THE CHARACTERIZATION OF A PUTATIVE MYC-INDUCED GENE

By

Justin Navada Lear

A thesis submitted in conformity with the requirements for the degree of Master of Science
Graduate Department of Molecular and Medical Genetics, Microbiology Program
University of Toronto

© Copyright by Justin Navada Lear (2000)
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

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

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

0-612-54122-3
THE CHARACTERIZATION OF A PUTATIVE MYC-INDUCED GENE

Justin Lear
M.Sc. Thesis, Department of Molecular and Medical Genetics, Microbiology Programme
University of Toronto, 2000

ABSTRACT

The c-myc proto-oncogene encodes a highly conserved nuclear protein that has been shown to be a key regulator of many cellular processes, including proliferation, differentiation and apoptosis. Deregulation of the normally rigid control of both gene and protein expression is also a frequent hallmark of tumor cell lines derived from a diverse range of cancers, and represents a potent proliferative stimulus to a cell.

The precise mechanism(s) utilized by Myc to implement its proliferation directive are unclear. The lack of a well-characterized set of Myc-induced genes is a major obstacle that must be overcome. To further this objective, I characterized a novel putative Myc-induced gene (MiG 1) previously identified in our laboratory. I have confirmed its nucleotide sequence and searched for additional sequences to explain the presence of two transcripts for this gene. I have also presented evidence indicating that this gene is induced rapidly in response to Myc activation, and demonstrated a correlation between constitutive ectopic Myc expression and increased mRNA levels of MiG 1. I have identified this novel gene as the rat homolog of alkB, (rABH; rat AlkB Homolog) a gene that was initially identified in E.coli that has been implicated in mitigating the toxic effect of Sn2 alkylating agents on cell survival. To improve our understanding of AlkB function, extensive computer modeling was performed to elucidate important characteristics of the predicted rABH protein, in addition to the identification of evolutionarily conserved regions. Lastly, I have mapped the human alkB homolog (hABH) to human chromosome 14q24 with greater accuracy than has previously been reported, to aid in the identification of diseases that might be associated with alterations in hABH.
I would like to extend my deepest and most sincere thanks and appreciation to my supervisor, Dr. Linda Penn, for your continued guidance, patience and unwavering support. You have gone well above and beyond what most other supervisors would do for a student who left before he should have, and who dropped in and out of your life on an unpredictable basis. I continue to feel welcome in your lab, in spite of the considerable demands on your time. I hope to one day be able to show you the kindness that you have shown me, and I wish you continued happiness and success in all aspects of your life. I have learned a lot from you, and thank you for sharing that with me. I would also like to thank Dr. Andy Bognar and Dr. Eleanor Fish for their input and direction throughout my studies, in addition to their willingness to help me see this chapter of my life through to its long-awaited conclusion.

I would also like to thank the members of the Penn lab, both past and present, for their friendship over the years. I would especially like to thank Wilson Marlin and Linda Facchini for tolerating my antics, and helping my development as both a student and a person, especially in the early days. I am glad that we all still keep in touch, and hope that we will remain friends for many years to come. I have yet to enjoy working as much as I enjoyed working with you. Thanks to all of my friends, especially Shaj, Noble, Eddy, and Scott for helping me laugh through the tougher times. I wouldn’t have made it without you.

My parents have taught me a great deal over the years, and I would like to thank them for sharing their knowledge. My thanks to my mother Margaret for stimulating my mind from early childhood with both questions and answers, and teaching me the joy of exploration and the thrill of discovery. I would like to thank my father Terry for checking my homework every night, and providing me with corrections that were to be done at the breakfast table. Thanks to my sister Chantelle who was my constant companion and co-discoverer of the world as we grew up together.

Lastly, I would not have been able to do this without my partner for life and best friend, Jacquie. You have made the pages that follow possible, and have helped me finish this degree, with love and support that I did not know existed until I met you. I cherish every moment that we have together, and promise more of them, now that this thesis is finally written. Only you truly understand the importance of this to me.

It may not be a conventional wedding present, but I would like to dedicate this thesis to you, with all my love.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Abstract</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xi</td>
</tr>
</tbody>
</table>

## CHAPTER 1. INTRODUCTION

1.1 A BRIEF HISTORY OF MYC INVESTIGATION ................. 1
1.2 THE MYC FAMILY .................................... 2
1.3 C-MYC GENE STRUCTURE ................................ 2
1.4 MYC PROTEIN STRUCTURE ................................ 3
1.5 THE MYC/MAK/MAD NETWORK ......................... 5
1.6 OTHER MYC BINDING PROTEINS INVOLVED IN REGULATING MYC FUNCTION ............... 6
1.7 MYC RERESSED GENES ................................ 9
1.8 MYC ACTIVATED GENES ............................. 11
1.9 THE cad GENE: THE EMERGENCE OF A BONA FIDE MYC TARGET GENE AND A MODEL SYSTEM FOR THE FUTURE STUDY OF MYC INDUCED GENES ......................... 14
1.10 THESIS OBJECTIVES .............................. 15

## CHAPTER 2. MATERIALS AND METHODS

2.1 MATERIALS ........................................... 17
   2.1.1 ANTIBIOTICS ................................... 17
   2.1.2 BACTERIAL STRAINS ............................. 17
   2.1.3 CELL LINES .................................... 17
   2.1.4 CLONING VECTORS ............................. 17
   2.1.5 ENZYMES ....................................... 17
2.2 METHODS..........................................................................................23

2.2.1 TISSUE CULTURE............................................................................23
  2.2.1.1 Revival and preparation of cryogenic cells.................................23
  2.2.1.2 Maintenance of cell lines.........................................................23
  2.2.1.3 Time course experiments........................................................24

2.2.2 ISOLATION OF DNA........................................................................24
  2.2.2.1 Cesium chloride method (large-scale preparation)......................24
  2.2.2.2 Alkali-lysis method (small-scale preparation).............................25
  2.2.2.3 Measurement of DNA yield......................................................26

2.2.3 RNA ISOLATION...............................................................................26
  2.2.3.1 Treatment of solutions and equipment to eliminate RNase Activity...............................................................26
  2.2.3.2 Preparation of total cellular RNA.............................................26
  2.2.3.3 Measurement of RNA yield......................................................27

2.2.4 DNA SUBCLONING TECHNIQUES....................................................28
  2.2.4.1 Preparation of vectors and DNA inserts......................................28
  2.2.4.2 Isolation of DNA fragments......................................................28
  2.2.4.3 Ligation of DNA fragments......................................................29
  2.2.4.4 Transformation..........................................................................29
  2.2.4.5 Diagnostic digests......................................................................30
  2.2.4.6 Preparation of Glycerol storage cultures....................................30

2.2.5 RANDOM PRIMING........................................................................31

2.2.6 cDNA LIBRARY SCREENING..........................................................31
  2.2.6.1 Preparation of plating bacteria..................................................31
  2.2.6.2 Plating the cDNA library..........................................................32
CHAPTER 2. METHODS

2.2.6.3 Primary screening.................................................................32
2.2.6.4 Secondary screening.............................................................33
2.2.6.5 Tertiary screening and \textit{in vivo} excision...............................33

2.2.7 \textbf{NORTHERN BLOT ANALYSIS}.............................................34
  2.2.7.1 Northern gel electrophoresis.............................................34
  2.2.7.2 Northern blot transfer.....................................................35
  2.2.7.3 Northern blot hybridization.............................................35

2.2.8 \textbf{RNASE PROTECTION ASSAY}............................................35
  2.2.8.1 RNA probe synthesis.....................................................35
  2.2.8.2 Sample preparation.......................................................36
  2.2.8.3 Hybridization...............................................................37
  2.2.8.4 RNase protection..........................................................37
  2.2.8.5 Radiolabelling of DNA molecular weight standards..............37

2.2.9 \textbf{POLYMERASE CHAIN REACTION}........................................38
  2.2.9.1 Primer design...............................................................38
  2.2.9.2 PCR.............................................................................38

2.2.10 \textbf{ISOLATION OF A GENOMIC hABH CLONE}..........................39

2.2.11 \textbf{FLUORESCENCE IN SITU HYBRIDIZATION}.........................39

2.2.12 \textbf{DNA NUCLEOTIDE SEQUENCING}....................................39

2.2.13 \textbf{COMPUTER RESOURCES}..................................................40

CHAPTER 3. RESULTS

3.1 \textbf{INTRODUCTION}..................................................................41
  3.1.1 THE IDENTIFICATION OF MYC-INDUCED GENES BY
       SUBTRACTIVE HYBRIDIZATION.................................................41
  3.1.2 \textit{MiG} 1: PREVIOUS DATA..................................................42
  3.1.3 THESIS OBJECTIVES AND RATIONALE....................................42

3.2 \textbf{RESULTS}............................................................................44
  3.2.1 CONFIRMING THE NUCLEOTIDE SEQUENCE OF \textit{MiG} 1........44
  3.2.2 DETAILED KINETIC ANALYSIS OF \textit{MiG} 1 RNA INDUCTION
       BY MYC.............................................................................50
  3.2.3 \textit{MiG} 1 RNA LEVELS ARE ELEVATED IN CELLS
       CONSTITUTIVELY EXPRESSING EXOGENOUS MYC.................53

vi
SEQUENCE ANALYSIS: Mig 1 is the rat homolog of AlkB..........................55

ALKB AND THE ADAPTIVE RESPONSE OF E. coli..........................58

PREDICTING THE STRUCTURE OF rABH USING COMPUTER MODELS...........59

PREDICTING THE STABILITY OF rABH...........................................63

PREDICTING THE SUBCELLULAR LOCALIZATION OF rABH..........................64

MOTIF SEARCHES...........................................................................66

SUMMARY OF THE INSIGHTS GAINED INTO rABH BY COMPUTER MODELING...........73

AN IN-DEPTH SEARCH FOR OTHER AlkB HOMOLOGS...............................74

AN IN-DEPTH SEQUENCE COMPARISON BETWEEN hABH AND rABH..................75

AN ALIGNMENT OF ALL THE AlkB HOMOLOGS, FOR THE PURPOSE OF IDENTIFYING CONSERVED SEQUENCES........80

REFINED MAPPING OF hABH................................................................84

CHAPTER 4. DISCUSSION

4.1 SUMMARY OF RESULTS.................................................................90

4.2 GENE SEQUENCE AND STRUCTURE...............................................91

4.3 IS rABH A MYC-INDUCED GENE?..................................................92

4.3.1 rABH INDUCTION OCCURS INDEPENDENTLY OF THE ESTROGEN RECEPTOR TAD.................................................................92

4.3.2 THE CHARACTERISTICS OF rABH EXPRESSION ARE CONSISTENT WITH THAT SEEN IN OTHER PUTATIVE MYC TARGET GENES.................................92

4.3.3 FURTHER EXPERIMENTS ARE REQUIRED TO DEFINITIVELY CONCLUDE THAT rABH IS A MYC TARGET GENE........................................94

4.4 rABH AND hABH: PROPOSED CHANGES TO THE PREDICTED hABH PROTEIN.................................................................96

4.5 COMPUTER MODELLING OF THE rABH PROTEIN...............................97

4.6 THE CONTRIBUTION OF rABH TOWARDS AN UNDERSTANDING OF AlkB..................................................................................99
## LIST OF FIGURES

### CHAPTER 3

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 3.1</td>
<td>The original MiG 1 cDNA sequence</td>
<td>46</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>The nucleotide and deduced amino acid sequences of MiG 1</td>
<td>47</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>A schematic representation of the probe fragments of MiG 1</td>
<td>48</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>A schematic summary of the cDNA clones identified from the screening of a 40d old rat jejunum cDNA library using the ORF probe of MiG 1</td>
<td>49</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>MiG 1 RNA levels are induced in response to Myc activation</td>
<td>52</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>MiG 1 RNA levels are elevated in cells constitutively expressing exogenous Myc protein</td>
<td>54</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>Results of a blastp algorithm search of the GenBank database, using the predicted MiG 1 polypeptide sequence as a query</td>
<td>57</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>rABH hydrophobicity plot</td>
<td>60</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>The MLRC algorithm can predict known regions of the secondary structure of the cMyc protein</td>
<td>61</td>
</tr>
<tr>
<td>Figure 3.10</td>
<td>Predicting the secondary structure of rABH</td>
<td>62</td>
</tr>
<tr>
<td>Figure 3.11</td>
<td>Alignment of the deduced human and rat AlkB protein sequences</td>
<td>77</td>
</tr>
<tr>
<td>Figure 3.12</td>
<td>Alignment of the putative rat, human, and murine AlkB cDNA sequences</td>
<td>78</td>
</tr>
<tr>
<td>Figure 3.13</td>
<td>Alignment of the deduced human and rat AlkB polypeptide sequences, after the hABH sequence was changed to match the rat and murine cDNA sequences</td>
<td>79</td>
</tr>
<tr>
<td>Figure 3.14</td>
<td>Alignment of the predicted AlkB polypeptide sequences of eight putative homologs identified to date</td>
<td>81</td>
</tr>
<tr>
<td>Figure 3.15</td>
<td>Primer design for the PCR screen of somatic cell hybrid and YAC panels</td>
<td>85</td>
</tr>
<tr>
<td>Figure 3.16</td>
<td>FISH analysis of hABH, using an anchored probe for reference</td>
<td>88</td>
</tr>
<tr>
<td>Figure 3.17</td>
<td>Combined results of hABH mapping to human chromosome 14</td>
<td>89</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

## CHAPTER 1

<table>
<thead>
<tr>
<th>Table 1.1</th>
<th>Putative Myc target genes</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>13</td>
</tr>
</tbody>
</table>

## CHAPTER 3

<table>
<thead>
<tr>
<th>Table 3.1</th>
<th>Summary of PSORT II prediction of subcellular localization for Myc and rABH</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3.2</th>
<th>Results of a search of the Prosite database using human cMyc as a query</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3.3</th>
<th>Results of a search of the Prosite database using rABH as a query</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3.4</th>
<th>Results of the GeneFIND database search system</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>71</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3.5</th>
<th>Results of the IDENTIFY database search system</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3.6</th>
<th>AlkB motifs</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>83</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3.7</th>
<th>Summary of PCR mapping of hABH</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>86</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

4-OHT 4-hydroxytamoxifen
α-MEM alpha-modified Eagle’s medium
AdMLP adenovirus major late promoter
BHI brain heart infusion
BR basic region
cad carbamoyl-phosphate synthase/aspartate carbamoyltransferase/dihyrocarnate
ddH2O distilled, deionized H2O
dhfr dihydrofolate reductase
DMS dimethyl sulfate
EIF-4E eukaryotic initiation factor 4E
ER estrogen receptor
FBS fetal bovine serum
FISH Fluorescence in situ hybridization
gadd45 growth arrest and DNA damage-inducible gene 45
gapdh glyceraldehyde phosphate dehydrogenase
GSK glycogen synthase kinase
HDAC histone deacetylase complex
HLH helix-loop-helix
IgEμ immunoglobulin heavy chain enhancer
Igκ immunoglobulin light chain kappa
Inr Initiator element
LZ leucine zipper
MAPK mitogen activated protein kinase
max Myc associated factor X
Mbl Myc homology box I
MblII Myc homology box II
MBD Myc binding domain
MiG Myc induced gene
MMS methyl-methanesulfonate
MNNG N-methyl N’-nitro-N-nitrosoguanidine
MNU methylnitrosourea
MycER Myc estrogen receptor fusion protein
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MycER™</td>
<td>Myc estrogen receptor transactivation mutant fusion protein</td>
</tr>
<tr>
<td>N-cor</td>
<td>transcriptional corepressor</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localization sequence</td>
</tr>
<tr>
<td><em>adc</em></td>
<td>ornithine decarboxylase</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PAC</td>
<td>P! artificial chromosome</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PDGF-βR</td>
<td>platelet derived growth factor β receptor</td>
</tr>
<tr>
<td>PR-</td>
<td>phenol red minus</td>
</tr>
<tr>
<td>PR+</td>
<td>phenol red plus</td>
</tr>
<tr>
<td>Rb</td>
<td>retinoblastoma protein</td>
</tr>
<tr>
<td>REF</td>
<td>rat embryo fibroblast</td>
</tr>
<tr>
<td>SV40</td>
<td>simian virus 40</td>
</tr>
<tr>
<td>TAD</td>
<td>transcription activation domain</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA binding protein</td>
</tr>
<tr>
<td>TRRAP</td>
<td>transformation/transcription-domain-associated protein</td>
</tr>
<tr>
<td>USF</td>
<td>upstream regulatory factor</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>YAC</td>
<td>yeast artificial chromosome</td>
</tr>
<tr>
<td>YY-1</td>
<td>yin yang 1</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION
INTRODUCTION

1.1 A BRIEF HISTORY OF MYC INVESTIGATION

Close to a quarter century has passed since the discovery of the avian myelocytomatosis (v-myc) gene as the gene responsible for the transforming capability of a number of avian retroviruses (Sheiness et al., 1978). Shortly afterwards, c-myc was identified as the cellular homolog to this gene (Vennstrom et al., 1982), and the localization of the human c-myc gene to chromosome 8q24 was almost immediately recognized as a location that is frequently altered in various human tumours, including translocation to one of three immunoglobulin loci that is associated with B cell lymphomas (Cole, 1986; Dalla-Favera et al., 1982; Marcu et al., 1992). Subsequent studies have shown that the normally rigid control of Myc expression is defective in a broad spectrum of tumours including those derived from breast, colon, and lung. Furthermore, alterations in Myc regulation or expression levels are among the most common genetic alterations seen in cancerous cells (Henriksson and Luscher, 1996), and may be involved in as many as one in seven cancer fatalities (Dang, 1999).

The realization that alterations in the expression and regulation of Myc have a potentially tumourigenic effect have resulted in intense scrutiny into Myc over the last two decades. Confirmation of Myc’s potent proliferation stimulus has been confirmed by the observation that overexpressing ectopic Myc can immortalize fibroblasts in tissue culture, prevent withdrawal from the cell cycle, and result in an inability to terminally differentiate (Kohl and Ruley, 1987; Mougneau et al., 1984). Furthermore, in the presence of a conflicting growth arrest signal such as the absence of sufficient nutrients, the strong proliferative stimulus of Myc results in the triggering of apoptosis (Prendergast, 1999). Although deregulated myc expression alone is insufficient to create a malignant phenotype, co-expression of myc with an activated ras oncogene in primary rat embryo fibroblasts results in malignant transformation of these cells (Land et al., 1983; Lee et al., 1985).

Our understanding of the molecular mechanism of Myc activity was advanced greatly by the discovery of Myc’s ability to regulate gene transcription through its interaction with DNA in a sequence specific manner via dimerization with another protein, named Max (Myc-associated protein X) (Blackwood and Eisenman, 1991). It is a tribute to the complex nature of Myc that despite the combined effort of numerous researchers, many questions remain unanswered regarding the precise mechanism(s) Myc employs in its role as a regulator of gene expression. There is also a great deal of uncertainty regarding the identity of specific Myc target genes and
their role in executing the wide variety of cellular processes associated with Myc. Improving our understanding of these issues will greatly advance our understanding of the cellular processes associated with Myc, and hopefully provide insight into tumourigenesis and the design of more effective treatments against cancers where Myc expression and regulation are altered.

1.2 THE MYC FAMILY

The c-myc gene is the most famous member of a family that includes the retroviral v-myc, and two closely related cellular genes, N-myc and L-myc, with N-myc sequences demonstrating the greatest similarity to c-myc. Two additional genes, S-myc and B-myc, are not as closely related, and corresponding proteins have not been isolated for these genes. Both N-myc and L-myc share the same three-exon gene structure as c-myc. Several regions of the N-Myc and L-Myc proteins share at least 80% amino acid conservation with c-Myc, and this homology is localized to the Myc box I (Mbi) and Myc box II (MbII) domains, in addition to several other motifs characteristic of the Myc family, namely the basic region (BR), helix loop helix (HLH), and leucine zipper (LZ) motifs. Myc protein structure is discussed in greater detail in Section 1.4. The expression patterns of N-myc and L-myc vary relative to c-Myc during development, and have even been shown to vary during the development of a single tissue (Mugrauer and Ekblom, 1991). N-myc and L-myc may also play important roles in neoplasia in some tissues, as evidenced by the observation that N-myc is frequently amplified in neuroblastomas, and L-myc is frequently amplified in small cell lung carcinoma (Marcu et al., 1992). The precise role of these other family members requires further study, but their roles appear to be generally more tissue-specific and of developmental importance, in contrast to the more global role played by c-myc.

1.3 C-MYC GENE STRUCTURE

The c-myc gene is comprised of three exons, but protein-encoding sequences are found only in exons two and three. The long, untranslated exon I contains two transcription start sites, designated P1 and P2, with the P2-derived transcript forming the majority of the transcript population in most tissues. Both P1 and P2 contain consensus TATA box elements, but only P2 contains a strong consensus initiator element (Inr). Transcription of c-myc is tightly controlled by an intricate and incompletely understood network that includes negative autoregulation by Myc protein (Facchini et al., 1997; Penn et al., 1990a; Penn et al., 1990b).
The P1 and P2-derived transcripts encode for two Myc proteins, with predicted molecular weights of 64kDa (Myc 2) and 67kDa (Myc 1). The larger Myc1 protein is the result of translation initiation from an in-frame, non-canonical CUG start codon that is 42 nucleotides 5' to the AUG start codon of Myc 2. Therefore, the Myc 1 protein differs from Myc 2 only by the presence of 14 additional amino acids prior to the N-terminal Methionine of Myc 2. In most tissues, the p64kDa Myc 2 predominates, and shall be synonymous with Myc throughout this thesis, unless stated otherwise. Furthermore, human Myc amino acid positions shall be used as the standard when referring to specific amino acid residues.

1.4 MYC PROTEIN STRUCTURE

Although the complete structure of Myc has not been solved, extensive experimental evidence and computer model predictions have provided a great deal of insight. From a functional perspective, Myc can be divided into two halves, linked together by an unstructured hinge region. The amino half of the protein contains sequences required for transactivation and repression of gene activity, and the carboxyl half of the protein is crucial to DNA binding and Max dimerization.

The first 143 amino acids of Myc have been shown to be the residues responsible for the activation of transcription when fused to the GAL4 DNA binding domain during in vitro assays, passing a key litmus test for acceptance of this region as a transcription activation domain (TAD) (Kato et al., 1990). The TAD can be divided into three components based on primary sequence and structure-function studies; a glutamine rich region located between amino acids 1-41, a proline rich region between amino acids 42-103, and an acidic domain between amino acids 104-143. Also found in the TAD are two highly conserved stretches of sequences between amino acids 45-63 (Mbl) and 129-143 (MbII). Both Mbl and MbII have been shown to be important to Myc function, with MbII being absolutely essential. Both Mbl and MbII have recently been shown to be important sites for interaction with other proteins that may play a role in Myc's ability to regulate gene activity.

Mbl is also a key region of post-translational regulation. Thr58 has been shown to be phosphorylated in vivo by glycogen synthase kinase (GSK), and Ser62 has been shown to be phosphorylated in vivo by both mitogen activated protein kinase (MAPK) and p34cdc2. However, the impact of phosphorylation at these sites on Myc's transactivation ability is under dispute (Gupta et al., 1993; Lutterbach and Hann, 1997). In contrast to this dispute is a strong consensus about the importance of the residues themselves. Ser62 has a positive role in regulating
Myc's proliferative capabilities, as evidenced by a drastic reduction in the capacity of Ser	extsuperscript{62} mutant Myc in cellular transformation assays (Henriksson et al., 1993; Lutterbach and Hann, 1994). Thr	extsuperscript{58} appears to be a negative regulatory site that is frequently mutated in Burkitt's lymphoma and other cancers (Henriksson and Luscher, 1996), presumably enhancing Myc's proliferative or tumourigenic properties.

The first half of Myc also contains regions that are essential to Myc's ability to repress transcription (Facchini and Penn, 1998). In particular, MblI has been demonstrated as essential and a mutation at Trp	extsuperscript{136} has been shown to eliminate both transforming and transcriptional repression activities (Brough et al., 1995; Lee et al., 1996; Penn et al., 1990b; Lee et al., 1996). Interestingly, the ability to repress transcription is separable from transactivation (Li et al., 1994; Xiao et al., 1998). The shared sequence requirements between the repression of transcription and transformation confirmed earlier indications of a link between these two biological activities (Penn et al., 1990b), and the ability of Myc to repress transcription has been suggested to be crucial to Myc-mediated transformation (Claassen and Hann, 1999).

The carboxyl half of Myc is much more thoroughly understood and contains sequences that are essential to both DNA binding and dimerization with the first identified and most important Myc-binding protein, Max (Blackwood and Eisenman, 1991). This critical region contains three different elements; the basic region (BR), the helix-loop-helix (HLH) and the leucine zipper (LZ) domain, and is commonly referred to as the BR-HLH-LZ domain.

The BR-HLH-LZ domain is characteristic of a class of proteins capable of forming dimers and binding to DNA sequences referred to as E-boxes, with a core consensus sequence of the palindromic hexamer CACGTG. Other proteins in this family include Max, the Mad and Mxi families of proteins, the upstream regulatory factor (USF), TFE3 and TFEB (Luscher and Larsson, 1999). Members of this extended family are not necessarily capable of forming hetero or homodimers with other BR-HLH-LZ proteins, and dimerization capabilities are controlled by sequences surrounding the key conserved residues. For example, Myc-Myc homodimers have not been observed in vivo. Dimerization is essential to the DNA binding capacity of BR-HLH-LZ proteins, and Myc is no exception. Dimerization is believed to facilitate binding to the major groove of DNA through the alignment of basic regions in the two proteins.

Also contained in the carboxyl half of the protein is a region between amino acids 290 and 318 that mediates non-specific DNA binding (Dang et al., 1989). These residues are believed to enhance the targeting of Myc to DNA once inside the nucleus. Indeed, there are two nuclear localization sequences (NLS) located in the carboxyl half of the protein; the primary NLS is located between residues 320-328, and has been shown to provide complete nuclear targeting. A
second, weaker NLS between residues 364-374 was found to only partially compensate for the absence of the primary NLS (Dang and Lee, 1988). Both NLS motifs are similar to the nuclear targeting sequence of the SV40 large T antigen.

1.5 THE MYC/MAX/MAD NETWORK

The discovery of Max marks a milestone in Myc study, and generated the first mechanistic model for Myc-mediated gene regulation (Blackwood and Eisenman, 1991). Max is similar to Myc in that it is a highly conserved, nuclear, BR- HLH-LZ phosphoprotein. Alternative splicing of the max gene results in two major protein isoforms; p21Max, and p22Max. The latter contains an additional 9 amino acids within the amino terminus. An NLS occurs in the 23 carboxy terminal amino acids, and the BR- HLH-LZ motif is found between amino acids 15-99. Max is also phosphorylated at Ser2 and Ser11 residues by casein kinase II and Mxi2 (Bousset et al., 1993; Koskinen et al., 1994; Zervos et al., 1993).

