being purified an extra 1µl glassmilk per 0.5µg extra DNA was added. This was left on ice for 5-10 minutes. The glassmilk was spun down for 5s in a microfuge and the NaI solution removed, this was repeated to remove any traces of NaI. The glassmilk was resuspended in 200µl NEW wash, pelleted by a 10s spin and the supernatant removed. This was repeated twice but for the final wash 400µl NEW wash was used. Any traces of ethanol from the NEW wash were removed by briefly (1 minute) freeze drying. To elute the DNA the pellet was resuspended in 25µl of H_2O at 55°C and incubated at 55°C for 5 minutes. The glassmilk was spun down for 1 minute and the supernatant removed to another eppendorf. This was repeated and the two elutes combined. This was spun briefly (30s) before use to make sure that any glassmilk accidentally carried over was pelleted.

2.2.6.3 Purification of a specific DNA fragment

If a specific fragment was desired a preparative agarose gel was run and the desired band excised with a scalpel. The bands were purified away from the agarose by either GENECLEAN® or Spin-X.

2.2.6.3.1. GENECLEAN® II (Bio 101)

The gel slice was melted at 55°C in 3 volumes NaI and purified as above.

2.2.6.3.2. Spin-X

The agarose slice was placed in the insert of a Spin-X tube (Costar). On spinning in a microfuge the DNA moved though the 0.22μm cellulose acetate filter at the bottom of the insert leaving the agarose behind. The tube was spun twice for 15 minutes, the agarose was redispersed and about 50μl TE buffer added between spins. The DNA in the elute was precipitated by adding 1/10th volume 3M NaAc pH5.2 and 2 volumes EtOH. This was incubated at -20°C for at least 30 minutes, the precipitate was spun down, dried, and resuspended in 20μl.

2.2.6.4 "Double GENECLEAN®" protocol for cloning PCR products

This is a strategy for cloning PCR products which involves firstly GENECLEANing directly or after purification on an agarose gel, then modification of the PCR ends and a second GENECLEAN. The whole PCR reaction was GENECLEANed using 25μl glassmilk and eluting in 20μl dH_2O. The second GENECLEAN used 10μl glassmilk and eluted in 25μl dH_2O.

2.2.7 Quantification of DNA fragments

The amount of a DNA fragment isolated was quantified by running an analytical agarose gel with 1-2μl of the DNA solution. In parallel, a range of known amounts of size markers were run at approximately the same concentration as the expected DNA concentration. The amount of DNA in the sample was estimated by comparison with the brightness of a size marker band of approximately the same size.

2.2.8 Enzyme modification of DNA and RNA

2.2.8.1 Restriction digest of plasmid DNA

Typically 1μg of DNA was digested per 10μl reaction volume with 1 unit of restriction enzyme (Boehringer Mannheim), in the restriction buffer recommended by the manufacturer and at the appropriate temperature. For most enzymes the appropriate temperature was 37°C. If the plasmid DNA had been prepared by the small scale plasmid preparation method 4mM spermidine was included in the reaction and 3μg RNase A (DNase free). The cutting of DNA was analysed using agarose gel electrophoresis.

For analytical digestion approximately 0.5-1μg DNA was digested for about 1 hour.

For preparative digestion approximately 10-20μg DNA was digested overnight. After digestion the enzyme was heat inactivated if appropriate.

2.2.8.2 Restriction digests of genomic DNA

Digests were set up containing 10μg mammalian DNA or the genomic equivalent as the size of the genome varied for the classes used.

10μg mammalian DNA or the genomic equivalent,

5mM spermidine,

1x restriction enzyme buffer,

3U restriction enzyme per 1μg DNA,

dH_2O to 50μl.

To ensure even digestion of the DNA, the reactions were set up minus the enzyme and left at 4°C for 1-2 hours occasionally stirring with a yellow pipette tip. The enzyme was then added, mixed thoroughly and the reactions left for another hour at 4°C and then at 37°C overnight. To check that the digestion was complete 0.5μg was run on a minigel before loading the rest of the sample on a maxi gel.

2.2.8.3 Production of radiolabelled DNA probes

2.2.8.3.1 Random primed labelling

The Boehringer Mannheim Random primed DNA labelling kit was used, utilising a technique developed by Feinberg and Vogelstein (1983). DNA fragments to be labelled were produced by restriction digest and purified using agarose gel electrophoresis and Spin-X.

25-30ng DNA in 9μl dH_2O was boiled for 5 minutes and immediately chilled on ice. To this 1μl 0.5mM dATP, dGTP, dTTP, 3μl 10xreaction buffer (containing 100μl hexanucleotide mixture), 5μl [α-^{32}P]dCTP (3000Ci/mmol, Amersham) and 2U Klenow enzyme was added. This was incubated overnight at RT or for 1 hour at 37°C. Incorporation of the label was assessed by precipitation of 0.5μl reaction mix on a GF/B Whatmann filter with approximately 3x15ml 5% TCA, where incorporation of 30% or more had occurred the labelled DNA was used as a probe.

Unincorporated nucleotides were removed by purification on a nick translation column (Pharmacia) containing Sephadex® G-50. The DNA eluted in 400µl TE buffer and the unincorporated nucleotides were retained on the column.

2.2.8.3.2 Specific primed labelling

This was used for labelling of the genomic zebrafish probe used for screening cDNA libraries. Random priming was not used due to the small size of the probe, instead the PCR primer C229 which was at the 5' end of the probe was used to prime the extension using Taq polymerase.

25ng DNA was denatured by boiling for 5 minutes in 6.8µl dH_2O and then chilled on ice. To this 3.2µl 10xTaq polymerase buffer (Promega), 1µl 0.5mM each of dATP, dGTP, dTTP, 1µl 0.1mg/ml primer, 5µl [α-^{32}P]dCTP, 1µl Taq polymerase (ABI) was added. This was then subjected to three rounds of; 72°C 25 minutes, 94°C 2 minutes, 55°C 1 minute, and a final extension of 72°C 30 minutes, using a Hybaid Omnigene thermal cycler. The incorporation of radiolabel was checked and the probe purified as for random primed probes.

2.2.8.4 Preparation of riboprobes

This was performed using the Boehringer Mannheim DIG RNA labelling kit, utilising the T7 and T3 promoter sites at either side of the cloning site.

The transcription reaction was set up by adding,

1µg of linearised DNA purified by phenol:chloroform extraction,

4µl 10x transcription buffer,

4µl 10mM dATP, dCTP, dGTP, 6.5mM dUTP, 3.4mM DIG-UTP,

1µl 1M DTT,

2µl RNase inhibitor,

2µl T7 or T3 RNA polymerase,

redistilled dH_2O to give total volume of 40µl.

This was incubated for 1 hour at 37°C whereupon a second dose of enzyme was added and the incubation repeated. To remove the DNA template, 2µl (20U) DNaseI and 1µl 10µg/µl tRNA was added and incubated at 37°C for 15 minutes. The reaction was stopped with 4µl 0.2mM EDTA pH8.0 and the RNA precipitated by adding 5µl 4M LiCl and 150µl EtOH. This was left overnight at -20°C, then spun at

4°C for 30 minutes. The RNA pellet was washed three times with 70% EtOH at -20°C and freeze dried. The RNA was resuspended in 20µl DEPC dH_2O and 1µl RNase inhibitor added.

The amount of riboprobe produced was assessed by comparison with the labelled RNA supplied with the kit, using the Boehringer Mannheim DIG Nucleic Acid Detection kit. Serial dilutions of the probe and standard were dot blotted onto Hybond N and fixed by UV exposure. The filters were then washed in buffer 1 for 1 minute at RT on a shaker, then in buffer 2 for 30 minutes. This was followed by another brief wash in buffer 1 and then filters were soaked in buffer 1 containing 1/5000 dilution of the anti-DIG-alkaline phosphatase conjugate for 30 minutes. The unbound antibody was removed by washing twice for 15 minutes in buffer 1, then equilibrated in buffer 3 for 2 minutes. This was replaced with buffer 3 containing 4.5µl NBT solution and 3.5µl BCIP (X-phosphate) and incubated in the dark until the dots became visible (about 20 minutes). The reaction was stopped by washing in buffer 4. In general a yield between 0.1-1µl/µg of RNA was obtained The probe was denatured prior to use in 20µl DEPC dH_2O at 80°C for 5 minutes and cooled on ice.

2.2.8.5 cDNA synthesis

This was performed using the Boehringer Mannheim cDNA synthesis kit (Gubler and Hoffman, 1983). The first DNA strand was synthesised by adding,

2µl buffer I,

0.5µl (12.5U) RNase inhibitor,

1µl 40mM dNTP's,

1µl 10ng/ml random hexamers,

1µl (20U) AMV-Reverse Transcriptase,

2µg of total RNA,

redistilled dH_2O to a total volume 10µl.

Incubation was at 42°C for 1 hour.

The second strand was synthesised by adding,

20µl buffer II,

0.5µl RNaseH,

2.5µl (12.5U) *E.coli* DNA polymerase I,

7µl redistilled dH_2O,

and incubating at 16°C for 1 hour, RT for 1 hour and 65°C for 10 minutes.

2.2.8.6 Modification of ragged ends of PCR products

For blunt ended cloning the non template directed extra base added by the Taq polymerase to the 3' end of the PCR product was chewed back. The reaction was set up as follows:

One half of a GENECLEANed PCR reaction,

10µl 10x *E.coli* DNA polymerase I buffer,

1µl 100mM rATP,

10U T4 polynucleotide kinase,

10U *E.coli* DNA polymerase I (BRL),

total volume of 100µl.

This was incubated for 1 hour at 37°C and the reaction stopped by the addition of 1µl 0.5M EDTA pH8.0.

2.2.8.7 Ligation of DNA fragments into vectors

Initially the pBluescript vector, pSKII+ (Stratagene) was used. This is a derivative of pUC19. It has a high copy number in *E.coli.* and contains a multiple cloning site (MCS) with unique restriction sites and the T7 and T3 promoters. It also has a partial *lacZ* gene and so blue /white selection to detect the presence of an insert can be used.

2.2.8.7.1 Cohesive end ligation

The DNA fragment and the vector were digested with the appropriate enzymes and gel purified. The vector and the insert were mixed at a molar ratio of about 3:1 (about 100ng DNA) in a volume of 7µl and warmed for 5 minutes at 45°C and then chilled on ice. 1µl of 10x ligation buffer, 1µl rATP, 1µl (1U) T4 ligase were added and incubated overnight at 16°C.

2.2.8.7.2 Blunt ended ligation of PCR products

pSKII+ was cut with EcoRV in the MCS to produce blunt ends. The PCR product was purified using the "Double GENECLEAN®" protocol, ragged ends were chewed back to produce blunt ends. 40ng of vector and insert were added to 1µl of 10x ligation buffer, 1µl rATP, 1µl (1U) T4 ligase and incubated at least overnight at 16°C.

2.2.8.7.3 Turbo cloning of PCR products

This is a very fast and non size selective method for cloning (Boyd 1993), and is explained in Chapter 4. pBS:lox is a manipulated version of pBluescribe (pBS) (Stratagene). A recognition site (lox site) for the Cre enzyme has been introduced at the PvuII site 5' to the MCS.

The ends of the GENECLEANed PCR products were made flush (section 2.2.8.6) and the modifying enzymes were inactivated by treatment at 75°C for 5 minutes, and the DNA was desalted by dialysis on a Millipore filter (section 2..1.2.2). The ligation mixture was set up;

7µl (10ng) PCR product,

1µl (100ng) SmaI cut and dephosphorylated pBS:lox (gift A.C. Boyd),

1.6µl 10x buffer,

6µl 40%(w/v) PEG 6000,

0.4µl(0.4U) T4 ligase.

This was incubated at 20°C for 15 minutes, the ligase was then inactivated by 10 minutes at 65°C. 64µl M restriction enzyme buffer (Boehringer Mannheim) and 0.5µl Cre enzyme (NEN) was added. This was incubated at 30°C for 30 minutes. The enzyme was heat inactivated as above.

2.3 Sequencing DNA

The method used is based on the original method of Sanger *et al.,* (1977) which involved extension of a primer with a polymerase. The DNA produced is labelled by incorporation of [α-^{35}S]dATP and reactions terminated by the incorporation of dideoxynucleotides. The Sequenase® II kit produced by USB was used which includes an engineered version of the T7 polymerase, Sequenase® enzyme. This increases the processivity of the enzyme and therefore increases the readability of

sequence obtained. The primers used were pBluescript primers or those originally designed for PCR.

2.3.1 Double stranded sequencing of plasmid DNA

1-5µg of plasmid DNA in 16µl dH_2O was denatured by adding 4µl 1M NaOH, 1mM EDTA and incubating at 65°C for 5 minutes before chilling on ice. The DNA was precipitated by adding 2µl 2M NH_4Ac pH4.5, 60µl cold EtOH and leaving on dry ice for 10 minutes. The DNA was pelleted at 4°C for 10 minutes and washed in 1 volume 80% EtOH, then dried.

The annealing mix was made up;

2µl 5xSequenase® buffer,

1µl (100µg/ml) sequencing primer,

7µl DNA pellet resuspended in dH_2O

This was heated to 65°C for 2 minutes and then at 37°C for 15 minutes.

The template mix was assembled;

1µl 0.1mM DTT,

2µl 1/5 dilution dGTP labelling mix,

0.5µl [α-^{35}S]dATP (<1000Ci/mmol, NEN),

1µl 1/8th dilution of Sequenase® in enzyme dilution buffer.

The template mix was added to the annealing mix, incubated at RT for 4 minutes and 3.4µl was added to 2.5µl of each termination mix (prewarmed at 37°C). This was incubated for 4 minutes at 37°C, the extension was stopped with 4µl formamide loading buffer. About 3µl of each termination reaction was loaded per lane of a sequencing gel.

2.3.2 Direct sequencing of PCR products

This is an adaption of the Sequenase® method for plasmid DNA sequencing, and involves the use of DMSO to retain the PCR template in a denatured state (Winship 1989).

The annealing mix was made up;

approximately 300ng of gel purified PCR fragment in 6µl dH_2O,

1µl 0.5µM primer,

2µl 5xSequenase® buffer,

1µl DMSO.

This was denatured at 100°C for 3 minutes and snap frozen on dry ice/EtOH.

The labelling mix was assembled;

1µl 0.1M DTT,

2µl 1/15th dilution dGTP labelling mix,

0.5µl [α-^{35}S]dATP,

1.5µl Sequenase® .

This was added to the annealing mix as it was thawing. This reaction mix was incubated at RT for 45s and then 3µl added to 2µl of each termination mix (containing 10% DMSO and prewarmed to 37°C). These were incubated at 37°C for 5 minutes and the reactions stopped by adding 4µl formamide dye. 4µl of each termination reaction was loaded per lane of a sequencing gel.

2.3.3 Polyacrylamide gel electrophoresis (PAGE) of sequencing reactions

Prior to using, the glass sequencing plates were washed with soap, dried and one plate coated in dichlorodimethylsilane, to prevent the gel adhering to this plate. The plates were taped together along the sides separated by 0.4cm spacers and gripped together by bulldog clips. The Sequagel system (National Diagnostics) was used for making up acrylamide. Just prior to use the appropriate amounts of concentrate, dilution buffer and running buffer were mixed to give the desired acrylamide concentration. To this was added 1/10 volume 10% ammonium persulphate and 50µl TEMED (N,N,N'N'-tetramethylethylenediamine), this was quickly poured between the plates while the mixture remained liquid. 0.5-0.8% acrylamide was used depending on the resolution required (Sambrock *et al.*, 1989).

Gels were run in 1xTBE using a limiting factor of 30W, the current and voltage being approximately 24mA and 1250V. Gels were typically run between 1.5-5 hours, using bromophenol blue and xylene cyanol dye fronts to estimate the positions of the fragments. A metal plate, clipped to the front of the plates, acted as a heat distributor to avoid 'smiling' of the gel. After the run was completed the gel plates were

separated and the gel fixed in 10% acetic acid, 10% methanol for 15-30 minutes. The gel was then transferred onto 3MM Whatman paper, covered with Saren wrap and dried under vacuum for 1 hour.

2.4 Detection of radioactive signal

2.4.1 Autoradiography

This was the most frequently used method for detecting radioactive signal. Filters hybridised to [α-^{32}P] labelled probes were exposed to Kodak XAR-5 film at -70^{o}C with an intensifier screen. In the first instance exposure was overnight except for PCR fragments which were exposed for 1-2 hours. Spots of Glow Juice (IBI) at the edge of the filters were used to line up the filters with the film.

The DNA sequence 'ladders' were visualised by exposure to Kodak XAR-5 film at RT, in the first instance overnight or longer if the signal was faint.

2.4.2 PhosphorImager analysis

This was used towards the end of the project, particularly for genomic DNA and Northern blots. The filter or sequencing gel was placed in a PhosphorImager cassette, and depending on the signal, left for an appropriate amount of time at RT. (1-2 hours for sequencing gels or overnight for blots). The PhosphorImage screen was then analysed by the PhosphorImager using a laser and an image produced and analysed using the ImageQuant^{TM} software (Molecular Dynamics).

2.5 Incubation and analysis of embryos

2.5.1 Rearing of *Xenopus* embryos

Xenopus eggs were obtained from Professor P.M. Gaze, Department of Zoology, University of Edinburgh. The embryos were kept in clean *Xenopus* rearing solution and fed on Nettle Powder (Philip Harris Scientific) and Complan (Boots). In order to stop the embryos swimming so that they could be staged (Niewkoop and Faber 1967), the tadpoles were anaesthetised by placing them in a very dilute solution of MS222. Prior to dissection the tadpoles were placed in a more concentrated solution.

2.5.2 Whole mount *in situ* hybridisation

This protocol is an adaptation by Lesley McInnis of that described by Wilkinson (1992).

2.5.2.1 Incubation and dissection of embryos

Hen eggs were supplied by the AFRC unit at Roslin, East Lothian. These were incubated at 37°C and turned every 24 hours for the first 3 days. Embryos were dissected at RT in sterile PBS. As many extra-embryonic membranes as possible were removed to avoid trapping of the probe/antibody. They were then placed on ice in PBS pH7.2 in Sterilin conical tubes. Embryos were staged according to the tables of normal development (Hamburger and Hamilton 1951).

2.5.2.2 Preparation of embryos

The embryos were treated as follows (all washes were for 10 minutes in 5-10ml);

1) Washed embryos twice in ice cold PBS.

2) Fixed embryos for at least 2 hours in freshly made 4% paraformaldehyde (PFA) in PBS at 4°C.

3) Washed, twice in PBT (PBS+0.1%Tween 20).

4) Dehydrated on ice through 25%, 50%, 75% and 100% (twice) methanol in PBS.

5) The embryos were stored at -20°C until required.

6) Rehydration on ice through the same percentages as above.

7) Washed three times in PBT.

8) Bleached with 6%H_2O_2 in PBT for 1 hour.

9) Washed three times in PBT.

10) Digested with 10µg/ml proteinase K in PBT for 15 minutes.

11) Washed in 2mg/ml glycine in PBT.

12) Washed twice in PBT.

13) Refixed in fresh 0.2% gluteraldehyde/4%PFA in PBT for 20 minutes.

14) Washed twice in PBT.

2.5.2.3 Hybridisation

Hybridisation of the riboprobe was performed as follows (in a volume of 1ml);

1) Washed in 50% hybridisation buffer:50% PBT at RT.

2) Washed in hybridisation buffer at RT.

3) Prehybridisation in hybridisation buffer containing 100µg/ml tRNA and 100µg/ml denatured salmon sperm DNA at 70°C for 1 hour.

4) Replaced solution with the same containing 500ng denatured riboprobe.

5) Tubes humidified by placing a 0.5ml eppendorf tube containing hybridisation buffer into the neck of the tube.

6) Hybridisation overnight at 70°C.

7) Washed twice in hybridisation buffer at 70°C for 10 minutes.

8) Washed twice for 5 minutes, and three times for 30 minutes at 65°C in post hybridisation wash.

9) Allowed embryos to cool to RT.

10) Washed in 1xTBST three times.

2.5.2.4 Antibody conjugate binding and staining

1) Diluted anti-DIG alkaline phosphatase conjugate 1 in 500 in1xTBST,1% heat inactivated sheep serum (Scottish Laboratories Products) and added 6mg/ml heat inactivated embryo powder from the appropriate species.

2) Preadsorbed antibody 1 hour at 4°C.

3) Blocked embryos with 1xTBST, 10% heat inactivated sheep serum for 1 hour at RT.

4) Replaced blocking solution the same solution containing 1 in 2000 dilution of preadsorbed antibody. Allowed antibody to bind overnight at 4°C.

5) Washed 3 times for 5 minutes and 3-5 times 30-60 minutes in1xTBST.

6) Equilibrated embryos in fresh alkaline phosphatase buffer by washing 3 times for 10 minutes

7) Replaced solution with alkaline phosphatase buffer containing 4.5µl NBT and 3.5µl BCIP in glass vials.

8) Allowed stain to develop in the dark and examined at regular intervals until a stain appeared.

9) Stopped reaction by rinsing in PBT containing 0.2M EDTA

10) Refixed the embryos in 4%PFA.

2.5.2.5 Photography of embryos

Stained embryos were photographed using a Wild Heerbrugg Photomakroskop M400 and a Wild Leitz WPS Photoautomat using 64 ASA colour film (Fuji).

2.6 Polymerase chain reaction (PCR)

2.6.1 Synthesis and purification of oligonucleotides

PCR primers were synthesised by either D. Chambers or A. Gallacher on an Applied Biosystems 381 A synthesiser. 350µl ammonium hydroxide stock of oligonucleotides was precipitated with 35µl 3M NaAc pH5.5 and 770µl EtOH. This was incubated at -20°C for 30 minutes and the precipitate spun down in a microfuge at 4°C for 15 minutes. The pellet was washed twice in 80% EtOH and then freeze dried. The pellet was resuspended in 200µl TE and the concentration estimated from the OD_{260nm}. An OD_{260nm} of 1 was estimated to be equivalent to 25µg/ml single stranded oligonucleotide.

2.6.2 Reaction conditions

The standard PCR mix contained;

1x Taq polymerase reaction buffer (Promega or nbl (Northumbria biolabs)),

200µM each dNTP,

5µg/ml each PCR primer,

2.5U Taq polymerase (Promega or nbl),

sterile dH_2O.

The reactions were set up in 0.5ml eppendorf tubes generally using either 25 or 50 µl reaction mix, and covered with 3-4 drops mineral oil. Amplification was performed by an initial 3 minute denaturation (94°C), and then 30 cycles; 50°C 2 minutes 72°C

1 minute 94°C 1 minute, followed by a final extension of 50°C 1 minute and 72°C 10 minutes.

If hot start PCR was used the reaction mix, minus the enzyme, was heated to 94°C for 3 minutes and the temperature was then held at 72°C while the Taq polymerase was added. The annealing temperature was subject to variation depending on the primers involved and whether they were 100% identical to the template or not. The extension time was increased to 2 minutes or more when long products (>1kb) were expected. For exon 1 PCR 10% glycerol was included to minimise amplification of spurious bands, these probably occurred due to the high GC content of the region and of the primers. Initially reactions were performed using Perkin Elmer Cetus DNA thermal cyclers and later Hybaid Omnigene thermal cyclers.

2.6.3 PCR template

Several different templates were used in the course of this project.

PCR analysis of bacterial colonies was performed by picking the colony and vortexing in 1ml dH_2O. The cell suspension was boiled for 1 minute and chilled on ice. 1µl was then used as a template.

For PCR directly from bacteriophage, 5µl phage suspension in SM buffer were placed in the balance dH_2O for the PCR reaction and boiled for 5 minutes, the rest of the PCR mix was then added.

If cDNA was used 1/10th-1/5 of the PCR reaction volume was added, equivalent to 25-50ng RNA per 50µl PCR reaction.

For PCR from genomic DNA, 0.3µg genomic DNA was added per 100µl reaction

2.7 Solutions

Bacteria and bacteriophage media

L-Broth per litre, pH7.2	10g tryptone (Difco), 5g yeast extract (Difco), 10g NaCl, 2.46g $MgSO_4$.
L-agar	contains L-both plus 15g agar/litre.

CY (select) agar per litre, pH7.2	10g casamino acids (Difco), 5g yeast extract, 3g NaCl, 10g agar.
CY top agarose	contains 6.5g/l agarose in place of agar.
Terrific Broth per 800ml.	12g tryptone, 24g yeast extract, 4ml glycerol.
KPO_4 buffer	170mM KH_2PO_4, 720mM K_2HPO_4. 1/10 volume added to Terrific broth.
SM buffer	100mM NaCl, 10mM $MgSO_4$, 50mM Tris-HCl pH7.5, 0.01% Gelatin. All reagents were sterilised by autoclaving before use.
Ampicillin stock solution	100mg/ml in dH_2O (filter sterilised).
IPTG stock solution	20mg/ml in dH_2O (filter sterilised).
X-Galactose	20mg/ml in dimethylformamide.

DNA manipulation and purification

Denaturing solution	1.5M NaCl, 0.5M NaOH.
Neutralising solution	1.5m NaCl, 0.5M Tris-HCl pH8.
20xSSC	3M NaCl, 0.3M Na_3citrate.
GTE	50mM Glucose, 25mM Tris-HCl pH8.0, 10mM EDTA.
TE	10mM Tris-HCl pH7.6, 1mM EDTA pH8.0.

Magic Minipreps™

Cell Resuspension Solution	50mM Tris-HCl pH7.5, 10mM EDTA, 100mg/ml RNase A.
Cell Lysis Solution	1%SDS, 0.2M NaOH.
Column Wash Solution	100mM NaCl, 10mM Tris-HCl pH7.5, 2.5mM EDTA, 50% EtOH.

Electrophoresis and hybridisation

1xTAE	40mM Tris-Acetate, 1mM EDTA pH8.0.
1xTBE	90mM Tris-Borate, 2mM EDTA pH8.0.
10x RNA gel running buffer	0.4M MOPS (morpholinopropansulphonic acid), 100mM NaAc, 10mM EDTA pH8.0.
6x glycerol loading buffer	0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol in dH_2O.
Formamide loading buffer	95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol.
Alkali transfer buffer	0.25 NaOH, 1.5M NaCl.
Church and Gilbert's solution	0.5M Na Phosphate pH7.2, 7%SDS.

Enzyme buffers

10x *E.coli* DNA polymerase I buffer	0.5M Tris-HCl pH7.5, 0.1M $MgCl_2$, 10mM DTT, 500µg/ml BSA (Fraction V), 200µM dNTPs.

10x T4 Ligation buffer	200mM Tris-HCl pH7.6, 50mM $MgCl_2$, 50mM DTT, 500µg/ml BSA (Fraction V).
10x Turbo cloning buffer	500mM Tris-HCl pH8, 5mM rATP, 5mM DTE, 50mM $MgCl_2$.
10x Taq polymerase buffer (Promega)	1.5mM $MgCl_2$, 50mM KCl, 10mM Tris-HCl pH 9 0, 1% Triton X-100.

Alkaline phosphatase colour reaction

Buffer 1	0.1M Tris-HCl pH 7.5, 0.15 M NaCl.
Buffer 2	buffer 1 containing 1% blocking reagent.
Buffer 3	0.1M Tris-HCl pH9.5, 0.1M NaCl, 50mM $MgCl_2$.
Buffer 4	10mM Tris-HCl pH8.0, 1mM EDTA.
NBT solution	75mg/ml nitroblue tetrazolium in dimethylformamide.
BCIP or X-phosphate solution	50mg/ml toluidinium 5-bromo-4-chloro-3-indolyl-phosphate in dimethylformamide.

Whole mount *in situ* hybridisation

Hybridisation buffer pH to 5 with 1M citric acid.	50% deionised formamide, 5xSSC, 50µg/ml Heparin, 0.1% Tween-20.
Post hybridisation Wash	2xSSC, 50% formamide, 0.1%Tween-20.

1xTBST	140mM NaCl, 3mM KCl, 25mM Tris pH7.5, 0.1% Tween-20.
Alkaline phosphatase buffer	100mM Tris-HCl pH9.5, 100mM NaCl, 50mM $MgCl_2$, 0.1% Tween-20.
Heat inactivated embryo powder	ground, acetone dehydrated, homogenised embryonic tissue, heated at 70°C for 30 minutes in 1xTBST
Xenopus rearing solution	60mM NaCl, 0.6mM KCl, 0.4mM $MgSO_4$, 0.3mM $Ca(NO_3)_2$, 0.8mM Na_2HPO_4, 0.15mM KH_2PO_4, 2.4mM $NaHCO_3$.

2.8 Computer analysis of DNA sequences

The University of Wisconsin's Genetics Computer Group have produced an extensive and expanding set of programs for DNA sequence analysis and manipulation (GCG 1991). Routine functions were performed using these programs. In addition sequence comparison of the *WT1* orthologs was performed using the gap program. The pepplot program was used to draw the protein structure prediction plots, motifs and helicalwheel programs were used to investigate protein structure motifs. The Blast program was used to search the EMBL, GenBank and SwissProt sequence databases using sequences isolated in this study (Altschul *et al.,* 1990, Karlin and Altshul 1990). Clustal V was used for phylogenetic analysis and multiple alignment of the *WT1* sequences (Higgins and Sharpe 1988, Higgins *et al.,* 1992), and is described in Chapter 5. Signal Scan was used to locate the presence of any possible transcription binding sites (Prestridge 1991). All the above programs were used via the Human Genome Mapping computing facility.

The Coils program (Lupas *et al.,* 1991) and three updates were used to analyse the potential of *WT1* to form coiled coils. A UNIX version (Yauso Ina 1993) of Li's 1993 program, lwl15, (Li *et al.,* 1985, Li 1993) was used for the calculation of K_S and K_A to investigate the rate of evolution of *WT1*.

Chapter 3

Identification of sequences cross hybridising to *WT1* in the vertebrates

3.1 Introduction

Initially, to determine whether sequences homologous to the *WT1* gene were present in species other than mouse and human, hybridisation of a *WT1* probe to a range of vertebrate DNA's was carried out at low stringency. This technique known as zoo blot hybridisation is often used in gene mapping to determine whether a single copy probe is conserved across species and therefore may correspond to exonic gene sequence (Monaco *et al.,* 1986). It has been used to demonstrate, for known genes, the conservation of homologous sequences in the genomes of different species, in some cases as far as *Drosophila* e.g. *PAX6* (Ton *et al.,* 1991b).

Comparison of nucleotide sequences of mouse and human *WT1* showed 81% overall identity of the entire cDNA and this increased in the protein coding region to 91%, and was even higher in the zinc finger region at 95% (Buckler *et al.,* 1991). This indicated that the *WT1* gene is highly conserved between mouse and human and therefore probably highly conserved within the rest of the mammalian lineage at least.

3.2 Detection of sequences cross hybridising to *WT1* in mammals

In order to determine the presence of *WT1* in other mammalian species apart from human and mouse, genomic DNA was obtained from human (*Homo sapiens*), mouse (*Mus domesticus*), rat (Wistar), guinea pig (*Cava porcellus*), pig (*Sos scrofa*), sheep (*Ovis aries*), and cow (*Bos tarsus*) (gifts R. Hill and A. Archibald). A mouse *Wt1* cDNA clone pK/S2 (gift Buckler *et al.*) containing all but the first 779bp of exon 1 was used to probe the EcoRI restricted DNA's at reduced stringency. The hybridisation temperature that was used was determined using the formula of 1-1.5^{o}C reduction in melting temperature (Tm) per percentage mismatch of the probe with the target DNA (Sambrook *et al.,* 1989). Washing was carried out in stages and autoradiography performed before the stringency was increased.

Bands were observed in the pig, sheep and cow samples after washing at 58^{o}C in 2xSSC. For human, mouse, rat and guinea pig, bands were seen after washing at 67^{o}C in 0.5xSSC (Figure 3.1).

The strongest band detected in mouse was 8kb and this corresponded with the band seen in C57/Bl6 mice (Buckler *et al.,* 1991) under high stringency conditions. In human genomic DNA a *WT1* cDNA probe detects 8 EcoRI fragments (13.6, 10.4, 7.2, 6.1, 5.8, 3.7, 3.1, 1.85kb) (Call *et al.,* 1990, Haber *et al.*, 1991). In this

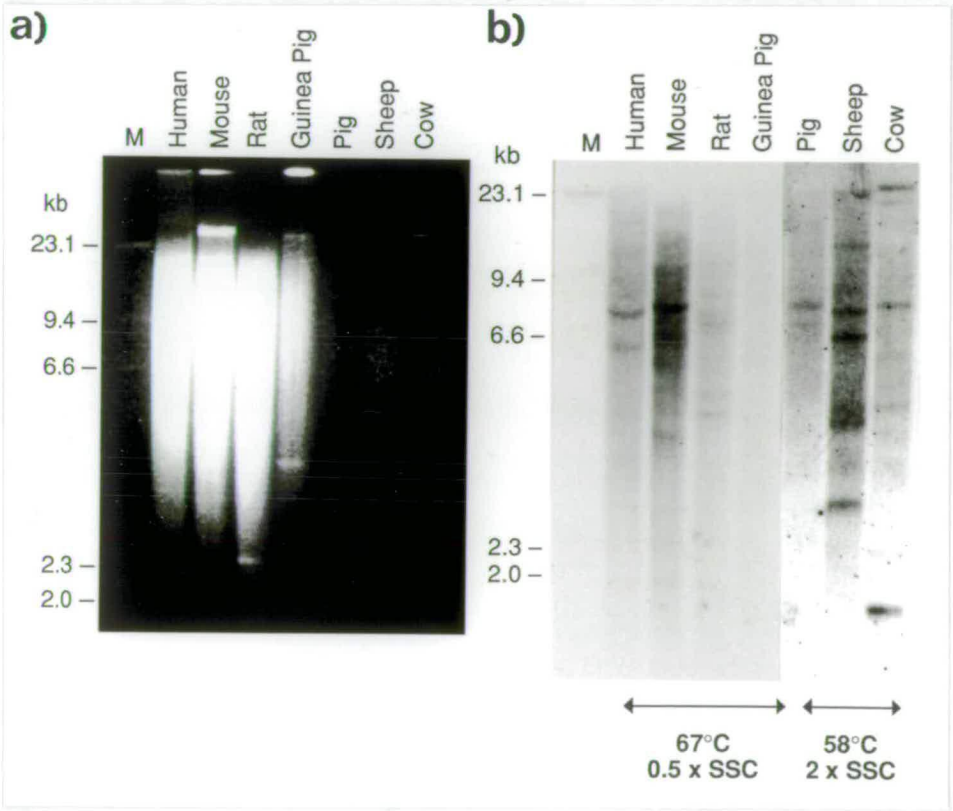
Figure 3.1

Detection of sequences that hybridise to *WT1* in a range of mammalian genomes.

a Ethidium bromide stained agarose gel of EcoRI digested genomic DNA.

b Autoradiographs after hybridisation at reduced stringency (58°C) with mouse *Wt1* cDNA (pK/S2).

Washing conditions at which specific bands were observed are indicated.



experiment the 3.7 and 3.1 kb bands were faint and the 6.1kb band was not detected but all the rest were seen (Figure 3.1b). Therefore it was thought that the conditions used were detecting *WT1* sequences and so it could be presumed that sequence homologous to *WT1* were present in rat, guinea pig, pig, sheep and cow genomes (Figure 3.1).

3.3 Detection of sequences cross hybridising to *WT1* in non mammalian vertebrates

A major problem with low stringency hybridisation with a probe containing a motif is that hybridisation may be occurring to paralogous or related genes also containing this motif, and therefore giving spurious signals. In the case of *TFIIIA*-like zinc fingers it has been estimated that there are 300-700 finger containing genes in the vertebrate genome (Crossley and Little 1989). In order to avoid cross finger hybridisation a 454bp NcoI restriction digest fragment of WT33 was subcloned (H454). This contained of part of exon 1 and exons 2-4 but not the zinc finger encoding region. WT33 was one of the first human *WT1* cDNA clones to be isolated, containing all but the 5' 630 bp (Call *et al.,* 1990).

Human, mouse, rat, pig, chick (*Gallus domesticus*), frog (*Xenopus laevis*) and trout (*Salmo*) genomic DNA was digested either with BamHI or EcoRI. The amount of DNA used for the digestion was calculated so that the equivalent number of haploid genomes was used for each species (Table 3). Hybridisation was performed at 55^{o}C with the human probe H454 and washing carried out until bands were observed (2xSSC). These conditions had previously been shown to detect cross species hybridisation for *SRY*-like sequences in vertebrate genomes (Tiersch *et al.,* 1991). Bands were observed in human, mouse, chick, frog and trout with the BamHI digest (Figure 3.2) and frog with EcoRI digest (data not shown).

3.4 Conclusions

Zoo blotting of vertebrate genomic DNA has demonstrated the presence of sequences capable of cross hybridisation to *WT1* probes in four of the five vertebrate classes (mammal, bird, amphibian, and fish). These may not be *WT1* orthologs but closely related sequences, however it was shown that mostly *WT1* specific bands were being detected in species where the sizes of *WT1* specific bands were known (Figure 3.1). Secondly the possibility of hybridisation to other zinc finger genes was reduced

by the use of a probe not containing the zinc finger encoding region. The sequence of the part of *WT1* sequence used was not similar to any known sequence on searching of the EMBL and GenBank databases. Therefore these results showed that it was likely that *WT1*-like genes were present in all vertebrate genomes and gave encouragement to attempt the cloning of *WT1* from various species.

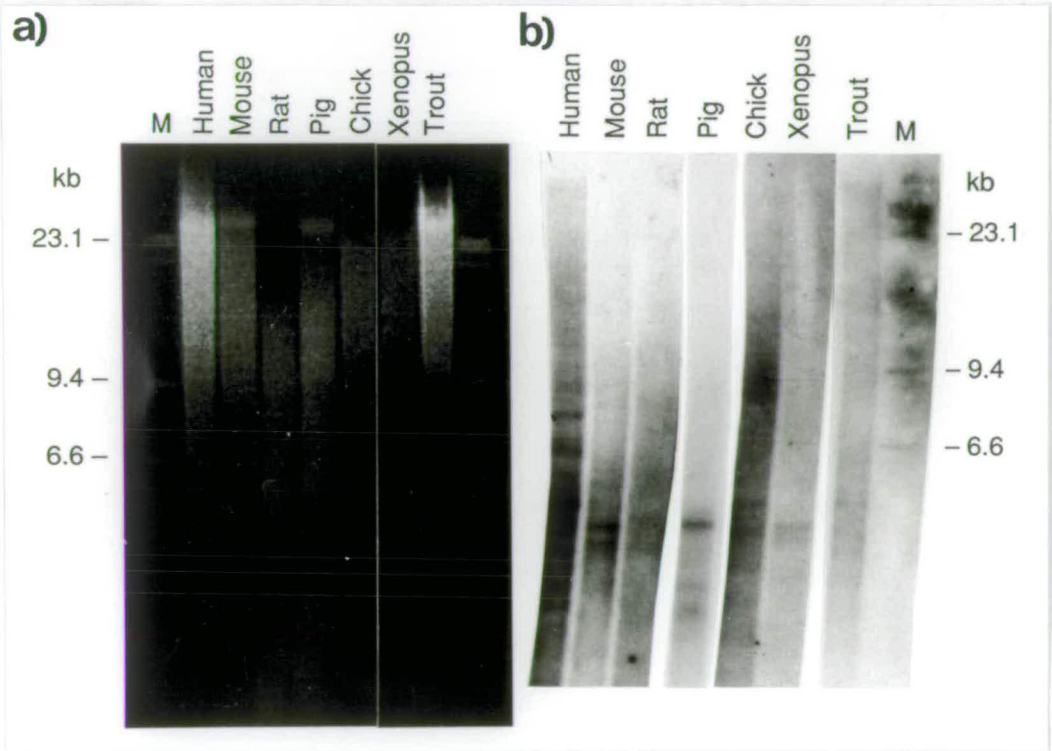
Table 3 Size of vertebrate genomes.

Organism	Nucleotide pairs $x10^9$ per haploid genome
Homo sapiens	3.4
Mus musculus	3.36
Rattus norvegicus	2.9
Gallus gallus	1.2
Xenopus laevis	6.2
Trout	2.5

Figure 3.2
Detection of sequences that hybridise to *WT1* in a range of vertebrate genomes.

a Ethidium bromide stained gel of BamHI digested genomic DNA

b Autoradiograph after hybridisation at reduced stringency (55°C) with a subclone of human *WT1* including the region 5' to the zinc fingers (H454).



Chapter 4

Isolation of *WT1* orthologs

4.1 Isolation by cross species PCR

4.1.1 Introduction

Since it had been shown, by reduced stringency hybridisation, that there were sequences within non mammalian vertebrate genomes which hybridised to a *WT1* probe, the cloning of sequences homologous to *WT1* within the vertebrates was attempted. The traditional method for the isolation of sequences homologous to a known gene, such as orthologs and paralogs, has been by reducing the stringency of hybridisation when screening libraries. This can be a very time consuming process and often many unrelated sequences are isolated.

In the last few years a major change in methodology has come about with the invention of the Polymerase Chain Reaction (PCR). PCR was developed in 1984 by the Cetus Corporation and is a method of amplification of specific DNA sequences in a cell-free system. PCR abrogates the need for the initial cloning of the DNA prior to amplification (Saiki *et al.,* 1985, Mullis and Faloona 1987). It is based on two oligonucleotide primers flanking the region of interest on opposite strands of the DNA, and cycles of annealing the primers to the template, extension of the primers by a DNA polymerase, and denaturation of the product. The presence of two primers instead of one results in an almost exponential, rather than linear, amplification of the target region (Mullis and Faloona 1987). The speed and sensitivity of PCR has been enhanced by microchip controlled heating blocks, and isolation of a heat stable DNA polymerase (Taq polymerase) from *Thermus aquaticus*. The simplicity of this technique lends itself to manipulation and therefore a wide range of uses (reviewed by Arnheim and Erlich 1992).

Although PCR was originally developed for situations where the template sequence was known and so complementary primers could be designed (e.g. for the analysis of mutations (Saiki *et al.,* 1985)), it quickly became clear that by lowering the stringency of the reaction a degree of mismatch between the primer and template could be tolerated. This has been exploited to amplify related genes, genes from different species and phyla (Orlandi *et al.,* 1989, Sarker and Sommer 1989, Pang *et al.,* 1990, Sarker *et al.,* 1990, Sommer *et al.,* 1992), or where only the amino acid sequence of the protein is known (Lee *et al.,* 1988). Once part of the target gene has been isolated this can be used as an homologous probe for library screening (Lee *et al.,* 1988).

For the isolation of *WT1* orthologs cross species PCR was particularly applicable for several reasons. The speed of the technique makes it the best

Table 4.1 Sources of RNA from which the isolation of *WT1* orthologs was attempted

Species	Taxon	Stage/Age	Tissue	Source
Mouse (BALB/c)	Mammal	Adult	Testis	
Rat (Wistar)	Mammal	Adult	Testis	
Pig (*Sos scrofa*)	Mammal	Juvenile	Testis	Testis - A. Archibald, AFRC Roslin, Midlothian
Chick (*Gallus domesticus*)	Bird	8-14 day Embryo	Mesonephros / Metanephros and Gonad combined	
Alligator (*Alligator mississippiensis*)	Reptile	Stage 20 Embryo	Gonad	RNA - AM. Coriat, University of Manchester
Frog (*Xenopus laevis*)	Amphibian	Stage 45 Tadpole	Whole tadpole	Eggs - P.M. Gaze, University of Edinburgh
Frog (*Xenopus laevis*)	Amphibian	Stage 53 Tadpole	Mesonephros	" " "
Zebrafish (*Brachydanio rerio*)	Fish Teleost	Larva	Whole larva	RNA - P. Ingham, ICRF, Oxford Reverse transcribed RNA - T. Jowett, University of Newcastle
Ascidian (*Herdmania momus*)	Tunicate	Tadpole	Whole tadpole	RNA - M. Levine, Queensland Institute of Medical Research, Australia

way of isolating multiple sequences and therefore for producing sequence comparisons. The zinc finger motif is one of the most highly conserved regions within *WT1* when human and mouse are compared and therefore is most likely to hybridise cross species. As mentioned in chapter 3, the numbers of this motif are high in the vertebrate genome (Crossley and Little 1991). Therefore if reduced stringency screening with the motif was used, non *WT1* orthologs especially the *EGR* family (the most similar fingers to *WT1*) could be isolated. The use of two primers in PCR allows for the selection of zinc fingers at one site and *WT1* specificity in the other primer.

4.1.2 Experimental design

The main aim of the design was to favour the amplification of *WT1*-like genes over any other genes in PCR i.e. to find the balance between specificity and degeneracy. If non specific amplification occurs, in addition to obscuring the desired product, it can inhibit its amplification by sequestering substrates.

4.1.2.1 Template

Three PCR strategies were considered. PCR from genomic DNA amplifying across introns (interexonic); amplifying each exon separately (intraexonic); and RT-PCR using RNA from a tissue likely to be expressing *WT1*. RNA was chosen as the starting material rather than genomic DNA. This was because the genomic structure of only the zinc finger region was known (Haber *et al.*, 1991). For both interexonic and intraexonic PCR a knowledge of the genomic structure is needed.

Total RNA was prepared wherever possible from tissue found to express *WT1* in mammals (Table 4.1). The gonads were frequently chosen as these are easily recognised and *WT1* expression had been demonstrated to be at a higher level than in the kidney of the post natal mouse. The species chosen represent every gnathostome class and one non-vertebrate chordate. Mammalian species were included mainly to act as a test system to show that cross species PCR of *WT1* was possible. Other species were chosen as the most studied example of their class and are therefore widely obtainable and the most useful for subsequent studies.

4.1.2.2 Primer design

Probably the most important feature of cross species PCR is the design of the oligonucleotide primers. It has been found that for successful priming mismatches are not tolerated in the three 3' bases for a 20bp primer (Sommer and Tautz 1989, Sarker *et al.*, 1990), although some priming can sometimes be obtained in longer primers

Figure 4.1 Predicted amino acid sequence of human *WT1* showing the exon structure and the positions of degenerate PCR primers used in this study

```
     Exon 1
1    MGSDVRDLNA LLPAVPSLGG GGGCALPVSG AAQWAPVLDF APPGASAYGS

                          B312→
51   LGGPAPPPAP PPPPPPPPHS FIKQEPSWGG AEPHEEQCLS AFTVHFSGQF

101  TGTAGACRYG PFGPPPPSQA SSGQARMFPN APYLPSCLES QPAIRNQGYS

     Exon 2                                          Exon 3
151  TVTFDGTPSY GHTPSHHAAQ FPNHSFKHED PMGQQGSLGE QQYSVPPPVY

                               Exon 4
                               B300→        B297→
201  GCHTPTDSCT GSQALLLRTP YSSDNLYQMT SQLECMTWNQ MNLGATLKGV

     Exon 5             Exon 6                C227→
251  AAGSSSSVKW TEGQSNHSTG YESDNHTTPI LCGAQYRIHT HGVFRGIQDV

     Exon 7                B620→               C229→
301  RRVPGVAPTL VRSASETSEK RPFMCAYPGC NKRYFKLSHL QMHSRKHTGE

     Exon 8                            Exon 9
351  KPYQCDFKDC ERRFSRSDQL KRHQRRHTGV KPFQCKTCQR KFSRSDHLKT

                Exon 10
     ←C228      ←B298
401  HTRTHTGKTS EKPFSCRWPS CQKKFARSDE LVRHHNMHQR NMTKLQLAL*
```

with a mismatch in the third base from the 3' end. Mismatches between primer and template in any position will affect the melting temperature of the primer template hybrid. In the 5' region mismatches are tolerated as shown by the inclusion of restriction enzyme recognition sites (Scharf *et al.,* 1986).

For *WT1*, primers were designed by comparison of the two known sequences, mouse (Buckler *et al.,* 1991) and human (Call *et al.,* 1990, Gessler *et al.,* 1990) using both the nucleotide and predicted amino acid sequences. Figure 4.1 shows the positions of the chosen primers relative to human *WT1*. To minimise the number of mismatches regions of greatest conservation were used and sequences selected containing amino acids with the least number of codon degeneracies. The downstream primer was made complementary to a region within the zinc fingers in the linker region between ZF3 and 4 just downstream of the KTS alternative splice. This linker is a highly conserved motif between the *TFIIIA*-like zinc fingers (Choo and Klug 1993). It was hoped that the proximity to a splice site might lead to greater conservation at the nucleotide level. The positioning of the upstream primers was not as obvious. Therefore two primers were made complementary to parts of the putative leucine zipper, which is methionine rich (methionine is encoded by only one codon). Further upstream the sequence is GC and proline rich. The former is known to inhibit PCR of *WT1* (W. Bickmore personal communication); the latter increases the possible degeneracy of the coding sequence and therefore increases the likelihood of mismatches and failure of amplification. One primer was made complementary to the sequence thought least likely to be degenerate in this region. When variation in nucleotide sequence could be predicted from codon degeneracy, or there were changes between human and mouse sequences, mixed bases or inosine were incorporated (Lee *et al.,* 1988, Sommer and Tautz 1989). Inosine is capable of complementing all four bases. Codon preference can sometimes be useful in defining whether codons are likely or not to appear in a species. For those vertebrate species used there only seems to be slight differences in codon preference. No codons were exclusively favoured or unused (Wada *et al.,* 1991). This is different to the situation in some organisms such as *E.coli* or yeast where some codons are used exclusively and some not at all, arising from differences in proportions of tRNA molecules for the different codons (Ikemura 1985). In humans the GC content of the genome is known to influence codon choice. The primers were designed to amplify all of the species selected rather than targeting one. At the same time primer degeneracy was kept to a minimum as the number of sequences to which primers can anneal increases very rapidly with increasing degeneracy. To produce specific amplification the right

Table 4.2 Sequence of degenerate primers for cross species PCR

Primer Number	Position in human sequence	Primer Sequence
B312	Exon 1 (572-599bp)	T G 5'CGCGTCGACCCCACTCITTCATCAAACA 3'
		Human CGCCGCCGCCTCACTCCTTCATCAAACA
		||||||| || ||||||||||||||
		Mouse CGCCGCCACCCCACTCCTTCATCAAACA
		aa sequence P P P H S F I K Q
B300	Exon 4 (1042-1065bp)	G 5'AGCGTCGACAATTTATACCAAATG 3'
		Human AGCAGTGACAATTTATACCAAATG
		||||||||||||||||||||||||
		Mouse AGCAGTGACAATTTATACCAAATG
		aa sequence S S D N L Y Q M
B297	Exon 4 (1075-1101bp)	C 5'TTGGTCGACATGACCTGGAATCAGATG 3'
		Human CTTGAATGCATGACCTGGAATCAGATG
		|||||||||||||||||||||||||
		Mouse CTTGAATGCATGACCTGGAATCAGATG
		aa sequence L E C M T W N Q M
B298	Exon 10 / ZF4 (1609-1634bp)	C C 5'TGCAAGCTTCAGCTGAAGGGTTTTTC 3'
		Human GGCCACCGACAGCTGAAGGGTTTTTC
		||||||||||||||||||||||||
		Mouse TGCCACCGACAGCTGAAGGGTTTTTC
		aa sequence P W R C S F P K E

Underlined sequences indicate changes made to produce restriction enzyme recognition sites

balance between selectivity and degeneracy must be found. To facilitate cloning SalI restriction enzyme recognition sites were designed three bases in from the 5' end of the upstream primers and a HindIII site in the downstream primer. Table 4.2 shows the primers and the sequence from which they were designed.

4.1.2.3 Reaction conditions

An estimate of the melting temperature (Tm) of a primer template hybrid can be calculated on the basis of the primer sequence (Sambrook *et al.,* 1989). This can be used as a guideline to define the annealing temperature of the PCR. With degenerate primers only an average value can be obtained as more than one the primer sequence is present. The Tm is affected by mismatches; the magnitude of the effect can depend on the position of the mismatch within the primer. Therefore conditions at which primers produced specific amplification of *WT1* were obtained empirically. The stringency of the reaction could be altered by changing either the annealing temperature or the magnesium concentration, both of which influence the binding of primer to the template.

4.1.3 Testing PCR primers

All three upstream primers (B312, B300, B297) were tested in combination with the downstream primer (B298) using either a mouse *Wt1* cDNA clone (pK/S2) and/or total RNA reverse transcribed by the first strand synthesis method (RT-PCR) (using the protocol described in Innis *et al.,* 1990). A range of conditions were tried the most successful, producing specific amplification of *Wt1,* were 94°C 1 minute, 49°C 2 minutes, 72°C 1 minute for 30 cycles and a final extension of 20 minutes at 72°C (in a magnesium concentration of 1.5mM). 'Hot start' PCR where the enzyme is not added until after the first denaturation step was also used. This is in order to avoid non specific extension by the enzyme that can occur at room temperature (Arnheim and Erlich 1992). That *Wt1* was represented within the PCR products was verified by Southern blotting and hybridisation to the mouse *Wt1* cDNA clone (Figure 4.2).

4.1.4 Isolation of rat, pig and alligator *WT1* orthologs

4.1.4.1 Cross species PCR

Using the conditions which were found to amplify mouse *Wt1*, RT-PCR was performed on rat, pig, chick, alligator and *Xenopus* RNA. Unfortunately no specific

amplification products were detected. RT-PCR was thought not to be as efficient as conventional PCR from double stranded DNA (I. Jackson personal communication). Therefore cDNA was produced from total RNA using a Boehringer Mannheim cDNA synthesis kit. After 30 cycles of PCR, bands on an agarose gel of the predicted size could be detected with the combination of primers B297-B298 with rat, pig and alligator cDNA templates and with B312-B298 in rat (Figure 4.3a shows a typical amplification with strong specific bands being produced). The possibility of that these PCR products were homologous to *WT1* was confirmed initially by Southern blotting and reduced stringency hybridisation with the probe *Wt1*. That non specific amplification was also occurring was evident from the appearance of other non hybridising bands. No specific amplification was detectable using chick or *Xenopus* cDNA as a template either by ethidium bromide staining or by hybridisation to *Wt1*. The same negative result was obtained with all combinations of primers, using a range of temperatures, primer and template concentrations.

4.1.4.2 Cloning PCR products

For all successful amplifications, specific *WT1* products were purified away from non-specific products, primers, nucleotides and Taq polymerase by excision of the band from an agarose gel followed by purification away from the agarose using GENECLEAN (Bio 101). If the amount of isolated DNA was insufficient for subsequent manipulations a second round of PCR was performed using the purified fragment. Products were digested with SalI and HindIII restriction enzymes to cut within the sites included in the ends of the PCR primers and then ligated into the appropriately digested and dephosphorylated pBluescript SKII+ vector. On transformation of JM101 or XL1-Blue cells no clones were obtained. The exact reason for this failure is not known, but it could be due to the enzymes not cutting these sites near the end of DNA primers. However both HindIII and SalI sites have been used successfully in this context before (Lee *et al.*, 1988 and personal experience). The second possibility became clear when these products were eventually cloned by another method. It was found that very few of the primers were full length, in particular the 5' end, which includes the restriction site, was often lost. If this was also the case in the uncloned product it would mean that many fewer product molecules would be substrates for the restriction enzymes. Therefore fewer molecules with compatible sticky ends would be generated, reducing cloning efficiency.

Blunt ended cloning following the 'double GENECLEAN' protocol (Bio101) was subsequently used. The ends of PCR products were chewed back and kinased using

Figure 4.3 Isolation of *WT1* orthologs

a Ethidium bromide stained agarose gel of PCR products amplified from cDNA from mouse, rat, chick, alligator and *Xenopus*. * indicates the presence of the expected size band on the gel. No amplification with these primers was obtained with *Xenopus* cDNA, and with chick cDNA only a smear of non-specific amplification was obtained.

b Reduced stringency hybridisation with *Wt1* (KS/2) to white colonies after transformation with rat PCR products ligated into pBluescript SKII+. Filters were hybridised and washed at 61°C and washed to 0.5xSSC (2hr exposure).

c PCR products resolved on an agarose gel. PCR analysis of the bacterial colonies giving a positive signal after hybridisation to *Wt1*. The forward and reverse M13 vector primers were used to amplify across the cloning site. The expected sizes of the products, depending on the PCR product cloned, are indicated by arrows. Most of the colonies were found to contain inserts of the expected size.

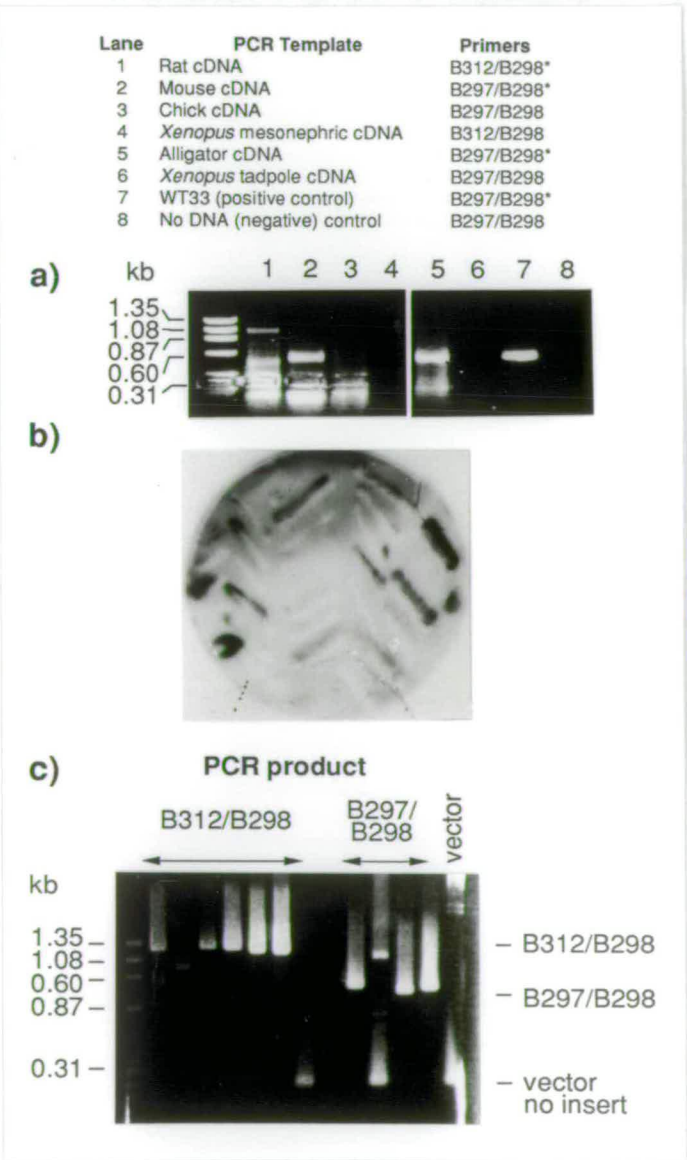
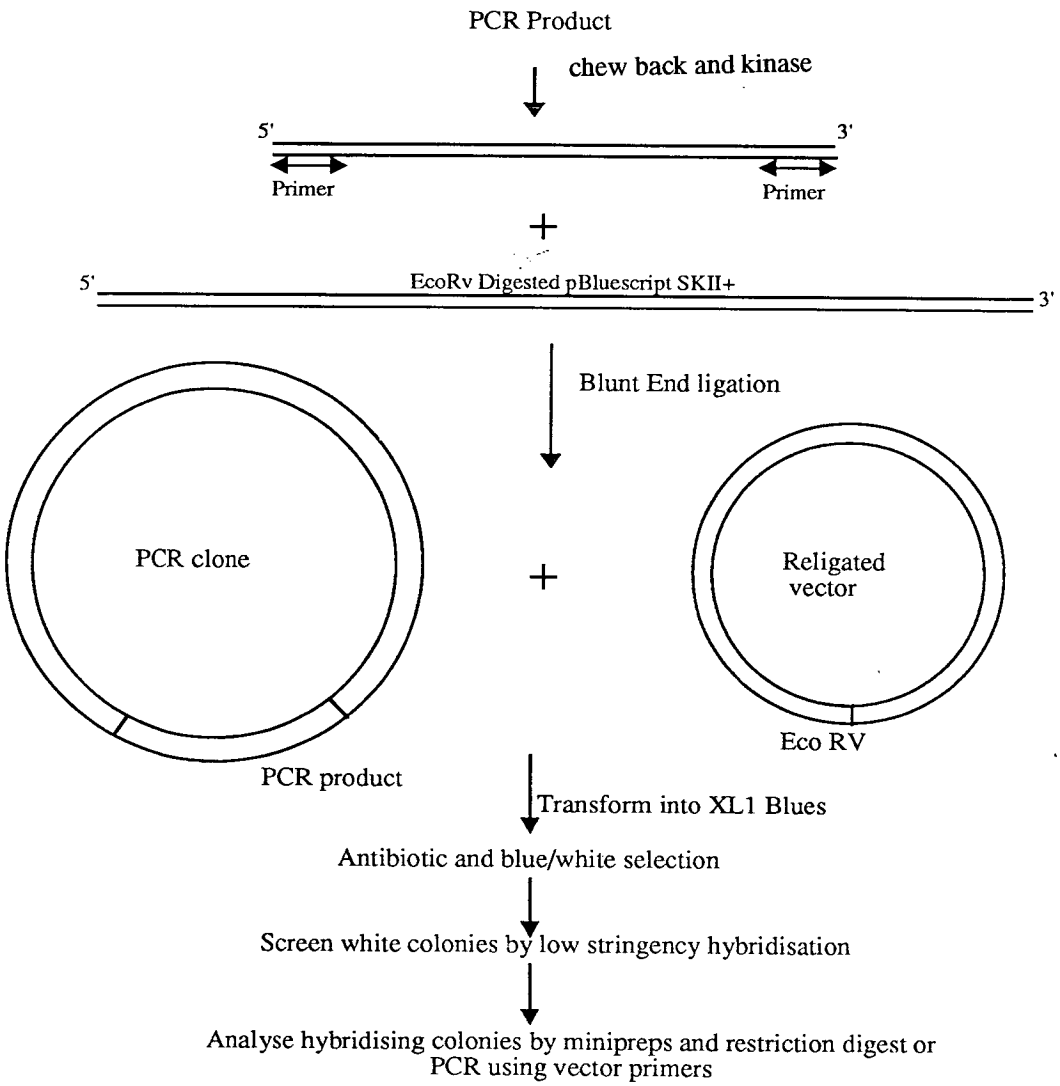


Figure 4.4 Blunt ended cloning strategy used for cloning PCR products

The ragged ends of the purified PCR products were chewed back and the 5' ends kinased before ligation into EcoRV cut pBluescriptSKII+. After transformation white colonies were selected and screened for *WT1* inserts by low stringency hybridisation. The inserts from positive colonies were analysed by PCR across the cloning site.



DNA polymerase I and polynucleotide kinase prior to ligation with EcoRV digested pBluescript SKII+ (Figure 4.4). Chewing back the DNA ends was performed because Taq polymerase, in common with several other polymerases can add a non template directed nucleotide (normally dATP) to the 3' OH of blunt ended double stranded DNA (Clarke 1988). This would interfere with blunt ended ligation. After transformation, recombinants were identified using the blue/white system which relies on the α complementation of β-galactosidase function. Recombinants were screened for *WT1*-like inserts by reduced stringency hybridisation to *Wt1* (Figure 4.3b). Plasmid DNA from positive clones was prepared by small scale plasmid preparations. The presence of an insert was then confirmed by restriction digest, or PCR across the cloning site using vector primers (Figure 4.3c). Inserts of the expected size were then sequenced.

All the clones sequenced from the three species (rat, pig, and alligator), contained inserts which encoded predicted open reading frames. These closely resembled but differed from the two known primary amino acid sequences of *WT1* and also differed from each other.[1] Table 4.3 shows the homology at the nucleotide and amino acid levels. These were named *RWT1*, (rat *WT1*), *PWT1* (pig *WT1*) and *AWT1* (alligator *WT1*). This high level of similarity at the nucleotide and amino acid level provided strong evidence that orthologs of *WT1* had been isolated. More than one clone was sequenced for each species as Taq polymerase has a higher rate of misincorporation as it lacks 3'5' exonuclease (proof-reading) activity. The error rate is currently estimated to be 8.5×10^{6} but is known to vary with the conditions (Fuchareon *et al.*, 1989). Therefore for a 200bp fragment an estimated 3% of cloned products will contain errors. This error rate was consistent with the very low level of variation between clones e.g. in alligator only one point of variation in three 500bp clones was found. On publication the sequence of the entire coding region of *WT1* from a rat cDNA (Sharma *et al.*, 1992) corresponded to that obtained here. Sequencing several clones also revealed variation (Figure 4.5) owing to the presence of the two alternative splices which have already been observed in mouse and human (Haber *et al.*, 1991).

[1]The sequences of all the clones and a nucleotide comparison are contained in Appendix A.

Table 4.3 Similarity of the PCR clones isolated to human and mouse *WT1* at the nucleotide and amino acid level.
Figures were calculated using the Gap program from GCG.

Species	Nucleotide % Identity		No. Amino Acids compared	Amino Acid % Similarity	
	Human	Mouse		Human	Mouse
Rat	91.5	95.2	335	99.1	99.1
Pig	95.1	91.9	165	98.7	97.0
Alligator	81.0	81.5	156	96.1	95.5
Chick	82.5	80.3	154	96.7	96.7
Xenopus	84.2	83.8	82	100	100
Axolotl	85.2	84.7	72	98.6	98.6
Zebrafish (predicted cDNA sequence)	82.7	82.2	72	97.3	97.3

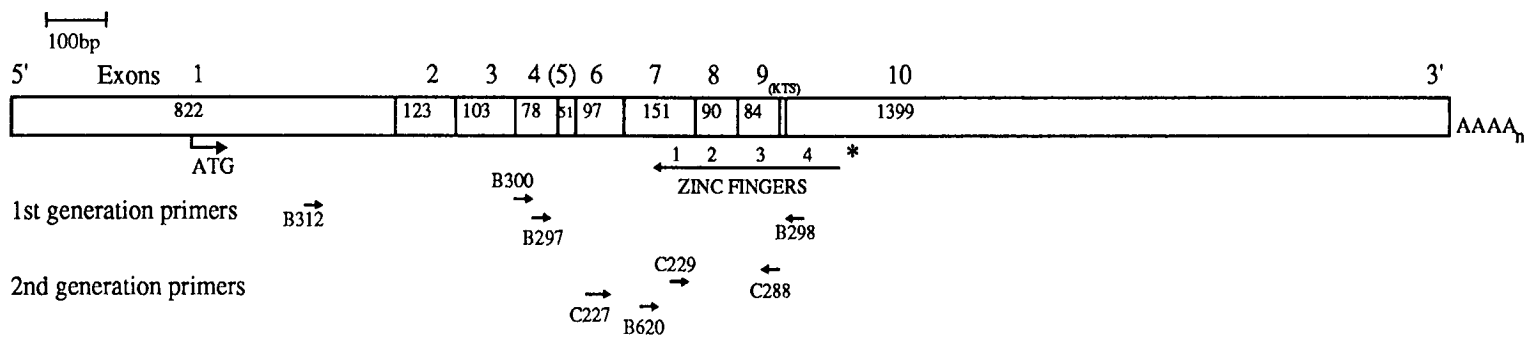
Table 4.4 PCR primers designed using sequence from PCR clones

Primer Number	Position in human sequence	Primer Sequence
C227	Exon 6 (1221-1244bp)	T T G 5'CTGCGGAGCCCAATACAGAATICA 3'
		Human CTGCGGAGCCCAATACAGAATACA
		||| || ||||| ||||||||||
		Mouse CTGTGGTGCCCAGTACAGAATACA
		|||||| ||||| ||||||||||
		Alligator CTGTGGAGCCCAATACAGAATACA
		aa sequence C G A Q Y R I H
B620	Exon 7 / ZF1 (1324-1352bp)	T 5'GAG_CAGCTGAGAAACGICCITTCATGTG 3'
		Human GAGACCAGTGAGAAACGCCCCTTCATGTG
		|| |||||||||||||| || |||||||
		Mouse GAAACCAGTGAGAAACGTCCTTTCATGTG
		|| || | ||| ||||||| |||||||
		Alligator GAGACAAATGAAAAACGTCCCTTCATGTG
		aa sequence E T S E K R P F M **C**
C229	Exon 7 / ZF1 (1384-1406bp)	T A T 5'AAGCTCTCCCACTTACAGATGCA 3'
		Human AAGCTGTCCCACTTACAGATGCA
		|||||||||||||||||||||||
		Mouse AAGCTGTCCCACTTACAGATGCA
		||| | |||| |||||||||
		Alligator AAGTTATCCCATTTACAGATGCA
		aa sequence K L S H L Q M **H**
C228	Exon 9 / ZF3 (1579-1599bp)	A (antisense) 5'ACCTGTATGAGTCCTIGTGTG 3'
		Human ACCTGTATGAGTCCTGGTGTG
		|||||||||||||||||||||
		Mouse ACCTGTATGAGTCCTGGTGTG
		|||||||||||||||||||||
		Alligator ACCTGTATGAGTCCTGGTGTG
		aa sequence G T **H** T R T **H**

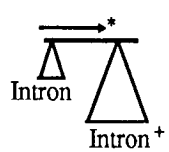
Underlined sequences indicate changes made to produce restriction enzyme recognition sites.
Bold letters indicate zinc chelating residues within zinc fingers.

