The Use of the Yeast Two-Hybrid System as a Means of Identifying Protein Interactors of the Human Protein, BMI-1

Robert George Edward Holgate

A thesis submitted for the degree of
Doctor of Philosophy
at the University of London

March 1999

Institute of Child Health
30 Guilford Street
London
WC1N 1EH
Abstract

In *Drosophila melanogaster*, the establishment of the expression patterns of the Homeotic complex (HOM-C) genes, involved in the anteroposterior organisation of the developing embryo, are initiated by maternal-effect genes and segmentation genes. These expression patterns defining differentially determined states need to be maintained over many cell generations after initiation and, to ensure this, two classes of genes have evolved, the Trithorax-Group and the Polycomb-Group (Pc-G), which act to maintain the spatially restricted expression patterns in a positive and negative manner respectively. Similar mechanisms to those found in *Drosophila* are thought to exist in vertebrates to regulate expression of the vertebrate Hox genes.

BMI-1 is a 45KDa nuclear protein with several features in common with other nuclear proteins which suggest a role in transcriptional regulation. Furthermore, BMI-1-specific motifs are conserved in a related gene in *Drosophila, posterior sex combs (Psc)*, a member of the Polycomb-Group of proteins. The mode of action of the Polycomb-Group proteins is thought to be through the formation of higher-order chromatin structures, and so it was postulated that BMI-1 would interact with several other proteins, with other Pc-G proteins seeming suitable candidates.

In order to identify such interactors, a truncated BMI-1 fragment was used as bait to perform several library screens in a GAL4-based yeast two-hybrid system. From these screens, several clones were identified as potential interactors. In particular, two non-Polycomb-Group clones were each identified from two different species, and were deemed suitable candidates for further analysis. One encodes a putative novel protein of unknown function, while the other encodes a protein, MCM6, involved in the control, or ‘licensing’ of DNA replication. Further work has been aimed at characterising these interactions further, and corroborating the interactions using two different techniques: GST-affinity capture and co-immunoprecipitation. In addition, further analysis of the novel protein has also been performed.

At present, the balance of evidence suggests that the novel protein is a false positive, and thus not a real interactor of BMI-1. Further work is required to investigate the MCM6-BMI-1 interaction and, if proved, could provide valuable insights into Polycomb-Group-mediated inheritance of repressive chromatin states. In addition, the interaction may also help in the understanding of how replication of DNA is able to proceed, despite the presence of higher order chromatin structures.
Acknowledgements

To start, I would like to thank CHRAT for funding this project over the last three years or so, and everyone who sent clones or strains or libraries that have assisted in this project. I would also like to thank my ‘boss’, Professor Pete Scambler for his comments on this thesis and his support during my time here (by the way, I hope Preston N.E. get promoted this year, Pete).

As for everyone else, there are too many to mention, but mainly amongst them are those in the lab, past and present, (Robin, Hannah, Gaetan, Judith, and Paola), my old flatmates (Little Phil, Big Al and Lucy), The Lamb, Tiffany, and my Mum, Dad and Brother. Many apologies to any of the numerous people that I have missed out.

Oh, and how could I possibly forget Matt (Thumper) and Pete (Rabid Wombat) and G. for the regular slappings they received at Chaos Quake.
List of Abbreviations

aa  amino acid
AD  Activation Domain (of a transcription factor)
Ade  Adenine
ANT-C  Antennapedia complex
AP  Alkaline Phosphatase
ATP  adenosine triphosphate
BD  Binding Domain (of a transcription factor)
β-Gal  β-Galactosidase
BLAST  Basic Local Alignment and Search Tool
bp  basepair
BMI-1  B-cell Moloney murine leukaemia virus Insertion region 1
BSA  bovine serum albumin
BX-C  Bithorax complex
CATCH22  Cardiac abnormalities, Abnormal facies, Thymic hypoplasia, Cleft lip, Hypocalcaemia associated with hemizygosity for human chromosome 22q11
CHX  Cycloheximide
Ci  Curie
Da  Dalton
ddH2O  distilled, deionised water
DGS  DiGeorge syndrome
DMSO  dimethylsulphoxide
DNA  deoxyribonucleic acid
dpc  days post coitum
DTT  Dithiothreitol
EDTA  Ethylene diamine- tetraacetic acid
EGTA  Ethylene glycol bis (β-aminoethyl)ether) N, N, N’, N’- tetraacetic acid
EST  Expressed Sequence Tag
fBMI-1  Full-length BMI-1 protein
g  gram
g  acceleration due to gravity
GST  glutathione-S-transferase
HGMP  Human Genome Mapping Project
hr  hour
IPTG  isopropylthiogalactoside
Kilo
K  Kilo
LacZ  the gene encoding β-Galactosidase
l  litre
LMP  low melting-point agarose
µ  micro
m  milli
M  Molar
MCM  minichromosome maintenance
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>MoMLV</td>
<td>Moloney murine leukemia virus</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>n</td>
<td>nano</td>
</tr>
<tr>
<td>NLS</td>
<td>Nuclear Localisation signal</td>
</tr>
<tr>
<td>OLB</td>
<td>oligonucleotide labelling buffer</td>
</tr>
<tr>
<td>ONPG</td>
<td>o-Nitrophenyl β-D-Galactopyranoside</td>
</tr>
<tr>
<td>ORC</td>
<td>Origin Recognition Complex</td>
</tr>
<tr>
<td>ORF</td>
<td>Open Reading Frame</td>
</tr>
<tr>
<td>p</td>
<td>pico</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate Buffered Saline</td>
</tr>
<tr>
<td>Pc-G</td>
<td>Polycomb-Group</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene Glycol</td>
</tr>
<tr>
<td>PEST</td>
<td>proline (P), glutamic acid (E), serine (S), threonine (T) motif</td>
</tr>
<tr>
<td>PEV</td>
<td>position effect variegation</td>
</tr>
<tr>
<td>PRE</td>
<td>Polycomb Response Element</td>
</tr>
<tr>
<td>PVDF</td>
<td>polyvinylidene difluoride</td>
</tr>
<tr>
<td>RACE</td>
<td>Rapid Amplification of cDNA Ends</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rt</td>
<td>room temperature</td>
</tr>
<tr>
<td>RT</td>
<td>reverse transcription</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>sec</td>
<td>second</td>
</tr>
<tr>
<td>ssDNA</td>
<td>salmon sperm DNA</td>
</tr>
<tr>
<td>STS</td>
<td>Sequence Tag Site</td>
</tr>
<tr>
<td>tBMI-1</td>
<td>Truncated BMI-1 fragment corresponding to the A 2330bp EcoRI fragment (C-ter 191 amino acids) of human BMI-1 (K562) cDNA</td>
</tr>
<tr>
<td>TCA</td>
<td>trichloroacetic acid</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-HCl/EDTA buffer</td>
</tr>
<tr>
<td>TEMED</td>
<td>N, N, N', N' - tetramethylethylenediamine</td>
</tr>
<tr>
<td>THC</td>
<td>Tentative Human Consensus</td>
</tr>
<tr>
<td>TRIS</td>
<td>2-amino-2-[hydroxymethyl]-propane 1,3 diol</td>
</tr>
<tr>
<td>Trx-G</td>
<td>Trithorax Group</td>
</tr>
<tr>
<td>UAS</td>
<td>Upstream Activating Sequence</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>V</td>
<td>volts</td>
</tr>
<tr>
<td>WWW</td>
<td>World Wide Web</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo, 4-chloro, 3-indolyl-D galactoside</td>
</tr>
</tbody>
</table>
Table of Contents

<table>
<thead>
<tr>
<th>Title</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>I</td>
</tr>
<tr>
<td>Abstract</td>
<td>II</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>III</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>IV</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>VI</td>
</tr>
<tr>
<td>List of Tables</td>
<td>XIII</td>
</tr>
<tr>
<td>List of Figures</td>
<td>XIII</td>
</tr>
</tbody>
</table>

Chapter 1  Introduction

1.1 The Role of Drosophila in the Understanding of Developmental Pathways 1
  1.1.1 Introduction to Drosophila melanogaster 1
  1.1.2 Drosophila Developmental Genetics 2
  1.1.3 The Establishment of the Anteroposterior axis 2
  1.1.4 Pair Rule Genes 3
  1.1.5 Segment Polarity Genes 4
  1.1.6 Homeotic Selector Genes 4

1.2 Polycomb 6
  1.2.1 The Identification of Polycomb 6
  1.2.2 Other Polycomb Genes in Drosophila 7
  1.2.3 Polycomb and Chromatin Silencing 8

1.3 Conservation in Mammals 9
  1.3.1 Vertebrate Hox Genes 9
  1.3.2 Vertebrate Polycomb Genes 10

1.4 Mechanistic Aspects of Polycomb Action 12
  1.4.1 Initiation of Polycomb-Group Protein Activity 12
  1.4.2 Mechanisms of Inactivation 13
  1.4.3 Polycomb-Group Proteins and Genetic Memory 18
  1.4.4 Polycomb-Group Proteins and Kinetochore – Functional Relevance? 18
  1.4.5 Polycomb and Trithorax 19
  1.4.6 Hox Gene and Polycomb-Group Gene Regulation. 20

1.5 Beyond Anteroposterior Patterning 22
  1.5.1 Polycomb-Group Proteins and Hematopoiesis 22
  1.5.2 Bmi-1 and Cell Cycle Control 24
  1.5.3 M33 and Sex Determination 25
  1.5.4 Other Functions of Polycomb-Group Proteins 25
### 1.5.4.1 Polycomb-Group Proteins in Plants

### 1.5.4.2 Cosuppression

25

### 1.6 Bmi-1

1.6.1 Cloning of the Gene

1.6.2 Posterior Sex Combs

1.6.3 Bmi-1 as a Vertebrate Polycomb-Group Member

26

### 1.7 Strategies for the Identification of Protein-Protein Interactions

1.7.1 Introduction

1.7.2 The Yeast Two-Hybrid System

1.7.2.1 Historical Perspectives and Theory

1.7.2.2 Yeast Two-Hybrid System – Methodology

1.7.2.3 Yeast Two-Hybrid System – Applications

1.7.2.4 Yeast Two-Hybrid System – Advantages and Limitations

30

### 1.7.3 Further Developments of the Yeast Two-Hybrid System

1.7.3.1 The One-Hybrid System

1.7.3.2 The Mammalian Two-Hybrid System

1.7.3.3 The Reverse Two-Hybrid System

1.7.3.4 Three-Hybrid Systems

1.7.3.5 Exhaustive Screening

33

### 1.8 Project Aim

34

### Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Reagents

2.1.2 Materials

2.1.3 Commercial Kits

2.1.4 Enzymes

2.1.5 Immunoglobulins

2.1.6 Radioisotopes

2.1.7 Size Markers

2.1.8 Bacterial Strains

2.1.9 Yeast Strains

2.1.10 DNA Clones and Libraries

2.1.11 Vectors

2.1.12 Oligonucleotides

2.1.13 Cell Lines

38

### 2.2 Solutions, Buffers and Media

2.2.1 DNA Related Solutions

2.2.1.1 General Solutions

2.2.1.2 Solutions for DNA Preps

2.2.1.3 Hybridisation/pre-hybridisation/Wash solutions

41
2.2.1.4 Random-Primming Buffer
2.2.1.5 DNA Electrophoresis Buffers
2.2.1.6 Gel Loading Buffer
2.2.1.7 Lambda Diluent
2.2.1.8 Restriction Enzyme Buffer
2.2.1.9 Ligation Buffer
2.2.1.10 Calf Intestinal Phosphatase Buffer
2.2.1.11 Taq PCR Buffer
2.2.1.12 Pfu Taq PCR Buffer
2.2.1.13 cDNA First Strand Synthesis Buffer
2.2.1.14 cDNA Second Strand Synthesis Buffer
2.2.1.15 Second Strand Enzyme Cocktail
2.2.1.16 Adaptor Ligation Buffer

2.2.2 Protein Related Solutions
2.2.2.1 Protein Gel Buffers
2.2.2.2 10X Protein Electrophoresis Buffer
2.2.2.3 2X Gel Loading Buffer
2.2.2.4 Fixing Solution
2.2.2.5 Staining Solution
2.2.2.6 Destaining Solution
2.2.2.7 Protein Transfer Buffer
2.2.2.8 Bradford Buffer
2.2.2.9 PVDF and Hybond C Membrane Blocking Solutions

2.2.3 Yeast Manipulation Solutions
2.2.3.1 Transformation Solutions

2.2.4 Bacterial Transformation Solutions
2.2.4.1 Electrocompetent Cell Preparation
2.2.4.2 Chemical Competent Cell Preparation

2.2.5 Media
2.2.5.1 Bacterial Media
2.2.5.2 Antibiotics
2.2.5.3 Colour Selection
2.2.5.4 Yeast Media

2.2.6 GST-Related and other Protein Buffers
2.2.6.1 GST Purification Buffer
2.2.6.2 In vitro Assay Buffer
2.2.6.3 Nuclear/Cytoplasmic Preparation Buffers

2.2.7 Immunoprecipitation Lysis Buffers
2.2.7.1 General Immunoprecipitation Buffers
2.2.7.2 Protease Inhibitors

2.2.8 Mammalian Cell Culture Solutions
2.2.8.1 Cell Culture Media
2.2.8.2 Cell Culture β-Gal Assay Buffer

2.2.9 Oligonucleotides and their Uses

2.3 Methods
2.3.1 Isolation and Purification of DNA
2.3.1.1 Plasmid Minipreps
2.3.1.2 Isolation of DNA from Mammalian Cell Cultures
2.3.1.3 Plasmid Maxipreps
2.3.1.4 Purification of Phage DNA

2.3.2 Analysis of DNA and RNA
  2.3.2.1 Quantitation of Nucleic Acid Concentrations
  2.3.2.2 Purification of Deprotected Oligonucleotides
  2.3.2.3 Restriction Analysis of DNA
  2.3.2.4 Electrophoresis of DNA
  2.3.2.5 Southern Blotting
  2.3.2.6 Phage Library Plating
  2.3.2.7 Transfer of Phage DNA to Nylon Membranes
  2.3.2.8 Preparation of DNA Probes
  2.3.2.9 Radioactive labelling of DNA Probes
  2.3.2.10 Hybridisation
  2.3.2.11 Multiple Tissue Northern (MTN) Blot Hybridisation

2.3.3 Subcloning DNA
  2.3.3.1 Preparation of Vectors and Inserts
  2.3.3.2 Ligation of DNA into Vectors
  2.3.3.3 Transformation of Bacteria with Plasmid DNA
  2.3.3.4 Preparation of Electrocompetent Bacterial Cells
  2.3.3.5 Preparation of Chemically Competent Bacterial Cells

