CLONING AND CHARACTERIZATION
OF "Unp"
A NOVEL NUCLEAR PROTO-ONCOGENE

KALPANA GUPTA
Ottawa, Canada, 1993

Thesis submitted to
The School of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Department of Biochemistry
University of Ottawa
Ottawa, Ontario
CANADA

© Kalpana Gupta, Ottawa, Canada, 1993
The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L’auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L’auteur conserve la propriété du droit d’auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ABSTRACT

The cloning and characterization of a novel gene, Unp (for Ubiquitin-specific Nuclear Protease), is described in this thesis. This gene was identified during a survey for candidate genes for the retroviral insertional mutation MpV 20. Unp is transcribed into two ubiquitously expressed mRNAs of approximately 3.5 and 3.7 kb. The two mRNAs vary at their 3' ends due to alternative usage of two degenerate polyadenylation signals. The 5' UTR of Unp contains several minicistrons, which suggest a growth related function for this gene. The Unp gene has been mapped to mouse chromosome 9 and human chromosome 3 in a region which is implicated in several types of human cancers based on loss of heterozygocity. This suggests a possible tumor suppressor function for this gene.

The predicted amino acid sequence of the Unp protein contains several conserved motifs, including p53 type nuclear localization signal, consensus for binding to the retinoblastoma protein, and the Cys and His domains conserved in yeast ubiquitin proteases. The pRb binding consensus is present in the region of Unp which shares homology to the recombinant human oncogene tre-2. Unp protein is localized in the nucleus and migrates at a higher molecular mass than expected, suggesting that this protein is post-translationally modified. Unp is tumorigenic in athymic mice showing that it is a proto-oncogene. The protein product of Unp is a ubiquitin-specific protease. The Unp gene is highly polymorphic in humans, and loss of Unp sequences is detected in 1 out of the 7 informative human lung tumors analyzed.
DEDICATION

I would like to dedicate this thesis
to my husband Sanjay

and my parents Swaroop and Sheela Gupta,

without their love, support and encouragement

this thesis would not have been possible.
ACKNOWLEDGEMENTS

There are a lot of people who have contributed to my scientific background in one way or another. It is impossible to acknowledge all of them, and will try to acknowledge the ones who have contributed directly.

First and foremost, I would like to thank my supervisor Dr. Douglas A. Gray who has not only provided me with an excellent working environment, but also made this project so very challenging. He has been extremely understanding and supportive throughout the course of my stay with him.

I am very grateful to Dr. John Bell for bailing me out when I was beginning to lose faith in science and for introducing me to Dr. Gray and recommending me to him. If it were not for Dr. Bell I probably would not have enrolled in this Ph.D. John has also provided me with a lot of helpful advice regarding the protein work presented in this thesis.

Dr. M. Chevrette and Dr. M. McBurney are thanked for putting up with me during the committee meetings and for keeping my brain awake during these meetings.

Beth Mason is thanked for all her help throughout my stay at University of Ottawa.

My appreciation and thanks also goes to the following people in the lab:

Manon Dubé for inspiring me to clean my bench once in a while and for providing me with technical assistance, especially during the past year. She was also a very entertaining bench mate to work with.

Josee Coulombe for being such an understanding and caring labmate, and for her soup recipes.

Marco Di Fruscio for his entertaining swing dances and for conducting a critical computer search.

Ricardo Marius and Elizabeth Duville for being brave enough to accept working opposite me in the lab. Congratulations Liz, you are the only person to survive working opposite me for the entire three years.
Brian Howell, Danny Afar, Peter Duncun Ninan Abrahim and Maria Jarmillo for their helpful advice with the protein work.

Trudy Marshall for helping me with preparation of this thesis.

A special thanks goes to my soon to be born daughter, Sneha Rani Gupta, who has been very cooperative throughout the final stages of my work and also during the writing of this thesis.
PREFACE

Some of the work presented in this thesis has been or will be submitted for publication as described below:

CHAPTER 2 (in part)


Chapters 2, 3 and 4 (some parts)


The experimental work presented in this thesis is entirely my own with the following exceptions:

CHAPTER 2

The initial screening of the cDNA libraries was done by Dr. D.A. Gray.

The chromosomal mapping of Unp to mouse chromosome 9 was performed by Dr. Nancy Jenkin's group.

CHAPTER 3

The peptides used to generate polyclonal antibodies were made by Ricardo Marius.

CHAPTER 4

The analysis showing ubiquitin-specific protease activity of Unp was done in collaboration with Dr. R.T. Baker.
TABLE OF CONTENTS

ABSTRACT ........................................ Page: i
DEDICATION ........................................ Page: ii
ACKNOWLEDGEMENTS ................................ Page: iii
PREFACE ............................................ Page: V
TABLE OF CONTENTS ............................... Page: Vi
LIST OF TABLES ..................................... Page: Xii
LIST OF FIGURES ................................... Page: Xiii
LIST OF ABBREVIATIONS ........................... Page: XVI

CHAPTER I GENERAL INTRODUCTION ............ Page: 1

1.1 RETROVIRAL INSERTIONAL MUTAGENESIS .... Page: 2

1.11 Insertional Mutagenesis ....................... Page: 3
1.12 Insertional Mutants ............................ Page: 4
1.13 Mpv 20 ....................................... Page: 5

1.2 PROTO-ONCOGENES ............................. Page: 6

1.21 Mechanisms Of Activation Of Oncogenes .... Page: 11

1.211 Activation Through Changes In Phosphorylation Of Specific Tyrosine Residues: .......... Page: 11
1.212 Through Transmission Of Signals Via GTPases: ........................................ Page: 12
1.213 Via Altered Regulation Of Transcription Factors ........................................ Page: 16

1.214 Through Proviral Insertion .............................................................................. Page: 18

1.215 Through Chromosomal Translocation .......................................................... Page: 20

1.3 THE RECOMBINANT HUMAN ONCOGENE tre ........................................ Page: 25

1.31 Cloning And Identification Of tre ................................................................. Page: 25

1.32 Transcription Of The Transforming Gene tre ............................................. Page: 26

1.33 Features Of The tre-2 cDNAs ....................................................................... Page: 27

1.34 Tumorigenicity Of tre-2 ................................................................................ Page: 29

1.4 TUMOR SUPPRESSOR GENES .................................................................... Page: 29

1.41 Rb-1, The Retinoblastoma Gene .................................................................. Page: 33

1.411 Cell Cycle Dependent Phosphorylation Of pRb ........................................ Page: 34

1.412 Binding Of pRb To Viral Oncoproteins ..................................................... Page: 36

1.413 Binding Of pRb To E2F ............................................................................ Page: 39

1.42 The p53 Gene ............................................................................................... Page: 42

1.5 UBIQUITINATION ......................................................................................... Page: 46

1.51 The Ubiquitin Pathway ................................................................................. Page: 47

1.52 Functions Of Ubiquitination ........................................................................ Page: 49

1.53 Substrates Of The Ubiquitin Proteolytic Pathway ........................................ Page: 53
CHAPTER 2  

Unp: A NOVEL GENE WITH SEVERAL  
CONSERVED DOMAINS AND  
INTERESTING FEATURES  

Page: 56

2.1 SUMMARY  

Page: 57

2.2 INTRODUCTION  

Page: 58

2.3 MATERIALS AND METHODS  

Page: 59

2.31 Screening And Isolation Of cDNA Clones  

Page: 59

2.32 Sequencing Of cDNA Clones  

Page: 59

2.33 Southern (DNA) Blot Analysis  

Page: 60

2.4 RESULTS  

Page: 60

2.41 Cloning Of cDNAs For The Mouse Unp Gene  

Page: 60

2.42 Absence Of A Consensus Polyadenylation In Unp  
cDNAs  

Page: 61

2.43 5' UTR Of Unp cDNAs Contains Several  
Minicistrons  

Page: 64

2.44 Features Of The Large Open Reading Frame Of  
Unp cDNAs  

Page: 66

2.45 Unp Is Located On Mouse Chromosome 9  

Page: 69

2.46 Unp Is A Member Of A Gene Family, Conserved  
Through Evolution  

Page: 72

2.47 Unp Shares Homology With The Recombinant  
Human Oncogene tre  

Page: 75

2.48 Unp Is Derived From A Gene Distinct From tre-  
2  

Page: 78
2.5 DISCUSSION ............................................................ Page: 80

CHAPTER 3  
THE UBIQUITOUSLY EXPRESSED Unp  
GENE CODES FOR A NUCLEAR  
PROTEIN ............................................................ Page: 83

3.1 SUMMARY ........................................................... Page: 84

3.2 INTRODUCTION ....................................................... Page: 85

3.3 MATERIALS AND METHODS ....................................... Page: 86

3.31 Constructs .......................................................... Page: 86

3.32 Tissue Culture And Cell Lines Used ....................... Page: 86

3.33 Northern (RNA) Blot Analysis ............................... Page: 88

3.34 Expression Of Bacterial Gst-Unp Fusion Protein .... Page: 89

3.35 Generation And Affinity Purification Of Unp  
Antibodies ............................................................ Page: 89

3.36 Western (Protein) Blot Analysis ............................. Page: 91

3.37 Nuclear Fractionation Of NIH 3T3 Extracts ............ Page: 92

3.38 In Vitro Transcription-Translation Reactions .......... Page: 92

3.4 RESULTS ............................................................... Page: 93

3.41 Unp mRNA Is Ubiquitously Expressed .................... Page: 93

3.42 In vitro Translation Of Unp cDNAs ....................... Page: 96

3.43 Detection Of Unp Protein in vivo ......................... Page: 96

3.44 Unp protein is localized in the nucleus ............... Page: 98

3.5 DISCUSSION ............................................................ Page: 101
CHAPTER 4

Unp Is A Proto-Oncogene And Encodes A Ubiquitin-Specific Protease

4.1 SUMMARY ........................................ Page: 105
4.2 INTRODUCTION ..................................... Page: 106
4.3 MATERIALS AND METHODS ....................... Page: 107
  4.31 Constructs ....................................... Page: 107
  4.32 Tissue Culture And Cell Lines Used .......... Page: 109
  4.33 Southern (DNA) Blot Analysis ................. Page: 109
  4.34 Tumorigenicity Assays ........................ Page: 110
  4.35 Western (Protein) Blot Analysis ............. Page: 110
4.4 RESULTS ........................................ Page: 111
  4.41 Unp Is Tumorigenic In Athymic Mice .......... Page: 111
  4.42 Unp Is Highly Polymorphic In Human Cell Lines Page: 116
  4.43 Southern Blot Analysis Of Human Lung Tumors Page: 118
  4.44 Unp Codes For A Ubiquitin Protease .......... Page: 122
4.5 DISCUSSION ................................... Page: 123

CHAPTER 5

DISCUSSION ........................................ Page: 125

5.1 Features Of Unp mRNAs .......................... Page: 126
5.2 Features Of The Predicted Unp Protein .......... Page: 127
5.3 Homology Between Unp And tre Oncogene .......... Page: 128

5.4 Unp Is A Proto-Oncogene ..................................... Page: 129

5.5 Unp Is A Ubiquitin-Specific Protease ................... Page: 129

5.6 Unp As A Candidate Tumor Suppressor Gene ........... Page: 130

5.7 Model For Unp Function ....................................... Page: 131

ORIGINAL CONTRIBUTIONS TO KNOWLEDGE ................. Page: 137

REFERENCES ......................................................... Page: 138
LIST OF TABLES

CHAPTER 1

TABLE 1.1 Small mammalian GTPases ....................... Page: 9
TABLE 1.2 Proto-oncogenes altered in human malignancies ....................... Page: 14
TABLE 1.3 Viral integration sites clustered on mouse chromosomes ....................... Page: 21
TABLE 1.4 Tumor suppressor genes ....................... Page: 31

CHAPTER 4

TABLE 4.1 Tumorigenicity of Unp in athymic mice ....................... Page: 113
LIST OF FIGURES

CHAPTER 1

FIGURE 1.1  Mouse Mpv 20 locus ................................ : Page: 7
FIGURE 1.2  p21\textsuperscript{ras} GTPase pathway ....................... Page: 15
FIGURE 1.3  Cellular proto-oncogenes vs corresponding viral oncogenes .............................................. Page: 17
FIGURE 1.4  Schematic representation of \textit{tre-2} cDNAs ................................................................. Page: 28
FIGURE 1.5  Retinoblastoma protein and the cell cycle ................................................................. Page: 35
FIGURE 1.6  Sequence comparison of the NLSs of p53 proteins ...................................................... Page: 43
FIGURE 1.7  Representation of the ubiquitin pathway ................................................................. Page: 48
FIGURE 1.8  Comparison of Cys and His domains of \textit{Ubp} proteins .................................................. Page: 50

CHAPTER 2

FIGURE 2.1  Schematic representation of mouse \textit{Unp} cDNAs isolated ........................................... Page: 62
FIGURE 2.2  5' UTR of \textit{Unp} cDNA ................................................................. Page: 65
FIGURE 2.3  Sequence of the longest mouse \textit{Unp} cDNA, T4 .................................................. Page: 67
CHAPTER 2

FIGURE 2.4 Characteristic features of the Unp cDNA ........................................ Page: 68

FIGURE 2.5 Comparision of pRb binding consensus sequences ............................... Page: 70

FIGURE 2.6 Comparision of Cys and His domains of Ubps and Unp .......................... Page: 71

FIGURE 2.7 Chromosomal localization of Unp to mouse chromosome 9 ................. Page: 73

FIGURE 2.8 Zoo blot analysis of the Unp gene ...................................................... Page: 74

FIGURE 2.9 Southern blot analysis of Unp gene at different stringencies .................... Page: 76

FIGURE 2.10 Homology between Unp and tre-2 .................................................. Page: 77

FIGURE 2.11 Southern blot analysis of hybrid cell lines ........................................ Page: 79

CHAPTER 3

FIGURE 3.1 Schematic representation of Gst-Unp fusion constructs .......................... Page: 87

FIGURE 3.2 Northern blot analysis showing ubiquitous expression of Unp ................ Page: 94

FIGURE 3.3 Northern blot analysis to show the Unp mRNA doublet .......................... Page: 95

FIGURE 3.4 in vitro translation of Unp protein ....................................................... Page: 97

FIGURE 3.5 Western blot analysis showing specificity of the antisera ....................... Page: 99
FIGURE 3.6 Detection of *in vivo* Unp protein by western blot analysis .......... Page: 100

FIGURE 3.7 Detection of *Unp* protein in the nuclear fraction of NIH 3T3 cells .......... Page: 102

CHAPTER 4

FIGURE 4.1 Constructs used for tumorigenicity assays ......................... Page: 108

FIGURE 4.2 Southern blot analysis of DNA from tumors obtained in athymic mice ........ Page: 114

FIGURE 4.3 Western blot analysis of tumors from athymic mice .................. Page: 115

FIGURE 4.4 Southern blot analysis showing Unp polymorphism ................... Page: 117

FIGURE 4.5 Detection of polymorphism in Unp using various enzymes ............ Page: 119

FIGURE 4.6 Southern blot analysis of human lung tumors ....................... Page: 120

FIGURE 4.7 Southern blot analysis of human lung tumor 10 ..................... Page: 121

CHAPTER 5

FIGURE 5.1 A possible model for *Unp* function .......................... Page: 132
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>Acute Lymphocytic Leukemia</td>
</tr>
<tr>
<td>AML</td>
<td>Acute Myelogenous Leukemia</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CLL</td>
<td>Chronic Lymphocytic Leukemia</td>
</tr>
<tr>
<td>CML</td>
<td>Chronic Myelogenous Leukemia</td>
</tr>
<tr>
<td>CR</td>
<td>Conserved Region</td>
</tr>
<tr>
<td>C-terminus</td>
<td>Carboxy-terminus</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle’s Medium</td>
</tr>
<tr>
<td>DMSO</td>
<td>Di-Methyl-Sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxy ribo Nucleic Acid</td>
</tr>
<tr>
<td>GAP</td>
<td>Guanosine triphosphate Activating Protein</td>
</tr>
<tr>
<td>GDP</td>
<td>Guanosine Di-Phosphate</td>
</tr>
<tr>
<td>GNRP</td>
<td>Guanine Nucleotide Release Protein</td>
</tr>
<tr>
<td>Gst</td>
<td>Glutathione-S-transferase</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine Tri-Phosphate</td>
</tr>
<tr>
<td>GTPase</td>
<td>Guanosine Tri-Phosphatases</td>
</tr>
<tr>
<td>HPV</td>
<td>Human Pappilloma Virus</td>
</tr>
<tr>
<td>kb</td>
<td>kilo bases</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>LOH</td>
<td>Loss Of Heterozygocity</td>
</tr>
<tr>
<td>LTR</td>
<td>Long Terminal Repeat</td>
</tr>
<tr>
<td>MAP</td>
<td>Multiple Antigen Peptide</td>
</tr>
<tr>
<td>MMTV</td>
<td>Mouse Mammary Tumor Virus</td>
</tr>
<tr>
<td>MPF</td>
<td>M-Phase Promoting Factor</td>
</tr>
<tr>
<td>NLS</td>
<td>Nuclear Localization Signal</td>
</tr>
<tr>
<td>N-terminus</td>
<td>Amino-terminus</td>
</tr>
<tr>
<td>ORF</td>
<td>Open Reading Frame</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate Buffered Saline</td>
</tr>
<tr>
<td>Pgtk</td>
<td>Phospho-glycerate kinase</td>
</tr>
<tr>
<td>PPT-1</td>
<td>Protein Phosphatase Type-1</td>
</tr>
<tr>
<td>pRb</td>
<td>Retinoblastoma protein</td>
</tr>
<tr>
<td>RAR</td>
<td>Retinoic Acid Receptor</td>
</tr>
<tr>
<td>Rb</td>
<td>Retinoblastoma</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribo Nucleic Acid</td>
</tr>
<tr>
<td>RSV</td>
<td>Rous Sarcoma Virus</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulphate</td>
</tr>
<tr>
<td>SH2 Domain</td>
<td>Src Homology Domain-2</td>
</tr>
<tr>
<td>SV40</td>
<td>Simian Virus type 40</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris Buffered Saline</td>
</tr>
<tr>
<td>TPA</td>
<td>12-O-Tetradecanoyl-Phorbol-13-Acetate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>tre</td>
<td>Transfection Recombined</td>
</tr>
<tr>
<td>Tyr</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>Ub</td>
<td>Ubiquitin</td>
</tr>
<tr>
<td>Ubp</td>
<td>Ubiquitin-specific protease</td>
</tr>
<tr>
<td>Unp</td>
<td>Ubiquitin-specific Nuclear Protease</td>
</tr>
<tr>
<td>UTR</td>
<td>UnTranslated Region</td>
</tr>
</tbody>
</table>
CHAPTER I

GENERAL INTRODUCTION
The *Unp* gene was discovered during our survey for candidate genes for the retroviral insertional mouse mutation Mpv 20 (Gray, unpublished data). Based on the nucleotide sequence of *Unp* cDNAs, the predicted amino acid sequence of the *Unp* protein was compared to entries in the PIR and SwissProt computer databases. *Unp* protein was found to have significant homology to a known recombinant human oncogene *tre* (Nakamura et al., 1992). The region of homology resides within the portion of *tre* which is essential for transformation. Based on this homology, we investigated the possibility that *Unp* is a proto-oncogene. *Unp* is located on mouse chromosome 9 in a region which is homologous to human chromosome region 3p. Recently in collaboration with Dr. R. Baker, we have discovered that *Unp* protein has a ubiquitin-specific protease activity. In this thesis, besides the sequence and expression pattern of *Unp*, I will also present evidence to show that *Unp* is a proto-oncogene which codes for a ubiquitin-specific protease and some preliminary evidence which suggests that *Unp* may be a tumor suppressor gene as well. Thus, as an introduction to this thesis, I will discuss the relevant topics, namely: retroviral insertional mutations in brief, proto-oncogenes, the recombinant human oncogene *tre*, tumor suppressor genes, and ubiquitination.

1.1 RETROVIRAL INSERTIONAL MUTAGENESIS

A major technical advance in biology has been the ability to introduce exogenous DNA into the germ line of mammalian cells that can give rise to transgenic animals
(Gridley et al., 1987). The introduced DNA can cause mutations by disrupting or affecting the expression of a gene or genes, and can later be used as a tag for molecular cloning of the affected gene. Besides identification and cloning of new genes involved in complex developmental processes, transgenic animals have proven to be very useful as animal model systems for the study of the mechanism of development, developmental regulation of genes and action of oncogenes. The project described in this thesis was initiated during the analysis of a retroviral insertional mutant mouse strain Mpv 20.

1.11 Insertional Mutagenesis

Insertional mutations in mice have been produced in three main ways, each of which involves the introduction of exogenous DNA into the germ line of mouse. The first transgenic mouse strain was produced by exposing early mouse embryos to infectious Moloney Murine Leukemia Virus (Jaenisch, 1976). A single copy of the provirus is inserted into the chromosomal DNA. Apart from a short duplication of host sequences at the site of integration, no major rearrangements have been observed. This makes it easier to isolate intact host sequences adjacent to the site of integration. Isolation of flanking host sequences is important in identifying and cloning of the gene disrupted by the integration of the proviral DNA.

Another way of generating insertional mutants is through the microinjection of exogenous DNA into one of the pronuclei of a fertilized egg (Palmiter and Brinster, 1986). Upon microinjection, several DNA molecules can integrate into the host genome.
Also rearrangements including deletions, duplications and translocations of the host sequences have been seen to occur at the site of integration (Covarrubias et al., 1986). This makes it harder to clone the flanking host sequences in order to identify the disrupted gene.

The third approach for generating mutant mouse strains has been the introduction of ES cells (embryonal stem cells), which have previously been transfected or infected with the DNA of interest, into mouse blastocysts. After injection into host blastocysts the ES cells retain their pluripotent character and can colonize the embryo to give rise to a chimeric animal (Bradley et al., 1984).

1.12 Insertional Mutants

A few of the mouse mutants generated by early embryo infections have been characterized in considerable details. Four of these are Mov 13 (Jaenisch et al., 1983), Mov 34 (Soriano et al., 1987), Mpv 17 (Weiher et al., 1990) and Mpv 20 (Di Fruscio, Gupta and Gray, unpublished data). Three of these (Mov 13, Mpv 17 and Mpv 20) have an early embryonic lethal phenotype. Mice homozygous for these mutations die within the first 13 days of gestation. The fourth mutation (Mpv 17) has an adult recessive lethal phenotype.