Figure 4.5 Schematic representation of the human *WT1* cDNA and the *WT1* clones isolated by degenerate PCR.

The primers with which successful amplifications were obtained are shown as well as the alternative splice forms observed. The clones which were subsequently used as probes are also indicated (*).



PCR Primers		Species	Alternative splices	
5'	3'		17aa	KTS
B312	B298	Rat	+	+
B297	B298		+	+
B297	B298	Pig	+	N/D
B300	C228*	Chick	-	N/A
C227	C228		N/A	N/A
C229	C228		N/A	N/A
B297	B298*	Alligator	-	+
B620	B298*	*Xenopus*	N/A	N/A
C229	C228		N/A	N/D
B620	C228	Zebrafish	N/A	N/A
C229	C228*		N/A	N/A



N/A = region not covered by clone
N/D = not observed but not large sample examined
* Probes used for library screens

+ Not to scale

4.1.5 Redesign of PCR primers

The extra sequence information provided by these different species, in particular the first non mammalian *WT1* to be isolated, *AWT1*, enabled several more primers to be designed from conserved regions. Upstream primers C227, B620 and C229 were synthesised, the latter two were complementary to ZF1. A second downstream primer C228 with the 3' end within one of the invariant zinc chelating histidines in ZF3 was also made (Table 4.4, Figure 4.1). These primers were tested on alligator cDNA and found to give specific amplification of *AWT1* (data not shown).

4.1.6 Isolation of chick, *Xenopus* and axolotl *WT1* orthologs

Amplification using chick cDNA from 8-14 day embryos was successful with all the combinations of new primers and the combination B300-C228. That B300-B298 combination failed to work but B300-C228 gave a strong specific signal suggests that there was a mismatch between the chick sequence and B298. Amplification with *Xenopus* cDNA was successful with only the primer combinations B620-B298 and C229-C228. Two sources of total RNA were used; either whole stage 45 tadpoles or mesonephros from stage 53 tadpoles (Nieuwkoop and Faber 1967). Only RNA from the mesonephros produced specific amplification, suggesting that in the whole tadpole RNA the concentration of template was too low. Figure 4.6 shows *WT1* specific amplification from chick and *Xenopus* cDNA. These PCR products were cloned and sequenced as above. Again sequences resembling but not identical to previously cloned *WT1* genes were obtained (Table 4.3, Appendix A). These were named *CWT1* and *XWT1* for chick and *Xenopus* respectively. In both cases only the shortest transcript was isolated (Figure 4.5) even though the 17aa region was covered in chick and the KTS region in *Xenopus*.

These primers, along with *XWT1,* were supplied to P.A. Tsonis at the University of Dayton, Ohio who subsequently amplified and cloned 216bp of *WT1* from the axolotl (*Ambystoma mexicanum*), *AXWT1* (Table 4.3).

4.1.7 Isolation of the zebrafish *WT1* ortholog

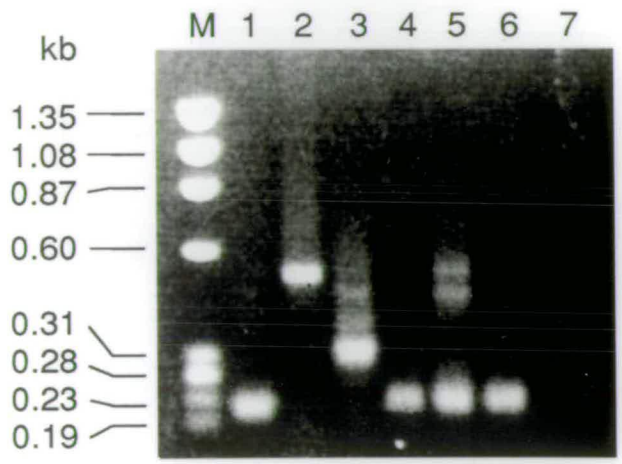
4.1.7.1 Degenerate PCR

The presence of *WT1* in the zebrafish, and the ascidian (a non-vertebrate chordate) was investigated by PCR from whole larva or tadpole cDNA. In both cases

Figure 4.6 Isolation of chick and *Xenopus WT1* orthologs

Resolution of PCR products on an agarose gel. PCR used the primers designed on the basis of the clones already obtained. Strong bands and a low level of non-specific amplification were observed.

Lane	PCR Template	Primers
1	Chick cDNA	C229/C228
2	Chick cDNA	B300/C228
3	*Xenopus* mesonephric cDNA	B620/B298
4	*Xenopus* mesonephric cDNA	C229/C228
5	Chick cDNA (RePCR)	C229/C228
6	Alligator cDNA	C229/C228
7	No DNA (negative) control	C229/C228



amplification was obtained, but on cloning these products either non *WT1*-like products or contaminants were found.

To avoid contamination problems, PCR on zebrafish was repeated at the source of zebrafish material in Newcastle. A range of conditions and primer combinations which had been known to produce *WT1* specific amplification were used with either whole embryo RNA (reverse transcribed by first strand synthesis of RACE reaction) or genomic DNA (both a gift from T. Jowett). Genomic DNA was used in case PCR from whole embryo RNA was not successful. The primers used were known to cross introns in the human genomic structure therefore PCR products would probably not be the same size as any contaminants, but this meant the size of the *WT1* specific products could not be predicted.

Very little amplification with the RNA derived template was seen even with nested PCR (Mullis and Faloona 1987). In contrast a wide range of bands were observed with genomic DNA (Figure 4.7a). To identify which of these bands might be *WT1* orthologs, Southern blotting and reduced stringency hybridisation to *AWT1* was performed. By comparison of the ethidium stained agarose gel and the autoradiograph two genomic bands were identified that hybridised to *AWT1* far more strongly than any others (Figure 4.7b). These corresponded to amplifications with primers B620-C228 and C229-C228. These primers were ones that had given amplification with *Xenopus WT1* (the most diverged *WT1* gene so far isolated). Both bands were very faint, this could well have been due to the large size, approximately 2.8kb, of the products. PCR is known to favour the amplification of smaller products as extension as far as the other primer is more likely to occur for shorter regions. The similar size of the two products was consistent with the human exon structure, in which B620 and C229 would be in the same exon separated by 60bp (Figure 4.9a).

4.1.7.2 Turbo cloning

The B620-C228 band was purified away from non-specific products as before (section 4.1.4.2). This was used as template in heminested PCR (Li *et al.,* 1990) with primers C229-C228 to produce sufficient material for cloning. It was found for successful rePCR of such a long product the fragment could not be directly visualised with ethidium bromide and UV light, as nicking of the DNA inhibited the second PCR resulting in a smear, or a ladder of smaller bands.

Most conventional cloning techniques can be size selective because the ends of smaller fragments have a greater chance of being ligated to the ends of the same vector DNA molecule. To avoid this turbo cloning was used (Boyd 1993) (Figure

Figure 4.7 Isolation of zebrafish *WT1*

a Agarose gel electrophoresis of zebrafish PCR products. RT-PCR and PCR from genomic DNA were tried using two annealing temperatures and a range of different primer combinations. Multiple bands were obtained with genomic DNA but none with RT-PCR after 30 rounds of amplification.

b Identification of *WT1* specific bands.
Autoradiograph of PCR products after hybridisation with *AWT1* at low stringency. Digestions of *WT1* clones were also run on the gel for comparison of the level of hybridisation to the probe (1.5hr exposure).
→ identifies two bands showing strong hybridisation to *AWT1*, but were only just visible on the ethidium bromide stained gel above.

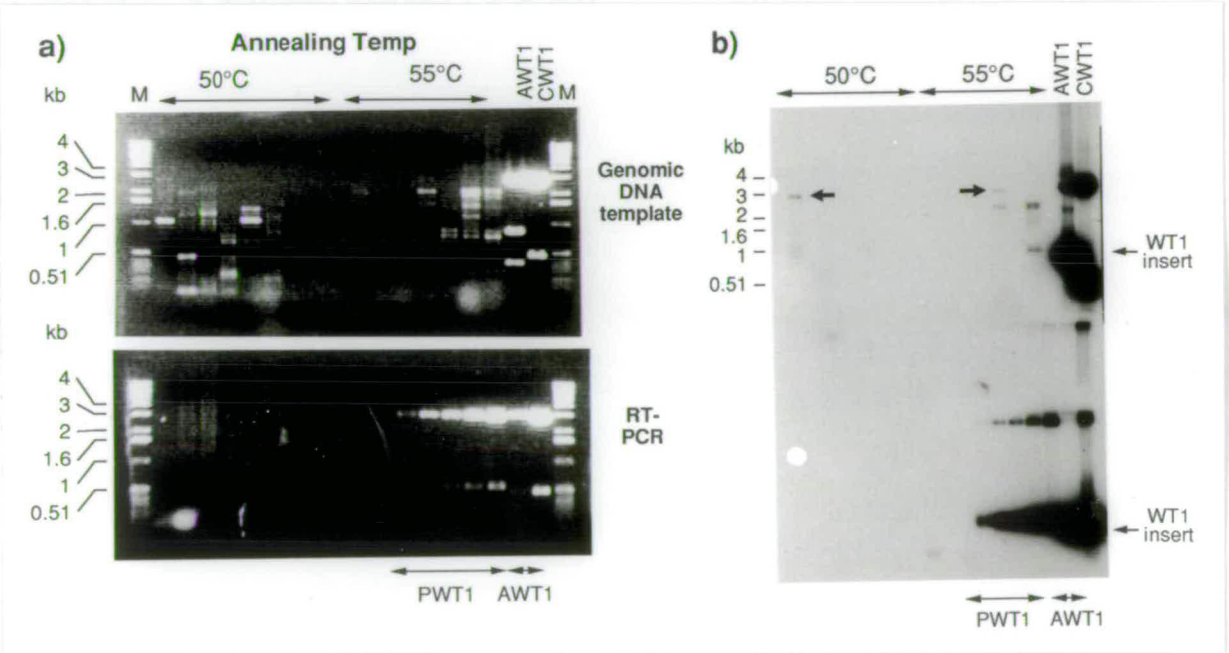
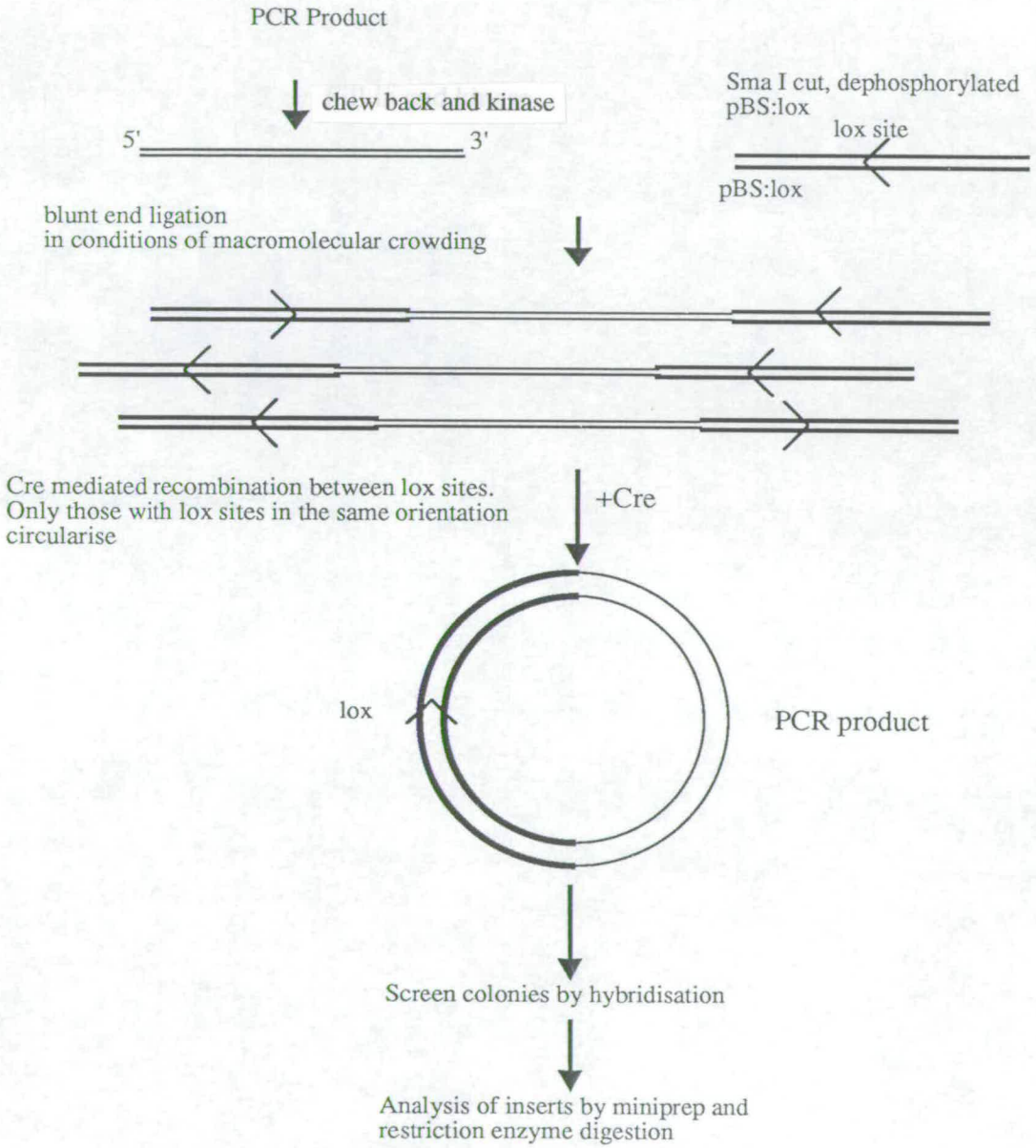


Figure 4.8 Turbo cloning of PCR products

The ragged ends of the PCR products were chewed back and the 5' ends kinased. These were ligated to SmaI digested pBS:lox in conditions of macromolecular crowding. To the ligation products Cre enzyme was added which catalysed recombination of the lox sites leading to circularisation of molecules where the lox sites were in the same orientation. After transformation, white colonies were screened for *WT1* inserts by low stringency hybridisation to A*WT1* and positive colonies were further analysed by restriction digest of the plasmid DNA.



4.8). This involves blunt ended ligation in conditions of macromolecular crowding into pBS:lox, where the ends of the insert can rapidly independently ligate to vector molecules. Circularisation is achieved afterwards by utilisation of the Cre:lox recombination system of P1 bacteriophage. Cre enzyme is added and recognises lox sites in the vector, and those with lox sites in the same orientation are circularised. After electroporation into XL1-Blue cells recombinants were screened for inserts with the blue/white selection system and for *WT1* specific inserts by hybridisation to *AWT1* at low stringency. About 10% of white colonies were found to contain inserts that hybridised to the probe. These were analysed by purification of the plasmid by 'Magic Minipreps' (Promega) and restriction digest. 50% of these had the same restriction pattern and the inserts were the expected size. Sequencing of the inserts using vector primers showed the presence of the PCR primers (C228/C229) at either end of the insert, confirming these as full length products.

From comparison with the human genomic structure it was expected that the clones would contain sequences homologous to the three exons 7, 8 and 9 encoding ZF1-3. Sequencing of the genomic clones revealed three regions of homology corresponding to these exons. These, as expected, were separated by unconserved sequences when compared with the human genomic sequence (Gessler *et al.,* 1992). The AT richness and repetitive nature of the sequences suggested that they were indeed introns. The regions of homology were found to extend just outside the exons to include the splice donor and acceptor sequences. From this it was predicted that the zebrafish *WT1* cDNA had a similar exon structure to human, although the actual sizes of the introns were completely diverged (Figure 4.9). The 60bp between primers B620 and C229 was amplified by heminested PCR using B620 and a ZF2 primer designed from the sequence of the cloned region. The product was sequenced directly (Winship *et al.,* 1989) and found to correspond to the part of ZF1 expected.

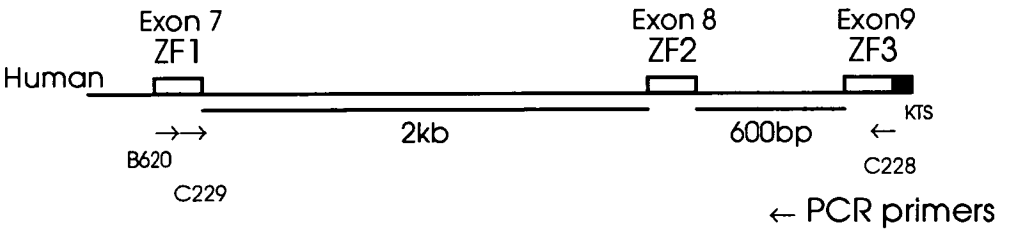
4.2 Isolation by library screening

4.2.1 Introduction

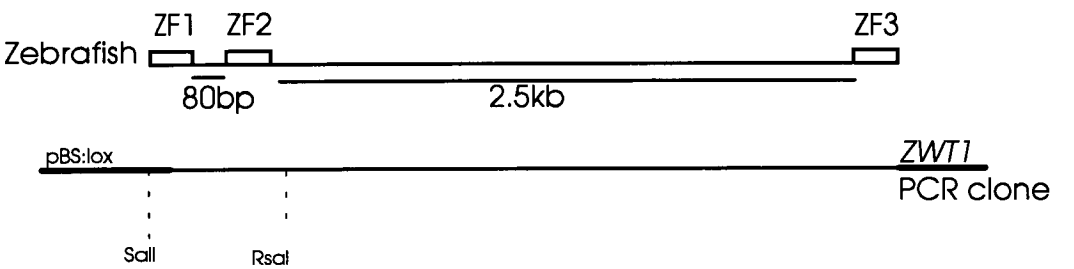
PCR is a fast and relatively easy method for obtaining conserved regions of a gene. From the sequence data obtained above it was clear that *WT1* has not evolved quickly within the vertebrates, especially within the zinc fingers (Table 4.3). This is consistent with mutation analysis of *WT1*, where almost all mutations found in *WT1* affect the zinc fingers. To isolate more of the coding sequence, to determine the extent and pattern of conservation within the proposed transregulatory domain,

Figure 4.9
Comparison of human and zebrafish *WT1* genomic structures in the zinc finger region

a Human genomic structure showing positions of primers producing *WT1* specific amplification in zebrafish.



b Genomic structure of zebrafish *WT1*



Comparison with the human intron/exon structure shows conservation of the organisation of the exons but not the intron sizes in the PCR clone *ZWT1*. The SalI/RsaI fragment used to probe the cDNA libraries is shown.

several strategies could be used. Library screening was chosen in preference to PCR based techniques such as 5' RACE (Rapid amplification of cDNA ends) because of the difficulty experienced amplifying this GC rich region in humans and the possibility of contamination. The major disadvantage to cross species isolation by library screening had been removed by the isolation of homologous probes.

4.2.2 Library screens

cDNA libraries in bacteriophage vectors were obtained either from the embryonic stages most likely to be expressing *WT1* mRNA or from adult testis (Table 4.5). These libraries were titred on the relevant host cells, and approximately $4x10^5$ clones screened following the recommended method supplied by Stratagene for λZAPII. Cloned PCR products were used to probe the library of the same species under high stringency conditions. Clones giving positive signals were purified with further rounds of screening. On isolation the phage were amplified and either converted into phagemids by *in vivo* excision (λZAP type vectors) or the inserts were subcloned into pBluescript SKII+ at the EcoRI site (λgt11 type vectors). The isolates were then sequenced. The library from the striped-faced dunnart (*Sminthopsis macroura*), a marsupial that resembles a mouse, was initially screened by S. Wilcox at La Trobe University, Australia with the human *WT1* probe (WT33) at reduced stringency. Five positives from this primary screen were then analysed for *WT1* by PCR directly from the bacteriophage and a single clone isolated by hybridisation at reduced stringency to *AWT1*.

4.2.3 *WT1* cDNA clones isolated

The results of the library screens are summarised in Table 4.5 and Figure 4.10. The percentage similarities to mouse and human *WT1* are shown in Table 4.6, calculated using the gap program of GCG[1].

A unique clone was isolated from the chick (C2.1), the alligator (AL1), and the marsupial mouse (Sc41). The three partial cDNA clones between them cover the entire coding region, the entire 3' non coding region and probably most of the 5' non coding region, with large amounts of overlap between them. The alligator and chick clones were found to contain the regions isolated by PCR and their sequence was found to be identical to the consensus sequences obtained from multiple PCR clones.

[1]All sequences are contained in Appendix A

Table 4.5 Summary of libraries screened for *WT1* and the clones isolated

Species	Taxon	Vector for Library construction	Source of cDNA	Clone Name / Size	Isolated Region covered	Source
Alligator (*Alligator mississippiensis*)	Reptile	λ ZAP II (Stratagene custom made)	Embryonic - Stage 20	AL1 / 1.3kb	Exon 2 to 3' untranslated region	Paul Sharpe University of Manchester
Chick (*Gallus domesticus*)	Bird	λgt11	Embryonic -Stage 20	C2.1 / 1.6kb	5' untranslated region to exon 9	Nick Platt University of Oxford
Frog (*Xenopus laevis*)	Amphibian	λgt10 λgt11	Neuralation - Stage 17 Embryonic - Stage 22-24	None None		Doug Melton Harvard University, USA Igor Dawid NIH, USA
Marsupial Mouse (*Sminthopsis macroura*)	Marsupial	λ ZAP	Testis	Sc41 / 2kb	Exon 3 to poly A tail	Stephen Wilcox, Jenny Marshall Graves La Trobe University, Australia
Zebrafish (*Brachydanio rerio*)	Fish	λ ZAP II λ ZAP II	Adult Post somitogenesis	Z1a / 3.5kb None	Genomic Intron 6 to 9	David Grunwald University of Utah, USA

Figure 4.10 Diagram of *WT1* clones isolated by library screening, in relation to the human *WT1* gene.

The presence of alternatively spliced sequences within the clones is indicated.

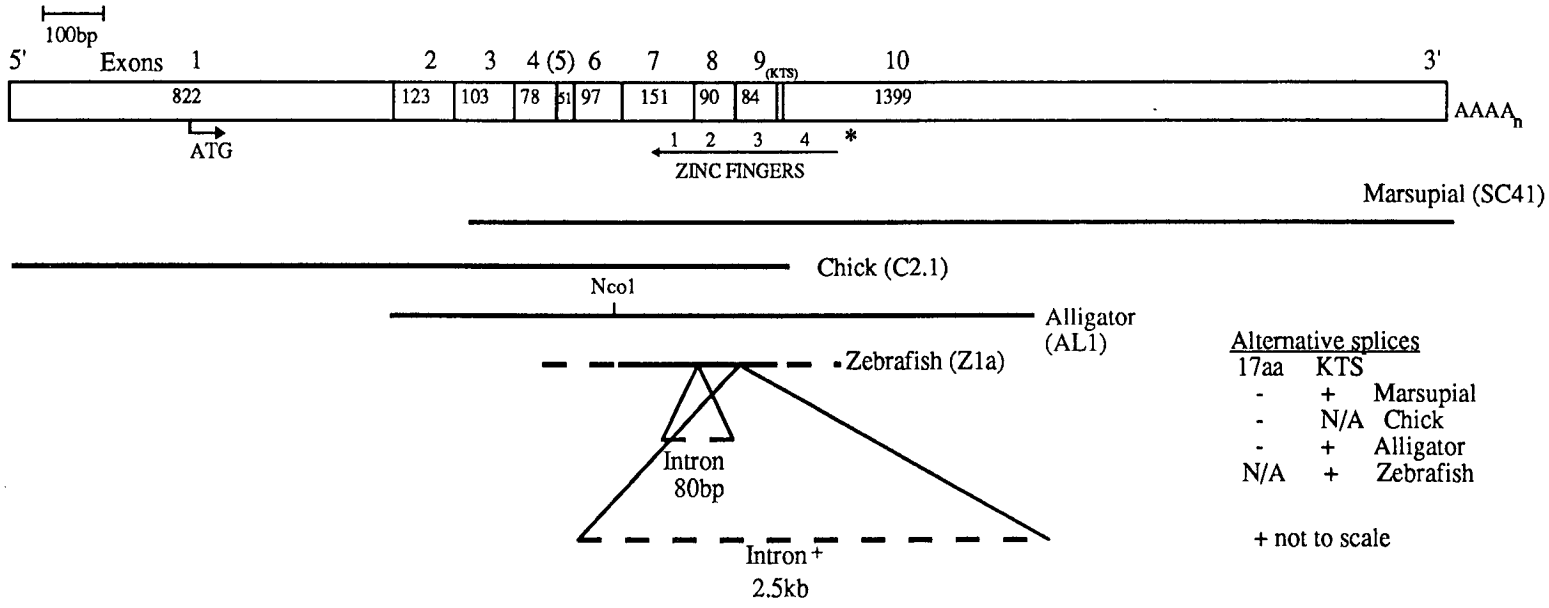


Table 4.6 Similarity of clones isolated from libraries to human and mouse *WT1* at the nucleotide and amino acid level.
Figures were calculated using the Gap program from GCG.

Species	Nucleotide % Identity		No. Amino Acids Compared	Amino Acid % Similarity	
	Human	Mouse		Human	Mouse
Marsupial Mouse	84.8	84.8	240	96.7	95.0
Chick	83.4	82.5	390	93.8	92.3
Alligator	84.0	84.0	289	95.8	95.8
Zebrafish (predicted cDNA sequence)	80.0	79.8	112	94.5	93.6

Although approximately $1x10^6$ clones were screened from both *Xenopus* cDNA libraries and rescreened at reduced stringency with the non finger region of the alligator clone (NcoI/SalI restriction digest fragment of AL1 covering exons 2-6, Figure 4.10) no positives were found.

4.2.4 Analysis of zinc finger 4 in chick

On sequencing the 3' end of C2.1 it became apparent that the 3' 100bp following ZF3 did not correspond to ZF4 as would have been expected. The sequence encoded a termination codon four residues downstream from the end of ZF3 (Figure 4.11a). The sequence did not correspond to any sequences in the GenBank or EMBL databases and so was not part of the bacteriophage vector accidentally included during subcloning.

This and the previous inability to amplify chick *WT1* with a ZF4 primer B298 (Section 4.1.6) led to doubts about the conservation of ZF4 in the chick. To test this another ZF4 primer (D502) was designed using the closely related alligator sequence (Figure 4.11a). PCR with a chick cDNA template showed specific amplification of the expected size band (Figure 4.11b). The PCR product was directly sequenced and was found to encode the 'missing' zinc finger. That this product was derived from chick and no other contaminating species was verified by comparing the sequence of ZF3 from the PCR product. The sequences of the chick PCR clones, C2.1 and the PCR product were found to be identical. On the sequencing gel it was found that once the boundary from ZF4 to ZF3 was crossed two sequences became superimposed on each other (Figure 4.11ci). This was found to be due to the presence of a 9bp insertion encoding the KTS alternative splice in some transcripts (Figure 4.11cii & iii).

The origin of the non ZF4 sequence is not clear but that it commences exactly at the position of a splice site points to it being due to an aberrant splicing event. The possibility that it is due to the intron at this position not being excised was considered. In human the first 9bp of this intron encode the KTS alternative splice. It has been shown that this sequence is also present in chick *WT1* transcripts but no such sequence was present in C2.1. This suggests that the unknown sequence is not this intron. Clone C2.1 could have been produced from contaminating genomic DNA, or from a pseudogene from which ZF4 had been lost. The presence of a pseudogene in the chick genome was investigated. Figure 4.11d shows a zoo blot of EcoRI genomic digests probed at high stringency with *CWT1*. Only one band was detected,

Figure 4.11 Isolation of ZF4 from chick

a Nucleotide and amino acid comparison of C2.1 and human *WT1* ZF3&4 showing a high level of similarity in ZF3 but not in the region expected to be ZF4.

b PCR amplification of chick ZF4
The three 3' primers were used to try to amplify *WT1* from chick cDNA in combination with 5' primer C227 (located in exon 6).
In accordance with earlier results amplification was seen with C228 (ZF3) but not B298 (ZF4), in addition amplification of a band the same size as the alligator was seen with the new ZF4 primer D502. This suggested that the unidentified sequence in C2.1 was a cloning or splicing artefact and that ZF4 is conserved in chick.
(D502 5' CATGTTGTGATGACGAACTAA(C/T)TC(A/G)TC 3')

ci Direct sequencing of the chick cDNA PCR product with D502
A single sequence is seen until the ZF3/4 boundary is crossed, and then the amount of cross banding increases. This was found to be due to two sequences superimposed on each other.
ii&iii Dissection of the two sequences showed that they were both identical to C2.1 in the ZF3 region and that they correspond to the + and - KTS forms of the mRNA, owing to the insertion of the 9bp at the exon 9/10 boundary. The splice site between ZF3&4 was also seen to be the position where the homology between C2.1 and *CWT1* ended indicating the probability that the mRNA transcript from which C2.1 was made was aberrently spliced.
← indicates the positions of the splice sites. * indicates the position of a mismatch between the chick sequence and primer B298 which probably inhibited amplification with this primer.

d PhosphorImager analysis of a zoo blot probed at high stringency with C*WT1* (PCR clone). Only a single band was detected indicating that *WT1* is single copy in the chick genome and that C2.1 could not have been derived from a pseudogene.
(B, bovine, C, chick, A, alligator)

a

Chick/Human

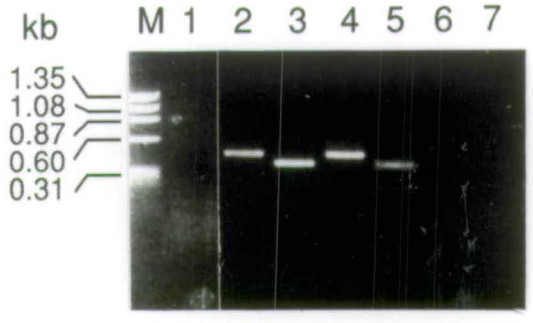
```
GATCATCTGAAGACTCATACCAGGACTCATACAGCTA....TAGTGAAATGAACATCATGGCCTCAGTCCACAAATGTCAACTCCATCTACTGTGTTCACTGAAAAAGGGCCTACAA
||  ||  ||||||||  ||  ||||||||||||||||  ||     |||||||  |  | |||           |  |  |||||     |  |     | |||    ||   | ||  ||
GACCACCTGAAGACCCACACCAGGACTCATACAGGTAAAACAAGTGAAAAGCCCTTCAGCTGTCGGTGGCCAAGTTGTCAGAAAAAGTTTGCCCGGTCAGATGAATTAGTCCGCCAT
D  H  L  K  T  H  T  R  T  H  T  A  I     V  K  *  T  S  W  P  Q  S  T  N  V  N  S  I  Y  C  V  H  *  K  R  A  Y  K
D  H  L  K  T  H  T  R  T  H  T  G  K  T  S  E  K  P  F  S  C  R  W  P  S  C  Q  K  K  F  A  R  S  D  E  L  V  R  H
```

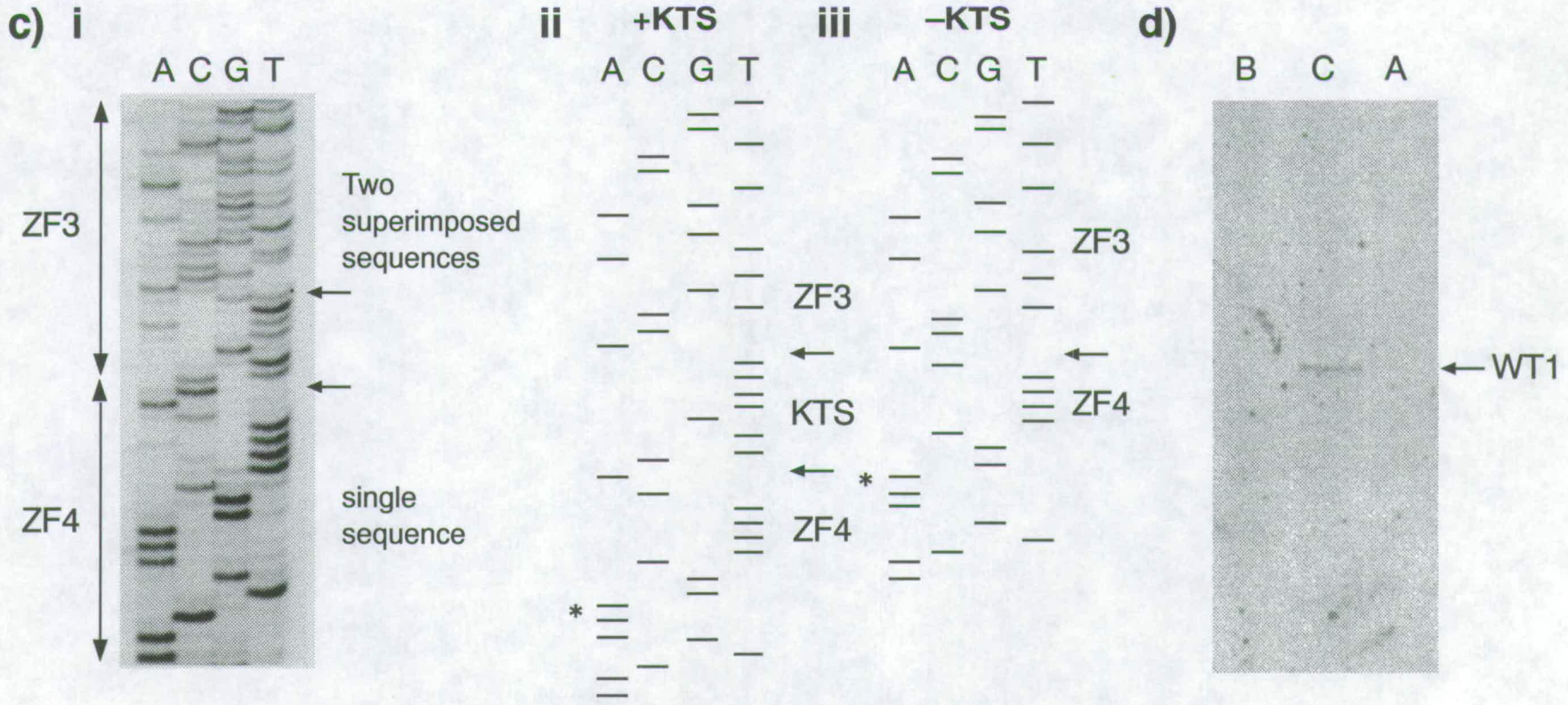
ZF3 | Human ZF4/Chick unidentified sequence

3' Primers ←C228 ←B298 ←D502

b

Lane	PCR Template	3' Primer
1	Chick cDNA	B298
2	Chick cDNA	D502
3	Chick cDNA	C228
4	Alligator cDNA	D502
5	Alligator cDNA	C228
6	No DNA (negative) control	D502
7	No DNA (negative) control	C228




c) i
A C G T
ZF3
ZF4
Two superimposed sequences
single sequence
ii +KTS
A C G T
ZF3
KTS
ZF4
iii –KTS
A C G T
ZF3
ZF4
d)
B C A
WT1

demonstrating the *WT1* gene to be single copy in the chick genome.

4.2.5 Isolation of a zebrafish *WT1* genomic clone (Z1a)

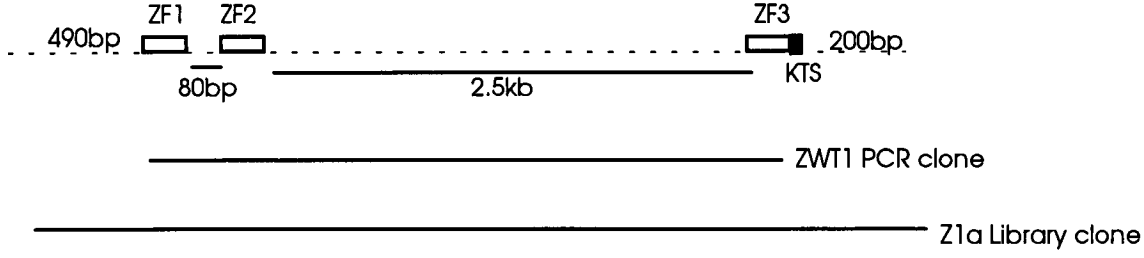
To screen the zebrafish libraries a SalI/RsaI fragment from a zebrafish genomic PCR clone was used that contained part of ZF1, ZF2 and the 80bp intron between them (Figure 4.9b). This excluded the long 2.5kb intron between ZF2 and ZF3. Two cDNA libraries were screened, a postsomitogenesis one and an adult one. One positive was obtained; from the adult library. On sequencing the isolated clone (Z1a) it was found to contain interruptions between the zinc finger encoding sequences identical to the predicted introns in the genomic PCR clones *ZWT1*. This could either be because introns had not been removed due to partial splicing of the pre-mRNA transcript or that the clone was generated from contaminating genomic DNA in the RNA preparation. The presence of four introns within the clone favoured the latter possibility (Figure 4.12a). The clone extends 5' and 3' of the PCR clone containing the whole of exons 7, 8 and 9 and therefore includes the region of intron 9 which encodes the KTS alternative splice in humans. In zebrafish the two alternative splice donor sites are very highly conserved suggesting it is very likely that both + and -KTS isoforms are expressed in fish (Figure 4.12b). Therefore genomic organisation of *WT1* with each zinc finger in a separate exon has been conserved through vertebrate evolution. In contrast the sizes of the introns are completely different; introns 7 and 8 being 2kb and 600bp respectively in human (Haber *et al.*, 1991) whereas in zebrafish they are 80bp and 2.5kb (Figure 4.9).

4.3 Discussion

PCR and library screening was an effective method for isolating a number of different *WT1* orthologs. Less sequence was isolated than was desired especially in the transregulatory domain. This was firstly due to PCR in this region not being successful, either because the sequence variation was too great for the primers used, or the GC content was too high. Secondly 5' incomplete clones were obtained from the library screens. This could be a result of the reverse transcription step not producing full length DNA; reverse transcriptase can be inhibited due to the presence GC rich regions and RNA secondary structure. The first clones isolated for mouse and human also lacked part of the 5' region.

Figure 4.12 Genomic structure of zebrafish *WT1*

a Comparison of region covered by the PCR and library clones. Z1a contains 460bp and 200bp of the introns flanking exon 7 (ZF1) and exon 9 (ZF3) an includes the region encoding the KTS alternative splice in human.



b Comparison of the sequence of Z1a with the human genomic sequence showing the conservation of the region containing the KTS alternative splice, indicating that there is a high probability that the inclusion of this alternative splice is conserved across 400Myr.

```
SPLICE SITE CONSENSUS                                  A/CAGGTAAGT
                                                                 A/CAGGTAAGT
ZEBRAFISH GTTTTCACGTTCAGACCACCTTAAGACCCACACGCGGACACATACAGGTAAAACAAGTGCGTAAACCTTTTCATTTTTTTCATGATTCGCTCCTCTCTT
          ||| || || || ||||||||||||||||||||  |||| |||||||||||||||||||||||||| |||       |||   | |       | | |
HUMAN     GTTCTCCCGGTCCGACCACCTGAAGACCCACACCAGGACTCATACAGGTAAAACAAGTGCGTAAACTTTTCTTCACATTTATTTTTCATTATTTTTTTA
          ____________________________________________|_______|_____________________________________________
                                EXON9/ZF3                 KTS                    INTRON
```

Much time was wasted because of PCR contamination. Cross species PCR is more prone to contamination than homologous PCR for several reasons.

i) Primers are designed from known and therefore amplified or cloned sequences. These are likely be more similar to the primers than the desired template and so in the first few rounds of amplification these will be preferentially amplified. This means that even if the template is initially in excess of the contaminant the final product may contain enough contaminant to obscure the real product or take over the reaction.

ii) If only one species is involved the presence of a contaminant can be assayed for by high stringency hybridisation with the suspected contaminant, but where many species are involved it may not be known which is the contaminant.

The precautions advised to avoid contamination were followed (Clackson *et al.*, 1991). It was found that consistent contamination free PCR was only possible by using a separate room for setting up the reactions.

The reason for failure to isolate cDNA sequences from zebrafish and ascidian (tunicate) by PCR may be because of a restriction of high levels of *WT1* mRNA to certain structures. In whole embryo preparations *WT1* message might be at too low a concentration for cross species PCR. This seems to have been the case for *Xenopus* as amplification was successful only with RNA isolated from the mesonephros (Section 4.1.6). There are no vertebrate-like kidneys present in the ascidian from which RNA could be isolated. The gonads may be a possible source of material but these gonadal structures may have a completely different origin to vertebrate gonads.

That no larger *Xenopus* clone was obtained from cDNA library screens was disappointing. This was maybe because *WT1* was not expressed in the early stages from which the libraries were made. In the mouse the first *Wt1* expression is observed at about 9 days, just before the onset of mesonephric differentiation (Armstrong *et al.*, 1992). In *Xenopus* the first evidence of pronephros differentiation is seen at stage 23 with the slight thickening of the mesoderm corresponding to the pronephric duct rudiment (Nieuwkoop and Faber 1967, Lynch and Fraser 1990). It was not possible to locate libraries later than stage 24. This is probably owing to the major interest in *Xenopus* concentrating in the earliest stages of development. The zebrafish postsomitogenesis library was thought likely to contain *WT1* clones, although none were detected. It may be that the intron within the probe reduced the affinity of the probe for a cDNA to a too large extent. The isolation of the genomic clones has allowed the comparison of *WT1* to be extended to genomic structure.

In *Xenopus* there are probably two copies of *WT1* in the genome, arising from a tetraploidisation event which occurred about 30 Myra (Tymowska and Fischberg 1973). In the three *Xenopus* PCR clones which were sequenced three differences were found in one clone relative to the others. It is not known if these arose from PCR errors or if the two forms were isolated. The number of differences was higher than for the three alligator clones sequenced, where one change was noted in clones twice the length of *XWT1*. If these clones do represent the two copies of *WT1* they are very highly conserved. This could be because substitutions within *WT1* zinc fingers are thought to have dominant negative properties, which has lead to selection for conservation.

The sequence information derived here will enable further species to be cloned more easily and also lead to suggestions about the structure and evolution of *WT1,* which are discussed in the next chapter.

Chapter 5

Structural and evolutionary comparison of *WT1* orthologs

5.1 Introduction

The sequence information from the *WT1* orthologs obtained can be used to analyse the conservation of structural features of the *WT1* protein. Conservation can then be used to make inferences about the structural and functional importance of these regions. This can be especially useful when combined with functional studies of the gene. The sequences can also be used to investigate the evolutionary changes that have occurred, such as the rate of change of both the nucleotide and the encoded amino acid sequence.

In the following discussion it was found that the most convenient way in which to define the regions referred to was to use the exon structure of the gene. Knowledge of the exon structure is based entirely on the information from the human and zebrafish, as no information has been as yet been obtained for other species. In addition all positions in the amino acid and nucleotide sequence refer to numbering of the human +17aa, +KTS isoform.

5.2 Structural analysis

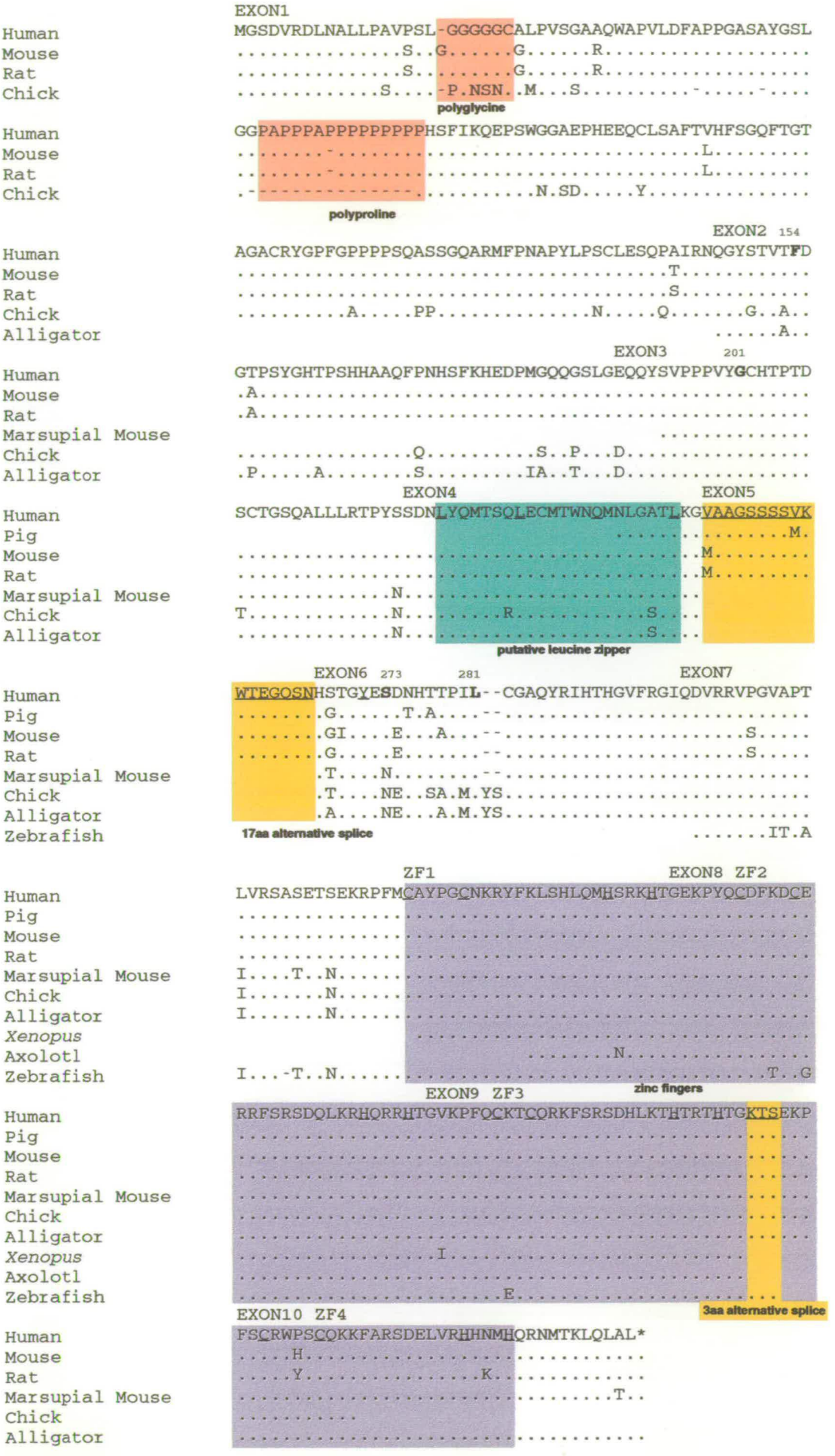
5.2.1 Sequence motifs

5.2.1.1 Structure prediction

The nucleotide sequences of *WT1* orthologs were translated to produce the predicted amino acid sequence of the protein (Figure 5.1). Most of the amino acid changes fall into the category of frequently accepted, conservative mutations which are due to single base pair mutations when compared against those in Dayhoff's matrix (Dayhoff *et al.*, 1972). Therefore since multiple changes have not occurred, mutations which result in an accepted change must occur infrequently. The sequences were analysed using the motifs program from GCG, but no other sequence motifs were found in addition to the zinc fingers. The hydrophobicity profiles (Kyte and Doolittle 1982) and secondary structure predictions (Chou and Fasman 1978) of each ortholog were compared and no great changes were observed in the different species (Appendix B). An increase in hydrophilicity was seen in the zinc finger region consistent with it interacting with DNA. Protein structure prediction from the primary amino acid sequence alone is still in its infancy. Only a low level of accuracy is obtained with the Chou and Fasman prediction (Creighton 1984) and so a newer and more reliable program was tried. This was Predict Protein (Sander and Schneider 1991) which relies on sequence comparison to predict a structure. The only region

Figure 5.1 Comparison of the predicted amino acid sequences of the *WT1* orthologs

The human exon structure is shown as are the positions of residues discussed in the text (bold). The alternative splices and known sequence motifs are highlighted.


EXON1
Human
Mouse
Rat
Chick
MGSDVRDLNALLPAVPSL-GGGGGCALPVSGAAQWAPVLDFAPPGASAYGSL
polyglycine
GGPAPPPAPPPPPPPPPHSFIKQEPSWGGAEPHEEQCLSAFTVHFSGQFTGT
polyproline
EXON2 154
AGACRYGPFGPPPPSQASSGQARMFPNAPYLPSCLESQPAIRNQGYSTVTFD
EXON3 201
GTPSYGHTPSHHAAQFPNHSFKHEDPMGQQGSLGEQQYSVPPPVYGCHTPTD
Marsupial Mouse
Pig
EXON4
EXON5
SCTGSQALLLRTPYSSDNLYQMTSQLECMTWNQMNLGATLKGVAAGSSSSVK
putative leucine zipper
EXON6 273 281
EXON7
WTEGQSNHSTGYESDNHTTPIL--CGAQYRIHTHGVFRGIQDVRRVPGVAPT
17aa alternative splice
ZF1
EXON8 ZF2
LVRSASETSEKRPFMCAYPGCNKRYFKLSHLQMHSRKHTGEKPYQCDFKDCE
zinc fingers
EXON9 ZF3
RRFSRSDQLKRHQRRHTGVKPFQCKTCQRKFSRSDHLKTHTRTHTGKTSEKP
3aa alternative splice
EXON10 ZF4
FSCRWPSCQKKFARSDELVRHHNMHQRNMTKLQLAL*

for which a comparison was obtained was the zinc fingers for which we already have structural information.

5.2.1.2 Post-translational modification

To investigate whether *WT1* protein may be post-translationally modified, searches were made for recognition sequences for N-linked glycosylation (Doolittle 1986) and phosphorylation (Kemp and Pearson 1990) sites. No putative glycosylation sites were found in any species, but putative phosphorylation sites were found. Recognition sites found included casein kinase I and II, proline dependent kinase and cAMP dependent kinase. Caution must be observed when drawing conclusions from the presence of recognition sites as it is known that not all sites are recognised and there can be synergy between different sites. Although these sites are present, no phosphorylation of *WT1* has been observed (Morris *et al.,* 1991). The conservation of these sites was found to be variable with some being present in all species and some in only one.

5.2.1.3 Exon structure

To investigate if the exon structure of *WT1* is conserved in all vertebrates will require the isolation of genomic sequences. Comparison of the sequences at the known human splice sites in the cDNAs isolated can only give limited information because the most invariant splice site sequences occur in the introns. The first and last two bases of an exon tend to be GT and AG. This is the case in some of the human *WT1* splice sites, but most are at least purine rich. Table 5.1 shows that most of these sites are conserved. Substitutions which have occurred increase the similarity to the consensus sequence, indicating that the exon structure may well have been conserved, as seen for exons 7-9 in the zebrafish.

5.2.2 5' Non coding region

Multiple transcription start points have been reported in the *WT1* promoters from mouse and human, approximately 500-400bp upstream of the predicted initiation codon (Pelletier *et al.,* 1991c, Hofmann *et al.,* 1993). Within the 5' non coding region there is still over 80% nucleotide sequence identity between mouse and human (Table 5.2). Alignment with the chick reveals a lower level of conservation but several regions with a higher than background level of identity. These may correspond to protein binding sites (Figure 5.2). *WT1* has been shown to bind this region in mouse (Rupprecht *et al.,* 1994) but since the true DNA consensus sequence for *WT1* is unknown comparison cannot reveal whether binding is conserved in chick. The C+G

Table 5.1 Conservation of splice sites in *WT1*

The sequence of the exons at the splice donor and acceptor sites in human *WT1* is compared with the consensus sequence and the sequence in the *WT1* orthologs.

Exon	Species	Splice acceptor sequence	Splice donor sequence
		GTA/GAGT	A/CAG
1	Human		GCAATCAGG
	Mouse		----C--A-
	Rat		----C--A-
	Chick		----C----
2	Human	GTTACAGC	GCTCGCTGG
	Mouse	-A------	---------
	Rat	-A------	---------
	Chick	-----G--	CA--T--A-
	Alligator	--------	CT--C--A-
3	Human	GTGAGCAG	CTACAGCAG
	Mouse	-C------	---------
	Rat	-C------	---------
	Marsupial mouse		-----A---
	Chick	-G--C---	-----A---
	Alligator	-A--C---	-----A---
4	Human	TGACAATT	CTTAAAGGG
	Mouse	--------	---------
	Rat	--------	---G-----
	Marsupial mouse	-------C	AC-G-----
	Chick	--------	GC-G-----
	Alligator	--------	A--------
5	Human	AGTTGCTG	GCAGAGCAA
	Mouse	-A-G----	---------
	Rat	-A-G----	---------
	Pig	--------	---------
6	Human	CCACAGCA	GGCATTCAG
	Mouse	----G-T-	---------
	Rat	----G---	---------
	Marsupial mouse	-----C--	--A--A--A
	Chick	---T-CG-	-----A--A
	Alligator	---TGCA-	--A--A--A
7	Human	GATGTGCG	AGCACACTG
	Mouse	--------	---------
	Rat	--------	---------
	Marsupial mouse	-----C--	----T----
	Chick	-----C--	---------
	Alligator	-----C--	---------
	Xenopus		-A-----A-
	Zebrafish	--------	-A-----A-
8	Human	GTGAGAAA	GACATACAG
	Mouse	--------	----C----
	Rat	--------	----C----
	Marsupial mouse	--------	----C----
	Chick	----A---	----C----
	Alligator	----A---	----C----
	Xenopus	-G--A---	-G--C----
	Zebrafish	-G------	----C----
9	Human	GTGTGAAA	CTCATACAG
	Mouse	--------	---------
	Rat	--------	---------
	Marsupial mouse	--------	---------
	Chick	--------	---------
	Alligator	--------	---------
	Xenopus	--A-C---	---------
	Zebrafish	-C--T--G	-A-------
9 (KTS)	Human		GTAAAACAA
	Mouse		---------
	Rat		---------
	Marsupial mouse		---------
	Chick		---------
	Alligator		---------
	Zebrafish		---------
10	Human	GTGAAAAG	
	Mouse	--------	
	Rat	--------	
	Marsupial mouse	--------	
	Chick	--------	
	Alligator	--------	

content of this region is high (over 70%) in all species; this drops along the sequence falling to 50% in exon 4. The percentage of CpG dinucleotides is normally under-represented in the vertebrate genome due to deamination of methylated cytosine (Bird 1986). In exon 1, CpG dinucleotides have been retained in all species. This would indicate the conservation of the mammalian CpG island in chick and probably transcriptional control of this region.

5.2.3 The predicted initiation codon is conserved

The clone C2.1 (chick cDNA clone - Figure 4.10) extended 370bp 5' to the predicted mammalian methionine initiation codon. Although it has not been demonstrated *in vivo* that this is the initiation codon the protein observed on Western blots is of the predicted size (Morris *et al.*, 1991, Telerman *et al.*, 1992). Table 5.2 compares the nucleotide identity between species in the coding and non coding regions of exon 1 and shows that the identity between all species increases immediately downstream of the putative initiation codon consistent with this being the initiation codon. The Kozak sequence is the optimum sequence required for initiation of translation in eukaryotes (Kozak 1987). It has been found that this is not the most frequent sequence at translation start sites and A/GNC most commonly precedes the initiation codon (Cavener and Ray 1991). This is not the case in all the *WT1* orthologs, but this sequence is found two bases further upstream, probably making it a favourable site of translation initiation (Table 5.3). Any upstream methionine initiation codons that are present in human, mouse and chick transcripts are followed by downstream termination codons indicating no N-terminal extended version of *WT1* can be produced.

5.2.4 Transregulatory domain

5.2.4.1 Homopolymer domains

The transactivating/repressing activity of *WT1* has been located to the N-terminal 324 amino acids (exons 1-7) using domain swap experiments (Madden *et al.* 1991). In the amino acid comparison the two most striking inter-species changes in this domain are those affecting two homopolymer sequences $(Proline)_{13}$ (54-68aa) and $(Glycine)_5$ (19-23aa) (Figure 5.1). The lack of conservation of these regions is especially noticeable in the context of rest of the transregulatory domain where only a few scattered amino acid changes are seen (Figure 5.1).

Figure 5.2

Multiple sequence alignment of the non coding region of exon 1 using CLUSTAL V

Sequence footprinted by both -/+KTS isoforms are indicated (overlined) as well as regions of greater than average similarity between the three species (highlighted).

```
                          _____________WT1 binding_____________________________
HUMAN     AGTG--AATGGAGCGGCCGAGCCTCC-TGGCTCCTCCTCTTCCCCGCGCCGCCGGCCCCT
MOUSE     TGTGTGAATGGAGCGGCCGAGCATCC-TGGCTCCTCCTCCTTCCC-TGCTGCCGGCCCCT
CHICK     CG-GCACG-AGGGCGCTGAGGCGGCTGTGGGTCCACCCCCTCCCTAT--------CCCCC
           * *       * ***      **   *  *** *** ** * * **            ****

          ____________________________
HUMAN     CTTATTTGAGCTTTGGGAAGCTGAGGGCAGCCAGGCAGCTGGGGTAAGGAGTTCAAGGCA
MOUSE     CTTATTTGAGCTTTGGGAAGCTGGGGGCAGCCAGGCAGCTGGGGTAAGGAGTTCAAGGCA
CHICK     C---------CCCCGGCACGC-----GCGGACAGACA----------------CACGGC-
          *          *   ** * **     ** * *** **                  ** ***

HUMAN     GCGCCCACACCCGGGGGCTCTCCGCAACCCGACCGCCTGTCCGCTCCCCCACTTCC----
MOUSE     GCGCCCACACCCGGGG-CTCTCCGCAACCCGACCGCCTGCCTGCCTCCCCCTTTCCTTTT
CHICK     -CGCACACAC----GGACACGCGGCCA--CGGACACGCGCGCACC---------------
           *** *****    ** * * * ** *  **  * *  *    *
                    __________WT1 binding_____________________
HUMAN     --CGCCCTCCCTCCCACCTACTCATTCACCCACCCACCCACCCAGAGCCGGGACGGCAGC
MOUSE     TTCCCCGCCCCTCCCTCCCACCCACTCATTCACCCACCCACCCAGAGAGAGGACGGCAGC
CHICK     ----CCACCCTGCGCTGCGAGCGG-------GCACAGC--CCCGGAG--------GCGAT
              **  **  * *  * *             * ** *  *** ***         **

HUMAN     CCAGGCGCCCGGGCCCCGCCGTCTCCTCGCCGCGATCCTGGACTTCCTCTTGCTGCAGGA
MOUSE     CCAGGAACCCGGGCCC-GCCGCCTCCTCGCCGCGATCCTGGACTTCCTCCTGTCGCAGGA
CHICK     CCTGGAGAACAA-CC------TTTCCCTGT-GCGCTGCTCGCCGCTCCCGCGG-GCTGGA
          ** **    *   **        ***  *  *** * ** * *   * *  *  ** ***

HUMAN     CCCGGCTTCCACGTGTGTCCCGGAGCCGGCGTCTCAGCACACGCTCCGCTCCGGGCCTGG
MOUSE     GCCGGCTTCCACGTGTGTCCCGGAGCCGGCGTCTCAGCACACGCTCCGCCGGGAGCCCGG
CHICK     CCCGGGGAGC---TGCGTCGCGGAGCAAACCTCTC--CAAGGGCTGAGCTCTCCGGCTGG
           ****    *   ** *** ******   * ****  **   ***  **     * * **

HUMAN     GTGCCTACAGCAGCCAGAGCAGCAGGGAGTCCGGGACCCGGGCGGCATCTGGGCCAAGTT
MOUSE     GTGCGTCCAGCAGCCGGAGCAACCTGGGGACCGAGGCCCCCGGAGCGCCTGGGCCAAGT-
CHICK     -TGCCT---GCAGCCGGAGCGTCCGG----CTGGCGGC--AGAGGCAGGTGTCCTGGAAT
           *** *   ****** ****  *  *    * *    *   *  **   **  *

HUMAN     AGGCGCCGCCGAGGCCAGCGCTGAACGTCTCCAGGGCCGGAGGAGCCGCGGGGCGTCCGG
MOUSE     --------------CCAGCGCCGAGAATCCGCAGGATCGCAGGAGCGGAGAACCGTCCGC
CHICK     AAGTGCGGCCTGGGGGGAGACTGGCTGCCC----GGCGGCAGG--CAGAAG---GTCTG-
                          * *     *     *   * ***  * *      *** *

HUMAN     GTCTGAGCCTCAGCAA
MOUSE     ATCCGAGCCGCACCTC
CHICK     ---CGAGC---AG--A
              ****   *
```

Table 5.2

Percentage nucleotide identity between *WT1* orthologs in the non coding and coding regions of exon 1

Non coding

	mouse	chick
human	81.4	45
mouse		47

Coding

	mouse	rat	chick
human	87.9	89.7	81.3
mouse		96.1	81.5
rat			79.0

Calculations performed using the gap program from the GCG

Table 5.3

Comparison of the Kozak sequence with the putative *WT1* initiation codon

	A
Kozak	GCCGCCGCC**ATG**G
Human	CCTCAGCAA**ATG**G
Mouse	CCGCACCTC**ATG**G
Rat	ACCTC**ATG**G
Chick	TGCGAGCAG**ATG**G

The initiation codon is highlighted and the Cavener and Ray (1991) sequence is underlined

---

Table 5.4 Percentage amino acid composition of the transregulatory domain

Species	Exon	% Proline	% Glutamine	% Serine
Human	1	20.5	9.6*	9.5
	2-3	13.8	9.2	13.3
	4	0.0	11.5	3.8
Chick	1	14.7	10.9*	13.4
	2-3	13.8	10.8	10.6
	4	0.0	11.5	7.2
Alligator	2-3	13.8	9.2	13.0
	4	0.0	11.5	7.2
Average for Vertebrates (Doolittle 1986)		5.6	4.3	7.0

* Calculation using only the second half of exon 1

(Proline)$_{13}$ In human exon 1 there is a region of 15 amino acids, 13 of which are proline (the other two being alanine). In mouse and rat this is one proline shorter, but in contrast it is completely absent in the clone C2.1. PCR with primers flanking this region gave a product 54bp smaller from chick genomic DNA and cDNA than from mouse genomic DNA (Figure 5.3a). Direct sequencing of the products confirmed that this was due to the absence of (Proline)$_{13}$ in chick. The presence of this motif was investigated in other mammals besides the primate and rodent orders by PCR. The PCR products in two artiodactyls (pig and cow) were the same size as that in the mouse (Figure 5.3b) and by direct sequencing the proline stretch was found to be present (data not shown).

(Glycine)$_5$ This stretch of five glycines (rat and human), or six (mouse), is also located in exon 1. In C2.1 the predicted amino acid sequence of this region is completely diverged due to point mutations at non-degenerate sites (Figure 5.1). Direct sequencing of this region from the PCR products above, showed in pig and cow that the glycine region was conserved in these species (data not shown).

(Proline)$_4$ A third homopolymer sequence was is also present PPPP (Proline)$_4$ (114-117aa) in exon 1. In the chick this was not 100% conserved, being APPP.

5.2.4.2 The amino acid composition is similar to other transregulatory domains

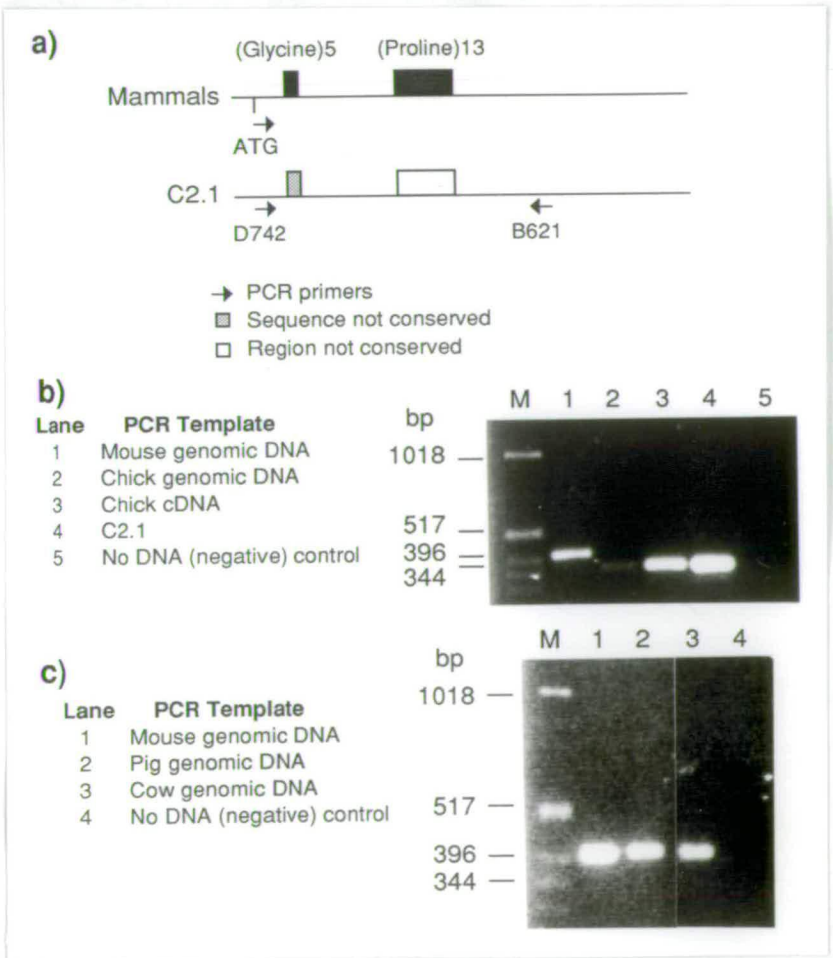
The proportion of proline in exon 1 is reduced to 15% in chick (mostly due to the deletion of (Proline)$_{13}$) as compared with 20% in humans, but this is still higher than the average value for vertebrates of 6.1% (Doolittle 1986) (Table 5.4). This is of importance because the high levels of proline and glutamine are thought to be characteristic of some transactivation domains, although the reason for this is not known (Mitchell and Tjian 1989). Although these levels are conserved in *WT1* and are more than double the average (Table 5.4) they are lower than the proline/glutamine rich regions of other transcription factors. The transactivation domains of *CTF* and human and *Xenopus Sp1* contain 20-30% proline and 25% glutamine respectively (Mitchell and Tjian 1989). *WT1* is more similar to another transcription factor, *Krox 20* (*EGR2*), where the tranactivation domain contain 12% proline and 13% serine. *Krox 20* also contains a conserved proline stretch, in this case it is seven residues long (Sukhatme *et al.,* 1988, Oxtoby and Jowett 1993). The proline and glutamine rich regions of *WT1* overlap but they are not are not identical. The proline rich region extends from exon 1 to 3 (amino acids 1-223), but the glutamine rich region covers the C-terminal half of exon 1 to exon 4 (amino acids 73-248). This may point to a modular arrangement of functions.