In contrast to Myc, Max does not possess a TAD, and is incapable of activating transcription. Furthermore, Max protein is more abundant than Myc, and max expression is unresponsive to mitogen stimulation of quiescent cells, remaining constant during quiescence and proliferation (Berberich et al., 1992; Blackwood et al., 1992). Expression levels in terminally differentiated cells appear to be cell type specific, with some cell types showing suppressed max expression (Delgado et al., 1995; Dunn et al., 1994), while others showing no change (Martel et al., 1995).

These observations led to the first proposed model of how Myc could regulate gene activity, with Myc protein levels as the central element in the regulatory paradigm. Under this model, when Myc levels are low, Max-Max homodimers occupy E-box sites and transactivation does not occur. As Myc levels rise, Myc-Max heterodimers accumulate and compete for the same E-box sites, resulting in transcription of E-box containing genes. Unfortunately, an inability to detect Max-Max homodimers in vivo conflicted with this model (Berberich et al., 1992; Koskinen et al., 1994; Sommer et al., 1998).

Revisions to the original model of Myc target gene regulation were made possible by the discovery of additional BR- HLH-LZ family members, the mad family of genes (mad1, mxi1, mad3, and mad4) and mnt (Ayer et al., 1993; Hurlin et al., 1997; Hurlin et al., 1995; Zervos et al., 1993). The expression patterns of these genes are the opposite of myc; low in proliferating cells and high in differentiated and growth arrested cells. Modest evidence in support of a role for mxi1 as a tumour suppressor is provided by the observation of frequent mutations in prostate
cancer (Eagle et al., 1995); however, there is little evidence in support of a tumour suppressor function for other mad genes.

The Mad and Mxi1 proteins contain an mSin3 binding domain (Ayer et al., 1995; Schreiber-Agus et al., 1998), and this observation has resulted in the development of a mechanistic model for the repression of Myc activated genes through Max-Mad/Mxi1/Mnt interaction with E-box elements. In this model, Max-Mad/Mxi1/Mnt interacts with mSin3, which recruits the transcriptional corepressor (N-Cor) and histone deacetylase (HDAC) to the complex at the E-box site. The resulting deacetylation of histones by HDAC prevents histones from dissociating with chromatin, thereby preventing transcription. Support for this model has been provided by reports that Mad family members and Mnt can efficiently block Myc + Ras cotransformation of rat embryo fibroblasts (Hurlin et al., 1997; Hurlin et al., 1995; Chin et al., 1995; Koskinen et al., 1994; Vastrik et al., 1995b), and block cells in the G1 phase of the cell cycle (Chen et al., 1995; Roussel et al., 1996). The Mad and Mnt proteins appear to be functionally redundant, despite having unique gene transcription profiles during differentiation (Hurlin et al., 1997; Chin et al., 1995; Vastrik et al., 1995a; Queva et al., 1998), and differing developmental consequences in mice null for mad1 or mxi1 (Schreiber-Agus et al., 1998; Foley et al., 1998). The recent discovery of a novel gene encoding a Max-like protein (mlx) that is only capable of interacting with Mad1 and Mad4 may help explain these differences and orchestrate the sequential silencing of Myc activated genes (Billin et al., 1999).

1.6 OTHER MYC BINDING PROTEINS INVOLVED IN REGULATING MYC FUNCTION

The Myc-Max paradigm has been the center of attention since the discovery of Max, but other Myc-binding proteins are beginning to emerge. Although the role of these additional proteins in Myc target gene regulation is currently poorly understood, it is an increasingly active area of research. In addition to the carboxyl half of Myc, Mbl and in particular MblI have emerged as key players in these additional Myc interactions.

C-terminal interacting proteins other than Max include Yin-Yang-1 (YY1), AP-2, and Miz1. YY-1 is a multifunctional zinc-finger DNA binding protein that has been shown to act as an initiator, activator, or repressor of transcription, and it was identified as a Myc-binding protein through a yeast-two hybrid screen with YY-1 as bait (Shrivastava et al., 1993). YY-1 and Max binding are mutually exclusive events, and YY-1 has been shown to interfere with efficient
Myc/Ras cotransformation (Austen et al., 1998). However, YY-1, cannot prevent Myc-Max heterodimer formation. Myc may interfere with YY-1's ability to regulate gene activity by hindering YY-1 interaction with TFIIB and or TBP, but more experimental evidence is required before a clear picture of the relationship between Myc and YY-1 can emerge.

The AP-2 transcription factor has been shown to bind to the BR-LHL-LZ of Myc without interfering with Myc-Max dimerization; however, AP-2 binding prevents Myc-Max DNA binding through an undetermined mechanism (Gaubatz et al., 1995). A second level of regulating Myc function may occur through direct interaction between AP-2 and its target sequence. One of the putative Myc induced genes, α-prothymosin, contains an AP-2 sites adjacent to and overlapping with Myc E-box elements, and binding of AP-2 has been shown to abrogate Myc-Max transactivation (Gaubatz et al., 1995). However, the uncertain status of α-prothymosin as a Myc target gene, and the absence of other putative target genes with AP-2 sites raise doubt as to the use of this negative regulatory system on a broad basis.

A third protein that has been shown to interact with the BR-HLH-LZ region of Myc is Miz-1, a novel zinc finger protein that was identified by a yeast two hybrid screen for proteins that interact with the carboxyl terminus of Myc (Schneider et al., 1997). Myc may play a role in negatively regulating transcription by binding to Miz-1 and preventing it from activating transcription of its targets, such as the adenovirus major late and Cyclin D1 promoters (Peukert et al., 1997). Overexpression of Miz-1 results in cell cycle arrest, but co-expression with Myc prevented cell cycle arrest and Miz-1 transactivation. Myc function may involve interaction with Miz-1 and the disruption of its activities, but there is very little experimental evidence to form a more in-depth model of the nature of this relationship.

The highly conserved MbI and MbII regions of the TAD have recently been recognized as sites of interaction between Myc and other proteins, and this area of study promises many new insights into the mechanism of Myc activity.

The retinoblastoma (Rb) family member p107 has been shown to interact in vivo with the MbII domain of Myc through its pocket domain, resulting in an inhibition of Myc transactivating properties (Beijersbergen et al., 1994; Gu et al., 1994; Hoang et al., 1995). p107 complexed with cyclin A/cdk2 has also been reported to inhibit the cell cycle dependent phosphorylation of MbI in vitro, suggesting a further regulatory role for p107 in Myc activity. The biological significance of this regulatory system in the deregulation of Myc in cancer has been questioned, since p107 has itself not been implicated in tumourigenesis, unlike Rb (Sakamuro and Prendergast, 1999). However, p107 is likely an important negative regulator of Myc function during the cell cycle.
The novel protein BIN-1 was identified by using MbII as a bait in a yeast two hybrid screen, but MbII is also required for BIN-1 interaction with full length Myc (Sakamuro et al., 1996). BIN-1 appears to be yet another protein that negatively regulates Myc function. Expression of BIN-1 inhibits Myc::Ras cotransformation of Rat embryo fibroblasts (REF) and requires binding to Myc (Elliott et al., 1999). The role of BIN-1 in regulating Myc target gene regulation is less clear, and appears to be both indirect and dependent upon the target gene. Activation from the odc promoter, but not the α-prothymosin promoter by Myc is affected by BIN-1 binding; however, the required region for this function did not include the Myc binding domain (MBD) of BIN-1. BIN-1 is either deleted or altered in several cancers, fitting the profile of a tumour suppressor (Bova and Isaacs, 1996; Cher et al., 1996). Furthermore, ectopic BIN-1 has been shown to selectively restore apoptosis in malignant breast cancer cells lacking endogenous BIN-1, and perhaps plays an important role in preventing tumourigenesis by mediating the elimination of cells where Myc expression has become elevated (Ge et al., 2000).

Interaction between the TATA Binding Protein (TBP) and Myc has been reported in vitro (Hateboer et al., 1993), but in vivo confirmation has not been obtained. However, Myc binding to YY-1 has been shown to block interactions between TBP and YY-1 or TFII-I. Recently, BIN-1 was shown to interact with TBP in vitro via sequences close to the MBD. There is insufficient evidence at this time to indicate a biological relationship between TBP and Myc that is mediated through direct interaction.

From a biological perspective, the most attractive candidate Myc binding protein is the Transformation / Transcription-domain-Associated Protein (TRRAP) TRRAP was purified by using DNA-bound Gal4-Myc fusion proteins as an affinity matrix, and a Gal4-ΔMbII Myc as a negative control (McMahon et al., 1998). TRRAP is a very large novel protein, with a molecular weight estimated at greater than 400kD. The only notable sequence similarity that was identified was to members of the PI-3 kinase family; however, TRRAP is almost certainly not functionally active for kinase activity since certain amino acids known to be essential to PI-3 kinase activity are lacking. As a result, TRRAP binding may prevent phosphorylation by other PI-3 like kinases and serve as a site for further protein-protein interactions. A yeast homolog (Tra1) has been identified, and was found to be a component of the major transcriptional regulatory complex SAGA (Saleh et al., 1998). This complex also contains histone acetyltransferases, which represents a tantalizing counterpart to the Mad/Max negative regulatory network that has been proposed to work through the recruitment of histone deacetylase activity to an E box element-containing promoter.
Evidence in support of the link between TRRAP, Myc and chromatin remodeling was recently provided when it was shown that Myc recruits the histone acetyltransferase activity of hGCN5 to DNA through its MbII-dependent interaction with TRRAP (McMahon et al., 1998). Further evidence that TRRAP plays a biological role through its interaction with Myc is found in the observation that ectopic expression of truncated TRRAP or antisense TRRAP RNA can specifically abolish Myc-Ras cotransformation Furthermore, titrating out TRRAP with dominant negative Myc mutants blocks transformation by wild type Myc. Lastly, a knock out of the yeast TRRAP homolog Tru1 resulted in growth arrest, suggesting that Tra1, and by extension TRRAP, is essential to proliferation (McMahon et al., 1999).

In summary, incremental gains have been made in recent years about the identity of proteins that associate with Myc and mediate its function. While many proteins such as AP-2, Miz-1, YY-1, and BIN-1 appear to negatively regulate Myc function, recent evidence has suggested that TRRAP may positively mediate Myc function by facilitating chromatin remodeling. Many potential binding proteins described here interact with Myc through MbI and especially MbII, and appear to directly impact Myc function. A clearer picture of these interactions is beginning to emerge, and the consequences of these interactions on Myc function are currently under investigation.

1.7 MYC REPRESSED GENES

The classical model of Myc function has centered on Myc's ability to activate the transcription of a set of target genes. Presumably, these target genes are the effectors of Myc, and perform various functions associated with Myc-mediated phenotypes, such as cell cycle progression, cellular transformation, and apoptosis. The theory that activation of transcription is the only method of Myc function has been put into question by the discovery of Myc target genes whose transcription is repressed, rather than activated by Myc. Furthermore, the discovery of a naturally occurring Myc variant, named MycS, has suggested that transcriptional activation of target genes may not be central to Myc function (Spotts et al., 1997; Xiao et al., 1998). MycS is derived from translation that initiates from an internal methionine codon located between MbI and MbII. MycS is incapable of activating transcription assayed from reporter constructs, yet retained the ability to repress transcription from reporter constructs of the Myc-repressed genes gadd45, gas1 and Adenovirus major late promoter (AdMLP). Furthermore, ectopic MycS behaved much like full-length Myc, including the induction of anchorage-independent growth and abrogating the long delay in the cell cycle of Myc null cells (Xiao et al., 1998). (Myc null
One caveat to the data generated by experiments involving MycS is that MycS is incapable of binding to TRRAP, and cannot cooperate with Ras to transform primary rodent cells (McMahon et al., 1998). The biological significance of MycS compared to Myc must take into consideration these important differences, particularly in light of the proposed importance of TRRAP in regulating Myc target gene activation.

Among the list of proposed Myc-repressed genes is c-myc (Facchini et al., 1997; Penn et al., 1990a; Penn et al., 1990b), the growth arrest and DNA damage inducible gene gadd45 (Marhin et al., 1997; Amundson et al., 1998), immunoglobulin (Ig) heavy chain enhancer (IgEμ), Ig light-chain kappa (Igk) (Sigvardsson et al., 1994), and the platelet derived growth factor β receptor (PDGF-βR) (Oster et al.). Repression of these targets occurs at the level of transcription, although the precise mechanism is unclear. Other potential target genes include gadd153, MHCII mmi-1 tsp-1, although it is unclear if this occurs through a transcriptional or post-transcriptional mechanism (Facchini and Penn, 1998).

Models of Myc-mediated repression of transcription differ from those of gene activation, since E-box elements have not been found in any of the putative Myc-repressed genes. Current models of Myc-mediated repression rely heavily on the association between Myc:Max interaction and other binding partners. These associations are suggested to block transactivation through an enhancer-dependent mechanism, or prevent the formation of transcription initiation complexes at those promoters. One model system involves the direction of Myc:Max to Inr promoter elements through TFII-I, preventing the binding of other proteins that are part of the basal transcription machinery. This model is complicated by a lack of data confirming an interaction between TFII-I and Myc, and the observation that the promoters of both gadd45 and PDGF-βR do not contain Inr elements.

Although the focus of this thesis is the study of Myc induced genes and a complete discussion of Myc gene repression is not presented here, it is important to note that Myc’s ability to repress transcription of specific target genes likely occurs at the same time as transcriptional activation, and likely plays an important role in the biological activities of Myc, most notably cellular transformation. As with Myc activated genes, the identification of additional targets and a greater understanding of Myc-binding proteins are needed to resolve this puzzle.
1.8 MYC ACTIVATED GENES

Perhaps the greatest quandary in Myc research today is the scarcity of proposed Myc induced genes, and an inability to account for the biological consequences of Myc activity with the current set of proposed target genes. The proposed set of target genes listed in Table 1.1 only contains those putative targets that have been shown to contain an E box element in or near the promoter and therefore meet an essential criterion as a direct target of Myc. The location of the responsive E box depends on the individual target gene, but in many cases is found 3' to the transcription start site. Overall, the picture that has emerged is that most of these putative target genes have not been definitively confirmed as bona fide targets, and there are many gaps in our understanding of the connection between transcriptional activation of these genes and the profound effects of Myc activity on cellular processes.

The changes in cellular processes associated with Myc almost certainly require the combined effects of changes in the expression of many direct and indirect target genes. This degree of complexity complicates the evaluation of individual target genes, since the removal of a single target gene’s function may not have a dramatic effect on a specific Myc-induced phenotype. Conversely, overexpression of the same gene is unlikely to produce the same phenotype as Myc overexpression. However, several Myc target genes have been demonstrated to play important roles in mediating cellular processes associated with Myc. For example, overexpressing ornithine decarboxylase (ODC) is sufficient to transform NIH 3T3 cells in culture (Hibshoosh et al., 1991; Moshier et al., 1993); furthermore, these cells are tumourigenic in nude mice (Auvinen et al., 1997). ODC has also been shown to mediate Myc induced apoptosis (Packham and Cleveland, 1994).

Most of the putative target genes listed in Table 1.1 demonstrate a similar expression to myc; downregulation upon withdrawal of growth factors, and upregulation in response to serum stimulation. Isolating the global proliferation signal of serum from an exclusive Myc growth signal is possible using the inducible MycER construct, and many putative targets (ODC, LDH-A, MrDb, and cdc25A) are activated upon the activation of MycER in the absence of de novo protein synthesis. However, it is also important to note that Myc is unlikely to be the sole regulator of these genes, based on the weak extent of induction that is normally observed in response to Myc.

The current list of Myc induced genes can be grouped by function into several broad categories: genes that encode rate-limiting enzymes involved in the production of compounds that are essential for the events of cell division and growth (ODC, cad, LDH-A), genes implicated in
regulating RNA structure and metabolism (*MrDb, EIF-4E, LDHA*), cell cycle regulatory genes (*cdc25A, ISGF3γ, p53*), and genes of unknown function (*α-prothymosin, ECA-39*)
<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>E box Location</th>
<th>In vivo footprinting?</th>
<th>Fold activation by MycER</th>
<th>*Misregulated in Myc null cells</th>
<th>References</th>
</tr>
</thead>
</table>
| cad  | Pyrimidine biosynthesis | Promoter | Yes | ND | Yes | Boyd and Farnham, 1997  
Boyd et al., 1998  
Eberhardy et al., 2000 |
| odc  | Polyamine biosynthesis | Intron 1 | Yes | 2-4 | No | Wagner et al., 1993  
Tavtigian et al., 1994  
Tsuneoka et al., 1997  
Eberhardy et al., 2000 |
| MrDb | RNA helicase | Protein coding sequence | ND | 2 | No | Grandori et al., 1996 |
| LDH-A | Metabolic enzyme | Promoter | ND | 2 | No | Shim et al., 1997 |
| α-prothymosin | unknown | Intron 1 | ND | 4 | No | Eilers et al., 1991  
Gaubatz et al., 1994 |
| edc-25A | Cell cycle phosphatase | Intron II | ND | 4 | No | Galaktionov et al., 1996 |
| p53  | Transcription factor | Intron 1 | ND | <2 | No | Reisman et al., 1993  
Tavtigian et al., 1994  
Shim et al., 1997 |
| elf-4E | mRNA cap-binding protein; translation initiation | Promoter | ND | ND | No | Jones et al., 1996 |
| ISG13y | Transcription factor | Promoter | ND | <2 | No | Weihsa et al., 1997 |
| RCC1 | Chromatin conformation | Intron IV, V | ND | 2 | No | Tsuneoka et al., 1997 |
| ECA-39 | unknown | 5' untranslated | ND | ND | No | Benvenist et al., 1997 |

*Bush et al., 1998

Table 1.1 Putative Myc target genes. Only Myc induced genes that have an E box element are shown. All data regarding the misregulation in Myc null cells comes from a single source (Bush et al., 1998) in vivo footprinting refers to the detection of Myc/Max binding to an E box element in vivo. ND = Not Determined.
1.9 THE cad GENE: THE EMERGENCE OF A BONA FIDE MYC TARGET GENE AND A MODEL SYSTEM FOR THE FUTURE STUDY OF MYC INDUCED GENES

Of the putative target genes listed in Table 1.1, cad has the greatest amount of evidence indicating it is a bona fide Myc induced gene where Myc is its key regulator. The cad gene encodes for a trifunctional enzyme (carbamoyl-phosphate synthase / aspartate carbamoyltransferase / dihydroorotase) that catalyse the first three rate-limiting reactions of de novo pyrimidine biosynthesis. In vivo binding of Myc / Max heterodimers to an E box element in the cad promoter has been reported (Boyd and Farnham, 1999; Boyd et al., 1998), although in vivo binding of Myc / Max to the odc and dhfr promoters has also recently been shown (Eberhardt et al.,; Boyd and Farnham, 1999). The responsive E box in the cad promoter is centered at residue +65 relative to the transcription start site. This is the closest location to the transcription start site of any putative target gene.

Cad is unique in that it is the only Myc induced gene that is downregulated in cells nullizygous for c-Myc (Bush et al., 1998). The c-myc null cell line was derived by targeted homologous recombination to delete all c-myc expression (Mateyak et al., 1997). These cell lines also lack detectable N-myc or L-myc expression, indicating that other Myc family members are not compensating for c-myc’s absence. The c-myc null cells are viable, but have extremely prolonged G1 and G2 phases, while S phase length is unaltered. The Myc-dependent misregulation of cad in Myc null cells indicates that Myc plays a central role in regulating cad gene expression; moreover, it suggests that the other proposed Myc targets do not contribute to the slow growth phenotype of Myc null cells and that Myc is not a crucial regulator of these genes. However, the relevance of cad in the development of Burkitt’s lymphoma and by extension proliferation and human neoplasia in general is unclear, since cad mRNA levels are not elevated in Burkitt’s lymphoma (Mac and Farnham, 2000).

The large body of evidence indicating that cad is a bona fide in vivo transcriptional target has allowed for its use as a model system for testing our current models of Myc target gene regulation. One issue that has plagued researchers is the criteria used by E box elements to selectively bind Myc / Max heterodimers instead of the other BR-HLH-LZ family members that have been shown to bind to E box elements, such as Mad / Max heterodimers or other proteins such as USF, TFE3, and TFEB. An analysis of Myc and USF1 in vivo binding patterns to the cad promoter via PCR amplification of DNA that had been cross-linked and then immunoprecipitated
revealed that USF1 binding remained constant after serum stimulation of quiescent cells, while Myc binding increased as Myc levels increased (Boyd et al., 1998). This result suggests that protein-protein interactions (i.e.: the amount of Myc / Max heterodimers) determine the amount of transcriptional activity from the cad promoter, rather than changes in the affinity of Myc to DNA binding. In further support of this hypothesis using stably integrated promoter constructs, USF and Myc were shown to bind to the exact same E box in the cad promoter, and site directed mutagenesis preventing Myc, but not USF binding, resulted in a lack of transcriptional activation (Boyd and Farnham, 1999).

The cad model system has also shed light on the questions raised by the reported binding of Myc / Max to non-consensus E box elements such as CATCTG, CACCTG, CATGCG, CACGCG, CAACGTG, and CACGAG (Grandori et al., 1996; Blackwell et al., 1993), demonstrating binding of Myc / Max to the non-consensus E box elements in the dhfr promoter. Dhfr and cad promoter constructs were also used to determine that the positioning of the E box was crucial to transcriptional activation, rather than flanking nucleotide sequences (Boyd et al., 1998).

The data presented above indicates that Myc transactivation is mediated by a post-DNA binding mechanism; the non-exclusive and location dependent nature of Myc / Max interactions with E box elements is consistent with the model of Myc-mediated transactivation described in Section 1.6. (i.e.: the binding of Myc – Max to E box elements located near transcription start sites results in the recruitment of histone acetylase activity through TRRAP. The acetylation of histones results in an opening of chromatin configuration, facilitating transcription). This model has very recently been tested using the chromatin immunoprecipitation system, and the experiments failed to show a correlation between Myc-mediated transcriptional activation and changes in histone acetylation for both the cad and odc promoters. This indicates that the histone acetylation model is insufficient to explain Myc-mediated transactivation, and more experimentation is required to determine the mechanism Myc utilizes to activate transcription (Eberhardy et al., ).

1.10 THESIS OBJECTIVES

The study of Myc induced genes is confounded by the absence of a strong link between gene regulation and biological function. This is exemplified by odc and cad. There is a great deal of evidence documenting the biological significance of ODC in promoting proliferation and transformation; however, odc gene expression patterns remain largely unaltered in c-myc null
cells, casting doubt on Myc's contribution to regulating this gene. The reverse case is true for *cad*, where despite a great deal of evidence indicating that Myc is a key regulator of this gene's expression, it is unlikely that *cad* expression alone is an important mediator of Myc associated transformation. The identification of additional Myc induced genes is required to elucidate the effector pathways that are activated by Myc, and may also assist in the creation of an improved model system for studying the mechanism of Myc-mediated gene regulation.

The overall objective of my thesis is to further our understanding of how the transcriptional activation of downstream target genes contributes to the various cellular processes associated with Myc, through the identification of additional Myc induced genes. This objective will be furthered through the study of a novel putative Myc target gene that was identified in our laboratory by a subtractive hybridization approach.
CHAPTER 2

MATERIALS AND METHODS
2.1 MATERIALS

2.1.1 ANTIBIOTICS

Ampicillin
Gentamycin sulphate
Kanamycin A
Tetracycline

2.1.2 BACTERIAL STRAINS

*Escherichia coli* DH5α

**Genotype:** *supE44 ΔlacU169(Hφ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*

*Escherichia coli* XL1 Blue MRF'

**Genotype:** Δ(mcrA) 183 Δ(mcrCB-hsdSMR-mrr) 173 endA1 supE44 thi-1 recA1 gyrA96 relA1

lac[F' proAB lacPZΔM15 Tn10(Tet')]

*Escherichia coli* SOLR™

**Genotype:** e14'(McrA') Δ(mcrCB-hsdSMR-mrr) 171 sbcC recB recI uvrC umuC::Tn5 (Kan') lac gyrA96 relA1 thi-1

endA1 λΔ5[F' proAB lacPZΔM15] Su' (nonsuppressing)

2.1.3 CELL LINES

The Rat-1 cell line is a subclone of the Fischer rat embryo fibroblast line F2408 (Lania et al., 1980). The Rat-1 MycER cell line has been described elsewhere (Facchini et al., 1997). The KP REF cell line is also derived from a rat embryo fibroblast cell line by spontaneous transformation. Rat 1 MycER™ cell lines were generated in the Penn laboratory (Oster et al., 2000).