Homozygous Mov 13 embryos show necrosis of erythropoietic and mesenchymal cells of the liver on day 12 of gestation and then die on day 13 following vascular rupture (Lohler et al., 1984). The Mov 13 locus has been cloned and the gene disrupted
by the provirus has been identified as the type-1(I) collagen gene. The provirus has integrated into the first intron of the gene inducing changes in the methylation pattern and chromatin conformation of the gene. This is thought to interfere with transcription and the synthesis of type-1 collagen in homozygous Mov 13 mice.

Homozygous Mov 34 mice die soon after implantation, at the post-blastocyst stage. The provirus has integrated within a gene that is expressed as an abundant ubiquitous transcript of 1.7 kb. The cDNA for this gene has been cloned and the putative protein encoded by this mRNA (39 kd in size) has no significant homology with proteins of known function (Soriano et al., 1987). Homozygous Mpv 17 embryos, on the other hand, appear normal phenotypically up to 2 to 3 months of age, but they later die of renal failure within 9 months of age (Weiher et al., 1990). The cDNA corresponding to a 1.7 kb mRNA from the disrupted gene has also been cloned. As in the case of Mov 34, the putative protein encoded by the cDNA has no homology to known proteins.

1.13 Mpv 20

The mutant mouse strain Mpv 20 was obtained by infection of 4 to 8 cell embryos with a defective Myeloproliferative sarcoma virus (Selinger et al., 1986, Weiher et al., 1987). Disruption of the gene (or genes) by the insertion of the provirus leads to a recessive embryonic lethal phenotype. Embryos homozygous for the Mpv 20 mutation die at the 8 cell stage as uncompacted morulae. In an effort to clone and identify the disrupted gene, DNA sequences flanking the viral integration site were cloned (figure
1.1) and a genomic fragment, highly conserved through evolution was identified by Southern blot analysis using DNA from different species. This fragment was used to screen a randomly primed testes cDNA library. The cloning and characterization of one of the genes (Unp) discovered during this screening is described in this thesis. The other candidate for the Mpv 20 gene is being characterized in the lab presently (Di Fruscio, Gupta and Gray, unpublished data).

1.2 PROTO-ONCOGENES

Oncogenes were initially thought to be components of retroviral genomes (Huebner and Todaro, 1969). However, based on several lines of evidence it is now clear that these oncogenes are mutated forms of cellular genes, referred to as proto-oncogenes. For example, proto-oncogenes are widely distributed among different species where as their retroviral counterparts are present in only one or very few species of retroviruses, with the homology between viral oncogenes and their cellular homologs is highest for the species in which the gene was first transduced (Stehelin et al., 1986, Frankel and Fischinger, 1977, Frankel et al., 1979). Moreover, the retroviral oncogenes represent an altered form (usually truncated form) of the homologous cellular gene (for example jun, fos and myb, discussed later in details). It is widely believed that retrovirus oncogenes arose by transduction of cellular proto-oncogenes, because transduction of cellular genes
**FIGURE 1.1**

Mouse Mpvl 20 locus

Mpvl 20 locus is shown along with the genomic phases flanking the retroviral insertion site. The evolutionarily conserved genomic fragment is shown by the hatched box within the EcoRI fragment. The black box shows the retroviral integration site and the arrow indicates the direction of transcription of the neo gene. The restriction endonucleases shown are: E; EcoRI, B; BamHI, K; KpnI, S; SalI, H; HindIII, Bc; BglII and Bg; BglIII. The provirus is shown at the top.
has been observed both in cell culture (Rapp and Todaro, 1978) and in animals (Neel et al., 1982, Wang et al., 1979). These transduced genes were then selected for during the course of evolution. Thus, by definition, proto-oncogenes are the normal cellular counterparts of transforming alleles or oncogenes of retroviruses.

Most of the proto-oncogenes are thought to be involved in processes which lead to normal growth and differentiation of cells. The functions of proto-oncogenes range from initiating oncogenic signals (for example, cell surface receptors in signal transduction pathways) to altering the pattern of expression of regulatory pathways (for example transcription factors). The list of proto-oncogenes that have been found to be altered in human malignancies is very long and by no means complete yet. Some of the proto-oncogenes which have been found to be altered in human malignancies are listed in table 1.1. The type of alterations listed in this table can lead to either changes in the expression pattern of the gene (or its target) or can change the biochemical properties of the gene product itself. The mutations in these proto-oncogenes can also lead to a change in the type of interactions of their gene products with other proteins in the cell, thus, changing the phenotype of the cell.

Mutant alleles of proto-oncogenes are usually dominant in nature because they can transform cells even when their normal counterparts are being expressed. However, in some cases it has been found that the mutant allele has to be expressed above a certain threshold level for it to be oncogenic. For example, in the case of RSV-
TABLE 1.1

Proto-oncogenes altered in human malignancies

A list of some of the proto-oncogenes which have been found to be altered in human malignancies. Modified from Bishop, M.J. (1991).
<table>
<thead>
<tr>
<th>Proto-Oncogene</th>
<th>Lesion</th>
<th>Neoplasm(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABL</td>
<td>Translocation</td>
<td>Chronic Myelogenous Leukemia</td>
</tr>
<tr>
<td>ERBB-1</td>
<td>Amplification</td>
<td>Squamous Cell Carcinoma, Astrocytoma</td>
</tr>
<tr>
<td>ERBB-2</td>
<td>Amplification</td>
<td>Adenocarcinoma of Breast, Ovary and Stomach</td>
</tr>
<tr>
<td>GIP</td>
<td>Point Mutations</td>
<td>Carcinoma of Ovary and Adrenal Gland</td>
</tr>
<tr>
<td>MYC</td>
<td>Translocation</td>
<td>Burkitt’s Lymphoma, Carcinoma of Lung, Breast and cervix</td>
</tr>
<tr>
<td></td>
<td>Amplification</td>
<td></td>
</tr>
<tr>
<td>L-MYC</td>
<td>Amplification</td>
<td>Carcinoma of Lung</td>
</tr>
<tr>
<td>N-MYC</td>
<td>Amplification</td>
<td>Neuroblastoma: Small Cell Carcinoma of Lung</td>
</tr>
<tr>
<td>H-RAS</td>
<td>Point Mutations</td>
<td>Melanoma, Carcinoma of Colon, Lung and Pancreas</td>
</tr>
<tr>
<td>K-RAS</td>
<td>Point Mutations</td>
<td>Melanoma, Acute Myelogenous and Lymphoblastic Leukemia, and Carcinoma of Thyroid</td>
</tr>
<tr>
<td>N-RAS</td>
<td>Point Mutations</td>
<td>Melanoma, Carcinoma of Thyroid and Genitourinary Tract</td>
</tr>
<tr>
<td>RET</td>
<td>Rearrangement</td>
<td>Carcinoma of Thyroid</td>
</tr>
<tr>
<td>K-SAM</td>
<td>Amplification</td>
<td>Carcinoma of Stomach</td>
</tr>
<tr>
<td>TRK</td>
<td>Rearrangement</td>
<td>Carcinoma of Thyroid</td>
</tr>
<tr>
<td>GSP</td>
<td>Point Mutations</td>
<td>Carcinoma of Thyroid, Adenoma of Pituitary Gland</td>
</tr>
<tr>
<td>SIS</td>
<td>—</td>
<td>Astrocytoma</td>
</tr>
<tr>
<td>SRC</td>
<td>—</td>
<td>Carcinoma of Colon</td>
</tr>
</tbody>
</table>

src gene expressed from the mouse mammary tumor virus (MMTV) promoter it was shown that the phenotype of the cells was dependent on the dose of the introduced src gene (Jakobovits et al., 1984). A distinct threshold expression was required to observe changes in cell morphology that were indicative of neoplastic potential of these cells.

The effect of the expression of an oncogene can also be altered by the expression of other cellular genes, thus inactivating the oncogene. This has been shown for the ras oncogene, where the expression of Kras-1 gene (a candidate tumor suppressor gene) was able to suppress the transformation of NIH 3T3 cells by the ras oncogene (Kitayama et al., 1989). The gene product of Kras-1 (Rap1A-p21) contains a region identical to the region of ras p21 protein which is thought to interact with GAP (Guanosine triphosphate activating protein). When tested, it was found that Rap1A-p21 bound GAP tightly in a GTP (Guanosine triphosphate) dependent manner, but it was not activated by GAP (French et al., 1990). Thus, Rap1A-p21 may suppress ras transformation by preventing the interaction between GAP and ras or ras related proteins.

It is thought that the potential of proto-oncogenes to act as oncogenes is due to the fact that their gene products are crucial parts of the biochemical pathways that dictate the phenotype of cells. These proteins can be classified into 4 major categories:

1) Growth factors - For example polypeptide hormones (like v-sis) that act on the surface of the cell.

2) Growth factor receptors - These are receptors for the growth factors (like v-erbB, v-kit, v-fms) which span the cellular membrane and act to transmit signals from
the outside of the cell to the inside.

3) Transducers - These are proteins that transfer signals from the receptors to the other parts of the cell where their effects are to be carried out. Examples of these are the cytosolic non-receptor kinases. Among these, pp60src is the best characterized prototype and will be discussed in more detail.

4) Transcription factors - These are nuclear proteins which govern the genetic changes (gene expression) in response to growth factor induced cascades.

1.21 Mechanisms Of Activation Of Oncogenes

There are several different ways in which proto-oncogenes can be activated to give rise to a transformed phenotype. These are discussed below using the most well studied genes as examples.

1.21.1 Activation Through Changes In Phosphorylation Of Specific Tyrosine Residues:

The first clue that phosphorylation at tyrosine residues can play a crucial role in converting proto-oncogenes into oncogenes came from the comparison of in vivo phosphorylation of pp60c-src and pp60v-src. It was found that in pp60c-src Tyr^{527} (a tyrosine residue near the carboxy-terminus but outside the kinase domain) is phosphorylated (Cooper et al., 1986). This tyrosine residue is absent in pp60v-src. Moreover, mutation of this Tyr^{527} to a phenylalanine increases the tyrosine kinase activity and also the transforming activity of pp60v-src (Cartwright et al., 1987, Piwnica-Worms et al., 1987,
Reynolds et al., 1987). Phosphorylation of pp60<sup>src</sup> is thought to be mediated through another protein-tyrosine kinase rather than auto-phosphorylation (Okada and Nakagawa, 1989).

A 1:1 complex has been shown to occur between pp60<sup>src</sup> and polyoma virus middle-T antigen and this complex formation activates its tyrosine kinase activity (Courtneidge and Smith, 1983). This is probably because the phosphorylation of Tyr<sup>527</sup> does not take place in the complex since the interaction between the two proteins occurs in a region proximal to Tyr<sup>527</sup> (Cheng et al., 1988a, Cheng et al., 1989). Other proto-oncogenes of the Src family also contain potential phosphorylation sites at locations analogous to Tyr<sup>527</sup> in c-src and some of these (like pp59<sup>src</sup> and pp62<sup>src</sup>) have been shown to associate with the polyoma middle-T protein (Cheng et al., 1988b, Kypta et al., 1988, Kornbluth et al., 1990). The association between these proteins and polyoma middle-T protein correlates with its transforming ability. Thus, a common mechanism for activation of these proto-oncogenes could be through the prevention or modification of phosphorylation of crucial tyrosine residues.

1.212 Through Transmission Of Signals Via GTPases:

GTPases are proteins that bind and hydrolyse GTP into GDP. The involvement of these proteins in tumorigenesis was first recognized through the study of the ras oncogenes which encode GTPases. It is now well documented that mutations in mammalian ras oncogenes cause neoplastic transformation. Among the GTPases listed
in table 1.2, p21<sup>ras</sup> is one of the best characterized GTPase. There are three mammalian ras genes (Ha-ras, Ki-ras and N-ras) which encode the p21<sup>ras</sup> protein (Barbacid, M., 1987). The GTP bound form of ras is capable of transforming fibroblasts, differentiating PC12 cells and causing maturation of Xenopus oocytes (Bar-Sagi and Feramisco, 1985, Birchmeier et al., 1985, Satoh et al., 1987). GNRPs (guanine nucleotide release proteins) and GAP (GTPase activating protein) specific for p21<sup>ras</sup> have been cloned and characterized (Wolfman and Macara, 1990, Downward et al., 1990, Trahey and McCormick, 1987).

The interactions between GTPase p21<sup>ras</sup>, GNRPs and GAP are shown in figure 1.2. The function of GNRPs is to increase $K_{\text{dis,GDP}}$ (rate of dissociation of GDP) and the function of GAPs is to increase $K_{\text{cat,GTP}}$ (rate of hydrolysis of bound GTP). The mutations in p21<sup>ras</sup> which are oncogenic show the importance of GTPase stimulation by GAP. One of the commonly found mutations of p21<sup>ras</sup> in human tumors locks the protein into its active GTP bound state. That is, that the protein can still bind GTP but it prevents GAP from increasing the $K_{\text{cat,GTP}}$ (Vogel et al., 1988). Another type of p21<sup>ras</sup> mutations (in vitro data only) have increased $K_{\text{dis,GDP}}$ (Walter et al., 1986, Feig and Cooper, 1988). This renders the protein oncogenic presumably by allowing rapid GTP induced activation.

Another level at which oncogenic regulation can take place is through mutations in G protein α chain regions. An example of this are mutations found in human pituitary tumors (Landis et al., 1989). These mutations occur in the α-chain region which is
### TABLE 1.2

**Small mammalian GTPases**

A table listing some of the mammalian GTPase proteins. Modified from Bourne et al. (1990).
### TABLE 1.2

<table>
<thead>
<tr>
<th>Small GTPase Subclass</th>
<th>Representative mammalian proteins</th>
</tr>
</thead>
</table>
| ras and ras like      | Ha-, Ki-, N-ras  
Rap proteins         |
| Ypt1 / Sec 4         | Rab 3                             |
| Rho                  | Rho C                             |

*Modified from Bourne et al., 1990*
FIGURE 1.2

p21\textsuperscript{mm} GTPase pathway

The p21\textsuperscript{mm} GTPase pathway is shown in its simplified form. Abbreviations used are: GTP; Guanosine triphosphate, GTPase; Guanosine triphosphatase, GDP; Guanosine diphosphatase, GAP; GTPase activating protein, GNRP; Guanine nucleotide release protein, K_{diss,GDP}; rate of dissociation of GDP, K_{cat,GTP}; rate of hydrolysis of bound GTP and P_i; free phosphate. Modified from Bourne et al. (1990).
FIGURE 1.2

Modified from Bourne et al., 1990
intimately involved in GTP hydrolysis. Thus, these mutations occur in the region which may serve as a built-in counter-part of the separate GAPs required for GTP hydrolysis by GTP-binding proteins like p21".

1.213 Via Altered Regulation Of Transcription Factors:

Some of the well studied transcription factors encoded by proto-oncogenes are rel, jun, fos, erb A, ets-1, ets-2, spi-1, myb and myc. Among these, c-jun and c-fos are part of the AP-1 transcription factor complex, a well characterized transcription factor (Lee et al., 1987*, Bohmann et al., 1987, Rauscher et al., 1988). The oncogenic forms of both these proto-oncogenes (v-jun and v-fos) are carried by retroviruses. v-jun is carried by the avian sarcoma virus which is capable of transforming chick embryo fibroblasts and can cause fibrosarcomas in chickens (Lewin, B., 1991). The v-fos gene is responsible for the transformation by the FBJ murine osteosarcoma virus (which induces chondrosarcomas) and by the chicken sarcoma virus (Lewin, B., 1991).

v-jun is derived from c-jun through deletion of a 27 amino acid region called the δ region, deletion of 3' UTR and some point mutations (figure 1.3A). Removal of the δ region of c-jun increases its transforming ability, the removal of the 3’ UTR increases the number of foci obtained (possibly by stabilizing the mRNA through removal of destabilizing sequences) and the introduction of point mutations does not seem to have much effect on the transforming ability of the c-jun protein (Bos et al., 1990).

Besides a few substitutions, the only major change between v-fos and c-fos is the
**FIGURE 1.3**

**Cellular proto-oncogenes vs corresponding viral oncogenes**

Schematic representation of the differences between 3 cellular proto-oncogenes and their corresponding viral oncogenes. A; *c-jun* vs *v-jun*, B; *c-fos* vs *v-fos* and c; *c-myc* vs *v-myc*. The functional domains (DNA binding, transactivation, dimerization, and negative regulation domains) present in the cellular proto-oncogenes along with the substitutions seen in corresponding viral oncogenes are also shown.
FIGURE 1.3

A

C-Jun
c-Jun
v-Jun

B

C-FOS
V-FOS

C

C-MYB
V-MYB (AMV)
V-MYB (E26)

3 Substitutions
Activation domains (A1 and A2)
Basic DNA-binding domain
Dimerization domain (Leu Zipper)

4 Substitutions
Basic region
Dimerization (Leu Zipper)

11 Substitutions
DNA-binding domains
Negative regulation domain
Activation domain
substitution of the c-terminal end (figure 1.3B). The basis for the oncogenic potential of v-fos is still unclear, but it is believed to be due to overexpression of the gene when it is carried by the retrovirus. Thus, part of the oncogenic potential of v-fos may be due to the upsetting of cellular balance between fos and jun proteins.

There are two retroviruses known to carry the mutated forms of c-myb gene (figure 1.3C), the avian myoblastosis virus and the E26 virus. The major difference between c-myb and the two v-myb forms is that c-myb contains a negative regulatory domain which is largely removed from both the v-myb proteins (Sakura et al., 1989, Weston and Bishop, 1989).

Thus, some of the ways in which proto-oncogenes coding for transcription factors have been found to be mutated so as to become oncogenic are 1) by changing the level of expression and thus upsetting the fine balance between different regulatory proteins (example c-jun, c-fos), 2) by removal of regulatory sequences from the UTRs or the coding regions of the gene (c-myb) and 3) by altering the properties of the transcription factor, that is, by converting it from a negative regulator to a positive regulator (c-rel, c-erb A, Zenke et al., 1988).

1.214 Through Proviral Insertion:

Although in theory every gene is a potential site for proviral integration and insertional mutagenesis, the number of genes which have been shown to be sites for proviral integration in tumor cells is limited. These genes include genes which had
previously been identified as components of acutely transforming viruses (like \textit{myc, myb, mos, fms, ras} and \textit{erb B}), genes whose properties seemed compatible with these genes being involved in oncogenesis (like \textit{lck, Il-2, IL-3, CSF-1, N-myc} and \textit{p53}) and novel genes that were identified on the basis of their proximity to viral sequences (for example \textit{wnt-1, int-2, pim-1} and \textit{wnt-3}). The proviral integration can activate cellular genes by either altering the function of their gene products or by changing their expression levels. The provirus can insert so as to produce fusion transcripts initiating either in the \textit{5'} or the \textit{3'} LTR (Long Terminal Repeat) using promoters present in the LTRs. This form of gene activation is called \textit{promoter insertion} and the first reported example of this is the activation of \textit{c-myc} by leucosis virus to give rise to B-cell lymphomas (Hayward et al., 1981).

In addition to promoters, the retroviral LTRs also contain enhancer elements which can regulate expression of genes downstream and upstream and in some instances in a tissue specific manner. A good example of this is the MMTV (mouse mammary tumor virus) LTR which directs the expression of genes driven by it in a hormone dependent manner specifically in mammary tissue. Since these LTR enhancers act in cis once integrated, they can affect cellular genes not only in the immediate vicinity but also genes located at large distances from the site of integration. Thus, the \textit{enhancer insertion} effect can occur over large distances in an orientation independent manner affecting transcription levels of multiple genes. The molecular mechanisms leading to these \textit{long distance effects} by retroviral LTRs are not fully understood, but it is thought that it
occurs through alterations in chromatin structure or through association with nuclear scaffold. This would also explain the clustering of common integration sites (Shih, et al., 1988). Table 1.3 lists some of these common clusters.

Oncogene activation can also occur via a combination of promoter insertion and change in the gene product. For example, in the case of *erb B* the proviral insertion dissociates the ligand binding domain from the intracellular kinase domain of the epidermal growth factor receptor (Nilsen et al., 1985). Also in the case of *myb* activation, truncation of both amino and carboxy terminal of the protein have been seen to occur (Kanter et al., 1988, Pizer and Hamphries, 1989). For some genes (rare) the preferred site for proviral insertion is the untranslated sequences. In the case of the *pim-I* gene (preferred site for integration is the 3' LTR) the integration leads to increased stabilization of the mRNA by dissociating RNA destabilizing or turnover sequences (Wingett, et al., 1991).

1.215 Through Chromosomal Translocation:

Two of the most exhaustively studied neoplasms, that show clear association between common chromosomal translocations and the location of cellular proto-oncogenes near the translocation breakpoints are, are leukemias and lymphomas. Cases of activation of oncogenes by chromosomal translocation in these human malignancies are t(9;22) translocation occurring in CML (chronic myelogenous leukemia), called the Philadelphia chromosome (Adams, 1985) and the t(8;14) translocation occurring in
TABLE 1.3

Viral integration sites clustered on mouse chromosomes

Some of the mouse chromosomes where clustering of virus integration sites have been seen are listed in this table, along with the insertion loci. Retroviruses tend to integrate into transcriptionally active regions of the chromosomes. Modified from Peter, G. (1990).
<table>
<thead>
<tr>
<th>Mouse Chromosome</th>
<th>Insertion Locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome 3</td>
<td>FIM-3, CSF-1, EVI-1, CB-1, IL2</td>
</tr>
<tr>
<td>Chromosome 4</td>
<td>DSI-1, LCK, MOS</td>
</tr>
<tr>
<td>Chromosome 7</td>
<td>H-RAS, INT-2, HST-1, FIS-1</td>
</tr>
<tr>
<td>Chromosome 9</td>
<td>FLI-1, ETS-1, SIC-1</td>
</tr>
<tr>
<td>Chromosome 11</td>
<td>EVI-2, p53, IL-3, WNT-3, HOX 2.4</td>
</tr>
<tr>
<td>Chromosome 15</td>
<td>MLVI-1, PVT-1, MIS-1, MLVI-4</td>
</tr>
<tr>
<td>Chromosome 17</td>
<td>PIM-1, PIM-2, INT-3</td>
</tr>
</tbody>
</table>

Burkitt's lymphomas (Croce and Nowell, 1986). These two translocations show two different types of mechanisms via which chromosomal translocations can lead to the transformed phenotype of cells.