Figure 5.3
Investigation into the conservation of homopolymer domains in exon 1

a Schematic comparison of the human, mouse and rat *WT1* amino acid sequence in exon 1 with the sequence derived from the chick clone C2.1. The two homopolymer domains $(Glycine)_5$ and $(Proline)_{13}$ were found not to be conserved in C2.1.

b Verification that the loss of the homopolymer domains is not a cloning artefact. PCR across the two domains with primers D742 and B621 showed that the band obtained with chick genomic and cDNA templates is the same size as from C2.1. These three bands are all approximately 50bp smaller than that amplified from mouse genomic DNA.
(B621 5' GGCTTGCAGGTAGGGCGC(A/G)TTIGG(A/G)AACA 3')
(D742 5' AGATGGGGTCCGACGTCCGGGACC 3')

c Investigation of exon 1 in other mammals. PCR using pig and cow genomic DNA templates shows that the product is the same size as the product from mouse genomic DNA.



5.2.4.3 Mutations

There are two instances where a somatic mutation in a Wilms' tumour occurs in the transregulatory domain. These are a phenylalanine to serine change at codon 154 and a glycine to aspartic acid change at codon 201 (Park *et al.,* 1993b&c, Table 1.3). Both are non conservative changes with an aliphatic residue being converted to a more hydrophilic residue. These changes may affect protein interactions, disrupt the hydrophobic core, or the folding of the protein. The G to D mutation changes the *in vitro* transrepression activity to transactivation. In both cases these two residues were found to be conserved; suggesting an important role in all species so far examined (Figure 5.1). The effect of the exclusion of exon 2 by alternative splicing in Wilms' tumour cell lines is to covert *WT1* into transactivator. Unlike wild type *WT1*, exon 2-less transcripts cannot repress the transformed phenotype of a Wilms' tumour cell line (Haber *et al.,* 1993). This therefore implicates exon 2 in both transcription repression and tumorigenesis, consistent with a high level of evolutionary conservation.

5.2.5 The conservation of putative leucine zipper in exon 4

The proposed leucine zipper motif, defined by the presence of leucines every seven residues, and the absence of helix breaking prolines, is encoded by exon 4. It may extend as far as the first proline in exon 6. Leucine zippers are a subset of helices that are amphipathic. The hydrophobicity of one face of the helix favours the interaction with another such helix on the hydrophobic side, forming a structure called a coiled coil. Figure 5.4a shows chick exon 4 and exon 6 as far as the first proline, drawn in a α helical format (helicalwheel program-GCG). This demonstrates that there is a grouping of hydrophobic residues on one side of the helix, and the amino acid substitutions observed are located outside this. It is also known that the polar residues are important stabilising the contact between the helices (Conway and Parry 1990). This may explain the high level of conservation of the non-hydrophobic residues.

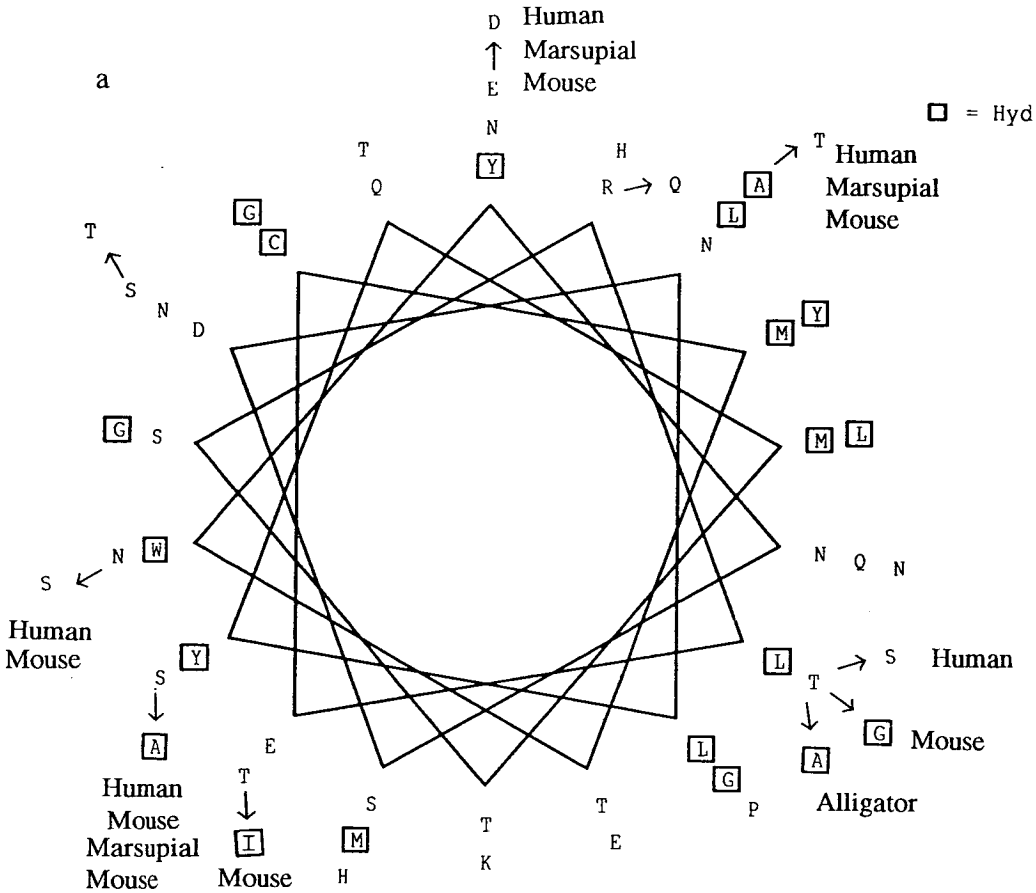
Although many sequences with a heptad repeat of leucines have been proposed to form coiled coil interactions it is doubtful that they all do (Lupas *et al.,* 1991). It is now possible to estimate statistically the likelihood of a coiled coil being formed by comparison with the positions of amino acids with proven coiled coil forming domains using the Coils program (Lupas *et al.,* 1991, and updated 1993, A Lupas, personal communication). The predicted likelihood of a coiled coil forming in this region of the *WT1* protein is very low. This program is known not to detect buried helices or

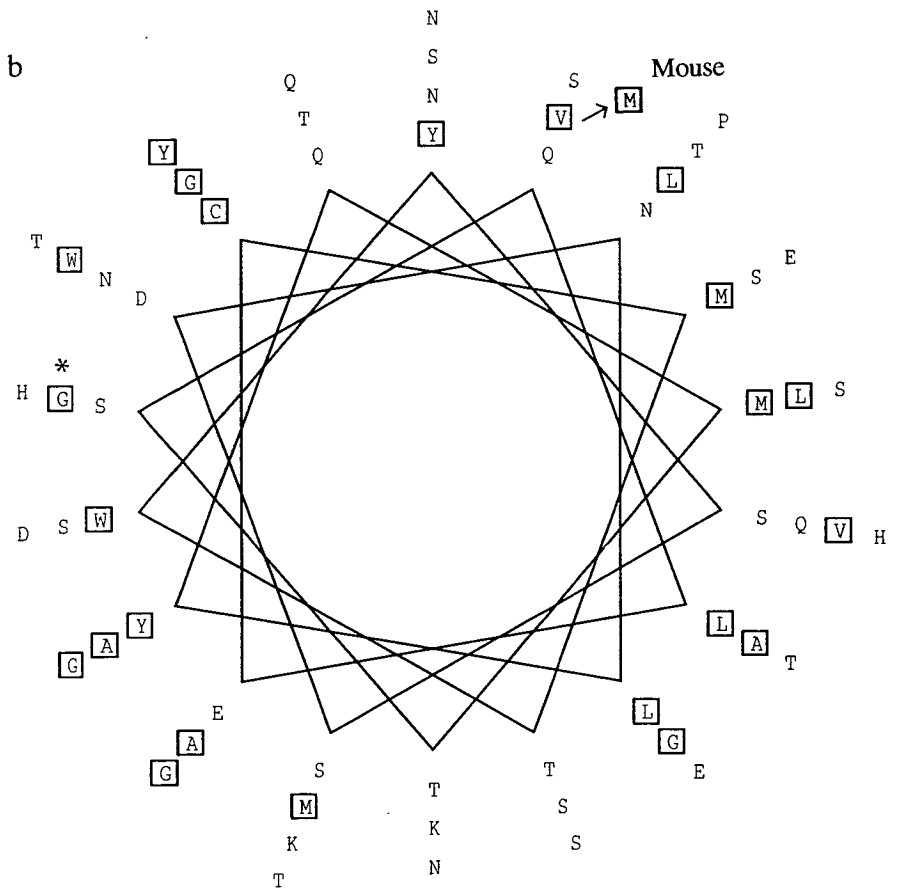
Figure 5.4 Helical representation of the putative leucine zipper

Sequence of chick and human *WT1* covering the putative leucine zipper region, exons 4-6 (221-279aa, human numbering). This was drawn in α helical format, using the helical wheel programme of GCG. Hydrophobic residues are boxed

a In the -17aa isoform (chick *WT1*) there is a grouping of hydrophobic residues and a lack of hydrophilic residues on one side of the helix. This is conserved in all species examined.
Cross species substitutions are indicated for all orthologs except pig. Most variability is located in the outer ring of the wheel which corresponds to exon 6 sequence (* = exon 6 onwards).

b When the 17aa alternative splice is included in mammals the localisation of hydrophobic residues is changed. (Cross species substitutions shown for the alternative splice only)


a
D Human
Marsupial
E Mouse
= Hydrophobic
Human
Marsupial
Mouse
Human
Mouse
Human
Mouse
Marsupial
Mouse
Mouse
Human
Mouse
Alligator


b
Mouse

those in bundles of more than three in antiparallel orientation. However if these types of helices were present in *WT1* their availability for the proposed protein-protein interaction function is probably limited.

5.2.6 Exon 5 - The 17aa alternative splice is not conserved in the archosaur lineage

Exon 5 is an alternatively spliced exon in human, mouse (Haber *et al.,* 1991) and rat (Sharma *et al.,* 1992) and the sequence is conserved in those species. In this study its presence has only been observed in the clones obtained from the mammalian species rat and pig (Figure 4.5). This demonstrates that although the PCR method by which the *WT1* orthologs were isolated did not exclude +17aa isoforms none were obtained from chick or alligator. To investigate whether this alternative splice occurs in the alligator and chick, PCR with primers flanking this region was performed with rat, chick and alligator cDNA template. A band was seen corresponding to the +17aa isoform only in the rat, but in all species the smaller isoform was detected (Figure 5.5). It may be that a low and therefore undetected level of the +17aa isoform is produced. Even this explanation points to a different role to mammals as at least 60% of transcripts in mouse and human are +17aa (Haber *et al.,* 1991).

5.2.7 Divergence and conservation of exon 6

Exon 6 is the first exon within which a clear difference between conserved and unconserved regions is apparent (Figure 5.6). The C-terminal end of exon 6 is perfectly conserved whereas there are many substitutions in the N-terminal end of the exon. The adjacent N-terminal region of exon 7 is also subject to variation, although these are only conservative changes. This may point to an important role for the C-terminal region of exon 6, although it contains no known motifs.

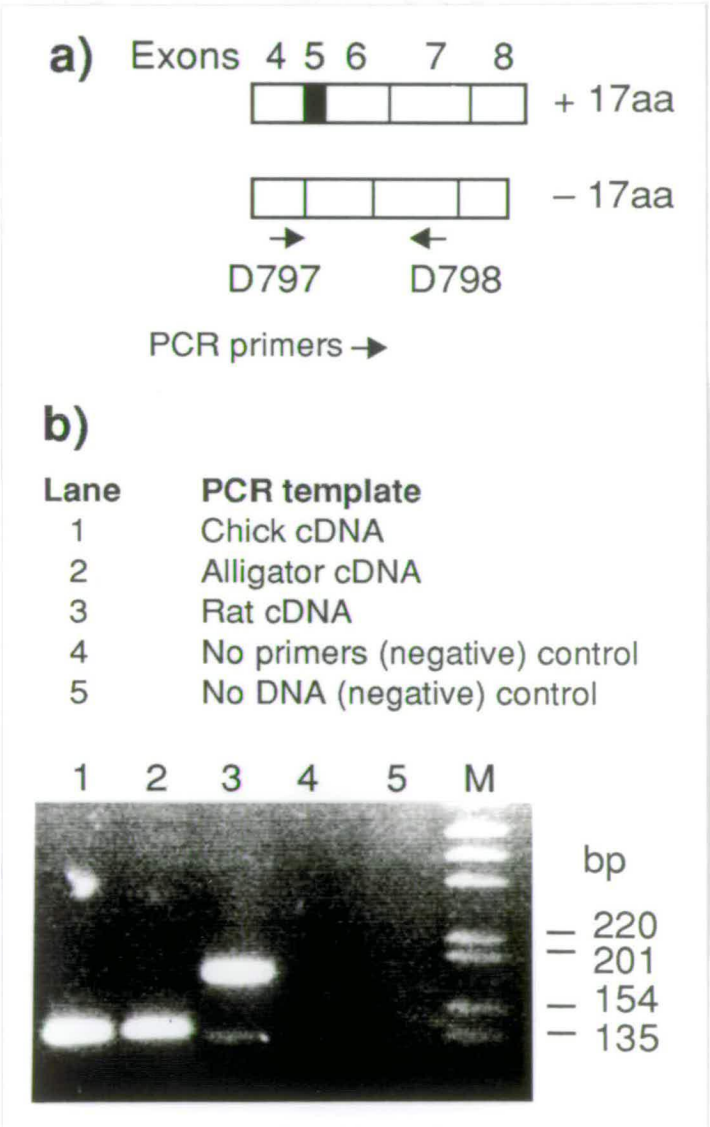
The first point mutation outside the zinc fingers in *WT1* was observed in a benign mesothelioma (Park *et al.,* 1993a). The tumour-specific homozygous A to G (1195nt) transition resulted in a serine to glycine (273aa) substitution which is considered to be a conservative change (Dayhoff *et al.,* 1972). This particular mutation was shown to have the capacity, *in vitro,* to change the transrepression activity of *WT1* to transactivation. This is surprising considering that in all species except the eutherian mammals, this position is found to be an asparagine. An asparagine to serine change is considered a non conservative substitution. This

Figure 5.5

Investigation into the conservation of the 17 amino acid alternative splice-exon 5

a Schematic representation of the + and - 17aa isoforms of *WT1* found in mammals showing the position of the PCR primers used.

b Amplification across exons 4-9 in chick, alligator and rat cDNA. Two bands 51bp apart corresponding to +/- 17aa isoforms are observed in rat. But in chick and alligator no larger band is detected.
(D797 5' AGCTTGAATGCATGACATGGAA 3')
(D798 5' TGGGTGTGTATTCTGTATAGGG 3')



substitution may have become fixed because other substitutions in this region compensate for, or neutralise, this change, or there is a functional difference between the species.

The region has also been implicated in type II RNA editing in rat and human (Sharma *et al.,* 1994). A T to C (1220nt) transition was found in the some mRNA transcripts, resulting in a leucine to proline (281aa) substitution, that was not encoded by the genomic DNA. Leucine was found to be conserved at this position in all species so far examined, including the rat. To try to assess the likelihood that editing at this position is conserved in other species the conservation of the nucleotide sequence was examined (Figure 5.6). RNA editing in mammals has been most closely studied in apolipoprotein B (Chan 1993). Dissection of the sequences required for editing has shown that an 11bp sequence, 5bp downstream from the edited position, is required and is not species-specific (Backus and Smith 1991). A species-specific sequence has been inferred further downstream. Downstream of the edited site in *WT1* there are certain conserved similarities to the apolipoprotein B non species-specific editing box. In alligator and chick immediately downstream from the affected nucleotide there is a 6 nucleotide insertion. This insertion results in a tyrosine, serine insertion in the gene product (Figure 5.6). Therefore the editing box in chick and alligator is in a different location relative to the putative edited nucleotide and so it is unlikely that the same nucleotide is edited in the reptilian lineage. The effect of a leucine to proline substitution in alligator and chick may be different to mammals because of presence of the YS insertion in the gene product. In the chick the apolipoprotein form produced by editing has not been detected; so editing of apolipoprotein B is thought not to be conserved in chick (Tarugi *et al.,* 1990).

5.2.8 The zinc finger region is the most highly conserved region of *WT1*

The region encoding the first three zinc fingers was the region for which the greatest number of sequences was obtained and therefore has the greatest depth to the comparison. In exons 7-10, which encode the zinc fingers of *WT1,* only 14 changes were observed in the 130 amino acid residues compared. Of these, seven were observed in the part of exon 7 which precedes ZF1, and these were conservative changes (Dayhoff *et al.,* 1972) or a deletion/insertion (Figure 5.1). Five of the seven substitutions found within the fingers were located either in the predicted β loop between the two zinc chelating cysteines or in the α helix between the two zinc chelating histidines (Figure 5.7). Another residue substituted in ZF2 of the zebrafish

```
                    MUTATED                          EDITED      NON SPECIES SPECIFIC
NUCLEOTIDE            G                                C           EDITING BOX
SEQUENCE              ↑                                ↑       ←5bp→TGATCANTATA
Human       GAGAGCGATAACCACACAACGCCCATCCT.....CTGCGGAGCCCAATACAGAATACA
Mouse       -----T--G--------GG-C-----------------T--T-----G----------
Rat         -----T--G----------C-----------------T--T-----G----------
Pig         ----------CG---G-C-------------------T--T-----G----------
Marsupial Mouse --A-AT----------G--T--G----------G--T----------T--------
Chick       --A-AT--G-------GTG-T--G--GT-ATACAG---T--G----------------
Alligator   --A-AT--G--------TG-T--G--GT-ATACAG---T------------------
```

```
PREDICTED          MUTATED  EDITED
AMINO ACID            G        P
SEQUENCE              ↑        ↑
Human       HSTGYESDNHTTPIL--CGAQYRIHTHGVFRGIQDVRRV
Mouse       .GI....E...A...--.....................
Rat         .G............--.....................
Pig         .G......T.A..........................
Marsupial Mouse .T....N........--.....................
Chick       .T....NE..SA.M.YS.....................
Alligator   .A....NE...A.M.YS.....................
```

Figure 5.6 Mutation, editing and conservation of exon 6

The changes in the exon 6 sequence due to mutation and editing in mammals are compared with substitutions in other species. The serine residue mutated in a benign cystic mesothelioma, causing a loss of transrepressor activity and gain of transactivator activity is not conserved outside of the eutherian mammals. The edited nucleotide in human and rat causes a leucine to proline change at a position conserved in all species. The nucleotide sequence does show some similarity to the 3' non species specific editing box in apolipoprotein but in alligator and chick this site is no longer 5bp downstream due to the insertion of 6bp directly 3' to the edited nucleotide.

WT1 is immediately after the second invariant cysteine in the β sheet. In the X-ray co-crystal structure of *EGR1(Zif 268)* all these residues are away from the DNA binding end of the finger (Figure 1.2b) and therefore probably have a limited role in binding to DNA (Pabo and Pavletich 1991). *EGR1* mutations at these sites had no effect on DNA binding (Wilson *et al.,* 1992).

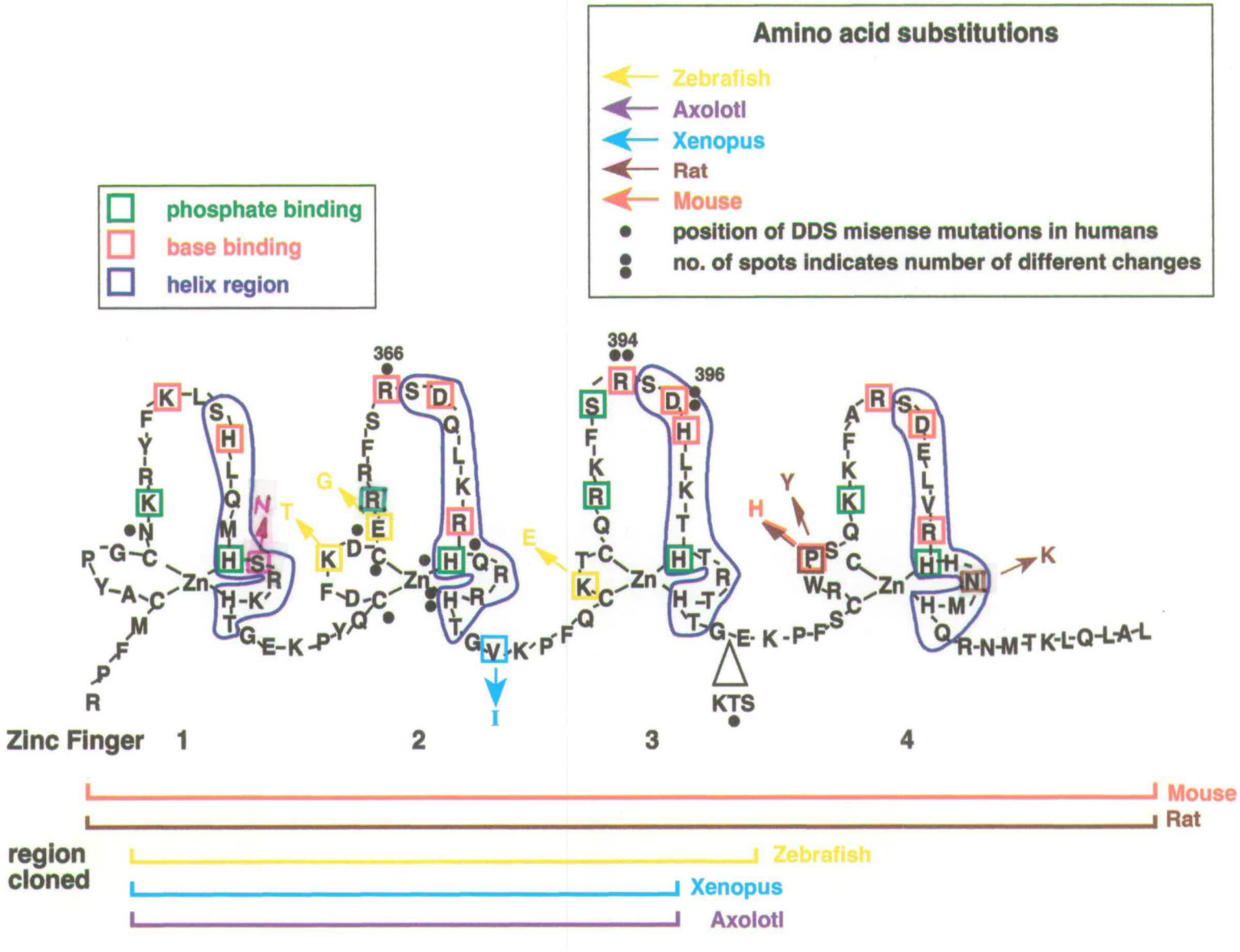
The one substitution found in *Xenopus* was in the linker region between ZF2 and 3. It is conservative substitution of one hydrophobic amino acid for another (valine for isoleucine). In a study of mutations which would change the function of the zinc finger proteins *EGR1* and *TFIIIA,* changes at this position were found to have no effect (Wilson *et al.,* 1992, Choo and Klug 1993). This is in contrast to the neighbouring lysine residue where mutation inhibits DNA binding activity (Wilson *et al.,* 1992). All the substitutions found are therefore located in sites already thought to have lesser importance in the function of the zinc fingers. This is in contrast to the positions at which pathogenic mutations have been found, which are all conserved residues.

One other change in the fingers apart from those mutations found in DDS and Wilms' tumours was found in one of the first *WT1* transcripts to be isolated, WT33 (Call *et al.,* 1990). This is a serine to phenylalanine change (365aa) in ZF2 in the C-terminal end of the β sheet (Figure 5.7); this is the part of the finger involved in DNA binding and the substitution has been shown to affect binding to some DNA target sequences (Bickmore *et al.,* 1992). It is not known whether this is a pathogenic mutation or a polymorphism, but as it has only been observed once it is unlikely to be a polymorphism. The library from which it was isolated was from a pre-B cell leukaemia cell line and so this change may be a pathogenic mutation involved in tumour progression (Bickmore *et al.,* 1992). This is also suggested because this position has been shown to be a conserved in *WT1* and in other homologous finger proteins.

The contribution made by ZF1 to the binding of *WT1* to its target DNA sequence is not known but a constitutional mutation of one of the zinc chelating cysteines results in DDS (Bruening *et al.,* 1992). The complete conservation through 400Myr, since the divergence of the fish lineage, supports an important function for ZF1. Although no missense mutations in ZF4 have been reported, this finger is conserved to the same degree as the other three fingers, indicating a functional role.

Figure 5.7 Schematic representation of the four zinc fingers of *WT1*.

The region cloned and sequenced in each species is shown and any substitutions relative to the human sequence are indicated. The positions of the missense DDS mutations are also indicated as are the important residues in DNA binding; inferred from *EGR1* and *Sp1* fingers (Pabo and Pavletich 1991, Kriwacki *et al.,* 1992). The substitutions found in the different species can be seen not to coincide with any mutated or any other functionally important sites.


Amino acid substitutions
Zebrafish
Axolotl
Xenopus
Rat
Mouse
position of DDS misense mutations in humans
no. of spots indicates number of different changes
phosphate binding
base binding
helix region
366
394
396
KTS
Zinc Finger 1
2
3
4
Mouse
Rat
region cloned
Zebrafish
Xenopus
Axolotl

5.2.9 The KTS alternative splice has been conserved through vertebrate evolution

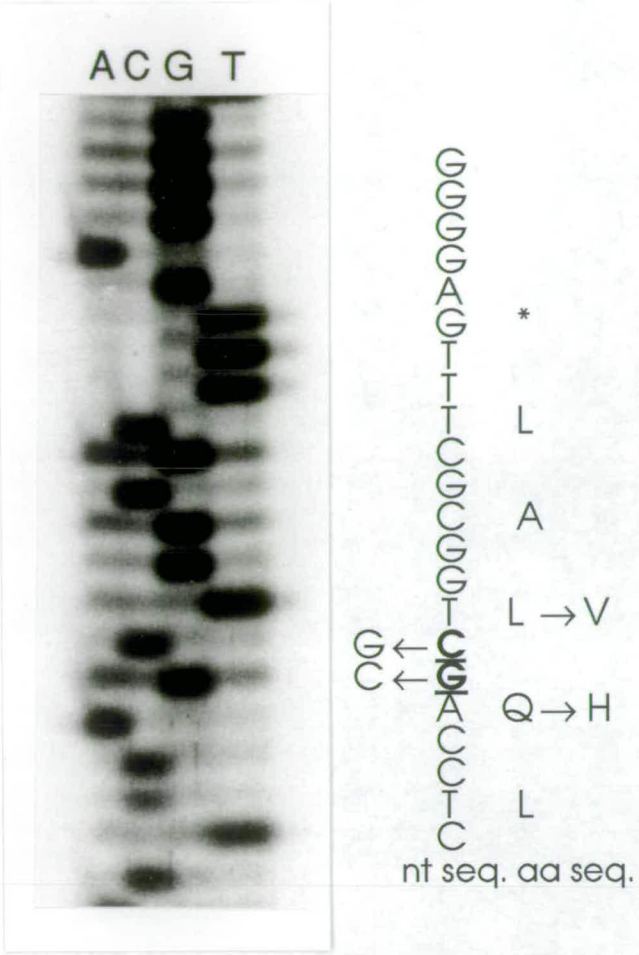
The role of the KTS alternative splice is not well understood. Previously the presence of +KTS isoform had been reported in mammalian species (human, mouse and rat). Transcripts containing the KTS isoform have been isolated from the non mammalian species, marsupial mouse, chick and alligator, and in the latter two both + and -KTS transcripts have been observed (Figure 4.5, Figure 4.11). In the zebrafish genomic clone, Z1a, the 9bp which encode the KTS are present at the 3' end of ZF3 and, in contrast to the lack of conservation in rest of the introns, both alternative splice donor sites are perfectly conserved between human and zebrafish. The conservation extends well past the minimal consensus sequence, making it very likely that +/-KTS isoforms are produced (Figure 4.12).

5.2.10 The C-Terminal region is not changed in mouse

C-terminal to the second conserved histidine of ZF4 there are only 11 amino acids before the stop codon in mammals, marsupials and alligator. In the published mouse and rat sequences (Buckler *et al.,* 1991, Sharma *et al.,* 1992) there is a two amino acid difference to the human sequence, QL becoming HV (448-9aa). This is not seen in either alligator or marsupial sequences and is attributable to an inversion from GC to CG (1715-1716nt) in human to mouse. To investigate if this a real sequence difference a mouse cDNA clone KS/2 (gift Buckler *et al.*) was sequenced and found to resemble the human (Figure 5.8). This has also been independently confirmed by Melissa Little (personal communication). This would explain why monoclonal antibodies raised against this region of the human sequence also bind the mouse *WT1* protein (V. van Heyningen personal communication). It is highly likely that the original sequencing error has been perpetuated in the sequence analysis of the rat. All calculations included in this study therefore have been based on the C-terminal end of mouse and rat *WT1* resembling human, marsupial and alligator.

5.2.11 3' Non coding region

Clones which extended past the stop codon were obtained for marsupial mouse and alligator. In the 3' untranslated region the level of identity between the clones became much lower, and an increase in the AT richness and repetitive character of the sequence was seen (Table 5.5, Appendix A). In the case of the marsupial mouse the



→ differences in the published mouse sequence

Figure 5.8 Sequence of the mouse *Wt1* cDNA clone, KS/2.

The sequencing gel of the mouse clone KS/2 shown above differs from the published mouse sequence (Buckler *et al.*, 1991) by the inversion of a GC to a CG and therefore a two amino acid difference QL to HV. The mouse sequence therefore does not differ from the human in the C-terminal region as previously supposed.

Table 5.5

Percentage similarities between coding and non coding regions of exon 10

Coding

	mouse	rat	marsupial mouse	alligator
human	95.0	92.6	88.5	83.6
mouse		95.9	86.9	81.3
rat			84.4	80.3
marsupial mouse				84.4

Non coding

	mouse	rat	marsupial mouse	alligator
human	72.0	73.4*	45.5*	45.2*
mouse		91.4*	48.2*	43.6*
rat			38.7*	41.3*
marsupial mouse				39.8*

Calculations performed using the Gap program from GCG.
Figures calculated using of the corrected mouse and rat sequences (see text for details)
* indicates incomplete sequences compared

entire 3' non coding was isolated as polyA tail was observed at the 3' end of the clone. The marsupial mouse untranslated region was approximately 1.3kb long (only 400bp were sequenced) which is the same as the length of this sequence in mouse and human. In the 3' untranslated region of the marsupial mouse the similarity to the human, mouse and rat regions was found to be as low as the alligator (Table 4.5). At the 3' end of the marsupial sequence a higher degree of similarity to the eutherian mammals was seen including a AUUAAA sequence approximately 15bp away from the beginning of the polyA tail (in bold in Appendix A). This is a weaker polyadenylation signal than the canonical AAUAAA (Proudfoot 1994) but its position and conservation in marsupial, human and mouse suggests it is the polyadenylation signal recognised. In humans and mouse a polyadenylation signal (AAUAAA) is located approximately 900bp from the 3' end of the mRNA. In alligator this sequence was repeated twice (890-899bp in AL1), only 20bp downstream from the stop codon, suggesting that in certain circumstances shorter transcripts could be produced if these sequences were recognised.

5.3 Evolutionary analysis of *WT1*

The main objectives of the comparison of *WT1* orthologs was to gain information at the structural and functional level, rather than to address the issues of rate of evolution and vertebrate phylogeny. Analysis of changes in the nucleotide sequence of *WT1* with time was performed mainly to investigate the reason for the high level of conservation of the protein sequence. At the same time it was tested if the data from these sequences were consistent with present notions of evolution and vertebrate phylogeny.

5.3.1 Rate of sequence evolution of *WT1*

5.3.1.1 Introduction

The structural analysis of *WT1* has shown that the predicted amino acid sequence has diverged very little in particular in the zinc finger domain. At the nucleotide level comparison of the of the coding versus the non coding regions (introns, 5' and 3' non coding regions, (Tables 5.2&5.5)) shows that the rate of change of the non coding region must be very much higher than the coding region. It is now known that the rate of change of coding sequences is dependent on both the actual rate of mutation and the amount of purifying selection exerted on the gene. For

protein coding sequences most of the purifying selection will operate at the level of the protein produced. Mutations which affect the nucleotide sequence without changing the encoded protein (synonymous sites) are effectively 'silent' and can be fixed in a population by such means as genetic drift. However these substitutions can have an effect at the level of GC content, codon preference, neighbouring base mutational effects, and RNA secondary structure (Eyre-Walker 1991). Therefore to quantify the evolution of a nucleotide sequence it is better to compare the 'silent' and non 'silent' sites separately rather than comparing the sequences simply on a percentage basis. A percentage comparison does not compensate for multiple, backwards and convergent substitutions. These can be corrected for, to some extent, by equations such as the Jukes and Cantor formula (Jukes and Cantor 1969).

Several programs for comparing the rate of mutation at synonymous and nonsynonymous sites have been produced. The method published by Li *et al.,* 1985 and updated in 1993 (Li 1993) was used to calculate the rate of synonymous substitutions per synonymous site (K_S) and the rate of nonsynonymous substitutions per nonsynonymous site (K_A) for *WT1*. The rate of nonsynonymous substitutions can be compared to the rate of synonymous substitutions by the formula K_A/K_S. If there is selection acting on the protein sequence the value for K_A will be much lower than the value for K_S.

The program classifies each site in the two sequences compared into 1) non-degenerate, 2) two-fold degenerate or 3) four-fold degenerate, depending on whether nucleotide substitutions at a site will change the amino acid sequence 1) always (non-degenerate), 2) one out of three times (two-fold) or 3) never (four-fold degenerate). The values obtained are then averaged between the two sequences. The sequences are then compared and the changes classified according to the above scheme. At two-fold sites transitions are taken as degenerate changes and transversions as non-degenerate. In the case of multiple substitutions in a codon the pathway within which the least number of non-degenerate changes would have occurred is given the higher weighting.

5.3.1.2 Analysis of *WT1*

The sequence information obtained for the *WT1* orthologs did not cover identical regions. It was suspected that analysing different regions for different species may well produce a bias in the figures, in particular for those clones where only the highly conserved zinc fingers had been obtained. Therefore exons 7-9 were

Table 5.6 Figures obtained for K_S and K_A using Li's program

a Exons 7-9

x100 K_S

x100 K_A

	H	M	R	P	MM	C	AL	X	AX
Human		22 (6)	24 (6)	10 (3)	55 (13)	58 (13)	60 (12)	114 (34)	62 (18)
Mouse	0.5 (0.5)		5 (2)	22 (6)	58 (12)	88 (19)	60 (12)	105 (30)	79 (23)
Rat	0.5 (0.5)	0.0 (0.0)		26 (7)	62 (13)	87 (19)	67 (14)	97 (25)	82 (24)
Pig	0.0 (0.0)	0.5 (0.5)	0.5 (0.5)		62 (14)	84 (19)	80 (19)	111 (32)	92 (44)
Marsupial Mouse	1.3 (0.8)	1.7 (0.9)	1.6 (0.9)	1.3 (0.7)		85 (19)	79 (16)	112 (34)	70 (19)
Chick	1.7 (0.9)	2.1 (0.9)	2.0 (0.9)	1.7 (0.8)	0.4 (0.4)		25 (6)	79 (20)	57 (14)
Alligator	1.7 (0.9)	2.1 (0.9)	2.0 (0.9)	1.6 (0.8)	0.4 (0.4)	0.0 (0.0)		99 (25)	47 (13)
Xenopus	0.9 (0.7)	0.8 (0.7)	0.8 (0.7)	0.8 (0.8)	0.8 (0.7)	0.8 (0.7)	0.8 (0.7)		75 (19)
Axolotl	0.9 (0.9)	0.8 (0.8)	0.8 (0.8)	0.8 (0.8)	0.8 (0.8)	0.8 (0.8)	0.8 (0.8)	1.7 (1.1)	

b Exons 3-10

x100 K_S

x100 K_A

	H	M	R	MM	C	A
H		35.3 (5.7)	34.2 (5.4)	72.6 (11.3)	95.4 (14.1)	100.4 (15.8)
M	1.8 (0.6)		10.7 (2.7)	81.6 (11.7)	93.3 (13.3)	89.2 (12.2)
R	1.7 (0.6)	0.8 (0.4)		86.2 (12.3)	94.7 (13.6)	95.1 (13.1)
MM	1.6 (0.6)	3.0 (0.8)	2.4 (0.8)		84.0 (12.9)	89.6 (14.1)
C	3.3 (0.8)	3.6 (0.9)	3.9 (0.9)	1.2 (0.6)		32.4 (5.3)
A	2.5 (0.7)	2.7 (0.7)	3.1 (0.8)	1.3 (0.5)	1.0 (0.5)	

Estimation of the rate of change of *WT1*

Calculations were performed for two regions of *WT1* using the program (Li 1993) for estimating the number of synonymous changes per synonymous site (K_S) and the number of nonsynonymous changes per nonsynonymous site (K_A). (Standard errors are in parenthesis)

Table 5.7 Comparison of K_S and K_A values for *WT1* with other genes

a Comparison of K_S and K_A between mouse and rat

Gene	K_S	K_A
Chaperonin	0.077	0.0007
αβ Crystallin	0.182	0.0000
Nucleolin	0.110	0.0281
GADPDH	0.167	0.0104
Average	0.144	0.0148
WT1 (exons 7-9)	0.050	0.0000
WT1 (exons 3-10)	0.107	0.0080

Data taken from Li 1993

b Comparison of K_A/K_S in the mammalian lineage

Gene	K_A/K_S
Aldolase A	0.019
Thyrotropin	0.017
ZFY	0.190
Histone 4	0.000
β globin	0.262
IFN γ	0.324
LDHA	0.040
Ig V_H	0.189
Insulin	0.323
IGF 2	0.225
WT1 (exons 7-9)	0.045
WT1 (exons 3-10)	0.051

Data taken from Li *et al.*, 1985* and Li 1993

*Figures calculated using old method which underestimates K_S and overestimates K_A to some extent.

compared for all the orthologs isolated (Table 5.6a) and exons 3-10 compared in the mammalian and reptilian lineages (Table 5.6b).

In general the increase in evolutionary distance correlated well with an increase in the values of K_A and K_S. The only major inconsistency was that the value for K_S in *Xenopus* was much higher than for axolotl, although both are amphibians and therefore similar values would have been expected. This could be owing to the small amount of sequence available and the large standard errors. The K_A values for *Xenopus* and axolotl were lower than for the reptiles although it is believed that amphibians diverged from common ancestors 60Myr earlier (Benton 1990). This is probably because the N-terminal sequence of exon 7 could not be included for *Xenopus* and axolotl.

For exons 7-9 K_A approaches zero as would be expected. The standard error therefore was very high (Table 5.6a). For exons 3-10 K_A was higher than for exons 7-9, consistent with a larger number of amino acid substitutions. In both regions the value for K_S was found to be much higher than the value for K_A, thus K_A/K_S is very low, which is consistent with purifying selection acting at the level of the protein sequence. The data obtained for *WT1* were compared with other proteins, the K_A, K_S and K_A/K_S values were in the same range as those calculated for other proteins (Table 5.7).

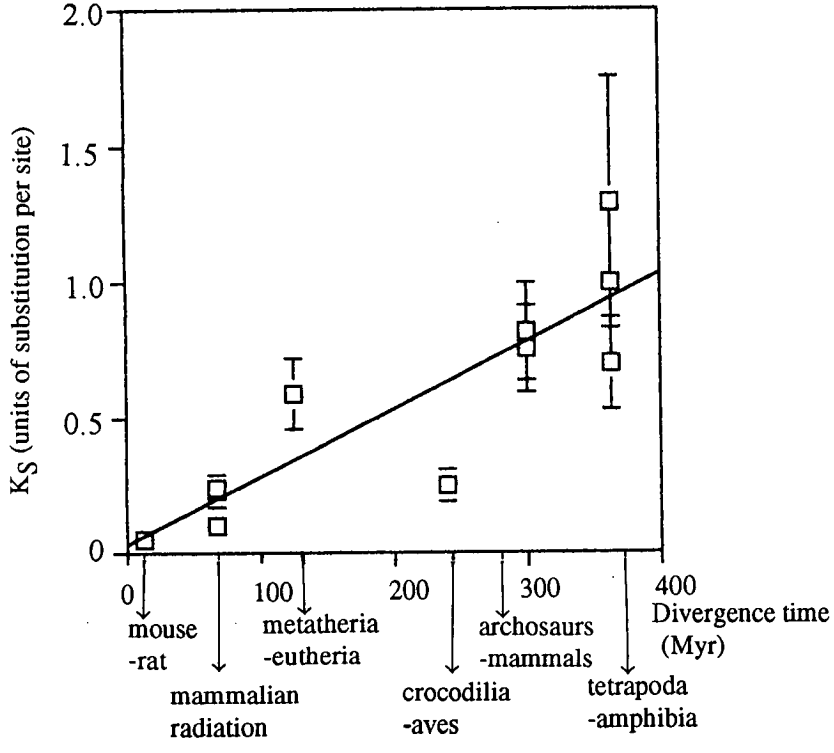
To investigate the relationship of K_S with time, the values for K_S within each clade were averaged and these values plotted against the estimated divergence times of the groups (Benton 1990). The relationship was found to show a linear trend (Figure 5.9). The rate of synonymous substitutions per site per year was found to be similar for both regions examined at $1\text{-}1.25\text{x}10^{-9}$ using the formula rate = $K_S/2T$ (T=divergence time) (Li and Graur 1991). This is lower than values obtained previously ($1.7\text{-}11.8\text{x}10^{-9}$) (Li *et al.*, 1985) but the updated method (Li 1993) is known to give a lower, more realistic, value for K_S. It was noticed from these graphs that the values obtained for comparisons with the chick and alligator sequences lay below the line (Figure 5.9). The K_S and K_A values were subjected to a relative rate test (Li and Graur 1991) and this suggested that the rate of evolution of *WT1* may be slower in the reptilian lineage. This could possibly be an artefact owing to a small data set. To discover the real reason behind the differences in the apparent rates a more extensive sequence data set will be required.

Figure 5.9 Graphs plotting K_S against divergence time

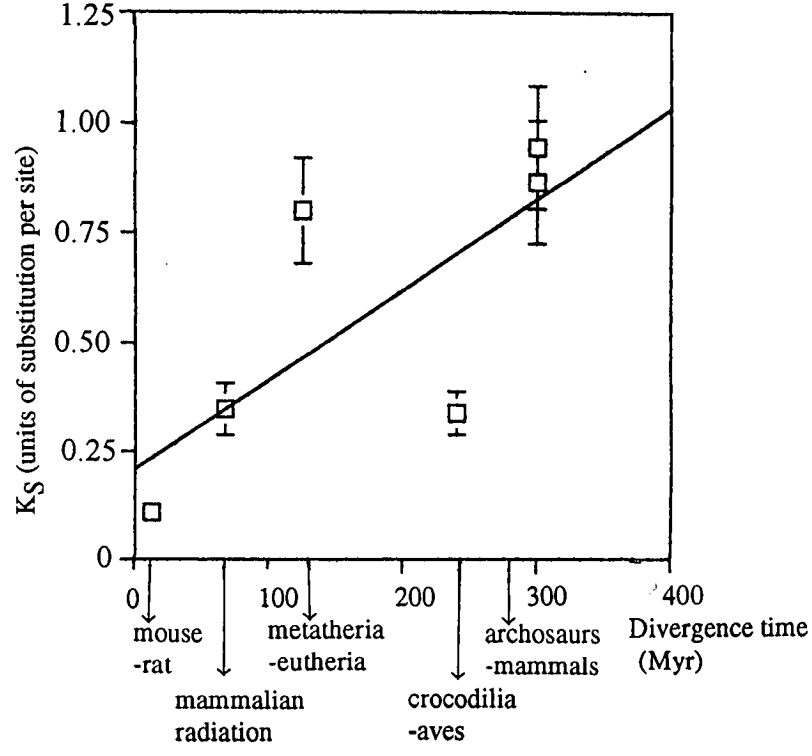
The K_S values (number of synonymous changes per synonymous site) for each species was calculated (Table 5.6) and these were averaged within each group; mammals (eutheria), marsupials, archosaur lineage (birds and some reptiles) and amphibians. The K_S values were then plotted against the divergence times for amphibians from tetrapods, archosaurs from the mammalian lineage, aves from crocodilians, eutheria from metatheria, mammalian radiation and mouse from rat.

No definitive conclusions can be drawn from this data as the errors are high, but points including data from the chick and alligator tended to lie below the line, maybe indicating a non constant rate of evolution of the nucleotide sequence of *WT1*.

a Exons 7-9



b Exons 3-10



5.3.2 Phylogenetic analysis

In order to calculate the divergence between sequences, from which to derive a phylogenetic relationship, multiple alignment of the nucleotide sequence of exons 7-9 was performed using the Clustal V program. The sequences are initially aligned using a fast comparison of the sequences, by producing k-tuple scores for the sequences, plotting these on a dot matrix and finding the diagonals with the most matches. Using the similarity scores a dendrogram was produced using the UPGMA (unweighted pair group method with arithmetic mean) method. The sequences were then aligned starting with the most closely related sequences. This alignment can be used to produce an unrooted phylogenetic tree using the neighbour joining method of Saitou and Nei 1987 (Figure 5.10). This is a distance method, as opposed to a parsimony method, where the percentage divergence is calculated from the alignment and then used to produce the tree. The tree was bootstrapped 500 times (i.e. samples of sequence were taken and from them a tree calculated, the number of times the same internal branches are found gives the degree of confidence in the tree). The weakest parts of the tree were found to correspond to the non mammalian branches. The major inconsistency with palaeontological relationships (Figure 1.7) was also found in this region. The reptiles are shown to form a clade with amphibians rather than with mammals (the amniotes). In addition, the axolotl sequence was found to diverge from the reptiles after *Xenopus,* whereas they would be expected to group together. The figures for the percentage divergence from which the tree was drawn, were plotted against time. A linear increase in divergence with time was seen (Figure 5.11). It was found that the values for the reptilian lineage were again relatively lower than expected, perhaps explaining why the amphibians and reptiles grouped together in the tree.

Figure 5.10 Unrooted phylogenetic tree for exons 7-9 of *WT1*.
This was produced by the neighbour joining method using a CLUSTALV alignment. Branch lengths are drawn approximately to scale and represent 10x percentage divergence. The level of confidence in each branch point as derived from bootstrapping is also indicated.

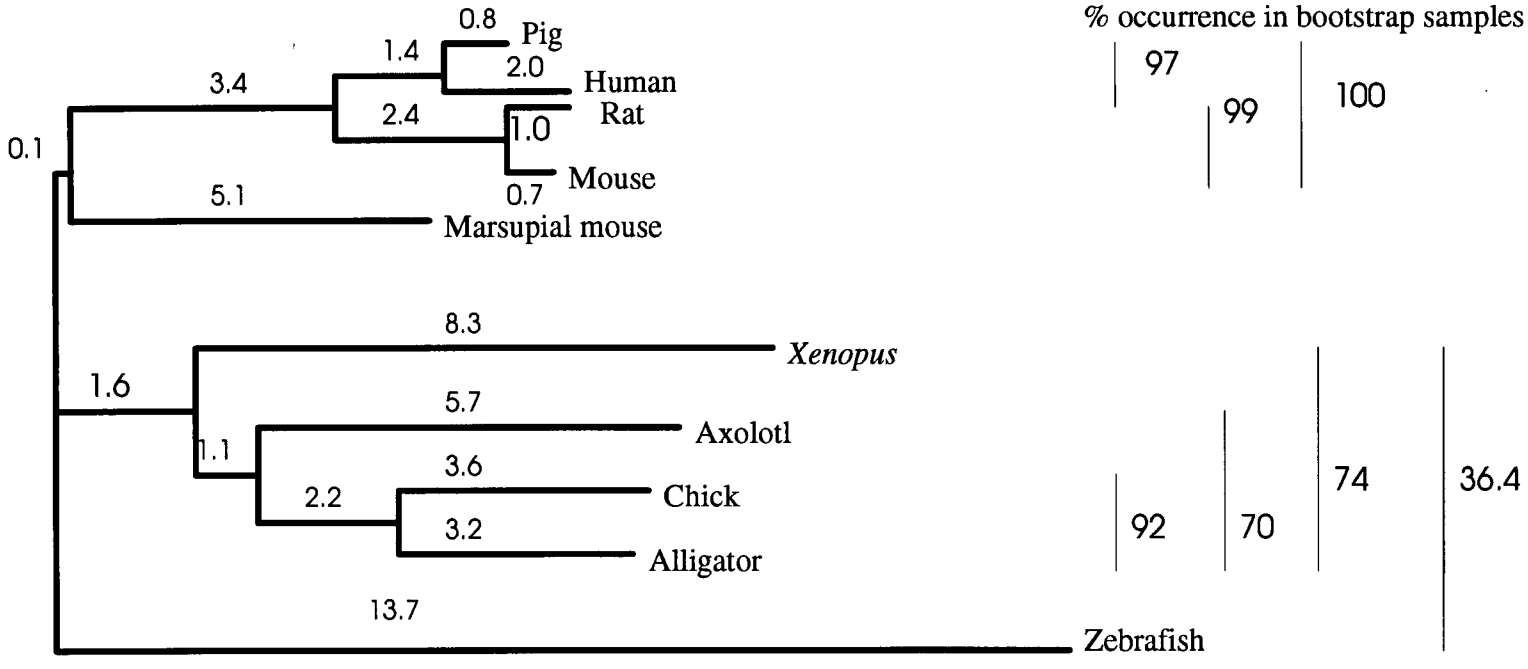
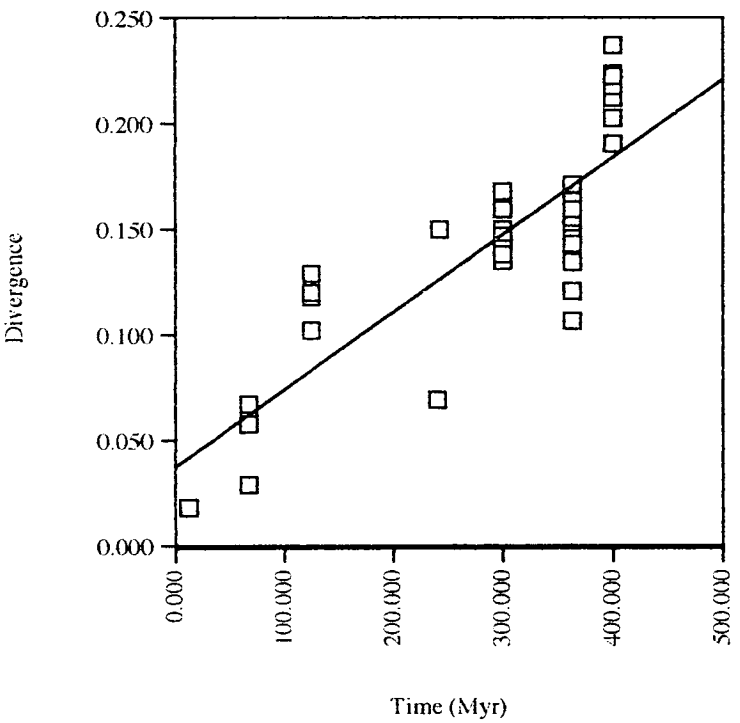


Figure 5.11 Graph of divergence of *WT1* against divergence time

Graph plotting percentage divergence, calculated using CLUSTAL V, against divergence time. An approximately linear increase in divergence with time was found. Outlying points were again found to be for reptilian species.



Chapter 6

Investigation of the Expression of *WT1*

6.1 Introduction

The cloning of *WT1* orthologs in the different classes of vertebrate will allow the investigation of the similarities and differences of *WT1* gene expression among them. Knowledge of normal expression patterns is an initial requirement before manipulation of a system can be attempted. Northern blotting and *in situ* hybridisation techniques were therefore used to investigate the size and expression pattern of *WT1* mRNA in several of the species from which *WT1* had been isolated so these could then be compared with that of mammals. The most extensive study was performed in chick and alligator embryos. The development of the genitourinary system in the reptilian lineage is of interest because of its intermediate nature between that of the mammals and lower vertebrates. In alligators the metanephros does form but contains relatively few nephrons while a functional mesonephros persists into post hatching stages (Forbes 1940). In reptiles, but not birds, the mechanism of sex determination is of interest because it is temperature dependent as opposed to being genetically determined (Ferguson and Joanen 1982). This mechanism is under investigation in alligators as it is a solely temperature dependent mechanism. In other reptiles there is also a genetic element. The cloning and determination of the function of genes involved in the early developing gonad will help elucidate this.

6.2 Northern analysis of *WT1* mRNA

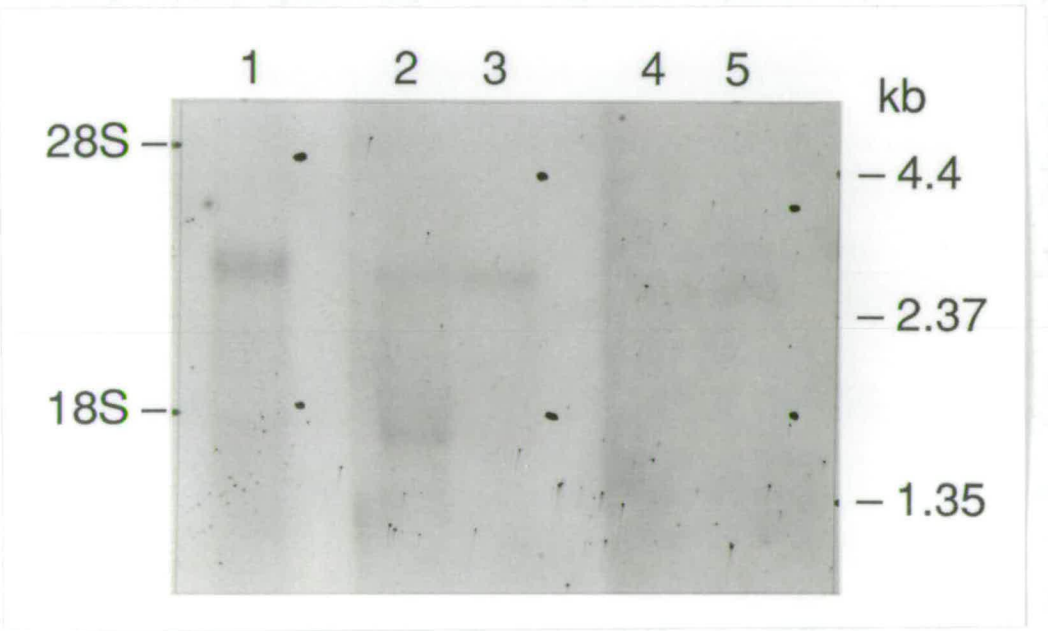
Total RNA was prepared from the genitourinary systems of pig, chick, alligator, and *Xenopus*. In order to determine the size of *WT1* mRNA transcripts in these species this RNA was probed with *WT1*. RNA was resolved on a denaturing agarose gel, and then transferred by vacublotting onto Hybond N (Amersham). The filters were then hybridised at high stringency with the appropriate homologous probe (Figures 6.1&6.2). Details of RNA samples and probes used are contained in Table 4.1 and Figure 4.5.

The major transcript detected in chick, alligator, and pig was approximately 3kb, the same size as that observed in human, mouse, and rat (Call *et al.*, 1990, Buckler *et al.*, 1991, Sharma *et al.*, 1992). An additional 2kb transcript was detected in chick RNA obtained from the complete developing genitourinary system of 11 day embryos, but not from RNA isolated from the mesonephros alone (Figure 6.1).

Total RNA from both whole *Xenopus laevis* tadpoles (stages 45-46) and mesonephros (stage 55) was probed with *XWT1*. Consistent with a higher level of

Figure 6.1 Detection of *WT1* mRNA in pig, chick and alligator

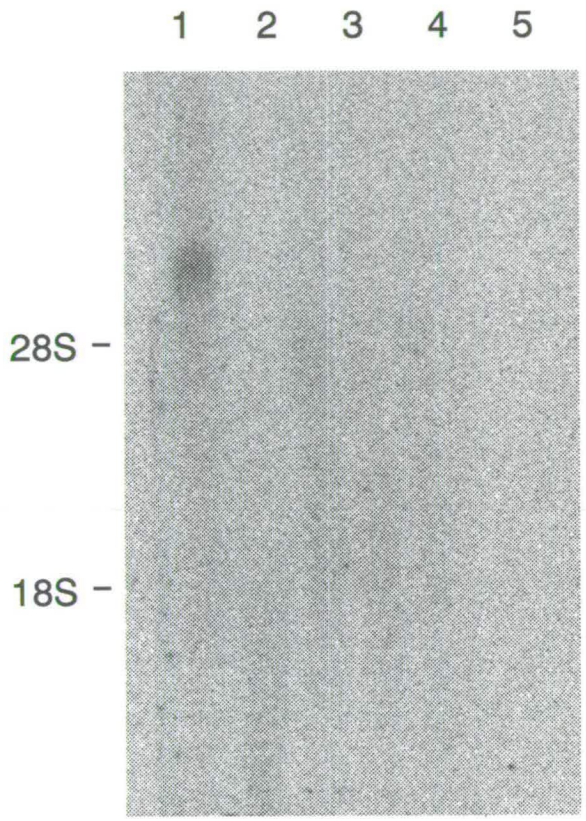
Lane	Species	Source of RNA Organ
1	Pig	Juvenile Testis
2	Chick	11d genitourinary system
3	Chick	13d mesonephros
4	Alligator	St. 20 male gonad (33°C)
5	Alligator	St. 20 female gonad (30°C)



Autoradiograph of a northern blot after high stringency hybridisation with the homologous PCR clones to total RNA from pig, chick and alligator. 3kb bands are present in all tracks demonstrating the conservation of the length of the mRNA transcript in mammals, birds and reptiles. A 2kb band is also detected in the sample from the 11 day chick genitourinary system.

Figure 6.2 Detection of *Xenopus WT1* mRNA

Lane	Source of RNA
1	St. 55 mesonephros
2	St. 45 tadpole
3	St. 45 tadpole
4	St. 46 tadpole
5	St. 46 tadpole



PhosphorImager analysis of a northern blot of total RNA from whole tadpole (St. 45-46) and mesonephros (St. St. 55) after high stringency hybridisation to *XWT1* (PCR clone). *XWT1* is only detected in the sample from mesonephros and the transcript size is 5kb.

WT1 expression in the mesonephros than elsewhere a 5kb transcript was detected only in mesonephric RNA (Figure 6.2).

6.3 *In situ* hybridisation

Recently, the technique of whole-mount *in situ* hybridisation has been developed which allows the expression pattern of a gene to be analysed in the whole embryo as opposed to having to reconstruct this from sections. The embryos can then be sectioned to identify more closely the sites of the expression. *In situ* hybridisations on sections can be used to obtain better histological detail than in whole-mounts. Whole-mounts also enable unexpected sites of expression to be identified which may not have been included in sections.

Antisense digoxigenin-dUTP labelled riboprobes were prepared from chick and alligator PCR clones *CWT1* and *AWT1* (probes described in Figure 4.5) by *in vitro* transcription, using the Boehringer Mannheim kit. Chick and alligator embryos were collected at various timepoints after incubation at the appropriate temperature. Albino *Xenopus laevis* embryos between stages 14-39 were generously donated by Richard Harland. Embryos were dissected away from the surrounding membranes, fixed, and hybridised with the appropriate riboprobe (as described in Materials and Methods). The signal was visualised by binding an alkaline phosphatase conjugated antidigoxigenin antibody to the annealed probe. When NBT (nitroblue tetrazolium salt) and BCIP (5-bromo-4-chloro-3-indolyl phosphate) were added, they were catabolised by the alkaline phosphatase to give a insoluble blue precipitate at sites of mRNA expression. To detect any non-specific signal produced due to trapping of the probe (in particular in the intraembryonic spaces), a sense riboprobe or vector-only riboprobe was hybridised in parallel to the antisense riboprobe.

6.3.1 Comparison of the developmental stages of amniotes

A guide to the equivalent stages of human, mouse, chick and alligator development based on somite number and limb development and other developmental features such as branchial arches, is provided in Table 6. Exact whole embryo comparison between species let alone across classes is impossible as different structures are found to develop at different stages and at different rates. This changing of developmental timing and rates (heterochrony) is probably one of the

Table 6 Development of the chick showing comparisons with human, mouse and alligator.

Comparisons based on somite number, limb development or branchial arch or other recorded features
Derived from Hamburger & Hamilton (1951), Romanoff (1960), Thelier (1972), Ferguson (1985), Wanek (1989), Kaufman (1992) and Moore and Persaud (1993).

Chick stage/age	Chick somite number	Alligator equivalent general features /somites	Mouse equivalent somites/limbs	Human equivalent	Chick developmental events
St. 1					Prestreak (no primitive streak)
St. 2/6-7h					Initial streak
St. 3				15d	Intermediate streak (primitive streak extending)
St. 4				16d	Definitive streak (primitive at maximun length)
St. 5			St. 11/7.5d		Head process appears
St. 6/23-24h					Head fold forms
St. 7	1 somite		St. 12/8d		Nephrogenic cord forms, mesonephros potentiality determined
St. 8	4 somites			18d	Neuralation
St. 9	8 somites		St. 13/8.5d		Pimordia of the pronephric tubules first appear as paired thickenings in the intermediate mesoderm in the region of somites 5-8
St. 10	10 somites	Eggs laid			Pronephric buds grow in a dorsolateral direction Posterior neuropore closes, tail bud stage
St. 11	13 somites	St. 1/*1*	St. 14/9d		Pronephric duct has lumen Closure of anterior neuropore
St. 12-13	16-19 somites				
St. 14/50-53h	21 somites	St. 2/**2**	St. 15/9.5d	24d	Degeneration of pronephros as far as somite 11
St. 15	24 somites	St. 3/**3**		25-26d	11 pairs of pronephric tubules 5-16 somite region limb primordia-prospective limb areas flat not yet determined
St. 16	26-28 somites			27-29d	Wing lifted off blastoderm, leg primordium is condensation of mesoderm
St. 17	29-32 somites	St. 4-6/**4**	St. 16/10d		Pronephric glomeruli first appear after the pronephros has begun to degenerate . The mesonephros starts to differentiate, condensing blastema from somites 13/14-20 limb buds apparent
St. 18/3d	30-36 somites	St. 7-8/**4**	**St. 15**		Anterior, non functional, mesonephric tubules have fused with the mesonephric duct Limb buds enlarged
St. 19	37-40 somites	St. 9/**5**	St. 17/10.5d **St. 16**	32d	Differentiation of principle mesonephric region (somites 20-30) occurring First sign of genital ridge development
St. 20	40-43 somites		St. 18/11d	32-33d	Tip of tail still unsegmented, faint eye pigment
St. 21	43-44 somites	St. 10	St. 19/11.5d **St. 17**	37d	
St. 22					Somites extend to tip of tail
St. 23/4d		St. 11	**St. 18-19**	5w	Pronephros degeneration is complete Ureteric bud appears shortly after the nephric duct opens into the cloaca First sign of Mullerian duct
St. 26/5d		St. 12-13	**St. 20**	41d	Mesonephros begins to function - evidence hydronephrosis in cut ducts and movement of dyes against gradient Ureteric bud turns anteriorly, metanephros is separated from the mesonephros by a region corrsponding to 2 somites developing no tubules Primary sex cords in indifferent gonad
6d		St. 14		4w	Appearence of genital tubercale
6.5d		St.21-22/31-40d	St. 21/13d	7w	Testis differentiation begins
7d		St. 22-23/36-45d	St. 22/13.5d	10w	Ovarian differentiation begins
8-9d			14d		Beginning of disintegration of mesonephros, loss of lumen in tubules Metanephric tubules present but as yet undifferentiated Male and right female Mullerian ducts stop developing and degeneration starts
10d		St. 24/45-50d	**St.25/16.5d**		Vascularisation of the mesonephros
11d				11w	Maximum function of mesonephros, metanephros starts to function Left female Mullerian duct reaches cloaca Rete cords of testis associate with the mesonephros
12d				11-13w	First loops of Henle appear in metanephros Male mullerian ducts completely regressed
15d					Major functioning transferred to metanephros
15-21d					Loss of 77% mesonephric weight
21d approx hatching		St. 28/65d hatching	19d birth	9 months birth	Mesnephros stops functioning

major mechanisms of evolution. One of the most obvious changes in the species investigated here is the difference in relative timing of somite and limb bud development. The onset of development of the limb buds in the chick and alligator occurs when a higher number of somites have developed than in mammals. Also the hind limb bud develops first in alligator, as opposed to the forelimb bud in mammals. These two characteristics are used for staging mammalian, chick and alligator embryos and so discrepancies can arise depending on which features of the embryos are compared.

If the developmental events in the genitourinary system are compared against other developmental events, mammalian pronephros differentiation and degeneration occurs at about the same time as in the chick, but mesonephros differentiation and degeneration occurs earlier. Mammalian mesonephros function has never been demonstrated, i.e. it is vestigial. This is in contrast to the chick and alligator where the mesonephros does function and develops as a much larger organ than that observed in mammals (Romanoff 1960).