2.3.4 Sequencing
  2.3.4.1 Manual Sequencing
  2.3.4.2 Automated Sequencing

2.3.5 Polymerase Chain Reaction
  2.3.5.1 General PCR
  2.3.5.2 RT-PCR and Template Production for RACE

2.3.6 Protein Analysis
  2.3.6.1 Protein Precipitation
  2.3.6.2 Measurement of Protein Concentration
  2.3.6.3 SDS-Polyacrylamide Gel Electrophoresis
  2.3.6.4 Visualisation of Proteins
  2.3.6.5 Autoradiography

2.3.7 Immunochemistry
  2.3.7.1 Immunoblotting
  2.3.7.2 Western Analysis
  2.3.7.3 Coupling Antibodies to Protein-A-Sepharose Beads
  2.3.7.4 Immunoprecipitation

2.3.8 In vitro Protein Analysis
  2.3.8.1 GST-fusion Protein Purification
  2.3.8.2 RNA Synthesis for In vitro Protein Synthesis
  2.3.8.3 In vitro Protein Synthesis
  2.3.8.4 TNT (Quik) Coupled Transcription/Translation Protein Synthesis
  2.3.8.5 In vitro GST-affinity Capture Assay
  2.3.8.6 Nuclear/Cytoplasmic Preparations

2.3.9 Yeast Analysis
  2.3.9.1 Yeast Cell Preparation
  2.3.9.2 Small Scale Yeast Transformation
  2.3.9.3 Library Scale Yeast Transformation
  2.3.9.4 Colony β-Gal Assay
  2.3.9.5 Liquid β-Gal Assay
  2.3.9.6 Recovery of Activation Domain Plasmids from Yeast
2.3.10 Cell Culture Techniques
- 2.3.10.1 Maintenance of Cells
- 2.3.10.2 Freezing and Thawing of Cells
- 2.3.10.3 Transfection of Cells
- 2.3.10.4 β-Gal Assay of Transfected Cells

2.3.11 Database Analysis
- 2.3.11.1 Computer Analysis

Chapter 3 Results

3.1 Work Preceding this Project

3.2 Preliminary Experiments
- 3.2.1 Planning
- 3.2.2 Subcloning
- 3.2.3 Autoactivation Analysis

3.3 Yeast Two-Hybrid Interaction Assays
- 3.3.1 Strategic Considerations
- 3.3.2 Library Screen 1
- 3.3.3 Library Screen 2
- 3.3.4 Library Screen 3
- 3.3.5 Summary of Library Transformations
- 3.3.6 Summary of Clones Obtained from Library Screens

3.4 Analysis of Results
- 3.4.1 False Positives
- 3.4.2 SRp20
- 3.4.3 MCM6
- 3.4.4 A7CO8
- 3.4.5 Other ESTs

3.5 Ranking of Clones in Order of Priority for Further Characterisation

3.6 Further Yeast Analysis
- 3.6.1 Y190
- 3.6.2 Y166
- 3.6.3 PJ69-4A

3.7 tBMI-1 and Mph1
- 3.7.1 Comparison of Truncated BMI-1 and Full-length Bmi-1 Interactions

3.8 Further Characterisation of Protein Interactions between BMI-1 and the Potential Interactors
Chapter 4 Discussion

4.1 Brief Summary of Results

4.2 Validity of the Positives as Real Interactors of BMI-1
  4.2.1 Known Bmi-1 Interacting Proteins
  4.2.2 fBmi-1 and Mph1 Interactions
  4.2.3 The Role of Electrostatic interactions?
  4.2.4 Domain Swap experiments
  4.2.5 Correlation Between Interaction Strength and Biological Significance
  4.2.6 Clone 16 (SRp20) as a Candidate Interactor of BMI-1
List of Tables

Table 1  List of Primers and their uses  51
Table 2  Recipes for SDS-PAGE separating gels  63
Table 3  Summary of yeast two-hybrid autoactivation analysis  77
Table 4  Summary of yeast two-hybrid library screen transformation results  85
Table 5  Summary of yeast two-hybrid library screen interactors  86
Table 6  Summary of yeast two-hybrid retransformation experiments  94
Table 7  Summary of GST-affinity capture assay results  113

List of Figures

Figure 1  Overview of the regulatory connections involved in antero-posterior specification in Drosophila melanogaster.  5
Figure 2  Simplified diagram illustrating the repressor model of homeotic control.  7
Figure 3  The heterochromatin model of Polycomb-Group protein action. 15
Figure 4  The looping model of Polycomb-Group protein action.  16
Figure 5  Expanded overview of the regulatory connections involved in anteroposterior specification in Drosophila melanogaster. 21
Figure 6  The expression of Pc-G proteins during the development of hematopoietic cells. 23
Figure 7  Representation of the human BMI-1 protein. 27
Figure 8  BLAST comparison of human BMI-1 against posterior sex combs, and suppressor 2 of zeste, at the amino-acid level. 28
Figure 9  The underlying principle of the yeast two-hybrid system. 32
Figure 10  Representation of the Human BMI-1 protein illustrating the EcoRI fragment used in yeast two-hybrid analysis. 75
Figure 11  Photograph of yeast (Y190) autoactivation analysis. 78
Figure 12  BLAST comparison of murine Bmi-1 against human BMI-1 at the amino-acid level. 82
Figure 13  Analysis of (mouse embryo) library screen 3 positives. 84
Figure 14  Alignment of the 5' ends of the different SRp20-encoding fragments. 90
Figure 15  Comparison of DNA and protein sequences of the 3' region of human MCM6, murine mis5, clone 54 and v45. 91
Figure 16  Photograph illustrating retransformation results in the yeast strain Y190. 95
Figure 17  Photographs illustrating retransformation results in the yeast strain PJ69-4A. 97
Figure 18  Liquid β-Galactosidase assay results for the yeast strain, PJ69-4A. 98
Figure 19  Photographs illustrating retransformation results comparing fBmi-1 and tBMI-1 interactions in the yeast strain PJ69-4A. 101
Figure 20  Liquid β-Galactosidase assay results comparing fBmi-1 and tBMI-1 interactions for the yeast strain, PJ69-4A. 102
Figure 21  The principle underlying the GST-affinity capture method for...
determining protein-protein interactions.

**Figure 22** SDS-PAGE analysis for GST-agarose purification of different pGEX-BMI-1 constructs.

**Figure 23** SDS-PAGE comparison of different methods of synthesising radiolabelled BMI-1.

**Figure 24** SDS-PAGE analysis for the GST-Affinity capture of $^{35}$S-methionine labelled BMI-1 by GST fusion proteins.

**Figure 25** SDS-PAGE analysis for the GST-Affinity capture of $^{35}$S-methionine labelled VP16 and VP16-BMI-1 by GST-fusion proteins.

**Figure 26** The principle underlying the (co-)immunoprecipitation method for analysing protein-protein interactions.

**Figure 27** Simple representation of BMI-1 with the peptides to which antibodies were raised shown.

**Figure 28** Western analysis of peptide-raised polyclonal antibodies (43583 and 43584) and the characterised F6 $\alpha$BMI-1 polyclonal antibody.

**Figure 29** Simple representation of MCM6 with the peptides to which antibodies were raised shown.

**Figure 30** Western analysis of peptide-raised polyclonal antibodies (83244 and 83245) and the characterised $\alpha$MCM6 polyclonal antibody.

**Figure 31** Western blot of 83244 anti-MCM6 antibody tested against *in vitro* translated MCM6.

**Figure 32** Western analysis showing the effect of blocking 83244 with peptide on the detection of MCM6.

**Figure 33** Comparison of different lipid-based methods of transfecting U2OS cells with pCMV-β-Gal.

**Figure 34** Western analysis of U2OS cells transfected with FLAG-25/MCM6 Expression Constructs.

**Figure 35** Western blot of positive control co-immunoprecipitation analysis of U2OS cells transfected with HA-BMI-1 and *myc*-RING1.

**Figure 36** Western blot of test co-immunoprecipitation analysis of U2OS cells transfected with HA-BMI-1 and either FLAG-25 or FLAG-MCM6.

**Figure 37** NIX output for BMI-1 yeast two-hybrid positive clone 25.

**Figure 38** BLAST output for BMI-1 yeast two-hybrid positive clone 25.

**Figure 39** THC175068 (from the TIGR database).

**Figure 40** Northern blot analysis of clone 25.

**Figure 41** Alignment of 5-prime ends of Human Fetal Brain library clones (HFB2, -5, and -6), THC175068 and clone 25.

**Figure 42** Summary of DNA sequences relating to BMI-1 yeast two-hybrid clone 25.

**Figure 43** Predicted protein sequence of BMI-1 yeast two-hybrid positive 25.

**Figure 44** Predicted potential Open Reading Frames encoded by BMI-1 yeast two-hybrid positive 25.

**Figure 45** Coomassie stained SDS-PAGE gel of GST and GST-25.
Figure 46  SDS-PAGE analysis of *in vitro* translation and immunoprecipitation of clone 25 constructs.

Figure 47  Amino acid composition for the predicted protein encoded by clone 25.

Figure 48  Crude intracellular localisation of FLAG-25 and FLAG-Pax3 in transfected U2OS cells.

Figure 49  Model of human Pc-G multimeric protein complexes.

Figure 50  Comparison of the Mphl interacting region of murine Bmi-1 and the fragment of human BMI-1 used as bait in yeast two-hybrid library screens.

Figure 51  Different conformations of BMI-1 could effect different protein interactions.

Figure 52  *GAL4* Domain Swap experiments alter two-hybrid BMI-1 self-association.

Figure 53  A model of DNA replication complexes during the Cell Cycle.

Figure 54  The principle of the scintillation proximity assay.
Chapter 1

Introduction

1.1 The Role of *Drosophila* in the Understanding of Developmental Pathways

1.1.1 *Introduction to Drosophila melanogaster*

The development of an adult organism from a single fertilised egg requires both the determination of many cell types and the organisation of these cells into an elaborate pattern. Development follows a programmed pathway, in which specific genes are turned on and off in a spatially and temporally restricted manner. The consequences of an error in this process can vary, from those errors where there is no obvious phenotypic difference, to those where there is a viable but noticeable phenotype, to those where the prognosis is lethality. The latter events are rare; and since they inevitably reduce the ability of the organism to perpetuate itself, they are selected against during evolution.

The occurrence of developmental errors is part of the genetic load carried as the price of maintaining the ability of the genetic apparatus to evolve by offering new, favourable variants for evolution. It is the existence of aberrations - genetic or epigenetic - in the developmental process that offers insight into the network of interactions that controls it: the identification of a developmental mutation opens the possibility of investigating the functions performed by the normal, wild-type allele.

It is in this regard that *Drosophila melanogaster* has been the model of choice in which to investigate developmental pathways. Not only does it have the advantage in that it has a very short developmental time-course (two weeks), it is also easy to manipulate in comparison to other model organisms, and the ethics and problems involved in the manipulation of higher organisms, e.g. genetically altering mice, and the collection of human tissues and samples, do not apply to *Drosophila*. For *Drosophila*, methods were developed which allowed the isolation of a comprehensive collection of late embryonic lethal mutations which affected the body plan (*Nüsslein-Volhard & Wieschaus, 1980; Nüsslein-Volhard et al., 1984*). Through this, and the advancement of molecular biology, new tools have been developed to allow genetically and embryonically important genes to be isolated.
1.1.2 Drosophila Developmental Genetics

*Drosophila melanogaster* is built up of an anteroposterior sequence of segments which fall into the three principle body regions of head, thorax and abdomen. The basic body plan is not specified initially, but is formed as a result of a hierarchy of temporal and spatial developmental decisions starting before fertilisation at oogenesis.

From examination of coordinate gene mutations, two independent systems have been found to exist, with one forming the dorsoventral axis and the other specifying the anteroposterior axis. For this project, only the anteroposterior axis will be described in detail.

1.1.3 The Establishment of the Anteroposterior Axis

The specification of the anteroposterior axis occurs in the egg chamber through the action of three pattern-forming systems with individual regions largely being determined independently. In this manner, the *anterior* system is responsible for the segmented region of head and thorax, the *posterior* system determines the segmented abdomen, and a third system, the *terminal* system determines the non-segmented acron and telson (*Nüsslein-Volhard et al., 1987*).

The *anterior* system determinant is a single sequence-specific DNA-binding gene, *bicoid* (*bcd*). *bcd* mRNA is transcribed during oogenesis and localised at the anterior egg pole. Following fertilisation, the mRNA is transcribed to produce *bcd* protein, which diffuses away from the source to set up an anterior-to-posterior *bcd* protein gradient (*Driever and Nüsslein-Volhard, 1988*). Through the use of promoters with a differing affinity to *bcd*, the gradient can be translated into discrete domains of gene expression (*Driever et al., 1989*), and indeed, the function of *bcd* has been shown to be to regulate the zygotic expression of the gap genes *hunchback* (*hb*) (*Struhl et al., 1989; Driever and Nüsslein-Volhard, 1989*) and *kruppel* into discrete domains.

The *posterior* system is a little more complex. During oogenesis, mRNA encoding the maternal effect gene *nanos* (*nos*) is deposited in the pole plasm of the egg (*Lehmann & Nüsslein-Volhard, 1991*). In contrast with the anterior system, this gene product does not regulate zygotic gene expression directly. Instead, *nos* protein is thought to diffuse anteriorly to where it is required for the transcriptional activation of the gap gene *knirps* through the elimination of a transcriptional repressor of *knirps*, thus establishing a
posterior expression domain for knirps. The repressor of knirps was identified as maternally derived hunchback (Tautz, 1988; Tautz & Pfeifle, 1989). The rationale for this double negative control is not obvious, and indeed, it has been shown that double mutant embryos hunchback+/nanos− are phenotypically normal (Hülskamp et al., 1989; Irish et al., 1989)).

The terminal system determinant is somewhat different from the other two in that the determinant is not of germ line origin, instead coming from the follicle cells surrounding the egg. The product of the gene torso (Sprenger et al., 1989) acts as a receptor for an extracellular signal that is produced at the two poles of the egg. Activation of the torso gene product at the egg poles results in a signal transduction chain that leads to a positive control of transcription of the zygotic target genes huckebein and tailless (Weigel et al., 1990). As activation of the torso ligand is restricted to the termini, transcription of these two genes is also restricted resulting in localisation of these proteins to the poles.