In Burkitt's lymphomas, the c-myc gene is juxtaposed to the immunoglobulin heavy-chain (IgH) locus (Klein and Klein, 1985). The c-myc protein produced in this case is normal but its expression is deregulated (overexpressed) due to its new proximity to cis-acting, enhancer like elements (juxtaposed either 5' or 3') present in the IgH locus (Erikson et al., 1983, and Hayday et al., 1984). It has been shown that transgenic mouse strains expressing an activated c-myc gene have a high frequency of B-lineage tumors (Adams et al., 1985). Constitutive high level expression of the translocated, but not the untranslocated c-myc allele was seen previously (ar-Rushdi, et al., 1983) and it has also been shown that the expression of the untranslocated c-myc gene in Burkitt's lymphoma cells can be induced by TPA (12-O-tetradecanoyl-phorbol-13-acetate). This induction was independent of nascent protein synthesis (Eick, and Bornkamm, 1989).

Results from nuclear run-on experiment suggest a mechanism involving block to elongation of transcription at the end of c-myc exon 1. The expression of the translocated c-myc allele on the other hand decreased by TPA treatment but only if cyclohexamidine (an inhibitor of protein synthesis) was present in the medium simultaneously. Differential regulation of the translocated vs the untranslocated c-myc allele could also play a part in the exerted phenotype. Along with these factors some of the Burkitt's lymphoma cells contain mutations and/or loss of the p53 gene (Gaidano et al., 1991). Thus, a
combination of loss and/or inactivation of the *p53* gene, coupled with *c-myc* activation and differential regulation could lead to the transformed phenotype.

There are other genes which have been found to be activated by the IgH locus. For example, the t(8;14) translocation found in lymphoblastic leukemias (Williams et al., 1984) is almost the same at the molecular level as the t(8;14) translocation described above in Burkitt’s lymphomas. In CLL (chronic lymphocytic leukemias) a t(11;14) translocation occurs which also involves translocation breakpoints clustering near the 5’ portions of *JH* segments (present in the IgH locus) on chromosome 14 (Erikson et al., 1984). Although it is not proven, the *bcl-1* gene is thought to be involved in the oncogenicity in these tumors. Similar results were found when breakpoints from a number of follicular lymphomas with t(14;18) translocations were cloned (Bakhshi et al., 1985). The transcriptional unit near the breakpoint on chromosome 18 was designated *bcl-2*, and like *c-myc*, this gene is also deregulated (overexpressed) in follicular lymphomas carrying the t(14;18) translocation (Tsujimoto et al., 1985).

In Philadelphia chromosome a reciprocal t(9;22) translocation occurs involving chromosome 9q34 and 22q11. This translocation involves regions on chromosome 9 found 5’ of the *c-abl* gene and the *bcr* locus on chromosome 22. An abnormal 8 kb mRNA fusion transcript resulting from the joining of *bcr* sequences 5’ of the *c-abl* coding sequences is produced (Adams, 1985, Shtivelman et al., 1985). Thus, an amino terminal truncated *c-abl* protein fused to *bcr* carboxy terminal deleted protein is produced. Sequences present in the *bcr* first exon are required for transformation by *bcr-
abl fusion protein. Introduction of sequences from the first exon of bcr, upstream of the 2nd exon of c-abl induces its tyrosine kinase activity (Muller et al., 1991), presumably by removing the negative regulation through the SH₂ domain of c-abl. These sequences (which contain two SH2 domains) can bind to the SH2 regulatory domain of abl specifically (Pendergast et al., 1991). Unlike other proteins which require phosphorylserine to bind to SH2 domains, bcr exon 1 sequences do not require phosphorylserine but require phosphoserine/ phosphothreonine residues to bind to the abl SH2 domain.

Another well characterized example of transformation mediated through chromosomal translocation is APL (Acute Promyelocytic Leukemia). APL is found to be consistently associated with balanced and reciprocal translocations between the long arms of chromosome 15 and chromosome 17 (Larson et al., 1984). Cloning of the breakpoint of this translocation revealed that the PML gene (present on chromosome 15) and the RAR-α gene (present on chromosome 17) are fused together to produce fusion transcripts PML-RAR-α and RAR-α-PML (Borrow et al., 1990, Alcalay et al., 1991). The dominant effect of the PML/RAR-α fusion protein over the normal RAR or the normal PML proteins can be explained through the altered properties of these proteins. Unlike normal RAR-α which can only heterodimerize with RXRs, the PML/RAR-α fusion proteins can homodimerize as well as heterodimerize with RXRs and PML proteins thereby causing the sequestration of RXRs (Kastner et al., 1992). Thus, APL shows a different type of mechanism where translocation generates a disease specific fusion
protein (as in the case of CML, AML and Ewing’s sarcoma) but in this case the fusion protein may function by sequestering the normal cellular proteins required for the controlled growth and differentiation of cells (Warrell et al., 1993).

1.3 THE RECOMBINANT HUMAN ONCOGENE tre

The tre oncogene was cloned and identified through co-transfection of Ewing’s sarcoma DNA into NIH 3T3 cells followed by tumorigenicity assays in nude mice (Nakamura et al., 1988). This kind of strategy has also identified activation of oncogenes human ros gene (Birchmeier et al., 1986) and mas (Young et al., 1986). Thus, this strategy can be considered fruitful for identification of new proto-oncogenes that can be activated by rearrangements or amplifications. The cloning and characterization of tre is described in brief below.

1.31 Cloning And Identification Of tre

EW1 DNA (DNA from an Ewing’s sarcoma cell line) was co-transfected into NIH 3T3 cells along with a selectable neo gene and then assayed for tumorigenicity in nude mice. Primary tumors with a latency of 3 to 4 weeks were obtained. Secondary and tertiary tumors arising from one of these primary tumors were found to retain related human sequences. When these transforming human sequences (tre, transfection recombined) were cloned, it was found that the cloned sequences were rearranged from three discontinuous regions of human chromosomes (Huebner et al., 1988).
The tripartite *tre* gene is made of 3 genetic elements the positions of which were identified by usage of rodent human hybrid cells. Of these, the 5' genetic element maps to human chromosomal region 5q 23-31, proximal to the *M-CSF* gene and distal to the *ECGF* gene. The internal genetic element maps proximal to the *bcl-2* gene on chromosomal region 18q 12, and the 3' genetic element is from chromosomal region 17q 21-22. It is interesting that a probe derived from the 3' genetic element of *tre* detects a family of related sequences mapping to at least two different regions of chromosome 17. This suggests that the gene from which these sequences are derived may be part of a multigene family.

1.32 Transcription Of The Transforming Gene *tre*

To look for transcripts containing sequences present in the recombinant oncogene *tre*, poly(A) selected RNA was extracted from transfected tumors and human cells of different types. Upon northern blot analysis of these RNAs using 4 probes from different regions of the *tre* gene (Nakamura et al., 1988), a 9 kb transcript was detected in each of the tertiary transfected tumors tested. No such transcript was detected in either the parental EW1 (Ewing’s sarcoma cell line) DNA or any of the other human cell line DNAs tested.

Later, cDNAs corresponding to a novel transcription unit of the *tre* oncogene called *tre*-2 (Nakamura et al., 1992) were cloned. When probes from *tre*-2 cDNA were used for northern blot analysis of *tre* positive ED1 cells, an approximately 8.2 kb
transcript (the same size as the full length \textit{tre-2} cDNA clone) was detected. This transcript, along with a few other transcripts, was also detected in RNA from Ewing's sarcoma cells and in all other types of human cancer cells, but was absent in RNA from non-transformed human cells.

The discrepancy in the results obtained in the two studies is thought to be due to the fact that the 3' genetic element probe used in the first study (Nakamura et al., 1988) contained 5' non-coding exon which may be absent in cancer cells because of alternative splicing or alternative transcription start sites. This assumption is supported by the finding that the \textit{tre-2} cDNA extends beyond the \textit{tre} boundary.

\textbf{1.33 Features Of The \textit{tre-2} cDNAs}

The structure of the \textit{tre-2} cDNAs is shown in figure 1.4 (Nakamura et al., 1992). Sequence comparison of these cDNAs suggests the presence of two ORFs in each of the cDNAs. The ORFs in cDNA clone 213 initiate at positions 1-3 and 1163-1165, and they terminate at positions 1129-1131, and 4430-4432, respectively. The ORFs in cDNA clone 210 start at positions 1-3 and 2672-2674, and these terminate at positions 2682-2684 and 4430-4432, respectively. At the time that these cDNAs were cloned, they did not share significant homology to any known sequences in the NRBF-PIR database. In a Southern blot analysis using DNAs from hybrid cell lines, the \textit{tre-2} sequences were detected only in chromosome 17 containing hybrids. This result was further verified by \textit{in situ} hybridization analysis.
FIGURE 1.4

Schematic representation of tre-2 cDNAs

A schematic representation of the two tre-2 cDNAs, 210 and 213. The start and ends of the open reading frames (Orf) are indicated in brackets. The black boxes indicate the regions that are deleted in the 210 cDNA clone. Modified from Nakamura et al. (1992).
FIGURE 1.4

Clone 213

Clone 210

Regions deleted in clone 210

Modified from Nakamura et al., 1992
1.34 Tumorigenicity Of tre-2

The biological activity of the cloned cDNAs was determined by transfecting individual ORFs from clones 210 and 213, as well as hybrid ORF 1 from clone 210 and ORF 2 from clone 213 into NIH 3T3 cells. Transfected cells were then assayed for tumorigenicity in nude mice. Tumors with a latency period of about five weeks were observed only in mice that were injected with cells transfected with ORF 1 of clone 210. None of the other ORFs either alone or in combination were found to be tumorigenic. Thus, it was concluded that ORF 1 of clone 210 had a tumorigenic potential and that the simultaneous expression of ORF 1 and ORF 2 may abrogate the tumorigenic potential of ORF 1. The reason for this abrogation is not clear.

1.4 TUMOR SUPPRESSOR GENES

Some of the first evidence that neoplastic transformation could result from the loss of function of normal genes comes from experiments involving somatic cell hybrids. In these studies hybrid cells obtained by fusion between tumor and non-tumor cells did not produce tumorigenic cells unless they had lost one or more specific chromosomes (Stanbridge, E.J., 1976, Stanbridge et al., 1982). These results suggested that there was something present in normal cells which was capable of suppressing the tumorigenicity of tumor cells. This suppression also suggested that recessive genetic changes (which could be complemented by the normal alleles present in normal cells) were needed for
the tumor phenotype. In some cases even hybrids obtained by fusion between two different types of tumor cell lines were found to be non-tumorigenic (Wiener et al., 1974). This not only reinforced the idea of recessive changes, but also suggested the multiple nature of these changes (Wiener et al., 1974), in other words that the function (functions) capable of repressing the tumor phenotype of tumor cells can be coded for by more than one gene. The genes that can replace the defective function in tumor cells and make them non-tumorigenic are called tumor suppressor genes. It has now been shown that various tumors are acted on by different tumor suppressors because deletion of different chromosomal regions is involved depending on the type of tumor under investigation (table 1.4).

The mechanism of suppression of the tumor phenotype is not well understood. Experiments where hybrid cells were injected into animals caused the cells to cease to proliferate (Stanbridge et al., 1981), suggest that suppression of tumorigenicity of hybrid cells can be a result of superimposition of a differentiation pathway expressed in normal cells. This also suggests that deletion or alterations in genes involved in differentiation are complemented by the normal cells. The negative effect of proposed tumor suppressor genes on cell proliferation is supported by observations made by Zhu et al. (1993) and Gu et al. (1993). The first group found that cell proliferation was inhibited by p107, a gene product related to a well documented tumor suppressor retinoblastoma gene product, although the mechanism via which the two proteins function are different. Gu et al. (1993) have shown that functional pRb is required for the production and maintenance
### TABLE 1.4

**Tumor suppressor genes**

A list of some of the tumor suppressor genes. The hatched line (---) indicates that the tumor gene at that locus has not yet been identified. Modified from Bishop, M.J. (1991).
<table>
<thead>
<tr>
<th>Tumor Suppressor Gene</th>
<th>Chromosomal Location</th>
<th>Neoplasm(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Rb1</em></td>
<td>13q14</td>
<td>Retinoblastoma, Osteocarcinoma, Carcinoma of Breast, Bladder and Lung</td>
</tr>
<tr>
<td><em>p53</em></td>
<td>17q12-13.3</td>
<td>Astrocytoma, Osteosarcoma, Carcinoma of Breast, Colon and Lung</td>
</tr>
<tr>
<td><em>WT1</em></td>
<td>11p13</td>
<td>Wilm’s Tumor</td>
</tr>
<tr>
<td><em>DCC</em></td>
<td>18q21</td>
<td>Carcinoma of Colon</td>
</tr>
<tr>
<td><em>NF1</em></td>
<td>17q11.2</td>
<td>Neurofibromatosis Type 1</td>
</tr>
<tr>
<td><em>FAP</em></td>
<td>15q21-22</td>
<td>Carcinoma of Colon</td>
</tr>
<tr>
<td><em>MEN-1</em></td>
<td>11q13</td>
<td>Tumors of Parathyroid, Pancreas, Pituitary and Adrenal Cortex</td>
</tr>
<tr>
<td>---</td>
<td>11p15.5</td>
<td>Wilm's Tumor</td>
</tr>
<tr>
<td>---</td>
<td>1p36.1</td>
<td>Neuroblastoma</td>
</tr>
<tr>
<td>---</td>
<td>3p21</td>
<td>Carcinoma of Lung</td>
</tr>
<tr>
<td>---</td>
<td>3p12-14</td>
<td>Carcinoma of Kidney</td>
</tr>
<tr>
<td>---</td>
<td>16q22.1-13.2</td>
<td>Carcinoma of Liver</td>
</tr>
</tbody>
</table>

of terminally differentiated phenotype of muscle cells.

It was possible to assign chromosomal locations to tumor suppressor genes using the hybrid cell system. However, identification of genes through this method was very difficult. Tumor suppressor genes like retinoblastoma, Wilm’s tumor and neurofibromatosis genes have been identified through the study of familial cancers where the affected individuals inherit a mutant tumor suppressor gene from the affected parent and then the normal allele from the other parent is somehow (through somatic mutation) inactivated (Knudson’s theory). Knudson’s "two-hit" theory proposes that all types of retinoblastomas involve two separate mutations carried by all retinoblastoma tumor cells (Knudson, 1971). In cases of sporadic retinoblastoma both mutations occur somatically in the same retinal precursor cell and in heritable retinoblastoma one of the mutations is germinal and the second mutation is somatic.

There is another way of identifying tumor suppressor genes which does not involve familial case studies. In this approach the DNA suspected of containing a tumor suppressor gene is introduced into a tumor cell line by microcell fusion (Fournier, 1981) or transfection. These cells are then injected into nude (athymic) mice. The DNA capable of suppressing the tumor phenotype can then be used to identify the tumor suppressor gene or genes responsible for the suppression. In discussing tumor suppression genes, I will focus on two of the best characterized tumor suppressor genes the Rb-I (retinoblastoma) and p53 genes.
1.41 \textit{Rb-1, The Retinoblastoma Gene}

Once the chromosomal region 13q14 had been identified as the region most likely to contain the retinoblastoma gene, chromosomal walking was performed using flanking markers. A gene (\textit{Rb}) was cloned that was transcribed into a 4.7 kb mRNA not only in retinal cells but also in most other tissues (Lee et al., 1987\textsuperscript{a}). The critical evidence that \textit{Rb} was the retinoblastoma gene was provided when internal deletions of \textit{Rb} in retinoblastomas were shown and later it was also shown that the \textit{Rb} protein (p110) is absent in all retinoblastoma cell lines and short term retinoblastoma tumor cultures (Horowitz et al., 1990).

Different types of genetic events including small and large deletions, splicing mutations and point mutations have been found to be responsible for inactivating \textit{Rb} alleles in tumors (Horowitz et al., 1989, Kaye et al., 1990, Bookstein et al., 1990\textsuperscript{b}). Besides retinoblastomas, other tumors (although in much lower proportions) such as small cell lung carcinomas, prostate and bladder carcinomas (but not colorectal carcinomas), sporadic and familial \textit{Rb} associated sarcomas and some breast carcinomas have also been found to contain inactivated \textit{Rb} alleles (Lee et al., 1988, Horowitz et al., 1990). The more direct proof that \textit{Rb} is a tumor suppressor gene is the fact that the reintroduction of a normal \textit{Rb} gene can reduce and delay the tumorigenicity in different tumor cell lines (Huang et al., 1988, Bookstein et al., 1990\textsuperscript{b}).

Both the \textit{Rb} mRNA and protein are expressed ubiquitously in all normal fetal and adult tissues (Friend et al., 1986). Despite the ubiquitous nature of \textit{Rb} expression its
alterations are seen only in certain types of tumors. This may be explained by the
different roles that Rb may be playing in different types of cells or by the restrictions in
the type of mutations that can occur in different tissues.

1.411 Cell Cycle Dependent Phosphorylation Of pRb

The Rb-I gene encodes a phosphorylated protein localized in the nucleus (Lee et
al., 1987) and a bipartite nuclear localization signal (NLS) has been identified in its
protein (Zachksenhaus et al., 1993). This sequence is both necessary for nuclear
localization of pRb and is sufficient to localize heterologous proteins to the nucleus. pRb
can be detected in both unphosphorylated (p110) and phosphorylated (pp114) states. The
NLS sequences seem to be required for hyperphosphorylation of pRb. Analysis of the Rb
protein at different stages of the cell cycle shows that the phosphorylation state of pRb
changes in a cell cycle dependent manner (Chen et al., 1989, Buchkovitch et al., 1989,
Decaprio et al., 1989). The protein is present in the unphosphorylated form in cell cycle
stages G₀ and G₁ and in progressively more phosphorylated forms in the S and G₂/M
stages (figure 1.5). Based on several lines of evidence (which will be discussed below),
it is generally believed that phosphorylation of pRb regulates its activity. A growth
suppression domain in pRb which is known to contain the regions required for
interactions with viral oncoproteins (Hu et al., 1990) was identified. This region extends
from amino acid residue 379-928 (Qin et al., 1992). Since genetic studies suggest that
Rb acts as a negative regulator of cellular proliferation then it follows that the
**FIGURE 1.5**

Retinoblastoma protein and the cell cycle

Rb (retinoblastoma protein) is shown at different stages of the cell cycle. P represents phosphate groups. Restriction point indicates the cell cycle stage at which cells become committed to cell division. X and Y are cellular Rb binding proteins which do not necessarily associate with Rb at the same time. E2F is a cellular transcription factor. Association of CDK (cyclin dependent kinase) is thought to be mediated through cyclins. PP1 is protein phosphatase type-1 which is one of the enzymes thought to dephosphorylate Rb. Modified from Hollingsworth et al., (1993).
FIGURE 1.5

Modified from Hollingsworth et al., 1993
unphosphorylated form of \( Rb \) (\( p110 \)) is the active form which holds the cells in \( G_0 \) and \( G_1 \) stages. When \( p110 \) is phosphorylated to \( p114 \) the cells are released from this negative control and they are able to proceed through DNA synthesis and mitosis. This is supported by observations that transfection and microinjection of \( Rb \) causes cell cycle arrest in mid to late \( G1 \) phase of the cell cycle (Goodrich et al., 1991). The observation that TPA or retinoic acid induced differentiation of human leukemia cells is associated with hypophosphorylation of \( pRb \) (Chen et al., 1989), suggests that \( Rb \) may play a role in induction of differentiation. This result is in agreement with the phenotypes observed in mice deficient of \( Rb \) (Jacks et al., 1992, Clarke et al., 1992). These mice fail to develop normally with defects in developing certain types of developing tissues. However the mice do not develop retinoblastomas, suggesting that other factors besides \( Rb \) may play a role in the development of retinoblastomas, or that there are species differences between mice and humans.

1.4.12 Binding Of \( pRb \) To Viral Oncoproteins

The proposal that the hypophosphorylated form of \( Rb \) acts as a negative regulator of cell growth is supported by the binding properties of \( pRb \). The \( Rb \) protein has been shown to associate with a number of transforming proteins of DNA tumor viruses. Among these are the adenovirus E1A, SV40 large T, polyoma large T and human papilloma virus (HPV) E7 proteins (Whyte et al., 1988*, Dyson et al., 1989, Decaprio et al., 1988, Munger et al., 1989). Mutational analysis of these proteins shows that the
regions that are needed for interaction with pRb are also the regions required for
transformation by these viruses, suggesting that pRb is a cellular target for transformation
by these viral proteins (Whyte et al., 1988a, 1988b). Except for E1A, which binds to both
the unphosphorylated and the phosphorylated forms of pRb, the other proteins generally
bind to the hypophosphorylated form of Rb. This again supports the claim that
hypophosphorylated form of pRb is the active form. One of the phosphatases which may
be involved in the dephosphorylation of pRb is protein phosphatase type-1 (PPT-1α),
because pRb can associate with the catalytic subunit of this phosphatase (Durfee et al.,
1993). This association occurs in a cell cycle dependent manner and requires sequences
similar to the sequences required for the interaction between SV40 large T antigen and
pRb.

In addition to binding to pRb, E1A can also form complexes with other cellular
proteins including cyclins, cyclin E-p33\textsuperscript{\alpha2} and the pRb related proteins p107, p300 and
p130 (Pines and Hunter, 1990, Tsai et al., 1991, Faha et al., 1993, and Whyte et al.,
1989). Two conserved regions (CR1 from amino acids 30 to 60, and CR2 from amino
acids 121 to 127) of the E1A protein are essential for binding to pRb. These regions are
also required for transformation by the viral protein (Egan et al., 1988). Mutants of E1A
deleted in the CR2 domain cannot coimmunoprecipitate pRb from infected cells.
However, E1A mutants with CR1 deletion are capable of coimmunoprecipitating pRb but
at a much lower level (Egan et al., 1988). Thus, the interaction of E1A with pRb when
both CR1 and CR2 are intact is much stronger that when only CR2 is intact. The reason
for this is still unclear. It is possible that the effect of CR1 may be to provide CR2 with a proper conformation so as to bind pRb efficiently or it may be that the affinity of CR2 and its interaction is much stronger than that of CR1. Thus, E1A-pRb complexes can only be detected with an intact CR2 in these assays. It has recently been shown through competition and binding assays that each of these regions can bind pRb independently but the affinity of CR2 peptide is higher than the affinity of CR1 peptide (Dyson et al., 1992).