6.3.2 *WT1* expression in chick embryos

Chick embryos were harvested between stages 15-20 (53 hrs-3.5 days) (Hamburger and Hamilton 1951) approximately equivalent to 9.5-11 day mouse (Table 6). These stages corresponded to the development of the mesonephros from the urogenital ridge and therefore with expression of *WT1* observed in mammals (Armstrong *et al.,* 1992). The analysis of *WT1* expression in chick embryos was prevented in the anterior region due to trapping of the probe. It was decided that the signal present in the anterior region but not the posterior region was non-specific because signal was also obtained with a vector only probe in the anterior but not the posterior region (Figure 6.3c). In early stages *WT1* expression was observed in the urogenital ridge region of the intermediate mesoderm, extending caudally from the level of the heart, and in later stages expression in the developing paired mesonephroi was seen. Figure 6.3a shows a stage 16 embryo probed with antisense *WT1* RNA. The paired regions of expression extend caudally from the heart along the ventral surface. At this stage no tubules were visible in this region and the mesonephric duct which induces their differentiation has just reached the posterior end of the embryo (Figure 6.3b), indicating that this may correspond to the early stage of nephron differentiation following induction by the mesonephric duct. By stage 19 tightly packed tubules are visible in mesonephros (data not shown). In the stage 20 embryo

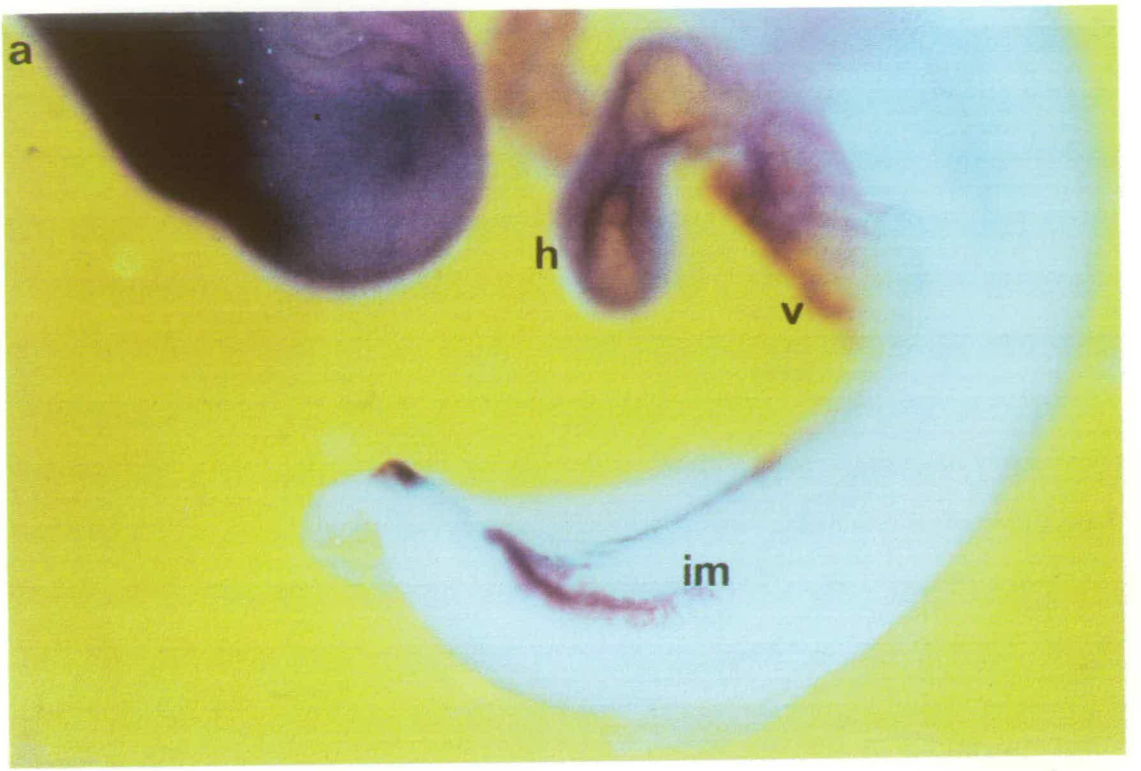
Figure 6.3 Expression of *WT1* in the developing chick

a Whole-mount *in situ* hybridisation of St. 16 chick embryo with an digoxigenin labelled antisense *WT1* RNA probe. Staining of paired ridges in the posterior end of the embryo corresponding to the differentiating intermediate mesoderm (urogenital ridge) is visible. The curve of the embryo hides the fact that the expression extends as far anterior as the heart. The most posterior end of the embryo was lost during dissection as it had not been completely delineated from the blastoderm (15x magnification).

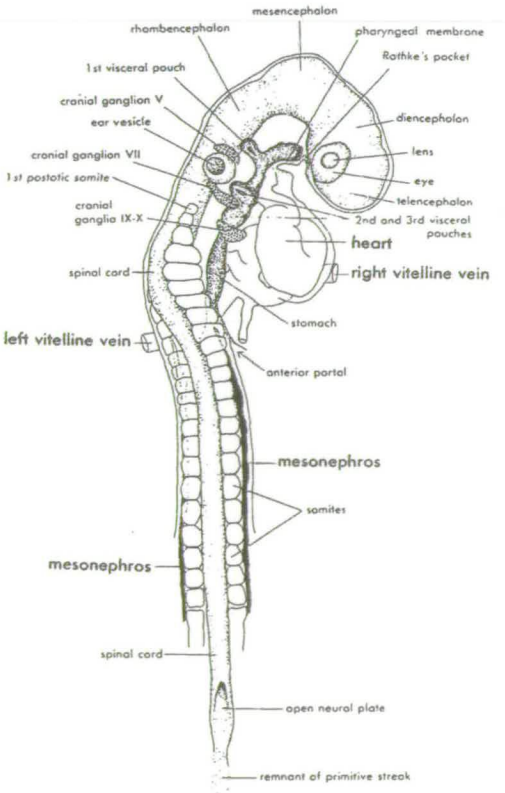
b Drawing of a St. 15 chick embryo (Balinsky 1981) showing the position of the intermediate mesoderm and mesonephros corresponds exactly with the region of *WT1* expression seen above.

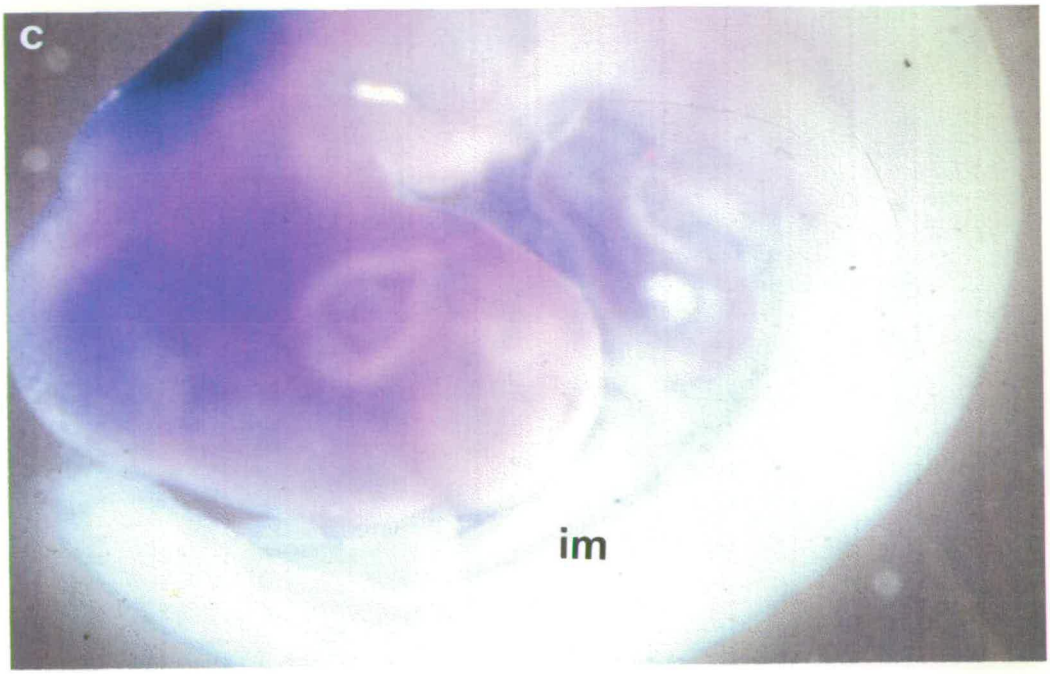
c Whole-mount *in situ* hybridisation, with a vector only probe, of a St. 16 chick. Staining is only seen in the anterior region, due to trapping of the probe in intraembryonic cavities (10x magnification).

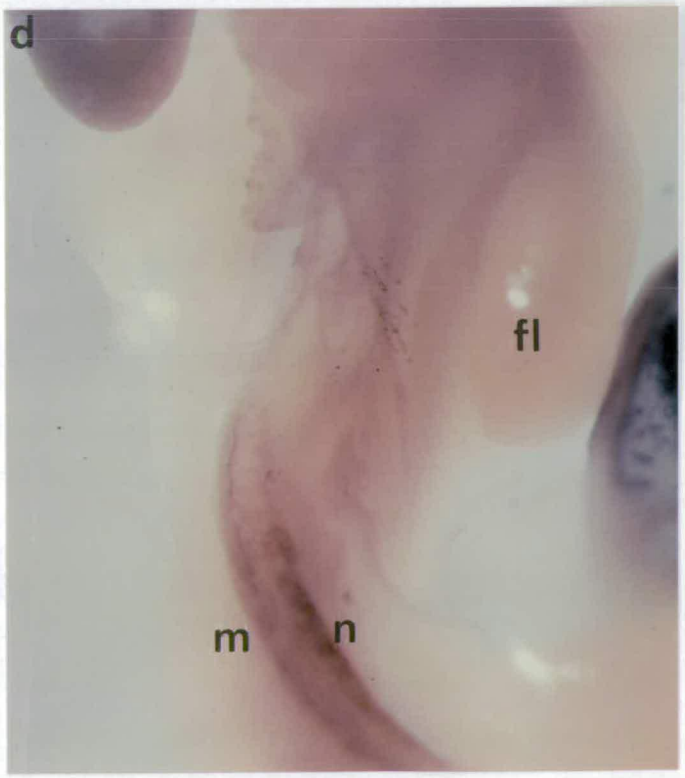
d A St. 20 embryo lying on its back to show the expression of *WT1* in the developing right mesonephros. The *WT1* signal does not appear to be uniform along the mesonephros and this may correspond to localisation of high expression levels in the developing nephrons. One of the forelimbs is also visible and no *WT1* expression was detectable (12x magnification).
(h heart, im intermediate mesoderm, fl forelimb bud, m mesonephros, n developing nephrons, v vitelline veins)


a
h
v
im

b


mesencephalon
rhombencephalon
pharyngeal membrane
Rathke's pocket
1st visceral pouch
cranial ganglion V
ear vesicle
diencephalon
cranial ganglion VII
lens
1st postotic somite
eye
telencephalon
cranial ganglia IX-X
2nd and 3rd visceral pouches
heart
spinal cord
right vitelline vein
stomach
left vitelline vein
anterior portal
mesonephros
somites
mesonephros
spinal cord
open neural plate
remnant of primitive streak


c
im


d
fl
m
n

the signal is discontinuous along the ventral edge of the mesonephros, consistent with the localisation of *WT1* to the differentiating nephrons, perhaps to the glomeruli which develop after the tubules are seen (Figure 6.3d).

6.3.3 *WT1* expression in alligator embryos

The collection of alligator eggs from Louisiana, USA, their incubation and *in situ* hybridisation was performed by AM Coriat at the University of Manchester as part of a collaborative study on early gonad development in alligators. Alligator embryos at stages 3-14 (3-17days) (Ferguson 1987) were investigated. These are equivalent to chick stages 15-29 (2-6days) (Hamburger and Hamilton 1951). Earlier embryos were not obtained because some development has occurred before oviposition and eggs cannot be immediately collected from nests.

No expression was detectable in alligator at stages 3 and 5 (Figure 6.4 a&b). The onset of *WT1* expression in the alligator occurs at stage 6 (Figure 6.4), equivalent to about stage 17 in chick (Table 6.1). The expression is localised to the anterior end of the intermediate mesoderm at the level of the heart (Figure 6.4c) and the roof of the rhomboencephalon (hind brain) (Figure 6c&d). In a number of embryos, signal was observed in the forebrain but this was probably due to trapping of the probe as a similar signal was observed in a stage 6 embryo probed with the sense control (Figure 6.4f). In stages 9-11, the expression of *WT1* extends down the intermediate mesoderm corresponding with the wave of differentiation of the urogenital ridge induced by the mesonephric duct (Figure 6.5). Two components of the developing circulatory system seem to express *WT1*. At the level of the developing forelimb bud, a paired region of expression ventral to the mesonephros is observed, this is thought to correspond to the vitelline veins which overlie the mesonephros and above this there is signal in the developing heart. At stages 10 and 11 the signal extends the length of the intermediate mesoderm and is also found in the fore and hind limb buds (Figure 6.5c-e). As the hand plate forms in stages 13-14 the limb bud expression becomes localised to the presumptive wrist and armpit regions and a region adjacent to this on the body wall (Figure 6.6). On either side of these embryos from the neck region downwards signal is also visible, this is thought to correspond to the lateral plate mesoderm, which is differentiating into the body wall. Sections of the mesonephros of stage 21 embryos show the tubules passing through the comma and S-shaped stages of differentiation (Figure 6.7). *In situ* hybridisation demonstrates localisation of *WT1* expression to the tubules.

Figure 6.4 Whole-mount *in situ* hybridisation of alligator embryos - onset of *WT1* expression

a&b St. 3 and 5 embryos probed with antisense *WT1* riboprobe, no expression of *WT1* is detectable.

c St. 6 embryo hybridised with antisense probe. The onset of *WT1* transcription is detected in anterior end of intermediate mesoderm beginning at the level of the heart

d. St. 6 showing that staining is also apparent in the roof of the rhomboencephalon, the hind brain.

e. St. 7 embryo showing persistence of the hind brain expression

f. St. 6 embryo probed with sense probe. Only non *WT1* specific signal was found in the fore brain
(fb fore brain, hb hind brain, h heart, im intermediate mesoderm)

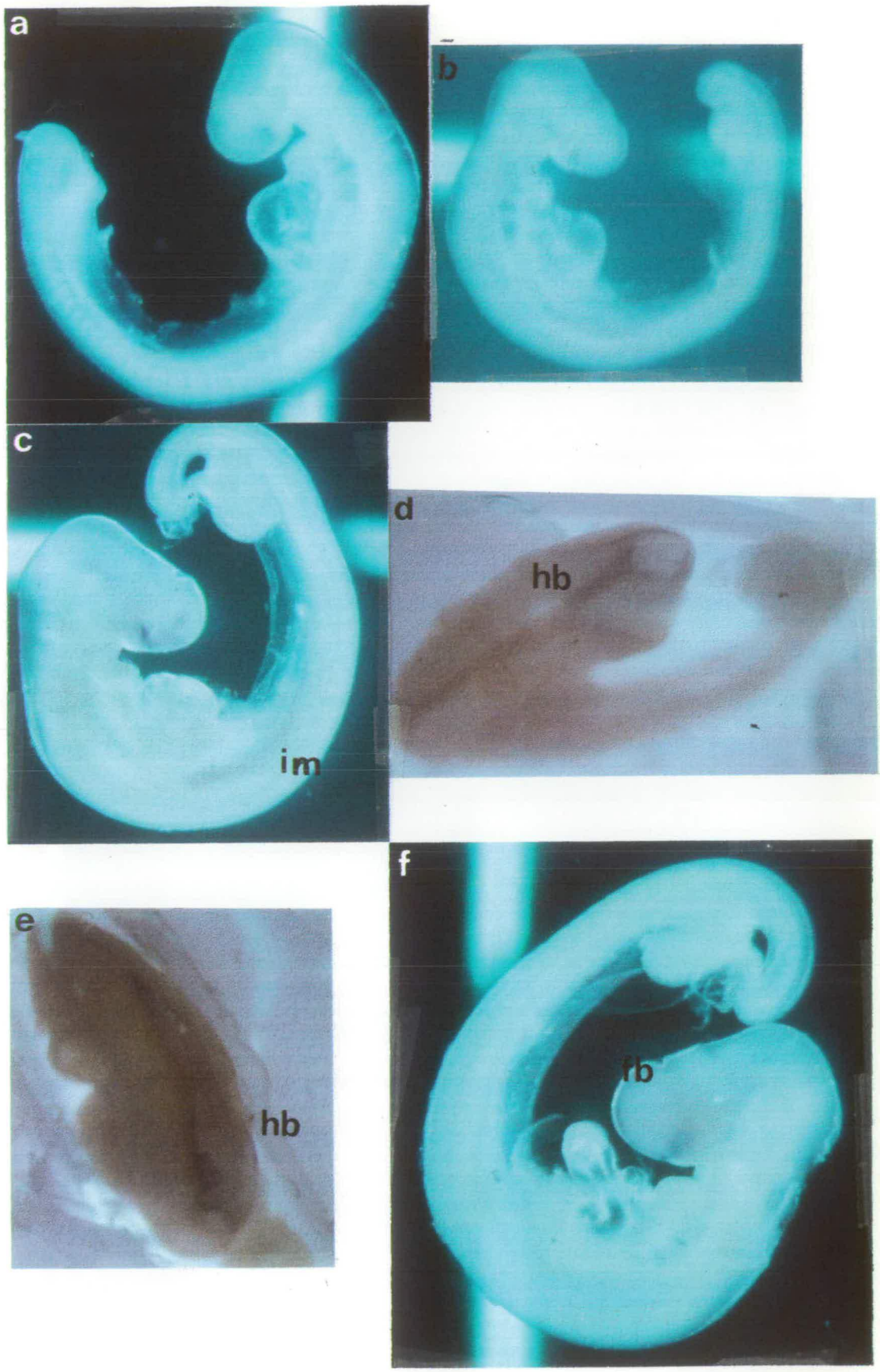

a
b
c
im
d
hb
e
hb
f
fb

Figure 6.5 *WT1* expression in stage 9-11 alligator embryos

a &b St. 9 embryos, left and right sides respectively, showing the extension caudally of *WT1* expression in the differentiating urogenital ridge as far as the cloaca. Expression in the heart and vitelline veins is apparent as is expression in the roof of the hind brain. Also at this stage the otocyst is stained, this is probably due to the trapping of the probe as well as in the forebrain.

c St. 10 embryo, expression of *WT1* is detected in the limb buds as well as the intermediate mesoderm.

d&e St. 11 embryos either lightly or heavily stained. Expression is maintained in the intermediate mesoderm, limb buds, heart and vitelline veins.
(fl forelimb bud, hl hind limb bud, h heart, hb hind brain, im intermediate mesoderm, o otocyst, v vitelline vein)

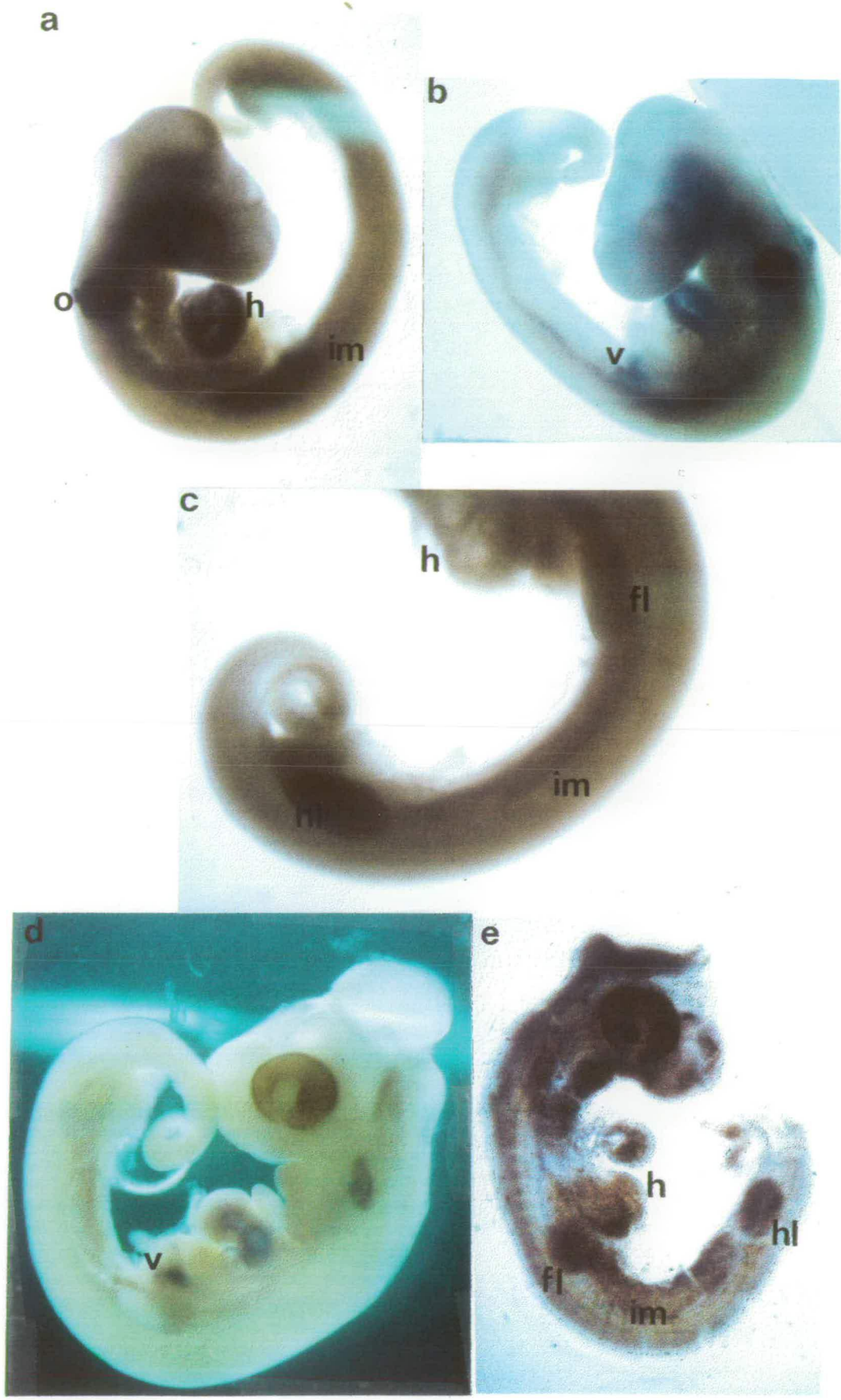
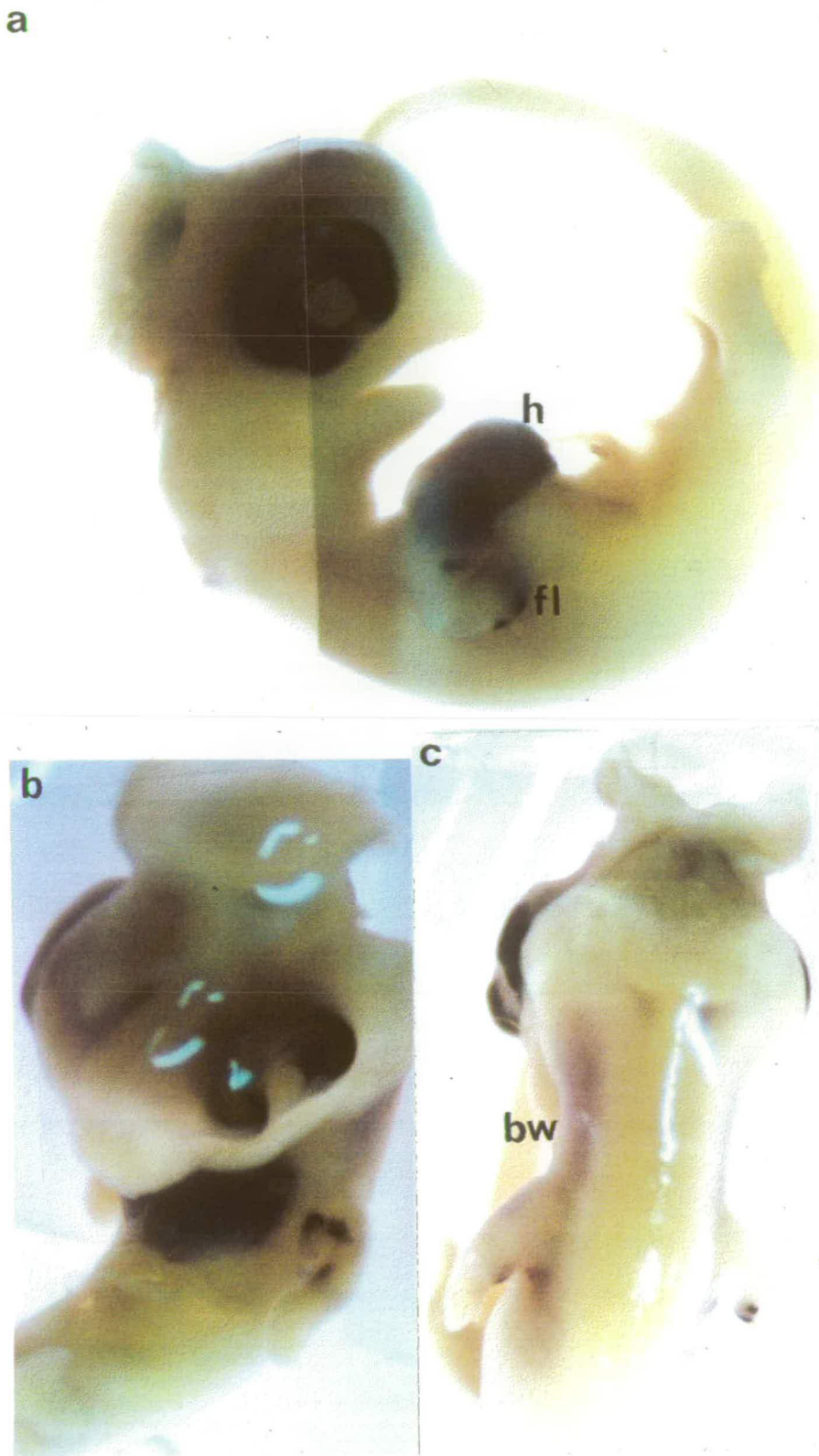
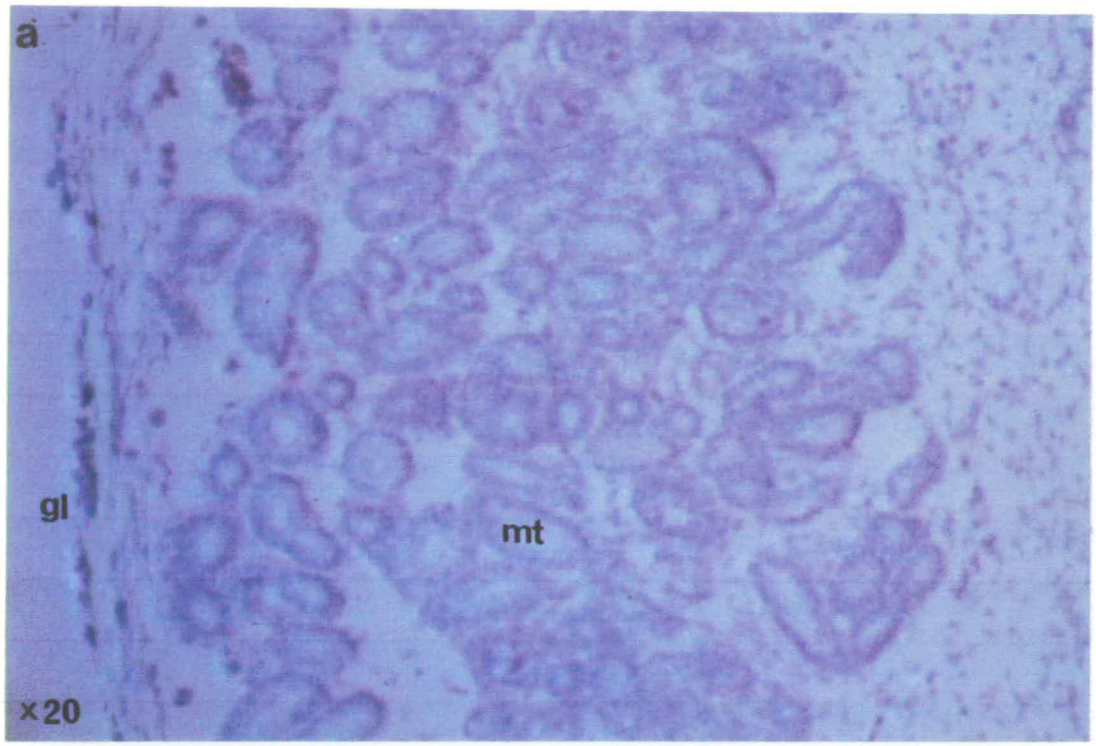
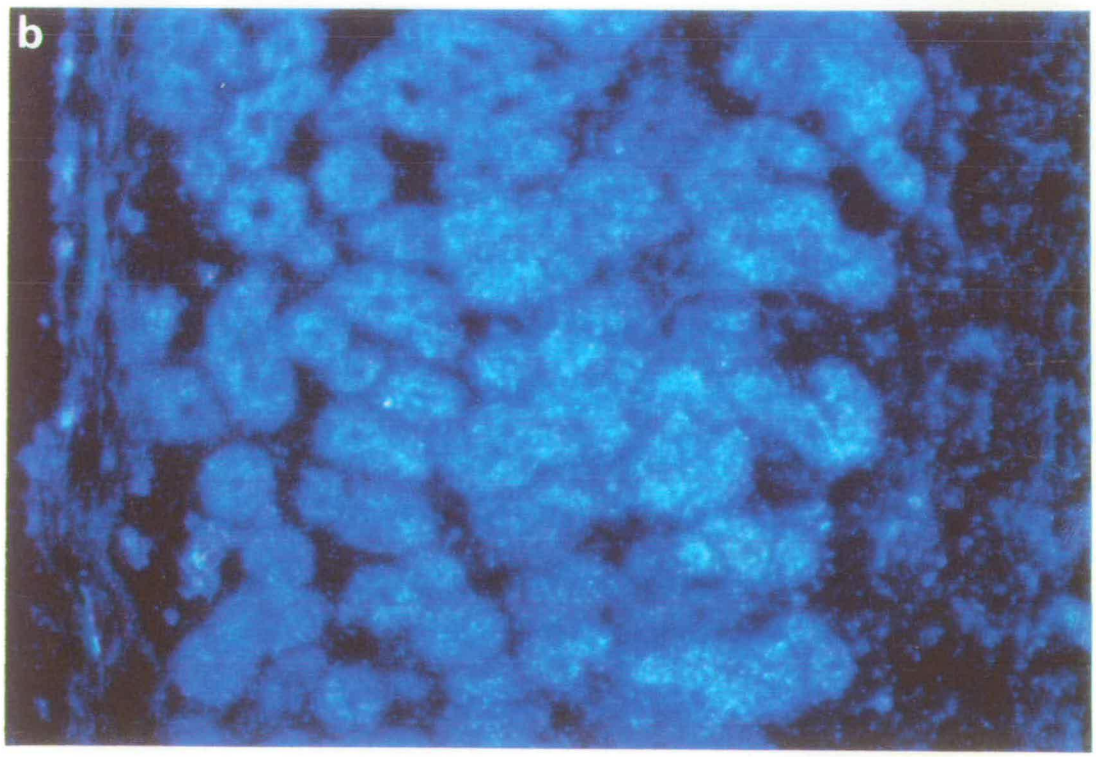

a
o
h
im
b
v
c
h
fl
im
hl
d
v
e
h
hl
fl
im

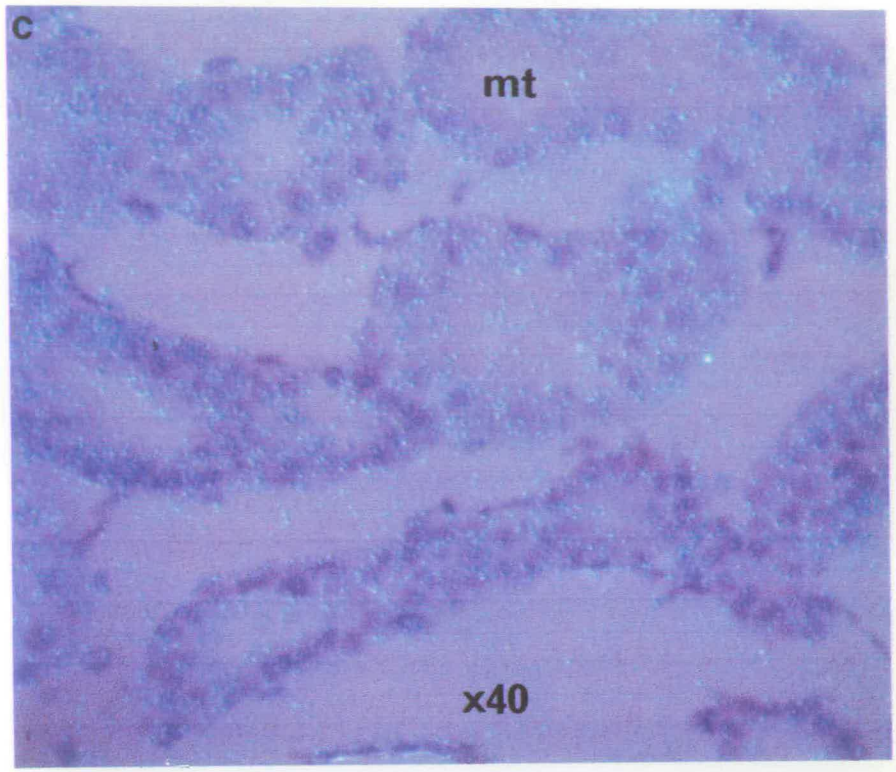
Figure 6.6 *WT1* in stage 13-14 alligator embryos

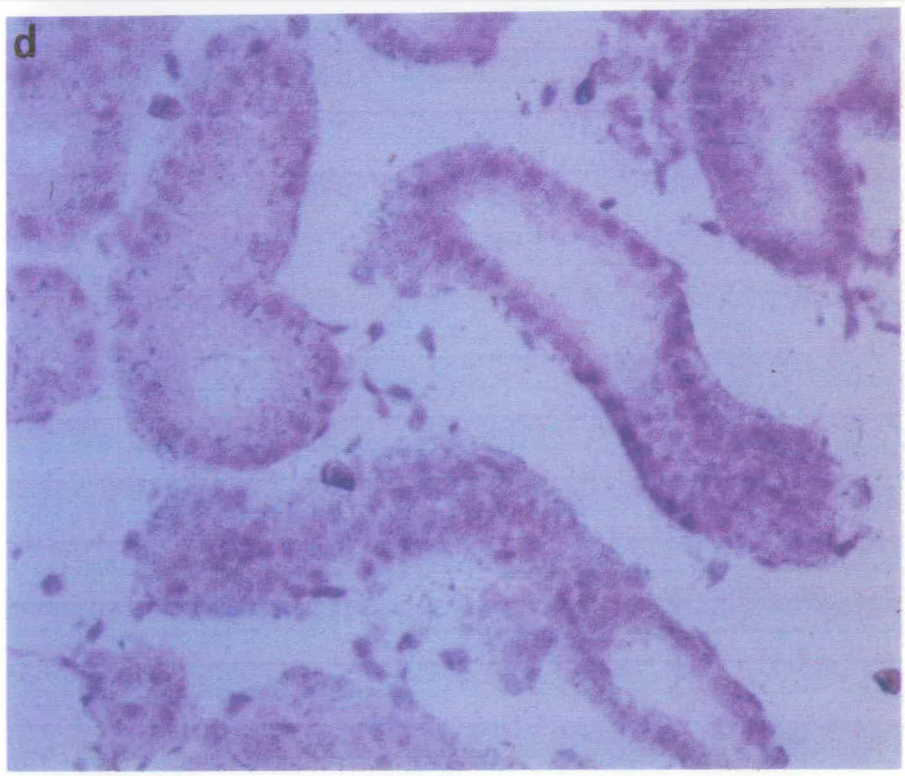
a-c. Stages 13-14. The abdominal body wall has formed so the mesonephros is no longer visible. *WT1* expression in the limbs has become to limited specific regions in the forelimbs.
The body wall is also expressing *WT1*, from the neck region down. The expression seen in the heart is probably due to a high level of expression in the mesothelium.
(bw body wall, fl fore limb, h heart)


a
h
fl
b
c
bw


a
gl
mt
×20


b


c
mt
x40


d

6.3.4 *Xenopus* whole-mount hybridisation

In the *Xenopus* embryos probed with an antisense *XWT1* riboprobe (transcribed from PCR clone *XWT1*) no convincing signal was detected, compared with the sense control. This could be because the probe that was used was too short. It was only 277bp long and the minimum recommended length is 500bp. The isolation of a larger region of *XWT1* had been attempted by library screening but no clones were obtained. Alternatively the stages used may have been too early. This probably not the case as in the oldest embryos used (stage 39) pronephric tubules were apparent and the nephric duct had extended along the length of the embryo (Lynch and Fraser 1990) and so the induction of the urogenital ridge would be expected to be underway and therefore *WT1* should be expressed. The dissection of *WT1* expression in *Xenopus* will have to await the isolation of a longer probe.

6.4 Comparison of *WT1* expression

In contrast to the detailed information about the chick genitourinary system, information on the timing of the onset of alligator nephrogenic differentiation is limited. The earliest reports observe the presence of what is thought to be the nephric duct rudiment in stage 1 embryos. In stage 3 embryos 5-6 nephric tubules are present at the level of the heart and this is therefore thought to be the onset of pronephric development (Reese 1908). This corresponds with the timing of pronephric development observed in chick. By stage 5, tubule formation has extended as far as the liver (Reese 1908). The metanephric anlagen has been observed at >40 somites (St. 12) (Forbes 1940) and many metanephric tubules are present by stage 20 (Reese 1908).

In the chick, expression was seen all along the urogenital ridge at stage 16. This is just before the first mesonephric differentiation is observed (stage 17), but after the mesonephric duct has reached the cloaca (stage 15) and therefore at the stage of induction of the mesonephric blastema (Romanoff 1960). The expression in chick just precedes the first observation of *WT1* expression in the anterior end of the urogenital ridge in the alligator. The stage 6 alligator in which this was seen was approximately equivalent to a stage 17 chick, and so *WT1* expression in the alligator is detectable around the times when pronephric and mesonephric differentiation is occurring in the chick, but with a slightly later onset than seen in the chick. Expression was detected in the region where the first tubules are observed in the alligator (somites 4-10), but after their differentiation has been reported. It maybe that in the pronephros, where

the nephrons have external glomeruli *WT1* is regulated in a slightly different manner, or that WT1 levels were below the detection threshold. In both species *WT1* expression extends the length of the urogenital ridge from the level of the heart (similarly to mouse) as far as the cloaca. In both species evidence of localisation to the developing nephrons has been found. These results agree with the timing and localisation of *WT1* expression in the mouse mesonephros (Armstrong *et al.*, 1992).

Several other sites of *WT1* expression were observed in alligator embryos. The possible neural expression is of interest. The expression of *WT1* in the roof of the developing hind brain has not been detected in any species apart from the alligator. Expression in the newborn rat (Sharma *et al.*, 1992), and in 15 day mouse brain and in neural tissue of the spinal cord of the developing human, mouse and rat (Pritchard-Jones *et al.*, 1990, Armstrong *et al.*, 1992, Sharma *et al.*, 1992) has been reported indicating that *WT1* can be expressed in tissues which are ectodermally derived in addition to those of mesodermal origin.

Expression is also present in the lateral plate mesoderm which will give rise to the connective tissue, blood vessels and musculature of the body wall. A transient localisation of signal in the developing striated musculature of the body wall has been observed in day 13 mice (Armstrong *et al.*, 1992). This is a particularly interesting property considering that ectopic striated muscle is histologically observed in Wilms' tumours whereas only smooth muscle is present in normal kidneys (Mierau *et al.*, 1987). The alligators in which this signal was observed were of a similar developmental stage to the mice indicating a possible conserved role for *WT1* in the developing body wall. The alligator expression in the limb buds and the vitelline veins represent new sites of *WT1* expression. The observed heart expression may be analogous to the mouse and human pattern where *WT1* is expressed in the overlying mesothelium which forms part of the pericardium (Pritchard-Jones *et al.*, 1990).

Chapter 7

Discussion

The evolutionary history of *WT1* in the vertebrate classes has been illuminated by the findings presented above.

- *WT1* orthologs have been isolated by a combination of PCR and library screening from representatives of all the vertebrate classes.
- A high degree of conservation, attributable to purifying selection, has been observed between the predicted protein sequences.
- Conservation is particularly stringent in the zinc fingers which are almost identical in all vertebrates.
- The human genomic structure which is one exon for each zinc finger domain has been conserved in zebrafish, including the presence in this species too of the alternative splice donor sequence that is responsible for the KTS insertion between ZF3 and ZF4.
- However two homopolymer domains in the transregulatory domain, the 17 amino acid alternative splice, and the N-terminal region of exon 6 are not highly conserved.
- All residues shown to be mutated in Wilms' tumours, both somatic and constitutional, are conserved.
- The *WT1* expression pattern has been shown to be conserved in the developing genitourinary ridge, in particular the developing nephric tubules.
- Expression of *WT1* outside of the intermediate mesoderm derived tissues has been found in the alligator as well as in mammals.

This study has made a significant start to the comparative analysis of *WT1* structure and function in the vertebrates.

7.1 The evolution of transcription factors in vertebrates

Comparison of the predicted amino acid sequences has revealed very few substitutions in the *WT1* orthologs, particularly in the zinc fingers. The conservation of other transcription factors for which a number of orthologs have been isolated was compared with *WT1*. Table 7.1 shows the percentage conservation for a selection of transcription factors within the vertebrate lineage. The figures were either taken from the quoted paper or calculated using the gap program of GCG from sequences in the GenBank database. The figures are similar to those obtained for *WT1* (Table 4.3 & 4.6), and with all of these there is a striking increase in homology within the

Table 7.1 Comparison of the conservation of transcription factors and their DNA binding domains

Gene	Species Compared	% Conserved Amino Acids in the Entire protein	Region of increased conservation	% Conserved Amino Acids in this Region	Reference
p53	Human-Mouse	85.5%	Regions 4&5	96%	Soussi *et al.*, 1990
	Human-Chick	69%		92%	
	Human-*Xenopus*	51%		100%	
	Mouse-*Xenopus*	57%		100%	
	Chick-*Xenopus*	69%		100%	
Krox-20	Human-Mouse	90%	Zinc fingers	100%	Oxtoby and Jowett 1993
	Human-Zebrafish	81.5%	Zinc Fingers	98.7%	
c-myc	Human-Marmoset	96%	Basic Helix loop helix	98%	Whitfield *et al.*, 1993
	Human-Mouse	94%	Basic Helix loop helix	94%	Schreiber-Agus *et al.*, 1993
	Human-Zebrafish	61%	Basic Helix loop helix	81%	
	Mouse-Zebrafish	63%	Basic Helix loop helix	83%	
	Xenopus-Zebrafish	65%	Basic Helix loop helix	83%	
	Trout-Zebrafish	77%	Basic Helix loop helix	98%	
max	Human-Mouse	98%	Basic Helix loop helix	100%	
	Human-Zebrafish	86%	Basic Helix loop helix	100%	
SRY	Human-Marmoset	61%	HMG box	95%	Whitfield *et al.*, 1993
	Human-Mouse	N/D	HMG box	80%	
SOX3	Human-Mouse	94%	HMG box	>94%	
Brachyury (T)	Mouse-Zebrafish	69.7%	DNA binding region	84.5%	Schult-Merker *et al.*, 1992
	Xenopus-Zebrafish	76%	DNA binding region	94%	
Snail	Mouse-*Xenopus*		Zinc fingers2-5	86%	Hammerschmidt and Nusslein-Volhard 1993
	Mouse-Zebrafish		Zinc fingers2-5	83%	
	Xenopus-Zebrafish		Zinc fingers2-5	88%	
	Zebrafish-Drosophila		Zinc fingers2-5	71%	
Pax 6	Human-Mouse	<99%			Ton *et al.*, 1992
	Human-Quail	98.6%	paired domain	100%	Martin *et al.*, 1992
	Human-Quail		homeodomain	100%	
	Human-Zebrafish	96%	paired domain	100%	Krauss *et al.*, 1991
	Human-Zebrafish		homeodomain	100%	
	Quail-Zebrafish	96%	paired domain	98.2%	
	Quail-Zebrafish		homeodomain	100%	

presumed DNA binding domains as compared to the rest of the protein. For both tumour suppressor genes (*WT1* and p53) the pathogenic mutations found have been almost all located in these conserved regions (Soussi *et al.,* 1990). This confirms the supposition that positions which have been found to be mutated in pathological states in humans are conserved in evolution. A high level of conservation has been observed for many transcription factors. Changes in the regulatory genes in development could be under a high level of selection because changes in these may affect more than one pathway so it may be a rare event that changes one of these proteins so that there is no deleterious effect on any of the pathways.

7.2 Rate of evolution of *WT1*

The comparison of the rate of change at synonymous and nonsynonymous sites demonstrated that the former was far greater than the latter, indicating that the slow rate of change of the protein sequence resulted not from an intrinsically slow rate of nucleotide mutation but was due to selection. By comparing the amount of sequence divergence with divergence time a linear trend was seen. The major anomaly seen was that the rate of change in the reptilian lineage seems to be lower than expected. It is difficult to make definite conclusions about rates in different lineages because of the considerable uncertainty about the true divergence times. Differences in the rate of evolution have been seen in a number of taxa, and it has been observed before that the rate of sequence change in the aves is lower than in some other lineages (Britten 1986). Attempts have been made to account for these differences in the framework of the theory of a molecular clock (Wilson *et al.,* 1987). The idea that local clocks exist as opposed to a single global clock, seems to have gained greater acceptance. There is probably a range of variables involved in the rate of sequence change, including generation time effect (Kohne 1970), DNA repair (Britten 1986), or metabolic rate (Martin and Palumbi 1993).

7.3 The transregulatory domain

The mechanism by which sequence specific transcription factors, such as *WT1*, act to activate or repress expression from their target genes is just beginning to be elucidated. There seems to be a common theme for the diverse regulators, in that they affect transcription from a wide variety of promoters and they can function synergistically and co-operatively with other factors. The stage at which these factors are now thought to act is at the initiation of transcription, by modulating the

formation of the complex between the RNA Polymerase II complex and its associated general transcription factors, which is now known to be a highly ordered multi-step process (Zawel and Reinberg 1993). Although the DNA binding regions in these transcription factors show a limited number of recognisable motifs e.g. zinc fingers, homeodomains, and bZip motifs, the only similarities between putative transregulatory domains is a richness in certain sorts of amino acids; the acidic domains or proline or glutamine rich (Mitchell and Tjian 1989). Experiments to define the functionally important regions within these domains have been tried. In several transcription factors different regions have been found to be acting independently to activate or repress transcription (Russo *et al.,* 1993, Chi and Carey 1993). In most experiments the effect of *WT1* has been to repress transcription from reporter genes and it seems that disturbing the structure of the transregulatory region of *WT1* always disturbs the repression effect of the protein (Madden *et al.,* 1991, Drummond *et al.,* 1992, Gashler *et al.,* 1992, Wang *et al.,* 1992, Werner *et al.,* 1993, Wang *et al.,* 1993a). It had been hoped that the pattern of conservation of this region of *WT1* would provide insight into the architecture of this region. No strikingly modular pattern of conserved and non-conserved residues was observed for *WT1* which would point to functional and non-functional regions. In some ways the high level of conservation of this region was surprising considering only a very small percentage of mutations occur in this region. When comparing mutated regions with evolutionarily conserved ones, evolutionary conservation probably provides a better picture of functional regions, as mutations will only be observed if the phenotypic change is large enough to be noticed but is not lethal. In addition only a subset of genitourinary abnormalities, mostly Wilms' tumours, have been analysed for *WT1* mutations. As yet only somatic mutations have been observed in the transregulatory domain and it may be that a constitutional change in this region of *WT1* may produce a different phenotype to constitutional zinc finger mutations.

The major difference observed was in the two homopolymer domains of polyproline $(Proline)_{13}$ and polyglycine $(Glycine)_5$. The conserved, higher than average levels, of proline, glutamine, and serine are consistent with a non random distribution of amino acids found in transcription factor transactivation domains. Also using the Chou and Fasman program for prediction plots of α helix, β sheet and β turns and the Kyte and Doolittle program for estimation of hydrophobicity no great changes between the different species was shown (Appendix B). The high level of conservation of the *WT1* orthologs points to the functional importance of this region suggesting that *WT1* is not merely acting as a structural antagonist of *EGR1*

(effectively blocking *EGR1* binding to its target promoters), and repressing growth in that manner.

7.3.1 Evolution of homopolymer domains

The mechanism by which the difference in these domains arose can be speculated on. $(Proline)_{13}$ is not present in the chick *WT1* ortholog and the region encoding $(Glycine)_5$ has not been conserved. It is not known whether or not the $(Proline)_{13}$ region has been inserted or deleted in *WT1,* since the mammalian/reptilian divergence. To answer this question will require the isolation of this region from more diverged species. Insertions or deletions can arise from slippage of the DNA polymerase during replication resulting either in replication of a region already replicated or skipping of a region. Slippage is known to happen in the regions of DNA containing contiguous short repeats due to the increased chance in mispairing between these repeats (Li and Graur 1991, Richards and Sutherland 1994). It could be speculated that the two regions $(Glycine)_5$ and $(Proline)_{13}$ in *WT1* have been produced by insertion or expansion of a trinucleotide repeat of CCN and GGN which encode proline and glycine respectively. $(Glycine)_5$ may have been produced prior to mammalian/reptilian divergence but was only conserved in the mammalian lineage. Both of these homopolymer motifs are conserved in three mammalian orders, primate, rodent and artiodactyls (Figure 5.3), although the number of residues is not completely stable, indicating that some replication slippage is occurring.

There is growing interest in these homopolymer motifs and their origin. This has been aroused by a new sort of pathogenic mutation, the expansion of trinucleotide repeats, called dynamic mutations (Richards and Sutherland 1992). The first incidence was discovered in the *FMR* gene. It was found that an increase in the number of CCG repeats in the gene was associated with Fragile X syndrome. Carriers are found to have a greater than normal number of repeats, and in affected individuals this is even further increased. The number of disease genes where this has been found to occur is expanding, e.g. X-Linked spinal and bulbar muscular atrophy (La Spada *et al.,* 1991), myotonic dystrophy (Brook *et al.,* 1992, Fu *et al.,* 1992, Mahadevan *et al.,* 1992), *FRAXE* site (Knight *et al.,* 1993) and Huntington's Chorea (The Huntington's Disease Collaborative Group 1993). The position within the gene is not the same for all diseases; expanded repeats are found in coding or non coding regions, suggesting that the expansions can interfere with gene expression or function in a variety of ways. Only two trinucleotides have been found to be expanded in this way, CCG and CAG;

Table 7.2 Genes containing homopolymer domains

Gene	Homopolymer Domain	Reference
EGR2	Proline	Joseph *et al.*, 1988
Androgen receptor	Glutamine	Tilley *et al.*, 1989
Yeast *ADR6*	Asparagine, Threonine	O'Hara *et al.*, 1989
Human calcineurin A	Proline	Guerini and Klee 1989
Glucocorticiod Receptor	Glutamine	Evans 1989
EBNA2	Proline	Horvath and Schwbach 1993
Mouse *Sry*	Glutamine	Gubbay *et al.*, 1990
GAL11	Glutamine	Suzuki *et al.* 1988
MCM1	Glutamine	Passmore *et al.* 1988

it is not known whether this is because these trinucleotides are more susceptible to expansion or whether it is due to the protein products produced. The mechanism by which these expansions occur has been elucidated in one gene: spinocerebella ataxia type 1. In this gene the normal repeat has a single base change interrupting it; if this is mutated back to perfect repeat it becomes unstable (Chung *et al.,* 1993). Instability of di, tri, and tetranucleotide repeats has now been seen in some cancers arising from mutations in the DNA repair mechanism (Richards and Sutherland 1994).

On searching databases many other genes have homopolymer domains which may also be attributable to the insertion or expansion of a trinucleotide repeat during evolution but in a much more minor way than that seen in disease (Table 7.2). As yet no expansions in *WT1* have been found but it may be speculated that if the region encoding the polyproline region became unstable, expansion could be a mechanism of gene inactivation.

7.3.2 Function of homopolymers

Homopolymer domains are thought to play a role in transcription regulation, since proline and glutamine homopolymers are mostly found in transcription factors. Fusion of these to *GAL4* DNA binding domains show that they can change the level of transcription from a reporter gene *in vitro* (Gerber *et al.,* 1994). The deletion of $(Proline)_{13}$ and $(Glycine)_5$ homopolymer domains from the human *WT1 in vitro* caused only a small reduction in the repression of transcription from reporter genes (Madden *et al.,* 1993, Wang *et al.,* 1993a). Although $(Proline)_{13}$ is interrupted at two positions by alanine residues, these are related to proline by only a single change in the nucleotide sequence i.e. CCN to GCN. This may be due to a selection for a functional domain as proline homopolymer domains have their maximum effect *in vitro* at 10 residues long (Gerber *et al.,* 1994).

From the Chou and Fasman structure prediction plots no structure is predicted for $(Proline)_{13}$ and $(Proline)_4$ (Appendix B). Studies on polyproline in solution have found that two structures, type I and type II helices can be formed, dependent on whether the α carbon is in the *cis* or *trans* conformation (Mandelkern 1967). Therefore in *WT1* these sequences may be forming this rigid rod type of structure if exposed to the solvent on the surface of the protein. The observation that *WT1* protein can exist with or without (in the chick) the $(Proline)_{13}$ sequence but without any other large sequence changes suggests that there is a certain degree of structural flexibility in this region. This would suggest a surface location, as seen for this type of

helix in other globular proteins (Adzhubei and Sternberg 1993). Therefore in mammals it may be involved in the type of protein-protein interactions seen with other proline rich sequences e.g. SH3 target binding sites, which have been shown to have a proline type II helix (Yu *et al.*,1994). The function of $(Glycine)_5$ is unknown. Glycine having no side chain is the residue with the greatest propensity for rotation and therefore this region may be expected to be flexible, which is consistent with the prediction of a high level of probability of a β loop in this region (Appendix B).

Why these regions have been conserved in mammals but not in birds may only be revealed by a comparison of the *in vivo* interactions made by the transregulatory domain with other proteins. It could be that there are other changes in the proteins with which it associates which compensate for the *WT1* differences. Another possibility is that the insertion or deletion event which produced the difference in $(Proline)_{13}$ is not as deleterious as substitutions occurring in this region so the region can be wholly conserved or not at all.

7.3.3 Conservation of the putative leucine zipper

At the C-terminal end of the transregulatory domain there is a putative leucine zipper. This has been proposed to be the means by which *WT1* protein molecules could dimerise. It is of particular interest because it is thought that the DDS phenotype arises as a result of a dominant-negative mutation, with the mutant *WT1* possibly interacting with wild type *WT1* through this domain. The conservation and lack of helix breaking prolines in all species would support the existence of leucine zipper. In conflict with this, structure prediction algorithms do not predict the presence of a coiled coil (leucine zipper) in this region. Therefore it may be that interaction of *WT1* protein molecules, leading to a dominant-negative phenotype, occurs in a different manner.

7.4 A region of *WT1* with species-specific function?

The most variable region between the orthologs is located in exon 6, immediately adjacent to a perfectly conserved region. Comparison of the chick and alligator orthologs reveals throughout the protein that the same positions are often substituted when compared with mammalian *WT1*. N-terminal to exon 6, the amino acids found at these substituted positions are often different in chick and alligator, which points to flexibility of the sequence in this region. In the highly substituted

region of exon 6 substituted amino acids are almost identical in chick and alligator. This suggests that some selection may be acting on this region which is different in the two lineages i.e. the sequence difference did not arise from sequence drift due to lack of function but a change in functional properties. This region is adjacent to the 17aa alternative splice (exon 5) which has not been detected in chick or alligator. This region is thought to have a function in the human for several reasons. A mutation in this region was found in a benign mesothelioma and is associated with a change in transrepression activity (Park *et al.,* 1993a). In the same Wilms' tumour cell line that excluded exon 2 from transcripts, exon 5 was also not included, suggesting a role for exon 5 in tumorigenesis (Haber *et al.,* 1993). Also the inclusion of exon 5, converts the +KTS isoform of *WT1* from a weak repressor into a strong repressor of *WT1* transcription *in vitro* (Rupprecht *et al.,* 1994).

Species differences in regulatory genes are very important in development as they may be the source of species-specific features. The genitourinary system in which *WT1* is expressed has many specialisations within the vertebrates; the most obvious of these being the development of the metanephros. In addition there are differences between classes in the organisation of the genital and urinary ducts. In particular the female genital ducts become highly specialised for viviparity and oviparity. Therefore the changes in exons 5 and 6 may give rise to subtle changes in *WT1* activity, which then plays a part in developmental variations in the intermediate mesoderm.

7.5 Selection of alternative splice sites

The conservation of the KTS alternative splice agrees with its importance as suggested by human mutation analysis. The importance of the + and - KTS isoforms being present at the correct ratio is underlined by the observation of two independent heterozygous mutations at the alternative splice site in DDS patients which completely inhibits the inclusion of this sequence. The perfect conservation of the zebrafish intron in the region adjacent to the alternative splice donor sites at the end of exon 9 which results in the inclusion of the KTS sequence, points to the context of a sequence influencing splice site choice. This seems to apply even when the two sites are only 9 bases apart, significantly this is almost one turn of the B-DNA helix. The mechanism by which alternative splicing occurs is unknown but it is thought to be closely related to constitutive splicing, the main elements of which are the snRNPs (small nuclear ribonucleoprotein particles) and the SR proteins (these contain

alternating serine and arginine residues in the C-terminal domain). There are several evolutionary conserved SR proteins, although the presence of only one seems to allow splicing to occur. Therefore the recruitment of different SR proteins may affect the splice site chosen (Zahler *et al.,* 1993). It may be that the sequence of intron 9 is conserved to favour the binding of particular SR proteins. The situation is complicated as there seem to be multiple ways by which a transcript can be alternatively spliced (McKeown 1992).

The conservation of the KTS alternative splice is consistent with information from other genes. Alternative splicing is found in many developmentally important transcription factors, in *PAX6* the same alternative splices have been detected in human as in quail (D. Englekamp personal communication) and the alternative splice 5a which inserts 14aa in the paired domain is identical in mouse and zebrafish (Glaser *et al.,* 1992). In *max* (a protein that dimerises with *c-myc*) the same alternative splice is seen in zebrafish and in human (Schreiber-Agus *et al.,* 1993).

To try to understand why exon 5 was not present in non mammalian transcripts the splice sites in the cDNAs were examined. Although most of the splice site consensus occurs in the intron there is a bias in the first two positions in the exon. Some of the least typical splice sites occur around exon 5. Non-consensus splice sites have been implicated in alternative splicing (Jackson 1991). Therefore these differences may be the reason why exon 5 is excluded from transcripts in chick and alligator. The isolation of the relevant genomic sequences will be able to resolve whether the sequence of exon 5 is actually conserved outside of mammals.

RNA editing is thought to be functionally associated with splicing as *in vitro* cross species editing sometimes results in correct editing but incorrect splicing. It may be that the differences in the editing of exon 6 (Section 5.2.7) and the splicing of exon 5 i.e. the 6bp insertion in the chick and alligator at the editing position in mammals and the lack of a +17aa isoform in the reptilian lineage, is in some way connected. The editing in rat was found to occur in a developmentally regulated manner and was first observed after birth when kidney development is virtually complete. RNA obtained for chick and alligator was only obtained from embryonic sources so it may be interesting to look at later stages to see if any editing could be detected.

7.6 Evolution of the zinc fingers

The data obtained here show that the rate of evolution of the zinc fingers of *WT1* is slow, and slower than the rest of the protein (Table 5.5). This is in agreement with evolutionary analysis of other zinc finger genes. The comparisons of *Krox 20* and *Krox 24* (*EGR2&EGR1*) (human to fish) (Lanfear *et al.,* 1991, Oxtoby and Jowett 1993); *snail* (mammals to *Drosophila*) (Sommer *et al.,* 1992, Hammerschmidt and Nusslein-Volhard 1993); *ZFY* family (mammals to birds) (Lanfear and Holland 1991), *Kruppel* and *hunchback* (arthropods) (Sommer *et al.,* 1992) show a pattern of substitution very similar to that obtained for *WT1*. The conservation between runs of fingers and homologous fingers has been analysed statistically (Jacobs 1992) and the substituted positions closely resembled the ones found in this study i.e. mostly between the two cysteines or the two histidines or after the second cysteine but not at the 'tip' of the finger. The pattern of substitution would suggest that the *WT1* orthologs retain the same DNA binding consensus sequence as human *WT1*. Although variability is limited to regions of no known function, other positions for which no function is known were also found to be invariant. At these positions mutations have not been observed and they are not conserved in the *EGR* fingers. The lack of mutations of these sites could arise from the rarity of *WT1* mutations, most occurring at CpG dinucleotides. Mutations at the invariant sites could be deleterious enough to be selected against but do not produce abnormalities that have been investigated. The fact that these residues are not present in the *EGR* zinc fingers but are conserved in *WT1* would suggest that these residues are acting to modify the properties of the fingers thereby making the DNA sequence specificity of *WT1* subtly different from that of *EGR1* and *2*.

Families of similar zinc fingers are now being recognised such as the *Kox* and *Fax* families (Nietfield *et al.,* 1993, Tunnacliffe *et al.,* 1993). The latter have similar zinc fingers but also contain the Fax motif. They are found in clusters in the vertebrate genome indicating that they have probably arisen by duplication of an ancestral gene. *WT1* belongs to the *EGR* family by virtue of the similarity of its fingers; this family includes *EGR1* (Sukhatme *et al.*, 1988), *EGR2* (Joseph *et al.*, 1988), *EGR3* (Patwardhan *et al.*, 1991), and *EGR4* (Crosby *et al.*, 1992). These have three fingers, ZF1 and ZF3 having a very similar consensus sequence in the DNA binding region (type 1) while ZF2 is slightly different (type 2) (Figure 1.2c). *WT1* fingers 2-4 are very similar to the EGR fingers in the DNA binding region but *WT1* has an extra zinc finger (type 3) which is most similar to the first finger in the *Sp1* family. The *Sp1* family are a set of genes which also have three fingers, similar to the

first three fingers of *WT1*. This includes *Sp1* and *Sp2* (Kadonaga *et al.*, 1987, Kingsley and Winoto 1992), *SPR-1* and *SPR-2* (Hagen *et al.*, 1992), *Drosophila buttonhead* (Wimmer *et al.*, 1993), and rat *BTEB* (Imataka *et al.*, 1992). The *EGR*-like second and third fingers are slightly different in that the first predicted backbone contacting residue is a lysine not an arginine, but this is a conservative substitution of one amino acid with a basic side chain for another. The type 1 finger is also found in two yeast genes, *MIG1* in *Saccharomyces cerevisiae* (Nehlin and Barnes 1990) and *Cre-A* in *Aspergillus nidulans* (Dowzer and Kelly 1991). Both genes are involved in the regulation of catabolite metabolism and have two zinc fingers in tandem, the second one resembling the type 1 finger.

The evolutionary relationship between these fingers is unknown; it is not clear whether it is due to homology arising from a common ancestral gene or whether the similarity has arisen as a result of adaptive changes producing convergent evolution for GC rich DNA binding sites. The most striking resemblance is not the homology of the individual fingers but the way in which the different fingers are in a similar order. Amongst these genes, for which the genomic structure is known, *WT1* is unique in having each finger in a separate exon as opposed to the others where the fingers are in a single exon. The intron/exon junctions occur at the same position in each finger of *WT1* and in the same phase, specifically within the glycine of the linker region. This is a similar genomic structure to that of *TFIIIA* in *Xenopus* (Miller *et al.*, 1985, Tso *et al.*, 1986), where all nine fingers are in separate exons with neighbouring fingers having similar exon junctions. From this arrangement it has been proposed that these fingers arose by duplications of an initial exon, as seen in other independent structural motifs such as kringles and growth factor domains (Patthy 1985). The process by which these modules of structure are duplicated and moved around is known as exon shuffling (Doolittle 1981). It has also been proposed that these genes have given rise to other genes with multiple fingers per exon (El-Baradi and Pieler 1991). This could have occurred by loss of introns or by reverse transcription of the mRNA transcripts and integration into the genome. Reverse transcription is a not an uncommon event, with the amount of the vertebrate genome produced by reverse transcription events estimated to exceed 10% (Temin 1985).

Therefore is *WT1* ancestral to the *EGR* or *Sp1* families? No similarities between *WT1* and the *EGR* or *Sp1* families are found in addition to the homology of the zinc fingers. Whereas within the *Sp1* and *EGR* families further homologous regions have been observed in the transactivation domain (Crosby *et al.*, 1992, Wimmer *et al.*, 1993). Therefore if they are related to *WT1* by ancestry there has been a complete

divergence of the rest of the gene or the gene was only partially duplicated. For *WT1* to be the ancestral sequence it would have to have evolved far earlier than so far detected as the *Sp1* homolog *buttonhead* has already been described in *Drosophila.* Intense search for, isolation and comparison of more orthologs will answer this question. The total genome sequencing programs that are underway in several species will permit the presence or absence of these genes to be definitively determined and provide information about the origin of these genes.

Convergent evolution may have been a major determinant of the similarity between these genes (Doolittle 1994). This could be a result of the limitations imposed on the sequence of the fingers by the fairly rigid structure holding it together and their interaction with a very invariant structure i.e. DNA. Database searching for *TFIIIA*-like fingers shows that the residues in common between the *WT1, EGR,* and *Sp1* fingers are very common in other fingers. This supports the fact that there are only a limited number of options allowing DNA binding. These similarities combined with the number of reported fingers (1340 in 1992, Jacobs 1992) means that many fingers may appear related but are not. Then again, it may be speculated that all fingers are related as they all may have arisen from a single ancestral finger. This ancestral finger probably arose after the divergence of pro- and eukaryotes, as only eukaryotes have been found to contain the *TFIIIA*-like zinc fingers.

7.7 Conservation of expression of the *WT1* gene in the intermediate mesoderm

PCR and Northern blot analysis of pig, chick, alligator and *Xenopus* RNA from the intermediate mesoderm derived organs revealed expression of *WT1.* This demonstrates that in addition to the conservation of *WT1* at the level of protein sequence there is also conservation in the pattern of gene expression. This would indicate that at least part of the function of *WT1* has been conserved in mammalian, reptilian and amphibian lineages. In the case of *Xenopus, WT1* expression in whole tadpole and mesonephros were compared. It was found that *WT1* expression was only detectable in the mesonephric RNA. This would suggest that expression is specific to the mesonephros. Unfortunately the samples compared were not from the same stages, so the presence of *WT1* could correspond to a general up-regulation of expression in the embryo, but it is more likely that it is due to localisation of *WT1* in the differentiating kidneys.

In pig, chick and alligator the size of *WT1* transcript was found to be conserved when compared to human and mice. In *Xenopus* the size of the mRNA transcript had

increased to 5kb. Until a more complete *Xenopus WT1* cDNA is cloned it will not be known whether this is due to a change in the coding or non coding regions.

In chick an additional 2kb transcript was also seen; its identity is unknown. Interestingly a smaller 2.5kb testis specific transcript has been detected in mouse and human (Pritchard-Jones *et al.,* 1990, Pelletier *et al.,* 1991c), and also in the developing rat kidney (Sharma *et al.,* 1992). Only the full length RNA was detected in pig testis of an equivalent age to when smaller transcripts are detected in mouse and human. In the chick in the mesonephric RNA alone no second transcript was detected so the small transcript may be gonad specific (Figure 6.1). The origin and functional importance of these smaller transcripts is not known; they may not be essential products in all species and could be transcriptional artefacts. These may result from a high level of transcription initiation with a second cryptic promoter being used or they maybe produced by the use of the stronger polyadenylation signal located in the 3' untranslated region in all species examined (Section 5.2.11).

7.7.1 Urinary system

In both the chick and alligator, expression of *WT1* was detected by *in situ* hybridisation in the developing genitourinary ridge and localised to the developing nephrons. This demonstrates that there is a role for *WT1* in nephrogenesis outside of the mammals. Consistent with a role in the earliest differentiation of the nephrons, the expression of *WT1* in chick is observed with the onset of mesonephric differentiation. The conservation of expression suggests that Wilms'-like tumours in other species (such as pig, rat, chick and Japanese eel) could also result from loss of *WT1* function disturbing nephrogenesis. This may now be investigated by looking in these tumours for overexpression of *WT1* and mutations, both of which occur in human tumours. This will permit the proportion of *WT1* mutants found to be compared to the 10% incidence seen in human sporadic tumours. Wilms'-like tumours can be experimentally induced in many vertebrates, so that a greater number of tumours, as compared with the relatively few human cases, can be analysed. Therefore do these treatments affect the *WT1* gene or another WT locus? Another unresolved issue is the relationship of *WT1* with *novC*, the *IGF1* binding protein-like gene implicated in chick virally induced tumours (Joliot *et al.,* 1992). Is this gene in the same pathway as *WT1*, either upstream or downstream?