It can be seen from the above examples from each of the three systems that through the action of maternally derived gradients, a state has been achieved whereby these gradients have been transformed into discrete spatial domains of non-maternal zygotic gap gene expression. These domains provide a course spatial prepatterning which is then further refined by the action and interaction of zygotic pattern genes.

1.1.4 Pair Rule Genes

The next level of patterning genes are the pair-rule genes, so called because mutations in pair rule genes cause deletions in every other segment. Two types of pair rule genes exist: the primary pair-rule genes e.g. hairy, runt and even-skipped; and the secondary pair-rule genes e.g. fushi tarazu.

Periodic patterns of primary pair rule gene expression are set up under the regulation of the gap and maternal genes, and these primary pair rule patterns are then further refined through their actions on secondary pair rule genes. During cellularisation, the pair rule stripes are formed by a combination of events: accumulation of products within the stripes and loss of pair rule products in the interstripes. Initially the products of most pair rule genes accumulate throughout the metameric region and it is thought that the
individual stripes and interstripes in the normal pattern are regulated independently of each other through complex processes. A common feature of the interstripes however, is a combination of transcriptional repression coupled with degradation of gene product.

1.1.5 Segment Polarity Genes

The segment polarity genes are responsible for maintaining certain repeated structures within each segment. Mutations in this group of genes cause a portion of each segment to be deleted and replaced by a mirror-image structure of another portion of the segment. For example, the engrailed gene is needed for the maintenance of the anterior-posterior boundary between segments.

1.1.6 Homeotic Selector Genes

After the segmental boundaries have been established, the characteristic structures of each segment are specified. As well as acting on the segment polarity genes, the maternal, gap and pair rule gene products all combine to activate the homeotic selector genes. Homeotic genes are defined as genes which when mutated cause transformations of one part of the body into another. For example, recessive mutations in the Antennapedia gene result in the transformation of mesothoracic leg into antennae (Struhl, 1981a) while dominant mutations in the Antennapedia gene has the opposite phenotype with mesothoracic leg growing instead of antenna (Schneuwly et al., 1987). While these segmentation genes are responsible for defining the number and locations of segments, it is the homeotic genes that impose the program determining the unique differential state of each segment. Many homeotic genes code for transcription factors, which as well as acting upon other homeotic genes, act on other target loci which determine the anatomy of the segment. However, the relationship between homeotic selector gene expression and terminal differentiation is one of the least well-understood processes of Drosophila development.

Most of the homeotic genes of Drosophila belong to two gene clusters called the Antennapedia complex (ANT-C) (Scott et al., 1983) and the Bithorax complex (BX-C) (Beachy et al., 1985) which both lie on the right arm of the third chromosome.
Maternal Genes

↓

Gap Genes

Primary Pair Rule Genes

Homeotic Genes

Secondary Pair Rule Genes

en and Segment Polarity Genes

Figure 1 – Overview of the regulatory connections involved in anteroposterior specification in Drosophila melanogaster.
1.2 Polycomb

1.2.1 The Identification of Polycomb

Through in-depth studies on the effect of gain- and loss-of-function mutants of the *Drosophila* BX-C on the transformation of body segments a model was developed to understand better the role of homeobox genes in development. In the model it was proposed that there is an anteroposterior gradient within the organism of BX-C gene action (for a review see Lewis, 1978). Several rules were observed based on this model:

1) the state of repression or activation of a given gene is controlled by *cis*-regulatory elements;
2) the genes tend to be individually, rather than co-ordinately, derepressed;
3) a gene derepressed in one segment is derepressed in all segments posterior thereto;
4) the more posterior the segment the greater the number of BX-C genes that are in the derepressed state; and
5) the more proximal the locus of a gene in the complex the more likely it is that the gene is in a derepressed state.

Rules 3), 4) and 5) suggest that two types of gradient are involved in BX-C regulation: an increasing anteroposterior gradient of BX-C gene products: and an increasing posterior-anterior gradient of a molecule that encodes a *cis*-regulatory repressor molecule (*Figure 2*).

The properties of *Polycomb* (*Pc*) (*Lindsley & Grell, 1968*), and of a more extreme allele *Pc*<sup>3</sup> suggested that *Pc* could act as such a repressor molecule (*Lewis, 1978*). While mutations in genes within the BX-C cause anterior transformations of body segments, mutations in the repressor would be expected to show an opposite phenotype, which was indeed observed, e.g. in *Drosophila* carrying mutations of the *Pc* gene, thoracic and upper abdominal segments underwent a posterior transformation towards the eighth abdominal segment. It was noted, however, that this transformation could only occur if at least one copy of BX-C is present indicating that BX-C is indeed the target of the *Pc* gene product.
1.2.2 Other Polycomb Genes in Drosophila

Further large-scale screens for embryonic-lethal mutations affecting the *Drosophila* body pattern yielded zygotic mutations in loci other than the *Pc* locus previously identified (Nüsslein-Volhard et al., 1984, Jürgens et al., 1984, Weischaus et al., 1984). These mutations resembled weak *Pc* mutations in both their dominant adult and recessive embryonic phenotype. In contrast to *Pc*, however, which causes posterior transformation of all body segments, an incomplete partial homeotic transformation is associated with several of the identified loci, e.g. *Polycomb-like* (*Pcl*) and *Posterior sex combs* (*Psc*) (Jürgens, 1985). However, embryos mutant for two or more of these genes show a much stronger transformation of all body segments similar to, or stronger than, that seen in *Pc* embryos, indicating that these genes act synergistically in normal development to control the spatial expression of the BX-C genes (Jürgens, 1985). On the basis of the mutant phenotypes, these genes collectively are referred to as Polycomb-group (*Pc-G*) proteins after its first known member.

The strong embryonic transformations caused by the *Pc-G* double mutants have been used as an assay for the occurrence of additional *Pc-G* genes, with evidence suggesting...
a figure of ∼40 Pc-G genes being encoded for within the *Drosophila* genome (Jürgens, 1985).

The reason why there may be so many Pc-G genes is not known, although an evolutionary hypothesis has been proposed. If the segment patterning observed within *Drosophila* were to be thought of as the result of evolutionary modification of a single repeated unit, slight genetic changes in *trans*-regulatory genes controlling the spatial expression of homeotic genes might gradually have built up the necessary genetic isolation between segments to ensure largely independent evolution of segment-specific features (Jürgens, 1985).

### 1.2.3 Polycomb and Chromatin Silencing

One of the fundamental features of the Pc-G complexes in *Drosophila* is that they preserve the state of repression for a given gene through many generations. This implies that early events alter the chromatin state in a heritable way determining either the open, active state, or the repressed, inactive state in subsequent cellular generations. Pc-G products are necessary not, it seems, for the establishment of the transcriptional states but only for the continuation of these states (with the exception of *extra sex combs*, (esc)).

Silencing by the Pc-G proteins is thought to be mediated by Polycomb-response elements (PREs), regulatory regions of several hundred nucleotides that are *in vivo* binding sites for Pc-G proteins (*Simon et al., 1993, Chan et al., 1994, Christen & Bienz 1994, and Chiang et al., 1995*). No distinct consensus sequences have been identified within the PREs, and until recently none of the known Pc-G proteins had been shown to bind DNA *in vitro*. Since then, a sequence-specific Pc-G DNA binding protein has been identified that interacts with a PRE from the *Drosophila engrailed* gene. This protein, *pleiohomeotic (pho)* is a homologue of the ubiquitous mammalian transcription factor Yin Yang-1 and it has been proposed that *pho* acts to anchor Pc-G protein complexes to DNA (*Brown et al., 1998*).

There are several other lines of evidence that exist demonstrating that the Pc-G complexes are associated with chromatin. Perhaps the most convincing is the observation that the *Pc* protein shares a homologous protein motif (the chromodomain) with the *Drosophila* heterochromatin-binding protein *HP1* encoded by the *Su(var)2-5* gene (*Paro & Hogness, 1991*). Mutations in the conserved chromodomain abolish the binding of *Pc*
to the target genes (Messmer et al., 1992). In addition, the chimeric protein produced by substitution of the Pc chromodomain into HP1 binds not only to heterochromatin but also Polycomb binding sites (Platero et al., 1995). Pc alone is sufficient to recruit other Pc-G proteins as shown by an experiment where Pc was fused with a GAL4 DNA-binding domain and targeted to a reporter gene containing a GAL4-responsive UAS (Müller, 1995). The reporter was subsequently silenced, presumably through the functional Pc-G complex that was recruited to the UAS. It may also be the case that somehow proteins interacting together in a complex together are able to generate a DNA-binding potential which as monomers they do not possess.

Further evidence for a role in chromatin regulation comes from immunologic studies with antibodies against several polycomb proteins. The protein products of the Pc, Ph and Psc Pc-G genes have been shown to co-localise at approximately 100 salivary gland polytene chromosomal loci (identified as PREs) (Zink et al., 1991, DeCamillis et al., 1992, and Martin & Adler, 1993), though not all Pc-G proteins have been found at all of the sites, indicating the possibility of different Pc-G complexes existing. In addition, a biochemical association has been identified with Pc and Ph being identified as constituents of a large multimeric protein complex containing somewhere between 10 and 15 different proteins (Franke et al., 1992).

1.3 Conservation in Mammals

1.3.1 Vertebrate Hox Genes

An important discovery in understanding homeotic gene function was made when it was found that a DNA probe corresponding to the homeobox is able to hybridise to multiple locations within the genomes of many species (McGinnis et al., 1984a, 1984b). This discovery that the homeobox motif is conserved throughout evolution has led to a situation where several of the lessons learnt from Drosophila were extrapolated to higher organisms. In all mammals studied to date it has been found that there exists a cluster of homeobox genes known as HOX (as opposed to the HOM-C in Drosophila). The mammalian cluster appears to have duplicated twice or three times to generate four copies of the cluster (Hox a-d) and there are 14 identifiable types of genes in the Hox complexes, called paralogues 1-13 and Evx. Parologue grouping is based on
homeodomain sequence similarities as well as position within the cluster. Studies on Hox gain and loss of function mouse mutants have demonstrated that Hox genes, like their *Drosophila* counterparts, act as key selector genes directing regional specification in the embryo, as well as assuming other roles in development (reviewed in Manak & Scott, 1994). The Hox genes are expressed in limited regions along the AP-axis, like the fly genes, although in more substantially overlapping domains. It is also often the case that corresponding paralogues in different clusters, in particular Hox-a and -d are often expressed in similar domains, possibly indicating a level of functional redundancy between clusters.

In a situation similar to that in the *Drosophila* HOM-C, several rules exist: genes are ordered along the chromosome in the same order as their expression and function along the anteroposterior axis, so called “colinearity”. Generally, more genes are expressed in more posterior regions; Loss of function leads to loss of structures or to the development of anterior structures where more posterior structures should be; Gain of gene function leads to posterior structures developing where more anterior structures would normally be found.

### 1.3.2 Vertebrate Polycomb Genes

With an apparent conservation of function of Hox genes between such diverse species as *Drosophila* and mouse, it was proposed that a similar mechanism might exist to regulate the expression of the Hox genes. While it appears that the mechanisms responsible for the initiation of Hox gene function differ radically between species, the mechanism responsible for the maintenance of the state of Hox gene expression shows several signs of being far more highly conserved.

One such example concerns the gene, *Pc*. In 1992, a murine homologue, *M33*, was identified through its similarity to the *Pc* chromodomain (Pearce et al., 1992). In addition to the chromodomain, *M33* also contains a C-terminal stretch of about 30 amino acids that is well conserved in *Pc* and is thought to recruit further Pc-G members to form a silencing complex (Franke et al., 1995, Müller et al., 1995). Müller et al. also showed that the *M33* protein is able to partially substitute for *Pc* in transgenic flies deleted for *Pc*. They proposed that if the fly *Pc* chromodomain recognises proteins on the DNA, it would follow that the mouse homologue can bind to these *Drosophila* proteins, suggesting that they have been conserved between mice and flies. In addition, if the C-ter
domain is there to recruit other proteins to form a multimeric complex, then the proteins
that are recruited may also be conserved.

A second case of conservation involves the two Pc-G members Su(z)2 and Posterior
sex combs (Psc). Proteins encoded by these two genes share similar regions with the
Bmi-I mammalian proto-oncogene (Brunk et al., 1991, and van Lohuizen et al., 1991b).
The regions of similarity are dispersed throughout much of the proteins, suggesting
conservation of a large domain or multiple domains (see later).

To test whether Bmi-I could indeed function as a mammalian Pc-G member, mutant
mice were generated that either over- or underexpress Bmi-I. Targeted deletion of Bmi-
I results in a posterior homeotic-like transformation along the complete anteroposterior
axis of the embryo (van der Lugt et al., 1994) as well as other defects. In contrast,
transgenic mice overexpressing Bmi-I show the opposite phenotype, that of a dose
dependent anterior transformation (Alkema et al., 1995). When Hox gene expression is
analysed, it was found that Bmi-I+ mice exhibit a slight anterior shift in the expression of
homeotic genes, and conversely, mice overexpressing Bmi-I show the opposite
phenotype, i.e. a slight posterior shift in the expression boundary of homeotic genes (van
der Lugt et al., 1996), indicating that in both cases the phenotypes observed were likely
to be a result of deregulation of Hox gene expression during embryonic development.

The incomplete posterior transformation seen with Bmi-I null mutants is similar to the
phenotype observed in Drosophila mutant embryos where only a partial posterior
transformation is observed (Jurgens, 1985, and Adler et al., 1989, 1991). This is in
contrast to some of the Drosophila Pc-G mutants where all body segments are
transformed into the most posterior abdominal segments (e.g. Pc null mutants). This
suggests that there may exist “weak” and “strong” Pc-G genes, where mutants in two or
more weak Pc-G genes would show a strong homeotic transformation of all body
segments similar to that seen in Pc-G embryos. This predicts that these genes would act
synergistically. The fact that only a subset of Hox genes belonging to different clusters
are altered by Bmi-I mutant mice hints that different Hox genes are under the control of
different Pc-G homologues. Multimeric Pc-G complexes of differing composition may
form, each having the ability to repress specific, or possibly overlapping, subsets of Hox
genes. In support of this idea, several Pc-G proteins have been shown to have an
overlapping subnuclear speckled distribution and the existence of a mammalian
multimeric Pc-G complex (comprising Bmi-1, M33, Mphl and other, as yet unidentified proteins) has been demonstrated (Alkema et al., 1997a and 1997b). This raises the possibility that the mechanism of action of the mammalian complexes is similar to that of Drosophila. The existence of a multimeric complex could also be used to explain the synergy seen between mutants. If such a complex existed, then a model of effect by mass action could be postulated.