By comparing the amino acid sequences homologous to these two regions in other pRb binding proteins, the consensus for binding to pRb is L H E X X L X C X E where X is any amino acid residue and X is a spacer region. Mutations in these motifs not only abolish pRb binding activity but also the transforming activity of the viral oncogene (Whyte et al., 1988). There is evidence that CR1 and CR2 homologous sequences mediate stable complex formation between adenovirus E1A and HPV-E7 and cellular proteins like p130, p107 and cyclins (Dyson et al., 1992). Their results suggest that these viruses use a common mechanism of transformation which is mediated through their interaction with these cellular proteins.

The regions in pRb required for binding to oncoproteins have also been characterized, though not in so much detail. Two non-contiguous regions of pRb required for complex formation with E1A and SV40 large T antigen are located between amino acids 393 and 572 and amino acids 646 and 772 (Hu et al., 1990). These two regions in pRb are also frequent sites for mutations in naturally occurring inactivating mutations of
the Rb gene. This observation strongly suggests that these regions are required for the normal function of Rb and that the viral oncoproteins target these regions so as to prevent pRb from performing its normal functions.

1.413 Binding Of pRb To E2F

Of all the proteins that pRb can bind to (for example protein phosphatase 1, p34cdc2/p58cyclinA, RBP1 and RBP2), the association of pRb to the transcription factor E2F (Chellappan et al., 1991) is the most direct link between pRb and the control of cell proliferation. E2F is a cellular transcription factor that is a nuclear phosphoprotein and can bind DNA. It was initially identified through its involvement in transcriptional activation of the adenovirus E2 promoter (Kovesdi et al., 1986). Cellular genes like DNA polymerase α, ribonucleotide reductase, c-myc, c-myb and dihydrofolate reductase (genes which have putative functions in G1 and S phases of the cell cycle) have been found to contain sequences homologous to E2F binding site upstream of their promoter (Hiebert et al., 1991). In some cases these E2F binding site sequences have been shown to be important for the transactivation of these genes, and E2F appears to be capable of transactivating these genes (Blake and Azizkhan, 1989). Evidence that E2F is a cellular protein which binds to pRb is that pRb-E2F complexes are seen to be present in a cell cycle regulated manner as is the phosphorylation of pRb, and E2F protein copurifies and coprecipitates with pRb (Chellappan, et al., 1991). Cyclin A (another E1A binding protein) is also involved in this regulation (Mudryj et al., 1991). Interestingly, the E2F-
pRb complexes are disrupted when pRb binding viral oncogenes are present suggesting that viral proteins can effectively compete and abolish the binding between pRb and E2F.

In 1991 (Defeo-Jones et al., 1991) two pRb binding proteins (RBP1 and RBP2) were cloned based on their property to bind pRb. Both of these proteins contained the bipartite pRb binding motif which is conserved among pRb binding viral oncoproteins. In vitro transcribed and translated RBP1 and RBP2 bound to pRb (on a Gst-pRb fusion protein column) and this binding was competed away by peptides containing the pRb binding motifs but not by non-specific peptides (peptides without the consensus binding site). These results show that these two pRBPs use the same motifs as E1A and HPV-E7 to interact with pRb. The role of these proteins in modulating cell growth and differentiation is still not clear.

E2F (or a protein with properties very similar to those of E2F) was cloned by two groups independently (Helin et al., 1992, Kaelin et al., 1992) and is known by two names: RBAP-1 and RBP3. Even though the 60 kd RBP3 protein does not contain the consensus pRb binding domains CR1 and CR2, it was shown to associate with pRb both in vitro and in vivo. The 284 amino-terminal amino acid residues were shown to provide the DNA binding specificity for E2F binding sequences which are required for transactivation. The functional domain of transactivation was found to be located within the 69 C-terminal residues of the pRBP3. The same region (72 residues at the C-terminal) were also found to be sufficient for Rb pocket binding, however the absence of the amino terminal DNA binding domain led to an increase in transactivation. The
reason for this is not clear. Like most viral oncoproteins, RBP3 seems to bind to the hypophosphorylated form of pRb. Because of all these characteristics of pRBP3, it is believed that the cloned gene codes for the E2F transcription factor or a factor very closely related to it.

Based on all these properties of pRb it is evident that this protein is involved in the control of cell cycle and their proliferation. In normal cells pRb binds to cellular proteins like the transcription factor E2F, which can transactivate genes involved in cell proliferation and differentiation. Phosphorylation regulates the time of binding of pRb to these proteins. One of the models to explain how pRb can keep cells in growth arrest is that growth regulatory cellular proteins (for example E2F) are bound by pRb in G₀ and G₁ stages when pRb is hypophosphorylated (Chellappan et al., 1991). These proteins (for example E2F) are at this time not available for their normal function like transactivation of genes involved in DNA synthesis (for example the DNA polymerase gene). In response to proper signals Rb becomes phosphorylated, E2F is free to transactivate and the cells can proceed to the next stage of the cell cycle. The transformation by viral oncoproteins like EIA which also bind pRb or the mutation of Rb gene depletes the cells of pRb and the cellular proteins which would normally be bound to pRb are available for transactivation. The cell proliferates in an uncontrolled manner, thus becoming transformed. This is of course a simplified model and other proteins are probably involved in the transformation of cells.
1.42 The p53 Gene

The 375 amino acid nuclear phosphoprotein p53 was originally discovered in extracts from SV40 transformed cells as a cellular protein that complexed and coimmunoprecipitated with SV40 large T antigen (Levine et al., 1991). It is now known to complex with other viral proteins like adenovirus E1B protein (Sarnow et al., 1982) and the E6 proteins of human papilloma virus type 16 and 18 (Werness et al., 1990). The complex formation of SV40 large T with p53 stabilizes p53 protein increasing its half life from 6-20 minutes to several hours. This kind of stabilization is characteristic of mutant p53 in tumor cells (Marshall, C.J. 1991). The nuclear localization of p53 is mediated by three nuclear localization signals (Shaulsky et al., 1990). Among these NLS1 (from amino acid 313 to 322) is the most highly conserved and it shares homology with consensus NLS sequences found in other nuclear proteins like nucleoplasmin, SV40 large T and polyoma large T proteins (figure 1.6). An interesting observation about the p53 protein is that in COS cells, even in the absence of these NLS, the p53 protein can still migrate to the nucleus by complexing with other nuclear proteins (Shaulsky et al., 1990). The nuclear localization of mutant p53 protein was found to be essential for enhancement of malignant transformation of partially transformed lymphoid cell line L12 (Shaulsky et al., 1990).

After p53 cDNAs were cloned, it was found that the p53 gene could cooperate with activated ras genes in transfection studies to transform primary rodent cells (Parada et al., 1984) and was alone sufficient to immortalize normal rat cells in culture (Robinski
**FIGURE 1.6**

<table>
<thead>
<tr>
<th>Protein</th>
<th>NLS1</th>
<th>NLS2</th>
<th>NLS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>p53-CD(M)</td>
<td>PPQKKKPLDG X(43)</td>
<td>LKTKKGQ X(3)</td>
<td>RHKKTM</td>
</tr>
<tr>
<td>p53-M11(M)</td>
<td>PPQKKKPLDG X(43)</td>
<td>LKTKKGQ X(3)</td>
<td>RHKKTM</td>
</tr>
<tr>
<td>p53-M8(M)</td>
<td>PPQKKKPLDG X(43)</td>
<td>PRAFQAL X(3)</td>
<td>ESPNC</td>
</tr>
<tr>
<td>p53-H19(H)</td>
<td>PQQKKKPLDG X(43)</td>
<td>LKSKKKGQ X(3)</td>
<td>RHKKLM</td>
</tr>
<tr>
<td>p53 (RAT)</td>
<td>PQQKKKPLDG X(43)</td>
<td>PTKKKKGQ X(3)</td>
<td>RHKKPM</td>
</tr>
<tr>
<td>p53-X.LAEVIS</td>
<td>PPLPKKRLVV X(44)</td>
<td>IKPKKG X(0)</td>
<td>KKLVK</td>
</tr>
<tr>
<td>SV40-T</td>
<td>TPPKKRRKVE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polyoma-T</td>
<td>TPPKKKREDP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV40-VP1</td>
<td>APTKRKGSC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleoplasmin</td>
<td>GQAKKKKLKD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sequence comparison of various NLSs (nuclear localization signals) present in the p53 proteins. NLS1 (the major NLS) is compared to other viral protein NLSs. From Shaulsky et al. (1990).
et al., 1988). Based on these observations p53 was classified as an oncogene. Later all the clones used in these studies turned out to be mutant forms of p53 (Hinds et al., 1989). It is now well documented that p53 genes that contain mutations can act as cooperating oncogenes, falling into the same class as c-myc and other nuclear oncogenes (Land et al., 1983). Wild type p53 on the other hand acts to suppress transformation when either E1A and ras or mutant p53 and ras are used (Eliyahu et al., 1989, Finlay et al., 1989). Wild type p53, like Rb, acts as a negative regulator of cell growth. Thus, the wild type p53 (the unmutated form) protein is now classified as a tumor suppressor gene where as mutant p53 is called an oncogene.

The p53 protein acts as a homodimer; mutations in only one allele of the gene can be enough to inhibit function of the normal allele (Stanbridge, 1990). Since only mutant p53 (not the wild type) complexes with the heat shock protein hsp 70 (Sturzbecher et al., 1988) it is believed that a mutant:wild type p53 dimer when formed can complex to hsp 70 and become neutralized. The mutant p53 in this way can act as a dominant negative mutation.

P53 can act as a transcriptional activator (Raycroft et al., 1990, Fields and Jang, 1990). The transcriptional activation domain of the protein has been mapped to the amino terminal 73 amino acids of the protein (Fields and Jang, 1990). However, mutations at amino acids 135 and 215 (found in naturally occurring tumors) also destroy the transcriptional activation potential of these mutant p53 proteins. Another role of p53 is thought to be in the control of initiation of DNA synthesis. This is based on the
observation that wild type p53 can block the binding of DNA polymerase α to SV40 large T antigen (Gannon and Lane, 1987). The replication of SV40 is also blocked. Wild type p53 can bind to sequences adjacent to the SV40 origin of replication (Bargonetti et al., 1991).

Like pRb, p53 is also regulated by phosphorylation and has been shown to be a substrate of the cdc2 kinase (Bischoff et al., 1990). Not only is the phosphorylation of p53 regulated in a cell cycle dependent manner but also the amount of the protein varies during the cell cycle, being maximal at the G₀ phase of the cell cycle. Recent evidence points to a role of p53 in apoptosis (programmed cell death) mediated by E1A (Debbas and White, 1993, Lowe and Ruley, 1993). Thus, wild type p53 may guard cells against unscheduled proliferation and tumor progression induced by E1A.

p53 has been implicated in several types of human malignancies including colorectal cancers, chronic myeloid leukemias, lung brain and breast tumors and neurofibrosarcomas (Takahashi et al., 1989, Sidrasky et al., 1992, Iggo et al., 1990, James et al., 1989, Mackay et al., 1988, Kelman et al., 1989). 75-80 % of all colon carcinomas contain mutations in both p53 alleles and most of the mutations are missense mutations which produce altered proteins (Baker et al., 1989, Nigro et al., 1989, Vogelstein et al., 1988). The mutations are seen to be clustered in specific conserved regions of the gene indicating that these regions are required for normal functions of cell control. Both p53 and pRb bind to the transforming proteins of the oncogenic viral proteins and mutations in the viral proteins which prevent binding to these cellular
proteins can cause the virus to become non-transforming. Thus, it is logical to assume that these interactions play a crucial role in the mechanism of transformation by these onco-proteins.

1.5 UBIQUITINATION

Ubiquitin (Ub) is a small (76 residue) abundant protein which is the most highly conserved protein known so far, with its amino acid sequence being identical among animals and only a 3 amino acid difference between animal and yeast ubiquitin forms (Ozkaynak et al., 1987). The protein contains two well characterized functional sites 1) a lysine residue at position 48 which serves as an acceptor site for other Ub protein molecules, thus leading to the formation of long chains of ubiquitin and 2) a site in the C-terminus of the protein which mediates the formation of iso-peptide bonds with lysine amino groups on acceptor proteins (Finley, 1991). These two sites are required for the activity of ubiquitin.

The process whereby acceptor proteins become joined to ubiquitin protein is called ubiquitination. Unlike phosphorylation, which is primarily used to change the functional state of proteins, ubiquitination appears to be mainly involved in marking proteins that are targeted for degradation. However, this is not always the case because some proteins (like Histone H2A and insect actin, Rechsteiner, 1991) are quite stable even in their ubiquitinated forms. None the less, there are good correlations between Ub
mediated conversion of a protein to high molecular weight conjugates and its rapid degradation, for example for proteins like cyclins (Glotzer et al., 1991) and p53 (Scheffner et al., 1990).

1.51 The Ubiquitin Pathway

Ub is an essential part of a major cellular proteolytic pathway. The biochemical reactions involved in Ub mediated degradation of proteins is outlined in figure 1.7. The first components of this pathway are the Ub activating enzymes designated E1 (for example UBA-I in yeast). These enzymes hydrolyse ATP to form a reactive thiol ester bond with the C-terminus of Ub. Ubiquitination cannot occur unless the C-terminus of ubiquitin is activated. Next, carrier proteins or ubiquitin conjugating enzymes, called E2 enzymes (for example UBCs in yeast) shuttle activated Ub between E1 and various acceptor proteins including ubiquitin.

The third component of the pathway is the ubiquitin-protein ligase or substrate recognition protein E3 (UBR in yeast). The E3 protein does not seem to be essential since in E3 deleted yeast mutant cells the overall proteolysis is not significantly inhibited and the growth rate of these cells is only slightly impaired (Bartel et al., 1990). E2 proteins, however, are necessary because both UBC4 and UBC5 (two of the yeast E2 genes) yeast have a 2-fold decrease in degradation of analog containing proteins (which are degraded through the Ub pathway) and their growth rate is markedly reduced (Seufert and Jentsch, 1990). Another interpretation of the results concerning E3' yeast strains is
FIGURE 1.7

Representation of the ubiquitin pathway

A schematic representation of the ubiquitin proteolytic pathway. Abbreviations used are Ub; ubiquitin, \( U_n \); multiple number of ubiquitin molecules. The abbreviation used are: S; substrate, E1; ubiquitin conjugating enzyme and E2; Ubiquitin conjugating enzyme. All enzymes shown are yeast enzymes. Modified from Finley and Chau (1991).
Ubiquitin Proteolytic Pathway

Ubiquitin + ATP + Amino acids → Ubiquitin (Ub) → Ubiquitin activating enzyme or E1 (UBA 1) → Amp - Ub → E1 S - Ub → E2 - S - Ub

Ubiquitin conjugate degrading enzyme (26S proteasome)

Ubiquitin C-terminal hydrolases or deubiquitinating enzymes (UBF1-3, YUH)

Ubiquitin + Protein-Ubₙ → Ub-Protein-Ubₙ → Ubiquitin-protein ligase or E3 (UBR1)

Modified from Finley, 1991

All enzymes listed are yeast enzymes
that there are other E3 enzymes in yeast cells which have not been cloned as yet and that is why the E3 yeast cells are still viable.

The ubiquitin-protein conjugates are then subject to degradation as well as deubiquitination, that is removal of ubiquitin from the conjugate. Deubiquitination is carried out by enzymes called ubiquitin C-terminal hydrolases or deubiquitinating enzymes or Ubiquitin-specific proteases (for example UBPI-3 and YUH-I) which process ubiquitin precursors as well as regenerate free ubiquitin from ubiquitin conjugates. UBPI enzymes contain two conserved domains, the Cys domain towards the N-terminus and the His domain towards the C-terminus (figure 1.8). Among these the Cys domain is better conserved and crucial for the activity of these proteins (Baker et al., 1992). The final step of the ubiquitin mediated proteolytic pathway is the degradation of ubiquitin protein conjugates which is done by ubiquitin-conjugate degrading enzyme called the 26S proteasome. The proteasome is a large multi-subunit protein complex present in both the cytoplasm and the nucleus (Arrigo et al., 1988) and can degrade ubiquitin-protein conjugates in vitro (Waxman et al., 1987). Seufert and Jentsch (1992) have used the yeast system to show that the proteasome is an essential part of the ubiquitin system and it mediates the degradation of ubiquitin-protein conjugates in vivo.

1.52 Functions Of Ubiquitination

Protein turnover is an essential function of cell growth and control because the levels of a lot of proteins have to be modulated rapidly and precisely in response to
### Cys Domains

<table>
<thead>
<tr>
<th>Ubp 1</th>
<th>100</th>
<th>IAGLVNDGNTCFMNSVLQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubp 2</td>
<td>727</td>
<td>PTGINNIGNTCYLNSLLQ</td>
</tr>
<tr>
<td>Ubp 3</td>
<td>459</td>
<td>PRGIINRANICFMSSVLQ</td>
</tr>
</tbody>
</table>

### His Domains

<table>
<thead>
<tr>
<th>Ubp 1</th>
<th>681</th>
<th>YSLRSVIVHYGTHN-YGHYIAFRKYR---GCWWRISDETVYVVDEAEVL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubp 2</td>
<td>1185</td>
<td>YSLFSVFIHRGEAS-YGHYWYIKDRNRNGIWRKYNDTISEVQEEEVF</td>
</tr>
<tr>
<td>Ubp 3</td>
<td>844</td>
<td>YKLTGVIYHHGVSDDGGHYTADVYHSEHNK-WYRIDDVNITELEDDDVL</td>
</tr>
</tbody>
</table>
FIGURE 1.8

Comparison of Cys and His domains of Ubp proteins

A comparison of the UBP (yeast ubiquitin-specific proteases) Cys and His domains. Cys; Cysteine domain and His; Histidine domain. The starting amino acids of each domain are indicated. | indicate amino acids that are conserved between all three proteins and : indicate amino acids that are conserved between only two of the three proteins.
specific inter or intracellular signals. In addition, defective proteins that can arise as a result of mutations, transcriptional errors, translational errors and chemical or physical damage to cell must be degraded as part of the defence system of the cell. That ubiquitination plays an important part in degrading defective proteins was shown (Ciechanover et al., 1984) by the use of a murine cell line containing a thermolabile mutation of E1 (ubiquitin activating enzyme). It was found that at the non-permissive temperature, when E1 was inactivated the half life of abnormal (amino acid analog containing) proteins as well as a lot of normally short lived proteins was increased dramatically in these cells.

There is a lot of circumstantial evidence for the involvement of the ubiquitin mediated proteolytic pathway in multiple cellular processes including DNA repair, stress response, ribosome and peroxisome biogenesis, transcription, viral infection, neural and muscular degeneration, differentiation, and cell cycle control (Finley and Chau, 1991). Among these functions the best understood functions of ubiquitination are in DNA repair and cell cycle control.

Most of the data linking ubiquitination to DNA repair comes from studies done using mutants of the *RAD6* gene, which codes for the major ubiquitin conjugating activities in yeast extracts (Jentsch et al., 1987). *RAD6* is induced by DNA damage and is expressed at a high level during meiotic cycle (Madura et al., 1990). Mutations in *RAD6* lead to hypersensitivity to ultraviolet light and other DNA damaging agents like X-rays and chemical mutagens as well as inability to convert DNA lesions into
mutations. Mutations of cysteine-88 of \textit{RAD6} protein (the single cysteine residue presumed to be the site of ubiquitin linkage) leads to complete loss of \textit{RAD6} ubiquitin conjugating activity \textit{in vitro} and these mutations behave as null alleles \textit{in vivo}, suggesting that the diverse functions of \textit{RAD6} are mediated through the ubiquitin conjugating activity (Sung et al., 1990). \textit{RAD6} can efficiently add multiple ubiquitin groups to purified mammalian histone proteins (Jentesch et al., 1987). Based on this finding, it is thought that DNA repair function of \textit{RAD6} may be mediated through localized ubiquitination of proteins flanking the site of DNA damage. This could lead to alterations of chromatin structure at the site of DNA damage, or it may mark the site for binding of repair enzymes, or it may expose the damaged DNA to repair enzymes. It is also possible that \textit{RAD6} ubiquitination is capable of activating DNA repair enzymes.

In the case of cell cycle control the evidence consists of studies done with cyclins. Cyclin is a principle subunit of MPF (M-phase promoting factor) whose activation induces mitosis. The end of mitosis (marked by transition from metaphase to anaphase) is induced by degradation of cyclin, which then leads to inactivation of MPF. Cyclin degradation stops as cells enter interphase which causes cyclin accumulation and following a complex series of events MPF is activated and mitosis begins. During interphase cyclin is still ubiquitinated, but at a low multiplicity of ubiquitin (only one or two per molecule). Cyclin mutants capable of activating MPF but incapable of degradation lead to arrest of the cell cycle at mitosis.

Two conserved regions in cyclin present between residues 13-66 were shown to
be required for ubiquitin mediated degradation of cyclins (Glotzer et al., 1991). One of these is a motif RXXLXXIXN (residues 42-50) which is termed the degradation box and is the proposed recognition site for the ubiquitin conjugating enzyme. The second motif is a stretch of four lysine residues (residues 60-66) which is thought to contain the ubiquitin ligation site. These two motifs seem to be essential and sufficient to target heterologous proteins for mitosis specific ubiquitin mediated proteolysis (Finley et al., 1991). Cyclin degradation at mitosis is accompanied by (as expected) the formation of ubiquitin-cyclin conjugates.

1.53 Substrates Of The Ubiquitin Proteolytic Pathway

Rapid change in concentrations of proteins in cells requires that either the proteins be synthesized very quickly or that the already present proteins be degraded. There is now a long list of proteins including transcription factors and oncoproteins are degraded through this pathway. This confirms the previously made suggestion that degradation of proteins plays a central role in regulatory mechanisms. Another function of proteolysis is to degrade abnormal or defective proteins. This function of ubiquitin mediated proteolysis was shown quite early on by Hershko et al (1982).

Proteolytic substrates are selected through the binding of E3 enzymes (ubiquitin-protein ligase) to specific classes of N-terminal residues of substrate proteins. In 1986 the N-end rule relating the in vivo half life of a protein to the identity of its N-terminal residue was discovered (Varshavsky, 1992) and distinct versions of this rule have since
been found to operate in all organisms tested (E. coli, yeast and mammalian reticulocyte lysates). In brief, the rule says that the N-terminal residue present in a ubiquitin-X-protein conjugate (where X is the N-terminal residue) dictates if the protein is going to be metabolically stable or unstable.