2.1.4 CLONING VECTORS

pBluescript II SK
pBluescript II KS

2.1.5 ENZYMES

Alkaline phosphatase (calf intestinal)
DNA polymerase (Klenow fragment)
DNase 1 (RNase free)
Proteinase K
Restriction endonucleases and buffers
RNase A
RNase T1
RNase block
T3 RNA polymerase
T4 DNA ligase

**VENDOR**

Sigma
Roussel Canada Inc.
Sigma
Sigma
Gibco/BRL
Stratagene
Stratagene
Stratagene
Boehringer Manheim
Boehringer Manheim
Boehringer Manheim
Gibco/BRL, Stratagene
Sigma
Sigma
Stratagene
Stratagene
Pharmacia
2.1.6 KITS

cDNA library (40d old Sprague Dawley rat jejunum)  
T7 DNA Sequencing Kit  
QiAquick® PCR Purification Kit (spin column purification)

2.1.7 LABWARE

Cryopreservation tubes (1.5ml)  
Electrophoresis apparatus and gel casting equipment  
Filter, 0.2μm  
Nitrocellulose membrane  
GeneScreen Nylon Membrane  
Needles, (18G11/2)  
Polypropylene tubes  
Eppendorf, 1.5ml  
50 ml conical tubes  
15ml conical tubes  
Syringes, 10cc, 5cc  
Tissue culture dishes, 100mm and 150mm  
3M Whatman paper  
X-ray film (KODAK X-OMAT AR)

2.1.8 MEDIA AND SERA

α-Modified Eagle’s Medium, with phenol red (α-MEM, PR+)  
α-Modified Eagle’s Medium, without phenol red (α-MEM, PR-)  
Brain Heart Infusion Broth (BHI)  
Fetal Bovine Serum (FBS)

2.1.9 NUCLEOTIDES, OLIGONUCLEOTIDES, MOLECULAR WEIGHT STANDARDS

Deoxynucleotide triphosphates  
Nucleotide triphosphates  
[32P]-dATP  
[32P]-GTP  
Random primers p(dN)₆  
100 bp ladder  
1 kb DNA ladder  
λ DNA digested with Hind III

2.1.10 REAGENTS
Acetic Acid (glacial)
Agar
Agarose (SeaKem LE)
Agarose, low melting point (NuSieve)
Ammonium acetate (NH₄OAc)
Ammonium persulphate
Antifoam A Emulsion
N,N’ Bisacrylamide (30:0:8)
Boric Acid
Bromophenol blue
Calcium chloride (CaCl₂)
Cesium chloride
Chloroform
Diethyl pyrocarbonate (DEPC)
Dithiothreitol (DTT)
Dry Ice
Ethidium bromide
Ethanol, anhydrous (EtOH)
Ethylene diaminetetraacetic acid (EDTA)
Formamide
Glycerol
Glycogen
Guanidine thiocyanate (GTC)
Iso-amyl alcohol
Hydrochloric Acid (HCl)
4-Hydroxytamoxifen (4-OHT)
N-Lauroylsarcosine
Magnesium chloride (MgCl₂)
Magnesium sulphate (MgSO₄)
β-Mercaptoethanol
3-[N-Morpholino] propanesulphonate acid (MOPS)
Nitrogen (liquid)
One Phor All buffer
Phenol, buffer saturated
Potassium acetate (KOAc)
Phosphate buffered saline (-Ca⁺, -Mg⁺)
Piperazine-N,N’-bis[ethanesulphonic acid] (PIPES)
Sodium acetate
Sodium chloride (NaCl)
Sodium citrate
Sodium dodecyl sulphate (SDS)
Sodium hydroxide (NaOH)
Spermidine, trihydrochloride
N,N,N’,N”-Tetramethylenediamine (TEMED)
Tris [hydroxyxymethyl] aminomethane
Triton X-100
Xylene cyanol
2.1.11 STOCK SOLUTIONS

Deionized formamide
Add 150g of AG 501-X8 ion exchange resin to 500 ml formamide.
Allow mixture to sit for 15 minutes, then filter through 0.2 μm filter

DNA loading buffer (6X)
0.25% Bromophenol Blue
0.25% Xylene cyanol
30% (v/v) glycerol
40 μg/ml RNase A

GTC lysis buffer, pH 7.0
4 M guanidine thiocyanate
25 mM sodium citrate
10% (w/v) N-lauroylsarcosine
100 mM β-Mercaptoethanol
2 to 3 drops Antifoam A

Messing Buffer (10X)
125 mM Tris-HCl, pH 7.4
25 mM MgCl2
12.5 mM DTT

Northern Loading Buffer
50 μl 20X MOPS
175 μl formaldehyde
500 μl deionized formamide

Northern Loading Dye
50% (v/v) Glycerol
1 mM EDTA
0.4% Bromophenol blue
0.4% Xylene cyanol
NOTE: 2 μl of loading dye was mixed with 14.5 μl of loading buffer, and this mixture is then added to samples.

Northern Running Buffer
50 ml 20X MOPS
125 ml formaldehyde
bring total volume to 1L with ddH2O

Phenol / chloroform
50% (v/v) buffer-saturated phenol
48% (v/v) chloroform
2% (v/v) iso-amyl alcohol
SDS Electrophoresis Buffer (5X)
125 mM Tris base
960 mM glycine
0.5% (w/v) SDS

SDS Polyacrylamide resolving gel
7.5% (w/v) 37.5:1 acrylamide/bis
375 mM Tris-HCl, pH 6.8
0.1% (w/v) SDS
0.03% (v/v) ammonium persulphate
0.05% (v/v) TEMED

SDS Polyacrylamide stacking gel
4% (w/v) 37.5:1 acrylamide/bis
125 mM Tris-HCl, pH 6.8
0.1% (w/v) SDS
0.05% (v/v) ammonium persulphate
0.1% (v/v) TEMED

SDS sample buffer (1X)
62.5 mM Tris, pH 6.8
2% (w/v) SDS
10% (v/v) glycerol
5% (v/v) β-mercaptoethanol
1 mM EDTA
0.001% Bromophenol Blue

Solution I
25 mM Tris-HCl, pH 8.0
10 mM EDTA
50 mM glucose

Solution II
0.2 N NaOH
1% (w/v) SDS

Solution III
60% (v/v) 5 M KOAc
11.5% (v/v) acetic acid (glacial)

SSC (20X), pH 7.0
3 M NaCl
300 mM sodium citrate

TAE (10X)
2 M Tris base
2 M acetic acid, (glacial)
50 mM EDTA
2.1.12 RNASE PROTECTION ASSAY SOLUTIONS (PREPARED IN DEPC-H₂O)

Denaturing polyacrylamide gel
6% (w/v) acrylamide:bis, 19:1
7.5 M urea
1X TBE

Elution buffer
500 mM NH₄Oac
0.1% (w/v) SDS
0.1 mM EDTA

Hybridization buffer
80% (v/v) formamide
40 mM PIPES
400 mM NaCl
1mM EDTA
0.1% (w/v) SDS

RNA Loading Dye
80% (v/v) formamide
50mM Tris-borate, pH 8.3
1mM EDTA
0.1% (w/v) xylene cyanol
0.1% (w/v) Bromophenol Blue

RNase Solution
10mM Tris, pH 7.5
5 mM EDTA
300 mM NaCl
40 μg/ml RNase A (in TE / NaCl)
1μg/ml RNase T1 (in TE)

TES
10mM Tris-HCl, pH 7.5
1mM EDTA
0.1% (w/v) SDS

Transcription Buffer (5X)
200 mM Tris-HCl, pH 7.5
30 mM MgCl₂
10 mM spermidine
2.2 METHODS

2.2.1 TISSUE CULTURE

2.2.1.1 Revival and preparation of cryogenic cells

Frozen vials were removed from liquid N₂ storage and thawed rapidly in a 37°C water bath, and added to a 100mm tissue culture dish. Cells were pooled to the base of the dish by tilting forward, and 10ml α-MEM 10% (v/v) FBS preheated to 37°C was added dropwise with constant, gentle agitation to the cells. Cells were incubated for 2 to 3 hours to allow for viable cells to adhere to the tissue culture dish. Dead cells and DMSO from the now diluted cryogenic medium were removed by aspirating the medium and washing the cells with PBS, and adding 10ml α-MEM 10% (v/v) FBS, as per the usual protocol for maintaining cell lines.

In preparation for freezing, cells were grown to confluence and seeded onto a 150mm tissue culture dish at a density of 4x10⁶ cells / dish. On the next day, these subconfluent, proliferating cells were washed, trypsinized, and resuspended in 3.6ml of 50% serum / 40% medium, and pooled to the base of the dish. DMSO was added dropwise with gentle agitation to a 1/10 (v/v) concentration. 1ml aliquots were quickly transferred to pre-labeled 1.5ml cryopreservation tubes, wrapped evenly in multiple layers of paper towels to facilitate gradual, even freezing, and placed on dry ice and transferred as soon as possible to -70°C. After an overnight incubation the tubes were unwrapped from the paper towel shroud and transferred to a liquid N₂ storage tank for long-term preservation.

2.2.1.2 Maintenance of cell lines

All cell lines were cultured as monolayers in α-modified Eagle’s medium (α-MEM) supplemented with 10% (v/v) fetal bovine serum (FBS). Media was also supplemented with 100µg /ml kanamycin A and 2µg/ml gentamycin sulphate. Cell lines were incubated in a humidified 37°C incubator with 5% CO₂ and had tested negative for mycoplasma infection. Rat-1 MycER™ cells were cultured in α-MEM without phenol red (PR-) indicator.

Cells were passaged by aspirating off the media and washing the cells with approximately 20ml of phosphate buffered saline (PBS), and then treated for 30 seconds with 1ml 1x trypsin-EDTA (0.05% (w/v) trypsin; 0.53mM EDTA), pre-warmed to 37°C. Excess trypsin was aspirated, and cells were incubated for an additional 2 minutes at 37°C. Cells were resuspended in 10ml of 37°C medium and plated at a 1/10 dilution.
2.2.1.3 Time course experiments

Rat-1 MycER cells were seeded on 150mm tissue culture dishes and grown to confluence. Upon reaching confluence, the media was aspirated; the cells were washed once with PBS and replaced with 0.1% FBS media. Cells were incubated for an additional 48 hours to become quiescent. 4-OHT was then added to the media to a final concentration of 100nM with gentle agitation to dilute the 4-OHT evenly in the medium. The addition of 4-OHT was set as the zero time point, and RNA was harvested from cells at various time points after the addition of the 4-OHT. Three 150mm plates were seeded for each time point, and all three were harvested at the same time.

2.2.2 ISOLATION OF DNA

2.2.2.1 Cesium chloride method (large-scale preparation)

A single bacterial colony was inoculated into 3ml BHI broth with the appropriate concentration of selective antibiotic and incubated for 8 hours at 37°C, 250 r.p.m. The entire 3 ml culture was then inoculated into 100 ml BHI broth with continued selection and incubated overnight at 37°C, 250 r.p.m. On the following day, the culture was centrifuged in a 50 ml conical polypropylene tube for 5 minutes at 4°C, 3,700g. Two 50 ml aliquots were centrifuged in the same tube, and the supernatant was discarded after each centrifugation. The bacterial pellet was resuspended in 7.5 ml of Solution I and incubated for 5 minutes at room temperature. Bacteria were lysed by the addition of 15 ml of Solution II that had been prepared immediately prior to use. The tube was inverted gently several times until the suspension became clear and viscous, and was then incubated on ice for 5 minutes. Chromosomal DNA and proteins were then removed from solution by the addition of 11.25 ml of cold Solution III, followed by gentle, but complete mixing by inversion and then incubation on ice for 5 minutes. The lysate was centrifuged for 10 minutes at 4,600g, 4°C, and the supernatant was transferred to a fresh tube containing 20 ml isopropanol, taking care not to transfer any debris. This tube was centrifuged for 10 minutes at 4,600g, 4°C, and the supernatant was discarded. The pellet was allowed to air dry for 15 minutes, and was then resuspended in 3.8 ml TE, and 4.5g of cesium chloride was added to the solution. The DNA / cesium mixture was added to a 5 ml polypropylene dome-top ultracentrifuge tube along with 80µl of ethidium bromide [10 mg/ml] using a syringe. Volume was adjusted to 5 ml using 1.55 g/ml of cesium chloride solution and the tube was sealed, prior to centrifugation in a Beckman Vti 65.2 rotor at 65,000 r.p.m., 25°C for 4 hours. Tubes were gently
removed from the rotor, and the lower EtBr – DNA band was extracted from the tube using a 5 ml syringe and 18G needle. This band of DNA was transferred to a clean 5 ml polypropylene dome-top ultracentrifuge tube and filled with 1.55 g/ml of cesium chloride solution, sealed and centrifuged overnight in a Beckman Vti 65.2 rotor at 55,000 r.p.m., 25°C. The EtBr – DNA band was collected as before from the tube to a maximum volume of 1.5 ml and transferred to a conical 15 ml polypropylene tube. The volume was adjusted to 4 ml with TE and the DNA was precipitated with 8 ml 100% EtOH, and incubated at −20°C for at least 30 minutes. The tube was warmed briefly at 37°C to dissolve any salts that had precipitated out of solution, and the tube was then centrifuged for 25 minutes at 4,600g, 20°C. The pellet was washed once with 5 ml of 70% (v/v) EtOH, centrifuged at 4,600g, 20°C for 5 minutes, and then allowed to air dry. The pellet was resuspended in 400μl TE, transferred to a 1.5ml eppendorf tube and 400μl phenol/chloroform was added. The mixture was vortexed for 1 minute, followed by centrifugation at 16,000g for 10 minutes. The top aqueous phase was transferred to a clean 1.5ml eppendorf tube containing 40μl 3M NaOAc and DNA was precipitated with 1ml 100% EtOH. The tube was incubated on dry ice for 15 minutes, and centrifuged at 16,000g for 25 minutes. The supernatant was discarded and the pellet was washed with 500μl 70% EtOH, and centrifuged for 5 minutes at 16,000g and allowed to air dry. The pellet was resuspended in TE and stored at −20°C.

2.2.2.3 Alkali-Lysis method (small-scale preparation)

A single bacterial colony was inoculated into a 3ml BHI broth with the appropriate concentration of selective antibiotic and incubated overnight at 37°C, 250 r.p.m. 1.5 ml of culture was transferred to a 1.5 ml eppendorf tube, and cells were pelleted by centrifugation at 4,000g for 5 minutes. The supernatant was aspirated, and the pellet was resuspended in 100μl Solution I and vortexing and incubating for 5 minutes at room temperature. 200μl of freshly prepared Solution II was added, and the tube was gently inverted five times to mix, and then incubated on ice for 5 minutes. 150μl of cold Solution III was added to the tube, and mixed thoroughly by inversion, followed by incubation on ice for 5 minutes. The tube was then centrifuged at 16,000g for 10 minutes, and 400μl of supernatant was transferred to a clean 1.5ml eppendorf tube. Proteins were extracted by the addition of 400μl of phenol/chloroform, followed by centrifugation at 16,000g for 10 minutes. The aqueous phase was transferred to a clean 1.5ml eppendorf tube and DNA was precipitated by adding 1ml of 100% EtOH, and incubating on dry ice for 10 minutes, followed by centrifugation at 16,000g for 25 minutes. The supernatant was discarded, and the pellet was
washed with 500 μl 70% (v/v) EtOH, and then centrifuged for 5 minutes at 16,000g and allowed to air dry. The pellet was resuspended in 35 μl TE, and stored at -20°C.

2.2.2.3 Measurement of DNA yield

To determine the yield and purity of the DNA isolated, 2.5 μl of dissolved DNA was added to 497.5 μl distilled, deionized H₂O (ddH₂O) and mixed thoroughly. The aliquot was transferred to a quartz spectrophotometer cuvette and the absorbance of the solution at 260 nm and 280 nm was measured, using ddH₂O as the calibration standard. The concentration of DNA in the original solution was calculated as follows:

\[
\text{[DNA]} \text{ in } \mu g/\mu l = (O.D. \text{ at } 260 \text{ nm of } 1/200 \text{ dilution}) (10).
\]

The ratio of DNA: protein in the sample was calculated as follows:

\[
\frac{O.D. \text{ at } 260 \text{ nm}}{O.D. \text{ at } 280 \text{ nm}}
\]

Relatively pure samples had a ratio of between 1.7 and 2.0

2.2.3 RNA ISOLATION

2.2.3.1 Treatment of solutions and equipment to eliminate RNase activity

To remove RNase activity from autoclavable solutions, 0.1% (v/v) diethyl pyrocarbonate (DEPC) was added for a minimum of 12 hours. DEPC was removed from the solutions by autoclaving at 121°C, 100kPa for 20 minutes. For non-autoclavable and Tris- containing solutions, powdered or liquid reagents were dissolved or diluted in DEPC-treated ddH₂O in RNase free glassware or disposable plastic tubes. Glassware was rendered free of RNase activity by soaking for several hours in ddH₂O containing 0.1% (v/v) DEPC, followed by autoclaving. Alternatively, glassware was baked at 150°C for at least 6 hours.

2.2.3.2 Preparation of total cellular RNA

Cells were seeded onto 150mm tissue culture dishes and RNA was harvested at the time dictated by the experiment's requirements. Plates were washed twice with cold PBS and incubated for 2 to 5 minutes with 2 ml cold GTC lysis buffer. The lysate was scraped with a
rubber policeman to the base of the plate and transferred to the next plate of cells to be lysed. Three plates were processed sequentially in this fashion. After the third plate was scraped, the lysate was transferred to a 15 ml polypropylene tube and stored at -70°C.

To purify RNA, samples were thawed on ice, and DNA was sheared by passing the lysate a minimum of ten times through an 18-gauge needle (Sambrook et al., 1989). The lysate was then layered delicately on top of a 1.5 ml cushion of 5.7 M cesium chloride/100 mM EDTA solution in 5 ml open-top polypropylene ultracentrifuge tubes. Samples were loaded into a Beckman SW 55 ultracentrifuge rotor, and centrifuged for at least 16 hours at 48,000 r.p.m., 4°C. Following centrifugation, supernatant was carefully aspirated and the base of the tubes was removed using a sterile scalpel. RNA pellets were washed 6 to 8 times with 300 μl cold EtOH, allowed to air dry briefly, and resuspended in 300 μl cold, DEPC-treated ddH2O. RNA was precipitated in clean 1.5 ml eppendorf tubes with 30 μl 3 M sodium acetate and 900 μl EtOH. Samples were vortexed briefly, incubated, incubated on dry ice for 15 minutes and RNA was pelleted by centrifugation at 16,000 g, 4°C for 25 minutes. The supernatant was discarded, pellets were washed with 500 μl cold EtOH and centrifuged as above for 5 minutes. The supernatant was discarded and the clear, glossy RNA pellets were allowed to air dry briefly. The RNA was resuspended in 300 μl cold DEPC-treated ddH2O and stored at -70°C.

2.2.3.3 Measurement of RNA yield

To determine the yield and purity of the RNA isolated by the above method, 3.0 μl of dissolved RNA was added to 297 μl DEPC-treated ddH2O or DEPC-treated TE and mixed thoroughly. The aliquot was transferred to a quartz spectrophotometer cuvette and the absorbance of the solution at 260 nm and 280 nm was measured, using the appropriate ddH2O or TE as the calibration standard. The concentration of RNA in the original solution was calculated as follows:

\[
[RNA] \text{ in } \mu\text{g/μl} = \frac{(O.D. \text{ 260nm of } 1/100 \text{ dilution})}{4}.
\]

The ration of RNA: protein in the sample was calculated as follows:

\[
\frac{\text{O.D. 260nm}}{\text{O.D. 280nm}}
\]
Relatively pure RNA isolations without much protein typically had desired values between 1.3 and 1.7.

2.2.4 DNA SUBCLONING TECHNIQUES

2.2.4.1 Preparation of vectors and DNA inserts

a) Restriction endonuclease digestion.
Approximately 1 to 2μg of vector DNA and enough plasmid DNA to yield 1 to 2μg of DNA insert were digested with the appropriate restriction endonucleases. The amount of plasmid DNA required for the digestion was calculated according to the following formula: Total amount of DNA required (μg) = [total plasmid size (kbp) ÷ [insert size (kbp)] [amount of insert required (μg)]. Reaction mixtures were assembled on ice in chilled 1.5ml eppendorf tubes and components were added in the following order: ddH2O, restriction endonuclease buffer, DNA, restriction enzymes. Buffer concentrations and total reaction volumes depended on the type and amount of enzymes used. In general, 2 to 5 U of enzyme were used per μg DNA, and the final enzyme concentration was never greater than 1/10 of the total reaction volume. Reactions were routinely incubated for 2 to 3 hours at the temperature required for optimal enzymatic activity.

b) Phosphatase treatment of vector DNA
To prevent recircularization of empty vector DNA, linearised vectors were treated with 1μl of calf intestinal phosphatase for 30 minutes at 37°C. The phosphatase was inactivated by the addition of 2μl 0.5 M EGTA pH 8.0 and incubated at 68μ°C for 15 minutes.

2.2.4.2 Isolation of DNA fragments

Digested DNA was mixed with 1/6 v/v of 6X DNA loading buffer and incubated at 37°C for 15 minutes. Samples were loaded into wells of 0.6% (w/v) low-melting temperature agarose gels in 1X TAE. Approximately 500ng of either λ / Hind III or 1kb DNA ladder was used as a molecular weight size marker. Samples were electrophoresed under 1X TAE buffer at 4°C, 70 V to 100V for 2 to 4 hours, or until the desired separation of fragments was achieved. After electrophoresis, the low-melting temperature agarose gel was placed on a clean UV lightbox. The desired band(s) were excised from the gel under longwave UV light, using a sterile scalpel blade.
for each fragment that was cut. The excised piece of gel was placed into a 1.5ml eppendorf tube, with an average total volume ≈40µl. Tubes were sealed with parafilm and stored at 4°C.

2.2.4.3 Ligation of DNA fragments

DNA fragment containing low-melting temperature agarose gel pieces were heated to 65°C for 2 to 3 minutes. During this time, ligation reaction mixtures were prepared on ice in 1.5ml eppendorf tubes. For each ligation, 2µl of molten agarose containing vector DNA and 4 to 8µl of molten agarose containing insert DNA were added to 1X One Phor All buffer (Pharmacia) with 1mM ATP. A reaction mixture containing only 2µl of vector DNA was prepared to determine the percentage of recircularized empty vector. As a positive control for the ligation reaction conditions, 2µl of a gel slice containing pBlueScript DNA that had been linearised with a single restriction endonuclease and untreated with phosphatase, was added to a separate reaction mixture. Approximately 2 to 4 U of T4 DNA ligase was added to each mixture and reactions were allowed to proceed for 4 to 16 hours at room temperature. Following ligation, reaction mixtures were incubated at 68°C for 15 minutes to inactivate the enzyme.

2.2.4.4 Transformations

a) Preparation of competent bacterial cells

A single colony of Eschericia coli DH5α cells was inoculated into 5ml TYM broth and grown for 2h at 37°C, 250 r.p.m. The culture was added to 100ml fresh TYM broth and incubated at 37°C, 250 r.p.m. for a further 2 to 3 hours until the O.D. 550nm measured 0.5. The culture was transferred to 2 x 50ml conical polypropylene tubes and the cells were pelleted by centrifugation at 3,000g, 0°C for ten minutes. The supernatant was discarded, the tubes were wiped dry, and the pellets were each thoroughly resuspended in 20 to 40 ml ice cold TFB I, pH 5.8. To maximize transformation efficiency, cells were incubated on ice for ten minutes. Following incubation, cells were gently pelleted by centrifugation at 1,700g, 0°C for ten minutes. Once again, the supernatant was discarded, the tubes dried, and the cells gently resuspended in a total of 4ml ice-cold TFB II. Competent cells were quickly aliquoted into pre-chilled sterile eppendorf tubes and stored at -70°C.
b) **Chemical transformation of competent bacterial cells**

Competent *E. coli* DH5α cells were thawed on wet ice. Cells were gently mixed, and 50μl aliquots were transferred to chilled 1.5ml eppendorf tubes. Any unused were discarded. To determine transformation efficiency, 0.05ng of pUC19 plasmid was added to one tube. To the remaining tubes, approximately 1 to 10ng of DNA (1 to 2μl of a single ligation reaction) was added and cells were incubated on ice for 30 minutes. Cells were heat-shocked at 37°C for 20 seconds and then returned to ice for 2 minutes. To each tube, 950μl of room temperature BHI broth was added and mixed by pipetting. The cells were then incubated at 37°C for 1 hour. For the pUC19 transformation control reaction, 100μl of a 1:100 dilution was spread onto a BHI agar plate containing 100μg/ml ampicillin (selection plate). For cells transformed with DNA from ligation reactions, 100μl of undiluted reaction was spread onto selection plates. The remaining cells were concentrated by centrifugation at 1,800g for 2 to 4 minutes, and 100μl of the concentrated cells were also plated on selection plates. Plates were allowed to dry, inverted, and incubated overnight at 37°C.

2.2.4.5 **Diagnostic digests**

To verify the plasmid construct that had been created, DNA was extracted from several single colonies arising from the plated ligation reactions as described in Section 2.2.2. Approximately 3 to 5 μg of DNA was analyzed by restriction endonuclease digestions as described in “restriction enzyme digestion”. Digested plasmids were visualized on 0.8 to 1.2% agarose gels, or 3% NuSieve agarose as described in “Isolation of DNA fragments”.

2.2.4.6 **Preparation of Glycerol storage cultures**

A single colony or a loopful of broth culture was inoculated into 3ml BHI broth containing the appropriate concentration of antibiotic, and incubated overnight at 37°C, 250 r.p.m. The following day, 800μl of log phase, overnight culture was transferred aseptically to each of two 1.5ml cryopreservation or eppendorf tubes. 200μl of sterile, 75% v/v glycerol was aseptically added to the cultures and mixed thoroughly by pipetting. Tubes were capped and immediately frozen in a dry ice / EtOH bath for 15 minutes. Tubes were then transferred to -70°C for long-term storage.
2.2.5 RANDOM PRIMING

Probe DNA was purified through restriction digestion and electrophoresis in low melt agar. The excised low melt agar gel slice (approximately 40μl, estimated to contain was heated at 65°C for 5 minutes, and 5μl of this was added to an eppendorf tube containing 1μl (1mg/1ml) of random primers p(dN)₆ and 14μl ddH₂O. Probe DNA was denatured by boiling for 5 minutes, then hybridized to the random primers at room temperature for 15 minutes. Radiolabelling was performed by the addition of 3μl of 10x Messing buffer, 1μl (33mM) dNTPs, 5μl (3,000Ci/m mole, 10μCi/μl) of α²³PdATP, 2μl of Klenow fragment DNA polymerase (1,000U/ml), and incubating at 37°C for 1 hour. The probe was purified using a spin column purification system (Qiagen) as per the manufacturer's instructions. Purified probe was eluted in TE. The probe's specific activity was measured, and generally ranged between 1x10⁷ and 1x10⁸ cpm/μg.

2.2.6 cDNA LIBRARY SCREENING

A random and oligo dT primed cDNA library generated from the jejunum of a 40d old Sprague Dawley rat (Rat Jejunum cDNA Library, Lambda Zap II vector, Stratagene, catalog # 936511) was used, and complete information regarding the characteristics and instructions for use are described in the package insert. The average insert size of the library was estimated to be 1.0kb, and inserts greater than 0.5kb were selected for during its construction. The library had been engineered so that EcoR1 restriction sites flanked the ends of each insert. The library had an estimated titer of 8.0 x 10¹⁰ pfu/ml with an estimated background of 2% nonrecombinants. The library had already been through one round of amplification. The recommended host strain of E.coli for this library is XL1-Blue MRF⁺.

2.2.6.1 Preparation of plating bacteria

E.coli XL1 Blue MRF⁺ cells were streaked from frozen stock onto an LB tetracycline (12.5μg/ml) agar plate, and incubated overnight at 37°C. A single colony from this plate was inoculated into a 5ml LB Broth, 0.2% w/v maltose, sterile 10mM MgSO₄. The tube was placed into a shaking (250r.p.m.) 37°C incubator for 4 to 6 hours. O.D.₆₀₀ readings were taken periodically to ensure that culture did not overgrow past O.D.₆₀₀ = 1.0. Bacteria were pelleted at
2,000 r.p.m. for 10 minutes, and gently resuspended in 2.5ml sterile 10mM MgSO$_4$. The cells were then diluted to an O.D.$_{600} = 0.5$ with sterile 10mM MgSO$_4$. These plating bacteria could then be used immediately or stored at 4°C for a maximum of 48 hours.