Since the initial descriptions of mammalian homologues, several other mammalian polycomb proteins have been identified. Of interest is the fact that some of the murine homologues of Drosophila Pc-G members exist in pairs e.g. M33 and MPc2 (Pc), Bmi-1 and Mel-18 (Psc), Mph1 and Mph2 (polyhomeotic, Ph) and Enx1 and Enx2 (enhancer of zeste, [E(z)]). Several of these proteins have also been found to exist in large complexes. For example, through studying the mammalian homologue of esc (Eed), two proteins, Enx1 and −2 were identified as part of an Eed complex. One notable observation was that no direct biochemical evidence was found to suggest an interaction between Eed/Enx and the M33/Bmi-1/Mph1 complex identified previously implying that different Pc-G complexes with different functions may exist (Sewalt et al., 1998, and van Lohuizen et al., 1998).

1.4 Mechanistic Aspects of Polycomb Action

1.4.1 Initiation of Polycomb-Group Protein Activity

Once the gap genes and pair rule genes have established a pattern of Hox gene activity along the AP axis of the Drosophila embryo (at ~2hr of embryogenesis), another system must set the mechanism of repression in action. Analysis of Polycomb-group mutants reveals that homeotic expression is established correctly, but subsequently expression spreads beyond normal boundaries, indicating that Pc-G proteins are not involved in the establishment of Hox gene expression zones (Struhl & Akam, 1985, Wedeen et al., 1986, Simon et al., 1992, Soto et al., 1995). At ~4hours of development, the gap and pair-rule proteins decay, while the Hox genes somehow maintain their expression. It is at this point that the Pc-G and the Trx-G proteins first become essential to the stable maintenance of gene expression through subsequent cell divisions. Even though the Polycomb-group gene products are present throughout the embryo, the Pc-G complexes...
must form, or at least function to repress the target genes only in embryonic regions where specific Hox genes are not expressed. A protein of interest in this regard is *extra sex combs* (*esc*) ([Struhl, 1981b](#)). Temperature-sensitive allele studies have shown that *esc* activity is required during the first 3-6 hours of embryogenesis ([Simon et al., 1995](#)). Absence of *esc* during this short period results in homeotic transformations similar to other Pc-G mutations. However, when *esc* is absent during earlier or, more importantly, later phases of development, no phenotypic defects are observed ([Struhl & Brower, 1982](#) and [Simon et al., 1995](#)). This suggests that *esc* may perform a different role than that of other Pc-G proteins, perhaps by acting as a bridge between early-acting gap proteins and later-acting Pc-G proteins. In relation to this, *esc* has been shown to bind to another Polycomb-Group protein, *E(z)* ([Tie et al., 1998](#)). *E(z)* is thought to be required for binding of other Polycomb-Group proteins to chromosomes, and so this suggests that the formation of an *E(z):esc* complex at Polycomb Response Elements may be an essential prerequisite for the establishment of silencing.

A more direct role for Pc-G genes has been suggested for the regulation of some segmentation genes, with the observation that mutations in several Pc-G genes can cause maternal-effect or zygotic segmentation defects ([McKeon et al., 1994](#) and [Pelegri & Lehmann, 1994](#)). With Pc-G genes being thought to control gene expression by regulating chromatin, a model has been suggested that imprinting at the chromatin level underlies the determination of anteroposterior polarity in the early embryo.

### 1.4.2 Mechanisms of Inactivation

One of the mysteries of Pc-G action is related to how the Pc-G group proteins mediate their repressive functions. Originally, through gene transfer studies using PREs to silence reporter genes, a perspective was adopted that a scenario analogous to Position Effect Variegation (PEV) might exist. PEV is a phenomenon where a gene is translocated from its wild type position in euchromatin to a new location close to heterochromatin. Even though the gene itself is unchanged, it is inactivated through the influence of its surroundings by a process of *cis* and *trans* spreading of a heterochromatin-like structure.

In relation to Pc-G action, the Pc-G complexes localised to PREs would similarly spread to flanking sequences to silence a large chromatin domain encompassing the Hox gene ([Figure 3](#)). Evidence for this model comes from the fact that as well as some Pc-G and Su(var) proteins having common motifs like the chromodomain and SET domain,
mutations in some Pc-G genes have also been shown to be modifiers of PEV (Enhancer of Polycomb, [E(Pc)](Sinclair et al., 1998), and Enhancer of zeste [E(z)] (Laible et al., 1997)), similar to the Su(var) phenotype seen in mutations of the chromodomain-containing HP1 gene (Eissenberg et al., 1990).

Recently, however, support for the heterochromatin model of Pc-G silencing has diminished. Domains initially thought to be shared between Su(var)s and Pc-G proteins, like the SET domain are also found to occur in activators of transcription (e.g. the SET domain occurs in both Pc-G proteins and trithorax-group proteins). Most Pc-G mutations are not Su(var)s and vice versa (Jenuwein et al., 1998). Improvements in cross-linking techniques have revealed that Pc-G proteins do not spread to coat the gene, as originally thought, and are instead linked primarily to the vicinity of known PREs, decreasing to nearly background levels within one or two kilobases (Strutt et al., 1997).

In addition, a compacted heterochromatin state associated with PEV would be expected to demonstrate a greatly reduced accessibility to transcription factors. However, studies with RNA polymerases as probes for accessibility showed that accessibility still remains potentially discounting the heterochromatin theory (McCall & Bender, 1996).

Bearing this in mind it is interesting to find that recently the nature of PEV itself has been brought into question. The conventional model of heterochromatin as a homogeneously condensed chromatin state that is inaccessible to the transcriptional machinery has been challenged with the discovery that heterochromatin and the proteins that bind it can actually promote, and in some cases are necessary for, the expression of certain types of genes (reviewed in Weiler & Wakimoto, 1995). As yet the functional implications of this regarding Pc-G protein action are unknown.
Figure 3 – The heterochromatin model of Polycomb-Group protein action. Pc-G complexes are recruited to PREs and spread to flanking sequences to silence a large chromatin domain encompassing the target gene.
If heterochromatinisation is not the method of action, then what is? Despite the localisation of Pc-G proteins to regions near the PREs, Pc-G complexes formed at a PRE can affect enhancers or promoters over distances of 20-30Kb. Transcriptional enhancers act at a distance by a mechanism of looping, bringing the enhancer-activator complex into close proximity with the promoter complex, and a similar model could be proposed for Pc proteins. Formation of a core complex at a PRE could then interact with, and stabilise, weaker transient complexes formed at frequently occurring but weak proto-PREs lying along the path between enhancer and promoter (Pirrotta, 1995) (Figure 4). It should be noted that this model does not provide an account for the actual method of silencing.

![Diagram](image)

**Figure 4** – *The looping model of Polycomb-Group protein action*, as proposed by Pirrotta, (1995).

An alternative mechanism might be provided by the fact that regions of the genome can be targeted to areas within the nucleus which are transcriptionally inactive. An example of this is provided by the HM mating-type loci and telomeric regions of the yeast *Saccharomyces cerevisiae* (Andrulis et al., 1998). Until recently it was unclear whether localisation to the nuclear periphery contributed to the formation of
heterochromatin, or was as a consequence of it. Andrulis et al. suggest that transcriptional silencing of the HM mating-type loci and telomeric regions occurs as a consequence of localisation towards the nuclear periphery. This was demonstrated by fusing integral membrane proteins to the GAL4 DNA-binding domain- hybrid and overexpressing the hybrid protein causing it to accumulate in the nuclear membrane. The hybrid proteins were expressed in a yeast strain carrying a HMR silencer with GAL4-binding sites replacing normal silencer elements, causing the silencer to become anchored to the nuclear periphery and leading to silencing of a nearby reporter gene.

By studying the subcellular three-dimensional distribution of three Polycomb-group (Pc-G) proteins - Pc, Ph and Psc - in fixed whole-mount Drosophila embryos it was found that all three proteins were localised in complex patterns of 100 or more loci throughout most of the interphase nuclear volume (Buchenau et al., 1998). This argues against a Pc-G-mediated sequestration of repressed target genes by aggregation into subnuclear domains. Similar results were found in mammalian cells where so-called Pc-G-bodies were observed in both transformed and primary cell lines (Saurin et al., 1998). However, it was found that Pc-G bodies vary in both size and number dependent on cell type, though any functional significance of such a varied Pc-G distribution is not known. In addition, in the Drosophila studies, there was a lack of correlation between the occurrence of Pc-G proteins and high concentrations of DNA, suggesting that the silenced genes are not targeted to heterochromatic regions within the nucleus (Buchenau et al., 1998).

One final possible mechanism has recently been suggested by Sewalt et al. (1999) who found that the Xenopus Pc and human Pc2 are capable of binding a protein of unknown function called the C-terminal Binding Protein (CtBP). CtBP has also been shown in other studies to bind the Histone Deacetylase, HDAC1, suggesting a promoter targeted CtBP-HDAC1 complex can silence transcription possibly through a deacetylation mechanism (Sundqvist et al., 1998).

Ultimately, it may well be the case that silencing is not dependent solely on one of the proposed mechanisms. It might be possible to rationalise several of the hypotheses and observations into one overall hypothesis which draws on aspects of many different cellular proteins and processes.
1.4.3 Polycomb-Group Proteins and Genetic Memory

A characteristic feature of Pc-G complexes is their self-replicating nature. The complex is not simply reconstituted *de novo* every cell cycle since the reassembly process must be able to discriminate between those sites that were inactive and those that were active in the previous generation. A “memory” system must exist to allow for this discrimination to be correctly performed. The role of DNA methylation has been discounted since DNA methylation is not thought to occur in *Drosophila*.

*Saurin et al. (1998)* observed that several Pc-G components remain chromatin-associated near centromeric regions throughout all stages of mitosis and, that around the prometaphase/metaphase stage of the mitotic cycle, the Pc-G complex appears to ‘redistribute’ so that each chromatid contains Pc-G complex. They suggest that by maintaining chromosome-associated Pc-G complexes during mitosis, the proteins constituting Pc-G bodies are ensured to be inherited by successive cell generations, thus providing the daughter cells with the necessary components for maintaining gene expression patterns. It is likely that Pc-G proteins are necessary during mitosis, in order to distinguish, in G1 phase following decondensation of the chromosome at the end of mitosis, the chromatin that is to remain condensed from the euchromatin. However, it may be the case that the more important point at which faithful inheritance of Pc-G proteins is determined occurs at a much earlier stage of the cell cycle. The most obvious point for the inheritance of markers would be at the time of DNA replication.

1.4.4 Polycomb-Group Proteins and Kinetochores – Functional Relevance?

As just mentioned, *Saurin et al., (1998)*, with their studies on the localisation of RING1, BMI-1 and hPc2 found that in three mammalian cell lines there was a direct heterochromatin association of the Pc-G complex in interphase to pericentromeric DNA sequences on human chromosome 1 and related pericentromeric sequences on other chromosomes. Similar results have also been obtained studying the *Psc* protein in *Drosophila (Buchenau et al, 1998)*.

One interesting possibility is that the Pc-G proteins might be involved in the maintenance of an epigenetic state that permits kinetochore activity. By studying a range of chromosomal translocations, it was found that chromosomes that contain two centromeres only have one active kinetochore (*Warburton et al., 1997*) (the kinetochore
is a complex of centromeric DNA and protein that links the chromosomes to spindle microtubules). Whilst the initial selection of which centromere to use is apparently random, once established, the position of the active kinetochore can propagate accurately through many cell generations, suggesting the existence of a type of memory system (for a review on epigenetic effects in kinetochore assembly see Wiens & Sorger, 1998). In addition, it seems likely that the location and function of kinetochores is determined by the position of centromeric chromatin. For example, in *S. pombe*, mutations in proteins required for chromatin-dependent gene silencing dramatically increase the rate of chromosome loss (Ekwall et al., 1997). These two facts, together with the association of Pc-G proteins to pericentromeric regions could suggest the existence of a Pc-G-based epigenetic system allowing for the propagation of centromeric heterochromatin and thus kinetochore positioning.

### 1.4.5 Polycomb and Trithorax

By performing similar analyses to those which led to the initial identification of the Pc-G class of proteins, another class of proteins was found which, in contrast to the Pc-G phenotype, gave a reciprocal phenotype of reduced expression of the homeotic genes (e.g. Ingham, 1985). This group, the *trithorax* group (Trx-G) is responsible for sustaining the active state of homeotic gene expression. While some Trx-G proteins may function as specific ‘anti-repressors’, directly modulating Pc-G-mediated repression (Kennison & Tamkun, 1988), others are likely to be involved in more general aspects of transcriptional activation, for example the mammalian Trx-G homologues which are members of chromatin opening complexes, e.g. SWI/SNF (Tamkun et al., 1992) and GAGA/NURF (Farkas et al., 1994).

Despite apparently having an opposite effect to each other, it appears that at a mechanistic level, the two groups of proteins are closely related. In *Drosophila*, the *trithorax* gene product (*trx*) colocalises on polytene chromosomes at Trx-G response elements (TREs) along with, at many sites, *Drosophila* Pc-G proteins. Furthermore, binding of the *trx* protein has been mapped to small DNA fragments that also contain binding sites for the Pc-G proteins, PREs (Chang et al., 1995, Chinwalla et al., 1995, and Gindhart & Kaufman, 1995). These studies eliminate the simple view that the expression status of a given gene is as a result of mutually competing activating and repressing complexes, since binding alone appears not to be a hallmark of function.
Additional evidence of the mechanistic similarity is provided by enhancer of zeste \([E(z)]\). Genetic evidence suggests that \(E(z)\) can be considered both a Pc-G and a Trx-G gene (LaJeunesse & Shearn, 1996). For example, embryos from \(E(z)\) mutant mothers have anterior to posterior segment transformations which are characteristic of embryos with loss of Pc-G gene function, while double heterozygous combinations of recessive loss-of-function \(E(z)\) and \(ash1\) (absent, small, or homeotic discs 1) alleles express homeotic transformation phenotypes similar to those expressed by double heterozygous combinations of recessive loss-of-function \(trithorax\) and \(ash1\) alleles.

1.4.6 Hox Gene Regulation and Polycomb-Group Gene Regulation

Most ongoing work has focussed on determining the mechanisms of regulation of the HOX genes by the Pc-G and Trx-G proteins. In contrast, the mechanisms responsible for the regulation of Polycomb-group gene expression are unknown, so any hypotheses are purely speculative.