The role of ubiquitin mediated proteolysis was first demonstrated in the case of phytochromes which are proteins used by plants for the measurement of day length. The next to be identified as a natural substrate for this pathway was MATα2, which is a repressor protein in yeast. Cyclins (discussed above) have now been well documented as substrates for ubiquitination. Besides the substrates mentioned above, the degradation of several oncogene products are also believed to be either directly or indirectly mediated through the ubiquitin pathway. These findings are based on studies done using rabbit reticulocyte lysates (an \textit{in vitro} system). Scheffner et al., (1990) have shown that the papillomavirus protein E6 promotes the degradation of p53 and this degradation is accompanied by the appearance of ubiquitin-p53 conjugates. Similarly, by using neutralizing antibodies against E1 (ubiquitin activating enzyme) it was shown that the proteolysis of several oncogene proteins (\textit{N-myc, c-myc, Fos and ElA}) was inhibited by the presence of these antibodies (Ciechanover et al., 1991). However, in this study no evidence for the accumulation of ubiquitin-oncoprotein conjugates was presented even in the absence of the antibodies, suggesting that the degradation of these proteins could be regulated by ubiquitin indirectly. Thus, it seems that ubiquitin may be able to effect the degradation of oncoproteins indirectly by regulating proteins involved in the degradation
of oncoproteins.
CHAPTER 2

Unp: A NOVEL GENE WITH SEVERAL CONSERVED DOMAINS AND INTERESTING FEATURES
2.1 SUMMARY

We have cloned cDNAs corresponding to mRNAs generated from a novel gene designated *Unp*. *Unp* cDNAs contain a large open reading frame which has a potential to code for a protein of 90 kDa. The predicted *Unp* protein contains several conserved domains including a putative p53 type nuclear localization signal and a bipartite consensus sequence for binding to the retinoblastoma gene product (pRb). The pRb binding consensus is present within a region with a strong homology to the human *tre-2* oncogene. *Unp* has been localized to human chromosome 3 and mouse chromosome 9 in a region of homology with human chromosome region 3p21. This region of human chromosome 3 has been implicated in a number of human malignancies.
2.2 INTRODUCTION

Retroviral insertional mutants have been used previously to identify novel genes like Mpv 17 (Weiher et al., 1990) and Mov 34 (Soriano et al., 1987). In each of these cases proviral DNA sequences were used to isolate intact host sequences adjacent to the site of integration of the retrovirus. These flanking sequences were then used to isolate and identify the gene present at that locus.

Mpv 20 is one such retroviral insertional mutation which was generated by infection of 4 to 8 cell embryos with a defective myeloproliferative sarcoma virus (Weiher et al., 1987). Mice homozygous for this mutation die at an 8 cell stage as uncompacted morulae. In order to clone the gene (genes) disrupted by the integration of proviral DNA, genomic fragments flanking the integration site were cloned (using retroviral LTR sequences as probe) and an evolutionarily conserved genomic fragment was identified (Gray, unpublished data). This conserved fragment was used to screen a randomly primed testes cDNA library and hybrid cDNAs were isolated, one of which led to the identification of a novel gene, Unp (for Ubiquitin-specific Nuclear Protease). The cloning of cDNAs from the Unp gene, main features of the corresponding mRNAs and the conserved domains in the predicted Unp protein are described in this chapter. The chromosomal location of the Unp gene in both mouse and human are also presented.
2.3 MATERIALS AND METHODS

2.31 Screening And Isolation Of cDNA Clones

A total of six independently derived cDNA libraries were screened. These libraries were: a randomly primed mouse testes cDNA library; oligo dT primed mouse testes, brain, F9 and P19 (embryonal carcinoma cell lines) cDNA libraries, and a randomly and oligo dT primed human frontal cortex cDNA library. The probe used was the evolutionarily conserved genomic fragment from the Mpv 20 insertion site for the initial screening of the randomly primed mouse testes cDNA library (Gray, unpublished data), and the hybrid cDNA clone t19 was obtained. The t19 cDNA clone was then used to screen the F9 and P19 cDNA libraries. For screening of the oligo dt primed mouse testes and brain cDNA libraries, the cDNA clone F9 isolated from the F9 cDNA library was used. The human frontal cortex cDNA library was screened with the largest mouse cDNA clone (T4) to obtain human Unp cDNA clones. Screening was carried out according to the procedure described by Maniatis et al., (1989).

2.32 Sequencing Of cDNA Clones

The cDNA inserts corresponding to the Unp message were cloned into the EcoRI site of plasmid pGEM4 (in the case of mouse Unp cDNAs) and into the EcoRI/XhoI sites of plasmid pBLUESCRIPT (in the case of human Unp cDNAs). Nested deletions were generated in the cDNAs using an Exonuclease III-Mung bean nuclease deletion kit (Stratagene). The deletion mutants were then sequenced by Sanger's di-deoxy sequencing
method (Sanger et al., 1977), using double stranded mini-prep plasmid DNA (Kraft et al., 1988). The T7 DNA polymerase sequencing kit from Pharmacia was used to perform the sequencing reactions in the presence of $^{35}$S-dATP.

2.33 Southern (DNA) Blot Analysis

DNAs were prepared by the usual method (Maniatis et al., 1989), restricted to completion using restriction endonucleases (Gibco, BRL and Boehringer Mannheim) and fractionated on a 1% agarose gel. Fractionated DNA was transferred and UV crosslinked to Hybond-N membrane (Amersham) as described by the manufacturer. Following this, the membranes were prehybridized in prehybridization solution containing 50 mM Hepes (pH 7.0), 0.75M NaCl, 50% formamide, 0.5% SDS, 2 mM EDTA (pH 8.0), 10X Denhardt ((Denhardt, 1966), 200 μg/ml Herring sperm DNA and 10 μg/ml poly(A). Hybridization was carried out in the prehybridization solution containing $^{32}$P-labelled DNA of interest. The blots were then washed at the stringencies described in the figure legends.

2.4 RESULTS

2.41 Cloning Of cDNAs For The Mouse Unp Gene

An evolutionarily conserved genomic fragment flanking the provirus insertion site in the Mpv 20 mutation was used to screen a randomly primed mouse testes cDNA library (Gray, unpublished data). Several hybrid cDNA clones were isolated and one of
these cDNAs (t19) was then used to screen cDNA libraries derived from P19 and F9 (embryonal carcinoma cell lines) poly(A)⁺ mRNA, to obtain cDNA clones F3-2, F9-1, P9B, and P11B. It is still not clear what (if any) the relationship between the Mpv 20 phenotype and *Unp* is, since some of the the t19 cDNA sequences (approximately 200 base pairs) have not been found in any of the *Unp* cDNA clones sequenced to date. Using the F9 cDNA fragment as a probe, additional cDNA clones T4 (from a mouse testes cDNA library) and 2.4B (from a mouse brain cDNA library) were isolated. The human cDNA clones (HL6-2, HL2 and HL1-1) were isolated by screening a human frontal cortex cDNA library (commercially available from Statagene) with the T4 cDNA fragment. A schematic representation of the mouse cDNA clones obtained is shown in figure 2.1.

2.42 Absence Of A Consensus Polyadenylation In *Unp* cDNAs

Initially, the cDNAs T4, F9 and 2.4B were sequenced at their ends to determine if the cDNAs represented a transcript from a novel gene or if they belonged to a previously cloned and characterized gene. The sequence obtained at the ends of these cDNAs showed no homology to any of the entries in the computer genebank databases, suggesting that *Unp* was a novel gene. These cDNAs were then fully sequenced. The sequence at the 3' end of the cDNAs (all of which were polyadenylated) showed that the 3' end of 2.4B and T4 was different from the 3' end of the F9 cDNA (figure 2.1), and contained additional 300 base pairs beyond the poly(A) tail of the F9 cDNA. The 3' UTR of the *Unp* cDNA was also unusual in that none of the cDNAs had a perfect match
FIGURE 2.1

Schematic representation of mouse Unp cDNAs isolated

A schematic representation of mouse Unp cDNAs that have been isolated and fully or partially (at the ends) sequenced. The lengths of the cDNAs is indicated in kb. A_{50} are the two degenerate polyadenylation signals. T4; cDNA isolated from a testes cDNA library, F9, F9-1 and F3-2; from F9 embryonal carcinoma cell line cDNA library, 2.4B; from brain cDNA library and P9B and P11B; from P19 embryonal carcinoma library.
<table>
<thead>
<tr>
<th>cDNA</th>
<th>Size (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4</td>
<td>3.7</td>
</tr>
<tr>
<td>F9</td>
<td>2.7</td>
</tr>
<tr>
<td>2.4B</td>
<td>2.4</td>
</tr>
<tr>
<td>F3-2</td>
<td>2.5</td>
</tr>
<tr>
<td>F9-1</td>
<td>1.8</td>
</tr>
<tr>
<td>P9B</td>
<td>1.7</td>
</tr>
<tr>
<td>P11B</td>
<td>2.0</td>
</tr>
</tbody>
</table>
to the consensus (AATAAA) polyadenylation signal (Proudfoot, N., 1991) preceding the poly(A) tails. Degenerate polyadenylation signals AGTAAA (in the case of F9) and GATAAA (in the case of T4 and B2.4) were present upstream of the poly(A) tails.

There were two possibilities with respect to the two types of 3' ends observed in these cDNAs. The first possibility was that Unp was transcribed into two types of mRNAs and that there was difference between the types of mRNAs produced in undifferentiated or partially differentiated cells (F9 and P19 cells) and adult tissue transcripts (testes and brain). The other possibility was that the difference observed between the cDNAs was due to cloning artifacts. To check for these two possibilities, other polyadenylated cDNAs (F3-2, F9-1, P9B, and P11B) were also sequenced at their ends.

The two 3' ends were represented in cDNAs obtained from both these libraries. This observation confirmed that the Unp gene is transcribed into two types of mRNA species which differed at their 3' ends. This result also rules out the possibility that undifferentiated cells (F9 and P19 embryonal carcinoma cells) express only the F9 cDNA type 3' end containing transcripts. RNase protection analysis (data not shown) showed the presence of both types of mRNAs in adult tissues including liver and brain. Preliminary data, based on comparison of genomic fragments containing the alternate 3' ends, suggested that the two 3' ends are contained within different exons. From the distribution of cDNA clones that have been obtained to date, and the intensity of the two bands obtained in RNase protection analysis, it seems that the two degenerate
polyadenylation signals are used at roughly similar frequencies.

2.43 5' UTR Of Unp cDNAs Contains Several Minicistrons

The 5' UTR of Unp is different from most conventional eukaryotic genes in that it is relatively large (more than 550 nucleotides) and it contains 11 short open reading frames (preceding the initiating AUG codon) ranging in size from 2 amino acids to 51 amino acids (fig 2.2). This is not in accordance with the scanning model for translation (Kozak, M. 1989). This model states that the 40S ribosomal subunit complex (containing Met-tRNA\textsuperscript{met} and initiating factors) binds to the 5' end of eukaryotic mRNAs and then migrates along it until it comes in contact with the first AUG codon in a favourable context where it initiates translation. Thus, the first AUG codon with the proper Kozak consensus (GCCGCCA/GCCAUGG) serves as the initiation codon.

These types of UTRs are usually present in growth regulatory genes like TGF-β and proto-oncogenes like Ick. In these genes it has been shown that removal of the upstream open reading frames from these genes leads to an increase in the translation efficiency of these genes (Marth et al., 1988, Arrick et al., 1991). The 5' UTR of Bcr/Abl also acts as an inhibitor of translation in vitro (Muller and Witte, 1989). Thus, the structure of the 5' UTR of Unp suggests that this gene may be involved in growth regulatory functions and that this gene may contain some tumorigenic potential.
FIGURE 2.2

5' UTR of Unp cDNA

A schematic representation of the open reading frames (Orf) present in the 5' UTR of the Unp gene. The length of the Orfs is indicated above the arrows which show the position of each Orf. Start of the major Orf is indicated by a dashed arrow.
Unp 5' Untranslated Region

FIGURE 2.2

←100 b→1
2.44 Features Of The Large Open Reading Frame Of Unp cDNAs

The cDNA clones T4, F9 and 2.4B, independently obtained from three cDNA libraries, were fully sequenced to obtain the sequence of Unp cDNA presented in figure 2.3. Protein analysis of the large open reading frame (DNA STAR) predicted a protein of 794 amino acids with a molecular mass of 90 kDa. The open reading frame extends from nucleotide 575 to nucleotide 2959. Although the predicted sequence of the Unp protein does not identify its function, the presence of known functional motifs (figure 2.4) in the sequence is consistent with a nuclear role of this protein. The first of these motifs is a putative nuclear localization signal which is very similar to the main nuclear localization signal (NLS1) found in the p53 protein (Shaulsky et al., 1990). The NLS1 of p53 protein can direct the localization of p53 as well as heterologous proteins to the nucleus.

The second motif is a bipartite pRb binding consensus amino acid sequence which is highly conserved among proteins (like adenovirus E1A, SV40 large T, and human papillomavirus type 16 and 18 E7 protein) that physically associate with the retinoblastoma protein (Hu et al., 1990, Phelps et al., 1988). The motif has a consensus LHE Xa LXCXE, where X is any amino acid and Xa is a variable spacer region (Defeo-Jones et al., 1991). The putative Rb binding motif is identical to that found in adenovirus 5 E1A and HPV16 E7 (figure 2.5). The presence of this motif in the predicted sequences strongly suggests that Unp may be a cellular pRb binding protein like RBP-1, RBP-2 and RBAP-2 (Defeo-Jones et al., 1991, Kaelin et al., 1992). The Unp protein also contains
FIGURE 2.3

Sequence of the longest mouse Unt cDNA, T4

DNA sequence of the longest mouse Unt cDNA clone T4. The sequence was deduced from nested deletions made using the Exonuclease III/ Mung bean deletion kit (Stratagene). The region of homology to the tre oncogene is indicated by ♦. The circled amino acids are the putative pRb binding pockets, CR1 and CR2. Boxed amino acids are the putative p53 type nuclear localization signal. The two degenerate polyadenylation signals are underlined. Nucleotide numbers are indicated on the left and the amino acid numbers on the right.
FIGURE 2.4

Characteristic features of the *Unp* cDNA

A schematic representation of the *Unp* cDNA and its predicted protein. The abbreviations used are listed at the bottom of the figure.
Structure Of The Unp cDNA

CR1 & CR2  \( Rb \) Binding Consensus
NLS     Nuclear Localization Signal
UTR     Untranslated Region
MRKKXXXL  p300 Binding Consensus
UPC    Ubiquitin Protease Consensus
A(n) Degenerate Polyadenylation Signals
consensus sequence for binding to p300 is the presence of the amino acid sequence MRKXXXL at the N-terminus of the protein, where X is any amino acid (Wang et al., 1993). Thus, Unp protein also has the potential to bind to the p300 protein.

The Unp open reading frame also contains the cysteine and histidine regions that are found to be conserved between the ubiquitin-specific proteases (UBP1, UBP2 and UBP3) of Saccharomyces cerevisiae (Baker et al., 1992, Tobias and Varshavsky, 1991), and are required for the activity of these proteins. A comparison of these regions among the proteins is shown in fig 2.6. The cysteine (Cys) domain extends from amino acid 136 to 151 in Unp and the histidine (His) region extends from amino acid 696 to 731 in Unp. Among the two domains, the cysteine domain is the one that is more highly conserved. The two domains are correctly positioned with relation to these domains in the Ubp proteins, with the cys domain being towards the N-terminus and the his domain towards the C-terminus. The presence of these conserved regions in the predicted Unp protein suggests that Unp may code for a ubiquitin-specific protease. If this is the case, then Unp would be the first non-yeast Ubp family member to be cloned.

2.45 Unp Is Located On Mouse Chromosome 9

The murine chromosomal location on Unp was determined in collaboration with Dr. Jenkins and co-workers by interspecific backcross analysis using progeny derived from matings of [(C57BL/6J × Mus spretus)F1 × C57BL/6J] mice (Copeland and Jenkins, 1991). The mapping results showed that Unp is located in the distal region of
Retinoblastoma protein (pRb) binding consensus amino acid sequences from *Unp* are compared to the pRb binding sequences present in two viral proteins known to associate with pRb. HPV; human papilloma virus.
FIGURE 2.5

<table>
<thead>
<tr>
<th></th>
<th>Ad5 E1A</th>
<th>HPV16 E7</th>
<th>UNP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HFEPPTLHE</td>
<td>HGDTPTLHE</td>
<td>AFILDGLHE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(...)</td>
<td>(...)</td>
<td>(...)</td>
</tr>
<tr>
<td></td>
<td>LTCH</td>
<td>LYCY</td>
<td>LVCP</td>
</tr>
<tr>
<td></td>
<td>(...)</td>
<td>(...)</td>
<td>(...)</td>
</tr>
</tbody>
</table>
**FIGURE 2.6**

Comparison of Cys and His domains of *Ubps* and *Unp*

A comparison between the Cys and the His domains of mouse *Unp* protein and yeast *UBP* (ubiquitin-specific protease) proteins. Cys; Cysteine domains and His; Histidine domains. | indicate the amino acids that are identical and : and . indicate conservative changes.
### Cys Domains

<table>
<thead>
<tr>
<th>Type</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubp</td>
<td>GLENLGNCFMNSELQ</td>
</tr>
<tr>
<td>Unp</td>
<td>GLGNLGNCFMNSELQ</td>
</tr>
</tbody>
</table>

### His Domains

<table>
<thead>
<tr>
<th>Type</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubp</td>
<td>YSLFGVIHHRGEASGGHYTAYIKDRNRNGIWRRIIDD</td>
</tr>
<tr>
<td>Unp</td>
<td>YDLIAVSNHYGAMGVGHYTAYAKNRLNGK-WYYFDD</td>
</tr>
</tbody>
</table>
Jenkins, 1991). The mapping results showed that \textit{Unp} is located in the distal region of mouse chromosome 9 linked to \textit{Ttf}, \textit{Gnai}-2 and \textit{Cck} genes (figure 2.7). The distal region of mouse chromosome 9 shares a region of homology with human chromosome 3, which suggests that the human homolog of \textit{Unp} will reside on chromosome 3 as well. Loss of heterozygosity (LOH) on chromosome 3p has frequently been detected in several human malignancies, including lung carcinomas (Kok et al., 1987), breast (Ali et al., 1989), ovary (Ehlen and Dubeau, 1990), testes (Lothe et al., 1989), kidney (Kovacs et al., 1988) and uterine cervical cancers (Kohno et al., 1993). A tumor suppressor gene(s) is thought to reside in this region of chromosome 3, but this (these) gene(s) remains to be cloned. The chromosomal location of \textit{Unp}, presence of a putative nuclear localization signal and an \textit{Rb} binding consensus in \textit{Unp} protein suggest an intriguing possibility that \textit{Unp} may be a tumor suppressor gene at this locus on chromosome 3.

2.46 \textit{Unp} Is A Member Of A Gene Family, Conserved Through Evolution

Insert from the T4 cDNA (longest mouse \textit{Unp} cDNA clone) was used to probe a Southern blot containing DNAs from different metazoan species (figure 2.8) including mouse, human, rat, dog, pig, chicken and Xenopus. The probe detected bands in mammalian DNA samples, for example in mouse (lane 1) and human (lane 2), as well as in the DNAs from avians (chicken, lane 6), and amphibians (Xenopus, lane 7). \textit{Unp}-related sequences were not detected in DNA from \textit{Drosophila melanogaster} (lane 9).

Interestingly, multiple bands were detected in mammalian DNAs and some of
FIGURE 2.7

Chromosomal localization of Unp to mouse chromosome 9

Mapping of the Unp gene to the distal region of mouse chromosome 9 by interspecific back-cross analysis. The segregation patterns of Unp and flanking genes in 151 back-cross animals that were typed for all loci are shown at the top of the figure. Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J X M. spreitus) F1 parent. ■ represent the presence of a C57BL/6J allele, and □ represent the presence of a M. spreitus allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. A partial chromosome 9 linkage map showing the location of Unp in relation to linked genes is shown at the bottom of the figure. Recombination distances between loci in centimorgans are shown to the left of the chromosome and the position of loci in human chromosome (where known) are shown to the right.
FIGURE 2.7

Trf  ■ ■  ■  ■  ■  ■  ■  ■  ■
Gnai-2  ■  ■  ■  ■  ■  ■  ■  ■  ■
Unp  ■  ■  ■  ■  ■  ■  ■  ■  ■
Cck  ■  ■  ■  ■  ■  ■  ■  ■  ■

69  59  4  1  3  0  8  7

9

3.6  Trf  3q21
1.8  Gnai-2  3p21
9.6  Unp

9.6  Cck  3pter-p21
FIGURE 2.8

Zoo blot analysis of the Unp gene

Southern blot analysis using DNA from various species. The DNAs were restricted with EcoRI. The probe used was a $^{32}$P-labelled T4 cDNA fragment and the blots were washed in 0.5X SSC and 0.1% SDS at 65°C. DNAs used were Lane 1; BALB/c mouse, lane 2; human primary fibroblast foreskin, lane 3; pig DNA, lane 4; dog, lane 5; rabbit, lane 6; chicken, lane 7; Xenopus, lane 8; lamprey, and lane 9; drosophila.
these bands were submolar in intensity, suggesting that Unp may belong to a family of related genes. To investigate this possibility, two parallel Southern blots containing mouse DNA cut with different restriction endonucleases (BamHI and EcoRI) were hybridized to the T4 cDNA fragment. These blots were then washed at different stringencies. At lower stringency (lanes 3 and 4), additional Unp-related bands were detected (figure 2.9). This result is consistent with the Unp gene being a member of a multigene family.