2.2.6.2 Plating the cDNA library

Enough library was plated to screen $1 \times 10^6$ plaques, at a density of 50,000 plaques per 150mm plate (20 plates). A sufficient quantity of library to generate the indicated density of plaques was added to 600μl of plating bacteria (O.D.$_{600} = 0.5$), and incubated at 37°C for 15 minutes with gentle shaking to allow the phage to attach to the cells. 6.5ml of NZY top agar (48°C) was added, mixed well and poured immediately onto an NZY agar plate (less than 2 days old). The top agar was allowed to set for 10 minutes, and the plates were then placed upside down in a 37°C incubator. Plaques normally started to become visible within 6 to 8 hours. Plaques were grown to a sufficient size that they were visible but did not overlap surrounding plaques. The plates were chilled for 2 hours at 4°C to prevent the NZY top agar from sticking to the nitrocellulose membrane. While wearing gloves and using forceps, nitrocellulose membrane (Gene Screen Plus Nylon Membrane, DuPont) was placed on top of the 150mm plate. Transfer to the membrane was allowed to continue for two minutes. Afterwards, a needle was used to prick through the membrane and agar for orientation. A duplicate transfer using a second membrane was carried out for 4 minutes. After transfer, the membrane was lifted from the plate and denatured by submerging in a 1.5M NaCl, 0.5M NaOH denaturation solution for 2 minutes. The membranes were also gently brushed with gloved fingertips to remove any NZY top agar. The membranes were neutralized for 5 minutes by submerging in a 1.5M NaCl, 0.5M Tris-HCl (pH 8.0) solution. The membrane was then rinsed for a maximum of 30 seconds by submerging in a 0.2M Tris-HCl (pH 7.5), 2x SSC solution. Membranes were blotted briefly on Whatman 3MM paper to remove excess liquid and cross-linked using the Stratalinker UV crosslinker (Stratagene) set at 120,000μJ of UV energy. Agar plates were stored at 4°C for use after screening.

2.2.6.3 Primary screening

Prehybridization solution was heated to 50°C prior to the addition of salmon sperm DNA. Salmon sperm DNA was preboiled for 10 minutes and added to the warm Prehybridization solution. A container that is as similar as possible to the shape of the filter was used to minimize the amount of solution required. Each membrane was wet quickly in the hybridization solution in
a plastic Tupperware container, placing each membrane on top of the previous membrane, until all membranes have been added. Approximately 2 to 3 ml of hybridization solution per membrane was required to ensure that all membranes were covered. High specific activity probe was generated by random priming (Section 2.2.5), and approximately $1 \times 10^7$ counts/membrane were used to screen the library. Both probe and salmon sperm DNA were boiled for 4 minutes prior to being added to hybridization solution that had been preheated to 42°C. The hybridization solution in the Tupperware container was decanted, and the probe-containing hybridization solution was added to the container. Hybridization was allowed to proceed overnight at 42°C with shaking at 60 r.p.m. On the following day, the hybridization solution was decanted, and the blots were washed with gentle shaking for five minutes at room temperature with 0.1xSSC, 0.1% (w/v) SDS, and then for 15 minutes in 0.1xSSC, 0.1% (w/v) SDS, that had been preheated to 65°C. After washing, the membrane was briefly rinsed in 2x SSC and blotted on Whatman 3MM paper to remove excess liquid. The membranes were placed between two sheets of plastic wrap in a cassette with intensifying screens and exposed to autoradiography film overnight at -70°C.

2.2.6.4 Secondary screening

Developed autoradiographs were oriented with their respective membranes and plates to determine the location of the strongest putative clones based on signal strength. A 1cm$^2$ hole was cut from the stock plate at the spot where the putative clone lined up and transferred to an eppendorf tube containing 1ml of SM buffer and 20μl of chloroform and vortexed. The cored putative clone was titered and plated with host cells on a 100mm NYZ agar plate with 3ml NZY top agar such that one plate had approximately 50 plaques and the second plate had approximately 450 plaques. Incubation, membrane transfer, and screening were carried out as described above.

2.2.6.5 Tertiary screening and in vivo excision

Tertiary screening was performed, following the same protocol as for secondary screening, and cored isolated plaques were stored in 0.5ml of SM buffer with 20μl chloroform. The tube was vortexed to release the phage particles into the SM buffer. Tubes were then placed at 4°C for storage. Overnight cultures of XL1-Blue MRF$^+$ and SOLR cells in 5ml LB broth were set up. Next day, 0.5ml of each culture was used to inoculate a fresh 50ml LB broth culture. This
culture was grown at 37°C for 2 to 3 hours to mid log phase (OD₆₀₀ = 0.2 to 0.5). XL1-Blue MRF⁺ cells were gently spun down at 1500g and resuspended to an OD₆₀₀ = 1.0 in 10mM MgSO₄. SOLR cells were allowed to continue growing until reaching an OD₆₀₀ = 0.5 to 1.0 and removed from the incubator and kept at room temperature until ready to use. 200μl of XL1-Blue MRF⁺ cells at OD₆₀₀ = 1.0 in 10mM MgSO₄, 250μl of phage stock and 1μl of ExAssist helper phage were combined in a 50ml conical tube and incubated at 37°C for 15 minutes. 3ml of LB broth was added to the tube and incubation continued for another 2 to 2½ hours. After incubation, the conical tube was heated to 70°C for 15 minutes and then centrifuged at 4,000g for 15 minutes. The supernatant, which contained the excised pBluescript phagemid packaged as filamentous phage particle, was then decanted into a sterile conical tube. Excised phagemids were recovered by mixing two eppendorf tubes of 200μl of freshly grown SOLR cells and 100μl and 10μl of phage supernatant, respectively. The tubes were incubated for 15 minutes at 37°C. 10-50μl from each eppendorf tube was plated on LB-ampicillin agar plates (50μg/ml) and incubated overnight at 37°C. Glycerol stocks of single colonies were made for long-term storage of each clone. The size of each insert was estimated by performing restriction digests using EcoRI.

2.2.7 NORTHERN BLOT ANALYSIS

2.2.7.1 Northern gel electrophoresis

1% w/v agarose gel was prepared by dissolving 1g of agarose in 55ml ddH₂O and 5ml 20X MOPS buffer, and then cooled to 50°C before adding to a 37°C mixture of 18ml (38%) formaldehyde in 22ml ddH₂O. The mixture was poured into a gel cast set up in a fume hood and allowed to solidify for 3 hours. The gel was then removed from its cast and placed in the electrophoresis apparatus, and overlaid with 1L of northern running buffer.

RNA samples (20μg of total RNA) were EtOH precipitated and washed with 70% EtOH. The pellet was allowed to air dry for 20 minutes, and was resuspended in 20μl of 0.1% v/v DEPC ddH₂O. 1/6 v/v of Northern loading buffer – dye was then added to the sample, mixed, and the RNA sample was denatured by heating the mixture to 65°C for 5 minutes and then placing immediately on ice. The samples were loaded onto the gel and electrophoresed for 18 hours at a constant voltage of 25V. The gel was removed from the apparatus and rinsed briefly in ddH₂O to remove excess formaldehyde. The gel was then base treated by submerging in 50mM NaCl, 100mM NaOH for 20 minutes, and then neutralized in 1mM Tris pH7.6.
2.2.7.2 Northern blot transfer

A solid support was placed in a reservoir containing 10X SSC. A wick of 3MM paper equilibrated in 10X SSC was placed on the solid support with its ends dipped in the 10X SSC reservoir. Three pieces of 3MM Whatman paper cut to the size of the gel were placed on the wick. The gel was then flipped so that the RNA was facing up, and placed on the Whatman paper. Air bubbles that could interfere with the RNA transfer were squeezed out from under the gel by rolling a glass pipette over the surface of the gel. GeneScreen membrane equilibrated in 10X SSC was placed on top of the gel, followed by three sheets of 3MM paper, topped with a 7-10 cm thick stack of paper towels. A weight was placed on top of the paper towels to aid in capillary action.

After transfer was complete 18 – 24 hours later, the paper towels and filter papers were removed. The position of the wells was marked on the membrane in pencil. The membrane was rinsed briefly in 2X SSC and then preserved by UV cross-linking (120,000 μJ of UV energy) in the Stratalinker UV crosslinker (Stratagene).

2.2.7.3 Northern blot hybridization

The procedure used was essentially the same as that described in Section 2.2.6.3, with the only a few changes. First, prehybridization and hybridization occurred overnight in sealed bags submerged in a water bath. Second, washes were performed as follows: (1) two washes of 2X SSC, 0.1% (w/v) SDS at room temperature for 5 minutes; (2) two washes of 0.2X SSC, 0.1% (w/v) SDS at 42°C for 15 minutes; (3) two washes of 0.1X SSC, 0.1% (w/v) SDS at 68°C for 15 minutes, and a brief rinse in 2X SSC prior to sealing the blot in a bag for autoradiography.

2.2.8 RNASE PROTECTION ASSAY

2.2.8.1 RNA probe synthesis

To synthesize radiolabelled, antisense RNA probes, 1 μg of template DNA was incubated with 1X transcription buffer, 10 mM DTT, 1/20 vol. RNase block, 500 μM ATP, 500 μM CTP, 500 μM TTP, 50 μCi [32P]-GTP and 10 to 20U T3 RNA polymerase in a total volume of 10 μl. The reaction was allowed to proceed at 37°C for 90 minutes. Transcription was stopped by adding 1 μg DNase I (RNase free), and incubating for 15 minutes at 37°C. The reaction volume was
adjusted to 100μl with DEPC-treated ddH₂O and proteins were extracted from the mixture by adding 100μl of phenol/chloroform. The mixture was vortexed for 1 minute and centrifuged for 10 minutes at 16,000g, 4°C. The aqueous (upper) phase was transferred to a clean 1.5ml eppendorf tube and RNA was precipitated with 1μl glycogen, 50 μl 5M ammonium acetate, and 450 μl 100% EtOH. Tubes were placed on dry ice for 10 to 15 minutes, and then centrifuged at 16,000g, 4°C for 25 minutes. The supernatant was decanted and the small pellet was washed with 500 μl of 70% EtOH. The tube was centrifuged at 16,000g, 4°C for 5 minutes, and the supernatant was decanted. The pellet was allowed to air dry, and was resuspended in 10 μl RNA loading dye. The tube was then placed in a boiling water bath for 5 minutes, chilled immediately on ice for 5 minutes, and loaded onto a 0.4 mm, 6% denaturing polyacrylamide gel. Samples were electrophoresed in 1X TBE buffer at a constant power of 65W, 1,800V for 1.5 hours. Following electrophoresis, the gel was covered with plastic wrap and exposed to pre-wrapped X-OMAT AR film for 45 seconds. The film was developed, re-oriented on the gel, and portions of the gel containing the purified riboprobes were excised, using the film as a guide, and transferred to clean 1.5ml eppendorf tubes. Probes were eluted from the gel in 400μl Elution buffer (500 mM NH₄OAc, 0.1% (w/v) SDS, 0.1mM EDTA) at 37°C, overnight. Polyacrylamide gel fragments were pelleted by centrifugation at 16,000g, 4°C for two minutes. The supernatant was transferred to a clean 1.5ml eppendorf tube and the riboprobe was precipitated as describe above, only with 2μl glycogen and 1ml 100% ethanol. The pellet was washed with 70% EtOH, allowed to air dry, and then resuspended in 100 μl TES. The specific activity of the probe was measured as counts per minute (c.p.m.)/μl. A probe with a specific activity greater than 10⁵ c.p.m./μl was considered acceptable.

2.2.8.2 Sample preparation

Purified RNA samples were thawed on ice and the RNA concentration was determined. An appropriate volume of RNA containing 10μg of RNA was mixed with 1μl glycogen, 10 μl 3M sodium acetate, and DEPC-treated ddH₂O to a final volume of 100μl, and precipitated with 300μl 100% EtOH. Samples were incubated on dry ice for 10 to 15 minutes and centrifuged for 25 minutes at 16,000g, 4°C. Pellets were washed with 500μl 70% EtOH, and centrifuged at 16,000g, 4°C for 5 minutes. The supernatants were discarded and the pellets were allowed to air dry thoroughly.
2.2.8.3 Hybridization

An excess of radiolabelled antisense RNA probe was prepared with hybridization buffer to a total volume of 32μl. The relative amounts of probe and hybridization buffer were determined by the specific activity of the particular probe(s) that had been prepared. The 32μl probe/hybridization buffer mixture was added to the RNA pellet. The tube was vortexed, placed in an 85°C water bath for 3 minutes. The tube was vortexed again, and returned to the 85°C water bath for 8 minutes. The tube was then transferred to a 52°C water bath and hybridization was conducted overnight.

2.2.8.4 RNase protection

After overnight hybridization, single stranded non-hybridized RNA was digested by adding 350 μl of RNase solution to each sample. Samples were inverted gently to mix, and incubated at 30°C for 1 hour. To stop the digestion, 10μl of 20% SDS (w/v) and 10μl of Proteinase K [10mg/ml in TE] and were added, sequentially and in order to each sample, and the samples were incubated at 37°C for thirty minutes. Proteins were extracted from the samples by the addition of 400μl phenol/chloroform, followed by vigorous vortexing and centrifugation at 16,000g, 4°C for ten minutes. 350μl of the aqueous phase was transferred to a clean 1.5ml eppendorf tube and protected RNA fragments were precipitated as previously described, with 1μl glycogen and 1ml 100% EtOH. Pelleted RNA was washed with 500μl 70% EtOH and allowed to air dry thoroughly. Pellets were resuspended in 4μl RNA loading buffer, boiled for 5 minutes, and then placed immediately on ice. Samples were loaded onto a 0.4mm 6% denaturing polyacrylamide gel and electrophoresed in 1X TBE buffer at constant power of 1,600V. Following electrophoresis, the gel was dried for 1 hour on a gel dryer (BioRad) and exposed to autoradiography film with an intensifying screen overnight at -70°C.

2.2.8.5 Radiolabelling of DNA molecular weight standards

To radioactively label DNA molecular weight standards, 1μg of 1kb DNA ladder was incubated with 2.5x Messing buffer, 33mM dCTP, 33mM dGTP, 33mM dITTP, 50μCi[32P]-dATP and 10 to 15 U Klenow fragment of DNA polymerase, at room temperature for 5 minutes. Following incubation, the volume of the reaction was adjusted to 100μl with ddH2O, and proteins
were extracted with 100μl phenol/chloroform. The aqueous phase was transferred to a clean tube and the DNA pellet was washed with 70% EtOH. The pellet was air dried and resuspended in 100μl RNA loading buffer. 2 to 3 μl of this final preparation was loaded onto the 0.4mm 6% denaturing polyacrylamide gel; slightly higher amounts were used if the radiolabelled dATP was particularly old or the radiolabelled DNA ladder was several weeks old.

2.2.9 POLYMERASE CHAIN REACTION

2.2.9.1 Primer design

Two PCR primer set sequences were selected from hABH sequences that differed from the rABH sequence sufficiently, and where possible, the last nucleotide of the primer sequence was selected to be a mismatch between rABH and hABH. One primer set was designed from sequences in the rBAH coding region, and the other set was designed with one primer located in the coding sequence of rABH, while the other primer was 3' to the stop codon at position 1165, relative to position 1 as the 5' end of the cDNA (see Figure 3.15). Primer pairs were checked by the OLIGO 4.0 software program to ensure that primer dimers and hairpin loops did not occur, in addition to predicting the Tm of each primer. Primers were designed to have similar Tm's, and were also checked against GenBank using the blast algorithm to ensure that primer sequence was similar only to hABH.

2.2.9.2 PCR

100ng of target DNA was added to a PCR microfuge tube, along with a sufficient volume of ddH2O to bring the total reaction volume to 50μl. Next, a mixture of 10X PCR buffer, 0.2μM PCR primers, 10μM mixture of dNTPs was added to the PCR microfuge tube. The tube was mixed thoroughly by pipetting and the tubes were placed inside a thermal cycler (Perkin Elmer) on a hold program at 4°C. Lastly, 2U of Taq DNA polymerase/reaction (kept on ice) were added to each reaction tube, the tubes were capped shut, and the lid on the thermal cycler was put in place. The following program was then used to amplify target DNA: 98°C, 10 minutes, followed by 50 cycles of 94°C for 40 seconds, 54°C for 40 seconds, 68°C 1 minute. A final ten minutes at 72°C and then indefinite hold at 4°C completed the amplification process. PCR products were
electrophoresed in a 3% agarose gel containing in 1X TBE for 45 minutes at a constant 100V. Products were visualized by ethidium bromide staining.

2.2.10 ISOLATION OF A GENOMIC hABH CLONE

A genomic hABH clone was obtained by screening a series of Southern blots containing EcoR1 digest of DNA from human chromosome 14 cloned into P1 Artificial Chromosomes (PACs). (Dr. Steve Scherer, Hospital for Sick Children). The Pst1/EcoR1 subclone fragment of rABH was labeled by random priming and used to probe the Southern blots under conditions of high stringency as described in Section 2.2.6.3.

2.2.11 FLUORESCENCE IN SITU HYBRIDIZATION (FISH)

Mapping of the hABH genomic clone was performed by FISH, essentially as described in (Lichter et al., 1990). In sum, FISH analysis was performed on a metaphase spread of normal human lymphocytes (NHLs) counterstained with 4',6-Diamidin-2-phenylindol-dihydrochloride (DAPI). The PAC 22 anchor probe was kindly provided by Dr. A.F. Roux (Dr. Diane Cox, Hospital for Sick Children) and was labeled with biotin and detected with FITC (green fluorescence). The hABH probe was labeled with digoxigenin and detected with rhodamine (red fluorescence). Separate images of DAPI counterstained and hybridization signals from the double digoxigenin/avidin were captured by a thermoelectrically cooled charge coupled camera (Photometrics, Tucson, AZ) and electronically overlaid and merged using Adobe Photoshop™ 3.0. The colors of the probe signals were digitally altered (yellow for FITC, and red for rhodamine).

2.2.12 DNA NUCLEOTIDE SEQUENCING

Direct nucleotide sequencing was performed with the T7 DNA Sequencing Kit (Pharmacia), according to the manufacturer's protocol. Briefly, 1.5 to 2.0µg of double stranded plasmid DNA from a minipreparation of DNA was denatured in 400mM NaOH for 10 minutes at room temperature. Denatured template was precipitated, resuspended in ddH2O and 2.5 to 5µM of primer was annealed at 37°C for 10 minutes. DNA was synthesized and labeled with [35S]-dATP by incubating with T7DNA polymerase at room temperature for 5 minutes. The labeling reaction was divided among 4 separate tubes, each containing a different dideoxynucleotide to
prevent further polymerization after the incorporated dideoxynucleotide, and incubated at 37°C for 3 minutes. Samples were mixed with loading dye and incubated at 80°C for 2 minutes. Electrophoresis of the sequencing reaction products was carried out on a 0.4mm, 6% denaturing polyacrylamide gel cast with sharks-tooth combs, that had been pre-run for at least 20 minutes to allow the gel to warm up to 45°C. Samples were run in duplicate; one set was loaded first and the second set was loaded for the remaining half of the time for electrophoresis, normally about 5 hours. Electrophoresis was carried out under constant voltage of approximately 1,800W and constant temperature of 45°C. After electrophoresis, sequencing products were visualized by autoradiographic film that had been exposed overnight at room temperature without intensifier screens. Alternatively, DNA sequencing was performed using a Perkin Elmer 310 instrument, as per the manufacturer’s instructions.

2.2.13 COMPUTER RESOURCES

Digital copies of autoradiographs were made using a 300dpi Hewlett Packard ScanJet IIc scanner. Image contrast and brightness were adjusted using Adobe PhotoShop 3.0 and imported into Illustrator 6.0 for text labeling. Figures were printed on a MacIntosh LaserWriter Select printer at 600 dpi or on a Canon BJC 4300 printer.
CHAPTER 3

RESULTS
3.1 INTRODUCTION

3.1.1 THE IDENTIFICATION OF MYC-INDUCED GENES BY SUBTRACTIVE HYBRIDIZATION

The current group of proposed Myc induced genes includes genes that are important to proliferation such as *odc, cad* and *cdc25A* (Cole and McMahon, 1999; Dang, 1999; Facchini and Penn, 1998); however, they fail to explain the mechanism whereby Myc exerts its extremely strong proliferative signal. Furthermore, these target genes do not explain Myc’s ability to trigger apoptosis in the absence of a strong survival signal. Based on these observations, the list of Myc induced genes is likely incomplete.

To isolate genes transcriptionally induced by Myc, our laboratory employed the inducible Rat-1 MycER system and a subtractive hybridization approach. (R. Lu, Post-doctorate fellow, Dr. Penn’s laboratory) Rat-1 cells were selected as the model system for the subtractive hybridization for several reasons. First, Rat-1 cells lack endogenous estrogen receptors, and are therefore unresponsive to βestradiol or its antagonists. Second, Rat-1 cells can be induced into quiescence by withdrawal of growth factors and contact inhibition; furthermore, they can re-enter the cell cycle after prolonged quiescence by the addition of growth factors to the medium.

The second element of the system is the MycER construct. The MycER fusion protein is the product of a chimeric gene composed of the human c-myc gene fused in-frame at its 3’ end to the hormone-binding domain of the human estrogen receptor. The MycER fusion protein is inactive for all of Myc’s functions, until the ER portion of the protein binds to βestradiol or one of its antagonists. The sudden de-repression caused by βestradiol binding to the ER leads to a rapid and dramatic increase in the amount of functionally active Myc protein in the Rat-1 MycER expressing cells. This system allows us to examine the effects of Myc activity alone, in the absence of the multitude of changes in gene regulation associated with mitogen stimulation.

In our subtractive hybridization approach, putative Myc induced cDNAs were generated by subtracting mRNA isolated from ethanol treated Rat-1 MycER cells (control) from a cDNA population generated from the βestradiol activated Rat-1 MycER population. These subtracted probes were then used to screen a cDNA library generated from the βestradiol-treated Rat-1 MycER population, to enrich for transcripts induced by Myc. Positive clones from this library screen were isolated and selected for further study.
3.1.2 \textit{MiG} 1: PREVIOUS DATA

One such cDNA isolated from this experiment was named \textit{MiG} 1 (Myc-induced Gene 1). Preliminary sequence data indicated that the \textit{MiG} 1 cDNA was 1681 bp long. A search for open reading frames (ORFs) indicated the presence of a large ORF at the 5' end of the clone. However, some of the sequence data were limited to one strand or was difficult to determine from the sequencing gels with absolute confidence. A search of the GenBank database at the time of discovery using the blastn algorithm did not reveal a statistical similarity to entries in the database.

A fragment consisting of the 5' ORF region hybridized to two transcripts on a northern blot of 2.0kb and 4.0kb. The 2.0kb transcript was not induced in response to Myc activation, as determined by northern blot analysis of RNA extracted at different time points after treatment of confluent, quiescent Rat 1 MycER cells treated with βestradiol. The 4.0kb transcript was induced at 3 hours in response to Myc activation, peaked at 6 hours, then decreased slightly by 24 hours. The 2.0kb transcript was induced in response to serum stimulation of confluent, quiescent, Rat-1 fibroblasts, while the 4.0kb transcript was not. (Gihane Wasfy, M.Sc. thesis, Penn laboratory)

The ORF probe was also used to screen a multiple tissue northern blot to investigate tissue distribution and expression levels of \textit{MiG} 1 transcripts. The probe detected the presence of both transcripts in all tissues examined: heart, brain, spleen, lung, liver, skeletal muscle, kidney, and testis. Overall, skeletal muscle and heart appeared to have the highest levels of transcript, although expression levels were relatively uniform in all the other tissues. The 4.0kb transcript was more abundant in all tissues, with the exception of testis, where the 2.0kb transcript predominated. (Gihane Wasfy, M.Sc. thesis, Penn laboratory)

\textit{MiG} 1 is likely a single copy gene, and has probably been highly conserved throughout evolution. These conclusions were made based on Southern hybridization experiments of the ORF probe to genomic DNA isolated from rat, mouse, human, monkey, dog, cow, and yeast. (Gihane Wasfy, M.Sc. thesis, Penn laboratory)

3.1.3 THESIS OBJECTIVES AND RATIONALE:

\textit{MiG} 1 was cloned as part of our laboratory's efforts to identify novel genes that are induced by Myc. The preliminary data summarized in the above section suggested that \textit{MiG} 1 was a single copy, novel gene, and one of its two transcripts was induced in response to Myc activation, while the other transcript was induced in response to serum. However, all of these
observations require substantiation and further investigation. In particular, little analysis had been performed on the nucleotide or predicted amino acid sequences. The overall objective of my thesis is to further characterize the MiG 1 cDNA, building on the preliminary data that have already been collected. Specifically, three areas of investigation will be emphasized:

1) Gene structure: As already mentioned, there is less than 100% confidence in the MiG 1 cDNA sequence. This matter must be resolved before any further sequence analysis can be attempted. Second, the ORF probe hybridizes to two transcripts of 2.0kb and 4.0kb. The observation of two transcripts on a northern blot suggests that MiG 1 may have splice variants. Alternatively, the cDNA clone identified in our laboratory may only be a partial clone. The first objective of my thesis is to further investigate these issues.

2) Regulation: Preliminary data indicated that MiG 1 transcription is induced in response to Myc. However, only three time points (3hr, 6hr and 24h post-Myc activation) were examined in the initial time course experiments. The second objective of my thesis is to further characterize this induction, and confirm its existence in the improved inducible MycER™ system that lacks a weak transcriptional activation domain normally found in the ER portion of the MycER protein.

3) Sequence analysis: The observation that MiG 1 is a novel gene was based solely on a brief search of GenBank using the blastn algorithm immediately after its isolation. Furthermore, no analysis had been performed on the putative MiG 1 polypeptide. The third objective of my thesis is to conduct a thorough analysis of the nucleotide and predicted polypeptide sequences of MiG 1, utilizing recent advances in database, algorithm, and Internet technologies.

Our understanding of MiG 1 will be greatly increased by meeting these objectives. Moreover, the identification of Myc-induced genes will greatly aid our understanding of Myc’s role in diverse cellular events such as proliferation, apoptosis and tumourigenesis.
3.2 RESULTS

3.2.1 CONFIRMING THE NUCLEOTIDE SEQUENCE OF MIG I

A putative MIG I nucleotide sequence was determined soon after MIG I was cloned in our laboratory, and is shown in figure 3.1. To verify this sequence, the original sequencing gels were reread to determine which regions, if any, needed to be resequenced. The areas in which the nucleotide sequence had not been determined in both strands or was ambiguous are also shown in Figure 3.1. To resolve these discrepancies and determine the correct nucleotide sequence of the cDNA with complete confidence, the region containing the disputed sequences was resequenced. Three corrections were made to the nucleotide sequence as a result of this resequencing: a cytosine residue was added after nucleotide 936, a thymine residue at nucleotide position 941 was removed, and a cytosine residue at position 1080 was removed. The complete, confirmed sequence of MIG I is shown in figure 3.2. The cDNA is 1682bp long, and contains a large ORF that begins with an ATG codon at position 7 and ends with a TAA stop codon at position 1165. The predicted polypeptide sequence corresponding to the ORF is also shown in Figure 3.2. The effect of these changes on the predicted amino acid sequence is not extremely significant. The frameshift caused by the deletion at nucleotide 936 is reversed by the addition of a single nucleotide only five residues later, relative to the initial sequence. The frameshift caused by the removal of a nucleotide after position 1080 allows for the extension of the proposed polypeptide to its current length

One of the most notable observations from the northern blot studies described in Section 3.1.2 was that the ORF probe hybridized to two transcripts of 2.0kb and 4.0kb. Furthermore, these transcripts appear to be differentially regulated, although they appear to be equally abundant in almost every tissue observed. A potential explanation for the presence of two transcripts is that they are splice variants of a single copy gene. An alternate hypothesis is that MIG I is a member of a multi-gene family; however, previous observations suggest that this is not the case. (Gihane Wasfy, M.Sc. thesis).