It seems reasonable to suggest that, as Hox gene expression is initiated by the actions of the gap and pair-rule genes, the initiation of Pc-G and Trx-G expression could be achieved through similar mechanisms. However with passing time, a strikingly similar situation is observed between the Hox genes and the Pc-G genes when the gap gene expression decays – in both cases Pc-G and HOX gene activity is still required for correct development.

One possibility is that the Polycomb-group and Trithorax-group genes, as well as repressing specific HOX gene expression, also regulate their own expression. The existence of different complexes could somehow regulate whether a Pc-G or Trx-G member is to be expressed or not within a particular cell type. If this were the case, then these decisions would have to be transmitted to subsequent cell generations, possibly through a memory mechanism identical to that of Hox gene expression.
Figure 5—Expanded overview of the regulatory connections involved in anteroposterior specification in Drosophila melanogaster. The maintenance pathways of the Trithorax-Group and Polycomb-Group proteins are superimposed.
1.5 Beyond Anteroposterior Patterning

Beyond regulation of homeotic pathways, increasing evidence points to Pc-G genes playing important lineage-specific roles throughout life in various somatic tissues, including the hematopoietic system and sex determination, as well as possible roles in the control of the cell cycle.

1.5.1. Polycomb-Group Proteins and Hematopoiesis

Evidence suggests that Hox gene expression is highest in the most primitive bone marrow cells with a markedly elevated expression of Hox genes located at the 3' end of the Hox gene clusters. As differentiation of bone marrow cells occurs, there is a gradual decrease in overall Hox gene expression, accompanied by a shift in expression towards the 5' of the Hox gene cluster (Sauvageau et al., 1994). Additionally, it has been suggested that expression of some Hox genes may be lineage-restricted, for example, as typified by the erythroid specific expression of several Hox-2 genes (Lawrence & Largman, 1992, Magli et al., 1991, and Mathews et al., 1991).

In support of this observation that there is a highly regulated program of Pc-G gene expression with mature bone marrow subpopulations showing much higher Pc-G gene expression levels relative to less differentiated precursors (Lessard et al., 1998). It has been suggested that Pc-G protein complexes present in primitive cells may differ from those found in mature bone marrow cells. For some of the Pc-G genes, e.g. Enx-1 and M31, the upregulation of expression occurs in the earliest stages of hematopoietic differentiation, whereas for others, e.g. Mel-18 and M33, increases in their expression levels coincides with later stages of differentiation. However, it was noted that Bmi-1 exhibits a strikingly different pattern of expression with high expression levels in primitive cells and very little expression found in mature bone marrow subpopulations (Lessard et al., 1998) (Figure 6).
Figure 6 – The expression of Pc-G proteins during the development of hematopoietic cells. The expression of Pc-G proteins, with the exception of Bmi-1, is thought to inversely correlate with the activity of Hox genes (Sauvageau et al., 1994, Lessard et al., 1998).

Other evidence for a role for Pc-G genes in hematopoiesis comes from the phenotypic alterations that have been reported in Pc-G mutants. B and T cell populations in M33⁻/⁻ mice exhibit a decreased proliferative response to plant agglutinin (Coré et al., 1997). Bmi-1⁻/⁻ mice display a progressive aplastic disease characterised by decrease in the number of bone marrow hematopoietic cells with the bone marrow space being replaced by adipocytes. A reduction in a subset of bone marrow colonies was also observed in heterozygotes, indicating a possible gene dosage effect (van der Lugt, 1994). In contrast, over expression of Bmi-1 causes development of lymphomas (Haupt et al., 1991, van Lohuizen et al., 1991a, and Haupt et al, 1993), and in humans, the
BMI-1 gene has been located to 10p13 (Alkema et al., 1993), a region involved in translocations in various leukemias (d’Alessandro et al., 1990, and Pui et al., 1987). An additional line of evidence suggesting a role for Polycomb group proteins in hematopoietic cell function is the finding of a direct interaction between the Pc-G protein Enxl (the murine homologue of the Drosophila enhancer of zeste [E(z)]) and Vav, a proto-oncogene that plays a critical role in hematopoietic signal transduction (Hobert et al., 1996).

All of these results suggest a role for the Pc-G proteins in regulating differentiation and/or proliferation of human hematopoietic cells by silencing Hox gene expression. A complex-constitution model has been proposed, in which newly expressed Pc-G gene products would progressively interact with existing Pc-G protein complexes, favouring novel interactions with target sequences. This, in turn, would allow a progressive packaging of DNA into a heterochromatin-like structure and, for the Hox genes, a progressive 3' to 5' closure of the clusters, allowing proper differentiation of the hematopoietic stem cells.

1.5.2 Bmi-1 and Cell Cycle Control

Jacobs et al. (1999), whilst investigating the role of Bmi-1 in cell proliferation, observed that Bmi-1 could alter the expression of the tumour suppressors p16 and p19, both encoded by the ink4a locus. Bmi-1 deficiency resulted in raised p16 and p19 expression, and conversely, Bmi-1 overexpression downregulated p16 and p19 expression. The ink4-encoded proteins (inhibitors of cdk4) are able to block cyclin-dependent kinase (cdk) activity (Kamb et al., 1994, Serrano et al., 1993, Chan et al., 1995, Hirai et al., 1995). Cyclin D-dependent kinases are proteins that regulate phosphorylation of the retinoblastoma protein (Rb) and exit from the G1 phase of the cell cycle. Removal of ink4a dramatically reduced the lymphoid and neurological defects seen in Bmi-1 deficient mice, indicating that ink4a is an in vivo target for Bmi-1. These observations suggest that Pc-G mediated repression is required not only for regulation of Hox gene expression, but also for the control of expression of critical cell cycle regulators.
**1.5.3 M33 and Sex Determination**

In a study by Katoh-Fukui et al., (1998) mice were generated in which the murine Pc homologue, M33, had the C-terminal domain deleted. More than half of the mice died before weaning, and of note was the fact that the surviving male offspring showed male-to-female sex reversal, i.e. XY M33 C-term − chromosomal males displayed female genitalia. Gonadal growth defects appeared near the time of expression of the Y-chromosome specific Sry gene, suggesting that M33 deficiency may cause sex reversal by interfering with steps either upstream, or downstream of Sry.

**1.5.4 Other Functions of Polycomb-Group Proteins**

**1.5.4.1 Polycomb-Group Proteins in Plants**

Over recent years studies of *Arabidopsis* have revealed a number of homeotic genes that play a role in plant development and in particular, flower development. Goodrich et al. (1997) described the isolation of the CURLY LEAF (CLF) gene which, through examination of clf mutants, is thought to be essential to repress the floral homeotic gene AGAMOUS (AG) (Mizukami & Ma, 1992). In particular it was observed that CLF is not required for the establishment of correct AG expression, rather it is required to keep AG turned off where it should not be expressed. Of interest is the observation that the CLF protein is very similar at the amino acid level to the polycomb-group of proteins, in particular to the product of the *Drosophila enhancer of zeste* gene [E(z)]. This high level of similarity suggests that CLF has a function that has been conserved between plants and animals.

**1.5.4.2 Cosuppression**

From transgenic analysis of many species a scenario exists where a positive correlation is normally observed between copy number and transgene expression. However, a set of phenomena has been identified whereby the presence of supernumerary (two or more) copies of a gene in the nuclear genome results in specific repression of expression of some or all copies of that gene. These have been collectively referred to as cosuppression (Jorgensen, 1995). Cosuppression involves two potentially separate processes, the first being a homology based trans-interaction that allows detection and location of supernumerary copies of a gene either through homologous genomic DNA sequences...
(reviewed in Wu et al., 1993) or through RNA-mediated, post-transcriptional mechanisms (Wassenegger et al., 1994). The second, cosuppression-associated repression, is also mechanistically heterogeneous, ranging from mechanisms that effect repression of transcription (Pal-Bhadra et al. 1997) to those that act at the level of post-transcriptional silencing through controlled RNA-degradation (Goodwin et al., 1996).

An interesting example comes from Pal-Bhadra et al. (1997) who introduced from two to six copies of a white promoter-Alcohol dehydrogenase (Adh) reporter fusion gene into the Drosophila genome. They observed that both endogenous and transgene Adh expression was progressively reduced with increasing copy number, but that when these fusions were tested in Drosophila that were heterozygote for either Polycomb or Polycomb-like the degree of cosuppression was reduced by approximately half, suggesting that silencing is dependent on Pc-G gene products.

1.6 Bmi-1

1.6.1 Cloning of the Gene

Bmi-1 was originally identified as an oncogene that cooperates with c-myc in B- and T-cell lymphomagenesis (Haupt et al., 1991, and van Lohuizen et al., 1991a). Transgenic mice overexpressing the c-myc gene in their lymphoid compartment by virtue of the immunoglobulin heavy chain enhancer (Eμ) were infected with Moloney murine leukemia virus (MoMLV). The rationale behind this was that retroviruses that lack an oncogene, such as MoMLV, promote tumorigenesis primarily by inadvertent insertion near or within cellular oncogenes, thereby enforcing their expression or altering their structure. Thus, the provirus tags relevant genes involved in tumorigenesis.

Two groups, working independently, found that in approximately 50% of independently induced pre-B cell lymphomas, the provirus was found to insert in or near the Bmi-1 gene (B lymphoma Mo-MLV Insertion Region 1), resulting in an enhanced transcription of the Bmi-1 gene, suggesting that an increase in the Bmi-1 gene product contributes to tumorigenesis. In wild-type mice, Bmi-1 mRNA was detected in most organs, with highest levels detected in thymus, brain and testis. The murine Bmi-1 gene consists of 10 exons and encodes a 324 amino acid, 45KDa nuclear protein which
contains a novel zinc finger motif (C$_3$HC$_4$), the RING finger, shared by a diverse set of proteins involved in gene regulation, DNA recombination and DNA repair.

The Bmi-1 gene is highly conserved in evolution, and Bmi-1 probes were used to clone both the BMI-1 in humans (Alkema et al., 1993), which maps to chromosome 10p13, and a related, though already cloned Drosophila gene, posterior sex combs (psc) (van Lohuizen et al., 1991b). The human BMI-1 protein contains 326 amino acids, two more than its murine homologue, and shows 98% identity at the protein level and 92.4% homology at the nucleotide level. The structural motifs are entirely conserved, and as well as the RING zinc finger motif, other motifs include acidic and basic domains, a potential nuclear localisation motif, a central domain exhibiting a putative helix turn helix motif, and PEST sequences (proline (P), glutamic acid (E), serine (S) and threonine (T)) which confer increased susceptibility to protein degradation (Figure 7) (Rodgers et al., 1986).

![Figure 7](image)

**Figure 7 – Representation of the human BMI-1 protein.** The major motifs present within the protein and their approximate amino acid numbers are shown.

### 1.6.2 Posterior Sex Combs

*Posterior sex combs* is a relatively large protein consisting of 1603 amino acids and predicted mass of 170KDa. The protein is encoded by a 6031bp cDNA. The Bmi-1 gene product shows distinct homology to a 200 amino acid region in the N-terminus of Psc, and slightly less homology to a related gene *Suppressor 2 of zeste (Su(z)2)*, with the zinc
finger domain being the most highly conserved between all three proteins (Brunk et al., 1991, and van Lohuizen et al., 1991b) (Figure 8).

Figure 8 - BLAST comparison of human BMI-1 against posterior sex combs (A), and suppressor 2 of zeste (B), at the amino-acid level. BLAST analysis performed using default page-settings for matrix BLOSUM62 at WWW reference: (http://www.ncbi.nlm.nih.gov/gorf/bl2.html)

Loss of function mutations in Psc are embryonic lethals displaying head defects, and partial posterior transformation of several segments as well as partial ventral-to-dorsal transformations (Jurgens, 1985, and Adler et al., 1989, 1991). The phenotype observed is characteristic of the Pc-G of proteins involved in maintaining the segment-specific repression of homeotic selector genes of the Antennapedia and Bithorax complexes and so, with Bmi-1 demonstrating significant homology to Psc, it was reasonable to suggest that Bmi-1 is the murine equivalent of Psc and a member of the vertebrate Pc-G; however, an immediate obvious discrepancy exists when the size of the two proteins are compared, with Psc being some 5 times larger.

1.6.3 Bmi-1 as a Vertebrate Polycomb-Group Member

Despite the size discrepancy, further evidence for a role in vertebrates for Bmi-1 in the regulation of homeobox containing genes comes in the form of genetic studies. Targeted deletion of Bmi-1 results in a posterior transformation along the complete
anteroposterior axis of the embryo (van der Lugt et al., 1994), as well as the hematopoietic problems mentioned previously. In contrast, transgenic mice overexpressing Bmi-1 show the opposite phenotype, that of a dose dependent anterior transformation (Alkema et al., 1995). When Hox gene expression was analysed, it was found that Bmi-1+ mice exhibit an anterior shift in the expression of homeotic genes (van der Lugt et al., 1996), and conversely, mice overexpressing Bmi-1 show the opposite phenotype, i.e. a posterior shift in the expression boundary of homeotic genes.

From this data, it was proposed that, in addition to the structural and functional conservation of homeotic selector genes through evolution, structural and functional conservation of the homeotic regulators has also occurred, with Bmi-1 as the first member of a vertebrate Pc-G complex that regulates segmental identity by repressing Hox genes throughout development. Interestingly, the fact that Bmi-1 was not found to alter the expression boundaries of all of the Hox genes examined, again supports the hypothesis that different Hox genes may be under the regulatory effect of different Pc-G complexes (van der Lugt et al., 1996).

At the protein level, as well as the conservation of certain Posterior sex combs features, Bmi-1 has also been found to be a constituent of a large multimeric protein complex along with other homologues of Drosophila Pc-G proteins such as Mph1 (Alkema et al., 1997a), M33, MPc2 (Alkema et al., 1997b) and other as yet unidentified proteins. These complexes are reminiscent of the multimeric complexes seen in Drosophila (Franke et al., 1992) and provide further evidence for a conservation of Polycomb-group function in higher organisms.
1.7 Strategies for the Identification of Protein-Protein Interactions

1.7.1 Introduction

Over the last couple of years, a trend has emerged within human molecular genetics to move away from the traditional areas of gene identification and cloning and into a more functional aspect. It has been realised that in order to understand better some of the mechanisms responsible for causing genetic disease, it is not sufficient to isolate a candidate gene, but more that it is necessary to understand how the protein(s) encoded by that gene is responsible for causing disease. Thus, the trend has been to develop new methods aimed at identifying and analysing protein interactions and from this constructing hypotheses regarding protein function. Several methods have been developed such as phage display, double tagging and the yeast two-hybrid system (for a general review of the yeast two-hybrid system see Fields & Sternglanz, 1994, and Allen et al., 1995).