2.47 Unp Shares Homology With The Recombinant Human Oncogene tre

When the predicted amino acid sequence of Unp was compared to entries in the PIR and SwissProt computer databases, it was found to have significant homology (between amino acids 96 and 393 of Unp protein) to the recombinant human oncogene tre, with 42.2% identity in a 297 amino acid overlap (figure 2.10). The tre oncogene has arisen as a result of recombination of transfected DNAs resulting in a fusion of portions of three human chromosomes: chromosomes 5, 17 and 18 (Huebner et al., 1988). The homology between Unp and tre resides in the portion of tre derived from human chromosome 17 (tre-2, Nakamura et al., 1992). This is the region of tre which is essential for transformation by this oncogene, suggesting that Unp may be a proto-oncogene. In Unp, the region of homology contains the putative Rb binding domain. The homology between the two genes raises two possibilities: one is that Unp is a murine homolog of the tre-2 oncogene, and the other possibility being that both Unp and tre are
Parallel Southern blots containing mouse DNA restricted with BamHI (lanes 1 and 3) and EcoRI (lanes 2 and 4). One of these blots (lanes 1 and 2) was washed in 0.1X SSC and 0.1% SDS at 65°C and the other blot (lanes 3 and 4) was washed in 2X SSC and 0.1% SDS at 65°C. Open circles (○) denote the fragments that are detected only at lower stringency and the closed circle (●) denotes a fragment whose intensity is increased at lower stringency.
<table>
<thead>
<tr>
<th>kb</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 2.10

Homology between *Unp* and *tre-2*

Alignment of *Unp* and *tre-2* amino acid sequences. The *Unp* sequence appears above, with the *tre-2* sequence below. Identical amino acids are repeated between, and conservative changes are indicated by dots.
related members of a gene family. The later possibility is supported by the observation that Southern blot analysis using both Unp (figure 2.9) and tre sequences (Huebner et al., 1988) suggest that they belong to gene families.

2.48 Unp Is Derived From A Gene Distinct From tre-2

The chromosomal mapping of Unp to a region of mouse chromosome 9 suggested that Unp will be present on human chromosome 3, rather than human chromosome 17 (the chromosomal location of tre-2). This led to the hypothesis that Unp gene is distinct from tre-2. To test this hypothesis, human Unp cDNAs were isolated by screening a commercially available human frontal cortex cDNA library. The identity of the cDNAs was confirmed by partial sequence analysis as well as by Southern blot analysis. An internal EcoRI fragment of the human Unp cDNA was used to hybridize a Southern blot containing DNAs from primary human foreskin fibroblast, mouse melanoma cell line (B78), and somatic cell hybrid cell lines (Chevrette, manuscript in preparation) containing either human chromosome 3 (the putative site of Unp) or human chromosome 17 (site of human oncogene tre-2) in a mouse melanoma (cell line B78) background. Human specific band was detected in the hybrid cell line DNA containing chromosome 3 (fig 2.11, lane MC185), but not in the hybrid cell line DNA containing human chromosome 17 (fig 2.11, lane MC197). This result clearly excludes the possibility that Unp is a murine homolog of tre-2.
FIGURE 2.11

Southern blot analysis of hybrid cell lines

Southern blot analysis of microcell hybrid DNAs restricted with BamHI. Lane B78 contains DNA from B78 mouse melanoma cell line, lane HFF contains DNA from cultured human primary foreskin fibroblast culture, MC185 lane contains DNA from B78 cells carrying human chromosome 3 and MC197 lane contains DNA from B78 cells carrying human chromosome 17. The blot was probed with a 32P-labelled 1.6 kb EcoRI fragment from the Unp cDNA. The arrow denotes a human specific Unp fragment that is present only in HFF and MC185 DNAs. A non-informative 6 kb fragment was common to both mouse and human DNAs.
FIGURE 2.11

MC 197 (Ch 17)
MC 185 (Ch 3)
HFF
1378

kb
- 12
- 8
- 6
2.5 DISCUSSION

I have described the cloning of cDNAs corresponding to a novel gene *Unp* (for Ubiquitin-specific Nuclear Protease). The sequence of *Unp* cDNAs (figure 2.3) shows two interesting features: 1) the presence of two types of cDNAs suggesting the usage of two different degenerate polyadenylation signals, which would generate two species of mRNAs differing in length by approximately 300 bases at their 3' ends. Such degenerate polyadenylation signals have been found in other eukaryotic genes (Brinstiel et al., 1985). It is still not clear what function (for example mRNA stability) these alternate 3' ends of the *Unp* gene mRNA serve.

The second feature of the *Unp* cDNAs is the presence of a long (over 550 base pairs) 5' untranslated region which contains several minicistrons some of which are present in a better Kozak consensus (Kozak, 1989) than the *Unp* protein initiation codon (figure 2.2). These types of 5' UTRs are generally found in genes which have a growth regulatory function (like *TGF-β*), or in proto-oncogenes (like *lck, abl, mos etc.*), or in recombinant oncogenes (like *bcr/abl* fusion transcript present in Philadelphia chromosome patients). In all these cases, the removal of the 5' UTR leads to an increase in the efficiency of translation of these mRNAs (Arrick et al., 1991, Marth et al., 1988, Muller and Witte, 1989). Thus, it will be very interesting to see if the 5' UTR serves similar functions in the *Unp* gene.

The predicted *Unp* protein is also very interesting because it contains several conserved amino acid sequences (figure 2.4), including a p53 type nuclear localization
signal (shaulsky et al., 1990), a bipartite motif for binding to the pRb protein (Dyson et al., 1992), and homology to a known recombinant human oncogene tre-2 (Nakamura et al., 1992). The homology between Unp and tre-2 (figure 2.10) lies within a region of tre-2 which is essential for the tumorigenicity of this gene. All of the above mentioned features of Unp (the 5' UTR, putative nuclear localization signal, Rb binding consensus and the homology to tre-2) point towards an intriguing possibility that Unp may be a proto-oncogene.

The other conserved sequences present in Unp are the Cys and the His domains found to be conserved in yeast UBP type ubiquitin-specific proteases (Baker et al., 1992). The positioning of these motifs is also similar to the one found in UBP proteins, that is that the Cys domain is present towards the N-terminus and the His domain is present towards the C-terminus. This feature of Unp suggest that a possible function of Unp may be deubiquitination.

Even though Unp and tre-2 share homology and Southern blot analysis shows that they both belong to gene families, I have clearly ruled out that Unp is a murine homolog of the human oncogene tre-2 by the use of somatic cell hybrids (figure 2.11). This analysis shows that the two genes are located on different human chromosomes. Although not proven, it is possible that they both belong to the same gene family and perform similar functions. This again supports the suggestion that Unp may be a proto-oncogene.

The chromosomal location of Unp on the distal region of mouse chromosome 9 (figure 2.7) in a region that shares homology with human chromosome region 3p21. This
suggests that the human homolog of Unp will be located in the human chromosome region 3p. This region of the human chromosome 3 is frequently deleted or rearranged in a number of human malignancies including lung (Kok et al., 1987), breast (Ali et al., 1989) and kidney (Kovacs et al., 1988). A yet to be identified tumor suppressor gene (genes) is thought to be present in this region. If the human homolog of Unp is located in this region then it is possible that Unp may be affected in these malignancies. This would make Unp a candidate for the tumor suppressor gene at this locus. All these possibilities remain to be tested. I have already shown that Unp is located on human chromosome 3, but its precise location remains to be determined.
CHAPTER 3

THE UBIQUITOUSLY EXPRESSED *Unp* GENE CODES FOR

A NUCLEAR PROTEIN
3.1 **SUMMARY**

The *Unp* gene which was identified during a survey for candidate genes for the Mpv 20 mutation is related to the recombinant human oncogene *tre-2*. This gene was found to be ubiquitously expressed in all mouse tissues and cell lines tested and at all stages of development. The *Unp* mRNA runs as a doublet at about 3.5 and 3.7 kb in size which is in agreement with the sequencing data presented in the previous chapter. Based on computer analysis the predicted size of the *Unp* protein is 90 kDa. Using a polyclonal antibody generated against a synthetic *Unp* peptide, the *Unp* protein detected migrates at about 180 kDa is detected. This protein is localized primarily in the nucleus. The difference between the predicted and the observed size of the *Unp* protein suggests that this protein may be post-translationally modified.
3.2 INTRODUCTION

During a survey for cDNAs derived from genes surrounding the proviral insertion site in the Mpv 20 mouse mutation (Gray, Weiher and Jaenisch, unpublished data), we have identified a novel gene designated Unp (for Ubiquitin-specific Nuclear Protein). The predicted Unp protein contains a putative p53 type nuclear localization signal (Shaulsky et al., 1990), the CR1 and CR2 motifs which mediate the interaction of several proteins to the retinoblastoma protein (DeFeo-Jones et al., 1991), and the cys and his domains which are conserved in yeast UBP type ubiquitin-specific proteases (Baker et al., 1992).

The region of Unp containing the pRB binding motifs shares homology to a recombinant human oncogene tre-2 (Nakamura et al., 1992). However based on the chromosomal localization of Unp (human chromosome 3) and tre-2 (human chromosome 17), it is clear that Unp is not a homolog of tre-2, but rather the two genes belong to a larger family of genes. In this chapter, I have described the expression pattern of the Unp gene, both at the level of mRNA and at the level of protein. The sequence of Unp cDNAs obtained from independent cDNA libraries suggested that Unp is expressed as two types of mRNA species differing at their 3' ends. We wanted to see if this were the case. Also, we wanted to know if the Unp protein is localized in the nucleus, as predicted by the presence of the putative nuclear localization signal.
3.3 MATERIALS AND METHODS

3.31 Constructs

For in vitro transcription and translation of the Unp protein the EcoRI fragment containing the entire Unp cDNA sequences (T4) or a 5' truncated Unp cDNA (F9) were cloned into the EcoRI site of plasmid pGEM4 (Stratagene) in both orientations.

Gst-Unp fusion constructs were made to produce a Gst-Unp fusion protein in bacteria (figure 3.1). Amino terminal deleted Unp open reading frame sequences (amino acids 230 to end of open reading frame) were cloned in frame with the Gst (bacterial glutathione transferase gene) coding sequences in the expression vector pGEX-3 (Smith and Johnson, 1988) as described below. An approximately 1.9 kb Nhe I, Dra I fragment was backfilled using the Klenow fragment of DNA Pol I and then ligated into the Sma I site of the vector pGEX-3. The ligation was transformed into competent TG-1 bacteria. Ampicillin resistant bacterial clones were analyzed for the presence and the orientation of the insert Unp sequences. Both sense and antisense orientation Gst-Unp constructs were used for analysis.

3.32 Tissue Culture And Cell Lines Used

Mouse NIH 3T3 cells were grown and maintained in Dulbecco’s modified Eagle medium (Gibco) with 10% heat inactivated calf serum (Gibco, BRL). Transfections were performed by the calcium phosphate precipitation method (Graham and Van der Eb, 1973). In brief, cell were seeded at $5 \times$
**FIGURE 3.1**

Schematic representation of *Gst-Unp* fusion constructs

*Gst-Unp* fusion constructs designed to produce fusion proteins in bacterial cells. *LacI* is the lactose repressor gene from *E. coli* which makes the production of these proteins inducible by IPTG. *pLac* is the *E. coli* lactose promoter and *X*, is factor X recognition site which can be used to cleave the fusion protein so as to obtain bacterially produced *Unp* protein. AS; antisense orientation of the fusion construct and S; sense orientation of the fusion construct.
**FIGURE 3.1**

\[
\text{Lac I} \quad \text{pLac} \quad \text{Gst} \quad \text{Xa} \quad \text{Unp (s)}
\]

\[
\text{Lac I} \quad \text{pLac} \quad \text{Gst} \quad \text{Xa} \quad \text{Unp (as)}
\]

\[
\text{Xa} \quad \text{Factor Xa recognition site}
\]

\[
\text{Gst} \quad \text{Glutathione-S Transferase}
\]

\[
\text{pLac} \quad \text{Lactose promoter of E. coli}
\]

\[
\text{Lac I} \quad \text{Lactose Repressor gene of E. coli}
\]
10^5 cells in a 100 mm tissue culture dish one day prior to transfections. Appropriate construct plasmid DNAs were combined in a ratio of 100:1 with selectable plasmid DNA and mixed in the transfection mix. The DNA mixture was allowed to precipitate at room temperature for 20 minutes, prior to adding it to the cells. After 8 hours, the precipitate containing medium was removed and replaced with fresh medium. The cells were then incubated for 24 hours (or until confluent). 5 X 10^5 transfected cells were seeded in selective medium containing either 400 µg/ml of G418 (when Pgk-neo plasmid was used as a selectable plasmid) or 2 µg/ml puromycin (if Pgk-puro plasmid was used as the selectable plasmid, Vara et al., 1986). The selectable plasmids Pgk-puro and Pgk-neo were generously provided by Dr. M.W. McBurney. The cells were fed twice a week with selection medium and resistant colonies were picked. Freezing medium for cells consisted of DMEM medium containing 10% serum and 10% DMSO.

3.33 Northern (RNA) Blot Analysis

Approximately 5 µg of total RNA, prepared by the usual method (Maniatis et al., 1989) from tissues and cell lines of interest was resolved on a 1 % agarose gel containing formamide. The gel was transferred and cross linked to Hybond-N membranes (Amersham) as described by the manufacturer. Prehybridization and hybridization was performed as described for Southern blot analysis in chapter 2. The blots were washed in 0.2X SSC and 0.1% SDS.
3.34 Expression Of Bacterial *Gst-Unp* Fusion Protein

Bacterial clones containing the vector sequences alone as well as the *Gst-Unp* sequences in the sense and antisense orientation were used (figure 3.1). To obtain protein extracts, bacteria containing the desired plasmids were grown to mid log phase (OD\_600 0.6 to 0.8) and then the expression of fusion proteins was induced by the addition of IPTG (isopropyl-B-D thiogalactosidase) to a final concentration of 0.1 mM. The bacteria were allowed to grow for 2 to 3 hours at 37°C. The cells were pelleted and resuspended in PBS (150 mM NaCl, 16 mM NaH\_2PO\_4, and 4 mM Na\_2HPO\_4) containing 1% triton X-100. The cells were lysed on ice by mild sonication and centrifuged at 10,000 X g for 5 minutes at 4°C. The pellet or supernatants containing the fusion protein of interest was used for western blot analysis. In some cases the fusion proteins were solubilized in 6M urea followed by gradual dialysis in decreasing concentrations of urea and then these soluble proteins were used for further analysis.

3.35 Generation And Affinity Purification Of Unp Antibodies

A multiple antigen peptide (MAP) representing eight copies of 14 amino acids (LGSFPGSDGGVKLS, corresponding to positions 764-777 of the predicted Unp protein) attached to an octameric lysyl core (Tam, J.P., 1988) was injected into sterile rabbits as described: a) On day 1, 500 μg of the Unp MAP peptide was mixed with an equal volume of Freund's complete adjuvant and injected intramuscularly. b) On day 7, 500 μg of Unp MAP peptide was mixed with an equal volume of incomplete adjuvant and
injected into the rabbit intraperitoneally. c) On day 21, 500 µg of the peptide was injected as on day 7. d) On day 45, 50 µg peptide immunization was given as on day 7.

The rabbits were bled six weeks from the day of start of injection of the peptide, the serum was collected and the immunoglobulins (polyclonal antiserum) were precipitated by incubation with 50% NH₄SO₄ overnight at 4°C. After centrifugation the pellet was resuspended in phosphate buffered saline (PBS) and then dialysed against PBS extensively. Unp-MAP peptide was coupled to cyanogen-bromide activated sephrose 4B (Pharmacia) to saturation, according to the manufacturer's instructions and a column was made. The column was washed with 10 column volumes of 10 mM Tris (pH 7.5), 10 column volumes of 100 mM glycine (pH 2.5), and 10 column volumes of 10 mM Tris (pH 8.8) until the pH of the wash reached 8.8. The column was then rinsed with fresh 100 mM triethylamine (pH 11.5) followed by wash with 10 mM Tris (pH 7.5) until the pH of the column reached 7.5. The polyclonal serum was then passed through the column at a slow rate three times to ensure complete binding. The column was washed with 20 bed-volumes of 10 mM Tris (pH 7.5), and then with 20 bed-volumes of 500mM NaCl, 10 mM Tris (pH 7.5).

Specific anti-Unp antibodies were eluted off the column by passing 10 bed-volumes of 100 mM glycine (pH 2.5). 10 fractions of the elute were collected in a tube containing 0.1 bed-volume of 1M Tris (pH 8.0). The eluted fractions containing protein (antibody) were identified using the Bradford's test (Bio-Rad). Antibody containing fractions were pooled and dialysed against PBS containing 0.02% sodium azide.
3.36 Western (Protein) Blot Analysis

NIH 3T3 cells were lysed in lysis buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2mM EDTA, 1% Triton, 200 µg/ml phenyl-methylsulfonyl fluoride, 2 µg/ml aprotinin and 5 µg/ml leupeptin) and sonicated on ice three times (20 seconds each time). The lysates were cleared by ultracentrifugation at 100,000g. Supernatants were then resolved on 8% SDS-polyacrylamide gels and then electro-transferred to nitrocellulose membrane (Schleicher and Schuell) in 25 mM Tris (pH 7.0), 200 glycine and 20% methanol buffer at 0.7 amps for 90 min. to 2 hours. Membranes were blocked in Tris buffered saline (TBS, 140mM NaCl, 10 mM Tris, pH 7.5) containing 0.05% tween 20 and 5% non-fat dry milk powder for one hour at room temperature. Affinity purified antibody, or antisera of interest was then added to the blocking buffer at the desired concentration and the membranes were incubated for another hour at room temperature. Blots were rinsed vigorously three times (5 minutes each time) in Tris-buffered-saline-tween (TBST) before incubating for 30 minutes with a 1:3000 dilution of a secondary antibody, conjugated to either alkaline phosphatase or to horse radish peroxidase (ICN Biologicals). The blots were once again washed vigorously in TBST as before and developed using the ECL western blotting detection solutions (Amersham) in the case of horse radish peroxidase conjugated antibody, or the blots were developed using BCIP and NBT reagents (Gibco, BRL) according to the manufacturer's instructions.
3.37 Nuclear Fractionation Of NIH 3T3 Extracts

NIH 3T3 cells were resuspended in hypotonic buffer (10mM HEPES pH 7.9, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 10 mM KCl, 200 µg/ml phenylmethylsulphonyl fluoride, 2 µg/ml aprotinin, and 5 µg/ml leupeptin) and placed on ice for 20 minutes. The cells were homogenized twice using the Dounce homogenizer. Nuclei were pelleted at 1200 rpm in a table top centrifuge at 4°C. The supernatant containing cytoplasmic proteins was cleared by centrifugation at 100,000g. The nuclear pellet was resuspended in lysis buffer and cleared as for the cytoplasmic fraction. The samples were analysed as in the case of whole cell lysates.

3.38 In Vitro Transcription-Translation Reactions

The plasmids containing the Unp cDNAs (T4 and F9) in both orientations were linearized with the enzyme Ssp I. 1 µg of each of these linearized plasmids were transcribed in vitro using the Riboprobe System II kit (Promega) as follows: 1 µg of the linearized plasmid was incubated with 500uM of each of the four NTPs, 10 mM DTT, 1 unit of Rnasin, and 10 units of T7 or the SP6 RNA polymerase (Promega), in the buffer supplied at 37°C for 1 hour. Two units of DNase I were added to the reaction and incubated for another 30 minutes. The reaction was then extracted with phenol-chloroform-isoamyl alcohol (25:24:1) and precipitated with 1M NH₄SO₄ and 2 volumes of ethanol at -20°C for 2 hours. Following centrifugation, the RNA pellet was rinsed with 70% ethanol and resuspended in DEPC (diethylpyrocarbonate) treated water. in vitro
transcribed RNA was translated in vitro using a rabbit reticulocyte lysate in vitro translation system (Bethesda Research Laboratories) in the presence of $^{35}$S-methionine. After a one hour incubation, the reactions were mixed in an equal volume of loading dye (50 mM Tris, pH 6.8, 10% glycerol, 5% SDS, 5% B-mercaptoethanol and 0.05% bromophenol blue) and resolved by SDS-polyacrylamide gel electrophoresis.

3.4 RESULTS

3.41 Unp mRNA Is Ubiquitously Expressed

The T4 cDNA fragment was used in northern blot analysis using RNA from different mouse tissues and cell lines (figure 3.2). The probe detected a pair of closely migrating mRNA species in all RNA samples. Based on ribosomal RNA markers, the estimated size of the Unp transcripts is approximately 3.5 to 3.7 kb. These results are in agreement with our previous results where we isolated two types of Unp cDNAs containing different lengths of the 3’ untranslated regions. The doublet can be seen better in figure 3.3 due to better resolution. The levels of Unp mRNA were similar in all tissues and cell lines except in testes, where the transcript is more abundant than other tissues (figure 3.2). The Unp mRNA was found to be ubiquitously expressed in all mouse tissues and cell lines examined and at all stages of development.
FIGURE 3.2

Northern blot analysis showing ubiquitous expression of Unp

Northern blot analysis using total RNAs from heart (lane 1), kidney (lane 2), brain (lane 3), lung (lane 4), testes (lane 5), and liver (lane 6). The blot was probed with the $^{32}$P-labelled T4 cDNA fragment (panel A). The same blot was stripped and rehybridized to a $^{32}$P-labelled fragment containing tubulin gene sequences (panel B), in order to standardize for the amount of RNA present in each lane. The blot was washed in both cases in 0.1X SSC and 0.1% SDS at 65°C.
FIGURE 3.2

1 2 3 4 5 6

a)

– 28S
– 18S

b)
FIGURE 3.3

Northern blot analysis to show the Unp mRNA doublet

Northern blot analysis of mouse polyA+ RNAs from undifferentiated P19 cells (lane 1), adult brain (lane 2) and adult heart (lane 3). The positions of the ribosomal RNA markers are indicated. This blot was hybridized to a $^{32}$P-labelled T4 cDNA fragment and washed at 65° C in 0.1X SSC and 0.1% SDS.
3.42 In vitro Translation Of Unp cDNAs

The T4 cDNA (containing the entire Unp coding sequences) or an amino truncated Unp cDNA (F9) were cloned into the vector pGEM4 (Promega) which contains the T7 and SP6 promoters, thus allowing for the production of both sense and antisense mRNAs of the Unp cDNAs. After linearization with the appropriate restriction endonuclease, the cDNAs were transcribed and translated in vitro in the presence of $^{35}$-S methionine. Based on computer analysis the size of the Unp protein produced by the T4 construct is 90 kDa. As expected protein products were seen only in the lanes containing reaction products from the translation of the sense mRNAs only (figure 3.4). It is interesting to note, however, that the translated protein migrates considerably higher (at approximately 120 kDa) than the predicted molecular mass of 90 kDa. The higher migration of the Unp protein could be due to post-translational modifications of the protein, for example, phosphorylation. These kinds of modifications have been seen to influence the migration of other proteins like the Wee-1 kinase (Tang et al., 1993).