Wilms' tumours are just one of many kidney diseases (Mierau *et al.,* 1987). The involvement of *WT1* in other kidney malfunctions, apart from the nephropathy seen in

DDS, is not known. The role of *WT1* in other conditions can now be investigated. As suggested earlier constitutional mutations of *WT1* in regions apart from the zinc fingers may produce different phenotypes. Rats are thought to be a considerably better model for kidney development and Wilms' tumour, as Wilms' tumours are seen in rats, but not in mice. Currently *RWT1* is being used to investigate *WT1* expression in the postnatal rat, with the specific aim of seeing whether there are changes in *WT1* expression in pathological states known to be associated with changes in renal growth. Kidney infection in early life (infantile pyelonephritis) which can interfere with renal growth can lead to permanent kidney damage, and is the greatest cause of kidney failure in later life. Preliminary results suggest that *WT1* expression is affected on infection, so that modulation of *WT1* activity on infection may help to avoid the permanent damage (E. Ostlund, Karolinska Hospital, Sweden).

7.7.2 Gonads

In stage 20 alligator gonads, just preceding the onset of sexual differentiation (stage 21-22 testis and stage 22-23 ovary), *WT1* expression was detected by Northern blotting and PCR. Expression appeared to be at approximately the same levels at both male and female producing temperatures. In mammals the indifferent gonad also expresses *WT1*.

The processes that lead to gonadal sex determination in vertebrates seem to be highly variable: sex can be determined genetically, environmentally or a mixture of both. Mechanisms can vary within classes. In mammals it is the presence of the Y chromosome carrying the *SRY* gene which leads to the development of the male phenotype. *SRY* is also known as TDF (testis determining factor) because it is the differentiation of the testis from the indifferent gonad which is important in sex differentiation, and which *SRY* controls. Therefore is *WT1* functionally related to *SRY*? *WT1* is important for gonad development in both sexes, homozygous deletion of *WT1* in mice (Kreidberg *et al.,* 1993) results in an inability to form gonads, as opposed to the sex reversal seen when *SRY* is deleted. *WT1* is expressed from the earliest stages of genital ridge development i.e. in the 9 dpc mouse (Armstrong *et al.,* 1992) and is followed by *SRY* expression at 10.5 dpc (Koopman *et al.,* 1990). Putative *WT1* binding sites have been reported in the *SRY* gene promoter (Behlke *et al.,* 1993) suggesting that *WT1* could lie upstream of *SRY* (Bogan and Page 1994). The localisation of *WT1* expression in males to the Sertoli cells, which are important for sex determination, further suggests this. *WT1* however appears not to affect

primary sex determination but it can perturb secondary sex determination. In humans changes in dose (WAGR) or activity (DDS) of *WT1* lead to genitourinary abnormalities; these are likely to be due to disturbances in the development of the gonad which affects hormone production on which secondary sex determination is dependent.

In alligator sex determination is thought to be wholly temperature dependent. In males sexual differentiation of the gonad occurs during the temperature dependent stage, and in females just afterwards (Ferguson and Joanen 1982). Testis determination, similarly to mammals, is thought to be dependent on Sertoli cell differentiation (Ferguson and Joanen 1982, Smith and Joss 1993). In alligators many HMG-box containing genes including *SOX* genes have been identified but not an *SRY* ortholog (Coriat *et al.,* 1993). *SRY* is known to be evolving at a much faster rate than expected when compared with other genes (Tucker and Lundrigan 1993, Whitfield *et al.,* 1993) suggesting that *SRY* may be involved in sex determination only in mammals. Mammalian gonad development is known to differ from other vertebrates. In mammals gonadal sex reversal is very rare, but in other vertebrates this is seen to occur much more frequently e.g. intersexes in amphibians, temperature sensitive reversal of genetic sex in axolotl (Dournon *et al.,* 1990) or by exogenous hormonal disturbances e.g. chicken (Elbrecht and Smith 1992) and alligator (Lance and Bogart 1991&1992). This may suggest that in other vertebrate classes the production of steroid hormones in development may have a greater capacity to influence gonadal sex determination. It is known that steroid hormones can be produced in very early gonadal cells (Haffen 1970). The conservation of *WT1* in alligators and chick makes it likely that its role in early gonad development, and therefore in the hormone producing cells, is conserved. Therefore it is possible, in some species at least, that *WT1* plays a role in sex determination. The *WT1* alligator ortholog is being used as a tool to probe early differentiation, first as a gonadal marker and maybe subsequently to isolate the genes involved in temperature dependent sex determination (AM Coriat, Guy's Hospital, London). One of the other sub-classes in which sex determination is being investigated is the metatheria (marsupials). This is because this sub-class shows most of the features of mammalian (eutherian) sex determination i.e. a testis determining Y chromosome, but also some effects of X chromosome dosage (Graves 1987). Therefore metatheria probably separated from the eutheria as the Y chromosome/*SRY* system of sex determination was evolving. Again in marsupials the predicted protein sequence of *WT1* is highly conserved indicating that the mechanisms that control early gonad development maybe similar to those in eutherian mammals.

7.8 Additional sites of expression

Similarly to mouse and human, sites of *WT1* expression in alligator were observed outside of the intermediate mesoderm, in derivatives of the lateral plate mesoderm and the ectoderm i.e. the limbs, body wall, heart, vitelline veins, and hindbrain. In the mouse expression has been observed in the lateral plate mesoderm, specifically the musculature of the body wall, and in defined areas of the developing spinal cord and brain (Armstrong *et al.,* 1992). The expression in the heart is probably a result of the surrounding mesothelium expressing *WT1*. The vitelline veins contain endothelial lining cells which are a type of epithelium that differentiate from the angioblastic mesoderm, derived from the splanchic mesoderm. These cells also undergo a mesenchymal-epithelial transition which may explain why *WT1* is detected (Balinsky 1981). These cells in mammals have not been observed to express *WT1* but, like podocytes, a major site of *WT1* expression, they are a subset of epithelia with special permeability properties. These results indicate that *WT1* expression may not be limited just to the intermediate mesoderm. The importance of most of these expression sites of is not known. No defects have been reported in the *Wt1* knockout mouse, in the limbs or body wall (Kreidberg *et al.,* 1993). This may be because either that there is a species difference, or a redundancy of function in the developmental pathway, or that defects were too subtle to observe, or an effect of genetic background. In all these cases further detailed analysis of *WT1* expression in histological sections will or will not confirm the preliminary results above. If expression is seen, detailed localisation will be informative.

7.9 *WT1* and limb development

The expression of *WT1* in the alligator limbs is of interest because of the known association of limb and genitourinary deformities in several species (e.g. human, mouse and axolotl). This implies that in these species there may be similarities between these systems. The *WT1* expression in the limb buds has not been reported before. That it is a *bona fide* result in alligators is suggested by the subsequent localisation of the expression to presumptive wrist and armpit regions at the time of hand plate formation and its disappearance from the hind limbs.

There is evidence for a connection between kidney and limb development. The limb buds develop in close proximity to the mesonephros, and the mesonephros is known to promote the normal differentiation of chick limb explants in culture, partly due to the production of *IGF1* (Geduspan and Solursh 1993). The limb buds are

produced by an initial proliferation of the lateral plate mesoderm beneath the epidermis, cells then condense and transform into chondrocytes to produce the skeletal elements (Gilbert 1991). It is thought that the muscular and dermal elements are contributed the somitic mesoderm (Balinsky 1981). The differential morphogenesis of the fore and hind limbs is thought to arise as a result of the action of regulatory factors in the control of processes common to all four limbs, such as fibronectin accumulation and mesenchyme condensation, but these are differentially regulated in the limbs (Downie and Newman 1994). The expression of *WT1* in only the fore limb may reflect a role in this.

Several genes have been found to be expressed in both lateral plate and intermediate mesoderm derived structures, specifically in the developing limbs and kidneys. The bone morphogenetic protein *BMP-2* is expressed in the mesonephros and the nephric duct (C. Tickle personal communication) as well as in the limb buds, tooth buds, and craniofacial mesoderm (Lyons *et al.,* 1990). The *BMP* family were originally isolated due to their inductive capabilities in bone formation and belong to the TGF-β superfamily. The *TGF*-β family is also known to be expressed in the developing kidneys (Hammerman *et al.,* 1992). *BMP-2* is now thought to have a wider role than just osteogenesis, being important in mesenchymal-epithelial interactions, like those required for kidney formation (Rosen and Thies 1992). The presence of bone in Wilms' tumours could therefore be due to the nephrogenic cells responding to the presence of *BMP-2* when the normal developmental pathway is disrupted.

Ld (*limb deformity*) mice were produced by the disruption of a gene due to the insertion of a *c-myc* transgene and have deformed limbs and renal aplasia (Woychik *et al.,* 1990). The gene affected is the *formin* gene and its expression in mouse and chick has been studied. In the chick, the gene is expressed in the pronephros, in the developing tubules and the glomeruli of the mesonephros (Trumpp *et al.,* 1992). The protein is detectable in the limb bud as it differentiates, and the expression becomes concentrated in the condensing mesenchymal cells destined to form the long bones and digits. In addition to those sites affected by disruption of the gene, expression is seen in the notochord, floor plate and ventral horns of the developing chick neural tube. The latter includes the small region in the spinal cord seen to express *WT1* in mouse (Armstrong *et al.,* 1992).

There is another mutation whose effect seems to be localised to the genitourinary system and the limbs. This is the *short toes (s)* mutation in the Mexican axolotl (*Ambystoma mexicanum*). This homozygote lethal mutation results in renal

aplasia which seems to affect mainly the glomeruli, which are reduced in number or absent. There is also an overall disorganisation of the kidney structure. At the molecular level there is known to be a defect in the disposition of laminin A chain which is randomly distributed in mutants but localised to the basement membrane in wild type kidneys (Washabaugh *et al.,* 1993). Laminin A chain production is thought to play a role in the conversion of the mesenchyme to epithelia in the kidney because it is first detected at the earliest stages of nephron development (Ekblom 1989). The mutation also affects the development and regeneration of the limbs but not of the tail. Differences in laminin A chain deposition have also been observed in mutant limbs (Del Rio-Tsonis *et al.,* 1992). The investigation of a possible role for *WT1* has been greatly facilitated by the isolation of the axolotl *WT1* gene (P.A. Tsonis personal communication). Expression patterns in the mutant and wild type are now being examined (K. Del Rio-Tsonis personal communication) and mutational analysis of the gene is also planned.

It may be that the similar set of genes expressed in the limb, kidneys and the CNS reflect a similarity in the developmental decisions and morphological changes being made in the cells. The *crocodilia*, to which alligators belong, closely resemble the primitive archosaurs from which some reptiles, birds, and mammals evolved. It may be that *WT1* expression in the limbs is a feature which has been lost in subsequent evolution of vertebrates. The patterning of the limb bud is due to the ZPA (zone of polarising activity located in the posterior region of the limb bud). This can be mimicked by notochord which is involved in patterning of the mesoderm (Ingham 1994) and is thought to play a role in determination of the urogenital ridge (Etheridge 1969). Therefore *WT1* expression may be elicited in both the limb and the intermediate mesoderm because the same spectrum of transcriptional regulators are present in both.

7.10 Evolution of the vertebrate kidney and *WT1*

One of the major reasons for isolating the *WT1* orthologs was to compare *WT1* functionally and structurally in the context of the evolution of the vertebrate kidney. The two major evolutionary innovations have occurred in the vertebrate kidney. First came the initial evolution of the nephric system, probably pronephros-like, at about the same time that the vertebrates appeared. The second was the formation of the ureteric bud and hence the metanephros at about the same time as the appearance of the reptiles. The mechanisms which underlie these changes are unknown. The

expression of *WT1* in the mesonephros of amphibians demonstrates that the mesonephros of anamniotes and amniotes have similarities at the molecular level. The molecular basis for the degeneration of the mesonephros in amniotes but its retention in anamniotes is also unknown. *WT1* missense mutations are involved in the degeneration of the glomeruli in DDS, therefore it may be that *WT1* is somehow involved in the normal degeneration process. The further investigation of *Xenopus WT1* will allow these questions to be answered.

The origin of the vertebrate genitourinary system is not known. In non-vertebrate chordates (i.e. those species with a notochord but no backbone) no nephrons or kidneys have been found; in fact the whole region corresponding to the intermediate mesoderm is not present. In the non-vertebrate chordates other methods of excretion are used such as nephridia (amphioxus) and storage kidneys (ascidia). It is thought that coelomoducts, i.e. mesodermally derived ducts which can lead from the coelom, could be the precursors of the nephrons as these are very similar to the nephrotomes seen in association with external glomeruli. These ducts are observed in many phyla (Barnes *et al.,* 1993). The one coelomoduct found in amphioxus, Hatscheck's pit, develops in the region, at the border of the lateral plate and segmental (somitic) mesoderm, the location of the intermediate mesoderm in vertebrates (Grove and Newell 1953). Therefore how did the intermediate mesoderm evolve and how did the differentiation of genitourinary system come about? All these structures basically develop from the differentiation of tubules in this region which would suggest that this was an initial development.

The specification of the vertebrate mesoderm is being unravelled. In the anterior-posterior direction the domains of expression of the *Hox* genes is thought to be important. These may be the factors controlling the extent of nephric differentiation and degeneration along the urogenital ridge, as many *Hox* genes are expressed in the kidney and its precursors (Bard and McConnell 1994). The dorsal-ventral axis is best understood in *Xenopus:* there is an initial dorsal-ventral specification where the dorsal vegetal pole induces the notochord and muscle and the ventral vegetal pole induces blood, mesenchyme and mesothelium. There is then thought to be a dorsalising signal (Dale and Slack 1987) which induces the formation of the intermediate mesoderm types, including the pronephros (Lettice and Slack 1993). In the chick, from fate maps, the lateral plate and somatic mesoderm give rise to the intermediate mesoderm (Lear 1994). At the molecular level the genes involved are just beginning to be isolated, i.e. *BMP-4* (Dale *et al.,* 1992) and *Xwnt8* (Christian and Moon 1993) have both been shown to have ventralising activity, and *goosecoid*

has a dose dependent dorsalising activity (Niehrs *et al.,* 1994). In contrast very little is known about axial-paraxial specification. During evolution it would not be hard to imagine that shifting the expression boundaries of one of the genes involved in these specifications may have taken place and so allowed an additional domain to be formed.

With the increase in molecular and morphological evolutionary data, the question of how morphological and molecular evolution are related is able to be addressed. Throughout the vertebrates, the association of *WT1* with a typically vertebrate structure raises the question of whether *WT1* arose around the time of vertebrate evolution and was a key to the initiation of kidney evolution. *WT1* is expressed during the differentiation of the coelomic epithelium into mesothelium, the gonads, and the nephrons; specifically during the transition from mesenchyme to epithelia in these tissues. The chordates belong to the deuterostomes, which are distinct from the protostomes. In the deuterostomes coeloms are derived as invaginations of the gut, in the protostomes coeloms are formed by splitting of the mesoderm (Gilbert 1991). Therefore *WT1* may be associated with coelomic development in all deuterostomes. The amplification by PCR of a non-vertebrate chordate (ascidian) *WT1* gene was attempted but was not successful. The information on the vertebrate sequence may be useful in achieving this.

The mesenchymal to epithelial cell type change was originally thought to be unusual (Pritchard-Jones *et al.,* 1990) but it is found in many systems. *WT1* is known to be involved only in some of the transitions in mouse development (Armstrong *et al.,* 1992). One developmental event in which the transition is being characterised is the differentiation of the mid gut of *Drosophila* (Tepass and Hartenstein 1994). The endodermal midgut precursors are induced to become epithelial by their interaction with the adjacent visceral mesoderm. This is especially interesting considering the relationship of the deuterostome coelom to the gut and it recently been proposed that there may be conserved elements in gut development of invertebrates and vertebrates due to the similar expression of rat *HNF-4* and its *Drosophila* homolog (Zhong *et al.,* 1993). Therefore if a *Drosophila WT1* ortholog was sought this may be the best place to start looking. This of course is speculation but recent data to some extent has hinted that evolutionary distant systems may have a similar molecular basis. This in particular has been found in the resemblance in the pattern specification genes, i.e. the expression patterns of the *hedgehog* genes in vertebrate and to the *hedgehog* gene in Arthropoda (Ingham 1994). There are now reports of the conservation of gene expression in structures with similar functions but distinct

evolutionary origins e.g. the expression of *PAX6* in the vertebrate eye and brain and in the sensory cells of the anterior region of *C. elegans* (Chisholm and Horovitz 1994, Dawkins 1994). This could be a by-product of genes being activated in similar regions, due to similar patterning mechanisms in these diverged species. These genes may have been co-opted later into specialised structures in a similar region. Therefore this raises the question is *WT1* involved in the formation a the functional homolog of the kidneys; the gut derived Malpighian tubules of insects?

7.11 The origin of *WT1*

In the evolutionary comparison of species and genes the question that arises at the same time as "why a gene has evolved?" is "how has it evolved?". One of the major ways that molecular evolution is thought to occur is by gene duplication. Duplication leads to the release of one copy from selection pressure. In many cases this will lead to the loss of function of one copy, as evidenced by the presence of unprocessed pseudogenes, but in other cases before function is lost a separate role for the second gene may be acquired and subsequently selected for. The increase in genetic information may then favour the acquisition of new characteristics. The most well known example is the duplication of the homeobox gene clusters that seems to have occurred around the time of chordate and vertebrate evolution (Pendleton *et al.*, 1993). The origin of *WT1* is not known. Using current technology there are no detectable homologies to any other genes outside the zinc fingers. Knowledge of the tertiary structure of the transregulatory domain may allow identification of structural homologies when sequence homology is lost, as is now happening with the protein structures of enzymes (Burley 1994). Only a limited number of protein folds have been detected in the enzymes suggesting that it is far easier to modify an existing fold rather than to create a new one. The presence and position of introns between the zinc fingers of *WT1* leads to the suggestion of exon duplication having given rise to these. If this is the case then there may exist in more diverged species *WT1* orthologs that have 3, 2 or even a single finger. Again the sequence analysis of whole genomes will allow the presence or absence of *WT1* to be unequivocally determined.

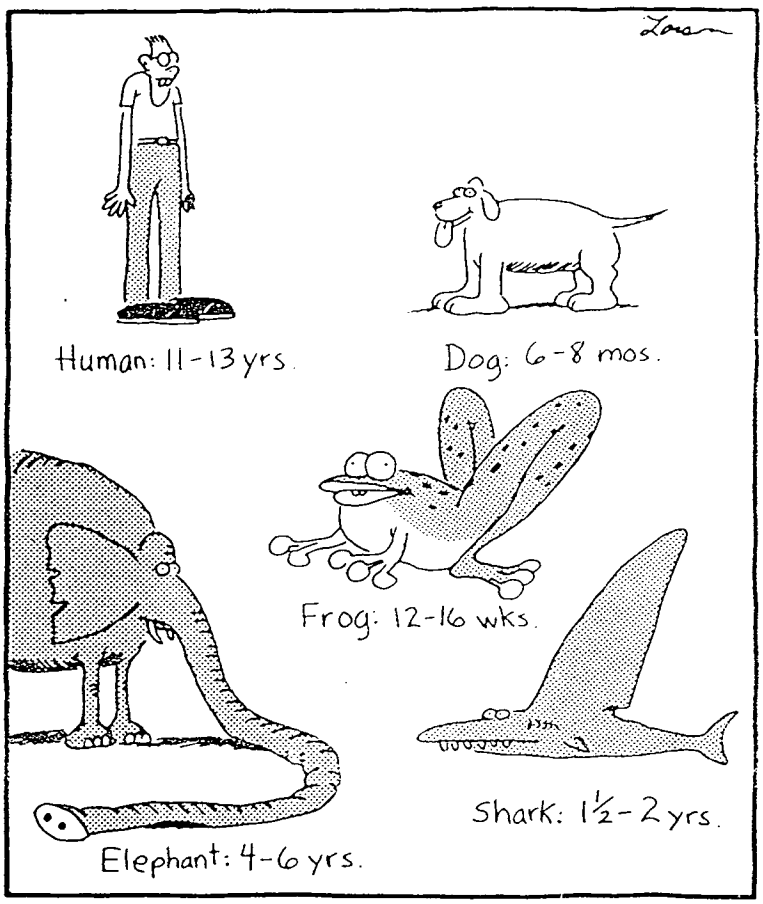
7.12 Future work

This project has only just begun the evolutionary analysis of *WT1*. It can now be extended at the structural, functional and expression levels. The immediate aims must be to consolidate this work by isolating complete sequences for the anamniotes,

giving greater depth to the comparison particularly in the transactivation domain and also to define exactly the cell types expressing *WT1* in whole-mount *in situ* hybridisation. The apparent conservation of nephrogenesis and nephroblastoma and the conservation of the *WT1* sequence and expression pattern suggest that study of *WT1* across the vertebrates will be an appropriate method for understanding *WT1* function in development and in Wilms' tumours. There are several projects that have been referred to above, in which the expression of *WT1* in various species is being investigated. In the future it will be particularly desirable to perform whole-mount *in situ* on the zebrafish, as these embryos are transparent and therefore particularly good for investigation of internal structures, especially as then it will be possible to fulfil the aim of looking at the pronephros and external glomeruli. The existence of a *Wt1* knockout mouse means that it is possible test whether these orthologs are functional in a different species by rescuing the lethal phenotype or perhaps more interestingly, there may be partial rescue. Also, would the introduction a *WT1* ortholog affect extent of differentiation and degeneration of the urogenital ridge. This would also test whether the unconserved regions are truly non-functional. One of the newest model organisms is the pufferfish (*Fugu*) which is known to have a genome one seventh of the size of the human genome, and genes are thought to cover correspondingly smaller genomic regions. Isolation of whole genomic regions of genes from pufferfish and introduction into other species would be a much simpler task than using orthologous genes from other vertebrates (Brenner *et al.*, 1993). This is now being pursued with the isolation of a *WT1* PCR clone for zinc fingers 1-3, which is three times smaller than the region in human and zebrafish (S. Macrae personal communication).

Finally the *in vitro* characterisation of *WT1* function may be facilitated by comparing the actions of the different orthologs. This may be useful in defining the important regions of the transregulatory domain, especially with reference to the homopolymer domains.

A Final Evolutionary Comparison



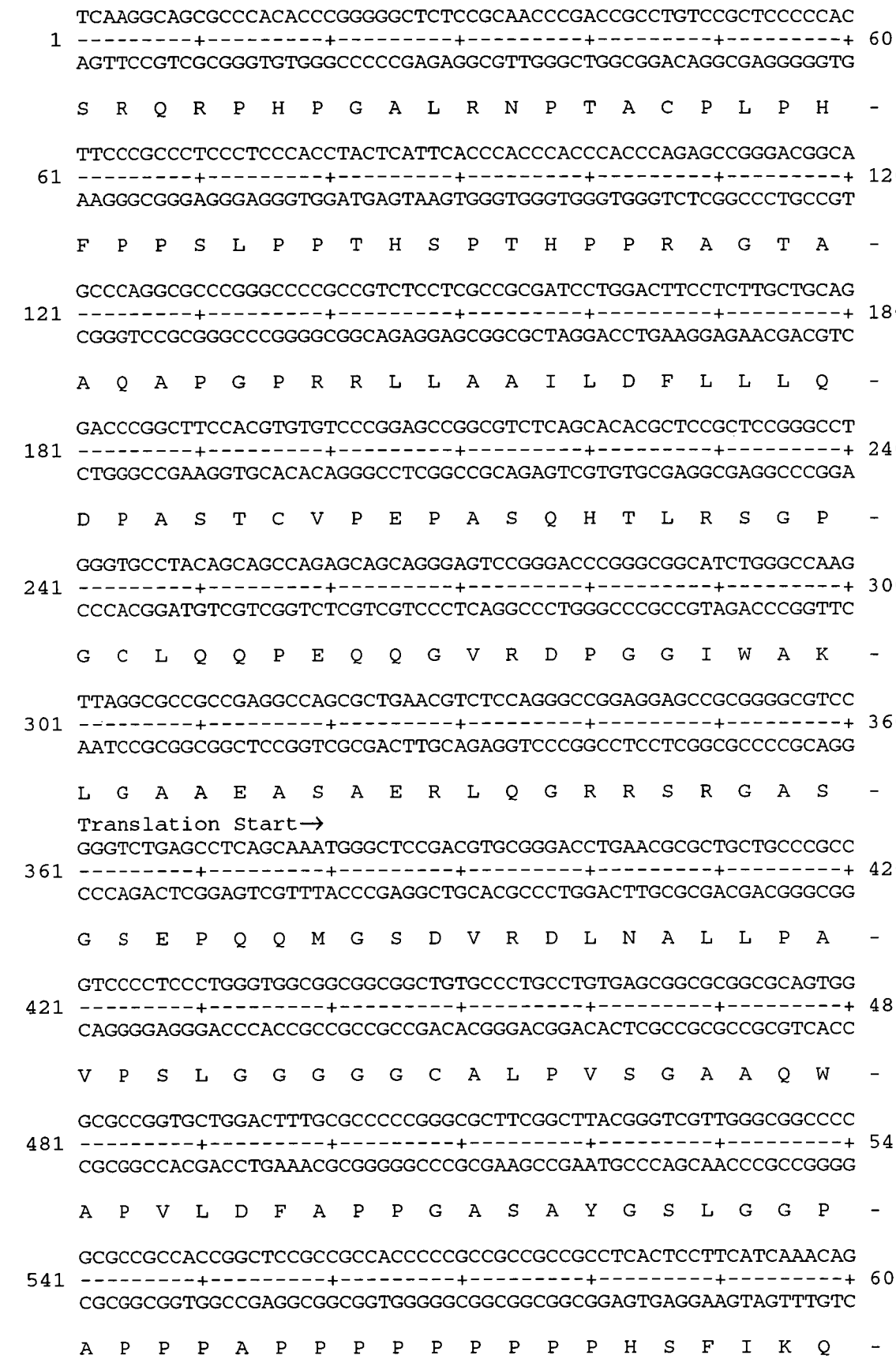
Awkward ages

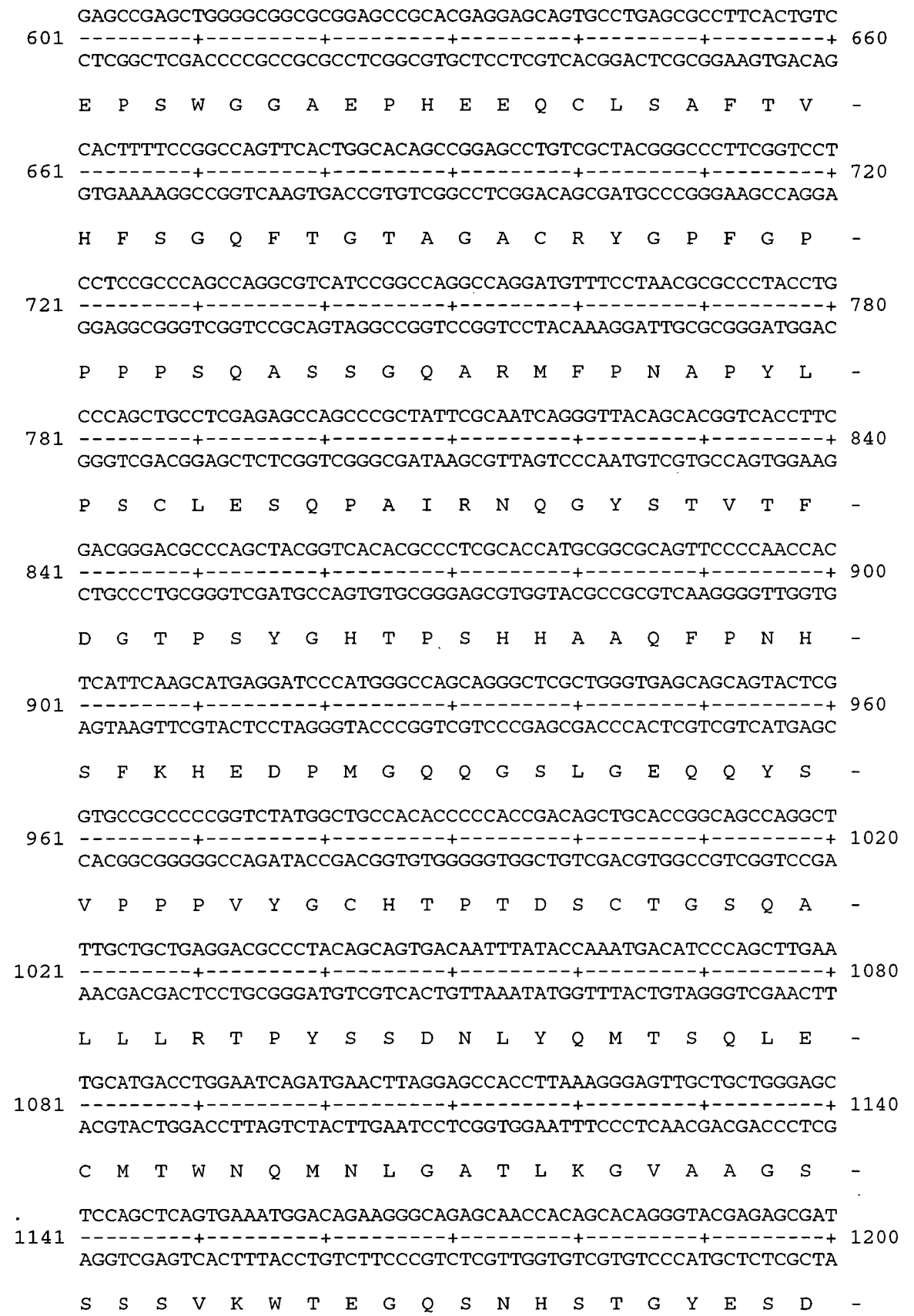
Appendix A
Sequences of *WT1* orthologs

The sequences of *WT1* orthologs obtained in this study are shown as well as the published human, mouse and rat sequences.
Where PCR clones were obtained in addition to library clones the region obtained by PCR is indicated.
For the genomic sequence obtained from zebrafish the predicted exons are shown in bold and the predicted cDNA sequence also shown.
In the 3' untanslated regions putative polyadenylation signals are highlighted

A line-up of the nucleotide sequence of all the *WT1* orthologs is also included (adapted from Clustal V output)

Human *WT1* cDNA (Gessler *et al.*, 1990)




GAGCCGAGCTGGGGCGGCGCGGAGCCGCACGAGGAGCAGTGCCTGAGCGCCTTCACTGTC
601 ---------+---------+---------+---------+---------+---------+ 660
CTCGGCTCGACCCCGCCGCGCCTCGGCGTGCTCCTCGTCACGGACTCGCGGAAGTGACAG
E P S W G G A E P H E E Q C L S A F T V -
CACTTTTCCGGCCAGTTCACTGGCACAGCCGGAGCCTGTCGCTACGGGCCCTTCGGTCCT
661 ---------+---------+---------+---------+---------+---------+ 720
GTGAAAAGGCCGGTCAAGTGACCGTGTCGGCCTCGGACAGCGATGCCCGGGAAGCCAGGA
H F S G Q F T G T A G A C R Y G P F G P -
CCTCCGCCCAGCCAGGCGTCATCCGGCCAGGCCAGGATGTTTCCTAACGCGCCCTACCTG
721 ---------+---------+---------+---------+---------+---------+ 780
GGAGGCGGGTCGGTCCGCAGTAGGCCGGTCCGGTCCTACAAAGGATTGCGCGGGATGGAC
P P P S Q A S S G Q A R M F P N A P Y L -
CCCAGCTGCCTCGAGAGCCAGCCCGCTATTCGCAATCAGGGTTACAGCACGGTCACCTTC
781 ---------+---------+---------+---------+---------+---------+ 840
GGGTCGACGGAGCTCTCGGTCGGGCGATAAGCGTTAGTCCCAATGTCGTGCCAGTGGAAG
P S C L E S Q P A I R N Q G Y S T V T F -
GACGGGACGCCCAGCTACGGTCACACGCCCTCGCACCATGCGGCGCAGTTCCCCAACCAC
841 ---------+---------+---------+---------+---------+---------+ 900
CTGCCCTGCGGGTCGATGCCAGTGTGCGGGAGCGTGGTACGCCGCGTCAAGGGGTTGGTG
D G T P S Y G H T P S H H A A Q F P N H -
TCATTCAAGCATGAGGATCCCATGGGCCAGCAGGGCTCGCTGGGTGAGCAGCAGTACTCG
901 ---------+---------+---------+---------+---------+---------+ 960
AGTAAGTTCGTACTCCTAGGGTACCCGGTCGTCCCGAGCGACCCACTCGTCGTCATGAGC
S F K H E D P M G Q Q G S L G E Q Q Y S -
GTGCCGCCCCCGGTCTATGGCTGCCACACCCCCACCGACAGCTGCACCGGCAGCCAGGCT
961 ---------+---------+---------+---------+---------+---------+ 1020
CACGGCGGGGGCCAGATACCGACGGTGTGGGGGTGGCTGTCGACGTGGCCGTCGGTCCGA
V P P P V Y G C H T P T D S C T G S Q A -
TTGCTGCTGAGGACGCCCTACAGCAGTGACAATTTATACCAAATGACATCCCAGCTTGAA
1021 ---------+---------+---------+---------+---------+---------+ 1080
AACGACGACTCCTGCGGGATGTCGTCACTGTTAAATATGGTTTACTGTAGGGTCGAACTT
L L L R T P Y S S D N L Y Q M T S Q L E -
TGCATGACCTGGAATCAGATGAACTTAGGAGCCACCTTAAAGGGAGTTGCTGCTGGGAGC
1081 ---------+---------+---------+---------+---------+---------+ 1140
ACGTACTGGACCTTAGTCTACTTGAATCCTCGGTGGAATTTCCCTCAACGACGACCCTCG
C M T W N Q M N L G A T L K G V A A G S -
. TCCAGCTCAGTGAAATGGACAGAAGGGCAGAGCAACCACAGCACAGGGTACGAGAGCGAT
1141 ---------+---------+---------+---------+---------+---------+ 1200
AGGTCGAGTCACTTTACCTGTCTTCCCGTCTCGTTGGTGTCGTGTCCCATGCTCTCGCTA
S S S V K W T E G Q S N H S T G Y E S D -

```
     AACCACACAACGCCCATCCTCTGCGGAGCCCAATACAGAATACACACGCACGGTGTCTTC
1201 ---------+---------+---------+---------+---------+---------+ 1260
     TTGGTGTGTTGCGGGTAGGAGACGCCTCGGGTTATGTCTTATGTGTGCGTGCCACAGAAG

     N  H  T  T  P  I  L  C  G  A  Q  Y  R  I  H  T  H  G  V  F  -

     AGAGGCATTCAGGATGTGCGACGTGTGCCTGGAGTAGCCCCGACTCTTGTACGGTCGGCA
1261 ---------+---------+---------+---------+---------+---------+ 1320
     TCTCCGTAAGTCCTACACGCTGCACACGGACCTCATCGGGGCTGAGAACATGCCAGCCGT

     R  G  I  Q  D  V  R  R  V  P  G  V  A  P  T  L  V  R  S  A  -

     TCTGAGACCAGTGAGAAACGCCCCTTCATGTGTGCTTACCCAGGCTGCAATAAGAGATAT
1321 ---------+---------+---------+---------+---------+---------+ 1380
     AGACTCTGGTCACTCTTTGCGGGGAAGTACACACGAATGGGTCCGACGTTATTCTCTATA

     S  E  T  S  E  K  R  P  F  M  C  A  Y  P  G  C  N  K  R  Y  -

     TTTAAGCTGTCCCACTTACAGATGCACAGCAGGAAGCACACTGGTGAGAAACCATACCAG
1381 ---------+---------+---------+---------+---------+---------+ 1440
     AAATTCGACAGGGTGAATGTCTACGTGTCGTCCTTCGTGTGACCACTCTTTGGTATGGTC

     F  K  L  S  H  L  Q  M  H  S  R  K  H  T  G  E  K  P  Y  Q  -

     TGTGACTTCAAGGACTGTGAACGAAGGTTTTCTCGTTCAGACCAGCTCAAAAGACACCAA
1441 ---------+---------+---------+---------+---------+---------+ 1500
     ACACTGAAGTTCCTGACACTTGCTTCCAAAAGAGCAAGTCTGGTCGAGTTTTCTGTGGTT

     C  D  F  K  D  C  E  R  R  F  S  R  S  D  Q  L  K  R  H  Q  -

     AGGAGACATACAGGTGTGAAACCATTCCAGTGTAAAACTTGTCAGCGAAAGTTCTCCCGG
1501 ---------+---------+---------+---------+---------+---------+ 1560
     TCCTCTGTATGTCCACACTTTGGTAAGGTCACATTTTGAACAGTCGCTTTCAAGAGGGCC

     R  R  H  T  G  V  K  P  F  Q  C  K  T  C  Q  R  K  F  S  R  -

     TCCGACCACCTGAAGACCCACACCAGGACTCATACAGGTAAAACAAGTGAAAAGCCCTTC
1561 ---------+---------+---------+---------+---------+---------+ 1620
     AGGCTGGTGGACTTCTGGGTGTGGTCCTGAGTATGTCCATTTTGTTCACTTTTCGGGAAG

     S  D  H  L  K  T  H  T  R  T  H  T  G  K  T  S  E  K  P  F  -

     AGCTGTCGGTGGCCAAGTTGTCAGAAAAAGTTTGCCCGGTCAGATGAATTAGTCCGCCAT
1621 ---------+---------+---------+---------+---------+---------+ 1680
     TCGACAGCCACCGGTTCAACAGTCTTTTTCAAACGGGCCAGTCTACTTAATCAGGCGGTA

     S  C  R  W  P  S  C  Q  K  K  F  A  R  S  D  E  L  V  R  H  -

     CACAACATGCATCAGAGAAACATGACCAAACTCCAGCTGGCGCTTTGAGGGGTCTCCCTC
1681 ---------+---------+---------+---------+---------+---------+ 1740
     GTGTTGTACGTAGTCTCTTTGTACTGGTTTGAGGTCGACCGCGAAACTCCCCAGAGGGAG

     H  N  M  H  Q  R  N  M  T  K  L  Q  L  A  L  *  G  V  S  L  -

     GGGGACCGTTCAGTGTCCCAGGCAGCACAGTGTGTGAACTGCTTTCAAGTCTGACTCTCC
1741 ---------+---------+---------+---------+---------+---------+ 1800
     CCCCTGGCAAGTCACAGGGTCCGTCGTGTCACACACTTGACGAAAGTTCAGACTGAGAGG

     G  D  R  S  V  S  Q  A  A  Q  C  V  N  C  F  Q  V  *  L  S  -

     ACTCCTCCTCACTAAAAAGGAAACTTCAGTTGATCTTCTTCATCCAACTTCCAAGACAAG
1801 ---------+---------+---------+---------+---------+---------+ 1860
```

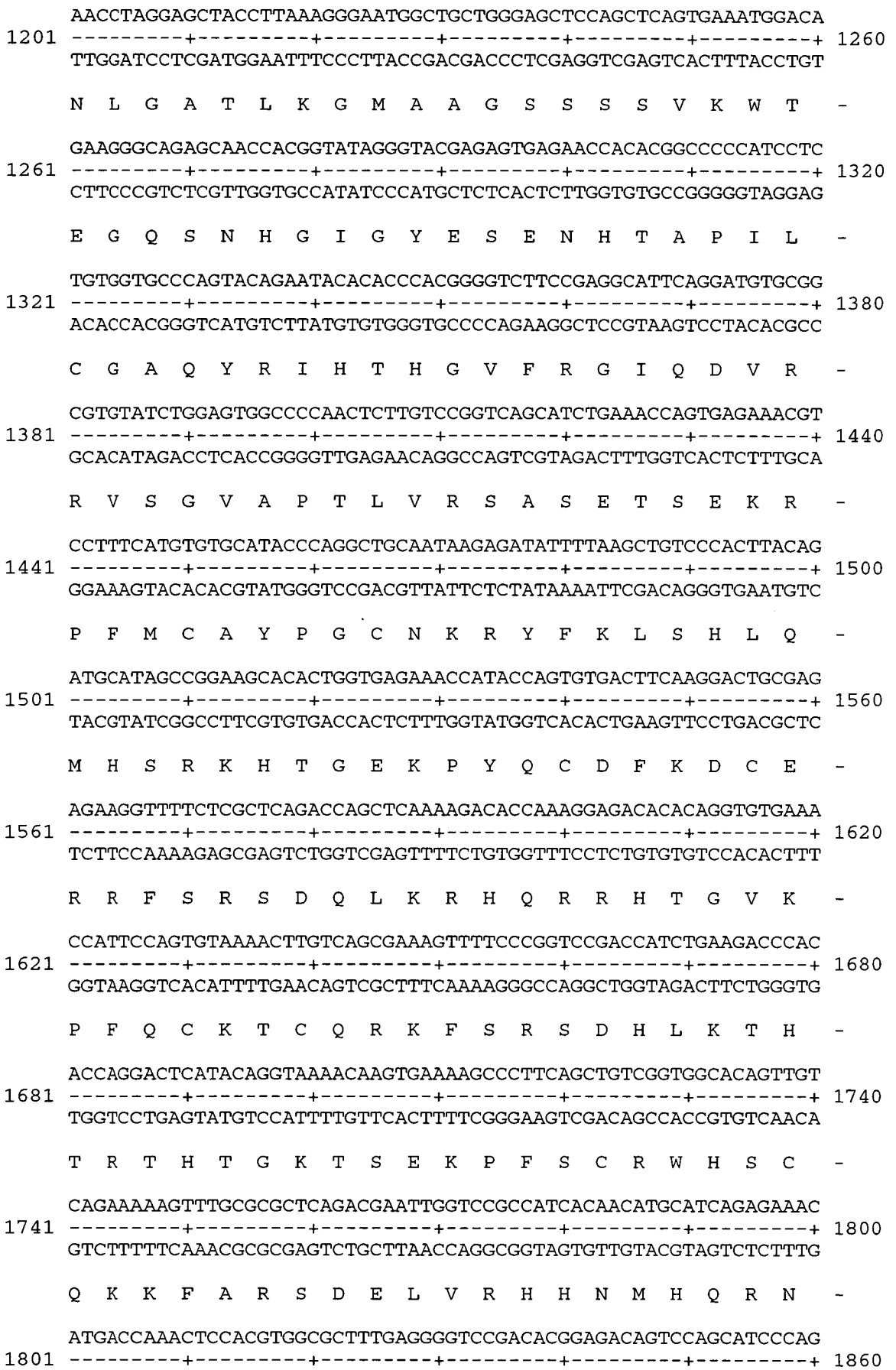
```
     TGAGGAGGAGTGATTTTTCCTTTGAAGTCAACTAGAAGAAGTAGGTTGAAGGTTCTGTTC

     T  P  P  H  *  K  G  N  F  S  *  S  S  S  S  N  F  Q  D  K  -

     ATACCGGTGCTTCTGGAAACTACCAGGTGTGCCTGGAAGAGTTGGTCTCTGCCCTGCCTA
1861 ---------+---------+---------+---------+---------+---------+ 1920
     TATGGCCACGAAGACCTTTGATGGTCCACACGGACCTTCTCAACCAGAGACGGGACGGAT

     I  P  V  L  L  E  T  T  R  C  A  W  K  S  W  S  L  P  C  L  -

     CTTTTAGTTGACTCACAGGCCCTGGAGAAGCAGCTAACAATGTCTGGTTAGTTAAAAGCC
1921 ---------+---------+---------+---------+---------+---------+ 1980
     GAAAATCAACTGAGTGTCCGGGACCTCTTCGTCGATTGTTACAGACCAATCAATTTTCGG

     L  L  V  D  S  Q  A  L  E  K  Q  L  T  M  S  G  *  L  K  A  -

     CATTGCCATTTGGTCTGGATTTTCTACTGTAAGAAGAGCCATAGCTGATCATGTCCCCCT
1981 ---------+---------+---------+---------+---------+---------+ 2040
     GTAACGGTAAACCAGACCTAAAAGATGACATTCTTCTCGGTATCGACTAGTACAGGGGGA

     H  C  H  L  V  W  I  F  Y  C  K  K  S  H  S  *  S  C  P  P  -

     GACCCTTCCCTTCTTTTTTTATGCTCGTTTTCGCTGGGGATGGAATTATTGTACCATTTT
2041 ---------+---------+---------+---------+---------+---------+ 2100
     CTGGGAAGGGAAGAAAAAAATACGAGCAAAAGCGACCCCTACCTTAATAACATGGTAAAA

     D  P  S  L  L  F  L  C  S  F  S  L  G  M  E  L  L  Y  H  F  -

     CTATCATGGAATATTTATAGGCCAGGGCATGTGTATGTGTCTGCTAATGTAAACTTTGTC
2101 ---------+---------+---------+---------+---------+---------+ 2160
     GATAGTACCTTATAAATATCCGGTCCCGTACACATACACAGACGATTACATTTGAAACAG

     L  S  W  N  I  Y  R  P  G  H  V  Y  V  S  A  N  V  N  F  V  -

     ATGGTTTCCATTTACTAACAGCAACAGCAAGAAATAAATCAGAGAGCAAGGCATCGGGGG
2161 ---------+---------+---------+---------+---------+---------+ 2220
     TACCAAAGGTAAATGATTGTCGTTGTCGTTCTTTATTTAGTCTCTCGTTCCGTAGCCCCC

     M  V  S  I  Y  *  Q  Q  Q  Q  E  I  N  Q  R  A  R  H  R  G  -

     TGAATCTTGTCTAACATTCCCGAGGTCAGCCAGGCTGCTAACCTGGAAAGCAGGATGTAG
2221 ---------+---------+---------+---------+---------+---------+ 2280
     ACTTAGAACAGATTGTAAGGGCTCCAGTCGGTCCGACGATTGGACCTTTCGTCCTACATC

     *  I  L  S  N  I  P  E  V  S  Q  A  A  N  L  E  S  R  M  *  -

     TTCTGCCAGGCAACTTTTAAAGCTCATGCATTTCAAGCAGCTGAAGAAAAGAATCAGAAC
2281 ---------+---------+---------+---------+---------+---------+ 2340
     AAGACGGTCCGTTGAAAATTTCGAGTACGTAAAGTTCGTCGACTTCTTTTCTTAGTCTTG

     F  C  Q  A  T  F  K  A  H  A  F  Q  A  A  E  E  K  N  Q  N  -

     TAACCAGTACCTCTGTATAGAAATCTAAAAGAATTTTACCATTCAGTTAATTCAATGTGA
2341 ---------+---------+---------+---------+---------+---------+ 2400
     ATTGGTCATGGAGACATATCTTTAGATTTTCTTAAAATGGTAAGTCAATTAAGTTACACT

     *  P  V  P  L  Y  R  N  L  K  E  F  Y  H  S  V  N  S  M  *  -

     ACACTGGCACACTGCTCTTAAGAAACTATGAAGATCTGAGATTTTTTTGTGTATGTTTTT
2401 ---------+---------+---------+---------+---------+---------+ 2460
     TGTGACCGTGTGACGAGAATTCTTTGATACTTCTAGACTCTAAAAAAACACATACAAAAA
```

```
     T  L  A  H  C  S  *  E  T  M  K  I  *  D  F  F  V  Y  V  F  -

     GACTCTTTTGAGTGGTAATCATATGTGTCTTTATAGATGTACATACCTCCTTGCACAAAT
2461 ---------+---------+---------+---------+---------+---------+ 2520
     CTGAGAAAACTCACCATTAGTATACACAGAAATATCTACATGTATGGAGGAACGTGTTTA

     D  S  F  E  W  *  S  Y  V  S  L  *  M  Y  I  P  P  C  T  N  -

     GGAGGGGAATTCATTTTCATCACTGGGACTGTCCTTAGTGTATAAAAACCATGCTGGTAT
2521 ---------+---------+---------+---------+---------+---------+ 2580
     CCTCCCCTTAAGTAAAAGTAGTGACCCTGACAGGAATCACATATTTTTGGTACGACCATA

     G  G  E  F  I  F  I  T  G  T  V  L  S  V  *  K  P  C  W  Y  -

     ATGGCTTCAAGTTGTAAAAATGAAAGTGACTTTAAAAGAAAATAGGGGATGGTCCAGGAT
2581 ---------+---------+---------+---------+---------+---------+ 2640
     TACCGAAGTTCAACATTTTTACTTTCACTGAAATTTTCTTTTATCCCCTACCAGGTCCTA

     M  A  S  S  C  K  N  E  S  D  F  K  R  K  *  G  M  V  Q  D  -

     CTCCACTGATAAGACTGTTTTTAAGTAACTTAAGGACCTTTGGGTCTACAAGTATATGTG
2641 ---------+---------+---------+---------+---------+---------+ 2700
     GAGGTGACTATTCTGACAAAAATTCATTGAATTCCTGGAAACCCAGATGTTCATATACAC

     L  H  *  *  D  C  F  *  V  T  *  G  P  L  G  L  Q  V  Y  V  -

     AAAAAAATGAGACTTACTGGGTGAGGAAATCCATTGTTTAAAGATGGTCGTGTGTGTGTG
2701 ---------+---------+---------+---------+---------+---------+ 2760
     TTTTTTTACTCTGAATGACCCACTCCTTTAGGTAACAAATTTCTACCAGCACACACACAC

     K  K  M  R  L  T  G  *  G  N  P  L  F  K  D  G  R  V  C  V  -

     TGTGTGTGTGTGTGTGTTGTGTTGTGTTTTGTTTTTTAAGGGAGGGAATTTATTATTTAC
2761 ---------+---------+---------+---------+---------+---------+ 2820
     ACACACACACACACACAACACAACACAAAACAAAAAATTCCCTCCCTTAAATAATAAATG

     C  V  C  V  C  V  V  L  C  F  V  F  *  G  R  E  F  I  I  Y  -

     CGTTGCTTGAAATTACTGTGTAAATATATGTCTGATAATGATTTGCTCTTTGACAACTAA
2821 ---------+---------+---------+---------+---------+---------+ 2880
     GCAACGAACTTTAATGACACATTTATATACAGACTATTACTAAACGAGAAACTGTTGATT

     R  C  L  K  L  L  C  K  Y  M  S  D  N  D  L  L  F  D  N  *  -

     AATTAGGACTGTATAAGTACTAGATGCATCACTGGGTGTTGATCTTACAAGATATTGATG
2881 ---------+---------+---------+---------+---------+---------+ 2940
     TTAATCCTGACATATTCATGATCTACGTAGTGACCCACAACTAGAATGTTCTATAACTAC

     N  *  D  C  I  S  T  R  C  I  T  G  C  *  S  Y  K  I  L  M  -

     ATAACACTTAAAATTGTAACCTGCATTTTTCACTTTGCTCTCAATTAAAGTCTATTCAAA
2941 ---------+---------+---------+---------+---------+---------+ 3000
     TATTGTGAATTTTAACATTGGACGTAAAAAGTGAAACGAGAGTTAATTTCAGATAAGTTT

     I  T  L  K  I  V  T  C  I  F  H  F  A  L  N  *  S  L  F  K  -

     AGGAAAAAAAAAAAAAAAA
3001 ---------+--------- 3019
     TCCTTTTTTTTTTTTTTTT

     R  K  K  K  K  K  ?  -
```

Mouse *WT1* cDNA (Buckler *et al.*, 1991)

```
    TGTGTGAATGGAGCGGCCGAGCATCCTGGCTCCTCCTCCTTCCCTGCTGCCGGCCCCTCT
  1 ---------+---------+---------+---------+---------+---------+ 60
    ACACACTTACCTCGCCGGCTCGTAGGACCGAGGAGGAGGAAGGGACGACGGCCGGGGAGA

    C  V  N  G  A  A  E  H  P  G  S  S  S  F  P  A  A  G  P  S  -

    TATTTGAGCTTTGGGAAGCTGGGGGCAGCCAGGCAGCTGGGGTAAGGAGTTCAAGGCAGC
 61 ---------+---------+---------+---------+---------+---------+ 120
    ATAAACTCGAAACCCTTCGACCCCCGTCGGTCCGTCGACCCCATTCCTCAAGTTCCGTCG

    Y  L  S  F  G  K  L  G  A  A  R  Q  L  G  *  G  V  Q  G  S  -

    GCCCACACCCGGGGCTCTCCGCAACCCGACCGCCTGCCTGCCTCCCCCTTTCCTTTTTTC
121 ---------+---------+---------+---------+---------+---------+ 180
    CGGGTGTGGGCCCCGAGAGGCGTTGGGCTGGCGGACGGACGGAGGGGGAAAGGAAAAAAG

    A  H  T  R  G  S  P  Q  P  D  R  L  P  A  S  P  F  P  F  F  -

    CCCGCCCCTCCCTCCCACCCACTCATTCACCCACCCACCCAGAGAGAGGACGGCAGCCCA
181 ---------+---------+---------+---------+---------+---------+ 240
    GGGCGGGGAGGGAGGGTGGGTGAGTAAGTGGGTGGGTGGGTCTCTCTCCTGCCGTCGGGT

    P  A  P  P  S  H  P  L  I  H  P  P  T  Q  R  E  D  G  S  P  -

    GGAACCCGGGCCCGCCGCCTCCTCGCCGCGATCCTGGACTTCCTCCTGTCGCAGGAGCCG
241 ---------+---------+---------+---------+---------+---------+ 300
    CCTTGGGCCCGGGCGGCGGAGGAGCGGCGCTAGGACCTGAAGGAGGACAGCGTCCTCGGC

    G  T  R  A  R  R  L  L  A  A  I  L  D  F  L  L  S  Q  E  P  -

    GCTTCCACGTGTGTCCCGGAGCCGGCGTCTCAGCACACGCTCCGCCGGGAGCCCGGGTGC
301 ---------+---------+---------+---------+---------+---------+ 360
    CGAAGGTGCACACAGGGCCTCGGCCGCAGAGTCGTGTGCGAGGCGGCCCTCGGGCCCACG

    A  S  T  C  V  P  E  P  A  S  Q  H  T  L  R  R  E  P  G  C  -

    GTCCAGCAGCCGGAGCAACCTGGGGACCGAGGCCCCCGGAGCGCCTGGGCCAAGTCCAGC
361 ---------+---------+---------+---------+---------+---------+ 420
    CAGGTCGTCGGCCTCGTTGGACCCCTGGCTCCGGGGGCCTCGCGGACCCGGTTCAGGTCG

    V  Q  Q  P  E  Q  P  G  D  R  G  P  R  S  A  W  A  K  S  S  -

                                                Translation Start→
    GCCGAGAATCCGCAGGATCGCAGGAGCGGAGAACCGTCCGCATCCGAGCCGCACCTCATG
421 ---------+---------+---------+---------+---------+---------+ 480
    CGGCTCTTAGGCGTCCTAGCGTCCTCGCCTCTTGGCAGGCGTAGGCTCGGCGTGGAGTAC

    A  E  N  P  Q  D  R  R  S  G  E  P  S  A  S  E  P  H  L  M  -

    GGTTCCGACGTGCGGGACCTGAACGCGCTGCTGCCCGCTGTGTCTTCGCTGGGCGGCGGC
481 ---------+---------+---------+---------+---------+---------+ 540
    CCAAGGCTGCACGCCCTGGACTTGCGCGACGACGGGCGACACAGAAGCGACCCGCCGCCG

    G  S  D  V  R  D  L  N  A  L  L  P  A  V  S  S  L  G  G  G  -

    GGCGGCGGCTGCGGGCTCCCTGTGAGCGGCGCACGGCAGTGGGCGCCCGTGTTGGACTTC
541 ---------+---------+---------+---------+---------+---------+ 600
    CCGCCGCCGACGCCCGAGGGACACTCGCCGCGTGCCGTCACCCGCGGGCACAACCTGAAG

    G  G  G  C  G  L  P  V  S  G  A  R  Q  W  A  P  V  L  D  F  -
```

```
    GCGCCTCCGGGCGCCTCGGCTTACGGGTCGCTGGGCGGTCCCGCGCCTCCTCCCGCTCCG
601 ---------+---------+---------+---------+---------+---------+ 660
    CGCGGAGGCCCGCGGAGCCGAATGCCCAGCGACCCGCCAGGGCGCGGAGGAGGGCGAGGC

    A  P  P  G  A  S  A  Y  G  S  L  G  G  P  A  P  P  P  A  P  -

    CCGCCGCCTCCGCCGCCACCCCACTCCTTCATCAAACAGGAGCCCAGCTGGGGCGGCGCC
661 ---------+---------+---------+---------+---------+---------+ 720
    GGCGGCGGAGGCGGCGGTGGGGTGAGGAAGTAGTTTGTCCTCGGGTCGACCCCGCCGCGG

    P  P  P  P  P  P  P  H  S  F  I  K  Q  E  P  S  W  G  G  A  -

    GAGCCACACGAGGAGCAGTGCCTGAGCGCCTTCACCTTGCACTTCTCGGGCCAGTTCACC
721 ---------+---------+---------+---------+---------+---------+ 780
    CTCGGTGTGCTCCTCGTCACGGACTCGCGGAAGTGGAACGTGAAGAGCCCGGTCAAGTGG

    E  P  H  E  E  Q  C  L  S  A  F  T  L  H  F  S  G  Q  F  T  -

    GGTACAGCCGGGGCCTGTCGCTACGGACCCTTCGGTCCTCCCCCGCCCAGCCAGGCGTCC
781 ---------+---------+---------+---------+---------+---------+ 840
    CCATGTCGGCCCCGGACAGCGATGCCTGGGAAGCCAGGAGGGGGCGGGTCGGTCCGCAGG

    G  T  A  G  A  C  R  Y  G  P  F  G  P  P  P  P  S  Q  A  S  -

    TCGGGCCAGGCCAGGATGTTCCCCAATGCGCCCTACCTGCCCAGCTGCCTGGAGAGCCAG
841 ---------+---------+---------+---------+---------+---------+ 900
    AGCCCGGTCCGGTCCTACAAGGGGTTACGCGGGATGGACGGGTCGACGGACCTCTCGGTC

    S  G  Q  A  R  M  F  P  N  A  P  Y  L  P  S  C  L  E  S  Q  -

    CCTACCATCCGCAACCAAGGATACAGCACGGTCACTTTCGACGGGGCGCCCAGCTATGGC
901 ---------+---------+---------+---------+---------+---------+ 960
    GGATGGTAGGCGTTGGTTCCTATGTCGTGCCAGTGAAAGCTGCCCCGCGGGTCGATACCG

    P  T  I  R  N  Q  G  Y  S  T  V  T  F  D  G  A  P  S  Y  G  -

    CACACGCCCTCGCATCACGCGGCGCAGTTCCCCAACCATTCCTTCAAACACGAGGACCCC
961 ---------+---------+---------+---------+---------+---------+ 1020
    GTGTGCGGGAGCGTAGTGCGCCGCGTCAAGGGGTTGGTAAGGAAGTTTGTGCTCCTGGGG

    H  T  P  S  H  H  A  A  Q  F  P  N  H  S  F  K  H  E  D  P  -

    ATGGGCCAGCAGGGCTCGCTGGGCGAGCAGCAGTACTCCGTGCCACCTCCGGTGTATGGC
1021 ---------+---------+---------+---------+---------+---------+ 1080
    TACCCGGTCGTCCCGAGCGACCCGCTCGTCGTCATGAGGCACGGTGGAGGCCACATACCG

    M  G  Q  Q  G  S  L  G  E  Q  Q  Y  S  V  P  P  P  V  Y  G  -

    TGCCACACCCCTACTGACAGTTGCACAGGCAGCCAGGCCCTGCTCCTGAGGACGCCCTAC
1081 ---------+---------+---------+---------+---------+---------+ 1140
    ACGGTGTGGGGATGACTGTCAACGTGTCCGTCGGTCCGGGACGAGGACTCCTGCGGGATG

    C  H  T  P  T  D  S  C  T  G  S  Q  A  L  L  L  R  T  P  Y  -

    AGCAGTGACAATTTATACCAAATGACCTCCCAGCTTGAATGCATGACCTGGAATCAGATG
1141 ---------+---------+---------+---------+---------+---------+ 1200
    TCGTCACTGTTAAATATGGTTTACTGGAGGGTCGAACTTACGTACTGGACCTTAGTCTAC

    S  S  D  N  L  Y  Q  M  T  S  Q  L  E  C  M  T  W  N  Q  M  -
```

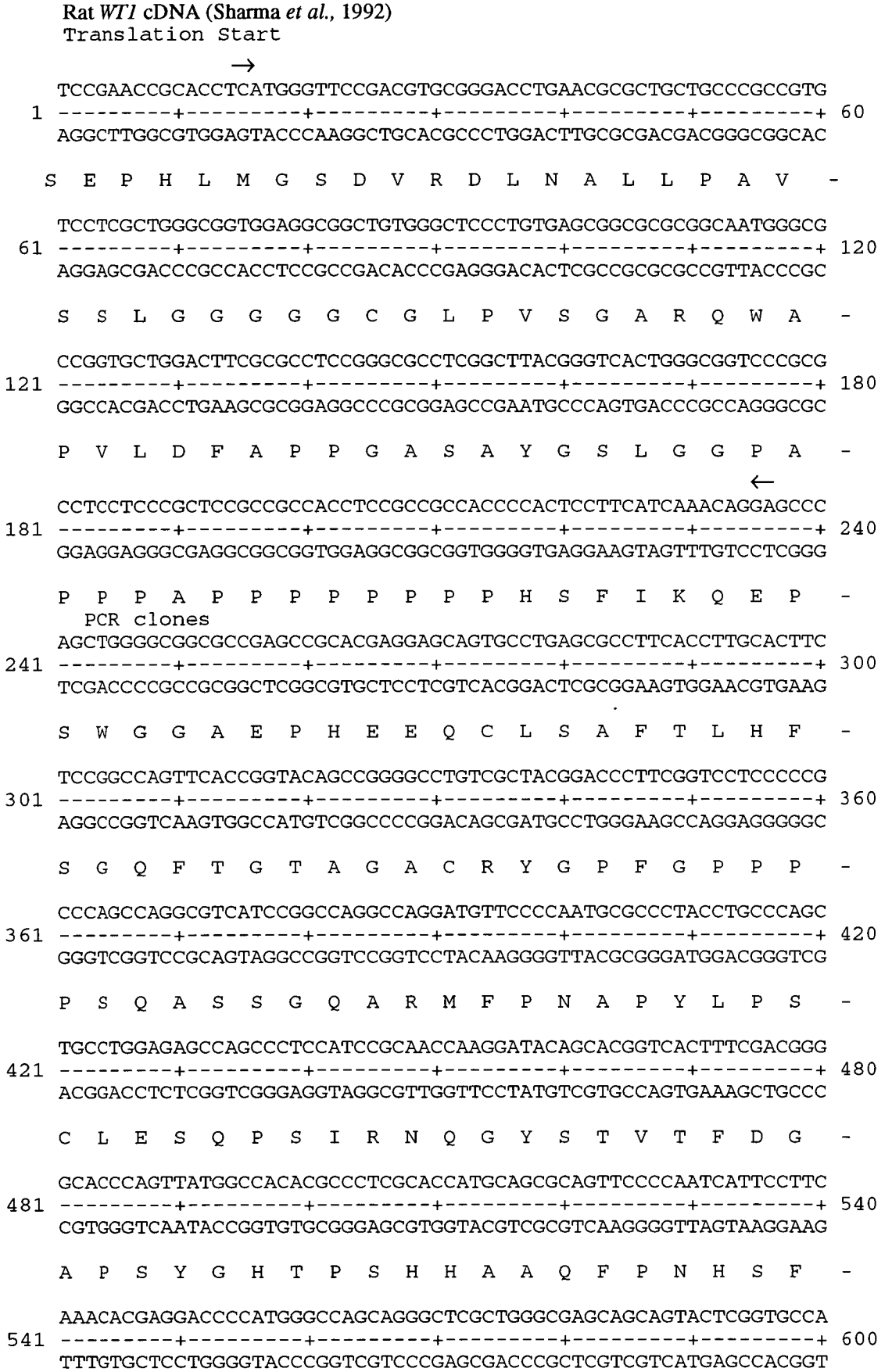

AACCTAGGAGCTACCTTAAAGGGAATGGCTGCTGGGAGCTCCAGCTCAGTGAAATGGACA
1201 --------+---------+---------+---------+---------+---------+---------+ 1260
TTGGATCCTCGATGGAATTTCCCTTACCGACGACCCTCGAGGTCGAGTCACTTTACCTGT
N L G A T L K G M A A G S S S S V K W T -
GAAGGGCAGAGCAACCACGGTATAGGGTACGAGAGTGAGAACCACACGGCCCCCATCCTC
1261 --------+---------+---------+---------+---------+---------+---------+ 1320
CTTCCCGTCTCGTTGGTGCCATATCCCATGCTCTCACTCTTGGTGTGCCGGGGGTAGGAG
E G Q S N H G I G Y E S E N H T A P I L -
TGTGGTGCCCAGTACAGAATACACACCCACGGGGTCTTCCGAGGCATTCAGGATGTGCGG
1321 --------+---------+---------+---------+---------+---------+---------+ 1380
ACACCACGGGTCATGTCTTATGTGTGGGTGCCCCAGAAGGCTCCGTAAGTCCTACACGCC
C G A Q Y R I H T H G V F R G I Q D V R -
CGTGTATCTGGAGTGGCCCCAACTCTTGTCCGGTCAGCATCTGAAACCAGTGAGAAACGT
1381 --------+---------+---------+---------+---------+---------+---------+ 1440
GCACATAGACCTCACCGGGGTTGAGAACAGGCCAGTCGTAGACTTTGGTCACTCTTTGCA
R V S G V A P T L V R S A S E T S E K R -
CCTTTCATGTGTGCATACCCAGGCTGCAATAAGAGATATTTTAAGCTGTCCCACTTACAG
1441 --------+---------+---------+---------+---------+---------+---------+ 1500
GGAAAGTACACACGTATGGGTCCGACGTTATTCTCTATAAAATTCGACAGGGTGAATGTC
P F M C A Y P G C N K R Y F K L S H L Q -
ATGCATAGCCGGAAGCACACTGGTGAGAAACCATACCAGTGTGACTTCAAGGACTGCGAG
1501 --------+---------+---------+---------+---------+---------+---------+ 1560
TACGTATCGGCCTTCGTGTGACCACTCTTTGGTATGGTCACACTGAAGTTCCTGACGCTC
M H S R K H T G E K P Y Q C D F K D C E -
AGAAGGTTTTCTCGCTCAGACCAGCTCAAAAGACACCAAAGGAGACACACAGGTGTGAAA
1561 --------+---------+---------+---------+---------+---------+---------+ 1620
TCTTCCAAAAGAGCGAGTCTGGTCGAGTTTTCTGTGGTTTCCTCTGTGTGTCCACACTTT
R R F S R S D Q L K R H Q R R H T G V K -
CCATTCCAGTGTAAAACTTGTCAGCGAAAGTTTTCCCGGTCCGACCATCTGAAGACCCAC
1621 --------+---------+---------+---------+---------+---------+---------+ 1680
GGTAAGGTCACATTTTGAACAGTCGCTTTCAAAAGGGCCAGGCTGGTAGACTTCTGGGTG
P F Q C K T C Q R K F S R S D H L K T H -
ACCAGGACTCATACAGGTAAAACAAGTGAAAAGCCCTTCAGCTGTCGGTGGCACAGTTGT
1681 --------+---------+---------+---------+---------+---------+---------+ 1740
TGGTCCTGAGTATGTCCATTTTGTTCACTTTTCGGGAAGTCGACAGCCACCGTGTCAACA
T R T H T G K T S E K P F S C R W H S C -
CAGAAAAAGTTTGCGCGCTCAGACGAATTGGTCCGCCATCACAACATGCATCAGAGAAAC
1741 --------+---------+---------+---------+---------+---------+---------+ 1800
GTCTTTTTCAAACGCGCGAGTCTGCTTAACCAGGCGGTAGTGTTGTACGTAGTCTCTTTG
Q K K F A R S D E L V R H H N M H Q R N -
ATGACCAAACTCCACGTGGCGCTTTGAGGGGTCCGACACGGAGACAGTCCAGCATCCCAG
1801 --------+---------+---------+---------+---------+---------+---------+ 1860

```
      TACTGGTTTGAGGTGCACCGCGAAACTCCCCAGGCTGTGCCTCTGTCAGGTCGTAGGGTC

      M  T  K  L  H  V  A  L  *  G  V  R  H  G  D  S  P  A  S  Q   -

      GCAGGAAAGTGTGCAAACTGCTTCCAAATCTGATTTTGAAATTCCTCCCACTCACCTTTC
1861  ---------+---------+---------+---------+---------+---------+  1920
      CGTCCTTTCACACGTTTGACGAAGGTTTAGACTAAAACTTTAAGGAGGGTGAGTGGAAAG

      A  G  K  C  A  N  C  F  Q  I  *  F  *  N  S  S  H  S  P  F   -

      AAAGGACACGACTGTGGATCTACATCCGACTTCCAAGACAGCACACCTGATTGACTGCAT
1921  ---------+---------+---------+---------+---------+---------+  1980
      TTTCCTGTGCTGACACCTAGATGTAGGCTGAAGGTTCTGTCGTGTGGACTAACTGACGTA

      K  G  H  D  C  G  S  T  S  D  F  Q  D  S  T  P  D  *  L  H   -

      CCTATCAGGTTTGCCGGAAGGAGTCGGTCCTCCGCCCACTTTTGATTAACTCACAGGCCT
1981  ---------+---------+---------+---------+---------+---------+  2040
      GGATAGTCCAAACGGCCTTCCTCAGCCAGGAGGCGGGTGAAAACTAATTGAGTGTCCGGA

      P  I  R  F  A  G  R  S  R  S  S  A  H  F  *  L  T  H  R  P   -

      GAAAAAAGTGGTTCAAGGTGTCTAGAAAGTCCAATTGTCTGAATTTTCTACTGTTAGAAG
2041  ---------+---------+---------+---------+---------+---------+  2100
      CTTTTTTCACCAAGTTCCACAGATCTTTCAGGTTAACAGACTTAAAAGATGACAATCTTC

      E  K  S  G  S  R  C  L  E  S  P  I  V  *  I  F  Y  C  *  K   -

      AACCATTGTTGATAATGCCCCCCGCCCCCCCCCCCCCCCGGGTTTCCTCTTCTCCTTTGTG
2101  ---------+---------+---------+---------+---------+---------+  2160
      TTGGTAACAACTATTACGGGGGGCGGGGGGGGGGGGGGCCCAAAGGAGAAGAGGAAACAC

      N  H  C  *  *  C  P  P  P  P  P  P  P  G  F  L  F  S  F  V   -

      ATCATTTCCCCAGGATTAGAGAGACTGTTACATTTTCTTTCATGGGATATTTATAGGCCA
2161  ---------+---------+---------+---------+---------+---------+  2220
      TAGTAAAGGGGTCCTAATCTCTCTGACAATGTAAAAGAAAGTACCCTATAAATATCCGGT

      I  I  S  P  G  L  E  R  L  L  H  F  L  S  W  D  I  Y  R  P   -

      GGGCATGTGTATGTGCCTGCTAATGTAAACTCTGTCATAGTTCCCATTTACTAACTGCCC
2221  ---------+---------+---------+---------+---------+---------+  2280
      CCCGTACACATACACGGACGATTACATTTGAGACAGTATCAAGGGTAAATGATTGACGGG

      G  H  V  Y  V  P  A  N  V  N  S  V  I  V  P  I  Y  *  L  P   -

      TAGAAAGAAATAAATCAGAGAGCAAGGCACCAGGCAAGAATCGTACAGAATTTCAGAGGT
2281  ---------+---------+---------+---------+---------+---------+  2340
      ATCTTTCTTTATTTAGTCTCTCGTTCCGTGGTCCGTTCTTAGCATGTCTTAAAGTCTCCA

      *  K  E  I  N  Q  R  A  R  H  Q  A  R  I  V  Q  N  F  R  G   -

      CTGGCTGCAAACCTGGAAACCTGGAAGGCCAGATGTAATTCTACAGGCGATTGTTAAAGC
2341  ---------+---------+---------+---------+---------+---------+  2400
      GACCGACGTTTGGACCTTTGGACCTTCCGGTCTACATTAAGATGTCCGCTAACAATTTCG

      L  A  A  N  L  E  T  W  K  A  R  C  N  S  T  G  D  C  *  S   -

      TCATAGGTTTTGAGTAACTGCATAGTAGGTTGGTATTAACTAGAACTCTGTATAGTTAGG
2401  ---------+---------+---------+---------+---------+---------+  2460
      AGTATCCAAAACTCATTGACGTATCATCCAACCATAATTGATCTTGAGACATATCAATCC
```