1.7.2 The Yeast Two-Hybrid system

1.7.2.1 Historical Perspectives and Theory

The basis for the two-hybrid was established in the mid-1980s when it was discovered that many eukaryotic transcription factors were modular in nature and were shown to comprise two distinct and separable domains: a specific DNA-binding domain (BD), and an acidic region that is required for transcriptional activation (AD) (Keegan et al., 1986). These two domains can be separated by large stretches of heterologous protein whilst still retaining function. The two-hybrid methodology exploits the ability of a pair of interacting proteins to bring these two domains into close proximity such that a reporter gene is thereby activated.

Work by Fields & Song (1989) resulted in the first general genetic assay for protein interactions when they measured the interaction between two yeast proteins involved in the regulation of the SUC2 gene, Snfl and Snf4, by expressing them as GAL4-derived chimeras. However, it was Chien et al., (1991) who first showed that the assay could be used to identify novel interacting proteins when they used Sir4 fused to the GAL4-BD as a bait. Using this system, they identified Sir4 itself and a novel Sir4-interacting protein, Sfi1.
**1.7.2.2 Yeast Two-Hybrid System - Methodology**

The method requires the construction of hybrid genes to encode: (1) a DNA-binding domain fused to a protein X (normally the protein of interest, i.e. the “bait”); and (2) an activation domain fused to a protein Y (a potentially interacting protein, or a cDNA library from a suitable source). The DNA-binding domain targets the hybrid protein to its binding site (located in the upstream activation sequence (UAS) of the reporter gene), however, because most proteins lack an activation domain, this DNA-binding hybrid does not activate transcription of the reporter gene. Likewise, the activation domain hybrid protein cannot activate transcription because it is not targeted to the UAS of the reporter gene. However, when both hybrid proteins are present and interact, the non-covalent interactions between X and Y tethers the binding domain to the activation domain and activates transcription. Conversely, when both hybrids are present and do not interact, e.g. X and Z, the components are not brought into proximity and transcription remains silent (see Figure 9).

The transcription factors most commonly used in two-hybrid systems are the **GAL4** and **LexA** proteins, and also the activation domain of the **Herpes simplex** VP16. The **GAL4** binds to the **GAL1** promoter as a dimer, and the reporter genes contain several copies of the **GAL1** binding site. Several reporter genes are used, however, one of the most common is the **E.coli lacZ** gene, which produces blue colonies on plates or filters containing the substrate X-Gal. Additionally, the addition of auxotrophic markers, such as genes involved in amino acid biosynthesis e.g. **HIS3**, allows selection for cells that grow on media lacking the relevant amino acids. The reporter genes may be integrated into the genome, or exist independently as an extrachromosomal reporter plasmid.

The two hybrid system should work in any eukaryotic cell, however, most work has been carried out in a yeast-based system, which has numerous advantages over other systems including relative ease of transformation, ease of plasmid retrieval, availability of nutritional markers and well-characterised reporter genes for direct selection. Also, endogenous yeast proteins are less likely to bind a mammalian target protein to prevent its interaction with a protein encoded by a library.
Figure 9 – The underlying principle of the yeast two-hybrid system. (A) – A transcription factor consists of two distinct domains, an Activation Domain and a DNA Binding Domain which can be separated by the fusion of proteins to either domain. (B) – The activity of the transcription factor can be reconstituted, and a reporter gene activated, through the association of two interacting proteins (X and Y), but not if two proteins do not interact (X and Z).
1.7.2.3 Yeast Two-Hybrid System - Applications

The two-hybrid system has three major applications. First, testing known proteins for interactions. Known proteins may be considered to interact with each other on the basis of genetic or biochemical data, sequence similarities or a range of other criteria and, using such data, hybrid-proteins can be created to test whether an interaction occurs. Second, once an interaction has been defined, the system can be used to delineate domains and residues responsible for mediating the interactions through the generation of deletion hybrids. Third, and probably the most important application, is the ability to rapidly isolate proteins that interact with a given bait protein. A gene of interest is fused to the DNA-binding domain and is used to screen an activation domain library prepared from a suitable source.

1.7.2.4 Yeast Two-Hybrid System - Advantages and Limitations

The success of the two-hybrid system lies in its sensitivity and as a result it can be used to detect interactions not necessarily revealed by other in vitro methods such as immunoprecipitation. The sensitivity is as a consequence of a number of factors. Firstly, the hybrid proteins are generally highly expressed from strong promoters on high copy-number plasmids, which favours protein complex formation. Secondly, the signal measured is roughly in proportion to the equilibrium concentration of the heterodimeric species. In contrast, physical methods, e.g. immunoprecipitation, generally rely on a low rate of disassociation since complexes must survive several washes necessary to prove specificity. Thirdly, the stability of the hybrid protein complex is probably enhanced by the interaction of the activation domain with proteins from the transcriptional machinery. Ternary complex formation increases the stability of the complex on DNA and further increases the probability of detection.

However, the system is not without its drawbacks and limitations. These can generally be classified into three groups:

**False positives:** these are interactions which do not represent a real in vivo interactions despite evidence from the two-hybrid system suggesting otherwise. This may be for several reasons such as one of the target proteins showing some transcriptional activation function. False positives may also arise if the GAL4 protein, which is a native yeast protein, interacts with an endogenous yeast protein resulting in activation of the reporter.
The effect of false positives can be reduced by using bacterial transcription factors such as *LexA* instead of *GAL4* which should not cross-react with any yeast proteins, and also the use of dual selection criteria using two distinct reporter genes e.g. both *HIS3* and *lacZ*.

**False negatives**: these are interactions which occur in vivo, yet are not detected by the two-hybrid system. For example, interactions which are dependent on, or are mediated by post-translational modifications may not be detected since the proteins are targeted to the nucleus and thus may avoid the endoplasmic reticulum where some of these modifications occur. Alternatively, the AD or BD may affect the proteins native conformation altering its biochemical properties.

**Toxic Proteins**: This problem arises when one of the hybrid genes is harmful or lethal when expressed. Given the basic conservation of cellular process between eukaryotes, it is apparent that some cellular systems are regulated by a careful balance of positive and negative factors, and it may be that the introduction of a hybrid protein disrupts this balance to the point of lethality. Steps toward overcoming this have recently focussed on using vectors that express the prey conditionally, for example, the expression of cDNA-encoded proteins from the intact yeast *GAL1* promoter which is only active in cells grown on galactose and repressed in cells grown on glucose.

1.7.3 Further Developments of the Yeast Two-Hybrid System

Since its original conception, several modifications to the system have been developed to aid further analysis of protein, and other, interactions.

1.7.3.1 The One-Hybrid System

This is used to isolate DNA-binding proteins that are able to bind to specific DNA sequences. A library is fused to the activation domain of, e.g. *GAL4* and is used to screen for proteins that are capable of binding to a specific sequence through the activation of a reporter gene (*Inouye et al., 1994* and *Li & Herskowitz, 1993*).

1.7.3.2 The Mammalian Two-Hybrid System

This is essentially the same as the yeast two hybrid system with the exception that the assay is performed in mammalian cells, and can be used to eliminate the possibility of a
false positive that may be an artifact of working in yeast cells. Instead of using nutritional markers such as those used in yeast, more commonly, the activation of the gene encoding chloramphenicol acetyl transferase (CAT) is measured (Fearon et al., 1992).

1.7.3.3 The Reverse Two-Hybrid System

The reverse two-hybrid system (Leanna & Hannink, 1996) enables genetic selection against a specific protein-protein interaction, and consequently is of use in the identification of mutant proteins that have lost their ability to associate with their partner protein.

The \textit{CYH2} gene is responsible for the sensitivity of yeast to the antibiotic cycloheximide (CHX). Mutation of \textit{CYH2} can produce a CHX-resistant strain, \textit{cyh2}. In the reverse two-hybrid system, a \textit{cyh2} yeast strain is used containing \textit{CYH2} under the control of the \textit{GAL1} promoter. However, because the \textit{CYH2} allele is dominant over the \textit{cyh2} allele, expression of the \textit{CYH2} gene product will confer CHX sensitivity to a CHX-resistant \textit{cyh2} strain. Thus localisation of the reconstituted \textit{GAL4} complex to the \textit{GAL1} promoter through association of the fusion proteins will activate transcription of \textit{CYH2} and confer sensitivity to CHX. Consequently a situation arises whereby a pair of interacting proteins confers sensitivity to CHX and the yeast will not grow. Contrast this with yeast containing a protein pair that do not interact (e.g. due to a mutation in one of the proteins) – these will not activate the \textit{CYH2} gene leaving the yeast CHX-resistant, allowing growth to occur. The potential of this system is enormous especially within the pharmaceutical industry as a method of screening for compounds that disrupt or prevent an interaction occurring.

1.7.3.4 Three-hybrid systems

Some protein-protein interactions cannot be detected with standard two-hybrid methods because they require a third molecule to be present in yeast. To overcome this problem, several variants of a three-hybrid system have been developed. Amongst these is the method developed by Van Aelst et al., (1993), which detects interactions through the expression of a third, bridging, protein. Alternatively, Osborne et al., (1996) expressed a tyrosine kinase, Lck, while performing interactor hunts with intracellular portions of receptors as baits. By this means, they were able to isolate proteins which interacted specifically with tyrosine phosphorylated forms of the receptors.
One other three hybrid method is that used to detect protein-RNA interactions. In the system devised by SenGupta et al., (1996), one of the three hybrid constructs utilised is LexA fused to a well-characterised sequence-specific RNA-binding protein from bacteriophage MS2. The second hybrid is an RNA molecule that contains both a bait sequence and a stem-loop structure recognised by MS2 and so associates with the LexA-MS2 protein. The third hybrid is a protein fused to an activation domain. If the protein of interest binds to the RNA sequence of interest, transcription of the reporter gene will be activated.

1.7.3.5 Exhaustive Screening

Through the various systems described, extensive knowledge has been accumulated relating to defined biochemical pathways. However, as yet, integration of these various pathways against the background complexity of a living cell has proved difficult with so many processes occurring simultaneously. Recently, an attempt was made to integrate some of these pathways in the yeast Saccharomyces cerevisiae. Using known pre-mRNA splicing factors as initial baits (Fromont-Racine et al., 1997), several interacting proteins were identified which, in turn, then became the new bait proteins for a second round of screening. Repeating this procedure several times led to the characterisation of a network of interactions between known splicing factors, the identification of new splicing factors and the revelation of novel potential functional links between cellular pathways, increasing further our understanding of biochemical processes.
1.8 Project Aim

As mentioned earlier, work carried out by Franke et al. (1992) identified the existence of a Drosophila multimeric Pc-G complex consisting of Pc, Ph and several other unidentified proteins. However, at the time this project started, the structural and functional similarities between mammalian and Drosophila Pc-G proteins were only just becoming apparent, and as such, no comparable complexes had yet been found to exist in higher organisms.

Despite this, with the rationale that mammalian Pc-G proteins might act in a similar fashion it seemed reasonable to suggest that BMI-1 most likely interacts with a variety of proteins, with other Pc-G members being amongst the most likely candidates. Thus, the aim of this project was to investigate whether, using the yeast two-hybrid system, BMI-1-interacting proteins could be identified. The yeast two-hybrid system had already proved useful for isolating protein partners and so seemed an almost ideal system to use in this regard. If any proteins were to be identified using the yeast two-hybrid system, follow-up work was to consist of techniques to corroborate the interactions, then attempt to place the interaction in a biologically meaningful context.

While this project has been continuing, several BMI-1 interactors have been identified in other laboratories, for example, HPH-1 and -2 (Gunster et al., 1997) and RING1 (Satijn et al., 1997). These interactions have all been recognised using the yeast two-hybrid system, confirming the suitability of the system for this project.
Chapter 2
Materials and methods

2.1 Materials

2.1.1 Reagents

All reagents used were of AnaLaR grade and were obtained from either British Drug Houses (BDH) or Sigma Aldrich, except for those described below. Bacto-tryptone, bacto-peptone, yeast extract, nitrogen base without amino acids, and agar were from Difco Laboratories. Low melting point (LMP) agarose TEMED and Lipofectin™ were from Gibco BRL. Effectene™ was obtained from Qiagen. Ficoll 600, sephadex G50, deoxyribonucleotides, ribonucleotides, random hexamer-oligonucleotides and Protein A-Sepharose CL-4B beads were from Pharmacia LKB Biotechnology. Tris-buffered Phenol and NaOH pellets were from Fisons, plc. Molecular Biology grade agarose was from Eastman Kodak. Sequagel and Protogel polyacrylamide gel products and buffers were obtained from National Diagnostics. YeastMaker™ Carrier DNA was from Clontech. Dried skimmed milk was from Safeway plc.

2.1.2 Materials

Hybond N, Hybond N+ and Hybond C nylon membranes were from Amersham International plc. PVDF membrane was from ICN. Biomax-MR autoradiography film from Kodak was used for 35S applications. For all other autoradiographic and chemiluminescent methods X-ograph Blue X-ray film from X-ograph Imaging Systems was used. ART-filter tips were from NBL. Cell culture flasks and all other plasticware were obtained from Falcon or Sterilin. MicroSample™ plates were obtained from Pharmacia

2.1.3 Commercial Kits

Sequenase™ Version 2.0 DNA sequencing kits were from Amersham. ABI™ Dye Terminator Cycle sequencing Ready Reaction Kit (FS enzyme) was purchased from Perkin Elmer. BioNick™ kits were obtained from Gibco BRL. The pCR-Script™ and
the *In-vitro* Express™ translation kit were obtained from Stratagene Ltd. TNT Quick-coupled/Coupled transcription-translation kits were obtained from Promega. Plasmid midi- and maxiprep Q-100/Q-500 tips were obtained from Qiagen, along with all the buffers and solutions required for the DNA preps. CDP-STAR™ chemiluminescent antibody detection kits were obtained from New England Biolabs. The Multiple Tissue Northern Blot and the Marathon cDNA Amplification kit were obtained from Clontech.

2.1.4 Enzymes

Restriction endonucleases, large fragment *E. coli* DNA polymerase I (Klenow), and *E. coli* DNA ligase were from Gibco BRL. Lyophilised RNase A was from Sigma. *Pfu* Polymerase was from Stratagene. *Taq* polymerase was from Bioline. Calf Intestinal Phosphatase and Pefabloc were from Boehringer Mannheim. SP6, T3 and T7 RNA polymerases were from Promega along with RNAsin.