3.43 Detection Of Unp Protein in vivo

For detection of endogenous Unp protein, polyclonal antisera against a synthetic MAP peptide (Tam, 1988) was generated in rabbits. Initial screening of the antisera was done using a bacterially produced Gst-Unp fusion protein. The antisera showed reactivity towards bacterial lysates expressing the sense orientation Gst-Unp fusion construct (figure 3.5A). No reactivity was observed either towards the antisense Gst-Unp fusion construct
in vitro translation of \textit{Unp} protein

\textit{in vitro} transcription/translation of \textit{Unp} cDNAs. Lanes 1 and 2 contain translation products using the sense (lane 1) and antisense (lane 2) orientation of the largest \textit{Unp} cDNA T4. Lanes 3 and 4 contain translation products using the sense (lane 3) and antisense (lane 4) orientations of an N-terminus deleted \textit{Unp} cDNA F9. The reactions were performed using the BRL translation kit in the presence of \textsuperscript{35}S-methionine.
FIGURE 3.4

1 2 3 4  

kDa
- 200
- 97
- 68
or the GST (vector alone) expressing bacterial lysates. Moreover, the immunoreactivity was totally abolished upon preincubation of the antisera with the synthetic Unp peptide used to generate the antisera (figure 3.5B). These results showed that the antisera was specific for the Unp protein. To increase the titer of the antibody, the antisera was affinity purified using a column containing the Unp MAP peptide coupled to cyanogen bromide activated sephrose 4B (Pharmacia).

Western blot analysis was performed using whole cell protein extracts from NIH 3T3 cells. The affinity purified antibody detected a protein which migrated at approximately 180 kDa (figure 3.6). The immunoreactivity to this protein was specifically competed with the synthetic peptide. The migration of the Unp protein detected by the antibody is much higher than the molecular mass predicted by the computer analysis or the molecular mass of the in vitro transcribed and translated Unp protein.

3.44 Unp protein is localized in the nucleus

The predicted amino acid sequence of the Unp protein contains a nuclear localization signal similar to the main nuclear localization signal (NLS1) of p53 (Shaulsky et al., 1990). In order to find out if the Unp protein is localized in the nucleus, protein extracts from nuclear and cytoplasmic fractions of NIH 3T3 cells were prepared. These extracts were analysed by western blot analysis using affinity purified anti-Unp polyclonal antibodies. The 180 kDa Unp protein was detected primarily in the nuclear
Western blot analysis showing the specificity of anti-\textit{Unp} antisera. Protein extracts from bacteria containing the \textit{Gst-Unp} fusion construct in either the sense (S) or the antisense (AS) orientation were used. Three parallel blots were probed as follows: blot on the left with a 1:1000 dilution of anti-\textit{Unp} antisera generated against a synthetic \textit{Unp} peptide, blot in the middle with preimmune sera at a dilution of 1:1000, and 3) blot in the right with anti-\textit{Unp} antisera preincubated with the \textit{Unp} peptide at a dilution of 1:1000.
FIGURE 3.5

A

B

C
FIGURE 3.6

Detection of *in vivo* Unp protein by western blot analysis

Western blot analysis of whole cell lysates from NIH 3T3 cells. The left blot was probed with affinity purified rabbit anti-Unp antisera generated against a synthetic Unp MAP peptide. The antisera was used at a 1:200 dilution. The right blot was hybridized to the anti Unp antisera which had been preincubated with the Unp peptide. In the right blot, the intensity of a smaller cross-reacting band is increased (this band is also detected by the preimmune sera).
\[ \alpha \text{ Unp} \quad \alpha \text{ Unp} \quad + \text{peptide} \]

\[ \text{kDa} \quad 200 \]

\[ \text{---} \quad 97 \]
fraction of NIH 3T3 cell lysates. (figure 3.7) Anti-tubulin antibodies were used to reprobe the same blot. The results obtained from the anti-tubulin antibody showed that the nuclear fraction of the cell lysates was enriched for nuclear proteins. Taken together these results show that the Unp protein is localized primarily in the nucleus.

3.5 DISCUSSION

The northern blot analysis performed using the T4 cDNA (largest Unp cDNA) shows that this gene is transcribed ubiquitously at all stages of development in all tissues and cell lines tested (figure 3.2). Moreover, in agreement with our sequencing data, there are two closely migrating mRNA species detected (figure 3.3). These mRNAs are about 3.5 and 3.7 kb in size. This result not only tells us that the cDNA T4 (3.7 kb in size) is full length, but it also supports the sequencing data where it was predicted that the difference between the two Unp mRNA species differing at their 3' ends would be roughly 300 base pairs. The function of the two mRNAs is currently unknown, however the sequence data suggests that they arise due to equal usage of two degenerate polyadenylation signals. There are examples in eukaryotic systems for the usage of unconventional polyadenylation signals (Birnstiel et al., 1985). It would be very interesting to see if the 3' UTR of Unp mRNA serves any regulatory functions such as influencing the mRNA stability and efficiency of polyadenylation. Recently, genetic complementation has revealed that the 3' UTR of some differentiation-specific mRNAs can also effect the growth and differentiation properties of cells (Rastinejad and Biau,
FIGURE 3.7

Detection of *Unp* protein in the nuclear fraction of NIH 3T3 cells

Western blot analysis of NIH 3T3 cell lysates fractionated into nuclear and cytoplasmic fractions. The blot in the left panel was probed with a 1:200 dilution of affinity purified anti-*Unp* antisera. In the right panel, the same blot was stripped and reprobed with an anti-tubulin antibody. Abbreviations used are C; cytoplasmic fraction of NIH 3T3 cell lysates and N; Nuclear fraction of NIH 3T3 cell lysates.
Polyclonal antibodies raised against a synthetic Unp multiple antigen peptide (MAP) detected an approximately 180 kDa protein in NIH 3T3 cell lysates (figure 3.6). The molecular mass of this protein was considerably larger than that deduced (90 kDa) from the amino acid sequence of the largest open reading frame present. The same result is seen with respect to the in vitro translated Unp protein, which also migrates significantly higher (120 kDa) than the predicted molecular mass of 90 kDa (figure 3.4). This could be due to extensive post-translational modifications of the primary translation product in vivo, some of which can also take place in vitro. There are precedents for this kind of shift in the motility of proteins on SDS PAGE gels, for example in the case of the Wee-1 protein kinase (Tang et al., 1993), phosphorylation of this protein makes it migrate at a molecular mass of 170 kDa as compared to the unphosphorylated protein which migrates at 107 kDa.

Fractionation of NIH 3T3 cell lysates into cytoplasmic and nuclear fractions shows that the Unp protein is primarily localized in the nucleus (figure 3.7). This confirms the prediction made on the basis of the presence of the nuclear localization signal in the predicted primary amino acid sequence of the Unp protein. Nuclear localization of Unp protein is interesting in light of the presence of CR2 element in this protein which may mediate interactions with the gene products of the retinoblastoma gene family members (Defeo-Jones et al., 1991, Hu et al., 1990, Phelps et al., 1988, Kaelin et al., 1992).
CHAPTER 4

Unp IS A PROTO-ONCOGENE AND ENCODES
A UBIQUITIN-SPECIFIC PROTEASE
4.1 SUMMARY

The *Unp* gene contains a region of homology to the *tre* oncogene, which suggested that *Unp* may be a proto-oncogene. Tumorigenesis assays performed in athymic mice show that *Unp* is tumorigenic in nude mice, which in turn shows that *Unp* is a proto-oncogene. In collaboration with Dr. R. Baker, we have also shown that the gene product of *Unp* is a ubiquitin-specific protease. Bacterially expressed *Unp* protein can cleave ubiquitin from synthetic ubiquitin-βgal substrates. The *Unp* gene is highly polymorphic in human cell lines and 1 out of 16 human lung tumors analysed show a deletion of *Unp* sequences in tumor tissue DNA compared to adjacent tissue DNA from the same patient.
4.2 INTRODUCTION

The primary amino acid sequence of the Unp protein contains several conserved domains which provide some clues towards the functions of this gene. It contains a nuclear localization signal similar to the p53 NLS1 (Shaulsky et al., 1990) and I have shown that the Unp protein is localized in the nucleus. The bipartite pRb binding consensus is present within a region that shares homology to the human oncogene tre-2. All of these features of Unp suggested a possibility that it may be a proto-oncogene. We have tested this possibility by assaying for tumorigenicity in athymic mice.

The Unp protein also contains the Cys and His domains which are conserved in yeast ubiquitin-specific proteases, UBPs (Baker et al., 1992). This suggests that the Unp gene product may be a ubiquitin-specific protease. We have tested this possibility using an in vitro system in collaboration with Dr. R.T. Baker.

The other interesting aspect of Unp is its chromosomal location in humans. I have already presented data (chapter 1) showing that Unp is located on chromosome 3 in humans. Although the exact position of Unp on human chromosome 3 is not yet determined, its location on mouse chromosome 9 is such that it is in a region which shares homology to human chromosome 3p21 region. Thus, it is very likely that Unp will be located in the 3p21 region on human chromosome. This region of chromosome 3 is implicated in several types of human malignancies including carcinomas of lung (Kok et al., 1987), breast (Ali et al., 1989) and kidney (Kovacs et al., 1988). This led us to speculate that Unp may be a tumor suppressor gene present in this region. We have
tested this possibility by looking for LOH of Unp in human lung tumors.

4.3 MATERIALS AND METHODS

4.3.1 Constructs

For assaying the tumorigenicity on Unp in athymic mice, two constructs, a pKJ-Unp construct (containing the entire open reading frame of Unp), and an amino terminal deleted myc-Unp construct, were made. The pKJ-Unp construct was made by ligating a SmaI/DraI fragment containing the entire Unp open reading frame into the SmaI site of plasmid pKJ1ΔR (McBurney et al., 1991). The myc-Unp fusion construct for tumorigenicity analysis (figure 4.1) was made by ligating a fragment containing amino acids 234 to the end of the open reading frame in frame into the Xba I site of the plasmid pBluescript (Stratagene) containing 6 epitopes of myc (pMTG) which can be recognized by the monoclonal antibody 9E10 (Evan et al., 1985). The myc tag sequences were positioned at the NH₂ terminal of the fusion protein being produced. A blunt ended Sal I, Not I fragment from this F9-MTG plasmid (containing both the myc and Unp sequences) was then introduced into the Sma I site of plasmid pKJ1ΔR containing both the Pgk promoter and polyadenylation sequences (McBurney et al., 1991), such that the myc-Unp fusion transcript was transcribed by the Pgk promoter. This plasmid was designated pKJMYC-orf 5'd.
FIGURE 4.1

Constructs used for tumorigenicity assays

The constructs used for the tumorigenicity assays in athymic mice. pKJT4 is a construct which contains the largest mouse Unp cDNA (T4) driven by the Pgtk (phosphoglycerate kinase) promoter. pKJmyc-orf5' Δ is a myc-Unp fusion construct, also driven by the Pgtk promoter. This construct lacks part of the coding sequences from the N-terminus of the Unp protein.
FIGURE 4.1

Constructs For Tumorigenicity Assay

A
pgkT4

B
pgjmyc-orf 5' d

Diagram of constructs for tumorigenicity assay. Each construct is labeled with their respective sequences: 5' to 3' (P-gal, unphosphorylated, phosphorylated), with myc protein shown as well.
4.32 Tissue Culture And Cell Lines Used

All cell lines were grown and maintained in DMEM (Gibco) supplemented with 10% heat inactivated calf serum and 10% heat inactivated fetal calf serum (Gibco, BRL). Freezing medium for the cells consisted of DMEM medium containing 10% fetal calf serum and 10% DMSO. The tumor cell lines were generously provided by Dr. D. Logan and the primary human foreskin fibroblast cell line was generously provided by Dr. M. Chevrette.

4.33 Southern (DNA) Blot Analysis

DNAs were prepared by the usual method (Maniatis et al., 1989), restricted to completion using restriction endonucleases (Gibco, BRL and Boehringer Mannheim) and fractionated on a 1% agarose gel. DNA was transferred and UV crosslinked to Hybond-N membrane (Amersham) as described by the manufacturer. Following this, the membranes were prehybridized in prehybridization solution containing 50 mM Hepes (pH 7.0), 0.75M NaCl, 50% formamide, 0.5% SDS, 2 mM EDTA (pH 8.0), 10X Denhardt ((Denhardt, D., 1966), 200 µg/ml Herring sperm DNA and 10 µg/ml poly(A). Hybridization was carried out in the prehybridization solution containing 32P-labelled DNA of interest. The blots were then washed at the stringencies described in the figure legends.
4.34 Tumorigenicity Assays

The constructs were cotransfected into NIH 3T3 cells along with a puromycin resistance plasmid (Vara et al., 1986). Either clones or populations of transfected puromycin resistant cells were injected into athymic CD1 mice subcutaneously. Each mouse was injected at 4 sites with $3 \times 10^6$ cells per site. The mice were assessed for the appearance of palpable tumors. Untransfected NIH 3T3 cells were also injected as control. Each mouse was injected into four sites with $3 \times 10^6$ cells per site. The tumors were allowed to grow until the first observed tumor reached a size of 1 cm in diameter. At this point the mice were sacrificed and the tumors were excised. DNA was isolated from the tumors by the usual method (Maniatis et al., 1989). The DNA was analysed for the presence of transfected DNA by Southern blot analysis. Protein extracts were also isolated from some of the tumors (where the expression construct contained the myc- Unp fusion protein), and this protein was used for western blot analysis using the anti-myc antibody 9E10 (Evan et al., 1985). As a control one mouse was injected with untransfected NIH 3T3 cells.

4.35 Western (Protein) Blot Analysis

Tumor tissues obtained from athymic mice were lysed in lysis buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2mM EDTA, 1% Triton, 200 µg/ml phenylmethylsulfonyl fluoride, 2 µg/ml aprotinin and 5 µg/ml leupeptin) and sonicated on ice three times (20 seconds each time). The lysates were cleared by ultracentrifugation at
100,000g. Supernatants were then resolved on 8% SDS-polyacrylamide gels and then 
electro-transferred to nitrocellulose (Schleicher and Schuell) in 25 mM Tris (pH 7.0), 200 
glycine and 20% methanol buffer at 0.7 amps for 90 min. to 2 hours. Membranes were 
blocked in Tris buffered saline (TBS, 140mM NaCl, 10 mM Tris, pH 7.5) containing 
0.05% tween 20 and 5% non-fat dry milk powder for one hour at room temperature. The 
monoclonal anti-myc antibody 9E-10 (Evan et al., 1985) was then added to the blocking 
buffer at the desired concentration (1:500) and the membranes were incubated for another 
hour at room temperature. Blots were rinsed vigorously three times (5 minutes each time) 
in Tris-buffered-saline-tween (TBST) before incubating for 30 minutes with a 1:3000 
dilution of a secondary antibody, conjugated to horse radish peroxidase (ICN 
Biologicals). The blots were once again washed vigorously in TBST as before and 
developed using the ECL western blotting detection solutions (Amersham) according to 
the manufacturer’s instructions.

4.4 RESULTS

4.41 Unp Is Tumorigenic In Athymic Mice

The similarity between the predicted amino acid sequence of the Unp protein and 
the tre-2 oncoprotein (Nakamura et al., 1992) suggests a possible role of the Unp protein 
in tumorigenicity. To assay for potential tumorigenicity of Unp, an assay similar to the 
one used for the identification of tre-2 (Nakamura et al., 1988) was used. Expression 
plasmids were made in which the phosphoglycerate kinase promoter (McBurney et al.,
1991) was used to drive the expression of either the T4 cDNA (containing the entire Unp open reading frame) or an amino truncated Unp cDNA (amino acids 234 to 794) fused in frame with six epitopes of c-myc gene. These myc epitopes can be recognized by the monoclonal antibody 9E10 (Evan et al., 1985).

The results obtained from the injection of cells into athymic mice are presented in Table 4.1. All of the athymic mice that received NIH 3T3 cells transfected with either of the constructs developed tumors. It is important to note, however, that not all of the tumors were scored in this assay, since the mice were sacrificed after the first observed tumor reached 1 cm in diameter. Thus, the number of tumors listed per mouse in the table are misleading because some of the sites which may have developed tumors went unscored as the tumors were rapid growing, usually reaching 1 cm diameter within one week after becoming noticeable.

All the tumors were analysed for the presence of transfected DNAs by Southern blot analysis. This analysis verified that the tumors arose from transfected NIH 3T3 cells (figure 4.2), because all of the DNAs from tumor tissues contained the band specific to transfected NIH 3T3 cells. Some of the tumors which were obtained from NIH 3T3 cells transfected with the myc-Unp fusion construct were also analysed for the presence of myc-Unp fusion protein using the anti-myc monoclonal antibody 9E10. These tumors were found to contain the fusion protein (figure 4.3). The myc-Unp specific fusion
**TABLE 4.1**

Tumorigenicity of *Unp* in athymic mice

Growth of transfected cells in athymic CD1 mice. Clones (C1A and C1B) or populations (popn) of transfected NIH 3T3 cells were injected into 4 sites of each mouse. Animals were sacrificed (Sac’d) when any tumor reached 1 cm in diameter. Thus any tumors which may have developed later on went unscored in our analysis.
<table>
<thead>
<tr>
<th>mouse #</th>
<th>inoculum</th>
<th>days to 1st tumor</th>
<th>sac'd</th>
<th>tumors at sac</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F9/3T3 cl B</td>
<td>31</td>
<td>31</td>
<td>1/3</td>
</tr>
<tr>
<td>2</td>
<td>F9/3T3 popn.</td>
<td>39</td>
<td>39</td>
<td>1/4</td>
</tr>
<tr>
<td>3</td>
<td>T4/3T3 popn.</td>
<td>25</td>
<td>43</td>
<td>3/4</td>
</tr>
<tr>
<td>4</td>
<td>F9/3T3 popn.</td>
<td>50</td>
<td>50</td>
<td>2/4</td>
</tr>
<tr>
<td>5</td>
<td>F9/3T3 cl A</td>
<td>106</td>
<td>110</td>
<td>2/4</td>
</tr>
<tr>
<td>6</td>
<td>3T3</td>
<td>--</td>
<td>120</td>
<td>0/4</td>
</tr>
</tbody>
</table>
Southern blot analysis to show the presence of transfected DNA in tumors obtained from athymic mice. DNAs from NIH 3T3 cells (lanes 1 and 2), NIH 3T3 cells transfected with the Unp construct (lanes 3 and 4), tumor 1 tissue (lanes 5 and 6) and tumor 2 tissue (lanes 7 and 8) were restricted with EcoRI (lanes 1, 3, 5, 7) and BamHI (lanes 2, 4, 6, 8). The probe used was a $^{32}$P-labelled T4 cDNA fragment and the blot was washed at 65$^\circ$C in 0.1X SSC and 0.1% SDS.
FIGURE 4.2

Analysis of Primary tumors
FIGURE 4.3

Western blot analysis of tumors from athymic mice

Western blot analysis to show the presence of Myc-Unp fusion protein in tumor tissue protein extract. The protein extracts used were 3T3; NIH 3T3, N28P; NIH 3T3 cells transfected with a Unp construct, P19; embryonal carcinoma cell line, and T1; tumor tissue extract. The primary antibody used was a commercially available monoclonal anti-myc antibody 9E10 at a dilution of 1:500.
protein band is seen only in lysates from the transfected NIH 3T3 cells and from the tumor tissues. This band is absent in lysates from NIH 3T3 cells or P19 cells. These results show that Unp has a potential to cause tumorigenicity in nude mice.

4.42 Unp Is Highly Polymorphic In Human Cell Lines

Based on the chromosomal location of Unp in the 3p21 region of human chromosome and the conserved domains present in its proteins (figure 2.4), we hypothesized that the Unp gene may be a tumor suppressor gene. Tumor suppressor genes are usually identified through deletion of their gene sequences in tumors. In the case of 3p21, chromosomal markers from the 3p21 region were found to be deleted in several tumor types (Kok et al., 1987, Kohno et al., 1993), indicating that tumor suppressor gene(s) in that region was also deleted in these tumors. To detect deletion of gene sequences in one of the alleles of the gene (LOH, loss of heterozygosity), restriction length polymorphism (RFLP) must be identified in order to distinguish between the two alleles. Also, the polymorphism must be present at a high enough frequency so that the number of tumors to be looked at can be kept to a minimum.

To look for polymorphisms in the Unp gene, DNA from several human tumor cell lines (including small cell lung carcinomas, non small cell lung carcinomas, breast carcinomas and colon carcinomas), and from primary human foreskin fibroblast cells was analysed by Southern blot analysis (figure 4.4). A polymorphism was seen to occur at a high frequency with the restriction endonuclease BamHI. To see if polymorphism in
FIGURE 4.4

Southern blot analysis showing *Unp* polymorphism

Southern blot analysis to look at polymorphism in the *Unp* gene in human tumor cell lines. DNAs used were from primary human foreskin fibroblast culture (HFF), Small cell lung carcinoma cell lines (LC3 and HTB119), non small cell lung carcinoma cell lines (HTB53, HTB56, HTB57 and HTB58), breast carcinoma cell lines (MCF7 and HTB22), and colon carcinoma cell lines (HTB37, HTB38, HTB39 and CL187). The probe used is a 32P-labelled 1.6 kb EcoRI fragment from the human *Unp* cDNA. The blot was washed at 65° C in 0.1X SSC and 0.1% SDS.
Unp is Polymorphic in Human Cell Lines

FIGURE 4.4
the *Unp* gene could be detected with other restriction endonucleases, five other enzymes were also tested in a similar analysis (figure 4.5), using two of the cell lines. Polymorphism was also detected with at least two other enzymes (EcoRI and KpnI). These results provided us with the tool to start looking for LOH in human tumors vs their matched normals.