Therefore, in an attempt to isolate the putative splice variants of MIG I, a cDNA library generated from a rat jejunum of a 40d old rat (Stratagene) was screened using the ORF probe. A schematic diagram demonstrating the relationship between the entire MIG I cDNA and the ORF probe is shown in Figure 3.3. Library inserts were both oligo dT and randomly primed. The average insert size was 1.0kb, and had been cloned into the EcoR1 site of the Lambda Zap® II
vector. Approximately $1 \times 10^6$ plaques were screened, and seven putative positive clones were identified. Secondary and tertiary screens confirmed their positive status. Plasmid DNA from the rescued clones was isolated, purified, and then digested by EcoR1 to estimate the clone sizes. The ends of each clone were sequenced, and these sequences were aligned with the existing MiG I nucleotide sequence, to indicate the expected size of the clone. This prediction was made assuming the absence of additional sequence in the newly identified clone. In all cases, the distance between the sequenced ends matched the predicted size from the EcoR1 restriction enzyme digest mapping, indicating that these clones represent partial clones of the originally identified MiG I cDNA. A summary of the identified clones is shown in Figure 3.4.

As Figure 3.4 demonstrates, all of the clones that were identified are partial clones of the initial MiG I clone. As a result, this screening did not identify any novel MiG I sequences that could account for the 2.0kb difference that is seen in the two transcripts.
The nucleotide sequence of *MiG* 1 based on initial sequence data. Regions where the sequence was uncertain after re-reading the sequencing gel autoradiographs are underlined in bold print. In these regions, the nucleotide sequence was either incomplete for both strands or impossible to determine with absolute confidence.
The nucleotide and deduced amino acid sequences of MiG 1.

Nucleotides that match the consensus sequence for mammalian translation initiation are capitalized, and the deduced amino acid sequence is shown beneath the corresponding nucleotide codon. The in-frame stop codon is in underlined bold type, and represented in the deduced amino acid sequence by an asterisk. The “rat identifier” nucleotide sequence is underlined.
Figure 3.3 A schematic representation of the probe fragments of *MiG* 1. The numbers indicate the nucleotide position below each arrow, with nucleotide 1 representing the 5' end of the cDNA sequence. Open reading frame sequences are represented by a box, and untranslated sequences are represented by a line. A. The entire *MiG* 1 cDNA. B. The PstI subclone fragment used as a probe in Southern blot, cDNA library screen, and northern blot analysis, referred to in the text as the "ORF probe". C. The RNAse protection probe generated from the PstI subcloned fragment by restriction enzyme digestion with DdeI.
Figure 3.4 A schematic summary of the cDNA clones identified from the screening of a 40d old rat jejumun cDNA library using the ORF probe of MiG 1.

The numbers indicate the nucleotide position below each arrow, relative to the 5' end of the MiG1 cDNA. Open reading frame sequences are represented by a box, and untranslated sequences are represented by a line. A. The entire MiG 1 cDNA. B. Schematic representations of the positive clones identified in the cDNA library screen, aligned with the original MiG1 cDNA. Positive clones were aligned with the original MiG1 cDNA by restriction digest mapping and sequencing the 5' and 3' ends of each clone.
3.2.2 DETAILED KINETIC ANALYSIS OF MiG I RNA INDUCTION BY MYC.

Previous experiments have shown that the 4.0kb MiG I transcript is induced in response to Myc activation. However, a thorough understanding of the induction kinetics did not exist, since the induction data was obtained by northern blot analysis on only a few time points that were hours apart. Determining the detailed kinetics of this induction is crucial to determining if MiG I is in fact transcriptionally regulated by Myc. An increase in MiG I transcript levels within the first few hours after Myc activation would suggest that Myc plays a key role in regulating MiG I. Therefore, to generate a more thorough picture of the induction of MiG I in response to Myc activation, a new time course was performed using our inducible system, with more time points selected for analysis. In addition to the increased number of time points in this assay, the new time course made use of improvements in the inducible Myc system.

The MycER construct has been improved during the interval subsequent to the cloning and preliminary induction kinetic studies of MiG I. The original MycER chimera contains a weak transcription activation domain (TAD) in the βestradiol-binding motif, which may contribute to transcriptional activity in the cell. To demonstrate that the ER TAD does not contribute to the induction of MiG I, but is due only to the presence of Myc, a time course was carried out in Rat 1 cells stably expressing the MycER™ (Transactivation Mutant) construct. The MycER™ construct has been engineered so that the TAD is not functional (Littlewood et al., 1995). Furthermore, activation is carried out by treatment with 4-hydroxy-Tamoxifen (4-OHT), an oestrogen receptor antagonist that cannot activate the ER TAD yet retains the derepressive effect on the Myc portion of the fusion protein.

To investigate MiG I RNA induction in the 4-OHT treated Rat 1 MycER™ cells, total RNA was isolated at the indicated points of this time course and analyzed by an RNase protection assay. The RNase protection assay was selected due to its greater sensitivity than northern blot analysis. Smaller amounts of RNA are required for an RNase protection assay, relative to the considerable amount of RNA required to generate polyA+ RNA required for MiG I northern blot analysis. The MiG I RNase protection probe was derived from the PstI ORF probe by restriction enzyme digestion with Ddel, and is schematically represented in Figure 3.3. The RNase protection probe was first tested by northern blot analysis to confirm probe specificity. The antisense probe readily detected both the 2.0kb and 4.0kb transcripts by northern blot analysis (data not shown). This compromises the probe’s ability to accurately determine the degree of induction. However, due to the increased number of time points in the induction
kinetics experiment and the large amount of tissue culture required, and the lower sensitivity of northern blot analysis, the RNase protection assay was still the better experimental method to use, from both the workflow and scientific perspectives.

Increases in the level of protected MiG I transcript fragments were visible as early as 1 hour after 4-OHT treatment (Figure 3.5, lane 3). As a control for equal lane loading, GAPDH levels were also examined. The induction of MiG I continued at the 2h and 3h time points, with MiG I RNA levels remaining at elevated levels until the end of the time course at the 18h and 24h time points. The induction observed represents a relatively weak induction, estimated not to exceed a 2-3-fold increase in both transcript levels.
Figure 3.5 MiG 1 RNA levels are induced in response to Myc activation.

RNase protection assay performed on 10μg of total cellular RNA harvested from confluent, quiescent Rat-1 MycER™ cells that were untreated (0hr timepoint), or after treatment with 4-OHT (indicated in hours above each lane). Probes for MiG 1 and gapdh were analysed concurrently.
3.2.3 *MiG I* RNA LEVELS ARE ELEVATED IN CELLS CONSTITUTIVELY EXPRESSING HIGH LEVELS OF MYC.

An overabundance of Myc is a frequent hallmark of cancer cells, and several putative Myc target genes are overexpressed in many types of cancers (Cole and McMahon, 1999). Since it is likely that the downstream effectors of Myc are overexpressed when Myc is overexpressed, I evaluated *MiG I* RNA levels in several cell lines constitutively expressing high levels of ectopic Myc.

Total RNA was extracted from cells infected with either a control retrovirus or a retrovirus containing *v-myc* sequences and assayed for *MiG I* RNA levels by northern blot analysis. The 1.0kb ORF probe described in Section 3.1.2 and shown in Figure 3.3B was used in these studies. In both the primary Rat Embryo Fibroblasts (REF) and non-transformed KP REF cell line, levels of both transcripts were significantly elevated in *v-myc* infected cells, relative to cells infected with the control virus (Figure 3.6, lanes 1-4). Interestingly, no increase was seen in the Rat-1 cells infected with *v-myc*. (Figure 3.6, lanes 5,6) Overall, however, the constitutive presence of high levels of Myc correlated with an increase in *MiG I* RNA levels.
Figure 3.6 MiG I RNA levels are elevated in cells constitutively expressing exogenous Myc protein.

RNA was extracted from asynchronous, proliferating subconfluent cells of the indicated cell lines (noted above each lane), that had been infected with a control retrovirus (-), or the same retrovirus containing v-myc sequences. The RNA (20µg) was then assayed by northern blot analysis for MiG I transcript levels using the 1kb ORF probe described in section 3.1.2.
3.2.4 SEQUENCE ANALYSIS: \textit{MiG} 1 IS THE RAT HOMOLOG OF \textit{ALK} B

Once the nucleotide sequence of \textit{MiG} 1 had been confirmed, our next step was to
thoroughly analyse the nucleotide and predicted amino acid sequences to determine if \textit{MiG} 1
represents a novel gene or is similar to previously identified genes. \textit{MiG} 1 contains a large ORF,
as previously mentioned, and the predicted polypeptide sequence is shown in Figure 3.2. The
ORF is 1158bp long, beginning with an ATG codon at nucleotide position seven. The sequence
surrounding this ATG codon is a strong match to the consensus sequence for mammalian
translation initiation, particularly at the crucial –3 and +4 residues relative to the initiation ATG.
This indicates that this ATG is an acceptable site for translation initiation, and might encode the
start methionine for the polypeptide(Kozak, 1987).

Our first search of the GenBank nr (non-redundant) database using the predicted
polypeptide sequence and the blastp and tblastx algorithms revealed a weak, yet statistically
significant similarity to AlkB, an \textit{E.coli} protein that is part of the adaptive response to DNA
alkylating agents(Altschul \textit{et al.}, 1997). (Kataoka and Sekiguchi, 1985; Kondo \textit{et al.}, 1986) The
region of similarity identified by blast was restricted to the carboxyl halves of the polypeptides.

Subsequent searches revealed similarities between \textit{MiG} 1 and other putative AlkB
homologs, as new entries were added to GenBank. There was an extremely high degree of
similarity (86% identity and 93% similarity over a 251 amino acid stretch) between the \textit{MiG} 1
predicted polypeptide and hABH (human \textit{alkB} homolog). (Wei \textit{et al.}, 1996) (Figure 3.7,
discussed in greater detail in Section 3.2.12) The "E value" for this match is extremely low (e-
131 or 1 x10^{-131}), and is indicative of a very high degree of sequence conservation between hABH
and \textit{MiG} 1. This match is extremely unlikely to have occurred by chance. The Expect value (E
value) is a statistical prediction for the randomness of a particular match between query and
subject sequences. In general, as the E value score approaches 0, the extent of sequence similarity
increases. As an E value score approaches 0, the likelihood of an evolutionary relationship
between a query sequence and a database entry increases. The E value for the match between
hABH and \textit{MiG} 1 indicates that 1 out of 1x10^{131} entries would match between \textit{MiG} 1 to the same
extent purely by chance. The similarity between \textit{MiG} 1 and the putative AlkB homologs in
\textit{A.thaliana} and \textit{S.pombe} was weaker than the similarity to hABH, but still possessed a very low E
value, indicating that these too are unlikely to have occurred by chance, and suggests that \textit{MiG} 1
shares an evolutionary history with these genes. In both cases the similarity was extensive,
particularly in the carboxyl half of the proteins. Lastly, the similarity between \textit{MiG} 1 and an AlkB
homolog in the Gram negative bacteria *Caulobacter crescentus* was very similar in extent and location to that observed in *E.coli*. (Colombi and Gomes, 1997; Kondo et al., 1986)

The initial search of the GenBank nr (non-redundant) nucleotide database using the blastn and tblastn algorithms performed immediately after the isolation of *MiG 1* did not reveal any matches of statistical significance. However, *MiG 1* was found to contain an 82bp “rat-specific transcript identifier sequence” repetitive element upstream of the polyA tail. (Figure 3.2) Subsequent blast searches confirmed the high degree of similarity in nucleotide sequence that was observed for the hABH protein. In fact, the E value for the hABH nucleotide alignment was 0.0, indicating that such an extensive match could not have statistically occurred by chance. The similarity between hABH and MiG 1 is examined in greater detail in section 3.2.12

The extremely high degree of conservation between *MiG 1* and hABH in both the nucleotide and amino acid sequences, in addition to the high similarity scores obtained for other *alkB* homologs strongly suggests that *MiG 1* is a member of the *alkB* family. Although no direct functional assay exists for AlkB, alkylation agent sensitivity experiments indicate that AlkB function has been conserved throughout evolution.(Chen et al., 1994; Colombi and Gomes, 1997; Wei et al., 1996) Therefore, *MiG 1* was renamed rABH (rat *alkB* homolog) at this point in the investigation, and will be referred to as rABH from hereon.
>ref|NP_006011.1| alkB homolog
pir|S64736 AlkB homolog protein ABH - human
emb|CAA63047.1 (X91992) alkB homolog protein [Homo sapiens]
prf|2208455A ABH gene [Homo sapiens]

Length = 299

Score = 468 bits (1192), Expect = e-131
Identities = 217/251 (86%), Positives = 236/251 (93%)

Figure 3.7 Results of a blastp algorithm search of the GenBank database, using the predicted MiG 1 polypeptide sequence as a query.

The predicted MiG 1 sequence is shown in the query line, and the hABH sequence is shown in the subject line. The numbers on either side of the sequence indicate amino acid position within each polypeptide. The middle line of sequence shows the amino acid, when the subject and query are identical, and a + sign when the two amino acids are similar in nature. Gaps represent residues where the two sequences differ. The alignment does not include residues beyond the regions of similarity that were identified by the blastp algorithm. Also shown above the alignment is the reference data for the GenBank entry that matched the query sequence, and statistical information regarding the number of residues that matched, and a numerical score for the match. Lastly, a statistical prediction for the randomness of this match is given, referred to as the Expect value. For this alignment, one out of 1x 10^{13} entries would be expected to match MiG 1 to the same extent as this entry.
3.2.5 ALKB AND THE ADAPTIVE RESPONSE OF E.coli

The adaptive response to DNA alkylation results in an increased resistance to the killing and mutagenic effects of alkylating agents such as N-methyl N'-nitro-N-nitrosoguanidine (MNNG), methylmethanesulfonate (MMS), and dimethyl sulfate (DMS). This resistance is characterized by a cell’s ability to survive an otherwise lethal exposure to an alkylating agent, following low-level exposure with the same agent (Samson and Cairns, 1977; Lindahl et al., 1988). In E. coli, the adaptive response is initiated by the product of the ada gene, a methyltransferase that directly repairs O6-methylguanine and O4-methylthymine residues by transferring the methyl group to its Cys-321 residue (Shevell et al., 1990; Margison et al., 1985). Ada is a “suicide enzyme”, because the active site cannot be regenerated after the Cys-321 residue accepts the methyl group from the alkylated base (Lindahl et al., 1982). Ada also becomes a transcriptional activator as a result of its ability to repair the S, stereoisomer of methylphosphotriesters by accepting a methyl group at its Cys-69 residue (Weinfeld et al., 1985). The Cys-69 alkylated Ada protein binds to a regulatory sequence in the promoters of the alkA and aidB genes, as well as the operon containing itself and the alkB gene, resulting in increased transcription (Hamblin and Potter, 1985; McCarthy and Lindahl, 1985). (Teo et al., 1986)

The alkA gene encodes a DNA glycosylase whose primary alkyl-nucleotide substrate is N3-methyladenine. N3-methyladenine is a lethal lesion that blocks DNA replication (Evensen and Seeberg, 1982; Karran and Marinus, 1982; Nakabeppu et al., 1984). AlkB catalyzes the first reaction in the base excision repair pathway, whereby the alkylated base is removed from the DNA strand and replaced (Freidberg et al., 1995).

AlkB was first identified as an E. coli mutant that showed substantially increased sensitivity to MMS, but not MNNG (Kataoka et al., 1983). The gene was subsequently cloned and shown to be transcribed as part of a single operon with the ada gene; the last adenine in the TAA stop codon of ada is the first nucleotide of the alkB ATG initiation codon (Kataoka and Sekiguchi, 1985; Kondo et al., 1986). The 24kDa E.coli AlkB protein has been isolated, but a function has not yet been assigned to the protein; it does not possess any observable DNA methyltransferase or DNA glycosylase activity (Kataoka and Sekiguchi, 1985; Kondo et al., 1986). An attempt to identify defects in the ability of alkB mutants to repair MMS induced lesions did not reveal any deficiencies in repairing the types of lesions examined (Dinglay et al., 1998). In spite of this, the ability of E. coli AlkB to protect human cells from MMS-associated cytotoxicity, along with the observation that MMS-treated phage λ survive better in alkB wild
type cells than in alkB mutants, suggest that AlkB possesses an intrinsic activity that helps a cell process DNA that has been alkylated by agents that react through an SN2 mechanism (Chen et al., 1994; Kataoka et al., 1983).

3.2.6 PREDICTING THE STRUCTURE OF rABH USING COMPUTER MODELS

Although over fifteen years have transpired since the AlkB mutant was discovered and characterized, our understanding of how it functions remains a mystery. This is in spite of the fact that several efforts have been made to experimentally identify the specific DNA lesions that are repaired by the protein (Dinglay et al., 1998; Kataoka and Sekiguchi, 1985).

I have attempted to gather as much information as possible on the predicted rABH protein through sequence analysis, in an effort to help determine AlkB’s function. A more detailed analysis of the predicted rABH protein might provide essential clues as to which regions of the protein are essential to function, and lead to a more complete understanding of AlkB function.

The 1158bp ORF translates into a 386 amino acid polypeptide with a predicted Mw of 43.6 kDa, and theoretical pI of 8.29, as calculated by algorithms on the Expasy website http://www.expasy.ch/tools/pi_tool-ref.html. A hydrophobicity plot for the predicted polypeptide was generated using the Kyte-Doolittle method at the Expasy website http://www.expasy.ch/cgi-bin/protscale.pl, and is shown in Figure 3.8. Of particular note is a very hydrophilic region centered approximately at residue 155. The polypeptide continues to be hydrophilic for approximately forty residues on either side of residue 155, with the amino side of residue 155 showing greatest hydrophilicity. The very hydrophilic nature of the sequences near residue 155 suggests that this region is likely exposed and may therefore be functionally important. Also of note is that the N-terminal 25 amino acids are very hydrophobic. This may be important in targeting the protein to a specific subcellular location, especially if the protein is membrane-associated.

Predicting a protein’s secondary structure can help identify functional motifs that have been conserved across broad evolutionary lines. To develop a model for rABH secondary structure, the MLRC algorithm was selected to perform the analysis (Guermeur et al., 1999). As a control, human Myc was first analyzed using these algorithms. (Figure 3.9) For the control Myc sequences, the MLRC algorithm identified the helical domains in Myc’s NLS, BR and LZ regions very clearly. Interestingly, the structure of the HLH region was not clearly predicted,
A hydrophobicity plot was generated for the putative rABH polypeptide using the Kyte-Doolittle method. The amino acid position is shown along the x-axis, and the relative hydrophobicity of each residue is shown along the y-axis. The data was generated using the Kyte-Doolittle scale. The numerical scale of the y-axis represents descending hydrophobicity. Hydrophobic amino acids have a high positive value, and hydrophilic residues have a low negative number.
Figure 3.9 The MLRC algorithm can predict known regions of the secondary structure of the cMyc protein.

The secondary structure predictions of the human c-Myc protein made using the MLRC algorithm are shown in text and graphic formats. A shows each amino acid with its predicted secondary structure indicated below. The letter indicating the type of secondary structure is defined in the Legend. B is a graphical representation of the algorithm’s results, showing the predicted structure for each residue. A predictive score was generated for each residue, indicating the likelihood of that particular residue being part of a helix, sheet, or coil. The prediction for a particular structure is highest when an amino acid has a score close to one for one of the structures and scores close to zero for the other two structures. Residue position is plotted on the x-axis, and probability is plotted along the y-axis. Separate lines for helix (black), β-sheet (grey) and coil (light grey) have been plotted. Shown above the graph is a summary of the most likely predicted secondary structure.
Figure 3.10 Predicting the secondary structure of rABH.

The secondary structure predictions of the rABH made using the MLRC algorithm are shown in text and graphic formats. A shows each amino acid with its predicted secondary structure indicated below. The letter indicating the type of secondary structure is defined in the Legend. B is a graphical representation of the algorithm’s results, showing the predicted structure for each residue. A predictive score was generated for each residue, indicating the likelihood of that particular residue being part of a helix, sheet, or coil. The prediction for a particular structure is highest when an amino acid has a score close to one for one of the structures and scores close to zero for the other two structures. Residue position is plotted on the x-axis, and probability is plotted along the y-axis. Separate lines for helix (black), βsheet (grey) and coil (light grey) have been plotted. Shown above the graph is a summary of the most likely predicted secondary structure.
perhaps indicating a weakness in the algorithm's ability to accurately predict all secondary structures. An alternate explanation lies in the algorithm. The sandwich-like nature of the HLH may confound the algorithm's ability to accurately isolate each element. The algorithm makes a continuous profile, and is unable to make sharp boundaries between regions. The presence of the loop component of the HLH may have blurred the ability of the algorithm to detect the presence of the helix components on either side, and vice versa.

The pattern that was predicted for the secondary structure of rABH is shown in Figure 3.10. Overall, the amino half contains more helical regions than the carboxyl half of the protein. These helical regions are clustered in the first fifty amino acids of rABH and several large helical regions spaced somewhat evenly between amino acids 100 and 200. The vast majority of the carboxyl half of the protein is predicted to be in a coiled format, with the exception of one large helical region extending from amino acid 325-350, and several other shorter helical and extended regions. Almost all of the protein's predicted extended or beta sheets were clustered between residues 200 and 275, and are surrounded by coiled regions.

3.2.7 PREDICTING THE STABILITY OF rABH

Although predicting secondary structure is a valuable tool in the evaluation of a novel protein, there are other predictive algorithms that look for different sequence characteristics related to other characteristics of the protein, such as stability. Among these is an algorithm that looks for PEST sequences. PEST sequences are regions that are rich in proline, glutamic acid, serine, and threonine, generally flanked by clusters of positively charged amino acids. Proteins that are rich in PEST sequences generally have a shorter half-life (Rogers et al., 1986). A search for PEST regions in rABH using the on-line server http://www.icnet.uk/LRITu/projects/pest/index.html did not identify any PEST sequences. When the short-lived Myc protein sequence was similarly analyzed as a control, no PEST sequences were identified.

The absence of PEST sequences does not mean that a protein is necessarily stable. Therefore, a more informative algorithm was selected that provides an "instability index" based on the relative abundance of certain dipeptides that are present at differing rates in stable vs. unstable proteins (Guruprasad et al., 1990). A protein whose instability index is smaller than 40 is predicted to be stable, while a value above 40 predicts that the protein may be unstable. Myc, an unstable protein with an estimated half-life of 30 minutes, had a score of 92.23 when the
sequence was entered into the algorithm at the on-line server www.expasy.ch/tools/protparam.html. rABH had a score of 56.94, indicating that it may be unstable, but is less likely than Myc to be unstable.

3.2.8 PREDICTING THE SUBCELLULAR LOCALIZATION OF rABH

Determining where a protein is localized often provides an indication as to its function. For example, proteins that regulate transcription by binding to DNA must be in the nucleus. Growth factor receptors must have an extracellular portion that binds to the ligand, as well as transmembrane and intracellular regions necessary to transduce that growth signal to the cell's intracellular signalling pathways. Since AlkB is presumed to perform its function in the nucleus (Kataoka et al., 1983), I set out to determine the most likely subcellular location of rABH.

To meet this objective, I utilised the PSORT II web server located at http://psort.nibb.ac.jp:8800/. The PSORT II program predicts the subcellular localization of proteins based on their amino acid sequences. It runs numerous algorithms on a protein sequence, each one searching for different identifying sequences that are associated with localization to various regions of the cell. It then combines all of the results of the individual algorithms, taking into consideration the limitations of each algorithm, such as database size, that reduces the overall accuracy of that algorithm. These weighted values are then used to generate a single prediction for the localization of the protein (Reinhardt and Hubbard, 1998). The results of the PSORT II search for rABH and Myc are summarized in Table 3.1.

As expected, PSORT II predicted that Myc is a nuclear protein, and it predicted Myc as being nuclear with a very high degree of confidence. Perhaps surprisingly, rABH was predicted to be cytoplasmic, although not with the same high degree of confidence that Myc was predicted to be nuclear. There are two other important observations about the results of the PSORT II search: rABH possesses four nuclear localization signal (NLS) motifs and rABH did not contain any of the selected 63 DNA binding motifs contained in the PROSITE database.
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Targeting sequence / Subcellular location</th>
<th>Result for Myc</th>
<th>Result for RABH</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSG, GvH</td>
<td>N-terminal signal peptide / extracellular</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
| NUCDISC  | Nuclear Localization Sequence (NLS) / nucleus | Pat 7: PAAKRVK at 320  
Pat 4: RKRR at 421 | Pat 4: KRPR at 156  
Pat 4: KRKR at 378  
Pat 7: PVKRKRL at 376 |
| KDEL      | ER retention motif in the C-terminus | None           | None            |
| SKL       | Peroxisomal targeting signals | None           | None            |
| SKL 2     | Second peroxisomal targeting sequence | None           | None            |
| VAC       | Possible vacuolar targeting motif | None           | None            |
| RNA-binding motif | None | None |
| Actin-binding motif | Type 1: None  
Type 2: None | Type 1: None  
Type 2: None |
| NMYR      | N-myristoylation pattern | None           | None            |
| memYQRL   | Transport motif from cell surface to Golgi | None           | None            |
| Tyrosine residues in tail | None | None |
| Dileucine motif in tail | None | None |
| PROSITE   | Search of 63 Prosite DNA binding motifs | 1) Leucine zipper pattern (PS00029) at 413  
2) Myc-type 'helix loop helix' dimerization domain signature (PS00038) at 391 | None |
| PROSITE   | Ribosomal protein motifs | None           | None            |
| NNCN      | Cytoplasmic/nuclear discrimination | Prediction: nuclear  
Reliability: 94.1 | Cytoplasmic  
Reliability: 70.6 |
| COIL      | Coil-coil regions | 37 residues: 402-438 | None |
| k-NN      | Combined prediction from all algorithms | 82.6%: nuclear  
8.7%: cytoplasmic  
4.3%: vacuolar  
4.3%: plasma membrane | 56.5%: cytoplasmic  
26.1%: nuclear  
4.3%: extracellular  
4.3%: Golgi  
4.3%: secretory vesicles  
4.3%: endoplasmic reticulum |

Table 3.1 Summary of PSORT II prediction of subcellular localization for Myc and rABH.
3.2.9 MOTIF SEARCHES

The PSORT II algorithm includes a partial search of the PROSITE database; however, it only searches 63 PROSITE entries related to protein-DNA interactions. Looking for sequence motifs that have been conserved throughout evolution provides yet another useful proteomics tool that can be used to help classify a novel protein and ascertain its function. To determine if rABH contains any previously identified motifs, I compared it to various motif databases using three different algorithms.