2.1.5 Immunoglobulins

Anti-BMI and anti-MCM6 polyclonal antibodies were made commercially by Research Genetics Inc. Anti-FLAG antibody was made by Eastman Kodak. Rabbit anti-mouse IgM antibody was from Zymed Laboratories. Anti-HA antibodies were from Babco. Anti-c-myc monoclonal antibody, rat anti-HA high affinity monoclonal antibody, and HA peptiide were obtained from Boehringer Mannheim. F6 (anti-BMI-1) antibody, raised against a GST-Bmi-1 (amino acids 1-202) fusion protein (*Alkema et al., 1997a*) was kindly provided by Maarten Van Lohuizen (The Netherlands Cancer Institute). Anti-MCM6, raised against a peptide sequence as described by *Holthoff et al., (1998)*, was kindly donated by Hiroshi Nojima (Osaka University).

2.1.6 Radioisotopes

[α-32P] dCTP (3000Ci/mmol) was obtained from ICN Biomedicals Inc. [α-35S] dATP (>600Ci/mmol) and L-[35S] Methionine (>1000Ci/mmol) were obtained from Amersham.
2.1.7 Size Markers

The 1Kb ladder purchased from Gibco BRL was used for all DNA analysis. Mid Range Protein Molecular markers were obtained from Promega, and Rainbow Protein Molecular markers were obtained from Sigma. Multi-Tag-Marker was obtained from Boehringer Mannheim.

2.1.8 Bacterial Strains

DH5α:
- F−, φ80, lacZΔM15, Δ(lacZYA-argF), U169, deoR,
- recA1, endA1, hsdR 17(rk+, mk+), supE44, λ−,
- thi-1, gyrA96, relA1

XL1B:
- recA1, endA1, gyrA46, thi-1, supE44, hsdR17,
  - relA1, lac [F′ proAB, lacI ΔlacZΔM15, Tn10(Tet')] ]

HB101:
- Δ(gpt-p10A)62, leuB6, thi-1, lacY1, hsdSB20,
  - recA, rpsL20(str+), ara14, galK2, xyl-5, mtl-1,
  - supE44, mcrB

BL21:
- E.coli B F−, dcm, ompT, hsdS(rB−mB−) gal

2.1.9 Yeast Strains

Y190
- MATa, gal4, gal80, his3-200, trp1-901, ade2-101
- ura3-52, leu2-3,-112 + URA3::GAL(UAS)→lacZ,
  - LYS2::GAL(UAS)→HIS3, cyh

Y166
- MATa, gal4, gal80, his3-200, trp1-901, ade2-101,
  - leu2-3,-112, + GAL(UAS)→URA3, GAL(UAS)→lacZ
  - LYS2::GAL(UAS)→HIS3.

PJ69-4A
- MATa, gal4, gal80, his3-200, trp1-901, ura3-52,
  - leu2-3,112, + GAL2→ADE2, LYS2::GAL1→HIS3,
  - met2::GAL7→LacZ
Both Y190 and Y166 were kindly provided by Stephen Elledge (Baylor College of Medicine) and PJ69-4A was kindly provided by Philip James (University of Wisconsin Medical School).

2.1.10 DNA Clones and Libraries

The BMI-1 (K562) clone in pSP72, Bmi-1 in pPC97 and Mph1 (C-ter fragment) in pPC67 were kindly provided by Maarten van Lohuizen.

The yeast two-hybrid B-cell library in pACT was kindly provided by Stephen Elledge (Baylor College of Medicine) and the mouse 9.5/10.5dpc embryo library in pVP16 was donated by Stan Hollenberg (Vollum Institute).

c-myc-RING1 was kindly provided by Paul Freemont (Imperial Cancer Research Fund).

The full-length MCM6 cDNA in pBluescript was kindly provided by Hiroshi Nojima.

The human fetal brain 5’ STRETCH PLUS cDNA library was from CLONTECH.

2.1.11 Vectors

pBluescript SK+, pCRscript and pCMVTag1 were from Stratagene Ltd., pAS2-1 was from Clontech Ltd., pET28a was from Novagen Ltd., pcDNA3 was from Invitrogen Ltd., and the pGEX vectors were from Pharmacia.

2.1.12 Oligonucleotides

Oligonucleotides were initially synthesised ‘in house’ using a Pharmacia LKB Gene Assembler Plus, and more recently, custom synthesised by Genosys UK.

2.1.13 Cell Lines

Two main cell lines were used:– U2OS and HEK-293, both of which were obtained from the ECACC (the European Collection of Cell Cultures, England)

2.2 Solutions, Buffers, and Media

Solutions and media were autoclaved at 15 pounds per square inch (psi) for 20 min. where necessary. Electrophoresis and blotting solutions were generally not autoclaved.
2.2.1 DNA Related Solutions

2.2.1.1 General Solutions

20xSSC: 3M NaCl, 0.3M tri-sodium citrate, pH 7.0

Denaturing solution: 0.5M NaOH, 1.5M NaCl

Neutralising solution: 1M Tris-HCl (pH 8.0), 1.5M NaCl

2.2.1.2 Solutions for DNA Preps

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

Solution II: 0.2M NaOH, 1% SDS

Solution III: 3M KAc diluted with glacial Acetic acid to pH 4.8

QBT: 750mM NaCl, 50mM MOPS (pH 7.0), 15% ethanol, 0.15% Triton X-100

QC: 1M NaCl, 50mM MOPS (pH 7.0), 15% ethanol

QF: 1.25M NaCl, 50mM Tris-HCl (pH 8.5), 15% ethanol

STET: 10mM Tris HCl (pH 8.0), 50mM EDTA, 8% Sucrose, 0.5% Triton X-100

2.2.1.3 Hybridisation/Pre-hybridisation/Wash solutions

General hyb/pre-hyb: 3x SSC, 10x Denhardt's, 0.1% SDS, 100μg/ml salmon sperm DNA

Denhardt's reagent (100x): 1% Ficoll, 1% polyriinylpyrrolidone, 1% BSA (fraction V)

Southern Blot wash buffer: 0.1-3xSSC, 0.1% SDS

MTN Wash solution 1: 2xSSC, 0.05% SDS

MTN Wash solution 2: 0.1xSSC, 0.1% SDS
2.2.1.4 Random-Priming Buffer

Oligo-labelling buffer solutions A, B and C are mixed in the ratio 1 : 2.5 : 1.5

(OLB):
Solution O: 1.25M Tris-HCl (pH 8.0), 125mM MgCl₂
Solution A: 1ml solution O, 18μl 2-mercaptoethanol
5μl each of 100mM stocks of dGTP, dATP, and dTTP
Solution B: 2M HEPES, pH 6.8
Solution C: 90 OD units/ml random hexamer nucleotides in 1mM Tris-HCl (pH 7.4), 1mM EDTA (pH 8.0)

2.2.1.5 DNA Electrophoresis Buffers

E-Buffer (TAE) (1x): 40mM Tris-Acetate, 1mM EDTA, pH 7.0
TBE (1x): 89mM Tris-Borate, 1mM EDTA, pH 8.3

2.2.1.6 Gel Loading Buffer

Bromophenol blue (10x): 0.5% bromophenol blue, 0.5% xylene cyanol, 50% Ficoll₄₀₀

2.2.1.7 Lambda Diluent

SM (for phage plating): 35mM Tris-HCl (pH 7.5), 0.1M NaCl, 10mM MgSO₄, 0.01% Gelatin

2.2.1.8 Restriction Enzyme Buffers

The buffers used in restriction digests were enzyme specific and supplied at 10x concentration by Gibco BRL and stored at -20°C.

1x BRL restriction buffer: 50mM Tris-HCl (pH 8.0), 10mM MgCl₂, 0-100mM NaCl, 0-100mM KCl

2.2.1.9 Ligation Buffer

Ligation Buffer was supplied at 5x and stored at -20°C.

1x BRL ligation buffer: 50mM Tris-HCl (pH 7.5), 7mM MgCl₂, 1mM DTT, 1mM ATP, 5% (w/v) PEG₈₀₀₀.
2.2.1.10 Calf Intestinal Phosphatase Buffer
CiP Buffer was supplied at 10x and stored at +4°C.
1x CiP buffer: 50mM Tris-HCl (pH 8.5), 0.1mM EDTA

2.2.1.11 Taq PCR Buffer
Taq PCR buffer was supplied at 10x concentration and stored at -20°C.
1x Taq PCR buffer: 67mM Tris-HCl pH 8.8, 16mM (NH₄)₂SO₄, 0.01% Tween-20, + 2mM MgCl₂ (if necessary).

2.2.1.12 Pfu PCR Buffer
Pfu PCR buffer was supplied at 10x concentration and stored at -20°C.
1x Pfu PCR buffer: 20mM Tris-HCl pH 8.8, 10mM KCl, 10mM (NH₄)₂SO₄, 2mM MgSO₄, 0.1% Triton X-100, 0.1mg/ml nuclease-free BSA.

2.2.1.13 cDNA First Strand Synthesis Buffer
5x Buffer: 250mM Tris-HCl (pH 8.5), 40mM MgCl₂, 150mM KCl, 5mM DTT

2.2.1.14 cDNA Second Strand Synthesis Buffer
5x Buffer: 100mM Tris-HCl (pH 7.5), 25mM MgCl₂, 500mM KCl, 50mM (NH₄)₂SO₄, 0.75mM β-NAD, 0.25mg/ml BSA

2.2.1.15 Second Strand Enzyme Cocktail
20x cocktail: E.coli DNA polymerase I (6U/µl), E.coli DNA ligase (1.2U/µl) and E.coli RNaseH (0.25U/µl)

2.2.1.16 Adaptor Ligation Buffer
5x ligation buffer: 250mM Tris-HCl (pH 7.8), 50mM MgCl₂, 5mM DTT, 5mM ATP, 25% (w/v) PEG₈₀₀₀.
2.2.2 Protein Related Solutions

2.2.2.1 Protein Gel Buffers
Separating gel buffer (4x): 0.5M Tris-HCl, pH 8.8, 0.4% SDS
Stacking gel buffer (4x): 1.5M Tris-HCl, pH 6.8, 0.4% SDS

2.2.2.2 10x Protein Electrophoresis Buffer
(1L):
30.2g Tris base, 144g glycine, 10g SDS.
Make up to 1L with ddH2O

2.2.2.3 2x Gel Loading Buffer
250mM Tris-HCl, pH 6.8, 4% SDS, 25% glycerol,
0.1% bromophenol blue, 5% β-mercaptoethanol
(added fresh)

2.2.2.4 Fixing Solution
50% Methanol, 9% Acetic Acid. Make up to 1L with ddH2O

2.2.2.5 Staining Solution
As Fixing solution, but with 0.25% Coomassie R250

2.2.2.6 Destaining Solution
25% Methanol, 10% Acetic acid, 65% H2O.

2.2.2.7 Protein Transfer buffer
0.5x Protein electrophoresis buffer, 20% Methanol

2.2.2.8 Bradford Buffer (for Protein concentration Determination)
Dissolve 100mg Coomassie Brilliant Blue G-250 in
50ml 95% Ethanol. Add 100ml 85% phosphoric acid and
make volume to 1L with ddH2O
2.2.2.9 PVDF and Hybond-C Membrane Blocking Solutions

Normal Blocking buffer
(for 100 ml): 10ml 10x PBS, 5g non-fat dry milk powder, 0.1ml Tween-20

Special Blocking buffer
(for 100 ml): 10ml 10x TBST, 3g BSA, 0.22g β-glycerophosphate

Normal and Special Wash buffers respectively: As blocking buffers but without milk powder and BSA

10x PBS: 1.37M NaCl, 27mM KCl, 43mM Na₂HPO₄·7H₂O, 14mM KH₂PO₄
10x TBST: 0.5M Tris-HCl (pH 8.0), 1.5M NaCl, 0.5% Tween-20
Alkali Wash buffer: 10mM Tris-HCl (pH 9.5), 10mM NaCl, 1mM MgCl₂

2.2.3 Yeast Manipulation Solutions

2.2.3.1 Transformation Solutions

10x TE: 100mM Tris-HCl (pH 8.0), 10mM EDTA
10x LiOAc: 1M LiOAc
LiSORB: 1x LiOAc, 1x TE, 1M Sorbitol
LiTEPEG: 1x LiOAc, 1x TE, 40% PEG4000

Yeast Lysis buffer: 10mM Tris-HCl (pH 8.0), 2% Triton X-100, 1% SDS, 100mM NaCl, 1mM EDTA

Z-Buffer: 60mM Na₂HPO₄·7H₂O, 40mM NaH₂PO₄·H₂O, 10mM KCl, 1mM MgSO₄·7H₂O

2.2.4 Bacterial Transformation Solutions

2.2.4.1 Electrocompetent Cell Preparation

10% Glycerol (cold)
2.2.4.2 Chemical Competent Cell Preparation

60mM CaCl$_2$, 15% Glycerol, 10mM PIPES (pH 7.0)

2.2.5 Media

2.2.5.1 Bacterial Media

The quantities given make 1 litre of medium.

L-broth (LB) : 10g bacto-tryptone, 5g yeast extract, 5g NaCl

SOB : 20g bacto-tryptone, 5g yeast extract, 10ml

1M NaCl, 2.5ml 1M KCl

SOC : as SOB + 10ml 2M MgCl$_2$, 10ml 2M Glucose

Minimal media : 200ml 5xM9 salts, 100ml 10x Dropout solution,

5ml 40% Glucose, 1ml 1M MgSO$_4$, 4ml 10mg/ml

Proline, 1ml 1M Thiamine, 2ml 1% Histidine,

2ml 1% Tryptophan, 1.5ml Ampicillin

5xM9 salts : 30g Na$_2$HPO$_4$.7H$_2$O, 15g KH$_2$PO$_4$,

5g NH$_4$Cl, 2.5g NaCl,

15mg CaCl$_2$ (optional)

10x Dropout solution : 300mg/L Isoleucine, 1500mg/L Valine,

(-L, -W, -H) 200mg/L Arginine-HCl, 300mg/L Lysine-HCl, 200mg/L

Methionine, 500mg/L Phenylalanine, 2000mg/L

Threonine, 300mg/L Tyrosine,

200mg/L Adenine, 200mg/L Uracil

Plates: Appropriate medium containing 15g agar

2.2.5.2 Antibiotics

Antibiotics were stored at -20C as 1000x stocks

Ampicillin (amp) : 50mg/ml, filter sterilised

Kanamycin (kan) : 15mg/ml, filter sterilised
2.2.5.3 Colour Selection
X-gal (1000x) : 25mg/ml in dimethyl formamide
IPTG (1000x) : 25mg/ml, filter sterilised

2.2.5.4 Yeast Media
The quantities given make 1 litre of medium.
YPAD: 20g bacto-peptone, 10g yeast extract, 50ml 40% Glucose, 10ml 0.2% Adenine,

Minimal media: 6.7g Nitrogen base 50ml 40% Glucose, 10ml 0.2% Adenine, 10ml 0.2% Uraic, 3ml 1% Lysine, (+5-50ml 1M 3-Aminotriazole, +2.5µg/ml Cycloheximide as necessary)

then add according to selection criteria;
3ml 1% Leucine, 2ml 1%Histidine, 2ml 1% Tryptophan, 2ml 1% Methionine

Plates: Appropriate medium containing 20g agar

2.2.6 GST-Related and other Protein Buffers
2.2.6.1 GST Purification Buffer
STE+T : 10mM Tris HCl (pH8.0), 150mM NaCl, 1mM EDTA, 10% Triton X-100
Other Purification buffers were as STE+T but with either sarcosine (1.5%), Tween-20 (1%) or SDS (0.05%) instead of Triton X-100.
2.2.6.2 In vitro Assay Buffer

TNT: 20mM Tris HCl (pH 8.0), 150mM NaCl, 0.2% Triton X-100, 0.2mg/ml pefabloc

Low/High Salt TNT: As TNT but with 50mM or 1M NaCl

2.2.6.3 Nuclear/Cytoplasmic Preparation Buffers

Buffer A 20mM Hepes (pH 7.6), 100mM KCl, 5% sucrose, 0.25mM EGTA, 0.5mM DTT.