**4.43 Southern Blot Analysis Of Human Lung Tumors**

DNA from 16 lung tumors along with their matched normals was analysed by Southern blot analysis using the restriction endonuclease BamHI. A representation of this analysis is shown in figure 4.6, for 5 of the tumors obtained. Of the 16 tumors analysed, only 7 were informative (heterozygous for the two *Unp* alleles, for example tumors 2 and 3 in figure 4.6), since LOH could not have been detected in the tumors homozygous for the two alleles of *Unp*, for example tumors 4A, 4B, and 5 in figure 4.6. Only one of the informative tumors (tumor 10) showed a loss of *Unp* specific band in tumor tissue when compared to the normal tissue. This change (LOH) was seen using all 3 enzymes, BamHI, EcoRI and KpnI (figure 4.7). Thus, a change in *Unp* gene was observed in 1 out of 7 informative tumors analysed, making it necessary to look at more tumors so that conclusions can be drawn. Also, there was only one small cell lung carcinoma (where the maximum frequency of 3p21 deletions is seen) among all of the tumors analysed and this tumor was non-informative. The fact that changes were not observed in the other tumors does not mean that *Unp* was not altered in these tumors. Minor changes (like
FIGURE 4.5

Detection of polymorphism in Unp using various enzymes

DNA from two human tumor cell lines (HTB56; even lanes, and HTB57; odd lanes) were restricted with various restriction endonucleases (indicated on the top of the figure) and analysed by Southern blot analysis. The blot was hybridized to $^{32}$P-labelled internal EcoRI fragment from the human Unp cDNA and washed in 0.1X SSC and 0.1% SDS at 65º C. The restriction enzymes used were HindIII (lanes 1 and 2), EcoRI (lanes 3 and 4), KpnI (lanes 5 and 6), SstI (lanes 7 and 8), XbaI (lanes 9 and 10), PstI (lanes 11 and 12), and BamHI (lanes 13 and 14).
FIGURE 4.5

Analysis of tumor cell lines
FIGURE 4.6

Southern blot analysis of human lung tumors

Southern blot analysis of five human lung tumors and their matched "normals" tissues. The numbers above indicate the tumor number. T; tumor tissue DNA, N; adjacent "normal" tissue DNA. The restriction endonuclease used was BamHI and the probe is a $^{32}$P-labelled, 1.6 kb EcoRI fragment from the human Unp cDNA. The blots were washed in 0.1X SSC and 0.1% SDS at 65°C.
FIGURE 4.7

Southern blot analysis of human lung tumor 10

Southern blot analysis using DNA from tumor 10 (which shows loss of Unp sequences). The restriction endonucleases used are shown on top. T; tumor tissue DNA and N; adjacent "normal" tissue DNA. The probe used was $^{32}$P-labelled, 1.6 kb EcoRI fragment from the human Unp cDNA. The blots were washed at 65°C, in 0.1X SSC and 0.1% SDS.
small deletions or point mutations) may have occurred but these would not have been detected due to the limitations of our analysis.

4.44 Unp Codes For A Ubiquitin Protease

The open reading frame of Unp contains two regions (the cys and the his domains) which are conserved among yeast ubiquitin-specific proteases (Baker et al., 1992). These two regions encompass the active sites of these proteins. The possibility that Unp protein may be a murine ubiquitin-specific protease was investigated in collaboration with Dr. R.T. Baker in Australia. This was an intriguing possibility because to date no members of the Urb family of ubiquitin cleavage enzymes have been isolated from any organism other than yeast. Dr. Bakers's analysis has proved by two separate assays, which are discussed below, that Unp is a ubiquitin-specific protease.

The first of these assays was to transform the T4 Unp cDNA into two strains of E. coli, one containing Ub-M-βgal (ubiquitin-methionine-β-galactosidase) fusion protein and the other containing Ub-R-βgal (ubiquitin-arginine-β-galactosidase) fusion protein. In the absence of exogenous ubiquitin cleavage activity, both these strains produce blue colonies, since bacterial cells lack ubiquitin protease activity. In the presence of exogenous ubiquitin cleavage activity, however, based on the N-end rule (Varshavsky, 1992) the Ub-M-βgal fusion construct containing E. coli still produce blue colonies but the Ub-R-βgal construct containing E. coli produce white colonies.

It was found that in the presence of T4 cDNA, the E. coli cells with Ub-R-βgal
construct gave rise to white colonies, showing that the T4 cDNA (Unp) did code for a ubiquitin cleavage enzyme. These results were also verified using extracts from the T4 cDNA expressing E. coli cells. This extract was found to be capable of cleaving a purified ubiquitin-glutathione-S-transferase fusion protein, once again proving that Unp codes for a ubiquitin-specific protease.

When a cDNA (F9) which lacks the conserved cysteine domain was used for similar analysis, it tested negative for ubiquitin cleavage activity. This experiment showed that the cys domain of Unp is essential for the protease activity of the protein. Dr. Baker has also made a site directed mutation in the Cys domain of Unp and this mutant does not show any ubiquitin-specific protease activity in his assays.

4.5 DISCUSSION

The homology between Unp and tre-2 along with its nuclear localization and presence of pRb binding consensus all suggested that Unp may be a proto-oncogene. The tumorigenicity data presented (Table 4.1) shows that Unp is a proto-oncogene because tumors were obtained only in mice injected with Unp transfected NIH 3T3 cells. It would be interesting to see if mutation in either the nuclear localization signal or the pRb binding consensus would have an effect on the tumorigenicity of Unp.

Dr. Baker’s analysis shows that Unp encodes a ubiquitin-specific protease and that mutation of the Cys domain cysteine leads to a loss of this activity. Experiments are ongoing to see if the protease activity is required for the tumorigenicity of this gene.
Based on the chromosomal location of Unp on chromosome 3 in humans, in a region where tumor suppressor gene(s) is thought to reside, we tested the possibility that Unp may be a tumor suppressor gene. Unp would not be the first gene capable of acting as both an oncogene as well as a tumor suppressor gene. p53 is a good example of a gene that is both a tumor suppressor gene (in its unmutated or wild type form) as well as an oncogene (in its mutated form). We have identified polymorphism in the Unp gene using at least three restriction endonucleases and DNA from a variety of different human tumor cell lines (figure 4.4). When we looked for LOH in human tumors using one of these restriction endonucleases (BamHI), 1 out of the 16 lung tumors analysed showed a loss of Unp specific sequences (figure 4.6). This loss of Unp sequences was verified with the other two restriction endonucleases as well (EcoRI and KpnI). Based on the pattern of bands seen in figure 4.7, it seems that the Unp gene has undergone some sort of rearrangement, rather than a simple deletion. The nature of this rearrangement is unclear at this point. The statistics of alterations of Unp sequences in human lung tumors is very low and more tumors need to be looked at. However, we cannot exclude the possibility that the tumors that did not show any alterations in our analysis may contain undetectable changes like small deletions or point mutations. Such mutations have been seen to occur in tumor suppressor genes like Rb and can produce a defective or non-functional protein (Horowitz et al., 1989). In these genes the mutations often occur in regions which are linked to the functional properties of the gene.
CHAPTER 5

DISCUSSION
5.1 Features Of Unp mRNAs

The Unp cDNAs that were sequenced were of two types which differed only at their 3' ends. None of these cDNAs contained a perfect match to the consensus polyadenylation signal (Proudfoot, 1991). Degenerate polyadenylation signals were present upstream of the poly(A) tails. The two types of cDNA clones were obtained at roughly equal frequencies, suggesting that there was no preference in the usage of the two degenerate polyadenylation signals. This is supported by the northern blot analysis data obtained where two closely migrating Unp mRNAs are detected in all tissues examined (figure 3.2 and 3.3). Thus, the Unp gene is transcribed ubiquitously into two mRNA species of approximately 3.5 and 3.7 kb. Mutations in the poly(A) signal AAUAAA have been shown to effect mRNA processing, that is cleavage of the mRNA at the 3’ end (Wickens and Stephanson, 1984).

The Unp cDNAs contain a large open reading frame (ORF) (figure 2.4) of 794 amino acids which would code for a protein of 90 kDa, as predicted by computer analysis. In addition to the large ORF, the 5' UTR of Unp cDNAs also contains several small ORFs (figure 2.2) which is an unusual feature for a eukaryotic gene. This kind of 5' UTR is characteristic of growth regulatory genes like TGF-β (Arrick et al., 1991), proto-oncogenes like Ick (Marth et al., 1988) or rearranged or recombinant oncogenes like bcr/abl (Muller and White, 1989). In all these genes the removal of the 5' UTR leads to an increase in the translational efficiency of the mRNAs. Removal of 5' UTR
oncogene by retroviruses (Garvis et al., 1988).

The relevance and function of the alternate 3' ends or the 5' UTR of Unp mRNAs remain unclear at this point. Experiments need to be done to see if the extended 3' UTR serves to increase or decrease the mRNA stability, or if the 5' UTR of Unp effects its translational efficiency.

5.2 Features Of The Predicted Unp Protein

Based on computer analysis, the predicted molecular mass of the Unp protein was 90 kDa. The in vitro transcribed and translated product from the largest Unp cDNA was found to migrate at 120 kDa (figure 3.4), which is significantly more than the predicted molecular mass. We went on to look for cellular Unp protein in NIH 3T3 cell lysates using polyclonal anti-Unp antisera generated in rabbits against a synthetic Unp MAP peptide. The antisera reacted specifically to bacterial Gsr-Unp protein and the immunopurified antisera detected a protein of 180 kDa in NIH 3T3 cell lysates (figure 3.6). The difference in the predicted molecular mass of Unp protein and the molecular mass of the Unp protein detected suggests that the Unp protein is post-translationally modified in vivo. One such modification could be phosphorylation which has been shown to have a significant effect on the migration of proteins (for example the wee I protein kinase, Tang et al., 1993).

The predicted Unp protein contains a putative p53 type nuclear localization signal which suggested that this protein could be localized in the nucleus. Upon fractionation
of NIH 3T3 cell lysates, the 180 kDa *Unp* protein was detected primarily in the nuclear fraction (figure 3.7). The nuclear localization of *Unp* is significant in light of the presence of the pRb binding consensus in proteins that associate with pRb (Dyson et al., 1992\(^b\)). This bipartite motif is required for efficient binding of these (for example E1A and SV40 large T antigen) proteins to pRb. Thus, an intriguing possibility is that *Unp* may be a cellular protein that binds to pRb. This possibility is currently under investigation in collaboration with Dr. B. Galley.

### 5.3 Homology Between *Unp* And *tre* Oncogene

In a computer search, the *Unp* protein was found to contain a region of homology to the recombinant human oncogene *tre*-2 (Nakamura et al., 1992). There is a 42% identity in a stretch of 288 amino acids (figure 2.10). This region of *tre* has been identified as being essential for the tumorigenicity of this gene. In *Unp* the region of homology contains the pRb binding motifs CR1 and CR2.

Our initial quarry was to see if *Unp* is a homolog of *tre*-2 (now designated *tre*-17 due to its chromosomal localization) which belongs to a family of genes, based on Southern blot analysis. This was an unlikely possibility since we knew from Dr. Jenkin's chromosomal mapping data (figure 2.7) that *Unp* would probably be located in the 3p21 region of chromosome 3 in humans and *tre*-2 is located on human chromosome 17. None the less, we made use of the somatic cell hybrids (provided by Dr. Chevrette) to show that *Unp* is located on chromosome 3 in humans (figure 2.11). Also our Southern blot
results (figure 2.9) show that like \textit{tre-2}, \textit{Unp} also belongs to a family of genes. Thus, we believe that \textit{Unp} is not a homolog of \textit{tre-2} but rather that the two genes belong to a family of genes and may serve similar functions. The homology between \textit{Unp} and \textit{tre-2}, along with the nuclear localization of \textit{Unp} and the presence of pRb binding consensus in its protein all suggested a very tempting possibility that \textit{Unp} may be a proto-oncogene.

5.4 \textit{Unp} Is A Proto-Oncogene

To investigate the possibility that \textit{Unp} is a proto-oncogene, we took the same approach that was used by Nakamura et al., (1988) for the identification of the \textit{tre} oncogene. Using this approach, we obtained tumors only in mice that were injected with NIH 3T3 cells transfected with a \textit{Unp} construct (table 4.1), showing that \textit{Unp} is tumorigenic in nude mice, that is, that \textit{Unp} is a proto-oncogene. Thus, the prediction that was made on the basis of the presence of conserved domains in the predicted protein of \textit{Unp} has been shown to be true. It would be very interesting to see what effect the deletion or mutation of either the nuclear localization signal or the pRb binding consensus domains will have on the tumorigenicity of this gene.

5.5 \textit{Unp} Is A Ubiquitin-Specific Protease

The predicted \textit{Unp} protein also contained amino acid sequences that are found to be conserved among yeast ubiquitin proteases (\textit{UBP1-3}, Baker et al., 1992). All these
enzymes contain two conserved domains (which are required for their activity). The Cys
domain is towards the N-terminus and the His domain is towards the C-terminus.

In collaboration with Dr. Baker, Unp was found to contain ubiquitin protease
activity, because it was able to cleave ubiquitin from synthetic ubiquitin-β'-galactosidase
substrates. Unp is the first UBP type mammalian ubiquitin protease to be discovered and
it will be very interesting to look at its substrates, especially in light of its tumorigenicity
and putative pRb binding properties.

5.6 Unp As A Candidate Tumor Suppressor Gene

Unp has been mapped to mouse chromosome 9 (figure 2.7) and human
chromosome 3 (figure 2.11), and its location on mouse chromosome 9 suggests that the
human homolog of Unp will be located in the 3p21 region of human chromosome.
Human chromosome region 3p21 has been implicated in several human malignancies
(identified by LOH) including cancers of lung (Kok et al., 1987, Kohno et al., 1993),
breast (Ali et al., 1989) and kidney (Kovacs et al., 1988). This raises the possibility that
Unp may be a tumor suppressor gene present at the 3p21 locus. The only gene that has
been cloned from the 3p21 locus as a candidate tumor suppressor gene is a gene with
sequence similarity to ubiquitin-activating enzymes (Kok et al., 1993). This raises the
possibility that there may be a cluster of genes involved in the ubiquitin pathway at this
locus and these genes may somehow interact together to lead to a phenotype of the cell.

The possibility that Unp may be a tumor suppressor gene is further supported by
the observation that the \textit{Gnai-2} marker (which has been found to be the marker closest to the \textit{Unp} gene on mouse chromosome 9) is also closest to the D3f15S2 marker in humans which shows the highest frequency of LOH in small cell lung cancers (Caritt et al., 1992, Kohno et al., 1993). Thus, \textit{Unp} seems to map in a region where a tumor suppressor gene is believed to reside.

We have already identified restriction fragment length polymorphisms (RFLPs) in the \textit{Unp} gene (figure 4.4). Although the number of tumors looked at so far is very small (16 tumors), and the number of informative tumors even smaller (7 tumors). Moreover, only one of these tumors was a small cell lung carcinoma, making our chances of detecting LOH even smaller, yet we have found loss of \textit{Unp} sequences in one of these tumors (figure 4.7). As more tumors will be analysed, LOH of \textit{Unp} may be detected in more tumors, this will strengthen our suggestion the \textit{Unp} is a tumor suppressor gene located at the 3p21 locus in humans.

\textbf{5.7 Model For \textit{Unp} Function}

I have mentioned so many features of \textit{Unp} and suggested that the ubiquitin protease encoding, proto-oncogene can act as both an oncogene as well as a tumor suppressor gene. I have presented data supporting both these aspects of \textit{Unp}, but the question that remains is, how does \textit{Unp} achieve all this? A model to explain a possible mechanism of action of \textit{Unp} is presented in figure 5.1. This model presumes that (although not proven) the \textit{Unp} protein can bind to pRb through its CR1 and CR2 motifs
FIGURE 5.1

A possible model for Unp function

This is a simplified model for the representation of Unp function, based on its association with pRb. Different shapes of protein molecules suggest conformational changes. E2F is a cellular transcription factor which does not contain the CR1 and CR2 pRb binding domain consensus. X and Y are cellular targets of E2F and may be acted upon at different times. CDK refers to pRb phosphorylating enzymes (for example cdc2), PP1 is a phosphatase which has found to be complexed with pRb and may dephosphorylate it, and P represents phosphate groups. The interaction of CDK (cyclin dependent kinase) is thought to be mediated through cyclins.
FIGURE 5.1
(motifs shown to mediate the binding of several proteins to pRb, Dyson et al., 1992)

In normal cells, pRb is phosphorylated and dephosphorylated in a cell cycle dependent manner. pRb is involved in maintaining the cells in G1 phase (in its hypophosphorylated form, active form), and then pRb is inactivated by phosphorylation allowing the cells to enter the S phase of the cell cycle (Chen et al., 1989, DeCaprio et al., 1992). The phosphorylation of pRb is thought to be carried out by cdc2 like kinases (Lees et al., 1991) and dephosphorylation by the PPT-I like phosphatases (Durfee et al., 1993), since these proteins can be coprecipitated with pRb. The interaction between cdc2 kinases and pRb is thought to occur through cyclins, which unlike the cdc2 kinases contain the CR motif in their protein (Matsushima et al., 1992). Cyclins are regulated in a cell cycle dependent manner through the ubiquitin pathway (Glotzer et al., 1991). Underphosphorylated pRb binds to cellular proteins like E2F (Chellappan et al., 1991), which is a cellular transcription factor that can activate the transcription of growth related genes carrying the E2F binding site upstream of their promoters (Heibert et al., 1991).

Based on the properties of pRb mentioned above, the model in figure 5.1 depicts the following theory. In the G1 phase of the cell cycle, when pRb is phosphorylated, it is bound to cyclins (Dowdy et al., 1993, Ewen et al., 1993) and thus the cdc2 like kinases, which phosphorylate pRb. After this the Unp protein (which may also be phosphorylated in a cell cycle dependent manner) deubiquitinates the cyclin molecules, thus causing a change in their conformation. This change in conformation may trigger
two events: 1) the dissociation of cyclins from pRb and 2) the priming of the cyclin molecules for degradation by the proteasome. Another possibility is that the affinity of Unp for pRb is greater than the affinity of cyclins for pRb, when pRb is hyperphosphorylated. In any case, Unp protein displaces the cyclin protein (along with the kinase) on the Rb protein. The phosphorylation of pRb would cease and it may begin to get dephosphorylated by the phosphatase. The dephosphorylation of pRb will increase its affinity to E2F and decrease its affinity to Unp, thus causing Unp to dissociate from pRb so E2F can bind to pRb. E2F would no longer be available to transactivate the growth related genes and, until some other protein (like viral E1A or HPV16-E7, or possibly Unp protein) can free E2F protein from pRb and allow for the pRb protein to be phosphorylated again. The cell cycle would then continue.

I have presented data showing that Unp expressing NIH 3T3 cells are tumorigenic. According to the model, when overexpressed, Unp may act as an oncogene by inhibiting the interaction between pRb and E2F, such that E2F is constitutively free to activate transcription and the cell continues to divide. This part of the model is based on the finding that microinjection of E2F cDNA into quiescent REF-52 cells can induce the entry of these cells into the S phase of cell cycle, while this effect is not seen with transactivation deficient mutants of E2F (Johnson et al., 1993). Unp may act as a tumor suppressor gene because it may be required for the degradation of cyclins. Thus, in the absence of Unp, cyclins may accumulate and remain bound to the pRb molecule allowing the kinases to continually phosphorylate pRb. This speculation is based on the observation
that cyclins (cyclin A, B1, D1 and E) have been found to be amplified and overexpressed in human breast cancer cell lines (Buckley et al., 1993). In this case again E2F would remain free to constitutively transactivate growth regulatory genes and the cells would remain mitotically active.

The entire model presented above is based on the assumption that Unp binds to pRb. Experiments are ongoing in the lab, in collaboration with Dr. B. Galley, to show an association between the two proteins. Another way to test this model is through mutagenesis of the domains required for the functioning of Unp. The pRb binding motif will be mutated through site directed mutagenesis CR2, the domain critical for binding to pRb in other proteins. These constructs will be transfected into NIH 3T3 cells and the transfected cells can be assayed for tumorigenicity in nude mice.

To test if the ubiquitin protease function is required for the tumorigenicity of Unp, mutations will be made in the Cys and His domains of Unp (ongoing experiments in collaboration with Dr. Baker) which are required for the protease activity of these proteins. Once again these mutants will be assayed for tumorigenicity in athymic mice, as described for the mutation in the pRb binding domain.

We have recently designed a system where by Unp proteins (both mutant and wild type) can be expressed in a retroviral expression system. These retroviruses expressing the Unp protein of interest can be used directly to test for in vivo tumorigenicity (by injecting them into mice) or they can be used to infect NIH 3T3 cells. The infected NIH 3T3 cells can then be injected into athymic mice to assay for tumorigenicity.
The effect of overexpression of *Unp* can also be tested in transgenic mice. Since the highest frequency of 3p deletions has been observed in breast and lung tumors, it is reasonable to assume that these tissues will be highly sensitive to changes in the amount of *Unp* protein. I have already made two constructs where in one case *Unp* is driven by the MMTV promoter (to drive mammary gland specific expression of *Unp*) and in the other construct *Unp* gene expression is driven by the SP-C promoter to drive lung specific expression of *Unp* (Wert et al., 1992). I would expect these transgenic mice to develop tumors in the targeted tissues.
ORIGINAL CONTRIBUTIONS TO KNOWLEDGE

The following findings which are presented in this thesis represent original contributions to knowledge:

1) The identification of a novel gene, *Unp*, including its cDNA sequence and the predicted amino acid sequence of its protein.

2) Detection of two types of *Unp* mRNAs varying in the length of their 3' UTR sequences due to usage of degenerate polyadenylation sequences.

3) *Unp* is part of a gene family suggesting that there are other ubiquitin proteases which may vary in their properties and substrate specificities.

4) Detection of cellular *Unp* protein (using polyclonal antibodies directed against a synthetic *Unp* MAP peptide), which migrates at a considerably higher molecular mass than that predicted by its primary amino acid sequence. This suggests post-translational modifications of the protein *in vivo*.

4) *Unp* is the first mammalian nuclear ubiquitin-specific protease to be discovered.

5) *Unp* is the first ubiquitin-specific protease that has been shown to be a proto-oncogene.

6) This gene may be the tumor suppressor gene located at the 3p21 locus in humans, and may be involved in several types of human cancers.
REFERENCES


62. Erikson, J., ar-Rushdi, A., Drwinga, H.L., Nowell, P.C., and Croce, C.M.


84. Hayday, A.C., Gillies, S.D., Saito, H., Wood, C., Wiman, K., Hayward, W.S.,


93. Hu, Q., Dyson, N., and Harlow, E. (1990). The regions of the retinoblastoma protein needed for binding to adenovirus E1A or SV40 large T antigen are common sites for mutations. EMBO J. 9: 1147-1155.


- 148 -


146. Muller, A.J., Young, J.C., Pendergast, A.-M., Pondel, M., Landau, N.R.,