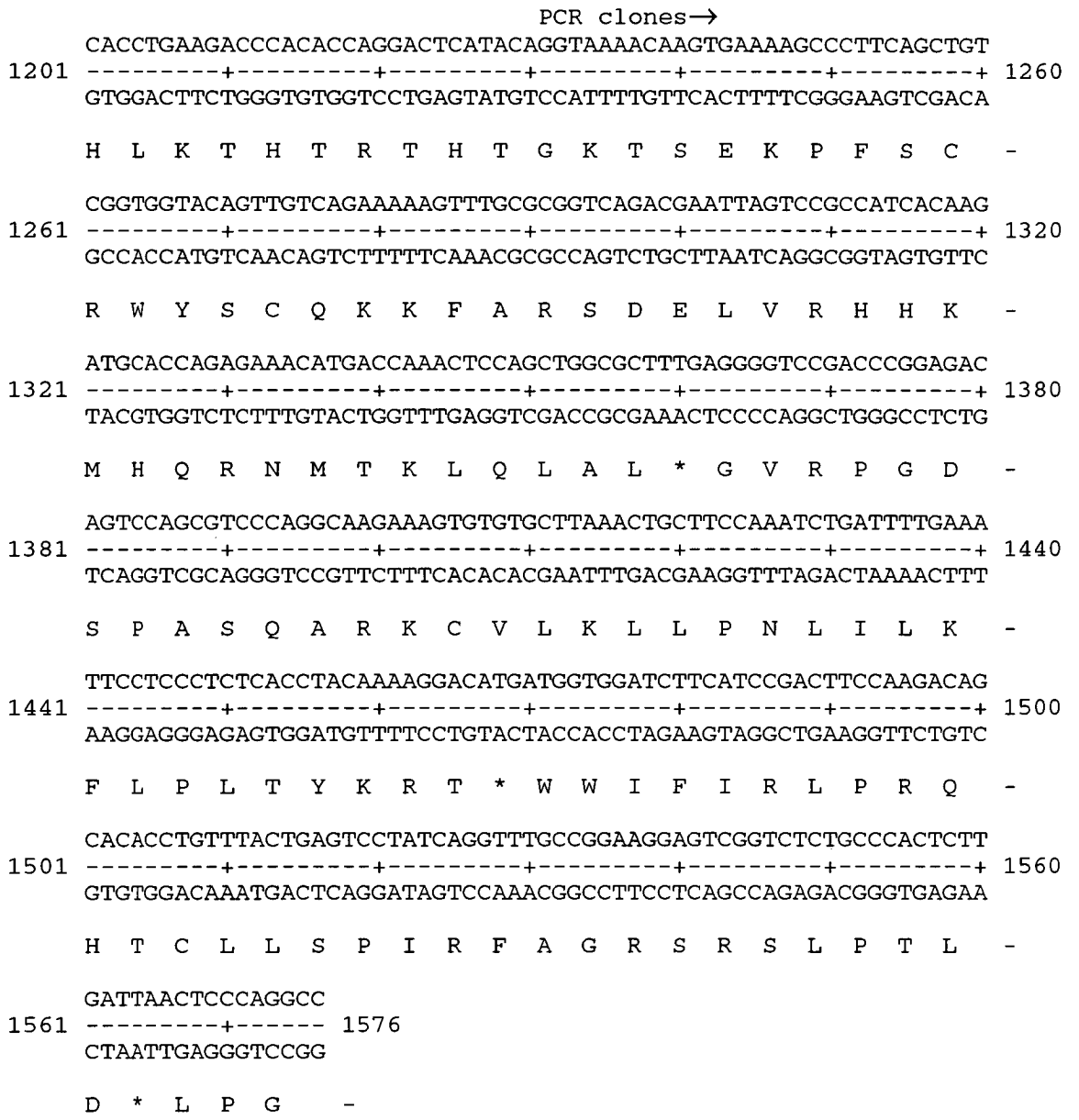
```
     S  *  V  L  S  N  C  I  V  G  W  Y  *  L  E  L  C  I  V  R  -

     ACGGAGAGGAGCCTTCCTGCTCAGCTATTCACTCTGAACACTAGCACTGGGCTCTTAAGA
2461 ---------+---------+---------+---------+---------+---------+ 2520
     TGCCTCTCCTCGGAAGGACGAGTCGATAAGTGAGACTTGTGATCGTGACCCGAGAATTCT

     T  E  R  S  L  P  A  Q  L  F  T  L  N  T  S  T  G  L  L  R  -

     AATGATGTTTTAAGAGCAGAGATCTTTTTTTAATGTCTTTGATTTATTTTTTAGTTGTAA
2521 ---------+---------+---------+---------+---------+---------+ 2580
     TTACTACAAAATTCTCGTCTCTAGAAAAAAATTACAGAAACTAAATAAAAAATCAACATT

     N  D  V  L  R  A  E  I  F  F  *  C  L  *  F  I  F  *  L  *  -

     TTAGGTACATCCTCAGAGATGTACTTTCCTCCTCTTGTGCAGGATGTGGAGGACTCGTTC
2581 ---------+---------+---------+---------+---------+---------+ 2640
     AATCCATGTAGGAGTCTCTACATGAAAGGAGGAGAACACGTCCTACACCTCCTGAGCAAG

     L  G  T  S  S  E  M  Y  F  P  P  L  V  Q  D  V  E  D  S  F  -

     CATCATCTGGGGCATCTTTAGAGTGTATAGACCACACTGGTTATGTGGCTTCAAGTTGTA
2641 ---------+---------+---------+---------+---------+---------+ 2700
     GTAGTAGACCCCGTAGAAATCTCACATATCTGGTGTGACCAATACACCGAAGTTCAACAT

     H  H  L  G  H  L  *  S  V  *  T  T  L  V  M  W  L  Q  V  V  -

     AAAATTAAAATGACTTTAAAAGAAACTAGGGGCTGGTCCAGGATCTCACTGGTAAGACTG
2701 ---------+---------+---------+---------+---------+---------+ 2760
     TTTTAATTTTACTGAAATTTTCTTTGATCCCCGACCAGGTCCTAGAGTGACCATTCTGAC

     K  I  K  M  T  L  K  E  T  R  G  W  S  R  I  S  L  V  R  L  -

     TTCTTAAGTAACTTAAGTATCTTTGAATCTGCAAGTATGTAGGGAAAAAAAAAAAGATAT
2761 ---------+---------+---------+---------+---------+---------+ 2820
     AAGAATTCATTGAATTCATAGAAACTTAGACGTTCATACATCCCTTTTTTTTTTTCTATA

     F  L  S  N  L  S  I  F  E  S  A  S  M  *  G  K  K  K  R  Y  -

     ATTATTGTGAGGAAATCCATTGTTTAAAGGTGTGCGTGTGTTGTTGTTGTTTTTTAAAGG
2821 ---------+---------+---------+---------+---------+---------+ 2880
     TAATAACACTCCTTTAGGTAACAAATTTCCACACGCACACAACAACAACAAAAAATTTCC

     I  I  V  R  K  S  I  V  *  R  C  A  C  V  V  V  V  F  *  R  -

     GAGGGAGTTTATTATTTACTGTAGCTTGAAATACTGTGTAAATATATATGTATATATATG
2881 ---------+---------+---------+---------+---------+---------+ 2940
     CTCCCTCAAATAATAAATGACATCGAACTTTATGACACATTTATATATACATATATATAC

     E  G  V  Y  Y  L  L  *  L  E  I  L  C  K  Y  I  C  I  Y  M  -

     ATGTGCTCTTTGTCAACTAAAATTAGGAGGTGTATGGTATTAGCTGCATCACTGTGTGGA
2941 ---------+---------+---------+---------+---------+---------+ 3000
     TACACGAGAAACAGTTGATTTTAATCCTCCACATACCATAATCGACGTAGTGACACACCT

     M  C  S  L  S  T  K  I  R  R  C  M  V  L  A  A  S  L  C  G  -

     TGTCAATCTTACAGTGTATTGATGATAATACTAAAAATGTAACCTGCATCTTTTTCCACT
3001 ---------+---------+---------+---------+---------+---------+ 3060
     ACAGTTAGAATGTCACATAACTACTATTATGATTTTTACATTGGACGTAGAAAAAGGTGA

     C  Q  S  Y  S  V  L  M  I  I  L  K  M  *  P  A  S  F  S  T  -
```

```
     TGGCTGTCAATTAAAGTCTATTCAAAAGG
3061 ---------+---------+--------- 3089
     ACCGACAGTTAATTTCAGATAAGTTTTCC

     W  L  S  I  K  V  Y  S  K  ?  -
```

Rat *WT1* cDNA (Sharma *et al.*, 1992)



```
    K  H  E  D  P  M  G  Q  Q  G  S  L  G  E  Q  Q  Y  S  V  P  -

    CCTCCAGTGTATGGCTGCCACACCCCTACCGACAGTTGCACAGGCAGCCAGGCCCTGCTC
601 ---------+---------+---------+---------+---------+---------+ 660
    GGAGGTCACATACCGACGGTGTGGGGATGGCTGTCAACGTGTCCGTCGGTCCGGGACGAG

    P  P  V  Y  G  C  H  T  P  T  D  S  C  T  G  S  Q  A  L  L  -

    CTGAGGACGCCCTACAGCAGTGACAATTTATACCAGATGACCTCCCAGCTTGAGTGCATG
661 ---------+---------+---------+---------+---------+---------+ 720
    GACTCCTGCGGGATGTCGTCACTGTTAAATATGGTCTACTGGAGGGTCGAACTCACGTAC

    L  R  T  P  Y  S  S  D  N  L  Y  Q  M  T  S  Q  L  E  C  M  -

    ACCTGGAACCAGATGAACCTCGGAGCTACCTTGAAGGGAATGGCTGCTGGGAGCTCCAGC
721 ---------+---------+---------+---------+---------+---------+ 780
    TGGACCTTGGTCTACTTGGAGCCTCGATGGAACTTCCCTTACCGACGACCCTCGAGGTCG

    T  W  N  Q  M  N  L  G  A  T  L  K  G  M  A  A  G  S  S  S  -

    TCAGTGAAATGGACAGAAGGGCAGAGCAACCACGGCACAGGGTACGAGAGTGAGAACCAC
781 ---------+---------+---------+---------+---------+---------+ 840
    AGTCACTTTACCTGTCTTCCCGTCTCGTTGGTGCCGTGTCCCATGCTCTCACTCTTGGTG

    S  V  K  W  T  E  G  Q  S  N  H  G  T  G  Y  E  S  E  N  H  -

    ACAACCCCCATCCTCTGTGGTGCCCAGTACAGAATACACACCCACGGTGTCTTCCGAGGC
841 ---------+---------+---------+---------+---------+---------+ 900
    TGTTGGGGGTAGGAGACACCACGGGTCATGTCTTATGTGTGGGTGCCACAGAAGGCTCCG

    T  T  P  I  L  C  G  A  Q  Y  R  I  H  T  H  G  V  F  R  G  -

    ATTCAGGATGTGCGGCGCGTATCTGGAGTGGCCCCAACTCTTGTCCGGTCAGCATCTGAA
901 ---------+---------+---------+---------+---------+---------+ 960
    TAAGTCCTACACGCCGCGCATAGACCTCACCGGGGTTGAGAACAGGCCAGTCGTAGACTT

    I  Q  D  V  R  R  V  S  G  V  A  P  T  L  V  R  S  A  S  E  -

    ACCAGTGAGAAGCGTCCTTTCATGTGTGCGTACCCAGGCTGCAATAAGAGATATTTTAAG
961 ---------+---------+---------+---------+---------+---------+ 1020
    TGGTCACTCTTCGCAGGAAAGTACACACGCATGGGTCCGACGTTATTCTCTATAAAATTC

    T  S  E  K  R  P  F  M  C  A  Y  P  G  C  N  K  R  Y  F  K  -

    CTGTCCCACTTACAGATGCACAGCCGGAAGCACACTGGTGAGAAACCATACCAGTGTGAC
1021---------+---------+---------+---------+---------+---------+ 1080
    GACAGGGTGAATGTCTACGTGTCGGCCTTCGTGTGACCACTCTTTGGTATGGTCACACTG

    L  S  H  L  Q  M  H  S  R  K  H  T  G  E  K  P  Y  Q  C  D  -

    TTCAAGGACTGCGAGAGAAGGTTTTCTCGCTCAGACCAGCTCAAAAGGCACCAAAGGAGA
1081---------+---------+---------+---------+---------+---------+ 1140
    AAGTTCCTGACGCTCTCTTCCAAAAGAGCGAGTCTGGTCGAGTTTTCCGTGGTTTCCTCT

    F  K' D  C  E  R  R  F  S  R  S  D  Q  L  K  R  H  Q  R  R  -

    CACACAGGTGTGAAACCATTCCAGTGTAAAACTTGTCAGCGAAAGTTTTCCCGGTCCGAC
1141---------+---------+---------+---------+---------+---------+ 1200
    GTGTGTCCACACTTTGGTAAGGTCACATTTTGAACAGTCGCTTTCAAAAGGGCCAGGCTG

    H  T  G  V  K  P  F  Q  C  K  T  C  Q  R  K  F  S  R  S  D  -
```


PCR clones→
CACCTGAAGACCCACACCAGGACTCATACAGGTAAAACAAGTGAAAAGCCCTTCAGCTGT
1201 ---------+---------+---------+---------+---------+---------+ 1260
GTGGACTTCTGGGTGTGGTCCTGAGTATGTCCATTTTGTTCACTTTTCGGGAAGTCGACA
H L K T H T R T H T G K T S E K P F S C -
CGGTGGTACAGTTGTCAGAAAAAGTTTGCGCGGTCAGACGAATTAGTCCGCCATCACAAG
1261 ---------+---------+---------+---------+---------+---------+ 1320
GCCACCATGTCAACAGTCTTTTTCAAACGCGCCAGTCTGCTTAATCAGGCGGTAGTGTTC
R W Y S C Q K K F A R S D E L V R H H K -
ATGCACCAGAGAAACATGACCAAACTCCAGCTGGCGCTTTGAGGGGTCCGACCCGGAGAC
1321 ---------+---------+---------+---------+---------+---------+ 1380
TACGTGGTCTCTTTGTACTGGTTTGAGGTCGACCGCGAAACTCCCCAGGCTGGGCCTCTG
M H Q R N M T K L Q L A L * G V R P G D -
AGTCCAGCGTCCCAGGCAAGAAAGTGTGTGCTTAAACTGCTTCCAAATCTGATTTTGAAA
1381 ---------+---------+---------+---------+---------+---------+ 1440
TCAGGTCGCAGGGTCCGTTCTTTCACACACGAATTTGACGAAGGTTTAGACTAAAACTTT
S P A S Q A R K C V L K L L P N L I L K -
TTCCTCCCTCTCACCTACAAAAGGACATGATGGTGGATCTTCATCCGACTTCCAAGACAG
1441 ---------+---------+---------+---------+---------+---------+ 1500
AAGGAGGGAGAGTGGATGTTTTCCTGTACTACCACCTAGAAGTAGGCTGAAGGTTCTGTC
F L P L T Y K R T * W W I F I R L P R Q -
CACACCTGTTTACTGAGTCCTATCAGGTTTGCCGGAAGGAGTCGGTCTCTGCCCACTCTT
1501 ---------+---------+---------+---------+---------+---------+ 1560
GTGTGGACAAATGACTCAGGATAGTCCAAACGGCCTTCCTCAGCCAGAGACGGGTGAGAA
H T C L L S P I R F A G R S R S L P T L -
GATTAACTCCCAGGCC
1561 ---------+------ 1576
CTAATTGAGGGTCCGG
D * L P G -

Pig *WT1* cDNA

```
    ACTTAGGAGCCACATTAAAAGGAGTTGCTGCTGGAAGCTCCAGCTCAATGAAATGGACAG
 19 -+---------+---------+---------+---------+---------+---------+-------- 78
    TGAATCCTCGGTGTAATTTTCCTCAACGACGACCTTCGAGGTCGAGTTACTTTACCTGTC

      L  G  A  T  L  K  G  V  A  A  G  S  S  S  S  M  K  W  T  E -

    AAGGGCAGAGCAACCACGGCACGGGGTACGAGAGCGATACGCACGCCACGCCCATCCTCT
 79 -+---------+---------+---------+---------+---------+---------+-------- 138
    TTCCCGTCTCGTTGGTGCCGTGCCCCATGCTCTCGCTATGCGTGCGGTGCGGGTAGGAGA

      G  Q  S  N  H  G  T  G  Y  E  S  D  T  H  A  T  P  I  L  C -

    GTGGTGCCCAGTACAGAATACACACCCACGGTGTCTTCAGGGGCATTCAGGACGTGCGAC
139 -+---------+---------+---------+---------+---------+---------+-------- 198
    CACCACGGGTCATGTCTTATGTGTGGGTGCCACAGAAGTCCCCGTAAGTCCTGCACGCTG

      G  A  Q  Y  R  I  H  T  H  G  V  F  R  G  I  Q  D  V  R  R -

    GTGTGCCTGGAGTAGCCCCGACTCTTGTGCGGTCGGCATCTGAGACCAGTGAGAAACGCC
199 -+---------+---------+---------+---------+---------+---------+-------- 258
    CACACGGACCTCATCGGGGCTGAGAACACGCCAGCCGTAGACTCTGGTCACTCTTTGCGG

      V  P  G  V  A  P  T  L  V  R  S  A  S  E  T  S  E  K  R  P -

    CCTTCATGTGTGCTTACCCGGGCTGCAATAAGAGATATTTTAAGCTGTCTCACTTACAGA
259 -+---------+---------+---------+---------+---------+---------+-------- 318
    GGAAGTACACACGAATGGGCCCGACGTTATTCTCTATAAAATTCGACAGAGTGAATGTCT

      F  M  C  A  Y  P  G  C  N  K  R  Y  F  K  L  S  H  L  Q  M -

    TGCATAGCCGGAAGCACACTGGTGAGAAACCATACCAGTGTGACTTCAAGGACTGTGAGA
319 -+---------+---------+---------+---------+---------+---------+-------- 378
    ACGTATCGGCCTTCGTGTGACCACTCTTTGGTATGGTCACACTGAAGTTCCTGACACTCT

      H  S  R  K  H  T  G  E  K  P  Y  Q  C  D  F  K  D  C  E  R -

    GAAGGTTTTCTCGTTCAGACCAGCTCAAAAGACACCAAAGGAGACACACAGGTGTGAAAC
379 -+---------+---------+---------+---------+---------+---------+-------- 438
    CTTCCAAAAGAGCAAGTCTGGTCGAGTTTTCTGTGGTTTCCTCTGTGTGTCCACACTTTG

      R  F  S  R  S  D  Q  L  K  R  H  Q  R  R  H  T  G  V  K  P -

    CATTCCAGTGTAAAACTTGTCAGCGAAAGTTCTCCCGGTCCGACCACCTGAAGACCCACA
439 -+---------+---------+---------+---------+---------+---------+-------- 498
    GTAAGGTCACATTTTGAACAGTCGCTTTCAAGAGGGCCAGGCTGGTGGACTTCTGGGTGT

      F  Q  C  K  T  C  Q  R  K  F  S  R  S  D  H  L  K  T  H  T -

    CCAGGACTCATACA
499 -+----------+-- 512
    GGTCCTGAGTATGT

      R  T  H  T  -
```

Marsupial mouse *WT1* cDNA, clone Sc41

```
    AGCGTGCCTCCCCCGGTCTATGGATGTCACACGCCCACCGACAGCTGCACCGGCAGCCAG
 22 --------+---------+---------+---------+---------+---------+- 81
    TCGCACGGAGGGGGCCAGATACCTACAGTGTGCGGGTGGCTGTCGACGTGGCCGTCGGTC

    S  V  P  P  P  V  Y  G  C  H  T  P  T  D  S  C  T  G  S  Q  -

    GCCCTGCTGCTCCGGACTCCCTACAACAGTGACAATCTGTACCAAATGACCTCGCAGCTA
 82 --------+---------+---------+---------+---------+---------+- 141
    CGGGACGACGAGGCCTGAGGGATGTTGTCACTGTTAGACATGGTTTACTGGAGCGTCGAT

    A  L  L  L  R  T  P  Y  N  S  D  N  L  Y  Q  M  T  S  Q  L  -

    GAATGCATGACCTGGAACCAGATGAACCTAGGGGCCACACTGAAGGGCCACACCACAGGA
142 --------+---------+---------+---------+---------+---------+- 201
    CTTACGTACTGGACCTTGGTCTACTTGGATCCCCGGTGTGACTTCCCGGTGTGGTGTCCT

    E  C  M  T  W  N  Q  M  N  L  G  A  T  L  K  G  H  T  T  G  -

    TATGAAAATGATAACCACACGACTCCGATCCTGTGTGGAGCCCAATATAGAATACACACT
202 --------+---------+---------+---------+---------+---------+- 261
    ATACTTTTACTATTGGTGTGCTGAGGCTAGGACACACCTCGGGTTATATCTTATGTGTGA

    Y  E  N  D  N  H  T  T  P  I  L  C  G  A  Q  Y  R  I  H  T  -

    CATGGCGTCTTTAGAGGAATACAAGATGTCCGGCGAGTGCCCGGGGTAGCCCCTACTATC
262 --------+---------+---------+---------+---------+---------+- 321
    GTACCGCAGAAATCTCCTTATGTTCTACAGGCCGCTCACGGGCCCCATCGGGGATGATAG

    H  G  V  F  R  G  I  Q  D  V  R  R  V  P  G  V  A  P  T  I  -

    GTGAGGTCAGCCACGGAGACCAATGAGAAACGCCCATTCATGTGTGCATATCCAGGCTGC
322 --------+---------+---------+---------+---------+---------+- 381
    CACTCCAGTCGGTGCCTCTGGTTACTCTTTGCGGGTAAGTACACACGTATAGGTCCGACG

    V  R  S  A  T  E  T  N  E  K  R  P  F  M  C  A  Y  P  G  C  -

    AACAAGAGATACTTTAAGCTGTCCCACTTACAGATGCATAGCAGGAAGCATACTGGTGAG
382 --------+---------+---------+---------+---------+---------+- 441
    TTGTTCTCTATGAAATTCGACAGGGTGAATGTCTACGTATCGTCCTTCGTATGACCACTC

    N  K  R  Y  F  K  L  S  H  L  Q  M  H  S  R  K  H  T  G  E  -

    AAACCCTACCAGTGTGACTTCAAGGACTGTGAACGGCGATTCTCTCGATCAGACCAACTC
442 --------+---------+---------+---------+---------+---------+- 501
    TTTGGGATGGTCACACTGAAGTTCCTGACACTTGCCGCTAAGAGAGCTAGTCTGGTTGAG

    K  P  Y  Q  C  D  F  K  D  C  E  R  R  F  S  R  S  D  Q  L  -

    AAAAGACACCAAAGGAGACACACAGGTGTGAAACCATTCCAGTGTAAAACCTGTCAGAGA
502 --------+---------+---------+---------+---------+---------+- 561
    TTTTCTGTGGTTTCCTCTGTGTGTCCACACTTTGGTAAGGTCACATTTTGGACAGTCTCT

    K  R  H  Q  R  R  H  T  G  V  K  P  F  Q  C  K  T  C  Q  R  -

    AAGTTCTCCCGGTCTGACCACCTGAAGACACACACCAGGACTCATACAGGTAAAACAAGT
562 --------+---------+---------+---------+---------+---------+- 621
    TTCAAGAGGGCCAGACTGGTGGACTTCTGTGTGTGGTCCTGAGTATGTCCATTTTGTTCA

    K  F  S  R  S  D  H  L  K  T  H  T  R  T  H  T  G  K  T  S  -
```

```
    GAAAAGCCCTTCAGCTGCCGGTGGCCCAGTTGTCAGAAAAAATTTGCCCGCTCAGATGAA
622 --------+---------+---------+---------+---------+---------+- 681
    CTTTTCGGGAAGTCGACGGCCACCGGGTCAACAGTCTTTTTTAAACGGGCGAGTCTACTT

    E  K  P  F  S  C  R  W  P  S  C  Q  K  K  F  A  R  S  D  E   -

    TTAGTGCGTCACCACAACATGCACCAAAGGAACATGACGAAGCTCCAGCTGACGCTTTAA
682 --------+---------+---------+---------+---------+---------+- 741
    AATCACGCAGTGGTGTTGTACGTGGTTTCCTTGTACTGCTTCGAGGTCGACTGCGAAATT

    L  V  R  H  H  N  M  H  Q  R  N  M  T  K  L  Q  L  T  L  *   -

    AAGGTCCAGGTCAAGGATGCATCTCTCTCCCAGGTAGAATGTTGTATGAACCCCGATGGA
742 --------+---------+---------+---------+---------+---------+- 801
    TTCCAGGTCCAGTTCCTACGTAGAGAGAGGGTCCATCTTACAACATACTTGGGGCTACCT

    K  V  Q  V  K  D  A  S  L  S  Q  V  E  C  C  M  N  P  D  G   -

    AATGGGCTTTTAGCTACCTTCAACTACCTTCCCAATAAGGAAACCCCAACCGACCACGAA
802 --------+---------+---------+---------+---------+---------+- 861
    TTACCCGAAAATCGATGGAAGTTGATGGAAGGGTTATTCCTTTGGGGTTGGCTGGTGCTT

    N  G  L  L  A  T  F  N  Y  L  P  N  K  E  T  P  T  D  H  E   -

    GACGAGTCACCCATCGGGATGGAGCGGGTTCCAGGTTGTCGATCATCACAGCAGTCTTCA
862 --------+---------+---------+---------+---------+---------+- 921
    CTGCTCAGTGGGTAGCCCTACCTCGCCCAAGGTCCAACAGCTAGTAGTGTCGTCAGAAGT

    D  E  S  P  I  G  M  E  R  V  P  G  C  R  S  S  Q  Q  S  S   -

    TTTTCTCCAATCTCTTCTTCACTCCTCAAGACTCAATTTCCATTCAGT
922 --------+---------+---------+---------+---------+--------- 969
    AAAAGAGGTTAGAGAAGAAGTGAGGAGTTCTGAGTTAAAGGTAAGTCA

    F  S  P  I  S  S  S  L  L  K  T  Q  F  P  F  S   -
```

Unsequenced region ~800bp

```
    TCTATCTGATAATGCATTGCTCATTGCCTTAACTTCAAAACAATTGCAAAGCGACTAAAT
  1 --------+---------+---------+---------+---------+---------+ 60
    AGATAGACTATTACGTAACGAGTAACGGAATTGAAGTTTTGTTAACGTTTCGCTGATTTA

    S  I  *  *  C  I  A  H  C  L  N  F  K  T  I  A  K  R  L  N   -
     L  S  D  N  A  L  L  I  A  L  T  S  K  Q  L  Q  S  D  *  I   -
      Y  L  I  M  H  C  S  L  P  *  L  Q  N  N  C  K  A  T  K  Y  -

    ACTTAGATGCGTTCAGATGTTACTGATGATGATGATAANNCTTAAGAATGTAATCTGCAT
 61 --------+---------+---------+---------+---------+---------+ 120
    TGAATCTACGCAAGTCTACAATGACTACTACTACTATTNNGAATTCTTACATTAGACGTA

    T  *  M  R  S  D  V  T  D  D  D  D  ?  ?  *  E  C  N  L  H   -
     L  R  C  V  Q  M  L  L  M  M  M  I  ?  L  K  N  V  I  C  I   -
      L  D  A  F  R  C  Y  *  *  *  *  *  ?  L  R  M  *  S  A  F  -

    TTTTCACCTTCAAATTTCA ATTAAA ACCCATTCAAAAAGGAAAAAAAAAAAAAAAA
121 --------+---------+---------+---------+---------+------ 176
    AAAAGTGGAAGTTTAAAGTTAATTTTGGGTAAGTTTTTCCTTTTTTTTTTTTTTTT

    F  S  P  S  N  F  N  *  N  P  F  K  K  E  K  K  K  K  ?   -
     F  H  L  Q  I  S  I  K  T  H  S  K  R  K  K  K  K  K  ?   -
      F  T  F  K  F  Q  L  K  P  I  Q  K  G  K  K  K  K  K   -
```

Chick *WT1* cDNA, clone C2.1

```
    CGGCACGAGGGCGCTGAGGCGGCTGTGGGTCCACCCCCTCCCTATCCCCCCCCCCGGCAC
  1 ---------+---------+---------+---------+---------+---------+ 60
    GCCGTGCTCCCGCGACTCCGCCGACACCCAGGTGGGGGAGGGATAGGGGGGGGGGCCGTG

    R  H  E  G  A  E  A  A  V  G  P  P  P  P  Y  P  P  P  R  H  -

    GCGCGGACAGACACACGGCCGCACACACGGACACGCGGCCACGGACACGCGCGCACCCCA
 61 ---------+---------+---------+---------+---------+---------+ 120
    CGCGCCTGTCTGTGTGCCGGCGTGTGTGCCTGTGCGCCGGTGCCTGTGCGCGCGTGGGGT

    A  R  T  D  T  R  P  H  T  R  T  R  G  H  G  H  A  R  T  P  -

    CCCTGCGCTGCGAGCGGGCACAGCCCCGGAGGCGATCCTGGAGAACAACCTTTCCCTGTG
121 ---------+---------+---------+---------+---------+---------+ 180
    GGGACGCGACGCTCGCCCGTGTCGGGGCCTCCGCTAGGACCTCTTGTTGGAAAGGGACAC

    P  C  A  A  S  G  H  S  P  G  G  D  P  G  E  Q  P  F  P  V  -

    CGCTGCTCGCCGCTCCCGCGGGCTGGACCCGGGGAGCTGCGTCGCGGAGCAAACCTCTCC
181 ---------+---------+---------+---------+---------+---------+ 240
    GCGACGAGCGGCGAGGGCGCCCGACCTGGGCCCCTCGACGCAGCGCCTCGTTTGGAGAGG

    R  C  S  P  L  P  R  A  G  P  G  E  L  R  R  G  A  N  L  S  -

    AAGGGCTGAGCTCTCCGGCTGGTGCCTGCAGCCGGAGCGTCCGGCTGGCGGCAGAGGCAG
241 ---------+---------+---------+---------+---------+---------+ 300
    TTCCCGACTCGAGAGGCCGACCACGGACGTCGGCCTCGCAGGCCGACCGCCGTCTCCGTC

    K  G  *  A  L  R  L  V  P  A  A  G  A  S  G  W  R  Q  R  Q  -

    GTGTCCTGGAATAAGTGCGGCCTGGGGGGAGACTGGCTGCCCGGCGGCAGGCAGAAGGTC
301 ---------+---------+---------+---------+---------+---------+ 360
    CACAGGACCTTATTCACGCCGGACCCCCCTCTGACCGACGGGCCGCCGTCCGTCTTCCAG

    V  S  W  N  K  C  G  L  G  G  D  W  L  P  G  G  R  Q  K  V  -
Translation Start→
    TGCGAGCAGATGGGGTCCGACGTCCGGGACCTGAACGCGCTGCTGCCCTCCGTGCCCTCC
361 ---------+---------+---------+---------+---------+---------+ 420
    ACGCTCGTCTACCCCAGGCTGCAGGCCCTGGACTTGCGCGACGACGGGAGGCACGGGAGG

    C  E  Q  M  G  S  D  V  R  D  L  N  A  L  L  P  S  V  P  S  -

    CTACCGGGCAACAGCAACTGCGCCATGCCGGTGAGCAGCGCGGCGCAGTGGGCTCCCGTC
421 ---------+---------+---------+---------+---------+---------+ 480
    GATGGCCCGTTGTCGTTGACGCGGTACGGCCACTCGTCGCGCCGCGTCACCCGAGGGCAG

    L  P  G  N  S  N  C  A  M  P  V  S  S  A  A  Q  W  A  P  V  -

    CTGGACTCTCCCCCGGGGGCCTCCTACGGCTCGCTGGGGCCGCACTCCTTCATCAAGCAG
481 ---------+---------+---------+---------+---------+---------+ 540
    GACCTGAGAGGGGGCCCCCGGAGGATGCCGAGCGACCCCGGCGTGAGGAAGTAGTTCGTC

    L  D  S  P  P  G  A  S  Y  G  S  L  G  P  H  S  F  I  K  Q  -

    GAACCCAGCTGGAACGGGTCGGACCCGCACGAGGAGCAGTACCTGAGCGCTTTCACCGTC
541 ---------+---------+---------+---------+---------+---------+ 600
    CTTGGGTCGACCTTGCCCAGCCTGGGCGTGCTCCTCGTCATGGACTCGCGAAAGTGGCAG

    E  P  S  W  N  G  S  D  P  H  E  E  Q  Y  L  S  A  F  T  V  -
```

```
     CACTTCTCCGGGCAGTTCACGGGCACGGCCGGGGCCTGCCGCTACGGGCCCTTCGGCGCC
601 ---------+---------+---------+---------+---------+---------+ 660
     GTGAAGAGGCCCGTCAAGTGCCCGTGCCGGCCCCGGACGGCGATGCCCGGGAAGCCGCGG

     H  F  S  G  Q  F  T  G  T  A  G  A  C  R  Y  G  P  F  G  A  -

     CCGCCGCCCAGCCAGCCGCCCTCCGGCCAGGCCAGGATGTTCCCCCAACGGCCCTACCTG
661 ---------+---------+---------+---------+---------+---------+ 720
     GGCGGCGGGTCGGTCGGCGGGAGGCCGGTCCGGTCCTACAAGGGGGTTGCCGGGATGGAC

     P  P  P  S  Q  P  P  S  G  Q  A  R  M  F  P  Q  R  P  Y  L  -

     CCCAACTGCCTGGAGAGCCAGCAAGCCATTCGCAACCAGGGTTACGGCACCGTGGCCTTT
721 ---------+---------+---------+---------+---------+---------+ 780
     GGGTTGACGGACCTCTCGGTCGTTCGGTAAGCGTTGGTCCCAATGCCGTGGCACCGGAAA

     P  N  C  L  E  S  Q  Q  A  I  R  N  Q  G  Y  G  T  V  A  F  -

     GACGGGACTCCGAGCTACGGCCACACGCCGTCTCACCACGCCGCGCAGTTTCAGAACCAC
781 ---------+---------+---------+---------+---------+---------+ 840
     CTGCCCTGAGGCTCGATGCCGGTGTGCGGCAGAGTGGTGCGGCGCGTCAAAGTCTTGGTG

     D  G  T  P  S  Y  G  H  T  P  S  H  H  A  A  Q  F  Q  N  H  -

     TCTTTCAAACACGAAGACCCCATGAGCCAGCAGCCATCTCTAGGGGACCAGCAGTACTCG
841 ---------+---------+---------+---------+---------+---------+ 900
     AGAAAGTTTGTGCTTCTGGGGTACTCGGTCGTCGGTAGAGATCCCCTGGTCGTCATGAGC

     S  F  K  H  E  D  P  M  S  Q  Q  P  S  L  G  D  Q  Q  Y  S  -

     GTGCCCCCTCCGGTGTACGGCTGTCACACCCCCACGGACACGTGCACGGGCAGCCAGGCC
901 ---------+---------+---------+---------+---------+---------+ 960
     CACGGGGGAGGCCACATGCCGACAGTGTGGGGGTGCCTGTGCACGTGCCCGTCGGTCCGG

     V  P  P  P  V  Y  G  C  H  T  P  T  D  T  C  T  G  S  Q  A  -
                                                  ← PCR clones
     CTGCTCCTGCGGACCCCCTACAACAGTGACAATTTGTACCAAATGACGTCACGACTGGAA
961 ---------+---------+---------+---------+---------+---------+ 1020
     GACGAGGACGCCTGGGGGATGTTGTCACTGTTAAACATGGTTTACTGCAGTGCTGACCTT

     L  L  L  R  T  P  Y  N  S  D  N  L  Y  Q  M  T  S  R  L  E  -

     TGCATGACATGGAACCAAATGAACCTGGGATCCACGCTGAAGGGCCATACGACAGGATAT
1021 ---------+---------+---------+---------+---------+---------+ 1080
     ACGTACTGTACCTTGGTTTACTTGGACCCTAGGTGCGACTTCCCGGTATGCTGTCCTATA

     C  M  T  W  N  Q  M  N  L  G  S  T  L  K  G  H  T  T  G  Y  -

     GAAAATGAGAACCACAGTGCTCCCATGTTATACAGCTGTGGGGCCCAATACAGAATACAC
1081 ---------+---------+---------+---------+---------+---------+ 1140
     CTTTTACTCTTGGTGTCACGAGGGTACAATATGTCGACACCCCGGGTTATGTCTTATGTG

     E  N  E  N  H  S  A  P  M  L  Y  S  C  G  A  Q  Y  R  I  H  -

     ACCCATGGAGTCTTTAGAGGCATACAAGATGTCCGACGAGTGCCAGGAGTAGCTCCGACT
1141 ---------+---------+---------+---------+---------+---------+ 1200
     TGGGTACCTCAGAAATCTCCGTATGTTCTACAGGCTGCTCACGGTCCTCATCGAGGCTGA

     T  H  G  V  F  R  G  I  Q  D  V  R  R  V  P  G  V  A  P  T  -
```

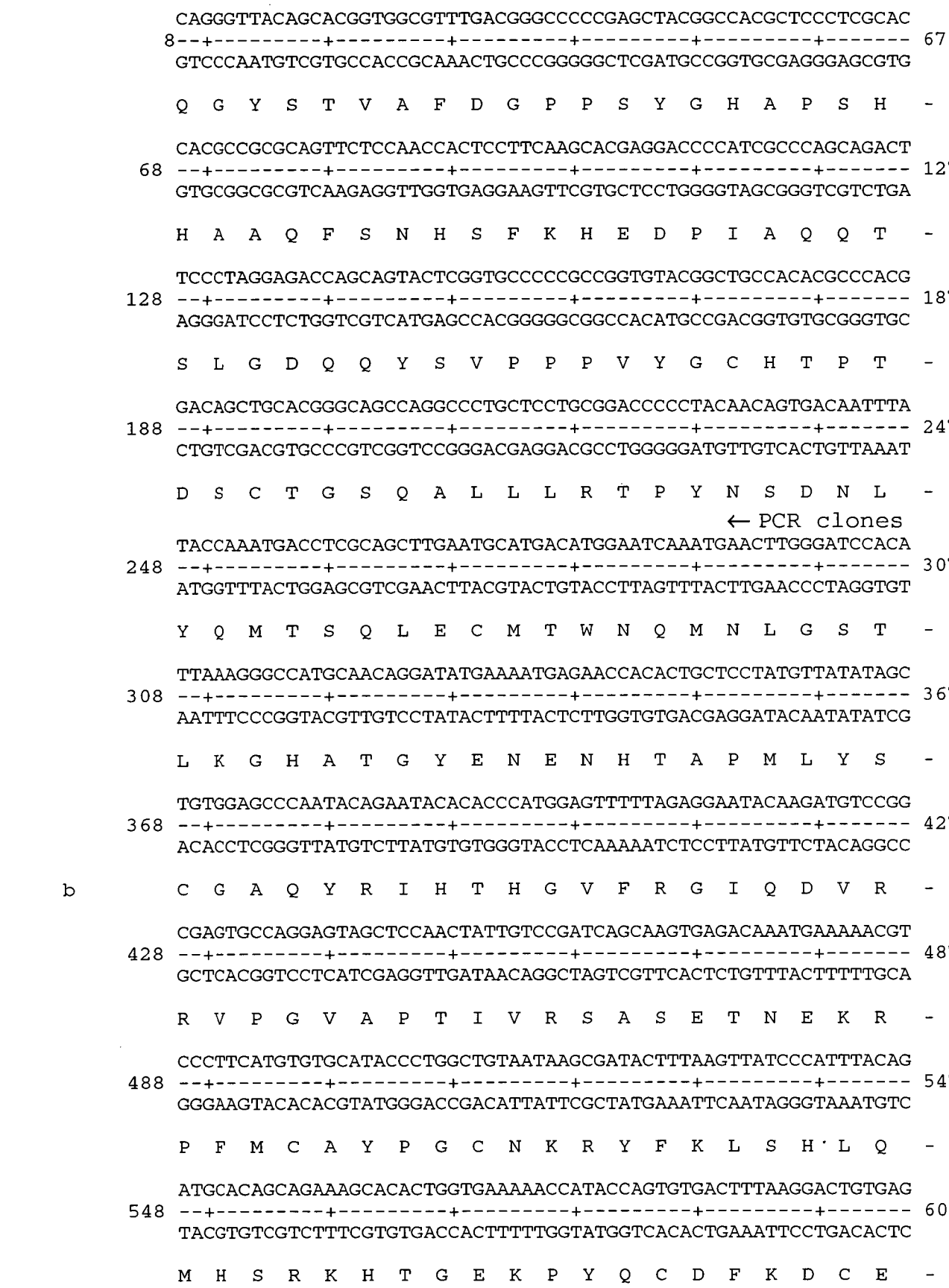
```
      ATTGTCCGATCAGCAAGTGAGACAAATGAAAAACGCCCCTTCATGTGTGCCTACCCCGGC
1201  ---------+---------+---------+---------+---------+---------+ 1260
      TAACAGGCTAGTCGTTCACTCTGTTTACTTTTTGCGGGGAAGTACACACGGATGGGGCCG

      I  V  R  S  A  S  E  T  N  E  K  R  P  F  M  C  A  Y  P  G  -

      TGCAACAAGCGATACTTCAAGCTGTCCCATCTACAGATGCACAGCAGAAAGCACACTGGT
1261  ---------+---------+---------+---------+---------+---------+ 1320
      ACGTTGTTCGCTATGAAGTTCGACAGGGTAGATGTCTACGTGTCGTCTTTCGTGTGACCA

      C  N  K  R  Y  F  K  L  S  H  L  Q  M  H  S  R  K  H  T  G  -

      GAAAAACCATATCAGTGTGATTTTAAGGACTGTGAACGAAGATTTTCTCGTTCAGACCAA
1321  ---------+---------+---------+---------+---------+---------+ 1380
      CTTTTTGGTATAGTCACACTAAAATTCCTGACACTTGCTTCTAAAAGAGCAAGTCTGGTT

      E  K  P  Y  Q  C  D  F  K  D  C  E  R  R  F  S  R  S  D  Q  -

      CTGAAACGGCACCAAAGACGACACACAGGTGTGAAACCCTTCCAATGTAAAACCTGTCAG
1381  ---------+---------+---------+---------+---------+---------+ 1440
      GACTTTGCCGTGGTTTCTGCTGTGTGTCCACACTTTGGGAAGGTTACATTTTGGACAGTC

      L  K  R  H  Q  R  R  H  T  G  V  K  P  F  Q  C  K  T  C  Q  -

                          PCR clones  →                    *
      AGAAAGTTCTCCAGATCTGATCATCTGAAGACTCATACCAGGACTCATACAGCTATAGTG
1441  ---------+---------+---------+---------+---------+---------+ 1500
      TCTTTCAAGAGGTCTAGACTAGTAGACTTCTGAGTATGGTCCTGAGTATGTCGATATCAC

      R  K  F  S  R  S  D  H  L  K  T  H  T  R  T  H  T  A  I  V  -

      AAATGAACATCATGGCCTCAGTCCACAAATGTCAACTCCATCTACTGTGTTCACTGAAAA
1501  ---------+---------+---------+---------+---------+---------+ 1560
      TTTACTTGTAGTACCGGAGTCAGGTGTTTACAGTTGAGGTAGATGACACAAGTGACTTTT

      K  *  T  S  W  P  Q  S  T  N  V  N  S  I  Y  C  V  H  *  K  -

      AGGGCCTACAAGAAAGAAATTTA
1561  ---------+---------+--- 1583
      TCCCGGATGTTCTTTCTTTAAAT

      R  A  Y  K  K  E  I  ?  -
```

Chick *WT1* PCR sequence of ZF3&4

```
                    *
      CAGGACTCATACAGGTAAAACAAGTGAAAAGCCTTTCAGTTGCCGATGGCCCAGCTGTCA
1479  +---------+---------+---------+---------+---------+--------- 1528
      GTCCTGAGTATGTCCATTTTGTTCACTTTTCGGAAAGTCAACGGCTACCGGGTCGACAGT

       R  T  H  T  G  K  T  S  E  K  P  F  S  C  R  W  P  S  C  Q  -

      AAAAAAATT
1529  +--------- 1537
      TTTTTTTAA

       K  K  ?  -
```

* Position in C2.1 where homology to *WT1* disappears

Alligator *WT1* cDNA, clone AL1



```
    AGAAGATTTTCCCGTTCAGACCAACTCAAACGACACCAAAGACGACACACAGGTGTGAAA
608 --+---------+---------+---------+---------+---------+---------+------- 667
    TCTTCTAAAAGGGCAAGTCTGGTTGAGTTTGCTGTGGTTTCTGCTGTGTGTCCACACTTT

    R  R  F  S  R  S  D  Q  L  K  R  H  Q  R  R  H  T  G  V  K   -

    CCCTTCCAGTGTAAAACCTGTCAGAGAAAGTTCTCTAGGTCTGATCATCTAAAGACTCAC
668 --+---------+---------+---------+---------+---------+---------+------- 727
    GGGAAGGTCACATTTTGGACAGTCTCTTTCAAGAGATCCAGACTAGTAGATTTCTGAGTG

    P  F  Q  C  K  T  C  Q  R  K  F  S  R  S  D  H  L  K  T  H   -

                PCR clones  →
    ACCAGGACTCATACAGGTAAAACAAGTGAAAAACCATTCAGCTGTCGATGGCCCAGCTGT
728 --+---------+---------+---------+---------+---------+---------+------- 787
    TGGTCCTGAGTATGTCCATTTTGTTCACTTTTTGGTAAGTCGACAGCTACCGGGTCGACA

    T  R  T  H  T  G  K  T  S  E  K  P  F  S  C  R  W  P  S  C   -

    CAAAAAAAATTTGCCAGATCTGATGAATTAGTTCGTCATCACAACATGCACCAGAGGAAC
788 --+---------+---------+---------+---------+---------+---------+------- 847
    GTTTTTTTTAAACGGTCTAGACTACTTAATCAAGCAGTAGTGTTGTACGTGGTCTCCTTG

    Q  K  K  F  A  R  S  D  E  L  V  R  H  H  N  M  H  Q  R  N   -

    ATGACTAAACTGCAGTTGGCCCTTTAGACAGTGAAACAAACG**AATAAATAAA**TGGGACTT
848 --+---------+---------+---------+---------+---------+---------+------- 907
    TACTGATTTGACGTCAACCGGGAAATCTGTCACTTTGTTTGCTTATTTATTTACCCTGAA

    M  T  K  L  Q  L  A  L  *  T  V  K  Q  T  N  K  *  M  G  L   -

    TATTGAAACTACTGCAGACTTTTTCAATCATCCTTGATCACATCTCTAGGAAAGCTACAG
908 --+---------+---------+---------+---------+---------+---------+------- 967
    ATAACTTTGATGACGTCTGAAAAAGTTAGTAGGAACTAGTGTAGAGATCCTTTCGATGTC

    Y  *  N  Y  C  R  L  F  Q  S  S  L  I  T  S  L  G  K  L  Q   -

    CTATCTTCATTACCCATTTGAGACAGGTTAGCTAAACCGTATGGATTCTGTTCCAGTTTT
968 --+---------+---------+---------+---------+---------+---------+------- 1027
    GATAGAAGTAATGGGTAAACTCTGTCCAATCGATTTGGCATACCTAAGACAAGGTCAAAA

    L  S  S  L  P  I  *  D  R  L  A  K  P  Y  G  F  C  S  S  F   -

    CTGACCTATGGACTTTAAAAAATGCTATTCAATGTCCTATTCAATAATGTTAAGTGGTTT
1028 --+---------+---------+---------+---------+---------+---------+------- 1087
    GACTGGATACCTGAAATTTTTTACGATAAGTTACAGGATAAGTTATTACAATTCACCAAA

    L  T  Y  G  L  *  K  M  L  F  N  V  L  F  N  N  V  K  W  F   -

    TAATTTGTAACTTTGTTATGGAATTACCAAAACTGTTACAGTTTAAATATTAATATAAAC
1088 --+---------+---------+---------+---------+---------+---------+------- 1147
    ATTAAACATTGAAACAATACCTTAATGGTTTTGACAATGTCAAATTTATAATTATATTTG

    *  F  V  T  L  L  W  N  Y  Q  N  C  Y  S  L  N  I  N  I  N   -

    TGTGTTGTTACTTATGTATTTTTTTTTCCTCAAAAAAGTTCTGTTTTTATTTTTACTTGA
1148 --+---------+---------+---------+---------+---------+---------+------- 1207
    ACACAACAATGAATACATAAAAAAAAAGGAGTTTTTTCAAGACAAAAATAAAAATGAACT

    C  V  V  T  Y  V  F  F  F  P  Q  K  S  S  V  F  I  F  T  *   -
```

```
     ACATTTCAGGTGGAAAGTACTGTGCCTTTTTCAGTAATACTGTGTGTTGGGTATGTGTAC
1208 --+---------+---------+---------+---------+---------+---------+------- 1267
     TGTAAAGTCCACCTTTCATGACACGGAAAAAGTCATTATGACACACAACCCATACACATG

     T  F  Q  V  E  S  T  V  P  F  S  V  I  L  C  V  G  Y  V  Y   -

     ATGTCAAAGACAAGCTTTGTGCTTGCATTCGTACTCGTGC
1268 --+---------+---------+---------+------- 1307
     TACAGTTTCTGTTCGAAACACGAACGTAAGCATGAGCACG

   .  M  S  K  T  S  F  V  L  A  F  V  L  V  ?  -
```

Xenopus WT1 cDNA

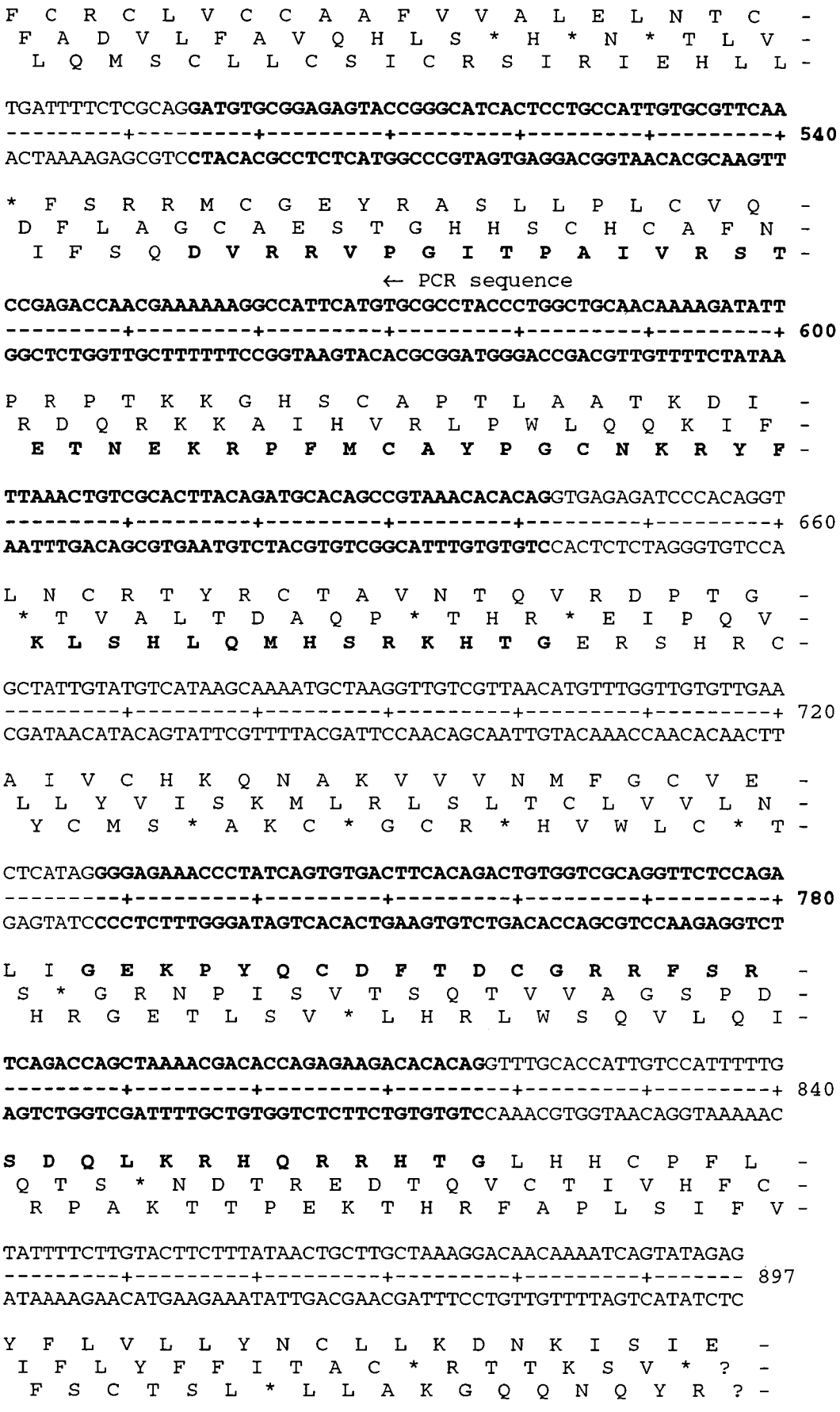
```
    TGCCTACCCAGGATGCAACAAGAGATACTTCAAATTGTCCCATTTGCAGATGCACAGCAG
 21 ---------+---------+---------+---------+---------+---------+ 80
    ACGGATGGGTCCTACGTTGTTCTCTATGAAGTTTAACAGGGTAAACGTCTACGTGTCGTC

     A  Y  P  G  C  N  K  R  Y  F  K  L  S  H  L  Q  M  H  S  R  -

    GAAACACACAGGGGAAAAACCGTACCAGTGTGATTTTAAAGACTGTGAGAGGCGATTTTC
 81 ---------+---------+---------+---------+---------+---------+ 140
    CTTTGTGTGTCCCCTTTTTGGCATGGTCACACTAAAATTTCTGACACTCTCCGCTAAAAG

     K  H  T  G  E  K  P  Y  Q  C  D  F  K  D  C  E  R  R  F  S  -

    TCGTTCAGATCAACTTAAAAGGCACCAAAGGAGGCACACAGGTATCAAACCTTTCCAATG
141 ---------+---------+---------+---------+---------+---------+ 200
    AGCAAGTCTAGTTGAATTTTCCGTGGTTTCCTCCGTGTGTCCATAGTTTGGAAAGGTTAC

     R  S  D  Q  L  K  R  H  Q  R  R  H  T  G  I  K  P  F  Q  C  -

    CAAAACCTGCCAACGAAAGTTCTCCAGGTCCGACCATCTGAAGACTCACACCAGGACTCA
201 ---------+---------+---------+---------+---------+---------+ 260
    GTTTTGGACGGTTGCTTTCAAGAGGTCCAGGCTGGTAGACTTCTGAGTGTGGTCCTGAGT

     K  T  C  Q  R  K  F  S  R  S  D  H  L  K  T  H  T  R  T  H  -

    TACAGGC
261 ------- 267
    ATGTCCG

     T  G  -
```

Axolotl *WT1* cDNA

```
    AAGTTGTCCCATTTGCAGATGCACAACCGAAAGCACACAGGTGAAAAGCCGTACCAGTGT
  1 --------+---------+---------+---------+---------+---------+- 60
    TTCAACAGGGTAAACGTCTACGTGTTGGCTTTCGTGTGTCCACTTTTCGGCATGGTCACA
    K  L  S  H  L  Q  M  H  N  R  K  H  T  G  E  K  P  Y  Q  C  -

    GATTTTAAGGATTGTGAGCGGCGATTTTCCCGTTCCGACCAACTCAAAAGACATCAAAGA
 61 --------+---------+---------+---------+---------+---------+- 120
    CTAAAATTCCTAACACTCGCCGCTAAAAGGGCAAGGCTGGTTGAGTTTTCTGTAGTTTCT

    D  F  K  D  C  E  R  R  F  S  R  S  D  Q  L  K  R  H  Q  R  -

    AGACACACAGGTGTGAAACCATTCCAATGTAAAACCTGTCAGAGAAAGTTCTCTCGATCT
121 --------+---------+---------+---------+---------+---------+- 180
    TCTGTGTGTCCACACTTTGGTAAGGTTACATTTTGGACAGTCTCTTTCAAGAGAGCTAGA

    R  H  T  G  V  K  P  F  Q  C  K  T  C  Q  R  K  F  S  R  S  -

    GATCATTTGAAGACCCACACCAGGACTCATACAGGT
181 --------+---------+---------+------- 217
    CTAGTAAACTTCTGGGTGTGGTCCTGAGTATGTCCA

    D  H  L  K  T  H  T  R  T  H  T  G  -
```

Zebrafish *WT1* genomic DNA, clone Z1a (exons in bold)

```
    CGGCACGAGATAACAGCTAATCTGAACATAACTGTCACCTACAAATCAAACGCTTGTGCT
  1 ---------+---------+---------+---------+---------+---------+ 60
    GCCGTGCTCTATTGTCGATTAGACTTGTATTGACAGTGGATGTTTAGTTTGCGAACACGA

    R  H  E  I  T  A  N  L  N  I  T  V  T  Y  K  S  N  A  C  A  -
     G  T  R  *  Q  L  I  *  T  *  L  S  P  T  N  Q  T  L  V  L  -
      A  R  D  N  S  *  S  E  H  N  C  H  L  Q  I  K  R  L  C  *  -

    AGCCTGCAGAGCTGTGTAGCCTAGCCTGCTGTGTTATCCTTGGGTGGGTCAAGGGAGACG
 61 ---------+---------+---------+---------+---------+---------+ 120
    TCGGACGTCTCGACACATCGGATCGGACGACACAATAGGAACCCACCCAGTTCCCTCTGC

    S  L  Q  S  C  V  A  *  P  A  V  L  S  L  G  G  S  R  E  T  -
     A  C  R  A  V  *  P  S  L  L  C  Y  P  W  V  G  Q  G  R  R  -
      P  A  E  L  C  S  L  A  C  C  V  I  L  G  W  V  K  G  D  V  -

    TCTGCCACTACAGTCGTAGGAACAGTGTATTTAGTTATCAATAGCACTTAACACTCATTC
121 ---------+---------+---------+---------+---------+---------+ 180
    AGACGGTGATGTCAGCATCCTTGTCACATAAATCAATAGTTATCGTGAATTGTGAGTAAG

    S  A  T  T  V  V  G  T  V  Y  L  V  I  N  S  T  *  H  S  F  -
     L  P  L  Q  S  *  E  Q  C  I  *  L  S  I  A  L  N  T  H  S  -
      C  H  Y  S  R  R  N  S  V  F  S  Y  Q  *  H  L  T  L  I  H  -

    ATTACTGTGCCAGCATACAGAAATGCCATTATGCTAAGCGCTAGTTAATTAAGAGCAGAT
181 ---------+---------+---------+---------+---------+---------+ 240
    TAATGACACGGTCGTATGTCTTTACGGTAATACGATTCGCGATCAATTAATTCTCGTCTA

    I  T  V  P  A  Y  R  N  A  I  M  L  S  A  S  *  L  R  A  D  -
     L  L  C  Q  H  T  E  M  P  L  C  *  A  L  V  N  *  E  Q  M  -
      Y  C  A  S  I  Q  K  C  H  Y  A  K  R  *  L  I  K  S  R  W  -

    GGGCTTCAGCAGCGGGTAAATGAGGGAAGCGGGAGGAACGGGGGAGCGAACGTGAAATAA
241 ---------+---------+---------+---------+---------+---------+ 300
    CCCGAAGTCGTCGCCCATTTACTCCCTTCGCCCTCCTTGCCCCCTCGCTTGCACTTTATT

    G  L  Q  Q  R  V  N  E  G  S  G  R  N  G  G  A  N  V  K  *  -
     G  F  S  S  G  *  M  R  E  A  G  G  T  G  E  R  T  *  N  K  -
      A  S  A  A  G  K  *  G  K  R  E  E  R  G  S  E  R  E  I  R  -

    GGTCTGCCTAACGTCGTGACTAACCGTCACAGTTAGAGCTCATATCAGCAGCACGGTGAT
301 ---------+---------+---------+---------+---------+---------+ 360
    CCAGACGGATTGCAGCACTGATTGGCAGTGTCAATCTCGAGTATAGTCGTCGTGCCACTA

    G  L  P  N  V  V  T  N  R  H  S  *  S  S  Y  Q  Q  H  G  D  -
     V  C  L  T  S  *  L  T  V  T  V  R  A  H  I  S  S  T  V  I  -
      S  A  *  R  R  D  *  P  S  Q  L  E  L  I  S  A  A  R  *  S  -

    CGCATGATCATTCACAGCTATCTCAGTATGACCATGCATGCTATGATGCTTTTTTTTTTT
361 ---------+---------+---------+---------+---------+---------+ 420
    GCGTACTAGTAAGTGTCGATAGAGTCATACTGGTACGTACGATACTACGAAAAAAAAAAA

    R  M  I  I  H  S  Y  L  S  M  T  M  H  A  M  M  L  F  F  F  -
     A  *  S  F  T  A  I  S  V  *  P  C  M  L  *  C  F  F  F  F  -
      H  D  H  S  Q  L  S  Q  Y  D  H  A  C  Y  D  A  F  F  F  F  -

    TTTTGCAGATGTCTTGTTTGCTGTGCAGCATTTGTCGTAGCATTAGAATTGAACACTTGT
421 ---------+---------+---------+---------+---------+---------+ 480
    AAAACGTCTACAGAACAAACGACACGTCGTAAACAGCATCGTAATCTTAACTTGTGAACA
```