First, Myc was compared against the entire PROSITE database using the server at http://www.expasy.ch/tools/scnpsite.html. and the results of this search are summarized in Table 3.2. The ability of the program to identify the HLH and LZ regions of Myc serve as a good control for the program’s performance. However, PROSITE failed to identify seven residues in Myc that are phosphorylated, including both Thr\textsuperscript{58} and Ser\textsuperscript{62} that have been shown to be phosphorylated \textit{in vivo} (Luscher \textit{et al.}, 1989; Lutterbach and Hann, 1994; Lutterbach and Hann, 1997). Thr\textsuperscript{58} has been shown to be phosphorylated by GSK and Ser\textsuperscript{62} has been shown to be phosphorylated by both MAPK and p34cdc2 (Alvarez \textit{et al.}, 1991; Henriksson \textit{et al.}, 1993; Pulverer \textit{et al.}, 1994; Seth \textit{et al.}, 1992). The inability of PROSITE to detect these phosphorylation sites casts serious doubt on the validity of its predictions regarding potential phosphorylation. Despite this apparent limitation, rABH was analyzed by the PROSITE algorithm, and the results of a search using rABH as the query are shown in Table 3.3. As indicated in Table 3.3, the search failed to identify the presence of any previously identified PROSITE motifs in rABH, other than post-translational modification sites. Primarily, sites for Ser/Thr kinases were identified that have short, weak consensus sequences and are therefore likely to occur simply by chance. However, it is still important to recognize the presence of these sites in rABH; they may be utilized \textit{in vivo} and may be important for function.
<table>
<thead>
<tr>
<th><strong>Prosite Pattern</strong></th>
<th><strong>Position</strong></th>
<th><strong>Sequence</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>N-glycosylation</td>
<td>4-7</td>
<td>NVSF</td>
</tr>
<tr>
<td></td>
<td>112-115</td>
<td>NQSF</td>
</tr>
<tr>
<td></td>
<td>200-203</td>
<td>NDSS</td>
</tr>
<tr>
<td>CAMP- and cGMP-dependent protein kinase phosphorylation site</td>
<td>156-159</td>
<td>RKDS</td>
</tr>
<tr>
<td></td>
<td>340-343</td>
<td>RKCT</td>
</tr>
<tr>
<td></td>
<td>355-358</td>
<td>KRRT</td>
</tr>
<tr>
<td></td>
<td>397-400</td>
<td>KKAT</td>
</tr>
<tr>
<td>Protein kinase C phosphorylation site</td>
<td>8-10</td>
<td>TNR</td>
</tr>
<tr>
<td></td>
<td>64-66</td>
<td>SRR</td>
</tr>
<tr>
<td></td>
<td>81-83</td>
<td>SLR</td>
</tr>
<tr>
<td></td>
<td>146-148</td>
<td>SEK</td>
</tr>
<tr>
<td></td>
<td>204-206</td>
<td>TPK</td>
</tr>
<tr>
<td></td>
<td>314-316</td>
<td>STR</td>
</tr>
<tr>
<td></td>
<td>315-317</td>
<td>TRK</td>
</tr>
<tr>
<td></td>
<td>329-331</td>
<td>SVR</td>
</tr>
<tr>
<td></td>
<td>344-346</td>
<td>SPR</td>
</tr>
<tr>
<td>Casein kinase II phosphorylation site</td>
<td>94-97</td>
<td>STAD</td>
</tr>
<tr>
<td></td>
<td>225-228</td>
<td>SSTE</td>
</tr>
<tr>
<td></td>
<td>248-251</td>
<td>TSSD</td>
</tr>
<tr>
<td></td>
<td>250-253</td>
<td>SDSE</td>
</tr>
<tr>
<td></td>
<td>252-255</td>
<td>SEE</td>
</tr>
<tr>
<td></td>
<td>315-318</td>
<td>TRKD</td>
</tr>
<tr>
<td></td>
<td>348-351</td>
<td>SDTE</td>
</tr>
<tr>
<td></td>
<td>415-418</td>
<td>SEED</td>
</tr>
<tr>
<td>Trosine kinase phosphorylation site</td>
<td>10-16</td>
<td>RNYDDLKY</td>
</tr>
<tr>
<td>N-myristolation site</td>
<td>88-93</td>
<td>GGGGSF</td>
</tr>
<tr>
<td></td>
<td>90-95</td>
<td>GGSFST</td>
</tr>
<tr>
<td></td>
<td>91-96</td>
<td>GSFSTA</td>
</tr>
<tr>
<td></td>
<td>280-285</td>
<td>GSPSAG</td>
</tr>
<tr>
<td>Amidation site</td>
<td>273-276</td>
<td>PGKR</td>
</tr>
<tr>
<td>Cell attachment sequence</td>
<td>83-85</td>
<td>RGD</td>
</tr>
<tr>
<td>Leucine Zipper</td>
<td>413-434</td>
<td>LISEEDLLRKRREQLKHKLKHKLEQL</td>
</tr>
<tr>
<td>Myc-type, helix loop helix dimerization domain signature</td>
<td>391-406</td>
<td>PKVVILKKKATAYILSV</td>
</tr>
</tbody>
</table>

Table 3.2. Results of a search of the Prosite database using human cMyc as a query.

The Prosite pattern identified in the search is listed in the left-hand column. The amino acid positions within Myc where the match to the Prosite pattern occurs is shown in the middle column, and the amino acid sequence is shown in the right-hand column.
<table>
<thead>
<tr>
<th>Prosite Pattern</th>
<th>Position</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-glycosylation</td>
<td>65-68</td>
<td>NVSS</td>
</tr>
<tr>
<td></td>
<td>337-340</td>
<td>NMTV</td>
</tr>
<tr>
<td>cAMP- and cGMP-dependent protein kinase phosphorylation site</td>
<td>31-34</td>
<td>RRGT</td>
</tr>
<tr>
<td></td>
<td>179-182</td>
<td>KKYS</td>
</tr>
<tr>
<td>Protein kinase C phosphorylation site</td>
<td>30-32</td>
<td>SRR</td>
</tr>
<tr>
<td></td>
<td>51-53</td>
<td>SPK</td>
</tr>
<tr>
<td></td>
<td>74-76</td>
<td>TSR</td>
</tr>
<tr>
<td></td>
<td>120-122</td>
<td>SQK</td>
</tr>
<tr>
<td></td>
<td>154-156</td>
<td>TKR</td>
</tr>
<tr>
<td></td>
<td>178-180</td>
<td>SKK</td>
</tr>
<tr>
<td></td>
<td>333-335</td>
<td>TAR</td>
</tr>
<tr>
<td></td>
<td>339-341</td>
<td>TVR</td>
</tr>
<tr>
<td>Casein kinase II phosphorylation site</td>
<td>68-71</td>
<td>SVTE</td>
</tr>
<tr>
<td></td>
<td>70-73</td>
<td>TEHD</td>
</tr>
<tr>
<td></td>
<td>133-136</td>
<td>TEEE</td>
</tr>
<tr>
<td></td>
<td>160-163</td>
<td>SLLLE</td>
</tr>
<tr>
<td></td>
<td>232-235</td>
<td>SELD</td>
</tr>
<tr>
<td></td>
<td>314-317</td>
<td>SLLLE</td>
</tr>
<tr>
<td></td>
<td>320-323</td>
<td>SVED</td>
</tr>
<tr>
<td></td>
<td>346-349</td>
<td>TGQD</td>
</tr>
</tbody>
</table>

Table 3.3. Results of a search of the Prosite database using rABH as a query.

The Prosite pattern identified in the search is listed in the left-hand column. The amino acid positions within rABH where the match to the Prosite pattern occurs is shown in the middle column, and the amino acid sequence is shown in the right-hand column.
The second system that was applied to rABH was GeneFIND. GeneFIND is an integrated database search system that combines results from BLAST and SSEARCH algorithm searches of PROSITE, PIR Superfamily, ProClass, and other “unclassified” entries. The potential to identify existing motifs is therefore greater, since numerous databases are screened (Wu et al., 1998). The GeneFIND output format identifies the degree of similarity that a query has to the identified database entry, and indicates the probability of the identification being a true, false, or equivocal positive. To perform the analysis, the GeneFIND server at Georgetown University, http://pir.georgetown.edu/gfcgi/genefind.pl, was used. Once again, Myc was analyzed as a positive control. The results of this analysis are shown in Table 3.4. The Myc control sequence was identified as being 100% similar to the identified human Myc protein by both Blast and Ssearch, and the GeneFIND algorithm identified this as a true positive. Furthermore, GeneBLAST runs a multiple alignment using ClustalW to confirm that the putative family member possesses the necessary consensus sequence. The algorithm used the HLH domain from residues 391-406 as the necessary signature, with 100% identity throughout the required 16 residues.

GeneFIND was unable to classify rABH as a PROSITE entry. An equivocal score was generated for two different “EGF-1 like domain signature 1”motif containing proteins by Blast and Ssearch; however, rABH failed the multiple alignment test for proteins that contain “EGF-1 like domain signature 1”motif. It did not contain the necessary conserved residues to be assigned to this group. This is reflected in the motif score of 0.0 that rABH received. (Table 3.4) Despite GeneFIND’s inability to classify rABH in PROSITE, it was able to identify that rABH is very likely to be an alkB homolog. GeneFIND identified both the E.coli and human alkB homologs by the Blast and Ssearch algorithms. In conclusion, GeneFIND did not provide any substantial indication that rABH contains a previously identified motif, but confirmed its status as a member of the alkB family, based on overall sequence similarity.

The last computer model that was used to identify motifs in rABH was the IDENTIFY system developed by Stanford University, with an internet server at http://dna.stanford.edu/cgi-bin/emotif/nph-identity. This system searches the BLOCKS and PRINTS databases, which differ somewhat from other databases. The entries of the BLOCKS database are generated by an algorithm that looks for the most highly conserved regions in groups of proteins in the PROSITE database. It is important to distinguish this information from the PROSITE pattern for a group of proteins. These blocks are then calibrated against the SWISS-PROT database to obtain an indication of the chance distribution of the blocks that have been generated. The blocks that pass through the filter and are unlikely to have occurred by chance form the BLOCKS database
entries. The BLOCKS database contains 4034 different blocks generated from 994 PROSITE groups. The PRINTS database contains protein fingerprints; a protein fingerprint is defined as a group of conserved motifs used to characterize a protein family. BLOCKS and PRINTS are frequently run together as an adjunct to PROSITE. The greater number of entries in BLOCKS and PRINTS is ideal when attempting to classify a protein such as rABH, which has not matched an assigned PROSITE entry. As with the other searches in this series, Myc was used as a control to examine the motifs recognized by the search engine. The results of the IDENTIFY algorithm searches are shown in Table 3.5. Only the highest order of stringency that identified results is shown.

The IDENTIFY algorithm reports only sequences that have a 100% match to the consensus sequence, and the database is much larger than the databases that have been examined previously. The IDENTIFY algorithm was the only algorithm used that identified all of the key motifs in Myc. The IDENTIFY algorithm identified eight sequence signatures in Myc at a level of stringency for which no false positive results were expected. The algorithm correctly detected the presence of Myc-identifying sequences in the HLH, BR, and the Mbol and MblI regions of the human cMyc protein.

The results for rABH were drastically different from those obtained for Myc. There were no sequence similarities for rABH at a stringency where zero false positive results were expected. Three positive matches occurred at stringency (1 in 10^5) where three false positive results were expected. Of those three matches, matches to P450 and Kelch repeat signature are potentially interesting because they occur in regions of rABH that have been highly conserved throughout evolution (Figure 3.14), and discussed in Section 3.2.13.
A

<table>
<thead>
<tr>
<th>Database</th>
<th>Global</th>
<th>Motif</th>
<th>Blast ID</th>
<th>SSearchID</th>
<th>Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosite</td>
<td>10.0</td>
<td>10.0</td>
<td>Myc-human</td>
<td>Myc-human</td>
<td></td>
</tr>
<tr>
<td>PIR Superfamilies</td>
<td>10.0</td>
<td>N/A</td>
<td>Myc transforming protein</td>
<td>Myc transforming protein</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>10.0</td>
<td>N/A</td>
<td>Avian myelocytomatosis virus</td>
<td>Avian myelocytomatosis virus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.5</td>
<td>N/A</td>
<td>Transforming protein B-myc</td>
<td>Transforming protein B-myc</td>
<td></td>
</tr>
<tr>
<td>Unclassified</td>
<td>4.5</td>
<td>N/A</td>
<td>Myc (<em>Asterias vulgaris</em>)</td>
<td>Myc (<em>Asterias vulgaris</em>)</td>
<td>N/A</td>
</tr>
<tr>
<td>Entries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Database</th>
<th>Global</th>
<th>Motif</th>
<th>Blast ID</th>
<th>SSearchID</th>
<th>Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosite</td>
<td>1.5</td>
<td>0.0</td>
<td>LMA1-mouse (EGF-1)</td>
<td>MUCL-rat (mucin-like intestinal protein)</td>
<td></td>
</tr>
<tr>
<td>PIR Superfamilies</td>
<td>2.5</td>
<td>N/A</td>
<td>AlkB (<em>E.coli</em>)</td>
<td>AlkB (<em>E.coli</em>)</td>
<td>N/A</td>
</tr>
<tr>
<td>Unclassified</td>
<td>9.0</td>
<td>N/A</td>
<td>HABH (Human alkB homolog)</td>
<td>HABH (Human alkB homolog)</td>
<td>N/A</td>
</tr>
<tr>
<td>Entries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4 Results of the GeneFIND database search system.

Myc was first analyzed using the algorithm as a control, and the results are shown in Table 3.4A. Results of the rABH search are presented in Table 3.4B. The global and motif scores range from 0-10, with 10 indicating a perfect match, and 0 indicating no match between the query and indicated database entry. Motif scores are reported only for the PROSITE database. The scores are a numerical calculation based on the degree of similarity between the query and entry from the blast and Ssearch algorithms. The names of the database entries uncovered by each algorithm are indicated in the next two columns. The last column, labeled flag, provides an indicator as to whether the match to the PROSITE entry contains no mismatches (T) contains mismatches to PROSITE (N) or is an uncertain match to a PROSITE pattern (?). A (?) indicates that the algorithm did not detect a match to the PROSITE entry listed, and is therefore likely to be a false positive result.
A

<table>
<thead>
<tr>
<th>Stringency FP</th>
<th>Description</th>
<th>Motif</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 in $10^{10}$</td>
<td>0 Myc-type HLH dimerization domain</td>
<td>er[kqr]r..[ilv].[as]f.[ast][kqr]</td>
<td>362-378</td>
</tr>
<tr>
<td></td>
<td>Myc-type HLH dimerization domain</td>
<td>[ekq][kr].[k]iv..[kr].a.y[iv].[ilv]</td>
<td>387-406</td>
</tr>
<tr>
<td></td>
<td>Myc proto-oncogene signature</td>
<td>p.ediwwkKel[ilv]p</td>
<td>44-57</td>
</tr>
<tr>
<td></td>
<td>Myc proto-oncogene signature</td>
<td>[ilv][ilv].[kqr]dcmws.fis[ast]</td>
<td>127-140</td>
</tr>
<tr>
<td></td>
<td>Myc proto-oncogene signature</td>
<td>bqhnyaaa.p</td>
<td>304-313</td>
</tr>
<tr>
<td></td>
<td>Myc proto-oncogene signature</td>
<td>d[st]e..[kr][kr][kr].h.[filvy]le</td>
<td>348-363</td>
</tr>
<tr>
<td></td>
<td>Myc proto-oncogene signature</td>
<td>r[ekq]r.n[de].[kr].f.[ast]lr[de]</td>
<td>348-363</td>
</tr>
<tr>
<td></td>
<td>Myc proto-oncogene signature</td>
<td>ka.kv.il.[kr].a..y</td>
<td>388-402</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Stringency FP</th>
<th>Description</th>
<th>Motif</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 in $10^{10}$</td>
<td>0 None</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1 in $10^{9}$</td>
<td>0 None</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1 in $10^{8}$</td>
<td>0 None</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1 in $10^{7}$</td>
<td>0 None</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1 in $10^{6}$</td>
<td>0 None</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1 in $10^{5}$</td>
<td>3 Arthropod defensins proteins</td>
<td>c. hc y..[fly]pg[filmv]</td>
<td>397-302</td>
</tr>
<tr>
<td></td>
<td>P450 Superfamily signature</td>
<td>[filmv].e.[fly]r.[st][filmv]</td>
<td>88-97</td>
</tr>
<tr>
<td></td>
<td>Kelch repeat signature</td>
<td>[eq].[ilv][filvy].[filvy]gg[filmv]</td>
<td>161-171</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>247-257</td>
</tr>
</tbody>
</table>

Table 3.5 Results of the IDENTIFY database search system.
Results for Myc are shown in Table 3.5A, and results for rABH are shown in Table 3.5B. The database search results are segregated into a series of stringency levels, based on the sequence requirements of the database entry. Matches between a query and an entry that has a poorly conserved consensus sequence are reported as lower stringency matches, and these stringencies are identified as the number of random matches expected. The next column over predicts how many false positives would be expected, based on the current number of entries in the database. The next two columns indicate the matched entry, as well as its consensus sequence. Lastly, the amino acid position within rABH where the sequence matched the indicated motif is shown.
3.2.10 SUMMARY OF THE INSIGHTS GAINED INTO rABH BY COMPUTER MODELING

The analysis of the rABH predicted polypeptide described in Sections 3.2.6 – 3.2.9 has provided a number of insights into some of the key properties of rABH; it has also failed to resolve the enigma surrounding how alkB functions to protect cells from SN2 alkylation damage. It also appears likely that AlkB does not belong to any previously identified protein family. The following is a brief summary of the data generated in the previous five sections.

The 1158bp ORF translates into a 386 amino acid polypeptide with a predicted Mw of 43.6kDa and pl of 8.29. The region surrounding residue 155 may be functionally important, since it was demonstrated to be the most hydrophilic, and therefore most likely accessible region of the protein. The secondary structure of rABH was also predicted by several algorithms. The N-terminal half of the protein was found to be very rich in helices, particularly between residues 1-50 and 100-150. The carboxyl half of the protein was predicted to have very few helical domains, and was predicted to be mostly coiled. RABH was predicted to be relatively unstable, although the algorithm results suggested it is more stable than the short-lived Myc control protein. Perhaps the most interesting prediction of all was that rABH is likely to be cytoplasmic, despite the presence of several NLS sequences.

Three different algorithms were applied to determine if rABH contained any known motifs. All were able to identify the control protein as Myc based on its numerous Myc-specific motifs, but they were unable to find any motifs in rABH. A number of potential phosphorylation sites and two potential Asparagine glycosylation sites were noted by PROSITE. However, the significance of PROSITE findings regarding phosphorylation should be considered with regard for the failure of PROSITE to identify both Thr58 and Ser62 in Mbl of Myc, both of which have been demonstrated to be phosphorylated in vivo. The GeneFIND algorithm was used to identify the human and E.coli AlkB homologs as highly similar to rABH, confirming our prior GenBank findings and our classification of the cDNA as a rat alkB homolog. Therefore, it appears that AlkB represents a novel protein that does not contain any previously identified motifs. Lastly, the IDENTIFY algorithm was used to search an extended database of protein identifying sequences, and two potential matches occurred in regions that have been highly conserved throughout evolution. However, the low level of stringency obtained for these matches raises doubts as to their validity, since three false positive matches were predicted to occur.
3.2.11 AN IN-DEPTH SEARCH FOR OTHER alkB HOMOLOGS

The absence of any identifiable motifs in rABH entrenches the enigma surrounding AlkB's mechanism of action. However, an important distinction needs to be made between the absence of existing motifs and the possibility that AlkB-specific motifs exist. The discovery of conserved residues and/or regions within AlkB has the potential to provide a great deal of insight into its mechanism of action, and could pave the way for a more focussed approach at elucidating its function. The capacity to perform this type of analysis has only become possible very recently, with the identification of additional AlkB homologs. To accomplish this, I set out to perform a multiple alignment of the predicted amino acid sequences of all the putative AlkB homologs identified to date. However, it was first necessary to perform a more thorough search for additional homologs that may not be a part of the nr GenBank database.

In Section 3.2.4, the results of a search of the GenBank database using the blastp algorithm were reported. Human, *Escherichia coli*, and *Caulobacter crescentus* homologs were identified in this search. (Colombi and Gomes, 1997; Kondo et al., 1986; Wei et al., 1996) In addition, similarity was also seen between rABH and the unpublished submission of potential *S. pombe* and *Arapodopsis thaliana* AlkB homologs.

In addition to GenBank, the Entrez server at http://www.ncbi.nlm.nih.gov/Entrez/nucleotide.html and http://www.ncbi.nlm.nih.gov/Entrez/protein.html was searched using "alkB" as a keyword. The entries in Entrez have been compiled from a variety of sources, including GenBank, EMBL, DDJB, PIR, SWISS-PROT, PRF, PDB and RefSeq. The Entrez search identified thirty nucleotide and thirty protein sequences containing the word alkB, that had been assigned to the entry based on ungapped blast alignments. In addition to the previously mentioned homologs, EST sequences from a putative murine homolog (Entrez Accession #AA560786) and *Gossypium hirsutum* (cotton plant) (Entrez accession # A1727846) were identified by this search.

In addition to the large databases that have been searched, there are a number of smaller, species-specific databases that have been developed to catalog the results of the various genome-sequencing projects. Several key species were selected for evaluation. A search of the *Caenorhabditis elegans* genome database http://www.sanger.ac.uk/DataSearch/omniblast.shtml identified a partial sequence of a potential homolog (Accession # Y51H7D 2980.D in the Sanger Center *C. elegans* genomic DNA database). The sequence similarity only included the 5' end of the rABH nucleotide sequence, and ended at the very end of the contig. Interestingly, the adjoining contig did not show the expected continuation of similarity to rABH, suggesting that
the sequence information is not complete for this particular region of the *C. elegans* genome. Searches of the *Saccharomyces cerevisiae* genome database, http://genome-www.stanford.edu/Saccharomyces/, and the *Drosophila melanogaster* genome database, http://www.fruitfly.org/blast/blastpage.html, using the nucleotide and amino acid sequences of *rABH* in conjunction with all of the blast algorithms, did not identify any sequence similarities of statistical significance, even at reduced stringency.

**3.2.12 AN IN-DEPTH SEQUENCE COMPARISON BETWEEN *hABH* AND *rABH***

The high degree of similarity between *rABH* and *hABH* was first identified by blast, and was mentioned in Section 3.2.4. However, the blast alignment report is limited in that it only shows a single stretch of sequence that has matched a GenBank entry; it does not show regions beyond the optimally matched sequences for that entry. It is therefore necessary to perform a separate alignment of the entire sequences to observe the degree of sequence conservation between two complete sequences. To achieve this objective, the predicted polypeptide sequences of *hABH* and *rABH* were analyzed by the clustal algorithm at http://bioweb.pasteur.fr/seqanal/interfaces/clustalw.html (Thompson *et al.*, 1994). The aligned sequences were then analyzed using the boxshade program at http://www.ch.embnet.org/software/BOX_form.html, which takes the aligned sequences and presents them in a format that facilitates visual analysis. The results of this process are shown in Figure 3.11.

Several characteristics about the amino acid sequence similarity between *rABH* and *hABH* were immediately noteworthy. First, the extent and fidelity of the conservation between the predicted amino acid sequences of *rABH* and *hABH* were exceptionally high. This is reflected in the extremely low E value that was generated by blast. The second observation was that this conservation begins and ends very abruptly, and there are additional amino acids at the N and C terminal ends of *rABH* relative to *hABH*. These unusual characteristics led us to investigate the similarity between *rABH* and *hABH* at the nucleotide level. To accomplish this, sequences of the human, rat, and murine *alkB* cDNAs were aligned using clustal and boxshade to determine if the nucleotide similarity ended in the same abrupt fashion. These three homologs were selected because they are the only three mammalian homologs identified to date, and are likely to be the most closely related. (Figure 3.12) Contrary to the amino acid alignment, the nucleotide sequences demonstrated a high degree of similarity between *rABH* and *hABH* throughout the entire length of the *rABH* coding sequence, with a dramatic loss of sequence
conservation shortly after the stop codon in rABH. This contrasts sharply to the absence of amino acid sequence similarity to hABH in the N and C terminal regions of the predicted rABH protein. Of particular interest was the nucleotide alignment at the position that corresponds to the divergence in the C-terminal amino acid sequences. Both the rat and murine homologs differ from the human homolog by the presence of an additional Thymine residue after nucleotide 1076 of rABH. The removal of this residue in hABH results in a frame shift in the putative amino acid sequence that extends the high degree of similarity to the predicted rABH polypeptide, up to and including a conserved stop codon. The nucleotide sequences of the three homologs diverge almost immediately after the conserved stop codon: significant gaps appear in the multiple alignment only 11 nucleotides downstream of the conserved stop codon.

At the 5' end of the nucleotide alignment, the removal of two nucleotides at position 100 and 134 and the addition of another after position 155 in hABH to match the rABH sequence result in a dramatic improvement in the conservation of the predicted amino acid sequences that more appropriately corresponds to the high degree of similarity observed between these two homologs at the nucleotide level. These changes are represented by the dashes in Figure3.12. Taken together, these sequence changes would result in the rABH and hABH polypeptides containing the same number of amino acids, with an extremely high degree of sequence conservation throughout the entire length of the putative proteins. An alignment of rABH and the altered hABH amino acid sequences is shown in Figure 3.13.
Figure 3.11. Alignment of the deduced human and rat AlkB protein sequences.

The amino acid sequences were first aligned using the clustal algorithm, and then processed using the boxshade program for improved visual presentation. The amino acid position is indicated immediately to the left of the residue that begins each line. Gaps in the alignment are indicated by a dash (-). Identical amino acids are boxed in black, and similar amino acids are blocked in gray. Amino acids in the rat homolog that are aligned with a gap are erroneously boxed in black. This is a default setting of the algorithm, and should be distinguished from a true match.
Figure 3.12 Alignment of the putative rat, human, and murine \textit{alkB} cDNA sequences.

The nucleotide sequences were first aligned using the clustal algorithm, and then processed using the boxshade program for improved visual presentation. The nucleotide position is indicated immediately to the left of the residue that begins each line. Gaps in the alignment are indicated by a dash (-). Conserved nucleotides are boxed in black. Similar nucleotides, ie: two purines or pyrimidines, are boxed in gray. The alignment was performed with the entire sequences of the putative homologs, but the last 239 residues of \textit{rABH} are not included in the figure. The weak degree of sequence conservation between the rat and human clones shown in the figure continued through to the end of the alignment. The stop codon that is conserved between the rat and murine cDNAs is indicated by asterisks below its sequence at position 1165 of \textit{rABH}. 
Figure 3.13. Alignment of the deduced human and rat polypeptide sequences, after the hABH cDNA sequence has been modified to match the rat and murine cDNA sequences.

The amino acid sequence of hABH was first translated based on the altered nucleotide sequence, and then aligned to the predicted rABH amino acid sequence using the clustal algorithm. The alignment was then processed using the boxshade program for improved visual presentation. This alignment did not contain any gaps. Identical amino acids are boxed in black, and similar amino acids are boxed in gray. In this particular alignment, both polypeptides contained the same number of amino acids.
3.2.13 AN ALIGNMENT OF ALL THE AlkB HOMOLOGS FOR THE PURPOSE OF IDENTIFYING CONSERVED SEQUENCES.

The absence of any AlkB motifs in the examined databases as described in section 3.2.9 does not preclude their existence. There has not been any mention in the literature of an attempt to identify conserved motifs or sequences based on an analysis of homologs identified to date. The recent identification of additional homologs, including rABH, makes such an attempt possible.

To identify potential residues / motifs that have been conserved throughout evolution, a clustal and boxshade analysis was performed on eight of the published and unpublished AlkB sequences. The putative murine homolog was not included due to its extremely truncated nature, in addition to the fact that it is virtually identical to the rABH sequence. For the human AlkB sequence, the “corrected” polypeptide that had been altered to match the predicted polypeptide sequence of rABH was used. The partial C. elegans AlkB polypeptide sequence was generated by translating contig Y51H7D 2980.D using the Expasy translation tool at http://www.expasy.ch/tools/dna.html, and selecting only the section that was displayed in the blast report of the C.elegans genome database screen as described in Section 3.2.11. The partial Gossypium hirsutum AlkB amino acid sequence was generated by translating the EST sequence AI727846 using the Expasy translation tool. The longest polypeptide generated from the six reading frames was confirmed to be similar to AlkB by comparing it to GenBank using the blastp algorithm. For the purposes of this alignment, the Met residue closest to the start of the cDNA is assumed to be the start of the polypeptide sequence. Amino acid 34 is represented by an X because the corresponding nucleotide sequence is NGC, and it is therefore not possible to definitively assign an amino acid to that residue. However, it could not be a stop codon, based on the identity of the second and third nucleotides.

The results of the multiple alignment are shown in Figure 3.14. The dominant feature of this alignment across a pool of evolutionarily distant homologs is the exceptional degree of sequence conservation in the C-terminal half of AlkB. The overall degree of conservation in the N-terminal half of the protein was considerably less. A fair degree of conservation was seen throughout the entire length of the eukaryote homologs, but this conservation was not as concentrated or as extensive as was observed in the C-terminal portion for all homologs.
<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>1</td>
<td>MAVAVSGLTEPGELEGFLRPEGSRSGTGA</td>
</tr>
<tr>
<td>Rat</td>
<td>1</td>
<td>MAVAVSGLTEPGELEGFLRPEGSRSGTGA</td>
</tr>
<tr>
<td>C.elegans</td>
<td>1</td>
<td>MEANVTELEKICRADAIP</td>
</tr>
<tr>
<td>S.pombe</td>
<td>1</td>
<td>NSUPAFKAEKLNRGSDWPG</td>
</tr>
<tr>
<td>C.crescentus</td>
<td>1</td>
<td>MESSANVSDADOZRAQPFNLYTEGKQSFKSRKFLFPI</td>
</tr>
<tr>
<td>E.coli</td>
<td>1</td>
<td>MESSANVSDADOZRAQPFNLYTEGKQSFKSRKFLFPI</td>
</tr>
<tr>
<td>A.thaliana</td>
<td>1</td>
<td>MESSANVSDADOZRAQPFNLYTEGKQSFKSRKFLFPI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>49</td>
<td>AQCKG------PAQIKISSQKSSGSTDRAFQ</td>
</tr>
<tr>
<td>Rat</td>
<td>49</td>
<td>TPQSP------FRHPFQFRPPNSRTEHDT</td>
</tr>
<tr>
<td>C.elegans</td>
<td>1</td>
<td>MEANVTELEKICRADAIP</td>
</tr>
<tr>
<td>S.pombe</td>
<td>36</td>
<td>NSUPAFKAEKLNRGSDWPG</td>
</tr>
<tr>
<td>C.crescentus</td>
<td>1</td>
<td>MESSANVSDADOZRAQPFNLYTEGKQSFKSRKFLFPI</td>
</tr>
<tr>
<td>E.coli</td>
<td>1</td>
<td>MESSANVSDADOZRAQPFNLYTEGKQSFKSRKFLFPI</td>
</tr>
<tr>
<td>A.thaliana</td>
<td>57</td>
<td>NYYNDGVLPEERKSVSDSPECKNGFGRFKTPDNLKKEKQGKRESLTSFPPRN</td>
</tr>
<tr>
<td>cotton</td>
<td>1</td>
<td>YLLKRYLSADAOQKVCRAGLGGGKTPX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>105</td>
<td>PGYGNVRWQKCLKLYQKPNVQCDTMSKEET</td>
</tr>
<tr>
<td>Rat</td>
<td>105</td>
<td>PGCCMNRLKQCLKYQKPNVQCDTMSKEET</td>
</tr>
<tr>
<td>C.elegans</td>
<td>1</td>
<td>TQPAQYTVETPR</td>
</tr>
<tr>
<td>S.pombe</td>
<td>87</td>
<td>KNLSQKQLPLGMKD---IMRKYNGDGEAIG---LGETPLTVDRNLK</td>
</tr>
<tr>
<td>C.crescentus</td>
<td>26</td>
<td>PSHTRTAYQPSVAAAGS</td>
</tr>
<tr>
<td>E.coli</td>
<td>46</td>
<td>FQKINTPTGGTMVSVAACIGS</td>
</tr>
<tr>
<td>A.thaliana</td>
<td>117</td>
<td>TNRHALGPIDDMSLARENVKVLQODLTNNWKAFTIPVDIEKATRRSCKSVSAVLYQ</td>
</tr>
<tr>
<td>cotton</td>
<td>35</td>
<td>TFRGDAKHLQSCMGDDPETGNTG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>165</td>
<td>GAPFLYRAKQFCLPFSDAEDSFGSFSQG</td>
</tr>
<tr>
<td>Rat</td>
<td>165</td>
<td>GAPFLYRAKQFCLPFSDAEDSFGSFSQG</td>
</tr>
<tr>
<td>C.elegans</td>
<td>17</td>
<td>TQPAQYTVETPR</td>
</tr>
<tr>
<td>S.pombe</td>
<td>135</td>
<td>KNSKQKQLPLGMKD---IMRKYNGDGEAIG---LGETPLTVDRNLK</td>
</tr>
<tr>
<td>C.crescentus</td>
<td>47</td>
<td>PSHTRTAYQPSVAAAGS</td>
</tr>
<tr>
<td>E.coli</td>
<td>67</td>
<td>FQKINTPTGGTMVSVAACIGS</td>
</tr>
<tr>
<td>A.thaliana</td>
<td>177</td>
<td>TNRHALGPIDDMSLARENVKVLQODLTNNWKAFTIPVDIEKATRRSCKSVSAVLYQ</td>
</tr>
<tr>
<td>cotton</td>
<td>62</td>
<td>TFRGDAKHLQSCMGDDPETGNTG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>222</td>
<td>PLGHHVLSLL-----------FSLPFLFS</td>
</tr>
<tr>
<td>Rat</td>
<td>222</td>
<td>PLGHHVLSLL-----------FSLPFLFS</td>
</tr>
<tr>
<td>C.elegans</td>
<td>73</td>
<td>KGSSRHRV</td>
</tr>
<tr>
<td>S.pombe</td>
<td>194</td>
<td>KNSKQKQLPLGMKD---IMRKYNGDGEAIG---LGETPLTVDRNLK</td>
</tr>
<tr>
<td>C.crescentus</td>
<td>101</td>
<td>PSHTRTAYQPSVAAAGS</td>
</tr>
<tr>
<td>E.coli</td>
<td>125</td>
<td>FQKINTPTGGTMVSVAACIGS</td>
</tr>
<tr>
<td>A.thaliana</td>
<td>237</td>
<td>TNRHALGPIDDMSLARENVKVLQODLTNNWKAFTIPVDIEKATRRSCKSVSAVLYQ</td>
</tr>
<tr>
<td>cotton</td>
<td>121</td>
<td>TFRGDAKHLQSCMGDDPETGNTG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>278</td>
<td>PSEILINA-----------FSLPFLFSEP</td>
</tr>
<tr>
<td>Rat</td>
<td>278</td>
<td>PSEILINA-----------FSLPFLFSEP</td>
</tr>
<tr>
<td>C.elegans</td>
<td>250</td>
<td>KSFGHYKMKFPR</td>
</tr>
<tr>
<td>S.pombe</td>
<td>250</td>
<td>KSFGHYKMKFPR</td>
</tr>
<tr>
<td>C.crescentus</td>
<td>157</td>
<td>PSHTRTAAYQPSVAAAGS</td>
</tr>
<tr>
<td>E.coli</td>
<td>181</td>
<td>FQKINTPTGGTMVSVAACIGS</td>
</tr>
<tr>
<td>A.thaliana</td>
<td>293</td>
<td>TNRHALGPIDDMSLARENVKVLQODLTNNWKAFTIPVDIEKATRRSCKSVSAVLYQ</td>
</tr>
<tr>
<td>cotton</td>
<td>181</td>
<td>TFRGDAKHLQSCMGDDPETGNTG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>329</td>
<td>SLKTAVYIVIVVLATQNFVFLEETEEKRDG</td>
</tr>
<tr>
<td>Rat</td>
<td>329</td>
<td>SLKTAVYIVIVVLATQNFVFLEETEEKRDG</td>
</tr>
<tr>
<td>C.elegans</td>
<td>180</td>
<td>YHLTHLHARTA</td>
</tr>
<tr>
<td>S.pombe</td>
<td>203</td>
<td>YHLTHLHARTA</td>
</tr>
<tr>
<td>C.crescentus</td>
<td>180</td>
<td>YHLTHLHARTA</td>
</tr>
<tr>
<td>E.coli</td>
<td>203</td>
<td>YHLTHLHARTA</td>
</tr>
<tr>
<td>A.thaliana</td>
<td>339</td>
<td>YHLTHLHARTA</td>
</tr>
<tr>
<td>cotton</td>
<td>339</td>
<td>YHLTHLHARTA</td>
</tr>
</tbody>
</table>
Figure 3.14 Alignment of the predicted AlkB polypeptide sequences of eight putative homologs identified to date.

The alignment was performed using the clustal algorithm. The alignment was then processed using the boxshade program for improved visual presentation. The amino acid position is indicated immediately to the left of the residue that begins each line.
The dramatic increase in the level of conservation that characterizes the second half of AlkB begins at amino acid 165 of rABH, with the presence of a leucine residue that is conserved amongst all eight homologs. This leucine residue is located in the middle of the p450 identifying sequence that was discovered using the IDENTIFY algorithm (Table 3.5). The subsequent six residues are also relatively well conserved, especially the two residues immediately following the universally conserved leucine. After a gap of two poorly conserved residues, the subsequent stretch of eight amino acids demonstrates a relatively high degree of conservation, but not to the same extent as the previous region. A third, weaker still area of sequence conservation occurs after a break of six amino acids. Interestingly, two proline residues are highly, although not universally, conserved in this stretch. The degree of conservation declines somewhat after this series until residue 210 of rABH. At this point the degree of similarity increases again, and remains at its highest level, with minor variation, until approximately amino acid 260 of rABH. This region also contains the Kelch repeat signature that was identified using the IDENTIFY algorithm (Table 3.5). Within this region there are several residues that are universally conserved, and most of the motifs identified by this search are found within this region. (Table 3.6) Another short gap in sequence conservation occurs after amino acid 260 in rABH, but there are two more regions of high sequence conservation that occur before the carboxyl terminus. The region from 272-288 contains three residues that are conserved throughout the seven homologs for which data are available, and two residues that are conserved for the six remaining homologs. Lastly, the region from amino acids 338-345 of rABH is highly conserved among five homologs, which interestingly include two mammalian and two procaryote homologs. Furthermore, the conserved sequences occur very close to the carboxyl terminus for the *E.coli*, *C.crescentus* and *A.thaliana* homologs.

The analysis described above has uncovered the AlkB sequences that have been conserved across an evolutionarily diverse group of homologs. Based on this analysis, several AlkB motifs can be proposed from the most highly conserved sequences. These putative alkB motifs identified by the multiple alignment are shown in Table 3.6.
<table>
<thead>
<tr>
<th>Homologs</th>
<th>Motif</th>
<th>Location in rABH</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>[rg][wp][vtsi][tsd][ldhrv][gar]</td>
<td>165 - 171</td>
</tr>
<tr>
<td>8</td>
<td>[yfp][dnrlg][wyi][dsntvp][stdpk][krie][kehdf][yf]</td>
<td>174 - 181</td>
</tr>
<tr>
<td>8</td>
<td>[pni][fmvid][ps][deas][lf]</td>
<td>188 - 193</td>
</tr>
<tr>
<td>8</td>
<td>[il][vli][n][yfr][yf]</td>
<td>218 - 222</td>
</tr>
<tr>
<td>7</td>
<td>[lm][gs][lipag][h][vqli][d][rkde][dsm]e</td>
<td>228 - 236</td>
</tr>
<tr>
<td>7</td>
<td>p[liv][liv][fli][vli][s][vli][m]</td>
<td>242 - 249</td>
</tr>
<tr>
<td>7</td>
<td>[ac][ivk][f][l][lrq][lif][y][g][td]</td>
<td>252 - 258</td>
</tr>
<tr>
<td>7</td>
<td>[sh][gd][vi][vmlc][v]</td>
<td>272 - 277</td>
</tr>
<tr>
<td>6</td>
<td>gj[fept][s][a][r][ke][la][c][fhy][h][ga]</td>
<td>280 - 288</td>
</tr>
<tr>
<td>5</td>
<td>r[viy][n][m][l][m][v][vl][f][r][qr]</td>
<td>338 - 345</td>
</tr>
</tbody>
</table>

Table 3.6 AlkB motifs.

The sequence of each motif was derived from the aligned AlkB homolog sequences in Figure 3.14. The number in the homolog column refers to the number of homologs for which sequence data was available to generate the motif sequence. Motif sequences are listed in the order which they occur from amino to carboxyl terminal, with the motif's position in rABH listed in the right-hand column. Residues located between the [ ] symbols represent the different residues that are found at that position in the different homologs. They are listed in decreasing order of frequency at that particular site. Gaps within the motif where a conserved amino acid(s) was not found are represented by a period (.)
3.2.14 Refined Mapping of hABH

The report describing the identification of hABH also described the mapping of hABH to chromosome 14q24 by FISH analysis (Wei et al., 1996). Determining the precise chromosomal localization of alkB is of particular importance if alkB is involved in eukaryote DNA repair mechanisms, as indicated by the apparent functional conservation between E. coli and human homologs of the gene. Many previously identified genes involved in DNA repair have been implicated in the development of disease (reviewed in chapter 14 of Freidberg et al., 1995), and a putative role for alkB in tumorigenises or other disease processes would be much easier to determine if the precise chromosomal location of hABH was known. The mapping of hABH to 14q24 by FISH analysis, as reported in Wei et al., is informative, but it does not provide enough resolution in hABH's genomic address to determine its potential involvement in cancer or any other disease. In order to achieve this objective, several different approaches were used to refine the chromosomal localization of hABH.

The initial approach used was a PCR-based screen of a somatic cell hybrid cell panel containing different segments of chromosome 14. First, two hABH-specific PCR primer sets were designed from nucleotide sequences within or near the suspected coding region of hABH (Figure 3.15). Their sequences, in addition to the predicted length of the PCR products, are shown in Table 3.7. Both primer sets were used as probes to detect hABH sequences in DNA extracted from a panel of somatic cell hybrids containing mapped segments of chromosome 14q. (Billingsley et al., 1994) They were first tested on murine and Chinese hamster DNA as a negative control. Next, the primer sets were used to amplify total DNA extracted from Wegroth-B3 cells. This cell line's human DNA component is restricted to a complete chromosome 14. Finally, the two primer sets were used in a PCR reaction with DNA extracted from the indicated hybrid cell lines as a template; the results are shown in Figure 3.15 Based on these results, the hABH gene was localized to between the markers PYGL, D14S978 (14q22) and D14S287 (14q31).
Figure 3.15 Primer design for the PCR screen of somatic cell hybrid and YAC panels.

Only a portion of the published hABH sequence is shown. Sequences used to generate PCR primers are blocked in black, and the PCR product in-between the primers is underlined. Capitalized nucleotides represent nucleotides that differ from rABH at the same position. The # symbol indicates the Thymine residue described in Section 3.2.12, that is absent in the mouse and rat homologs, and creates a gap in the nucleotide alignment shown in Figure 3.12. The three asterisks represent the stop codon suggested by Wei et.al as the translation end point for the hABH protein.
A

<table>
<thead>
<tr>
<th>Primer Set</th>
<th>Primer Sequence (5'-3')</th>
<th>Predicted length of PCR product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer 1</td>
<td>GAGGATGAAAAAAAAAGAGACA</td>
<td></td>
</tr>
<tr>
<td>Primer 2</td>
<td>AAAAAGGATGGGATCTCCCA</td>
<td>119 bp</td>
</tr>
<tr>
<td>Primer 3</td>
<td>TGGTCTTCAAAGGGATGAG</td>
<td></td>
</tr>
<tr>
<td>Primer 4</td>
<td>CTTCCCTTCTGGATTTGG</td>
<td>125 bp</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Parental cell lines (negative controls)</th>
<th>Chinese hamster, murine DNA</th>
<th>P1, P2</th>
<th>P3, P4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive control</td>
<td>Wegroth B-3</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Somatic Cell Hybrid Line</td>
<td>KHY-8B</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Somatic Cell Hybrid Line</td>
<td>W53-5-B</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Somatic Cell Hybrid Line</td>
<td>Rag-G03-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YAC contig spanning</td>
<td>773a8 (CEPH YAC address)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D14S298 – D14S277</td>
<td>762g10 (CEPH YAC address)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>924d6 (CEPH YAC address)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>938e10 (CEPH YAC address)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>804g10 (CEPH YAC address)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>788h12 (CEPH YAC address)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3.7 Summary of PCR mapping of hABH.

The PCR primers were first determined to be chromosome 14 specific, by using the Wegroth B-3 somatic cell hybrid, which contains all of chromosome 14 as the only source of human DNA. PCR was also performed on Chinese hamster and murine DNA as a negative control. PCR was then performed on a panel of hybrid cell lines whose human contributions contained various deletions of chromosome 14 as described previously (Billingsley et al., 1994). The results obtained from the hybrid panel screen place hABH between PYGL, D14S52 (14q22) and D14S13 (14q32.2) A YAC contig spanning the region between D14S298 and D14S277 was also screened, and a PCR product was not obtained with either primer set. Based on these results, hABH has been localized to between PYGL, D14S978 and D14S298, or between D14S277 and D14S287.
To further localize hABH within this region, PCR analysis was performed on a YAC (Yeast Artificial Chromosome) contig that spans from D14S298 to D13S277. The same PCR primer sets from the somatic cell hybrid screen were used in these experiments. No PCR product was detected for either primer set with DNA extracted from the contig components. (Table 3.7) The absence of a PCR product in this contig indicates that the region between D14S298 and D14S277 can be excluded as the region containing the gene, and the gene can be placed between PYGL, D14S978 and D14S298, or between D14S277 and D14S287.

The PCR based approach to localize hABH confirmed the assignment of the gene to chromosome 14q24 by FISH analysis as described in Wei et al.; however, it did not provide any new insights. Therefore, to localize hABH with greater resolution than the hybrid and YAC panel screenings, FISH analysis was performed using an hABH probe and a second “anchored” probe whose precise location is known. To generate an hABH FISH probe, the ORF probe fragment described in 3.1.2 was used to screen a P1-derived artificial chromosome (PAC) library to obtain a genomic clone of hABH. This genomic clone was then used in a dual FISH probing with a second anchored probe, PAC BP22 68F6, that has been previously described. (Roux et al., 1997). PAC BP 68F6 has been fine mapped to between the markers 2h10-L and 1-69 on chromosome 14q24. These markers are centromeric to the marker D14S277, but telomeric to D14S298. The results of the FISH analysis are shown in Figure 3.16. The PAC BP22 68F6 probe was labelled with biotin and detected with FITC (yellow) and the hABH probe was labelled with digoxigenin and detected with rhodamine (red). The hABH signal (red) was located telomeric to the BP22 68F6 signal, and the close, yet distinct signals indicates that hABH is a minimum of 1-2Mb telomeric to BP22 68F6. Therefore, the region between PYGL, D14S978 and D14S298 can be ruled out as the location of hABH. and hABH is located between D14S277 and D14S287. (Figure 3.17).
FITC-labeled Cos 22
Rhodamine-labeled MiG 1
Figure 3.16 FISH analysis of \textit{hABH}, using an anchored probe for reference

A genomic clone of \textit{hABH} was generated by screening a P1-derived artificial chromosome (PAC) panel and used as a probe in a FISH analysis of a metaphase spread of normal human lymphocytes (NHL). A second probe, PAC BP22 68F6, was used as an anchored reference point, whose precise location is known. The PAC BP22 68F6 probe was labelled with biotin and detected with FITC (yellow) and the \textit{hABH} probe was labelled with digoxigenin and detected with rhodamine (red). The \textit{hABH} signal was located telomeric to the BP22 68F6 signal, and the close, yet distinct signals indicates that \textit{hABH} is a minimum of 1-2Mb telomeric to BP22 68F6.
Figure 3.17 Combined results of hABH mapping to chromosome 14.

An ideogram of chromosome 14 is shown, with corresponding banding patterns indicated immediately to the left. To the right of the ideogram, the three columns of dashed lines represent the relative positions of various markers that have been mapped, with those in the third column labelled. The locations of the anchored PAC BP22 68F6 and hABH are shown, based on combining the results of the FISH and PCR analysis. HABH has been localized between the markers D14S277 and D14S287.
CHAPTER 4

DISCUSSION
4.1 SUMMARY OF RESULTS

In this thesis I have continued to investigate a cDNA clone that was identified in our laboratory during a search for potential Myc induced genes through subtractive hybridization cloning. Our efforts were focussed on three key aspects of this newly identified cDNA: its gene sequence and structure, its status as a Myc-induced gene, and its identity and function. By addressing these objectives, our understanding of Myc's role in diverse cellular processes such as proliferation, apoptosis, and oncogenesis would be advanced.

First, the nucleotide sequence of the cDNA was clarified, as a necessary first step to all further study. I also attempted to isolate an alternatively spliced cDNA of the gene, in an effort to explain the presence of two hybridization signals observed in northern blot studies. However, new sequence data were not identified in support of this hypothesis. I have confirmed that the gene is induced in response to Myc activation in an inducible system; furthermore, I determined that the transactivation domain present in the ER portion of the MycER fusion protein used to isolate the gene does not cause this induction. I observed elevated levels of both transcripts in cells that were constitutively expressing high levels of exogenous Myc. These results confirmed our initial observations regarding the gene's induction by Myc and characterized that induction in greater detail. A detailed analysis of the cDNA sequence has identified it as a rat homolog of the alkB gene that is part of the adaptive response of E.coli to DNA alkylating agents, and I have proposed the name rABH (rat AlkB homolog) for the gene. To elucidate regions likely important to the protein's elusive mechanism of action, detailed analysis was performed on rABH's putative amino acid sequence. The protein is predicted to be cytoplasmic, despite the presence of four potential NLS sites. The protein's secondary structure was also predicted, in addition to plotting the putative protein's hydrophobicity. An exhaustive search of various databases showed that the protein does not contain any previously identified motifs, with the exception of several commonly found post-translational modification sites. However, I have proposed several potential AlkB-specific motifs that were generated by comparing the predicted polypeptide sequences of multiple AlkB homologs. Lastly, a genomic clone of the recently identified human AlkB homolog, hABH, was identified, and fine mapped the gene to between the markers D14S277 and D14S287 on human chromosome 14q.
4.2 GENE SEQUENCE AND STRUCTURE

Previous experiments showed that the 1.0kb ORF probe hybridized to two distinct transcripts of approximately 2.0kb and 4.0kb by northern blot analysis which appear to be differentially regulated (G. Wasfy, M.Sc. thesis). These northern blot studies were conducted under conditions of high-stringency to minimize non-specific hybridization, so it is highly unlikely that either signal could be dismissed due to non-specific hybridization. Southern blot analysis (G. Wasfy, M.Sc. thesis), and the evidence presented in this thesis of the gene’s identity as alkB indicate that the gene is a single copy gene. The above evidence indicates that it is highly improbable that the presence of two hybridization signals is due to a second member in a gene family. Therefore, alternative splicing of a single gene is the most likely explanation for the two, differentially regulated hybridization signals seen in northern blot studies. To isolate a cDNA clone in support of this hypothesis, I screened a cDNA library generated from 40d old adult rat jejunum. I first attempted to derive a fragment of the cDNA that hybridized to only the 4.0kb transcript, in order to improve our chances of recovering a cDNA corresponding to the splice variant. Unfortunately, I was unable to do so. All prospective probes hybridized to both transcripts. More than 1 million plaques were screened, and seven positive clones were identified. All seven corresponded to various regions of the existing clone, as determined by sequencing the ends of the clones and restriction mapping.

One possible cause for the library screen’s failure to identify any new sequence data is that the average insert size of the 40d old rat jejunum library was 1.0kb, and larger transcripts were likely poorly represented. The isolation of six partial fragments of the original cDNA clone and only one cDNA similar in size to the 1.7kb rABH clone supports this hypothesis. The selection of a 40d old rat jejunum library was likely not a factor in the library screen’s failure. Previous studies showed that the 4.0kb transcript is present in all tissues in an equal or greater abundance than the 2.0kb transcript, with the possible exception of testis (G. Wasfy, M.Sc. thesis). The jejunum is composed largely of smooth muscle tissue, and therefore should contain both transcripts, based on previous data showing that both transcripts are expressed in muscle tissue. In conclusion, if screening a cDNA library to isolate a splice variant were to be repeated, a library that is size selected for very large transcripts should be used, regardless of the tissue source.

A more definitive way to answer the questions surrounding the structure of rABH will be the identification and sequencing of a genomic clone. The determination of the complete structure of the gene will provide us with all of the tools necessary to determine the identity of the 4.0kb transcript. An alignment between the existing cDNA sequence and the genomic sequence will
rapidly detect any additional exons that comprise the 4.0kb transcript, in addition to their position in the gene. Addition regulatory sequences present in mature transcripts will also be elucidated. Lastly, this data will also provide clues as the cause of the different regulation patterns between the 2.0kb and 4.0kb transcripts, and is essential to determining if rABH is in fact regulated by Myc.