F C R C L V C C A A F V V A L E L N T C -
F A D V L F A V Q H L S * H * N * T L V -
L Q M S C L L C S I C R S I R I E H L L -
TGATTTTCTCGCAGGATGTGCGGAGAGTACCGGGCATCACTCCTGCCATTGTGCGTTCAA
481 ---------+---------+---------+---------+---------+---------+ 540
ACTAAAAGAGCGTCCTACACGCCTCTCATGGCCCGTAGTGAGGACGGTAACACGCAAGTT
* F S R R M C G E Y R A S L L P L C V Q -
D F L A G C A E S T G H H S C H C A F N -
I F S Q D V R R V P G I T P A I V R S T -
← PCR sequence
CCGAGACCAACGAAAAAAGGCCATTCATGTGCGCCTACCCTGGCTGCAACAAAAGATATT
541 ---------+---------+---------+---------+---------+---------+ 600
GGCTCTGGTTGCTTTTTTCCGGTAAGTACACGCGGATGGGACCGACGTTGTTTTCTATAA
P R P T K K G H S C A P T L A A T K D I -
R D Q R K K A I H V R L P W L Q Q K I F -
E T N E K R P F M C A Y P G C N K R Y F -
TTAAACTGTCGCACTTACAGATGCACAGCCGTAAACACACAGGTGAGAGATCCCACAGGT
601 ---------+---------+---------+---------+---------+---------+ 660
AATTTGACAGCGTGAATGTCTACGTGTCGGCATTTGTGTGTCCACTCTCTAGGGTGTCCA
L N C R T Y R C T A V N T Q V R D P T G -
* T V A L T D A Q P * T H R * E I P Q V -
K L S H L Q M H S R K H T G E R S H R C -
GCTATTGTATGTCATAAGCAAAATGCTAAGGTTGTCGTTAACATGTTTGGTTGTGTTGAA
661 ---------+---------+---------+---------+---------+---------+ 720
CGATAACATACAGTATTCGTTTTACGATTCCAACAGCAATTGTACAAACCAACACAACTT
A I V C H K Q N A K V V V N M F G C V E -
L L Y V I S K M L R L S L T C L V V L N -
Y C M S * A K C * G C R * H V W L C * T -
CTCATAGGGGAGAAACCCTATCAGTGTGACTTCACAGACTGTGGTCGCAGGTTCTCCAGA
721 ---------+---------+---------+---------+---------+---------+ 780
GAGTATCCCCTCTTTGGGATAGTCACACTGAAGTGTCTGACACCAGCGTCCAAGAGGTCT
L I G E K P Y Q C D F T D C G R R F S R -
S * G R N P I S V T S Q T V V A G S P D -
H R G E T L S V * L H R L W S Q V L Q I -
TCAGACCAGCTAAAACGACACCAGAGAAGACACACAGGTTTGCACCATTGTCCATTTTTG
781 ---------+---------+---------+---------+---------+---------+ 840
AGTCTGGTCGATTTTGCTGTGGTCTCTTCTGTGTGTCCAAACGTGGTAACAGGTAAAAAC
S D Q L K R H Q R R H T G L H H C P F L -
Q T S * N D T R E D T Q V C T I V H F C -
R P A K T T P E K T H R F A P L S I F V -
TATTTTCTTGTACTTCTTTATAACTGCTTGCTAAAGGACAACAAAATCAGTATAGAG
841 ---------+---------+---------+---------+---------+------- 897
ATAAAAGAACATGAAGAAATATTGACGAACGATTTCCTGTTGTTTTAGTCATATCTC
Y F L V L L Y N C L L K D N K I S I E -
I F L Y F F I T A C * R T T K S V * ? -
F S C T S L * L L A K G Q Q N Q Y R ? -

Unsequenced region ~2kb

```
    AAAAGGTGTGATAGCGGTCTGCACTGTTGACTTGTCAACAAAGTTCACTTTCACTCGCGT
  1 ---------+---------+---------+---------+---------+---------+ 60
    TTTTCCACACTATCGCCAGACGTGACAACTGAACAGTTGTTTCAAGTGAAAGTGAGCGCA

    K  R  C  D  S  G  L  H  C  *  L  V  N  K  V  H  F  H  S  R  -
     K  G  V  I  A  V  C  T  V  D  L  S  T  K  F  T  F  T  R  V  -
      K  V  *  *  R  S  A  L  L  T  C  Q  Q  S  S  L  S  L  A  F  -

    TCCAAGTTTCTGAATCCTCAATCAAACACATTTAGCGTGAGTGTTTGCTGATGTATGGCT
 61 ---------+---------+---------+---------+---------+---------+ 120
    AGGTTCAAAGACTTAGGAGTTAGTTTGTGTAAATCGCACTCACAAACGACTACATACCGA

    S  K  F  L  N  P  Q  S  N  T  F  S  V  S  V  C  *  C  M  A  -
     P  S  F  *  I  L  N  Q  T  H  L  A  *  V  F  A  D  V  W  L  -
      Q  V  S  E  S  S  I  K  H  I  *  R  E  C  L  L  M  Y  G  *  -

    GAACAGGCAGACTCTCTTGGTCTCACATACTAATTTTTTTTTTTTTTTTTTTTGTGCTGA
121 ---------+---------+---------+---------+---------+---------+ 180
    CTTGTCCGTCTGAGAGAACCAGAGTGTATGATTAAAAAAAAAAAAAAAAAAAACACGACT

    E  Q  A  D  S  L  G  L  T  Y  *  F  F  F  F  F  F  L  C  *  -
     N  R  Q  T  L  L  V  S  H  T  N  F  F  F  F  F  F  C  A  D  -
      T  G  R  L  S  W  S  H  I  L  I  F  F  F  F  F  F  V  L  T  -

    CACATTTGGCTAGGCGTTTCTGACCTGTGCCTCGTTCTGTTCTTGTTGTGTGCTTGTGTG
181 ---------+---------+---------+---------+---------+---------+ 240
    GTGTAAACCGATCCGCAAAGACTGGACACGGAGCAAGACAAGAACAACACACGAACACAC

    H  I  W  L  G  V  S  D  L  C  L  V  L  F  L  L  C  A  C  V  -
     T  F  G  *  A  F  L  T  C  A  S  F  C  S  C  C  V  L  V  C  -
      H  L  A  R  R  F  *  P  V  P  R  S  V  L  V  V  C  L  C  V  -

    TGTGCGACCACACTTGTGTGTTCCTGTTGTTGTGTGTGTGTAGGCGTTAAGCCATTCCAG
241 ---------+---------+---------+---------+---------+---------+ 300
    ACACGCTGGTGTGAACACACAAGGACAACAACACACACACATCCGCAATTCGGTAAGGTC

    C  A  T  T  L  V  C  S  C  C  C  V  C  V  G  V  K  P  F  Q  -
     V  R  P  H  L  C  V  P  V  V  V  C  V  *  A  L  S  H  S  S  -
      C  D  H  T  C  V  F  L  L  L  C  V  C  R  R  *  A  I  P  V  -

                                    PCR sequence      →
    TGTGAAACCTGTCAGAGAAAGTTTTCACGTTCAGACCACCTTAAGACCCACACCCGGACA
301 ---------+---------+---------+---------+---------+---------+ 360
    ACACTTTGGACAGTCTCTTTCAAAAGTGCAAGTCTGGTGGAATTCTGGGTGTGGGCCTGT

    C  E  T  C  Q  R  K  F  S  R  S  D  H  L  K  T  H  T  R  T  -
     V  K  P  V  R  E  S  F  H  V  Q  T  T  L  R  P  T  P  G  L  -
      *  N  L  S  E  K  V  F  T  F  R  P  P  *  D  P  H  Q  D  S  -

    CATACAGGTAAAACAAGTGCGTAAACCTTTTCATTTTTTTCATGATTCGCTCCTCTCTTT
361 ---------+---------+---------+---------+---------+---------+ 420
    GTATGTCCATTTTGTTCACGCATTTGGAAAAGTAAAAAAAGTACTAAGCGAGGAGAGAAA

    H  T  G  K  T  S  A  *  T  F  S  F  F  S  *  F  A  P  L  F  -
     I  Q  V  K  Q  V  R  K  P  F  H  F  F  H  D  S  L  L  S  F  -
      Y  R  *  N  K  C  V  N  L  F  I  F  F  M  I  R  S  S  L  S  -

    CCACTCTTTGATTTAAAGTTGAATTTTTACAGGTCACTTTTTTATCTCAGCTTGTACTGG
421 ---------+---------+---------+---------+---------+---------+ 480
```

```
    GGTGAGAAACTAAATTTCAACTTAAAAATGTCCAGTGAAAAAATAGAGTCGAACATGACC

    P  L  F  D  L  K  L  N  F  Y  R  S  L  F  Y  L  S  L  Y  W  -
     H  S  L  I  *  S  *  I  F  T  G  H  F  F  I  S  A  C  T  G  -
      T  L  *  F  K  V  E  F  L  Q  V  T  F  L  S  Q  L  V  L  G -

    GAAGTTATTAAAAAAATATTTGCAGTGGTTTTTTTATTGATAGCCACTGGTTTTAGTATT
481 ---------+---------+---------+---------+---------+---------+ 540
    CTTCAATAATTTTTTTATAAACGTCACCAAAAAAATAACTATCGGTGACCAAAATCATAA

    E  V  I  K  K  I  F  A  V  V  F  L  L  I  A  T  G  F  S  I  -
     K  L  L  K  K  Y  L  Q  W  F  F  Y  *  *  P  L  V  L  V  F  -
      S  Y  *  K  N  I  C  S  G  F  F  I  D  S  H  W  F  *  Y  S -

    CTCATTCAACTGTAA
541 ---------+----- 555
    GAGTAAGTTGACATT

    L  I  Q  L  *  -
     S  F  N  C  ?  -
      H  S  T  V  ? -
```

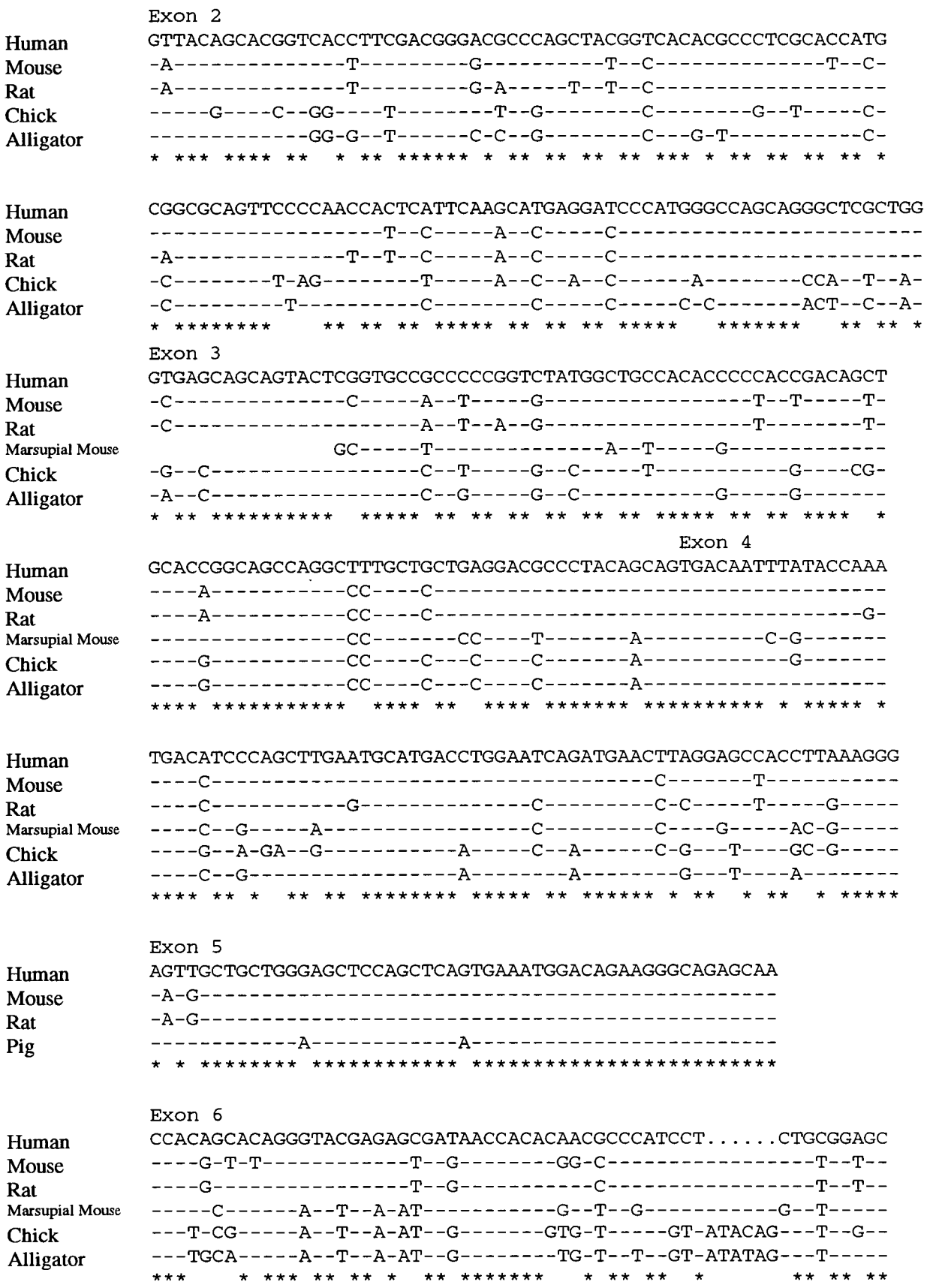
Zebrafish *WT1* predicted cDNA sequence

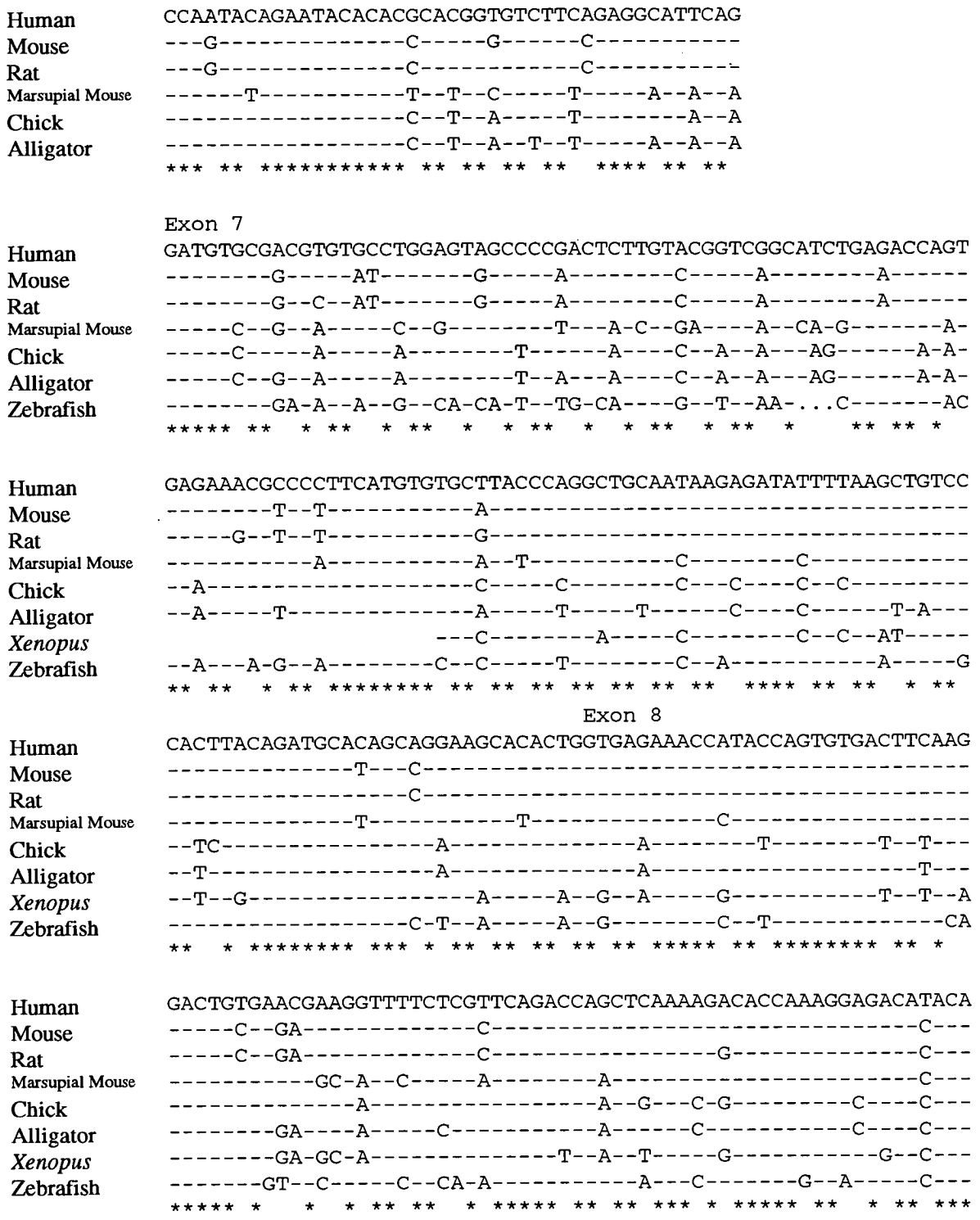
```
    GATGTGCGGAGAGTACCGGGCATCACTCCTGCCATTGTGCGTTCAACCGAGACCAACGAA
  1 ---------+---------+---------+---------+---------+---------+ 60
    CTACACGCCTCTCATGGCCCGTAGTGAGGACGGTAACACGCAAGTTGGCTCTGGTTGCTT

    D  V  R  R  V  P  G  I  T  P  A  I  V  R  S  T  E  T  N  E  -
                       ← PCR sequence
    AAAAGGCCATTCATGTGCGCCTACCCTGGCTGCAACAAAAGATATTTTAAACTGTCGCAC
 61 ---------+---------+---------+---------+---------+---------+ 120
    TTTTCCGGTAAGTACACGCGGATGGGACCGACGTTGTTTTCTATAAAATTTGACAGCGTG

    K  R  P  F  M  C  A  Y  P  G  C  N  K  R  Y  F  K  L  S  H  -

    TTACAGATGCACAGCCGTAAACACACAGGGGAGAAACCCTATCAGTGTGACTTCACAGAC
121 ---------+---------+---------+---------+---------+---------+ 180
    AATGTCTACGTGTCGGCATTTGTGTGTCCCCTCTTTGGGATAGTCACACTGAAGTGTCTG

    L  Q  M  H  S  R  K  H  T  G  E  K  P  Y  Q  C  D  F  T  D  -

    TGTGGTCGCAGGTTCTCCAGATCAGACCAGCTAAAACGACACCAGAGAAGACACACAGGC
181 ---------+---------+---------+---------+---------+---------+ 240
    ACACCAGCGTCCAAGAGGTCTAGTCTGGTCGATTTTGCTGTGGTCTCTTCTGTGTGTCCG

    C  G  R  R  F  S  R  S  D  Q  L  K  R  H  Q  R  R  H  T  G  -
                                                PCR sequence      →
    GTTAAGCCATTCCAGTGTGAAACCTGTCAGAGAAAGTTTTCACGTTCAGACCACCTTAAG
241 ---------+---------+---------+---------+---------+---------+ 300
    CAATTCGGTAAGGTCACACTTTGGACAGTCTCTTTCAAAAGTGCAAGTCTGGTGGAATTC

    V  K  P  F  Q  C  E  T  C  Q  R  K  F  S  R  S  D  H  L  K  -

    ACCCACACCCGGACACATACAGGTAAAACAA
301 ---------+---------+---------+- 331
    TGGGTGTGGGCCTGTGTATGTCCATTTTGTT

    T  H  T  R  T  H  T  G  K  T  -
```

Alignment of the nucleotide sequence of the coding region of the *WT1* orthologs

- indicates no sequence change
. indicates a deletion/insertion
* indicates a conserved position in all species sequenced

```
                Exon 1
Human           ATGGGCTCCGACGTGCGGGACCTGAACGCGCTGCTGCCCGCCGTCCCCTCCCTGGGTGGC
Mouse           -----T-----------------------------------T--GT-T--G-----C---
Rat             -----T-------------------------------------GT----G-----C--T
Chick           -----G--------C------------------------T----G--------AC...CG
                ***** ******** ************************ * **  * ** **

                Glycine 5 encoding region
Human           GG...CGGCGGCTGTGCCCTGCCTGTGAGCGGCGCGGCGCAGTGGGCGCCGGTGCTGGAC
Mouse           --CGG---------C-GG--C--------------ACG------------C---T-----
Rat             -----A----------GG--C---------------CG---A------------------
Chick           --CAA-A--AA---C---A----G------A----------------T--C--C------
                **     **  *** *   * ** ****** ****   *** ***** ** **  *****

                                                                Proline 13
Human           TTTGCGCCCCCGGGCGCTTCGGCTTACGGGTCGTTGGGCGGCCCCGCGCCGCCACCGGCT
Mouse           --C-----T--------C---------------C-------T--------T--T--C---
Rat             --C-----T--------C--------------AC-------T--------T--T--C---
Chick           -C-...---------G--C--...C-----C---C----.....................
                *     ** ***** ** **    ***** **  ****

                encoding region
Human           CCGCCGCCACCCCCGCCGCCGCCGCCTCACTCCTTCATCAAACAGGAGCCGAGCTGGGGC
Mouse           --------G--T--------A--...C----------------------C---------
Rat             -----------T--------A--...C----------------------C---------
Chick           ......................---G--------------G-----A--C------AA-
                                       ************** ***** ** ******  *

Human           GGCGCGGAGCCGCACGAGGAGCAGTGCCTGAGCGCCTTCACTGTCCACTTTTCCGGCCAG
Mouse           -----C-----A-----------------------------CT-G-----C--G------
Rat             -----C-----------------------------------CT-G-----C---------
Chick           --GT----C----------------A---------T-----C--------C-----G---
                **  * ** ** ************* ********* *****  * ***** ** ** ***

Human           TTCACTGGCACAGCCGGAGCCTGTCGCTACGGGCCCTTCGGTCCTCCTCCGCCCAGCCAG
Mouse           -----C--T--------G--------------A--------------C------------
Rat             -----C--T--------G--------------A--------------C------------
Chick           -----G-----G-----G-----C-----------------CG-C--G------------
                ***** ** ** ***** ***** ******** ********  * ** ************

Human           GCGTCATCCGGCCAGGCCAGGATGTTTCCTAACGCGCCCTACCTGCCCAGCTGCCTCGAG
Mouse           -----C--G----------------C--C--T----------------------G---
Rat             -------------------------C--C--T----------------------G---
Chick           C--C-C--------------------C--CC-ACG--------------A------G---
                 ** * ** ***************** **  *   ************** ****** ***

Human           AGCCAGCCCGCTATTCGCAATCAGG
Mouse           --------TA-C--C-----C--A-
Rat             ---------T-C--C-----C--A-
Chick           -------AA--C--------C----
                *******   * ** ***** ** *
```


Exon 2
Human GTTACAGCACGGTCACCTTCGACGGGACGCCCAGCTACGGTCACACGCCCTCGCACCATG
Mouse -A----------T---------G----------T--C----------T--C-
Rat -A----------T---------G-A-----T--T--C----------
Chick -----G----C--GG----T---------T--G---------C---------G--T-----C-
Alligator ----------GG-G--T------C-C--G---------C---G-T----------C-
* *** **** ** * ** ****** * ** ** ** ** *** * ** ** ** ** *
Human CGGCGCAGTTCCCCAACCACTCATTCAAGCATGAGGATCCCATGGGCCAGCAGGGCTCGCTGG
Mouse ----------T--C-----A--C-----C----------
Rat -A----------T--T--C-----A--C-----C----------
Chick -C---------T-AG--------T-----A--C--A--C------A---------CCA--T--A-
Alligator -C---------T----------C---------C-----C-----C-C-------ACT--C--A-
* ******** ** ** ** ***** ** ** ** ***** ******* ** ** *
Exon 3
Human GTGAGCAGCAGTACTCGGTGCCGCCCCCGGTCTATGGCTGCCACACCCCCACCGACAGCT
Mouse -C----------C-----A--T-----G----------T--T-----T-
Rat -C----------A--T--A--G----------T--------T-
Marsupial Mouse GC-----T----------A--T-----G----------
Chick -G--C----------C--T-----G--C-----T----------G----CG-
Alligator -A--C----------C--G-----G--C----------G-----G-------
* ** ********** ***** ** ** ** ** ** ** ***** ** ** **** *
Exon 4
Human GCACCGGCAGCCAGGCTTTGCTGCTGAGGACGCCCTACAGCAGTGACAATTTATACCAAA
Mouse ----A----------CC----C----------
Rat ----A----------CC----C----------G-
Marsupial Mouse ----------CC-------CC----T-------A----------C-G-------
Chick ----G----------CC----C---C----C-------A----------G-------
Alligator ----G----------CC----C---C----C-------A----------
**** ********** **** ** **** ******* ********** * ***** *
Human TGACATCCCAGCTTGAATGCATGACCTGGAATCAGATGAACTTAGGAGCCACCTTAAAGGG
Mouse ----C----------C-------T----------
Rat ----C----------G----------C---------C-C-----T-----G-----
Marsupial Mouse ----C--G-----A----------C---------C----G-----AC-G-----
Chick ----G--A-GA--G----------A-----C--A------C-G---T----GC-G-----
Alligator ----C--G----------A--------A--------G---T----A--------
**** ** * ** ** ******** ***** ** ****** * ** * ** * *****
Exon 5
Human AGTTGCTGCTGGGAGCTCCAGCTCAGTGAAATGGACAGAAGGGCAGAGCAA
Mouse -A-G----------
Rat -A-G----------
Pig ----------A----------A----------
* * ******** ********** **********
Exon 6
Human CCACAGCACAGGGTACGAGAGCGATAACCACACAACGCCCATCCT.....CTGCGGAGC
Mouse ----G-T-T----------T--G--------GG-C----------T--T--
Rat ----G----------T--G----------C----------T--T--
Marsupial Mouse -----C------A--T--A-AT----------G--T--G----------G--T-----
Chick ---T-CG-----A--T--A-AT--G-------GTG-T-----GT-ATACAG---T--G--
Alligator ---TGCA-----A--T--A-AT--G--------TG-T--T--GT-ATATAG---T-----
*** * *** ** ** * ** ******* * ** ** * ** ** **

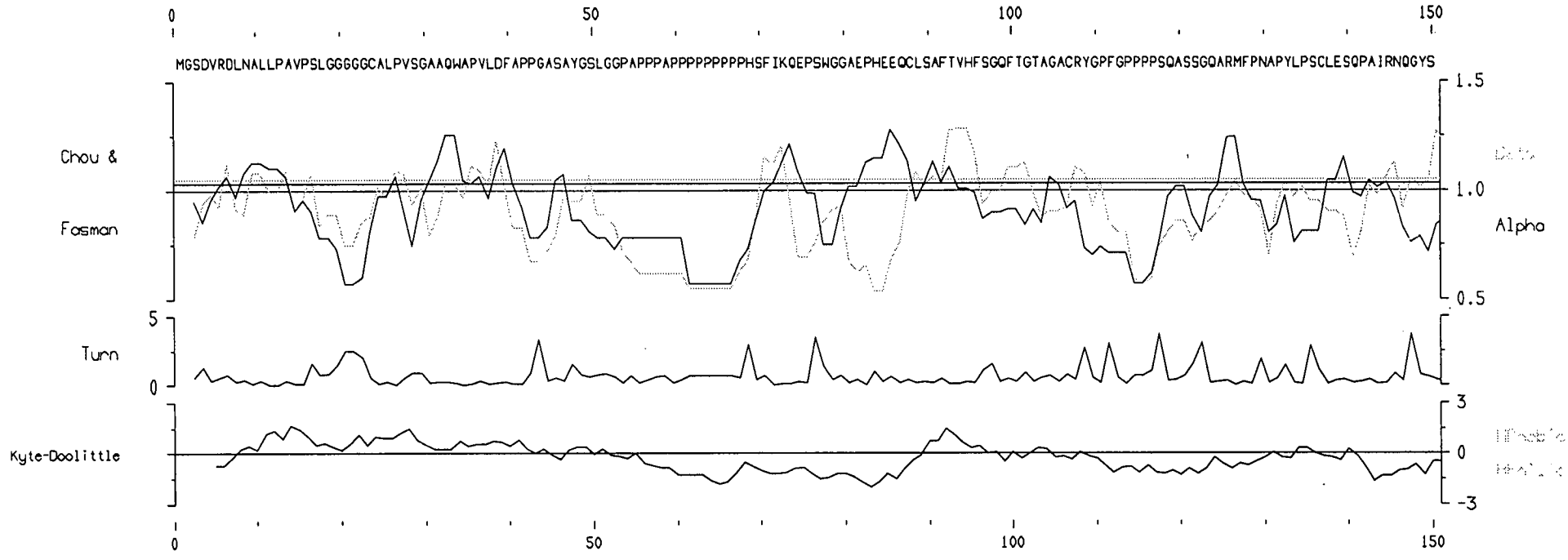

Human
Mouse
Rat
Marsupial Mouse
Chick
Alligator
Exon 7
Zebrafish
Xenopus
Exon 8

```
                Exon 9
Human           GGTGTGAAACCATTCCAGTGTAAAACTTGTCAGCGAAAGTTCTCCCGGTCCGACCACCTG
Mouse           ----------------------------------------T-------------T---
Rat             ----------------------------------------T----------------
Marsupial Mouse ------------------------C------A---------------T---------
Chick           ----------C-----A--------C------A-----------A-A--T--T--T---
Alligator       ----------C--------------C------A----------TA----T--T--T--A
Xenopus         ---A-C-----T-----A--C-----C--C--A-----------A----------T---
Zebrafish       --C--T--G-----------G----C------A-------T--A--T--A--------T
                **  * ** ** ***** **  **** ** **  ******* **  * ** ** ** **

                                    ---KTS---
Human           AAGACCCACACCAGGACTCATACAGGTAAAACAA
Mouse           ----------------------------------
Rat             ----------------------------------
Marsupial Mouse -----A----------------------------
Chick           -----T--T------------------------
Alligator       -----T----------------------------
Xenopus         -----T------------------
Zebrafish       ------------C----A----------------
                ***** ** **  **** ****************

                Exon 10
Human           GTGAAAAGCCCTTCAGCTGTCGGTGGCCAAGTTGTCAGAAAAAGTTTGCCCGGTCAGATG
Mouse           ---------------------------AC-------------------G--C-----C-
Rat             ---------------------------TAC------------------G--------C-
Marsupial Mouse -------------------C--------C--------------A--------C-------
Alligator       -------A--A-----------A-----C--C-----A-----A------A-A--T----
                ******* ** ******** ** ***   ** ***** ****  *****  * ** ** *

Human           AATTAGTCCGCCATCACAACATGCATCAGAGAAACATGACCAAACTCCAGCTGGCGCTTTGA
Mouse           ----G-------------------------------------------------------
Rat             -------------------G-----C-------------------------------------
Marsupial Mouse -------G--T--C-----------C--A--G--------G--G---------A------A-
Alligator       -------T--T--------------C-----G--------T-----G---T----C------
                **** ** ** ** ***** ***** ** ** ******** ** ** *** ** * **** *
```

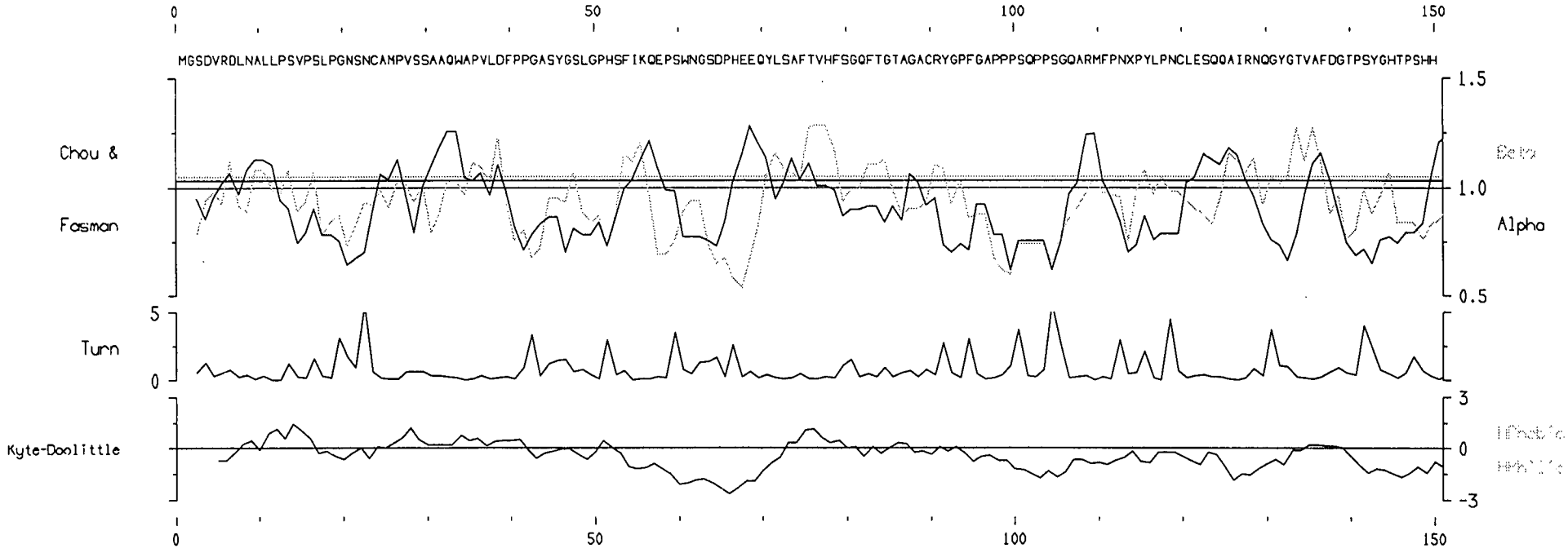
Appendix B
Structure prediction plots for *WT1* orthologs

Chou and Fasman prediction plots of α helix, β strand, and β turns were produced for *WT1* orthologs. Hydrophobicity profile using Kyte and Doolittle algorithms also were generated using pepplot (GCG). The plots for human and the most diverged sequences are shown for exon 1, exons 2-6 and exons 7-10.

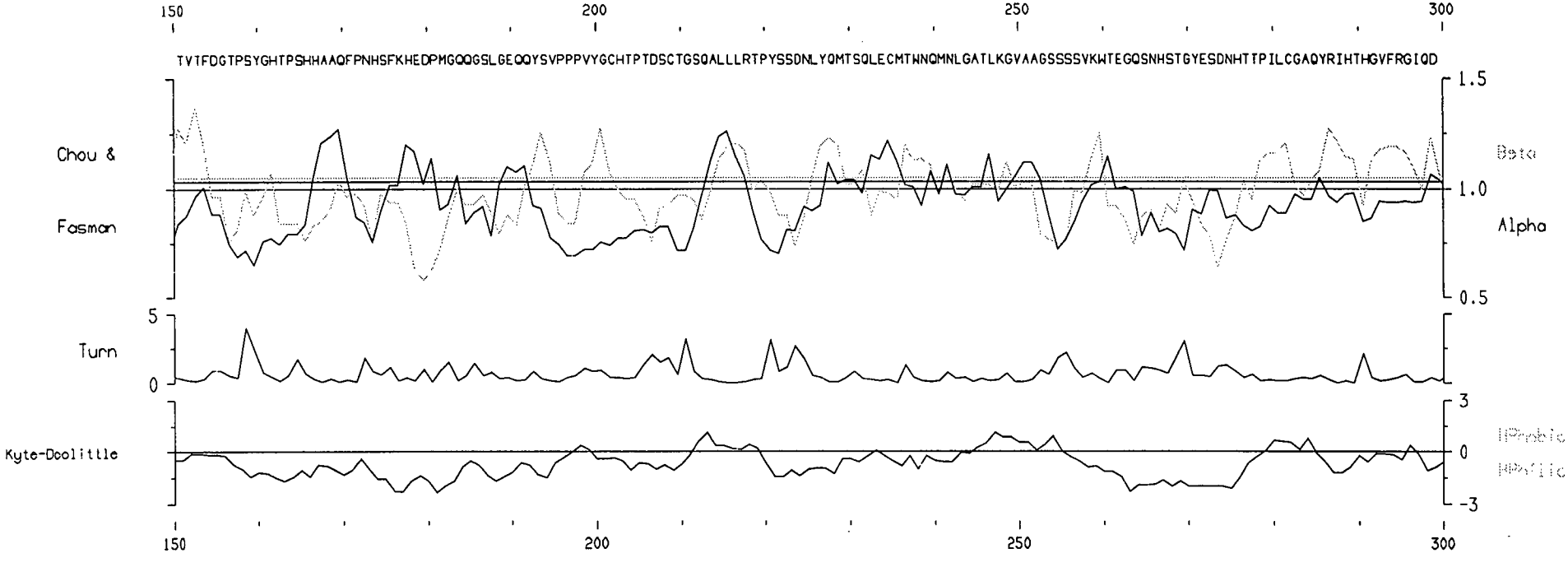
Structure Prediction Profile for Human WT1, Exon 1



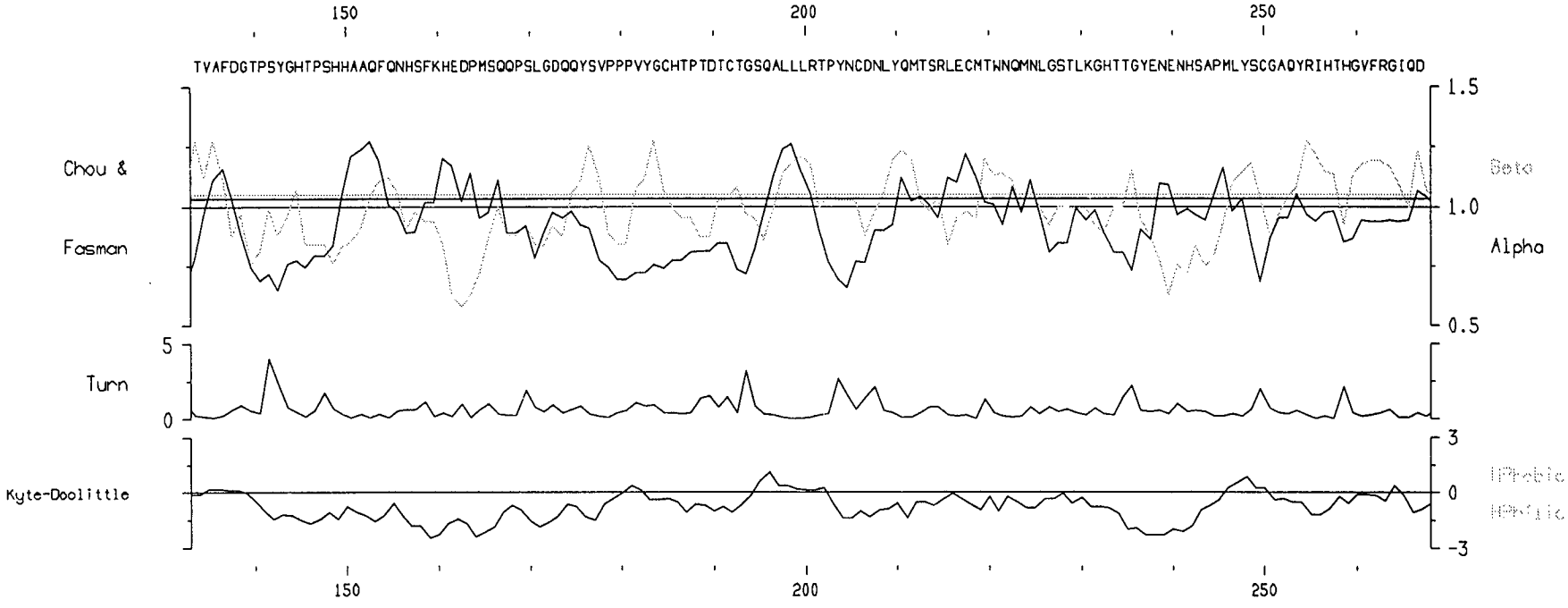
Structure Prediction Profile for Chick WT1, Exon 1



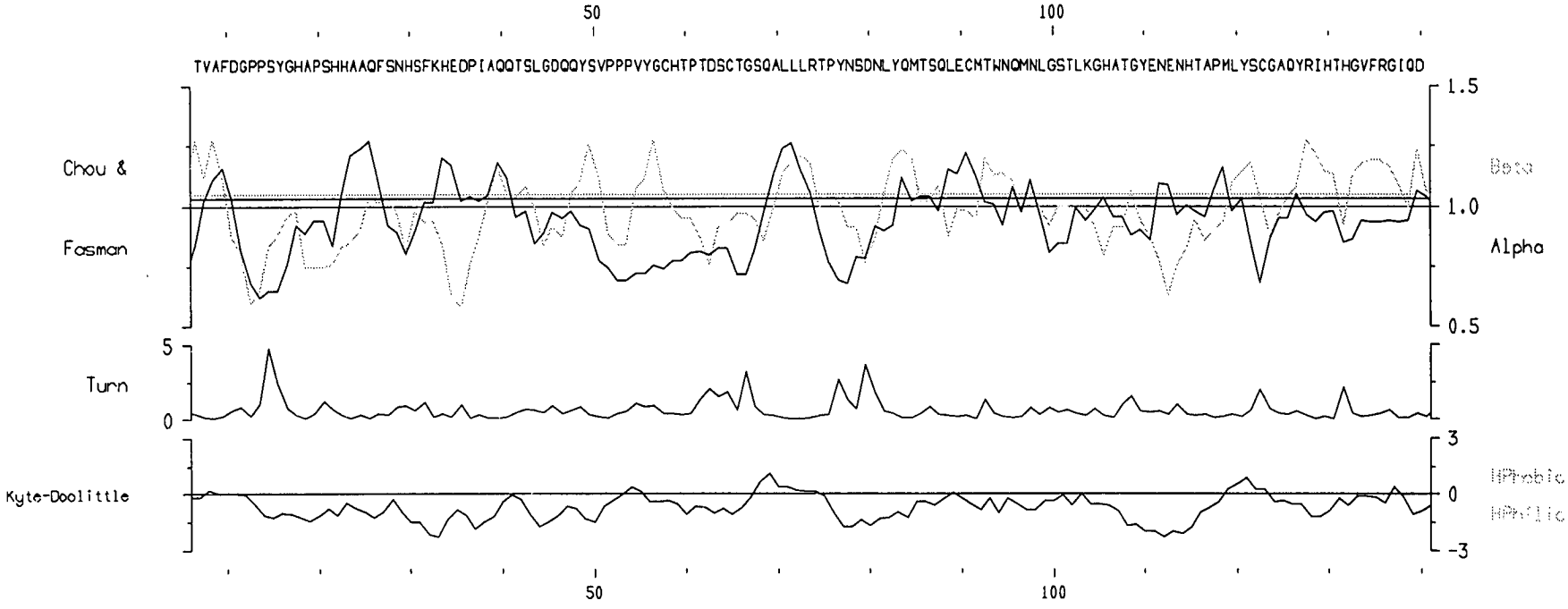
Structure Prediction Profile for Human WT1, Exons 2-6



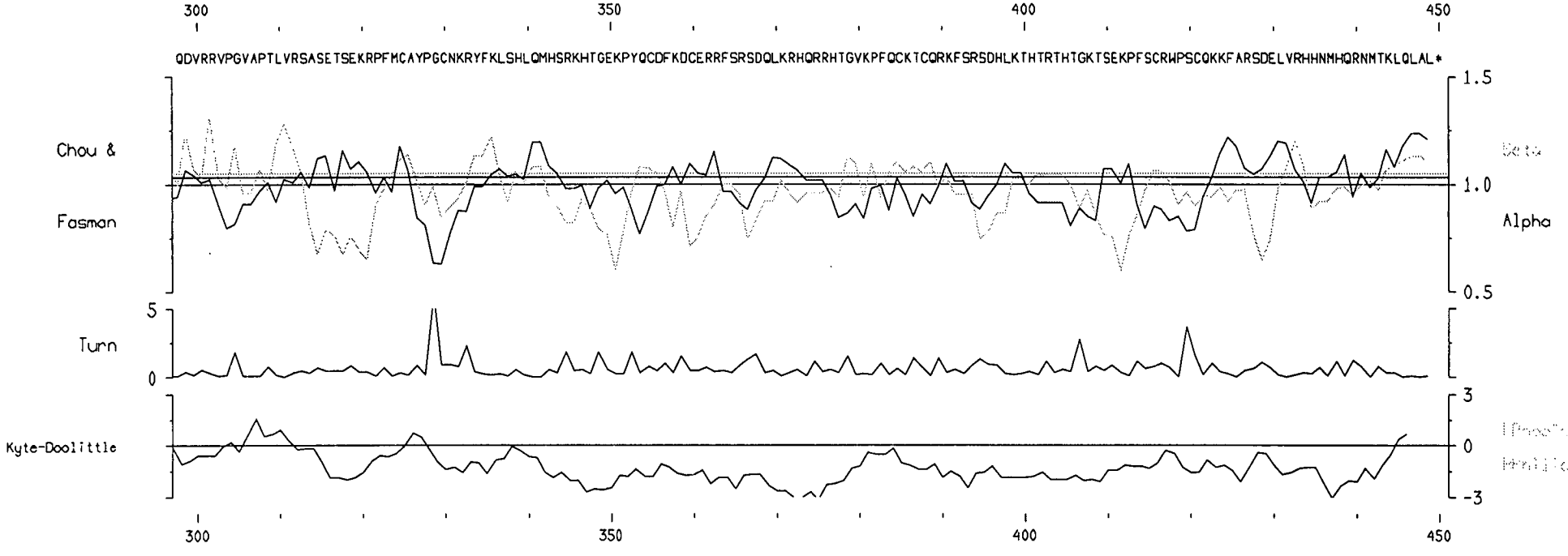
Structure Prediction Profile for Chick WT1, Exons 2-6



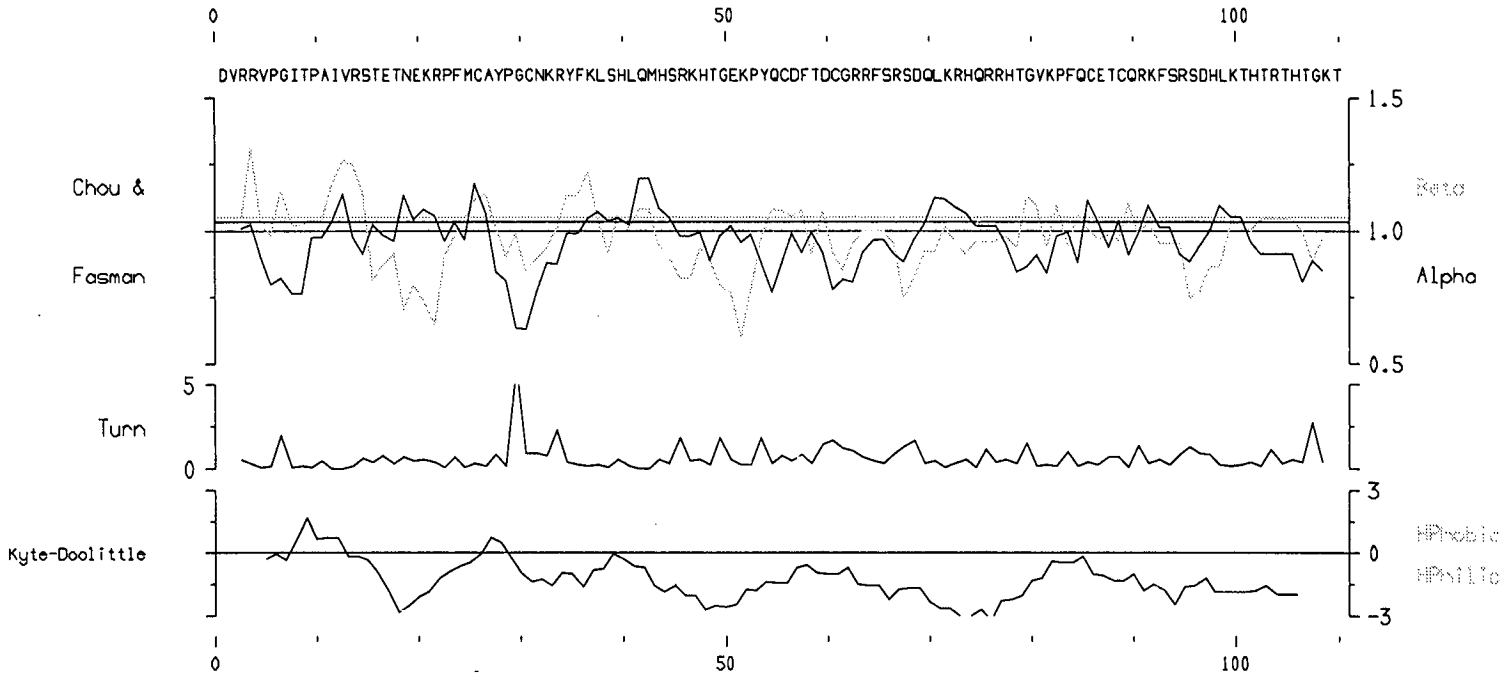
Structure Prediction Profile for Alligator WT1, Exons 2-6



Structure Prediction Profile for Human WT1, Exons 7-10



Structure Prediction Profile for Zebrafish WT1, Exons 7-9



Bibliography

Adzhubei A. A. and Sternberg M. J. (1993) Left-handed polyproline II helices commonly occur in globular proteins. *J.Mol.Biol.* **229**, 472-493.

Akasaka Y., Kikuchi H., Nagai T., Hiraoka N., Kato S., and Hata J. (1993) A point mutation found in the *WT1* gene in a sporadic Wilms' tumor without genitourinary abnormalities is identical with the most frequent point mutation in Denys-Drash syndrome. *FEBS Lett* **317**, 39-43.

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

Armstrong J. F., Pritchard-Jones K., Bickmore W. A., Hastie N. D., and Bard J. B. L. (1992) The expression of the Wilms' tumour gene, *WT1*, in the developing mammalian embryo. *Mech. Dev.* **40**, 85-97.

Arnheim N. and Erlich H. (1992) Polymerase chain reaction strategy. *Annu.Rev.Biochem.* **61**, 131-156.

Aterman K. (1989) Extrarenal nephroblastomas. *Journal of Cancer Research and Clinical Oncology* **115**, 409-417.

Backus J. W. and Smith H. C. (1991) Apolipoprotein B mRNA sequences 3' of the editing site are necessary and sufficient for editing and editosome assembly. *Nucleic Acids Res.* **19**, 6781-6786.

Baird P. N. and Cowell J. K. (1993) A novel zinc finger mutation in a patient with Denys-Drash syndrome. *Hum. Mol. Genet.* **2**, 2193-2194.

Baird P. N., Groves N., Haber D. A., Housman D. E., and Cowell J. K. (1992a) Identification of mutations in the *WT1* gene in tumours from patients with the WAGR syndrome. *Oncogene* **7**, 2141-2149.

Baird P. N., Santos A., Groves N., Jadresic L., and Cowell J. K. (1992b) Constitutional mutations in the *WT1* gene in patients with Denys-Drash syndrome. *Hum. Mol. Genet.* **1**, 301-305.

Balinsky B. I. (1981) *An Introduction to Embryology, 2nd ed.* Saunders, London.

el-Baradi T. and Pieler T. (1991) Zinc finger proteins: what we know and what we would like to know. *Mech. Dev.* **35**, 155-169.

Bard J. B. L. and McConnell J. E. (1994) Towards a genetic basis for kidney development. *Mech. Dev.* **In press**.

Bard J. B. L. and Woolf A. S. (1992) Nephrogenesis and the development of renal disease. *Nephrol.Dial.Transplant* **7**, 563-572.

Bardseey N., Falkoff D., Petruzzi M.-J., Nowak N., Zabel B., Adam M., Aguiar M. C., Grundy P., Shows T., and Pelletier J. (1994) Anaplastic Wilms' tumour, a subtype displaying poor prognosis, harbours p53 gene mutations. *Nat.Genet.* **7**, 91-97.

Barnes R. S. K., Calow P., and Olive P. J. W. (1993) *The Invertebrates, 2nd ed.* Blackwell, Oxford.

Beckwith J. B. (1969) Macroglossa, omphalecle, adrenal cytomegly, gigantism,and hyperplastic visceromegaly. *Birth Defects* **5**, 188-196.

Beckwith J. B. (1983) Wilms' tumours and other renal tumours of childhood: a selective review from the National Wilms' Tumour Study pathology centre. *Hum. Path.* **14**, 481-492.

Beckwith J. B., Kiviat N. B., and Bonadio J. F. (1990) Nephrogenic Rests, Nephroblastomatosis, And The Pathogenesis Of Wilms' Tumor. *Pediatr. Pathol.* **10**, 1-36.

Behlke M. A., Bogan J. S., Beer-Romero P., and Page D. C. (1993) Evidence that the *SRY* protein is encoded by a single exon on the human Y chromosome. *Genomics* **17**, 736-739.

Behringer R. R., Cate R. L., Froelick G. J., Palmiter R. D., and Brinster R. L. (1990) Abnormal sexual development in transgenic mice chronically expressing mullerian inhibiting substance. *Nature* **345**, 167-170.

Bennington J. L. and Beckwith J. B. (1975) *Atlas of Tumour Pathology, Second Series: Tumours of the kidney, renal pelvis and ureter.* Armed Forces Institute of Pathology, Washington D.C.

Benton M. J. (1990) Phylogeny of the major tetrapod groups: morphological data and divergence dates. *J.Mol.Evol.* **30**, 409-424.

Bickmore W. A., Porteous D. J., Christie S., Seawright A., Fletcher J. M., Maule J. C., Couillin P., Junien C., Hastie N. D., and van Heyningen V. (1989) CpG islands surround a DNA segment located between translocation breakpoints associated with genitourinary dysplasia and aniridia. *Genomics* **5**, 685-693.

Bickmore W. A., Oghene K., Little M. H., Seawright A., van Heyningen V., and Hastie N. D. (1992) Modulation of DNA Binding Specificity by Alternative Splicing of the Wilms Tumor *WT1* Gene Transcript. *Science* **257**, 235-237.

Bierkamp C. and Campos-Ortega J. A. (1993) A zebrafish homologue of the *Drosophila* neurogenic gene *Notch* and its pattern of transcription during early embryogenesis. *Mech. Dev.* **43**, 87-100.

Bird A. P. (1986) CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213.

Bogan J. and Page D. C. (1994) Ovary? Testis? - a mammalian dilemma. *Cell* **76**, 603-607.

Bonetta L., Kuehn S. E., Huang A., Law D. J., Kalikin L. M., Koi M., Reeve A. E., Brownstein B. H., Yeger H., Williams B. R. G., and Feinberg A. P. (1990) Wilms Tumor Locus on 11p13 Defined by Multiple CpG Island-Associated Transcripts. *Science* **250**, 994-997.

Bove K. E. and McAdams A. J. (1976) The nephroblastomatosis complex and its relationship to Wilms' tumour: a clinicopathologic treatise. *Perspectives in Pediatric Pathology* **3**, 185-223.

Boyd A. C. (1993) Turbo cloning: a fast, efficient method for cloning PCR products and other blunt-ended DNA fragments into plasmids. *Nucleic Acids Res.* **21**, 817-821.

Brenner B., Wildhardt G., Schneider S., and Royer-Pokora B. (1992) RNA polymerase chain reaction detects different levels of four alternatively spliced *WT1* transcripts in Wilms' tumors. *Oncogene* **7**, 1431-1433.

Brenner S., Elgar G., Sandford R., Macrae A., Venkatesh B., and Aparicio S. (1993) Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* **366**, 265-268.

Breslow N. and Beckwith J. B. (1982) Epidemiological features of Wilms' tumour: results of the National Wilms' Tumour Study. *J.Nat.Cancer Inst.* **68**, 429-436.

Breslow N., Beckwith B. J., Ciol M., and Sharples K. (1988) Age Distribution of Wilms' Tumor: Report from the National Wilms' Tumor Study. *Cancer Res.* **48**, 1653-1657.

Britten R. J. (1986) Rates of DNA sequence evolution differ between taxonomic groups. *Science* **231**, 1393-1398.

Brook J. D., McCurrach M. E., Harley H. G., Buckler A. J., Church D., Aburatani H., Hunter K., Stanton V. P., Thirion J. P., Hudson T., Sohn R., Zemelman B., Snell R. G., Rundle S. A., Crow S., Davies J., Shelbourne P., Buxton J., Jones C., Juvonen V., Johnson K., Harper P. S., Shaw D. J., and Houseman D. E. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**, 799-808.

Brown K. W., Watson J. E., Poirier V., Mott M. G., Berry P. J., and Maitland N. J. (1992) Inactivation of the remaining allele of the *WT1* gene in a Wilms' tumour from a WAGR patient. *Oncogene* **7**, 763-768.

Bruening W., Gros P., Sato T., Stanimir J., Nakamura Y., Housman D., and Pelletier J. (1993) Analysis of the 11p13 Wilms' tumor suppressor gene (*WT1*) in ovarian tumors. *Cancer Invest.* **11**, 393-399.

Bruening W., Bardeesy N., Silverman B. L., Cohn R. A., Machin G. A., Aronson A. J., Housman D., and Pelletier J. (1992) Germline intronic and exonic mutations in the Wilms' tumour gene (*WT1*) affecting urogenital development. *Nat.Genet.* **1**, 144-148.

Buckler A. J., Pelletier J., Haber D. A., Glaser T., and Housman D. E. (1991) Isolation, characterization, and expression of the murine Wilms' tumor gene (*WT1*) during kidney development. *Mol.Cell. Biol.* **11**, 1707-1712.

Buehr M., Gu S., and McLaren A. (1993) Mesonephric contribution to testis differentiation in the fetal mouse. *Development* **117**, 273-281.

Burley S. K. (1994) Forward to the fundamentals. *Nat. Struct. Biol.* **1**, 8-10.

Byskov A. G. (1986) Differentiation of the mammalian embryonic gonad. *Physiological Reviews* **66**, 71-117.

Call K. M., Glaser T., Ito C. Y., Buckler A. J., Pelletier J., Haber D. A., Rose E. A., Kral A., Yeger H., Lewis W. H., Jones C., and Housman D. E. (1990) Isolation and Characterization of a Zinc Finger Polypeptide Gene at the Human Chromosome 11 Wilms' Tumor Locus. *Cell* **60**, 509-520.

Campbell C. E., Huang A., Gurney A. L., Kessler P. M., Hewitt J. A., and Williams B. R. G. (1994) Antisense transcripts and protein binding motifs within the Wilms tumour (*WT1*) locus. *Oncogene* **9**, 583-595.

Cavenee W. K., Drya T. P., Phillips R. A., Benedict W. F., Godbout R., Gallie B. L., Murphree A. L., Strong L. C., and White R. (1983) Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**, 779-784.

Cavener D. R. and Ray S. C. (1991) Eukaryotic start and stop translation sites. *Nucleic Acids Res.* **19**, 3185-3192.

Chan L. (1993) RNA editing: exploring one mode with apolipoprotein B mRNA. *Bioessays* **15**, 33-41.

Chao Lian-Y., Huff V., Tomlinson G., Riccardi V. M., Strong L. C., and Saunders G. F. (1993) Genetic mosaicism in normal tissues of Wilms' tumour patients. *Nat.Genet.* **3**, 127-131.

Charlesworth B. (1993) More mutations in males? *Curr. Biol* **3**, 466-467.

Chi T. and Carey M. (1993) The ZEBRA activation domain: Modular organisation and mechanism of action. *Mol.Cell. Biol.* **13**, 7045-7055.

Chisholm A. and Horvitz H. R. (1994) Homeobox and paired box genes in *C.elegans* pattern formation. Abstract, British Society for Developmental Biology Spring Meeting, Evolution of Development.

Choo Y. and Klug A. (1993) A role in DNA binding for the linker sequences of the first three zinc fingers of *TFIIIA*. *Nucleic Acids Res.* **21**, 3341-3346.

Chou P. Y. and Fasman G. D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**, 45-148.

Christian J. L. and Moon R. T. (1993) Interactions between *Xwnt-8* and Spemann organizer signaling pathways generate dorsoventral pattern in the embryonic mesoderm of *Xenopus*. *Genes Dev.* **7**, 13-28.

Chung M. Y., Ranum L. P., Duvick L. A., Servadio A., Zoghbi H. Y., and Orr H. T. (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat.Genet.* **5**, 254-258.

Clark J. M. (1988) Novel non-templated nucleotide addition reactions catalysed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* **16**, 9677-9686.

Clarkson P. A., Davies H. R., Williams D. M., Chaudhary R., Hughes I. A., and Patterson M. N. (1993) Mutational screening of the Wilms's tumour gene, *WT1*, in males with genital abnormalities. *J.Med.Genet.* **30**, 767-772.

Clarkson T., Gussow D., and Jones P. T. (1991) General application of PCR to gene cloning and manipulation, in *PCR: A Practical Approach* (McPherson M. J., Quirke P., and Taylor G. R., eds) IRL Press, Oxford.

Coles H. S. R., Burne J. F., and Raff M. C. (1993) Large-scale normal cell death in the developing rat kidney and its reduction by epidermal growth factor. *Development* **118**, 777-784.

Compton D. A., Weil M. M., Bonetta L., Huang A., Jones C., Yeger H., Williams B. R., Strong L. C., and Saunders G. F. (1990) Definition of the limits of the Wilms tumor locus on human chromosome 11p13. *Genomics* **6**, 309-315.

Conway J. F. and Parry D. A. D. (1990) Structural features in the heptad substructure and longer range repeats of two-stranded alpha-fibrous proteins. *Int.J.Biol.Macromol.* **12**, 328-334.

Coppes M. J., Bonetta L., Huang A., Hoban P., Chilton-MacNeill S., Campbell C. E., Weksberg R., Yeger H., Reeve A. E., and Williams B. R. G. (1992a) Loss of Heterozygosity Mapping in Wilms Tumor Indicated the Involvement of Three Distinct Regions and a Limited Role for Nondisjunction or Mitotic Recombination. *Genes Chromosom.Cancer* **5**, 326-334.

Coppes M. J., Liefers G. J., Higuchi M., Zinn A. B., Balfe J. W., and Williams B. R. G. (1992b) Inherited *WT1* Mutation in Denys-Drash Syndrome. *Cancer Res.* **52**, 6125-6128.

Coppes M. J., Liefers G. J., Paul P., Yeger H., and Williams B. R. G. (1993a) Homozygous somatic *WT1* point mutations in sporadic unilateral Wilms tumor. *Proc.Natl.Acad.Sci.USA* **90**, 1416-1419.

Coppes M. J., Ye Y., Rackley R., Zhao X. L., Liefers G. J., Casey G., and Williams B. R. (1993b) Analysis of *WT1* in granulosa cell and other sex cord-stromal tumors. *Cancer Res.* **53**, 2712-2714.

Coriat A. M., Muller U., Harry J. L., Uwanogho D., and Sharpe P. T. (1993) PCR amplification of SRY-related gene sequences reveals evolutionary conservation of the *SRY*-box motif. *PCR Methods Appl.* **2**, 218-222.

Cowell J. K., Wadey R. B., Haber D. A., Call K. M., Housman D. E., and Pritchard J. (1991) Structural rearrangements of the *WT1* gene in Wilms' tumour cells. *Oncogene* **6**, 595-599.

Cowell J. K., Groves N., and Baird P. (1993) Loss of heterozygosity at 11p13 in Wilms' tumours does not necessarily involve mutations in the *WT1* gene. *Br.J.Cancer* **67**, 1259-1261.

Creighton T. E. (1984) *Proteins: Structural and Molecular Properties*. W.H. Freeman and Company, New York.

Crosby S. D., Veile R. A., Donis-Keller H., Baraban J. M., Bhat R. V., Simburger K. S., and Milbrandt J. (1992) Neural-specific expression, genomic structure, and chromosomal localization of the gene encoding the zinc-finger transcription factor *NGFI-C. Proc.Natl.Acad.Sci.USA* **89**, 4739-4743.

Crossley P. H. and Little P. F. R. (1991) A cluster of related zinc finger protein genes is deleted in the mouse embryonic lethal mutation t^{w18}. *Proc.Natl.Acad.Sci.USA* **88**, 7923-7927.

Dale L. and Slack J. M. W. (1987) Regional specification within the mesoderm of early embryos of *Xenopus laevis*. *Development* **100**, 279-295.

Dale L., Howes G., Price B. M. J., and Smith J. C. (1992) Bone morphogenetic protein 4: A ventralizing factor in early *Xenopus* development. *Development* **115**, 573-585.

Dawkins R. (1994) The eye in a twinkling. *Nature* **368**, 690-691.

Dayhoff M. O., Eck R. V., and Park C. M. (1972) A model of evolutionary change in proteins, in *Atlas of protein sequence and structure Vol. 5* (Dayhoff M. O., ed), pp. 89-99. National Biomedical Research Foundation, Silver Springs, Md.

DeChiara T. M., Robertson E. J., and Efstratiadis A. (1991) Parental Imprinting of the Mouse Insulin-like Growth Factor II Gene. *Cell* **64**, 849-859.

Del-Rio S., Menezes S. R., and Setzer D. R. (1993) The function of individual zinc fingers in sequence-specific DNA recognition by transcription factor IIIA. *J.Mol.Biol.* **233**, 567-579.

Del Rio-Tsonis K., Washabaugh C. H., and Tsonis P. A. (1992) The mutant axolotl Short toes exhibits impaired limb regeneration and abnormal basement membrane formation. *Proc.Natl.Acad.Sci.USA* **89**, 5502-5506.

Denys P., Malvaux P., van den Berghe H., Tanghe W., and Proesmans W. (1967) Association d'un syndrome anatomo-pathologique de pseudohermaphrodisme masculin, d'une tumeur de Wilms, d'une nephropathie parenchymateuse et d'un mosaicisme XX/XY. *Arch. Franc. Ped.* **24**, 729-739.

Deuel T. F. (1987) Polypeptide growth factors: roles in normal and abnormal cell growth. *Annu.Rev.Cell Biol.* **3**, 443-492.

Doolittle R. F. (1981) Similar amino acid sequences: chance or common ancestry. *Science* **274**, 149-159.

Doolittle R. F. (1986) *Of UFS and ORFS. A primer on how to analyze derived amino acid sequences*. University Science Books, Mill Valley, CA.

Doolittle R. F. (1994) Convergent evolution: the need to be explicit. *Trends Biochem. Sci.* **19**, 15-19.

Dournon C., Houillon C., and Pieau C. (1990) Temperature sex-reversal in amphibians and reptiles. *Int.J.Dev.Biol.* **34**, 81-92.

Downie S. A. and Newman S. A. (1994) Morphogenetic differences between fore and hind limb precartilage mesenchyme: relation to mechanisms of skeletal pattern formation. *Dev.Biol.* **162**, 195-208.

Dowzer C. E. A. and Kelly J. M. (1991) Analysis of the creA gene, a regulator of carbon catabolite repression in Aspergillus nidulans. *Mol.Cell. Biol.* **11**, 5701-5709.

Drash A., Sherman F., Hartmann W. H., and Blizzard R. M. (1970) A syndrome of pseudohermaphroditism, Wilms' tumor, hypertension, and degenerative renal disease. *J.Pediat* **76**, 585-593.

Dressler G. R. and Douglass E. C. (1992) *Pax-2* is a DNA-binding protein expressed in embryonic kidney and Wilms tumor. *Proc.Natl.Acad.Sci.USA* **89**, 1179-1183.

Dressler G. R., Deutsch U., Chowdhury K., Nornes H. O., and Gruss P. (1990) *Pax2*, a new murine paired-box-containing gene and its expression in the developing excretory system. *Development* **109**, 787-795.

Dressler G. R., Wilkinson J. E., Rothenpieler U. W., Patterson L. T., Williams-Simons L., and Westphal H. (1993) Deregulation of *Pax-2* expression in transgenic mice generates severe kidney abnormalities. *Nature* **362**, 65-67.

Drummond I. A., Madden S. L., Rohwer-Nutter P., Bell G. I., Sukhatme V. P., and Rauscher III F. J. (1992) Repression of the Insulin-Like Growth Factor II Gene by the Wilms Tumor Suppressor *WT1*. *Science* **257**, 674-678.

Echelard Y., Epstein D. J., St-Jacques B., Shen L., Mohler J., McMahon J. A., and McMahon A. P. (1993) Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**, 1417-1430.

Ekblom P. (1981) Determination and differentiation of the nephron. *Med. Biol.* **59**, 139-160.

Ekblom P. (1989) Developmentally regulated conversion of mesenchyme to epithelium. *FASEB J.* **3**, 2141-2150.

Elbrecht A. and Smith R. G. (1992) Aromatase enzyme activity and sex determination in chickens. *Science* **255**, 467-470.

Elkan E. (1963) Three different types of tumours in *Salientia*. *Cancer Res.* **23**, 1641-1645.

Etheridge A. L. (1969) Determination of the mesonephric kidney. *J.Exp.Zool.* **169**, 357-370.

Evans R. M. (1989) Molecular characterization of the glucocorticoid receptor. *Recent Prog.Horm.Res.* **45**, 1-22.