4.3 IS rABH A MYC INDUCED GENE?

In this thesis I have furthered our understanding of the induction kinetics of rABH and confirmed, with much greater detail, earlier observations that rABH is induced in response to the activation of Myc in the inducible MycER system. (G. Wasfy, M.Sc. thesis) Various aspects of the induction and their relevance to the status of rABH as a transcriptional target of Myc are discussed below, in addition to insights from studies of other putative Myc target genes.

4.3.1 rABH INDUCTION OCCURS INDEPENDENTLY OF THE ESTROGEN RECEPTOR TAD

I have determined that rABH induction occurs in the absence of the weak TAD found within the ER portion of the MycER fusion protein. Furthermore, the extent of the induction is not reduced in the absence of ER TAD activity, suggesting that the ER TAD does not significantly contribute to rABH transcriptional activation in response to MycER activation. This was accomplished utilizing the modified MycER™ (Transactivation Mutant) construct in the induction kinetics time course experiment (Figure 3.5), in place of the MycER construct that was used in the subtractive hybridization cloning of rABH (Littlewood et al., 1995). Furthermore, the fusion protein is only responsive to the estrogen antagonist 4-OH Tamoxifen (4-OHT).

4.3.2 THE CHARACTERISTICS OF rABH EXPRESSION ARE CONSISTENT WITH THAT SEEN IN OTHER PUTATIVE MYC TARGET GENES

The extent of the induction that was observed was relatively weak (approximately two-three fold, Figure 3.5), although in our RNAse protection assay the induction may be somewhat muted by the presence of protected transcript fragments from the 2.0kb transcript that is not induced in response to Myc activation. As mentioned in Section 4.2, I have not been able to isolate a cDNA fragment that hybridizes specifically to the 4.0kb transcript and generate from
this fragment a more accurate probe for quantitating induction. Isolating and sequencing a genomic clone will provide invaluable assistance towards meeting the goal of identifying a 4.0kb transcript-specific probe.

The relatively weak induction of rABH is not particularly surprising, since other putative Myc induced genes demonstrate similarly weak degrees of induction. (see Table 1.1)

The induction kinetics that were determined are consistent with rABH transcription being induced by Myc. An increase in transcription was seen within 1 hour after MycER activation by RNase protection assay, which is similar to what has been observed for other putative Myc target genes in the same assay (Wagner et al., 1993; Grandori et al., 1996).

In addition to an increase in rABH transcription in response to a sudden increase in the amount of functional Myc within a cell, I have reported that cells constitutively overexpressing ectopic Myc show elevated levels of rABH (Figure 3.6). Interestingly, both transcripts were elevated in the two cell lines that demonstrated an increase in rABH RNA levels. Once again, this is consistent with some of the other Myc target genes (Cole and McMahon, 1999). Numerous aspects of this experiment are different from the induction experiments that demonstrated different regulatory patterns for the two transcripts, and these differences are likely responsible for the observation that both transcripts are elevated in the Myc expressing cells. The RNA isolated from cells constitutively expressing ectopic Myc comes from a population of asynchronously proliferating cells. Furthermore, these cells are rapidly proliferating in a richly supplemented environment in the presence of overabundant, deregulated Myc. These conditions differ significantly from the induction kinetics experiments, where G0 synchronized, quiescent cells were being held in growth arrest in a minimal nutrient environment and then triggered at the same moment to enter the cell cycle via either MycER activation or the addition of serum.

It is important to note that only two of the three cell lines examined showed misregulation of rABH in response to stable ectopic expression of Myc. Furthermore, and perhaps paradoxically, it is Rat-1 cells, the parental line of the Rat-1 MycER cells used to isolate rABH, that did not demonstrate a noticeable difference in rABH expression levels. However, basal levels of rABH appear to be higher in Rat-1 cells than the other two related cell lines that were examined. The cause of this is unknown, but it suggests that there are multiple factors involved in regulating rABH expression. This also suggests that the kinetics of induction should be further analyzed in the more responsive cell types.

In sum, our observations that rABH expression is elevated in response to the overexpression of Myc further suggests a link between Myc and rABH regulation, and more
importantly, it indicates that \( rABH \) may play a role in mediating the phenotype of Myc overexpressing cells.

4.3.3 FURTHER EXPERIMENTS ARE REQUIRED TO DEFINITIVELY CONCLUDE THAT \( rABH \) IS A MYC TARGET GENE

All of the results discussed above are consistent with \( rABH \) being a transcriptional target of Myc. However, they have not definitively determined that \( rABH \) is a direct target of Myc transactivation. Additional experiments are required to establish the regulatory role of Myc in \( rABH \) expression.

First of all, I have not presented any evidence to indicate that \( rABH \) induction occurs in the absence of protein translation. For example, induction of \( rABH \) in 4-OHT treated, confluent, quiescent Rat-1 MycER™ cells, in the presence of cyclohexamide, would indicate that induction does not require additional protein synthesis. This would make the induction a direct consequence of transcriptional activation by factors present at that time, presumably the newly activated MycER fusion protein. Furthermore, secondary gene cascades would also be eliminated as being responsible for the induction. Our ability to perform this experiment was hampered by various intricacies of working with MycER that were discovered while attempting to establish this and other experiments. First among these difficulties is that the MycER fusion protein is often "leaky", and cells harbouring the fusion protein can display characteristics of Myc activation in the absence of 4-OHT. This phenomenon becomes particularly noticeable with extended periods of cell culture that are necessary to maintain the cell line and seed cells for time course experiments. In addition to the problem of leakiness, the MycER fusion protein is very short-lived, with protein levels declining to negligible levels within 2 hours of activation in the presence of cyclohexamide. In spite of these difficulties, adaptations such as the use of early passage cell lines will hopefully allow for the successful completion of these experiments in the near future.

The recently developed homozygous Myc null cell line is another tool that will help determine the role of \( rABH \) in Myc-mediated growth and cell cycle progression (Mateyak et al., 1997). These experiments represent a loss of function approach to examine the expression patterns of \( rABH \), in opposition to the overexpression model that has been used to date. An evaluation of the proposed Myc target genes in this environment revealed that only \( cad \), \( c-myc \) and the Myc-repressed gene \( gadd45 \) showed a corresponding misregulation in the absence of Myc (Bush et al., 1998). This suggests that of the putative Myc target genes identified to date, only
gadd45 and cad could contribute to the slow growth phenotype of c-myc null cells. The corollary hypothesis of this is that only cad could be contributing to the proliferation seen in response to Myc, and that Myc is not required for the basal expression of the other putative target genes. This does not mean that cad and gadd45 are the only Myc target genes; however, it does indicate that Myc is not an essential regulator of most putative Myc target genes in the Myc null cell line. From a biological perspective, a target gene's relevance to promoting cellular proliferation is reduced if its expression is unaltered in a cell with a slow growth phenotype. The relevance of the current proposed Myc target genes to proliferation and cellular transformation has been put into question through this loss of function approach. It will be interesting to evaluate rABH expression patterns in the novel Myc-null system to determine the impact of a loss of Myc regulation on rABH gene expression.

The c-myc null cells also allow for improved experimentation with the inducible MycER construct. Reconstituting a c-myc null cell with an inducible allele will allow for complete control of Myc activity, and allow us to develop a more accurate picture of the regulatory role Myc plays in the expression of putative target genes, including rABH. For example, a highly accurate measure of the extent of a target gene's induction in response to MycER activation in the presence of cyclohexamide can now be obtained. The recent development of the MycER™ construct, as well as improvements in protocols when working with MycER™ cell lines, or the development of a completely novel inducible system represent major advances in our ability to assay Myc target genes, when combined with the Myc null cell line.

Our current understanding of Myc target gene activation is far from clear; however, one definite requirement is that Myc-Max heterodimers be shown to physically bind to an E-box sequence, resulting in an increase in transcription from that gene's promoter (Boyd et al., 1998). It is an ultimate prerequisite for adding rABH to the list of definite Myc target genes that an E box be identified within the gene, and that it be demonstrated to be capable of activating transcription through Myc-Max association with the E box sequence in vivo. There are no E box sequences in the rABH cDNA, which is not surprising, since the sequence is from a cDNA that likely contains additional sequences 5' to the current rABH sequence. Our understanding of rABH gene regulation, and Myc target gene regulation, will be greatly enhanced by the sequencing of a genomic clone, and the analysis of Myc's interaction with an E box, assuming one is found.

To summarize, although the data collected to date are indicative and consistent with our hypothesis that rABH is induced by Myc, it is not possible at this time to definitively add rABH to the list of candidate Myc induced genes. However, the data that have been presented herein clearly warrants further study.
**4.4 rABH and hABH: PROPOSED CHANGES TO THE PREDICTED hABH PROTEIN**

I have clearly demonstrated that the cDNA that was cloned in our laboratory is a rat homolog of alkB, a gene that is part of the adaptive response of *E. coli* to DNA alkylating agents. A detailed comparison between the rat and human homologs revealed that the nucleotide conservation was extremely high throughout the cDNA; however, the proposed amino acid sequences showed a single region of very high conservation, bounded by regions of almost no sequence conservation. I have proposed a single change to the hABH sequence in the interest of preserving amino acid conservation between the two proteins and the carboxyl terminal. This single deletion in hABH resulted in a frame shift that brought the predicted protein in the same frame as rABH, including a stop codon that is conserved among the human, rat and murine cDNA clones. I also proposed three changes in the 5' end of hABH to harmonize the rat and human sequences and maintain amino acid conservation. Furthermore, I have demonstrated that the sequence surrounding the ATG at position 7 in both rABH and hABH is a more attractive target for translation initiation than has been previously proposed for hABH (Wei *et al.*, 1996).

Potential support for our hypothesis may come from experiments that have been done examining cross-species functional complementation of alkB and hABH. AlkB from *E. coli* has been shown to protect human cells from MMS alkylation killing much more effectively (fifteen times increased survival) (Chen *et al.*, 1994) than hABH capacity to protect alkB- cells (four times increased survival) (Wei *et al.*, 1996). A quick experiment to determine if the proposed corrections to hABH are correct would be to repeat the cross-species complementation experiment using the full-length cDNA that has been previously reported (Wei *et al.*, 1996) instead of the partial subclone that lacked the first 200 nucleotides. The results of an experiment using the MMS sensitive HeLa cell line used in the *E. coli* cross-species complementation experiment and the partial and full-length hABH clones are another, and perhaps more direct, route to address this question (Chen *et al.*, 1994).

To definitively resolve the dispute, the human and rat proteins will have to be isolated and at least the N terminal region will have to be sequenced. A preliminary answer to this debate could be resolved through Western blot studies to determine the approximate molecular weight of hABH. In sum, it is our contention that it is unlikely that the human and rat nucleotide sequences would have retained such a high degree of similarity if the amino acid sequences had not also been conserved. However, the possibility remains that the human and rat sequences are correct as
is, and the rat and human genes have simply diverged during evolution, while retaining a common core section that is essential to function.

4.5 COMPUTER MODELING OF THE rABH PROTEIN

In this thesis I have presented numerous predictions on various characteristics of the putative rABH protein. The computer models used to generate these predictions are not meant to replace experimental investigation, however, they can suggest a course of action for future experimentation. The combined results of multiple algorithms used in our prediction of secondary structure have been shown to increase the accuracy of the predictions that result from this analysis (Guernier et al., 1999). Using Myc as a “control sequence” in generating our models provided an increased level of confidence in the predictions that were generated, perhaps with the exception of protein structure, since Myc’s structure has not been solved yet. However, the Myc protein has been thoroughly modeled, and the model generated from the algorithm that was selected corresponded with what has been previously reported (Marcu et al., 1992; Facchini and Penn, 1998).

The putative rABH protein has a predicted Mw of 43.6kDa, and is composed of 386 amino acids. A highly hydrophilic region is focussed at amino acid 155, suggesting that this region of the protein is likely exposed and perhaps crucial to function. This area is also potentially important for determining the subcellular localization of rABH, since NLS signatures are found at amino acids 155 and 156. The high degree of conservation that characterizes the carboxyl half of alkB homologs begins very close by, at position 165 of rABH.

The prediction of very few β sheets and the prediction of many α helices in rABH is consistent with the assumption that rABH is a globular, soluble protein. Other than this observation, there are few concrete conclusions that can be made based on the predicted secondary structure alone, and solving rABH protein structure will provide the necessary confirmation for the highly probable predictions, as well as clarify the many regions in rABH where secondary structure predictions are uncertain.

I have predicted that rABH is likely to be an unstable protein, although the control Myc protein, which is known to be highly unstable with a half-life of less than thirty minutes, had a much higher “instability score”. Again, Myc provided an excellent control, since the prediction matches what has been confirmed through protein-half life experiments. The observations of the transient nature of the adaptive response in E.coli, and the regulation of alkB transcription to coincide with the onset of DNA replication in C.crescentus suggest that alkB gene expression is
tightly regulated (Samson and Cairns, 1977; Colombi and Gomes, 1997). Therefore, it would not be surprising to find that the longevity of the protein is also tightly controlled through determining the half-life of rABH. A search of the PROSITE database revealed that rABH has the potential to be phosphorylated at numerous sites by Ser/Thr kinases, in addition to the presence of two potential Asn glycosylation sites. These post-translation modifications have been shown to be important in regulating protein function and protein targeting, respectively. These potentially important sites will be evaluated in the future to determine if they play a role in rABH function.

The most surprising result from our computer modeling is that rABH is predicted to be found in the cytoplasm. This prediction was made despite the presence of four potential NLS sequences at positions 155, 156, 376, and 378. In contrast to this, the NNCN algorithm used by the PSORT prediction system predicted rABH to be cytoplasmic; however, this neural network has an overall accuracy rate of 66% with only 1/3 of its predictions having a prediction accuracy greater than 82% (Reinhardt and Hubbard, 1998). The conflicting results of the NNCN calculation and the presence of four NLS sequences algorithms reduces the overall confidence in the algorithm's prediction that rABH is cytoplasmic. Myc, on the other hand, contains two NLS sequences, was predicted to be nuclear by the NNCN calculation with high confidence, and possesses two known DNA binding motifs. The combined results of these algorithms predict Myc to be nuclear with a high level of confidence. The prediction that rABH is a cytoplasmic protein is surprising, since alkB has been demonstrated to possess an intrinsic ability to protect cells from alkylation induced cytotoxicity, and the most direct explanation for this phenomenon is that alkB directly interacts with and thereby repairs alkylated DNA lesions. However, since the mechanism of alkB function remains a mystery and no DNA binding capacity has been found for alkB, the possibility that alkB functions in the cytoplasm and doesn't directly bind to DNA cannot be ruled out.

A possible explanation for the cytoplasmic prediction rests in the fact that alkB is a novel protein and does not belong to any identified family of proteins. I have already demonstrated that a detailed search for any known motif failed to find a match with rABH. Furthermore, the presence of a sequence that matches the criteria of an NLS does not guarantee that the putative NLS is used in vivo. In fact, our understanding of NLS function is far from complete (Moroianu, 1999). With the absence of any indications to the contrary, other than the presence of an NLS, the cytoplasm is likely the default destination for a protein in this series of algorithms. This may have been the case with alkB. The subcellular localization of rABH will be determined once antibodies to rABH have been isolated and used to probe for the subcellular localization of rABH by immunohistochemistry.
The computer modeling that was conducted has provided additional insights into potentially important regions of rABH, as well as generated some uncertainty as to where the protein is located within the cell. Several lines of evidence indicate that the region near amino acid 155 may be important for function. Some basic facts, such as predictions of molecular weight, are important pieces of information that will be crucial in resolving the dispute over the translation start site of rABH. Our inability to find any known motifs within rABH, despite an exhaustive search of numerous databases underscores the unique nature of this protein. Computer modeling is most useful when the sequence being studied is at least remotely related to previously identified proteins. However, as our understanding of various subjects such as post-translational modification and protein targeting increases, the predictive value of the computer models used to predict the characteristics of a novel protein will improve.

4.6 THE CONTRIBUTION OF rABH TOWARDS AN UNDERSTANDING OF ALKB

The greatest problem in studying AlkB is the lack of a direct assay for AlkB activity. The specific lesion(s) that AlkB either repairs or prevents from becoming cytotoxic are not known, despite the fact that S\textsubscript{N}2 alkylating agents and the lesions they create have been thoroughly characterized (Dinglay \textit{et al.}, 1998). In the absence of an assay for the mechanistic action of rABH that would allow for the investigation of a link between rABH induction following Myc activation and DNA alkylation repair, I directed our attention towards an analysis of the predicted amino acid sequence. Although some regions that have been extremely well conserved throughout evolution were discovered, in addition to several potential post-translational modification sites, no other significant clues as to AlkB mechanistic function were elucidated. Furthermore, our search for existing motifs within rABH confirmed that AlkB is completely novel. It does not contain any known catalytic domains associated with DNA repair; furthermore, if it does bind directly to alkylated DNA as part of its function, it does so via a motif(s) that has (have) not been previously identified.

In the absence of any similarity to other protein families, I also focussed my efforts on determining which amino acids or groups of amino acids have been conserved among the various AlkB homologs. This approach allows for the determination of regions of AlkB that have been conserved throughout evolution, and are therefore likely crucial to function. The carboxyl half of the protein was found to be highly conserved, and many residues were conserved across most or even all of the homologs analyzed. As a result of this process, I have proposed several AlkB
specific motifs that may prove useful in the identification of other homologs through such techniques as degenerate PCR screening. The identification of these motifs has further pinpointed specific residues and groups of residues that are almost certainly critical for function.

The degree of conservation is very impressive because it extends across the vast evolutionary distance between prokaryotes and mammals. Therefore, the carboxyl half of AlkB has remained relatively unchanged throughout evolution. This degree of conservation across such broad evolutionary lines is frequently seen in proteins required for basic metabolism and other essential functions, and is exemplified by the extremely high degree of conservation seen in hsp70 (Gupta and Singh, 1994; Gupta, 1998). The highly conserved carboxyl half of AlkB is likely the portion that is responsible for mitigating the effects of DNA alkylation, as evidenced by cross-species complementation experiments (Chen et al., 1994; Wei et al., 1996).

It is interesting to note that the Kelch repeat and p450 motif matches that were obtained using the IDENTIFY algorithm are located in regions of high conservation, suggesting that these matches may be relevant to function. The combination of database searches and multiple alignments is a demonstration of the synergistic abilities for data analysis when the results of different algorithms are examined together.

The Kelch motif was first discovered as a repeated element in the Drosophila ORF1 protein (Xue and Cooley, 1993), and has subsequently been found in many proteins (Adams et al., 2000). The discovery of Kelch motifs in the crystal structure of fungal galactose oxidase revealed that each Kelch motif forms a single, four-stranded antiparallel β-sheet, and between four and seven such structures combine to form a structural motif called a β-propeller, with each Kelch motif representing one of the propeller’s “blades” (Ito et al., 1994; Bork and Doolittle, 1994). The precise function of Kelch repeats has not been determined; however, recent reports have suggested that it is important in protein-protein interactions, particularly with cytoskeletal proteins such as actin (Adams et al., 2000). It is unlikely that the match between rABH and the Kelch repeat is a true positive, since only one repeat element was found, and the β-propeller structure cannot form without multiple units.

The match between rABH and a signature sequence for the p450 superfamily of enzymes also occurs in a highly conserved region of rABH. The p450 superfamily contains over one thousand genes that have been identified to date, and this superfamily can be further divided into 70 different families (Peterson and Graham, 1998). The absence of a match to any of the subgroups in the p450 superfamily makes it difficult to determine that this match is a true match. However, AlkB may represent a novel family within this larger group, although further sequence and function analysis are required to test this hypothesis. It is interesting to note that the
degradation and detoxification of various compounds is a function performed by many p450 superfamily members (Chang and Kam, 1999), and the detoxification of $S_{N2}$ alkylating agents may represent an alternative mechanism of action to the direct repair of alkylated nucleotides. To test this hypothesis, I propose that the ability of $S_{N2}$ alkylating agents to alkylate DNA should be tested after exposure to purified AlkB. This hypothesis would also explain the intrinsic nature of AlkB function that allows AlkB from one species to function in another species.

The core region of AlkB in the carboxyl half of the protein has clearly been conserved throughout evolution. However, the amino acids preceding position 165 of rABH show very little conservation with other homologs, and this is interesting to note for several reasons. It suggests that there are species specific aspects to each homolog. This in turn suggests that AlkB may interact with other proteins that are specific to each cellular environment, or has incorporated various elements essential to its function in various organisms. For example, it is interesting to note that the four NLS sequences found in rABH are located either before or after the highly conserved region, suggesting that these elements have been added onto the existing AlkB platform, perhaps to assist in coping with the increased cellular complexity of eukaryotes.

Key hurdles that must be overcome before a thorough understanding of AlkB can be developed are the determination of the particular DNA lesions that are repaired by AlkB, and the precise mechanistic action of AlkB. In the absence of a mechanistic assay, the use of an assay for survival in response to $S_{N2}$ alkylating agents provides a useful approach for studying the various domains that have been identified as likely crucial to function. The first experiment that should be conducted is to confirm that rABH performs the same function as other AlkB homologs. This can be achieved by examining the effect of overexpressing the full length ORF of rABH on Rat-1 cell sensitivity to $S_{N2}$ alkylating agents, and the production of alkylated DNA bases. Conversely, the impact of abrogating rABH expression via antisense RNA should also be examined. To determine a potential role for rABH as a mediator of Myc activation, the impact of overexpressing the full length ORF of rABH on various Myc associated functions such as the cell cycle, and apoptosis should also be examined. The importance of the evolutionarily conserved region of AlkB should also be evaluated through the expression of only the conserved domains, or through site directed mutagenesis of universally conserved residues, and then assaying for $S_{N2}$ alkylating agent sensitivity. The results of these experiments will confirm the functional role of rABH, and help determine a precise mechanism of action by identifying crucial regions and individual amino acids.
4.7 REFINED MAP OF hABH

The first and only published report on hABH included data on the mapping of the gene to chromosome 14q by in situ hybridization (Wei et al., 1996). I have provided a more precise genomic localization of between the makers D14S 277 and D14S 287. Of the several genes that have been mapped to this region, the most well known is the Presenilin 1 (AD3) gene, associated with familial early on-set Alzheimer’s disease (Van et al., 1992). There have been reports in the literature demonstrating a correlation between the loss of chromosome 14q and cancer. In particular, loss of 14q in the region of hABH has recently been observed to occur more frequently in carcinomas of the distal esophagus than in the gastric cardia, and may prove important in distinguishing tumours in the gastroesophageal junction (GEJ) as either esophageal or gastric in origin (van et al., 1999). Further studies are required to determine whether or not the loss or mutation of hABH plays a role in tumourigenesis or other disease processes.

4.8 MYC, AlkB, AND DNA ALKYLATION

Myc and AlkB are very similar in that they remain enigmatic, despite many years of dedicated effort to elucidate their precise mechanism of action. Although several lines of investigation must be completed before it can be definitively said that alkB is a Myc target gene, the preliminary evidence presented here supports this hypothesis. If this is indeed so, the question arises as to why a gene involved in DNA repair would be regulated by Myc. Another DNA repair gene p53, has already been suggested as a Myc-induced gene, although it is not yet clear if p53 is a bona fide target (Reisman et al., 1993; Tavtigian et al., 1994; Shim et al., 1997). Interestingly, the C. crescentus alkB homolog is regulated in a cell cycle dependent manner in a similar fashion to what would occur if Myc did in fact regulate hABH and rABH. The C. crescentus alkB gene is differentially regulated to correlate with progression through the cell cycle and preparation for DNA replication (Colombi and Gomes, 1997). The mechanism regulating this expression has been characterized, and it has been suggested that an increase in the amount of AlkB in this species of bacteria occurs as a sentinel against potential alkylation events in the extra DNA that has resulted from replication. Also of interest is the fact that C. crescentus does not appear to have an adaptive response; mammalian cells also lack this response (Colombi and Gomes, 1997). It is possible that the adaptive response has been replaced by a more sophisticated regulatory system that has been tied into cell cycle progression.
A more pressing question still is if and how AlkB contributes to the many biological functions that have been associated with Myc. Under normal proliferation circumstances, Myc induction of alkB may serve the same function as the regulatory network in *C. crescentus*, and serve as a sentinel to protect the increased levels of DNA prior to mitosis from alkylation. The retention of alkB in a Myc-amplified tumour may also be a positive selector for the progression of cancers towards chemotherapeutic resistance and an indicator of poor prognosis. AlkB deregulation would allow cells to cope with increased alkylation and thereby generate resistance to chemotherapeutics, since many chemotherapeutics work through alkylation of DNA. In the absence of a direct functional assay, this hypothesis could be tested by comparing resistance to alkylation of chemotherapeutics with AlkB levels in various tumour cell lines. Furthermore, the examination of tumour cells for relative levels of Myc and AlkB, and correlating those findings to the sensitivity of the tumour cell to SN2 alkylation agents, may eventually lead to improved treatment of advanced cancers.

4.9 CONCLUDING REMARKS

There remains a great deal of mystery surrounding target gene activation and the role of Myc in cellular proliferation and transformation. Since no compelling target genes have been isolated that can mediate Myc’s profound impact upon cell cycle progression, apoptosis, and cellular transformation, our laboratory initiated a search for Myc-induced genes. In this thesis, I have investigated a cDNA that was isolated in our lab via a subtractive hybridization approach using the inducible MycER construct. I have confirmed its sequence and concluded that it is a rat homolog of the alkB gene. I have provided further evidence through detailed induction kinetics and constitutive Myc overexpressing cells that alkB is induced by Myc. I also discovered the presence of several highly conserved motifs in the AlkB homologs identified to date, and determined that AlkB does not contain any previously identified motifs that can account for its function. The motifs that were identified are likely crucial to the undiscovered mechanism of how AlkB mitigates the toxic effects of Sn2 DNA alkylation agents, and should be primary targets for further structure – function studies. Lastly, I have mapped the human homolog of alkB to chromosome 14q with greater accuracy than has previously been reported.

The development of recent models such as the c-myc null cell, in addition to the discovery of new cofactors linking Myc to transcriptional regulatory complexes should allow for the elucidation of new target genes, and an improved understanding of the current list of proposed targets, including alkB. AlkB may be a bona fide target, but more work needs to be done to
ascertain its status as a Myc induced gene, and the role that it plays in the cellular processes that have been linked to Myc. The isolation of this gene through an overexpression / induction model has indicated the value of screening for target genes in such a fashion, and complements the lack of function approach of c-myc null cells for further evaluation. If alkB is in fact a true Myc target gene and we can unravel the mystery of how it works to protect cells from the toxicity of alkylating agents, our understanding of DNA alkylation and repair, and the related issues of carcinogenesis and cancer therapy will be greatly advanced.
CHAPTER 5

REFERENCES
5.1 REFERENCES


independent prostate cancer detected by comparative genomic hybridization and allelotyping. *Cancer Res.* 56, 3091-3102.