Eyre-Walker A. C. (1991) An analysis of codon usage in mammals: selection or mutation bias? *J.Mol.Evol.* **33**, 442-449.

Fairall L., Schwabe J. W. R., Chapman L., Finch J. T., and Rhodes D. (1993) The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* **366**, 483-487.

Fantes J. A., Bickmore W. A., Fletcher J. M., Ballesta F., Hanson I. M., and van-Heyningen V. (1992) Submicroscopic deletions at the WAGR locus, revealed by nonradioactive in situ hybridization. *Am. J. Hum. Genet.* **51**, 1286-1294.

Fearon E. R., Vogelstein B., and Feinberg A. P. (1984) Somatic deletion and duplication of genes on chromosome 11 in Wilms' tumours. *Nature* **309**, 176-178.

Feinberg A. P. and Vogelstein B. (1983) A technique for radiolabelling DNA restriction fragments to high specific activity. *Anal. Biochem.* **132**, 6.

Ferguson M. W. J. (1987) Post-laying stages of embryonic development for Crocodilians, in *Wildlife management: crocodiles and alligators* (Webb G. J. W., Manolis C., and Whitehead P. J., eds) Surrey Beatty and Sons Pty Limited.

Ferguson M. W. J. and Joanen T. (1982) Temperature of egg incubation determines sex in Alligator mississippiensis. *Nature* **296**, 850-853.

Forbes T. R. (1940) Studies on the reproductive system of the alligator. IV Observations on the development of the gonad, adrenal cortex, and the Mullerian duct. *Contributions to embryology* **28**, 129-155.

Fox H. (1977) The urogenital system of reptiles, in *Biology of the reptilia* (Gans C. and Parsons T. S., eds) Academic Press, Inc, New York.

Franke U., George D. L., Brown M. G., and Riccardi V. M. (1977) Gene dose effect: intraband mapping of LDH A locus using cells from four individuals with different interstitial deletions of 11p. *Cytogenet. Cell Genet.* **19**, 197-207.

Francke U., Holmes L. B., Atkins L., and Riccardi V. M. (1979) Aniridia-Wilms' tumor association: evidence for specific deletion of 11p13. *Cytogenet. Cell Genet.* **24**, 185-192.

Frazier G. C., Wu Y. J., Hewitt S. M., Maity T., Ton C. C. T., Huff V., and Saunders G. F. (1994) Transcriptional regulation of the human Wilms' tumour gene (*WT1*) cell type-specific enhancer and promiscuous promoter. *J. Biol. Chem.* **269**, 8892-8900.

Friend S. H., Horowitz J. M., Gerber M. R., Wang X.-F., Bogenmann E., Li F. P., and Weinberg R. A. (1987) Deletions of a DNA sequence in retinoblastomas and mesenchymal tumours: organisation of the sequence and its encoded protein. *Proc.Natl.Acad.Sci.USA* **84**, 9059-9063.

Fu Y. H., Pizzuti A., Fenwick R. G. Jr., King J., Rajnarayan S., Dunne P. W., Dubel J., Nasser G. A., Ashizawa T., de-Jong P., Wieringa B., Korneluk R., Perryman M. B., Epstein H. F., and Caskey C. T. (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**, 1256-1258.

Fucharoen S., Fucharoen G., Fucharoen P., and Fukumaki Y. (1989) A novel amber mutation in the β-thalassemia gene of a Thai. *J. Biol. Chem.* **264**, 7780-7783.

Gansler T., Furlanetto R., Gramling T. S., Robinson K. A., Blocker N., Buse M. G., Sens D. A., and Garvin A. J. (1989) Antibody to type I insulinlike growth factor receptor inhibits growth of Wilms' tumor in culture and in athymic mice. *Am.J.Pathol.* **135**, 961-966.

Gashler A. L., Bonthron D. T., Madden S. L., Rauscher III F. J., Collins T., and Sukhatme V. P. (1992) Human platelet-derived growth factor A chain is transcriptionally repressed by the Wilms tumor suppressor *WT1*. *Proc.Natl.Acad.Sci.USA* **89**, 10984-10988.

Geduspan J. S. and Solursh M. (1993) Effects of the mesonephros and insulin-like growth factor I on chondrogenesis of limb explants. *Dev.Biol.* **156**, 500-508.

Genetics Computor Group (1991) *Program Manual for GCG Package, Version 7, April 1991.* 575 Science Drive, Madison, Wisconsin.

Gerber H.-P., Seipel K., Georgiev O., Hofferer M., Hug M., Rusconi S., and Schaffner W. (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263**, 808-811.

Gessler M., Thomas G. H., Couillin P., Junien C., McGillivray B. C., Hayden M., Jaschek G., and Bruns G. A. P. (1989) A Deletion Map of the WAGR Region on Chromosome 11. *Am. J. Hum. Genet.* **44**, 486-495.

Gessler M., Poustka A., Cavanee W., Neve R. L., Orkin S. H., and Bruns G. A. P. (1990) Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping. *Nature* **343**, 774-778.

Gessler M., Konig A., and Bruns G. A. P. (1992) The Genomic Organisation and Expression of the *WT1* Gene. *Genomics* **12**, 807-813.

Gessler M., Konig A., Moore J., Qualman S., Arden K., Cavenee W., and Bruns G. (1993) Homozygous Inactivation of *WT1* in a Wilms' Tumor Associated With the WAGR Syndrome. *Genes Chromosom.Cancer* **7**, 131-136.

Gilbert S. F. (1991) *Developmental Biology, 3rd ed.* Sinauer Associates, Inc, Sunderland, Mass.

Glaser T., Lewis W. H., Bruns G. A. P., Watkins P. C., Rogler C. E., Shows T. B., Powers V. E., Willard H. F., Gougen J. M., Simola K. O. J., and Houseman D. E. (1986) The β-subunit of follicle stimulating hormone is deleted in patients with aniridia and Wilms' tumour allowing a further definition of the WAGR locus. *Nature* **321**, 882-887.

Glaser T., Jones C., Douglass E. C., and Housman D. (1989) Constitutional and Somatic Mutations of Chromosome 11p in Wilms' Tumor. *Cancer Cells* **7**, 253-277.

Glaser T., Lane J., and Housman D. (1990) A mouse model of the aniridia-Wilms tumor deletion syndrome. *Science* **250**, 823-827.

Glaser T., Walton D. S., and Maas R. L. (1992) Genomic structure, evolutionary conservation, and aniridia mutations in the *PAX6* gene. *Nat.Genet.* **2**, 232-239.

Goodrich D. W. and Lee W. H. (1990) The molecular genetics of retinoblastoma. *Cancer Surv.* **9**, 529-554.

Graves J. A. M. (1987) The evolution of mammalian sex chromosomes and dosage compensation: clues from marsupials and monotremes. *Trends Genet.* **3**, 252-256.

Grobstein C. (1955) Inductive interaction in the development of the mouse metanephros. *J.Exp.Zool.* **130**, 319-340.

Grove A. J. and Newel G. E., eds (1953) *Animal Biology, 4th ed.* University Tutorial Press, London.

Grundy P., Koufos A., Morgan K., Li F. P., Meadows A. T., and Cavenee W. K. (1988) Familial predisposition to Wilms' tumour does not map to the short arm of chromosome 11. *Nature* **336**, 374-376.

Gubbay J., Collignon J., Koopman P., Capel B., Economou A., Munsterberg A., Vivian N., Goodfellow P., and Lovell-Badge R. (1990) A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**, 245-250.

Gubler U. and Hoffman B. J. (1983) A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263-269.

Guerini D. and Klee C. B. (1989) Cloning of human calcineurin A: evidence for two isozymes and identification of a polyproline structural domain. *Proc.Natl.Acad.Sci.USA* **86**, 9183-9187.

Haber D. A., Park S., Maheswaran S., Englert C., Re G. G., Hazen-Martin D. J., Sens D. A., and Garvin A. J. (1993) *WT1*-mediated growth suppression of Wilms tumor cells expressing a *WT1* splicing variant. *Science* **262**, 2057-2059.

Haber D. A., Buckler A. J., Glaser T., Call K. M., Pelletier J., Sohn R. L., Douglass E. C., and Housman D. E. (1990) An Internal Deletion within an 11p13 Zinc Finger Gene Contributes to the Development of Wilms' Tumor. *Cell* **61**, 1257-1269.

Haber D. A., Sohn R. L., Buckler A. J., Pelletier J., Call K. M., and Housman D. E. (1991) Alternative splicing and genomic structure of the Wilms tumor gene *WT1*. *Proc.Natl.Acad.Sci.USA* **88**, 9618-9622.

Haber D. A., Timmers H. T. M., Pelletier J., Sharp P. A., and Housman D. E. (1992) A dominant mutation in the Wilms tumor gene *WT1* cooperates with the viral oncogene *E1A* in transformation of primary kidney cells. *Proc.Natl.Acad.Sci.USA* **89**, 6010-6014.

Habib R., Loirat C., Gubler M. C., Niaudet P., Bensman A., Levy M., and Broyer M. (1985) The nephropathy associated with male pseudohermaphroditism and Wilms' tumor (Drash syndrome): a distinctive glomerular lesion - report of 10 cases. *Clinical Nephrology* **24**, 269-278.

Haffen K. (1970) Biosynthesis of steroid hormones by the embryonic gonads of vertebrates. *Adv. Morphog.* **8**, 285-306.

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

Hamburger V. and Hamilton H. L. (1951) A series of normal stages in the development of the chick embryo. *J. Morph.* **88**, 49-92.

Hammerman M. R., Rogers S. A., and Ryan G. (1992) Growth factors and metanephrogenesis. *Am J Physiol* **262**, F523-F532.

Hammerschmidt M. and Nusslein-Volhard C. (1993) The expression of a zebrafish gene homologue to *Drosophila snail* suggests a conserved function in invertebrate and vertebrate gastrulation. *Development* **119**, 1107-1118.

Hanson I. M., Fletcher J. M., Jordan T., Brown A., Taylor R. J., Adams R. J., Punnett H. H., and van Heyningen V. (1994) Mutations at the *PAX6* locus are found in heterogenous anterior segment malformations including Peters' anomaly. *Nat.Genet.* **6**, 168-173.

Hard G. C. (1984) Comparative oncology. II Nephroblastoma in domesticated and wild animals, in *Wilms' Tumour: Clinical and Biological Manifestations* (Pochedly C. and Baum E. S., eds) Elsevier Science Publishing Company Inc., New York.

Hard G. C. and Noble R. L. (1981) Occurrence, transplantation, and histological characterics of nephroblastoma in the Nb hooded rat. *Invest. Urol.* **18**, 371-376.

Harrington M. A., Konicek B., Song A., Xia X. L., Fredericks W. J., and Rauscher F. J. 3d. (1993) Inhibition of colony-stimulating factor-1 promoter activity by the product of the Wilms' tumor locus. *J. Biol. Chem.* **268**, 21271-21275.

Harris H., Miller O. J., Klein G., Worst P., and Tachebana T. (1969) Suppression of malignancy by cell fusion. *Nature* **223**, 363-368.

Heine V., de The G., and Ishiguro H. (1962) Multiplicity of cell response to the BAI strain A (myeloblastosis) avian tumour virus. II. Nephroblastoma (Wilms' tumour): ultrastructure. *J.Nat.Cancer Inst.* **29**, 41-105.

Hendricks J. D., Scanlan R. A., and Williams J. (1980) The carcinogenicity of N-methyl-N'-nitro-N-nitroguanidine to the livers and kidneys of rainbow trout (*Salmo garidneri*) exposed as embryos. *J.Nat.Cancer Inst.* **64**, 1511-1519.

Henry I., Grandjouan S., Couillin P., Barichard F., Huerre-Jeanpierre C., Glaser T., Philip T., Lenoir G., Chaussain J. L., and Junien C. (1989) Tumor-specific loss of 11p15.5 alleles in del11p13 Wilms tumor and in familial adrenocortical carcinoma. *Proc.Natl.Acad.Sci.USA* **86**, 3247-3251.

Henry I., Bonaiti-Pellie C., Chehensse V., Beldjord C., Schwartz C., Utermann G., and Junien C. (1991) Uniparental paternal disomy in a genetic cancer-predisposing syndrome. *Nature* **351**, 665-667.

Henry I., Puech A., Riesewijk A., Ahnine L., Mannens M., Beldjord C., Bitoun P., Tournade M. F., Landrieu P., and Junien C. (1993a) Somatic Mosaicism for Partial Paternal Isodisomy in Wiedemann-Beckwith Syndrome: A Post-Fertilization Event. *Eur. J. Hum. Gen.* **1**, 19-29.

Henry I., Hoovers J., Barichard F., Bertheas Marie-F., Puech A., Prieur F., Gessler M., Bruns G., Mannens M., and Junien C. (1993b) Pericentric Intrachromosomal Insertion Responsible for Recurrence of del(11)(p13p14) in a Family. *Genes Chromosom.Cancer* **7**, 57-62.

Heppe R. K., Koyle M. A., and Beckwith J. B. (1991) Nephrogenic rests in Wilms tumor patients with the Drash syndrome. *J.Urol* **145**, 1225-1228.

Herskowitz (1987) Functional inactivation of genes by dominant negative mutations. *Nature* **329**, 219-222.

Higgins D. G. and Sharpe P. M. (1988) CLUSTAL: a package for performing multiple sequence alignments on a microcomputor. *Gene* **73**, 237-244.

Higgins D. G., Bleasby A. J., and Fuchs R. (1992) CLUSTAL V: improved software for multiple sequence alignment. *Comput Appl Biosci* **8**, 189-191.

Hofmann W., Royer H. D., Drechsler M., Schneider S., and Royer-Pokora B. (1993) Characterization of the transcriptional regulatory region of the human *WT1* gene. *Oncogene* **8**, 3123-3132.

Holland P. W. H. (1990) Homeobox genes and segmentation: co-option, co-evolution, and convergence. *Sem.Dev.Biol.* **1**, 135-145.

Horvath G. C. and Schubach W. H. (1993) Identification of the Epstein-Barr virus nuclear antigen 2 transactivation domain. *Biochem Biophys Res Commun* **191**, 196-200.

Huang A., Campbell C. E., Bonetta L., McAndrews-Hill M. S., Chilton-MacNeill S., Coppes M. J., Law D. J., Feinberg A. P., Yeger H., and Williams B. R. (1990) Tissue, developmental, and tumor-specific expression of divergent transcripts in Wilms tumor. *Science* **250**, 991-994.

Huff V., Compton D. A., Chao L. Y., Strong L. C., Geiser C. F., and Saunders G. F. (1988) Lack of linkage of familial Wilms' tumour to chromosomal band 11p13. *Nature* **336**, 377-378.

Huff V., Miwa H., Haber D. A., Call K. M., Housman D., Strong L. C., and Saunders G. F. (1991) Evidence for *WT1* as a Wilms Tumor (WT) Gene: Intragenic germinal deletion in Bilateral WT. *Am. J. Hum. Genet.* **48**, 997-1003.

Huff V., Reeve A. E., Leppert M., Strong L. C., Douglass E. C., Geiser C. F., Li F. P., Meadows A., Callen D. F., Lenoir G., and Saunders G. F. (1992) Nonlinkage of 16q Markers to Familial Predisposition to Wilms' Tumor. *Cancer Res.* **52**, 6117-6120.

The Huntington's Disease Collaberative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971-983.

Ikamura T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol.Biol.Evol.* **2**, 13-34.

Imataka H., Sogawa K., Yasumoto K., Kikuchi Y., Sasano K., Kobayashi A., Hayami M., and Fujii-Kuriyama Y. (1992) Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-4501A1 gene. *Embo J.* **11**, 3663-3671.

Ingham P. W. (1994) Hedgehog points the way. *Curr. Biol* **4**, 347-350.

Innis M. A., Gelfand D. H., Sninsky J. J., and White T. W. (1990) *PCR Protocols. A Guide to Methods and Applications*. Academic Press, Inc., San Diego.

Ishiguro H., Beard D., and Sommer J. R. (1962) Multiplicity of cell response to the BAI strain A (myeloblastosis) avian tumour virus. I. Nephroblastoma (Wilms' tumour): Gross and microscopic pathology. *J.Nat.Cancer Inst.* **29**, 1-39.

Jacks T., Fazeli A., Schmitt E. M., Bronson R. T., Goodell M. A., and Weinberg R. A. (1992) Effects of an Rb mutation in the mouse. *Nature* **359**, 295-300.

Jackson I. J. (1991) A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* **19**, 3795-3798.

Jacobs G. H. (1992) Determination of the base recognition positions of zinc fingers from sequence analysis. *Embo J.* **11**, 4507-4517.

Jasanoff A. and Weiss M. A. (1993) Aromatic-histidine interactions in the zinc finger motif: structural inequivalence of phenylalanine and tyrosine in the hydrophobic core. *Biochemistry* **32**, 1423-1432.

Jasmin G. and Riopelle J. L. (1970) Nephroblastomas induced in ovariectomized rats by dimethylbenzanthracene. *Cancer Res.* **30**, 321-326.

Jirasek J. E. (1977) Morphogenesis of the genital system in the human. *Birth Defects: Original Article Series* **XIII**, 13-39.

Joliot V., Martinerie C., Dambrine G., Plassiart G., Brisac M., Crochet J., and Perbal B. (1992) Proviral rearrangements and overexpression of a new cellular gene (*nov*) in myeloblastosis-associated virus type 1-induced nephroblastomas. *Mol.Cell. Biol.* **12**, 10-21.

Jones D. S. C. and Schofield J. P. (1990) A rapid method for isolating high quality plasmid DNA suitable for DNA sequencing. *Nucleic Acids Res.* **18**, 7463-7464.

Jordan T., Hanson I., Zaletayev D., Hodgson S., Prosser J., Seawright A., Hastie N., and van Heyningen V. (1992) The human *PAX6* gene is mutated in two patients with aniridia. *Nat.Genet.* **1**, 328-332.

Joseph L. J., LeBeau M. M., Jamieson G. A., Acharya S., Shows T. B., Rowley J. D., and Sukhatme V. P. (1988) Molecular cloning, sequencing, and mapping of *EGR2,* a human early growth response gene encoding a protein with "zinc-binding finger" structure. *Proc.Natl.Acad.Sci.USA* **85**, 7164-7168.

Jukes T. H. and Cantor C. R. (1969) Evolution of protein molecules III, in *Mammalian Protein Metabolism* (Munro H. N., ed), pp. 21-132. Academic Press, New York.

Jurgelski W., Hudson P. M., and Falk H. L. (1976) Embryonal neoplasms in the opossum; a new model for solid tumours of infancy and childhood. *Science* **193**, 328-332.

Kadonaga J. T., Carner K. R., Masiarz F. R., and Tjian R. (1987) Isolation of a cDNA encoding transcription factor *Sp1* and functional analysis of the DNA binding domain. *Cell* **51**, 1079-1090.

Kalscheuer V. M., Mariman E. C., Schepens M. T., Rehder H., and Ropers H. H. (1993) The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat.Genet.* **5**, 74-78.

Karlin S. and Altschul S. F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc.Natl.Acad.Sci.USA* **87**, 2264-2268.

Kaufman M. H. (1992) *The Atlas of Mouse Development.* Academic Press, London.

Kemp B. E. and Pearson R. B. (1990) Protein kinase recognition sequence motifs. *Trends Biochem. Sci.* **15**, 342-346.

Kent G. C. (1987) *Comparative Anatomy of the Vertebrates, 6th ed.* Times Mirror/Mosby College Publishing, St. Louis.

Kerr L. D., Ransone L. J., Wamsley P., Schmitt M. J., Boyer T. G., Zhou Q., Berk A. J., and Verma I. M. (1993) Association between proto-oncoprotein *Rel* and TATA-binding protein mediates transcriptional activation by *NF-kappa B. Nature* **365**, 412-419.

Kimura M. (1968) Evolutionary rate at molecular level. *Nature* **217**, 624-626.

Kimura M. (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275-276.

Kimura M. (1983) *The neutral theory of molecular evolution.* Cambridge University Press, Cambridge.

Kingsley C. and Winoto A. (1992) Cloning of GT box-binding proteins: a novel *Sp1* multigene family regulating T-cell receptor gene expression. *Mol.Cell. Biol.* **12**, 4251-4261.

Knight S. J., Flannery A. V., Hirst M. C., Campbell L., Christodoulou Z., Phelps S. R., Pointon J., Middleton-Price H. R., Barnicoat A., Pembrey M. E., Holland J., Oostra B. A., Bobrow M., and Davies K. E. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74**, 127-134.

Knudson J. A. G. (1971) Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc.Natl.Acad.Sci.USA* **68**, 820-823.

Knudson J. A. G. and Strong L. C. (1972) Mutation and Cancer: A Model for Wilms' Tumor of the Kidney. *J.Nat.Cancer Inst.* **48**(2), 313-324.

Kohne D. E. (1970) Evolution of higher-organism DNA. *Quart. Rev. Biophys.* **33**, 327-375.

Konig A., Jakubiczka S., Wieacker P., Schlosser H. W., and Gessler M. (1993) Further evidence that imbalance of *WT1* isoforms may be involved in Denys-Drash syndrome. *Hum. Mol. Genet.* **2**, 1967-1968.

Koopman P., Gubbay J., Vivian N., Goodfellow P., and Lovell-Badge R. (1991) Male development of chromosomally female mice transgenic for *Sry. Nature* **351**, 117-121.

Koufos A., Hansen M. F., Copeland N. G., Jenkins N. A., Lampkin B. C., and Cavenee W. K. (1985) Loss of heterozygosity in three embryonal tumours suggests a common pathogenetic mechanism. *Nature* **316**, 330-334.

Koufos A., Grundy P., Morgan K., Aleck K. A., Hadro T., Lampkin B. C., Kalbakji A., and Cavenee W. K. (1989) Familial Wiedemann-Beckwith Syndrome and a Second Wilms Tumor Locus Both Map to 11p15.5. *Am. J. Hum. Genet.* **44**, 711-719.

Kouzarides T. and Ziff E. (1988) The role of the leucine zipper in the fos-jun interactions. *Nature* **336**, 646-651.

Kozak M. (1987) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J.Mol.Biol.* **196**, 947-950.

Kramer S., Meadows A. T., and Jarrett P. (1984) Racial Variation in incidence of Wilms' tumour: relationship to congenital abnormalities. *Med. Paed. Onc.* **12**, 401-405.

Krauss S., Johanson T., Korzh V., and Fjose A. (1991) Expression pattern of zebrafish pax genes suggests a role in early brain regionalization. *Nature* **353**, 267-270.

Krauss S., Concordet J.-P., and Ingham P. W. (1993) A functionally conserved homolog of the *Drosophila* segment polarity gene *hh* is expressed in tissues with polarising activity in zebrafish embryos. *Cell* **75**, 1431-1444.

Kreidberg J. A., Sariola H., Loring J. M., Maeda M., Pelletier J., Housman D., and Jaenisch R. (1993) *WT1* Is Required for Early Kidney Development. *Cell* **74**, 679-691.

Kriwacki R. W., Schultz S. C., Steitz T. A., and Caradonna J. P. (1992) Sequence-specific recognition of DNA by zinc-finger peptides derived from the transcription factor *Sp1*. *Proc.Natl.Acad.Sci.USA* **89**, 9759-9763.

Kumada Y., Benson D. R., Hillemann D., Hosted T. J., Rochefort D. A., Thompson C. J., Wohlleben W., and Tateno Y. (1993) Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes. *Proc.Natl.Acad.Sci.USA* **90**, 3009-3013.

Kushner L., Cozzolino A., Sherman J., and Rich M. A. (1992) Expression of a Wilms tumor gene in porcine kidney during compensatory renal growth. *J.Urol.* **148**, 555-558.

Kyte J. and Doolittle R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J.Mol.Biol.* **157**, 105-132.

Lamb P. and McKnight S. L. (1991) Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerisation. *Trends Biochem. Sci.* **16**, 417-422.

Lance V. A. and Bogart M. H. (1991) Tamoxifen 'sex reverses' alligator embryos at male producing temperature, but is an antiestrogen in female hatchlings. *Experientia* **47**, 263-266.

Lance V. A. and Bogart M. H. (1992) Disruption of ovarian development in alligator embryos treated with an aromatase inhibitor. *Gen. Comp. Endocrinol.* **86**, 59-71.

Landshultz W. H., Johnson P. F., and McKnight S. L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759-1764.

Lanfear J. and Holland P. W. H. (1991) The molecular evolution of *ZFY*-related genes in birds and mammals. *J.Mol.Evol.* **32**, 310-315.

Lanfear J., Jowett T., and Holland P. W. H. (1991) Cloning of fish zinc-finger genes related to *Krox-20* and *Krox-24*. *Biochem Biophys Res Commun* **179**, 1220-1224.

La Spada A. R., Wilson E. M., Lubahn D. B., Harding A. E., and Fischbeck K. H. (1991) Androgen receptor gene mutations in spinal and bulbar muscular atrophy. *Nature* **352**, 77-79.

Lear P. V. and Stern C. D. (1994) Fate map of the intermediate mesoderm of the chick embryo. Abstract, British Society for Developmental Biology Spring Meeting, Evolution of Development.

Lee C. C., Wu X., Gibbs R. A., Cook R. G., Muzny D. M., and Caskey C. T. (1988) Generation of cDNA probes directed by amino acid sequence: cloning of urate oxidase. *Science* **239**, 1288-1291.

Lee M. G. and Nurse P. (1987) Complementation used to clone a human homologue of the fission yeast cell cycle control gene *cdc2*. *Nature* **327**, 31-35.

Lee W.-H., Brookstein R., Hong F., Young L.-J., Shew J.-Y., and Lee E. Y.-H. P. (1987) Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science* **235**, 1394-1399.

Lemoine N. R., Hughes C. M., and Cowell J. K. (1992) Aberrant Expression Of The Tumour Suppressor Gene p53 Is Very Frequent In Wilms' Tumours. *J. Pathol.* **168**, 237-242.

Lettice L. A. and Slack J. M. W. (1993) Properties of the dorsalizing signal in gastrulae of *Xenopus laevis*. *Development* **117**, 263-271.

Levine A. J. (1992) The p53 tumour suppressor gene and product. *Cancer Surv.* **12**, 59-79.

Lewis W. H., Yeger H., Bonetta L., Chan H. S. L., Kang J., Junien C., Cowell J., Jones C., and Dafoe L. A. (1988) Homozygous Deletion of a DNA Marker from Chromosome 11p13 in Sporadic Wilms Tumor. *Genomics* **3**, 25-31.

Li H., Cui X., and Arnheim N. (1990) Direct electrophoretic detection of the allelic state of single DNA molecules in human sperm by using polymerase chain reaction. *Proc.Natl.Acad.Sci.USA* **87**, 4580-4584.

Li W. H. (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J.Mol.Evol.* **36**, 96-99.

Li W.-H. and Graur D. (1991) *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, Mass.

Li W.-H., Gojobori T., and Nei M. (1981) Pseudogenes as a paradigm of neutral evolution. *Nature* **292**, 237-239.

Li W.-H., Wu C.-I., and Luo C.-C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol.Biol.Evol.* **2**, 150-174.

Liebelt A. G., Sass B., Sobel H. J., and Werner R. M. (1989) Spontaneous nephroblastoma in a strain CE/J mouse. A case report. *Toxicol Pathol* **17**, 57-61.

Little M., van Heyningen V., and Hastie N. (1991) Dads and disomy and disease. *Nature* **351**, 609-610.

Little M. H., Prosser J., Condie A., Smith P. J., van Heyningen V., and Hastie N. D. (1992a) Zinc finger point mutations within the *WT1* gene in Wilms tumor patients. *Proc.Natl.Acad.Sci.USA* **89**, 4791-4795.

Little M. H., Dunn R., Byrne J. A., Seawright A., Smith P. J., Pritchard-Jones K., van Heyningen V., and Hastie N. D. (1992b) Equivalent expression of paternally and maternally inherited *WT1* alleles in normal fetal tissue and Wilms' tumours. *Oncogene* **7**, 635-641.

Little M. H., Clarke J., Byrne J., Dunn R., and Smith P. J. (1992c) Allelic Loss on Chromosome 11p is a Less Frequent Event in Bilateral than in Unilateral Wilms' Tumours. *Eur. J. Cancer* **28A**, 1876-1880.

Little M. H., Williamson K. A., Mannens M., Kelsey A., Gosden C., Hastie N. D., and van Heyningen V. (1993) Evidence that *WT1* mutations in Denys-Drash syndrome patients may act in a dominant-negative fashion. *Hum. Mol. Genet.* **2**, 259-264.

Lundin L. G. (1993) Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* **16**, 1-19.

Lupas A., Van-Dyke M., and Stock J. (1991) Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164.

Lynch K. and Fraser S. E. (1990) Cell migration in the formation of the pronephric duct in *Xenopus laevis*. *Dev.Biol.* **142**, 283-292.

Lyons K. M., Pelton R. W., and Hogan B. L. M. (1990) Organogenesis and pattern formation in the mouse: RNA distribution patterns suggest a role for Bone Morphogenetic Protein-2A (BMP2A). *Development* **109**, 833-844.

Madden S. L., Cook D. M., Morris J. F., Gashler A., Sukhatme V. P., and Rauscher III F. J. (1991) Transcriptional Repression Mediated by the *WT1* Wilms Tumor Gene Product. *Science* **253**, 1550-1553.

Madden S. L., Cook D. M., and Rauscher III F. J. (1993) A structure-function analysis of transcriptional repression mediated by the *WT1*, Wilms' tumor suppressor protein. *Oncogene* **8**, 1713-1720.

Mahadevan M., Tsilfidis C., Sabourin L., Shutler G., Amemiya C., Jansen G., Neville C., Narang M., Barcelo J., O'Hoy K., et al. (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**, 1253-1255.

Maheswaran S., Park S., Bernard A., Morris J. F., Rauscher III F. J., Hill D. E., and Haber D. A. (1993) Physical and functional interaction between *WT1* and p53 proteins. *Proc.Natl.Acad.Sci.USA* **90**, 5100-5104.

Mandelkern L. (1967), in *Poly-a-amino acids* (Fasman G. D., ed) Marcel Decker, New York.

Mannens M., Slater R. M., Heyting C., Bliek J., de Kraker J., Coad N., de Pagter-Holthuizen P., and Pearson P. L. (1988) Molecular nature of genetic changes resulting in loss of heterozygosity of chromosome 11 in Wilms' tumours. *Hum. Genet.* **81**, 41-48.

Martin A. P. and Palumbi S. R. (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proc.Natl.Acad.Sci.USA* **90**, 4087-4091.

Martin P., Carriere C., Dozier C., Quatannens B., Mirabel M. A., Vandenbunder B., Stehelin D., and Saule S. (1992) Characterization of a paired box- and homeobox-containing quail gene (*Pax-QNR*) expressed in the neuroretina. *Oncogene* **7**, 1721-1728.

Martinerie C., Viegas-Pequignot E., Guenard I., Dutrillaux B., Nguyen V. C., Bernheim A., and Perbal B. (1992) Physical mapping of human loci homologous to the chicken nov proto-oncogene. *Oncogene* **7**, 2529-2534.

Masahito P., Ishikawa T., Okamoto N., and Sugano H. (1992) Nephroblastomas in the Japanese eel, *Anguilla japonica* Temminck and Schlegel. *Cancer Res.* **52**, 2575-2579.

Matsunaga E. (1981) Genetics of Wilms' tumour. *Hum. Genet.* **57**, 231-246.

Maw M. A., Grundy P. E., Millow L. J., Eccles M. R., Dunn R. S., Smith P. J., Feinberg A. P., Law D. J., Paterson M. C., Telzerow P. E., Callen D. F., Thompson A. D., Richards R. I., and Reeve A. E. (1992) A Third Wilms' Tumor Locus on Chromosome 16q. *Cancer Res.* **52**, 3094-3098.

McKeown M. (1992) Alternative mRNA splicing. *Annu.Rev.Cell Biol.* **8**, 133-155.

Mermod N., O'Neill E. A., Kelly T. J., and Tjian R. (1989) The proline-rich transcriptional activator of *CTF/NF-I* is distinct from the replication and DNA binding domain. *Cell* **58**, 741-753.

Mierau G. W., Beckwith J. B., and Weeks D. A. (1987) Ultrastructure and histogenesis of the renal tumours of childhood: an overview. *Ultra. Path.* **11**, 313-333.

Miller J., McLachlan A. D., and Klug A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *Embo J.* **4**, 1609-1614.

Miller R. W., Fraumeni J. F., and Manning M. D. (1964) Association of Wilms' tumour with aniridia, hemihypertrophy and other congenital abnormalities. *N. Engl.J.Med.* **270**, 922-927.

Mitchell P. J. and Tjian R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-378.

Miwa H., Tomlinson G. E., Timmons C. F., Huff V., Cohn S. L., Strong L. C., and Saunders G. F. (1992a) RNA Expression of the *WT1* Gene in Wilms' Tumors in Relation to Histology. *J.Nat.Cancer Inst.* **84**(3), 181-187.

Miwa H., Beran M., and Saunders G. F. (1992b) Expression of the Wilms' Tumor Gene (*WT1*) in Human Leukemias. *Leukemia* **6**, 405-409.

Monaco A. P., Neve R. L., Colletti-Feener C., Bertelson C. J., Kurnit D. M., and Kunkel L. M. (1986) Isolation of candidate cDNAs for the portions of the Duchenne muscular distrophy gene. *Nature* **323**, 646-650.

Moore K. L. and Persaud T. V. N. (1993) *The Developing Human: clinically orientated embryology, 5th ed.* Saunders, Philadelphia.

Morris J. F., Madden S. L., Tournay O. E., Cook D. M., Sukhatme V. P., and Rauscher III F. J. (1991) Characterisation of the zinc finger protein encoded by the *WT1* Wilms' tumor locus. *Oncogene* **6**, 2339-2348.

Mugrauer G. and Ekblom P. (1991) Contrasting expression patterns of three members of the myc family of protooncogenes in the developing and adult mouse kidney. *J.Cell Biol.* **112**, 13-25.

Mullis K. B. and Faloona F. A. (1987) Specfic synthesis of DNA in vitro via a polymerase-catalysed chain reaction. *Methods Enzymol.* **155**, 335-350.

Mundlos S., Pelletier J., Darveau A., Bachmann M., Winterpacht A., and Zabel B. (1993) Nuclear-localisation of the protein encoded by the Wilms' tumour gene *WT1* in embryonic and adult tissues. *Development* **119**, 1329-1341.

Nehlin J. O. and Ronne H. (1990) Yeast *MIG1* repressor is related to the mammalian early growth response and Wilms' tumour finger proteins. *Embo J.* **9**, 2891-2898.

Nei M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Neuhaus D., Nakaseko Y., Nagai K., and Klug A. (1990) Sequence-specific [1H]NMR resonance assignments and secondary structure identification for 1- and 2-zinc finger constructs from SW15. A hydrophobic core involving four invariant residues. *FEBS Lett* **262**, 179-184.

Nicholson G. W. (1931) A embryonic tumour of the kidney in a foetus. *J.Pathol.Bacteriol.* **34**, 711-730.

Niehrs C., Steinbeisser H., and De Robertis E. M. (1994) Mesodermal patterning by a gradient of the vertebrate homeobox gene *goosecoid. Science* **263**, 817-820.

Nietfeld W., Conrad S., van-Wijk I., Giltay R., Bouwmeester T., Knochel W., and Pieler T. (1993) Evidence for a clustered genomic organization of FAX-zinc finger protein encoding transcription units in *Xenopus laevis. J.Mol.Biol.* **230**, 400-412.

Nieto M. A., Bradley L. C., and Wilkinson D. G. (1991) Conserved segmental expression of *Krox-20* in the vertebrate hindbrain and its relationship to lineage restriction. *Development*(Suppl. 2), 59-62.

Nieuwkoop P. D. and Faber J., eds (1967) *Normal Table of Xenopus laevis (Daudin), 2nd ed.* North-Holland Publishing Company, Amsterdam.

Nordenskjold A., Friedman E., and Anvret M. (1994) *WT1* mutations in patients with Denys-Drash syndrome: a novel mutation in exon 8 and paternal allele origin. *Hum. Genet.* **93**, 115-120.

Ogawa O., Eccles M. R., Yun K., Mueller R. F., Holdaway M. D. D., and Reeve A. E. (1993a) A novel insertional mutation at the third zinc finger coding region of the *WT1* gene in Denys-Drash syndrome. *Hum. Mol. Genet.* **2**, 203-204.

Ogawa O., Eccles M. R., Szeto J., McNoe L. A., Yun K., Maw M. A., Smith P. J., and Reeve A. E. (1993b) Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. *Nature* **362**, 749-751.

Ogawa O., Becroft D. M., Morison I. M., Eccles M. R., Skeen J. E., Mauger D. C., and Reeve A. E. (1993c) Constitutional relaxation of insulin-like growth factor II gene imprinting associated with Wilms' tumour and gigantism. *Nat.Genet.* **5**, 408-412.

O'Hara P. J., Horowitz H., Eichinger G., and Young E. T. (1988) The yeast *ADR6* gene encodes homopoymeric amino acid sequences and a potential metal-binding domain. *Nucleic Acids Res.* **16**, 10153-10169.

Ohlsson R., Nystrom A., Pfeifer-Ohlsson S., Tohonen V., Hedborg F., Schofield P., Flam F., and Ekstrom T. J. (1993) *IGF2* is parentally imprinted during human embryogenesis and in the Beckwith-Wiedemann syndrome. *Nat.Genet.* **4**, 94-97.

Olson J. M., Breslow N. E., and Beckwith J. B. (1993) Wilms' tumour and parental age: a report from the National Wilms' Tumour Study. *Br.J.Cancer* **67**, 813-818.

O'Rahilly R. (1977) The development of the vagina in the human. *Birth Defects: Original Article Series* **XIII**, 123-136.

Orkin S. H., Goldman D. S., and Sallen S. E. (1984) Development of homozygosity for chromosome 11p markers in Wilms' tumour. *Nature* **309**, 172-174.

Orlandi R., Gussow D. H., Jones P. T., and Winter G. (1989) Cloning immunoglobulin variable domains for expression by the polymerase chain reaction. *Proc.Natl.Acad.Sci.USA* **86**, 3833-3837.

Oxtoby E. and Jowett T. (1993) Cloning of the zebrafish *krox-20* gene *(krx-20)* and its expression during hindbrain development. *Nucleic Acids Res.* **21**, 1087-1095.

Pal N., Wadey R. B., Buckle B., Yeomans E., Pritchard J., and Cowell J. K. (1990) Preferential loss of maternal alleles in sporadic Wilms' tumour. *Oncogene* **5**, 1665-1668.

Pang C. P., Crossley M., Kent G., and Brownlee G. G. (1990) Comparative sequence analysis of mammalian factor IX promoters. *Nucleic Acids Res.* **18**, 6731-6732.

Park S., Schalling M., Bernard A., Maheswaran S., Shipley G. C., Roberts D., Fletcher J., Shipman R., Rheinwald J., Demetri G., Griffin J., Minden M., Houseman D. E., and Haber D. A. (1993a) The Wilms tumour gene *WT1* is expressed in murine mesoderm-derived tissues and mutated in a human mesothelioma. *Nat.Genet.* **4**, 415-420.

Park S., Bernard A., Bove K. E., Sens D. A., Hazen-Martin D. J., Garvin A. J., and Haber D. A. (1993b) Inactivation of *WT1* in nephrogenic rests, genetic precursors to Wilms' tumour. *Nat.Genet.* **5**, 363-367.

Park S., Tomlinson G., Nisen P., and Haber D. A. (1993c) Altered trans-activational properties of a mutated *WT1* gene product in a WAGR-associated Wilms' tumor. *Cancer Res.* **53**, 4757-4760.

Passmore S., Maine G. T., Elble R., Christ C., and Tye B.-K. (1988) Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MATα cells. *J.Mol.Biol.* **204**, 593-606.

Patthy L. (1985) Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. *Cell* **41**, 657-663.

Patwardhan S., Gashler A., Siegel M. G., Chang L. C., Joseph L. J., Shows T. B., Le-Beau M. M., and Sukhatme V. P. (1991) *EGR3*, a novel member of the Egr family of genes encoding immediate-early transcription factors. *Oncogene* **6**, 917-928.

Pavletich N. P. and Pabo C. O. (1993) Crystal structure of a five-finger *GLI*-DNA complex: new perspectives on zinc fingers. *Science* **261**, 1701-1707.

Pavletich N. P. and Pabo C. O. (1991) Zinc Finger-DNA Recognition: Crystal Structure of a *Zif268*-DNA Complex at 2.1A. *Science* **252**, 809-817.

Pelletier J., Schalling M., Buckler A. J., Rogers A., Haber D. A., and Housman D. (1991c) Expression of the Wilms' tumor gene *WT1* in the murine urogenital system. *Genes Dev.* **5**, 1345-1356.

Pelletier J., Bruening W., Li F. P., Haber D. A., Glaser T., and Housman D. E. (1991a) *WT1* mutations contribute to abnormal genital system development and hereditary Wilms' tumour. *Nature* **353**, 431-434.

Pelletier J., Bruening W., Kashtan C. E., Mauer S. M., Manivel J. C., Striegel J. E., Houghton D. C., Junien C., Habib R., Fouser L., Fine R. N., Silverman B. L., Haber D. A., and Housman D. (1991b) Germline Mutations in the Wilms' Tumor Suppressor Gene Are Associated with Abnormal Urogenital Development in Denys-Drash Syndrome. *Cell* **67**, 437-447.

Pendleton J. W., Nagai B. K., Murtha M. T., and Ruddle F. H. (1993) Expansion of the *Hox* gene family and the evolution of chordates. *Proc.Natl.Acad.Sci.USA* **90**, 6300-6304.

Perlman M., Goldberg G. M., Bar-Ziv J., and Danovitch G. (1973) Renal hamartomas and nephroblastomatosis with fetal gigantism: a familial syndrome. *J.Pediat* **83**, 414-418.

Plachov D., Chowdhury K., Walther C., Simon D., Guenet J. L., and Gruss P. (1990) *Pax8*, a murine paired box gene expressed in the developing excretory system and thyroid gland. *Development* **110**, 643-651.

Porteus D. J., Bickmore W., Christie S., Boyd P. A., Cranston G., Fletcher J. M., Gosden J. R., Rout D., Seawright A., Simola K. O. J., van Heyningen V., and Hastie N. D. (1987) HRAS-1 selected chromosome transfer generates markers that colocalise aniridia-and genitourinary dysplasia-associated translocation breakpoints and the Wilms' tumour gene within band 11p13. *Proc.Natl.Acad.Sci.USA* **84**, 5355-5359.

Poulat F., Morin D., Konig A., Brun P., Giltay J., Sultan C., Dumas R., Gessler M., and Berta P. (1993) Distinct molecular origins for Denys-Drash and Frasier syndromes. *Hum. Genet.* **91**, 285-286.

Powers D. A. (1991) Evolutionary Genetics of Fish. *Adv. Genet.* **29**, 119-228.

Prestridge D. S. (1991) SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. *Comput Appl Biosci* **7**, 203-206.

Pritchard-Jones K. and Hastie N. (1990) Wilms' tumour as a paradigm for the relationship of cancer to development. *Cancer Surv.* **9**, 555-578.

Pritchard-Jones K. and Fleming S. (1991) Cell types expressing the Wilms' tumour gene (*WT1*) in Wilms' tumours: implications for tumour histogenesis. *Oncogene* **6**, 2211-2220.

Pritchard-Jones K., Fleming S., Davidson D., Bickmore W., Porteous D., Gosden C., Bard J., Buckler A., Pelletier J., Housman D., van Heyningen V., and Hastie N. (1990) The candidate Wilms' tumour gene is involved in genitourinary development. *Nature* **346**, 194-197.

Proudfoot N. J. (1994) Chasing your own poly(A) tail. *Curr. Biol* **4**, 359-361.

Quek H. H., Chow V. T., and Tock E. P. (1993) The third zinc finger of the *WT1* gene is mutated in Wilms' tumour but not in a broad range of other urogenital tumours. *Anticancer Res.* **13**, 1575-1580.

Rauscher III F. J., Morris J. F., Tournay O. E., Cook D. M., and Curran T. (1990) Binding of the Wilms' Tumor Locus Zinc Finger Protein to the EGR-1 Consensus Sequence. *Science* **250**, 1259-1262.

Reese A. M. (1908) The development of the American alligator (*A. mississippiensis*). *Smithson. misc. Collns* **51**, 3-66.

Reeve A. E., Housiaux P. J., Gardener R. J. M., Chewings W. E., Grindley R. M., and Millow L. J. (1984) Loss of a Harvey ras allele in sporadic Wilms' tumour cells. *Nature* **309**, 174-176.

Reeve A. E., Eccles M. R., Wilkins R. J., Bell G. I., and Millow L. J. (1985) Expression of insulin-like growth factor-II transcripts in Wilms' tumour. *Nature* **317**, 258-260.

Reeve A. E., Sih S. A., Raizis A. M., and Feinberg A. P. (1989) Loss of Allelic Heterozygosity at a Second Locus on Chromosome 11 in Sporadic Wilms' Tumor Cells. *Mol.Cell. Biol.* **9**, 1799-1803.

Riccardi V. M., Sujansky E., Smith A. C., and Francke U. (1978) Chromosomal imbalance in the aniridia-Wilms' tumour association: 11p interstitial deletion. *Pediatrics* **61**, 604-610.

Richards R. I. and Sutherland G. R. (1992) Dynamic mutations: a new class of mutations causing human disease. *Cell* **70**, 709-712.

Richards R. I. and Sutherland G. R. (1994) Simple repeat DNA is not replicated simply. *Nat.Genet.* **6**, 114-116.

Riddle R. D., Johnson R. L., Laufer E., and Tabin C. (1993) Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401-1416.

Roberts D. J., Haber D., Sklar J., and Crum C. P. (1993) Extrarenal Wilms' tumors. A study of their relationship with classical renal Wilms' tumor using expression of *WT1* as a molecular marker. *Lab.Invest.* **68**, 528-536.

Romanoff A. L. (1960) *The Avian Embryo*. Macmillan, New York.

Romer A. S. and Parsons T. S. (1977) *The Vertebrate Body, 5th ed.* Saunders, Philadelphia.

Rose E. A., Glaser T., Jones C., Smith C. L., Lewis W. H., Call K. M., Minden M., Champagne E., Bonetta L., Yeger H., and Housman D. E. (1990) Complete Physical Map of the WAGR Region of 11p13 Localizes a Candidate Wilms' Tumor Gene. *Cell* **60**, 495-508.

Rosen V. and Thies R. S. (1992) The BMP proteins in bone formation and repair. *Trends Genet.* **8**, 97-102.

Rothenpieler U. W. and Dressler G. R. (1993) *Pax-2* is required for mesenchyme-to-epithelial conversion during kidney development. *Development* **119**, 711-720.

Royer-Pokora B., Ragg S., Heckl-Ostreicher B., Held M., Loos U., Call K., Glaser T., Housman D., Saunders G., Zabel B., Williams B., and Poustka A. (1991) Direct Pulsed Field Gel Electrophoresis of Wilms' Tumors Shows That DNA Deletions in 11p13 Are Rare. *Genes Chromosom.Cancer* **3**, 89-100.

Ruppert J. M., Kinzler K. W., Wong A. J., Bignor S. H., Kao F.-T., Law M. L., Seuanez H. N., O'Brien S. J., and Vogelstein B. (1988) The *GLI-Kruppel* Family of Human Genes. *Mol.Cell. Biol.* **8**, 3104-3113.

Rupprecht H. D., Drummond I. A., Madden S. L., Rauscher III F. J., and Sukhatme V. P. (1994) The Wilms' tumour suppressor gene *WT1* is negatively autoregulated. *J. Biol. Chem.* **269**, 6198-6206.

Russel M., Kidd S., and Kelley M. R. (1986) An improved filamentous helper phage for generating single-stranded plasmid DNA. *Gene* **45**, 333-338.

Russo M. W., Matheny C., and Milbrandt J. (1993) Transcriptional activity of the zinc finger protein *NGFI-A* is influenced by its interaction with a cellular factor. *Mol.Cell. Biol.* **13**, 6858-6865.

Saiki R. K., Scharf S., Faloona F., Mullis K. B., Horn G. T., Erlich H. A., and Arnheim N. (1985) Enzymatic Amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-1354.

Saitou N. and Nei.M (1987) A new method for reconstructing phylogenetic trees. *Mol.Biol.Evol.* **4**, 406-425.

Sakai A., Tadokoro K., Yanagisawa H., Nagafuchi S., Hoshikawa N., Suzuki T., Kohsaka T., Hasegawa T., Nakahori Y., and Yamada M. (1993) A novel mutation of the *WT1* gene (a tumor suppressor gene for Wilms' tumor) in a patient with Denys-Drash syndrome. *Hum. Mol. Genet.* **2**, 1969-1970.

Sambrook J., Fritsch E. F., and Maniatis T. (1989) *Molecular cloning A Laboratory Manual, 2nd ed.* Cold Spring Habor Laboratory Press, Cold Spring Habor, NY.

Sander C. and Schneider R. (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* **9**, 56-68.

Sanger F., Nicklen S., and Coulson A. R. (1977) DNA sequencing with chain terminating inhibitors. *Proc.Natl.Acad.Sci.USA* **74**, 5463-5467.

Santos A., Osorio-Almeida L., Baird P. N., Silva J. M., Boavida M. G., and Cowell J. (1993) Insertional inactivation of the *WT1* gene in tumour cells from a patient with WAGR syndrome. *Hum. Genet.* **92**, 83-86.

Sariola H., Ekblom P., and Henke-Fahle S. (1989) Embryonic neurons as *in vitro* inducers of differentiation of nephrogenic mesenchyme. *Dev.Biol.* **132**, 271-281.

Sarkar G. and Sommer S. S. (1989) Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. *Science* **244**, 331-334.

Sarkar G., Koeberl D. D., and Sommer S. S. (1990) Direct sequencing of the activation peptide and the catalytic domain of the factor IX gene in six species. *Genomics* **6**, 133-143.

Saxen L. (1970) Failure to demonstrate tubule induction in a heterlogous mesenchyme. *Dev.Biol.* **23**, 511-523.

Saxen L. (1987) *Organogenesis of the Kidney*. Cambridge University Press, Cambridge.

Scharf S. J., Horn G. T., and Erlich H. A. (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science* **233**, 1076-1078.

Schneider S., Wildhardt G., Ludwig R., and Royer-Pokora B. (1993) Exon skipping due to a mutation in a donor splice site in the *WT-1* gene is associated with Wilms' tumor and severe genital malformations. *Hum. Genet.* **91**, 599-604.

Schreiber-Agus N., Horner J., Torres R., Chiu F. C., and DePinho R. A. (1993) Zebra fish *myc* family and max genes: differential expression and oncogenic activity throughout vertebrate evolution. *Mol.Cell. Biol.* **13**, 2765-2775.

Schroeder W. T., Chao Lian-Y., Dao D. D., Strong L. C., Pathak S., Riccardi V., Lewis W. H., and Saunders G. F. (1987) Nonrandom Loss of Maternal Chromosome 11 Alleles in Wilms Tumors. *Am. J. Hum. Genet.* **40**, 413-420.

Schulte-Merker S., Ho R. K., Herrmann B. G., and Nusslein-Volhard C. (1992) The protein product of the zebrafish homologue of the mouse *T* gene is expressed in nuclei of the germ ring and the notochord of the early embryo. *Development* **116**, 1021-1032.

Schwartz C. E., Haber D. A., Stanton V. P., Strong L. C., Skolnick M. H., and Housman D. E. (1991) Familial predisposition to Wilms tumor does not segregate with the *WT1* gene. *Genomics* **10**, 927-930.

Scott J., Cowell J., Robertson M. E., Priestley L. M., Wadey R., Hopkins B., Pritchard J., Bell G. I., Rall L. B., Graham C. F., and Knott T. J. (1985) Insulin-like growth factor-II gene expression in Wilms' tumour and embryonic tissues. *Nature* **317**, 260-262.

Shannon R. S., Mann J. R., Harper E., Harnden D. G., Morten J. E. N., and Herbert A. (1982) Wilms' tumour and aniridia: clinical and cytogenetic features. *Arch. Disease. Childhood* **57**, 685-690.

Sharma P. M., Yang X., Bowman M., Roberts V., and Sukumar S. (1992) Molecular cloning of rat Wilms' tumor complementary DNA and a study of messenger RNA expression in the urogenital system and the brain. *Cancer Res.* **52**, 6407-6412.

Sharma P. M., Bowman M., Madden S. L., Rauscher III F. J., and Sukumar S. (1994) RNA editing in the Wilms' tumour susceptibility gene, *WT1*. *Genes Dev.* **8**, 720-731.

Short J. M., Fernandez J. M., Sorge J. A., and Huse W. D. (1988) λ-ZAP: a bacteriophage l expression vector with *in vivo* excision properties. *Nucleic Acids Res.* **16**, 7583-7600.

Sidow A. (1992) Diversification of the *Wnt* gene family on the ancestral lineage of vertebrates. *Proc.Natl.Acad.Sci.USA* **89**, 5098-5102.

Slack J. M. W., Holland P. W. H., and Graham C. F. (1993) The zootype and the phylotypic stage. *Nature* **361**, 490-492.

Smith C. A. and Joss J. M. P. (1993) Gonadal sex differentiation in *Alligator mississippiensis*, a species with temperature-dependent sex determination. *Cell Tissue Res.* **273**, 149-162.

Sommer R. and Tautz D. (1989) Minimal homology requirements for PCR primers. *Nucleic Acids Res.* **17**, 6749.

Sommer R. J., Retzlaff M., Goerlich K., Sander K., and Tautz D. (1992) Evolutionary conservation pattern of zinc-finger domains of *Drosophila* segmentation genes. *Proc.Natl.Acad.Sci.USA* **89**, 10782-10786.

Soussi T., Caron-de-Fromentel C., and May P. (1990) Structural aspects of the p53 protein in relation to gene evolution. *Oncogene* **5**, 945-952.

Stanbridge E. J. (1976) Suppression of malignancy in human cells. *Nature* **260**, 17-20.

Strahle U. and Ingham P. W. (1992) Flight of fancy or a major new school? *Curr. Biol* **2**, 135-138.

Strum R. A. and Herr W. (1988) The POU domain is a bipartite DNA-binding structure. *Nature* **336**, 601-604.

Su L. K., Johnson K. A., Smith K. J., Hill D. E., Vogelstein B., and Kinzler K. W. (1993) Association between wild type and mutant *APC* gene products. *Cancer Res.* **53**, 2728-2731.

Su W., Jackson S., Tjian R., and Echols H. (1991) DNA looping between sites for transcriptional activation: self-association of DNA-bound *Sp1*. *Genes Dev.* **5**, 820-826.

Sukhatme V. P. (1992) The *Egr* transcription factor family: from signal transduction to kidney differentiation. *Kidney Int* **41**, 550-553.

Sukhatme V. P., Cao X., Chang L. C., Tsai-Morris Chon-H., Stamenkovich D., Ferreira P. C. P., Cohen D. R., Edwards S. A., Shows T. B., Curran T., Le Beau M. M., and Adamson E. D. (1988) A Zinc Finger-Encoding Gene Coregulated with *c-fos* during Growth and Differentiation, and after Cellular Depolarization. *Cell* **53**, 37-43.

Suzuki Y., Nogi Y., Abe A., and Fukasawa T. (1988) *Gal11* protein, an auxiliary transcription activator for genes encoding galactose-metabolising enzymes in Saccharomyces cerevisiae. *Mol.Cell. Biol.* **8**, 4991-4999.

Tadokoro K., Fujii H., Ohshima A., Kakizawa Y., Shimizu K., Sakai A., Sumiyoshi K., Inoue T., Hayashi Y., and Yamada M. (1992a) Intragenic homozygous deletion of the *WT1* gene in Wilms' tumor. *Oncogene* **7**, 1215-1221.

Tarugi P., Albertazzi L., Nicolini S., and Calandra S. (1990) Absence of apolipoprotein B-48 in the chick, *Gallus domesticus*. *J.Lipid Res.* **31**, 417-427.

Telerman A., Dodemont H., Degraef C., Galand P., Bauwens S., van Oostveldt P., and Amson R. B. (1992) Identification of the cellular protein encoded by the human Wilms' tumor (*WT1*) gene. *Oncogene* **7**, 2545-2548.

Temin H. M. (1985) Reverse transcription in the eukaryotic genome: retrovirus, pararetrovirus, retrotransposons, and retrotranscripts. *Mol.Biol.Evol.* **2**, 455-468.

Tepass U. and Hartenstein V. (1994) Epithelium formation in the *Drosophila* midgut depends on the interaction of the endoderm and mesoderm. *Development* **120**, 579-590.

Thelier K. (1978) *The House Mouse. Atlas of Embryonic Development.* Springer-Verlag, New York.

Tiersch T. R., Mitchell M. J., and Wachtel S. S. (1991) Studies on the phylogenetic conservation of the *SRY* gene. *Hum. Genet.* **87**, 571-573.

Tilley W. D., Marcelli M., Wilson J. D., and McPhaul M. J. (1989) Characterization and expression of a cDNA encoding the human androgen receptor. *Proc.Natl.Acad.Sci.USA* **86**, 327-331.

Ton C. C. T., Miwa H., and Saunders G. F. (1992) Small eye (Sey): cloning and characterization of the murine homolog of the human aniridia gene. *Genomics* **13**, 251-256.

Ton C. C T., Huff V., Call K. M., Cohn S., Strong L. C., Housman D. E., and Saunders G. F. (1991a) Smallest region of overlap in Wilms tumor deletions uniquely implicates an 11p13 zinc finger gene as the disease locus. *Genomics* **10**, 293-297.

Ton C. C. T., Hirvonen H., Miwa H., Weil M. M., Monaghan P., Jordan T., van Heyningen V., Hastie N. D., Meijers-Heijboer H., Drechsler M., Royer-Pokora B., Collins F., Swaroop A., Strong L. C., and Saunders G. F. (1991b) Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell* **67**, 1059-1074.

Torrey T. W. (1965) Morphogenesis of the vertebrate kidney, in *Organogenesis* (DeHaan R. L. and Ursprung H., eds), pp. 559-579. Holt, Rinehart, and Winston, New York.

Travers A. (1993) *DNA - protein interactions*. Chapman and Hall, London.

Trumpp A., Blundell P. A., de-la-Pompa J. L., and Zeller R. (1992) The chicken limb deformity gene encodes nuclear proteins expressed in specific cell types during morphogenesis. *Genes Dev.* **6**, 14-28.

Tsarfaty I., Sing R., Resau J. H., Rulong S., Pinto da Silva P., and Vande Woude G. F. (1994) The *met* proto-oncogene mesenchymal to epithelial cell conversion. *Science* **263**, 98-101.

Tso Y., Van Den Berg D. J., and Korn L. J. (1986) Structure of the gene for Xenopus transcription factor *TFIIIA*. *Nucleic Acids Res.* **14**, 2187-2200.

Tucker P. K. and Lundrigan B. L. (1993) Rapid Evolution of the sex determining locus in Old World mice and rats. *Nature* **364**, 715-717.

Tunnacliffe A., Liu L., Moore J. K., Leversha M. A., Jackson M. S., Papi L., Ferguson-Smith M. A., Thiesen H. J., and Ponder B. A. (1993) Duplicated *KOX* zinc finger gene clusters flank the centromere of human chromosome 10: evidence for a pericentric inversion during primate evolution. *Nucleic Acids Res.* **21**, 1409-1417.

Turleau C., de Grouchy J., Tournade M.-F., Gagnadoux M.-F., and Junien C. (1984) Del 11p/aniridia complex: report of three patients and review of 37 observations from the literature. *Clin. Genet.* **26**, 356-362.

Tymowska J. and Fischberg M. (1973) Chromosome complements of the genus *Xenopus*. *Chromosoma* **44**, 335-342.

van Heyningen V. (1994) One gene - four syndromes. *Nature* **367**, 319-320.

van Heyningen V., Boyd P. A., Seawright A., Fletcher J. M., Fantes J. A., Buckton K. E., Spowart G., Porteous D. J., Hill R. E., Newton M. S., and Hastie N. D. (1985) Molecular analysis of chromosome 11 deletions in aniridia-Wilms' tumour syndrome. *Proc.Natl.Acad.Sci.USA* **82**, 8592-8596.

van Heyningen V., Bickmore W. A., Seawright A., Fletcher J. M., Maule J., Fekete G., Gessler M., Bruns G. A. P., Huerre-Jeanpierre C., Junien C., Williams B. R. G., and Hastie N. D. (1990) Role for the Wilms' tumor gene in genital development? *Proc.Natl.Acad.Sci.USA* **87**, 5383-5386.

Varanasi R., Bardeesy N., Ghahremani M., Petruzzi M.-J., Nowak N., Adam M. A., Grundy P., Shows T. B., and Pelletier J. (1994) Fine structure analysis of the *WT1* gene in sporadic Wilms tumours. *Proc.Natl.Acad.Sci.USA* **91**, 3554-3558.

Velasco S., D'Amico D., Schneider N. R., Timmons C., Chappell E., Lee D., and Nisen P. D. (1993) Molecular and cellular heterogeneity of Wilms' tumor. *Int J Cancer* **53**, 672-679.

Wada K., Wada Y., Doi H., Ishibashi F., Gojobori T., and Ikemura T. (1991) Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res.* **19 Suppl**, 1981-1986.

Wadey R. B., Pal N., Buckle B., Yeomans E., Pritchard J., and Cowell J. K. (1990) Loss of heterozygosity in Wilms' tumour involves two distinct regions of chromosome 11. *Oncogene* **5**, 901-907.

Wanek N., Muneoka K., Holler-Dinsmore G., Burton R., and Bryant S. V. (1989) A staging system for mouse limb development. *J.Exp.Zool.* **249**, 41-49.

Wang Z. Y., Madden S. L., Deuel T. F., and Rauscher III F. J. (1992) The Wilms' Tumor Gene Product, *WT1*, Represses Transcription of the Platelet-derived Growth Factor A-chain Gene. *J. Biol. Chem.* **267**, 21999-22002.

Wang Zhao-Y., Qiu Qing-Q., and Deuel T. F. (1993a) The Wilms' Tumor Gene Product *WT1* Activates or Suppresses Transcription through Separate Functional Domains. *J. Biol. Chem.* **268**(13), 9172-9175.

Wang-Wuu S., Soukoup S., Bove K., Gotwals B., and Lampkin B. (1990) Chromosome Analysis of 31 Wilms' tumours. *Cancer Res.* **50**, 2786-2793.

Washabaugh C. H., Del Rio-Tsonis K., and Tsonis P. A. (1993) Variable manifestations in the short toes (s) mutation of the axolotl. *J. Morph.* **218**, 107-114.

Waziri M., Patil S., Hanson J., and Bartley S. A. (1983) Abnormality of chromosome 11 patients with features of Beckwith-Wiedemann syndrome. *J.Pediat* **102**, 873-876.

Weidemann H. R. (1964) Complexe malformaltif familial avec hernie ombilicale et macroglossie: un syndrome nouveau? *J.Genet.Hum.* **13**, 223-232.

Wang Z. Y., Qiu Q. Q., Enger K. T., and Deuel T. F. (1993b) A second transcriptionally active DNA-binding site for the Wilms tumor gene product, *WT1*. *Proc.Natl.Acad.Sci.USA* **90**, 8896-8900.

Weissman B. E., Saxon P. J., Pasquale S. R., Jones G. R., Geiser A. G., and Stanbridge E. J. (1987) Introduction of a normal human chromosome 11 into a Wilms' tumour cell line controls its tumorigenic expression. *Science* **236**, 175-180.

Weksberg R., Shen D. R., Fei Y. L., Song Q. L., and Squire J. (1993) Disruption of insulin-like growth factor 2 imprinting in Beckwith-Wiedemann syndrome. *Nat.Genet.* **5**, 143-150.

Werner H., RE G. G., Drummond I. A., Sukhatme V. P., Rauscher III F. J., Sens D. A., Garvin A. J., LeRoith D., and Roberts J. C. T. (1993) Increased expression of the insulin-like growth factor I receptor gene, IGF1R, in Wilms tumor is correlated with modulation of IGF1R promoter activity by the *WT1* Wilms tumor gene product. *Proc.Natl.Acad.Sci.USA* **90**, 5828-5832.

Whitfield L. S., Lovell-Badge R., and Goodfellow P. N. (1993) Rapid sequence evolution of the mammalian sex-determining gene *SRY*. *Nature* **364**, 713-715.

Wilkins R. J. (1988) Genomic imprinting and carcinogenesis. *Lancet* **i**, 329-331.

Wilkinson D. G., ed (1992) *In situ hybridisation: a practical approach.* IRL, Oxford.

Williams J. C., Brown K. W., Mott M. G., and Maitland N. J. (1989) Maternal Allele Loss In Wilms' Tumour. *Lancet* **i**, 283-284.

Wilson A. C., Ochman H., and Prager E. M. (1987) Molecular time scale for evolution. *Trends Genet.* **3**, 241-247.

Wilson T. E., Day M. L., Pexton T., Padgett K. A., Johnston M., and Milbrandt J. (1992) In vivo mutational analysis of the *NGFI-A* zinc fingers. *J. Biol. Chem.* **267**, 3718-3724.

Wimmer E. A., Jackle H., Pfelfle C., and Cohen S. M. (1993) A *Drosophila* homologue of human *Sp1* is a head-specific segmentation gene. *Nature* **366**, 690-694.

Winship P. R. (1989) An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Res.* **17**, 1266.

Woychik R. P., Maas R. L., Zeller R., Vogt T. F., and Leder P. (1990) 'Formins': proteins deduced from the alternative transcripts of the limb deformity gene. *Nature* **346**, 850-853.

Wu C.-I. and Li W.-H. (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc.Natl.Acad.Sci.USA* **82**, 1741-1745.

Yanisch-Perron C., Vieira J., and Messing J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Yeger H., Cullinane C., Flenniken A., Chilton-MacNeill S., Campbell C., Huang A., Bonetta L., Coppes M. J., Thorner P., and Williams B. R. G. (1992) Coordinate Expression of Wilms' Tumor Genes Correlates with Wilms' Tumor Phenotypes. *Cell Growth & Differentiation* **3**, 855-864.

Young P. H. and Miller R. W. (1975) Incidence of malignant tumours in U.S. children. *J.Pediat* **86**, 254-258.

Young R. A. and Davis R. W. (1983a) Efficient isolation of genes by using antibody probes. *Proc.Natl.Acad.Sci.USA* **80**, 1194-1198.

Young R. A. and Davis R. W. (1983b) Yeast RNA polymerase II genes: isolation with antibody probes. *Science* **222**, 778-789.

Yu H., Chen J. K., Feng S., Dalgarno D. C., Brauer A. W., and Schreiber S. L. (1994) Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933-945.

Yun K., Fidler A. E., Eccles M. R., and Reeve A. E. (1993) Insulin-like growth factor II and *WT1* transcript localization in human fetal kidney and Wilms' tumor. *Cancer Res.* **53**, 5166-5171.

Zahler A. M., Neugebauer K. M., Lane W. S., and Roth M. B. (1993) Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* **260**, 219-222.

Zawel L. and Reinberg D. (1993) Initiation of transcription by RNA polymerase II: a multi-step process. *Prog.Nucleic Acid Res.Mol.Biol.* **44**, 67-108.

Zhong W., Sladek F. M., Darnell Jr. J. E. (1993) The expression pattern of a *Drosophila* homolog to the mouse transcription factor *HNF-4* suggests a determinative role in gut formation. *EMBO J.* **12**, 537-44

Zhou J., Rauscher III F. J., and Bondy C. (1993) Wilms' tumor (*WT1*) gene expression in rat decidual differentiation. *Differentiation* **54**, 109-114.

Zwart P. (1970) A nephroblastoma in a fire bellied newt, *Cynops pyrrhogaster. Cancer Res.* **30**, 2691-2694.