Buffer B As Buffer A + 0.8% NP40

2.2.7 Immunoprecipitation Lysis Buffers

2.2.7.1 General Immunoprecipitation Buffers

RIPA lysis buffer: 50mM Tris HCl (pH 7.6), 150mM NaCl, 1% NP-40, 0.5% Deoxycholate, 0.1% SDS

ELB lysis buffer: 50mM HEPES (pH 7.6), 250mM NaCl, 0.1% NP-40, 5mM EDTA, 0.5mM DTT

(The NaCl concentration could be varied if necessary between 100mM and 400mM)

2.2.7.2 Protease Inhibitors

For most protein techniques, Protease Inhibitors were added at the following final concentrations:

PMSF (1mM)
Aprotinin (1µg/ml)
Leupeptin (1µg/ml)
Pepstatin (1µg/ml)

2.2.8 Mammalian Cell Culture Solutions

2.2.8.1 Cell Culture Media

All media and PBS used were obtained from Gibco BRL. All cell lines were routinely grown in Iscove's Modified Dulbecco's medium with Glutamax II. OPTIMEM™ medium was used in transient transfections.
Supplements: 
Foetal Bovine Serum (FBS) 
L-Glutamine (200mM) 
Penicillin (10000IU/ml) 
Streptomycin (10000UG/ml)

Other: 
PBS 
Trypsin-EDTA Solution

2.2.8.2 Cell Culture β-gal Assay Buffer

2x βgal assay buffer: 0.2M Na Phosphate (pH 7.4), 2mM MgCl₂, 
0.1M β-mercaptoethanol, 1.5mg/ml ONPG
2.2.9 Oligonucleotides and their Uses.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>DNA sequence (5’→3’)</th>
<th>Melting Temperature (T_m) (°C)</th>
<th>Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3</td>
<td>ATTAACCCCTACTAAAGG GA</td>
<td>48</td>
<td>General PCR and sequencing of clones in pBluescript and derivatives thereof.</td>
</tr>
<tr>
<td>T7</td>
<td>TAATACGACTCATAAGG GG</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>SP6</td>
<td>TATTAGGGTGACACTATA G</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Matchmaker BD 5’</td>
<td>TCATCGGAAGAGAGTAG</td>
<td>50</td>
<td>Sequencing of inserts in pAS2-1</td>
</tr>
<tr>
<td>Matchmaker AD 5’</td>
<td>ATACCACTACAATGGATG</td>
<td>52</td>
<td>PCR and sequencing of pACT library inserts</td>
</tr>
<tr>
<td>Matchmaker AD 3’</td>
<td>AGTTGAAAGTGAACCTTGCG GGG</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>VP16F</td>
<td>GAGTTTGCAGATGT G A</td>
<td>52</td>
<td>PCR and sequencing of VP16 library inserts</td>
</tr>
<tr>
<td>VP16R</td>
<td>GGTGTAACCGACGGCA GT</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>BMISalFl</td>
<td>CGGGTCGACTCATGCATC GAACAACGAG</td>
<td>50</td>
<td>PCR of BMI-1 from K562 cDNA clone for cloning into pGEX4T1 and pET28a</td>
</tr>
<tr>
<td>BMISalR1</td>
<td>CAAGTCGACCTTGCCCTGC TTTCCAAG</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>pGEXseq</td>
<td>GGCAAGCCACGTGGT G T</td>
<td>54</td>
<td>Sequencing of clones in pGEX vectors</td>
</tr>
<tr>
<td>MCMBglF</td>
<td>ACTGAGATCTCTTAAATGGAC CTCGGGGCG</td>
<td>56</td>
<td>PCR of MCM6 ORF for cloning into pCMV-Tag1</td>
</tr>
<tr>
<td>MCMBglR</td>
<td>ACTGAGATCTGCCACAGT TCCTCAGC</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>3’ RACE 1</td>
<td>GATTTCTCCTATCGACCCA ACATG</td>
<td>66</td>
<td>Used for Rapid Amplification of cDNA Ends (RACE) for unknown EST clone 25</td>
</tr>
<tr>
<td>3’ RACE 2</td>
<td>GAAGTGGTAGCTTCAGCC CAG</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>5’ RACE 1</td>
<td>TCCACCAGAAACCAACTGC TCCTCC</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>5’ RACE 2</td>
<td>GTCTCCCTGGACTCCGCTC GTCC</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>AP1</td>
<td>CCACTCTAATACGACTCA CTATAGGGC</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>AP2</td>
<td>ACTCACTATAGGGCTCGA GCGGC</td>
<td>77</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 - List of primers and their uses relating to this project.
2.3 Methods

2.3.1 Isolation and Purification of DNA

2.3.1.1 Plasmid Minipreps

Two techniques were employed. The standard alkaline lysis plasmid miniprep technique (described in Sambrook et al., 1989) is a simple and efficient technique used for the routine checking of recombinant plasmids. The boiling technique (Holmes & Quigley, 1981) is somewhat quicker, although the quality of DNA obtained is generally lower, though still sufficient on which to perform restriction analysis.

Alkaline lysis technique - 5ml of an overnight bacterial culture was centrifuged at 12,000g for 5 min. The supernatant was removed and the cell pellet was resuspended in 100μl of solution I. 200μl of solution II was then added, the suspension inverted several times and left on ice for 5 min. After the addition of 150μl of solution III, the sample was again inverted and kept on ice for 5 min. The sample was then centrifuged at 4°C, at 12,000g for 5 min to remove cell debris. After a phenol:chloroform (1:1) extraction, the DNA was precipitated by the addition of 0.7 volume of isopropanol and left on ice for 20 min. The pellet was pelleted by centrifugation, 4°C, 12,000g for 20 min, washed in 70% ethanol and resuspended in 50μl TE or H2O.

Boiling technique - 1.5ml of an overnight bacterial culture was centrifuged at 12,000g for 5 min. The supernatant was removed and the cell pellet was resuspended in 200μl of STET. 20μl fresh lysozyme (10mg/ml in STET) was added, and the suspension vortexed. The sample was then boiled for 2 min at 100°C, centrifuged for 5 min at 12,000g, and the cell debris pellet removed with a toothpick. DNA was precipitated by the addition of 100μl 6M NH₄Ac and 1ml ethanol and left on ice for 20 min. The DNA was pelleted by centrifugation, 4°C, 12,000g for 20 min, washed in 70% ethanol and resuspended in 50μl TE or H2O.

2.3.1.2 Isolation of DNA from Mammalian Cell Cultures

Harvested cell pellets were washed in PBS before resuspension in 1x STE (lysis buffer) containing 200μg/ml proteinase K. Proteins were digested overnight by incubation at 55°C. The samples were then extracted twice with an equal volume of phenol, once with
phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform. The DNA was ethanol precipitated, removed with a Pasteur pipette and washed in 70% ethanol. After air-drying, the DNA was resuspended in TE and stored at 4°C.

2.3.1.3 Plasmid Maxiprep

Large scale DNA preps were performed according to the Qiagen handbook, using the Qiagen-100 or 500 tip. 200ml bacterial cultures were harvested for 5 min at 3000g, and the cell pellet was resuspended in 10ml Solution I containing RNase A (50μg/ml). 10ml of Solution II was added, the solution mixed gently by inversion, and then incubated on ice for 5 min. 10ml Solution III was then added, the sample mixed by inverting several times and left on ice for 20 min. The lysed cells were then centrifuged at 12,000g for 20 min at 4°C to remove cell debris. During this time, a Qiagen-500 tip was equilibrated with 10ml QBT buffer under gravity flow. The cleared lysate was then added to the pre-equilibrated column. Due to the salt and pH conditions of the lysate, the DNA will selectively bind to the resin, while the degraded RNA and cellular proteins can be removed by two washes using the QC buffer. DNA is then eluted from the tip by the addition of 15ml of QF buffer, and the DNA precipitated using 0.7 x vol. of isopropanol. The DNA is then centrifuged at 12,000g for 25 min at 4°C. After washing with 70% ethanol, the pellet is air dried and resuspended in 200μl H2O.

2.3.1.4 Purification of Phage DNA

Phage DNA was isolated by the method of Chisholm (1989). Plaques were isolated using a 3ml plastic pasteur pipette, transferred to 500μl SM containing 20μl chloroform and phage particles were allowed to elute overnight at 4°C. A 10ml overnight culture of host cells grown in the presence of 0.2% maltose was pelleted and the cells resuspended in 4ml 10mM MgSO₄. 50μl of eluted phage were added to 500μl of host cells and incubated at 37°C for 30 min.

2.3.2 Analysis of DNA and RNA

2.3.2.1 Quantitation of Nucleic Acid Concentrations

Nucleic acid concentration was determined by measuring the absorbance of the solution at 260nm. A value of $A_{260} = 1$, correlates to a concentration of ~50μg/ml for a
solution of double stranded DNA, and 40μg/ml of RNA. The A$_{260:280}$ ratio was also determined to ensure the sample is relatively free from proteins. The A$_{260:280}$ ratio should have a value of 1.8 for a pure DNA solution.

Oligonucleotide concentrations were determined by calculating their molar extinction coefficient, where this value is equal to the A$_{260}$ of a 1M solution of primer;

\[
\text{molar concentration} = \frac{A_{260}}{(# \text{ G} \times 12010) + (# \text{ A} \times 15200) + (# \text{ T} \times 8400) + (# \text{ C} \times 7050)}
\]

Where # G is the number of Guanine nucleotides in the oligonucleotide sequence etc.

2.3.2.2 Purification of Deprotected Oligonucleotides

Phosphoramidate-synthesized oligonucleotides require a final step involving concentrated ammonia to effect their complete deprotection. The resulting organic byproducts, e.g. benzamides, need to be removed from the oligos before they may be used in enzymatic reactions. The deprotected oligonucleotides manufactured in house were purified to remove such contaminants according to the procedure of Sawadogo & Van Dyke (1991). 100μl of the deprotected oligonucleotide solution was vortexed vigorously in an eppendorf with 1ml of n-butanol for 30 sec. After centrifugation for 1 min at 12,000g, the single H$_2$O-containing n-butanol phase was removed, the pellet dried and then resuspended in 100μl H$_2$O. The concentration was then calculated as described above.

2.3.2.3 Restriction Enzyme Analysis of DNA

DNA was digested according to the manufacturer's guidelines using the appropriate reaction buffer for a given enzyme, e.g. React Buffer 3 for EcoRI. Digestion of cloned DNA was allowed to proceed at 37°C (except SmaI, at 30°C and PvulII at 55°C) for 1.5-2 hr, while genomic DNA digested for 5-6 hr in the presence of 1mM Spermidine.
2.3.2.4 Electrophoresis of DNA

DNA fragments in the size range 100bp to 10kbp were separated by horizontal agarose gel electrophoresis. Gels were generally 1% agarose dissolved in 1x E buffer containing 0.5μg/ml ethidium bromide, added immediately before casting.

Preparative gels for the purification of DNA fragments were generally 1% LMP agarose. For sizing DNA fragments, a 1Kbp ladder was used as a molecular marker. Gels were run in 1x E buffer, with time and voltage being dependent on the size and resolution required of the fragments. DNA was visualised by exposure to UV light and gels were photographed using standard equipment.

2.3.2.5 Southern Blotting

Southern Blotting (Southern, 1975) was performed by standard techniques using Hybond-N membranes, following denaturation and neutralisation of the DNA within the gel. Transfer was generally allowed overnight. The positions of the wells were pencilled onto the membrane which was then lifted off, rinsed in 3x SSC and then baked at 80°C for 2-3 hr.

2.3.2.6 Phage Library Plating

Phage libraries were titred and plated according to Sambrook et al. (1989). Initial screening was generally of one-half to one million clones at a density of 50,000 plaques per plate. Library plating cells were grown overnight in LB containing 0.2% maltose. Cells were pelleted and resuspended in ½ volume of 10mM MgSO₄. A small volume of library (e.g. 10μl) was added to 200μl SM in a 30ml Sterilin. 500μl of plating cells were added to the library and incubated at 37°C for 20 min. Molten LB agarose was maintained at 48°C and 8ml was added to the library plating mix with gentle mixing before plating evenly onto 150mm LB plates which had been dried and prewarmed to 37°C.

2.3.2.7 Transfer of Phage DNA to Nylon Membranes

Library plates were replicated by the method of Benton & Davies (1977). Hybond N circles were labelled, laid onto library plates and orientation marks were made with a sterile needle. When replicating primary plaques, Hybond N+ membranes were used.
Filter A was removed after 1 min and a replica filter B was laid on the plate. This filter was removed after 2 min. Filters were lifted off and laid plaque side up on 3MM paper soaked in denaturing solution. After 3 min they were transferred to 3MM paper soaked in neutralising solution and left for 3 min. They were then rinsed in 3xSSC and baked for 2 hr at 80°C.

**2.3.2.8 Preparation of DNA Probes**

DNA probes were obtained from LMP agarose gels by extracting twice with an equal volume of phenol, once with phenol : chloroform (1:1) and once with chloroform. The DNA was then ethanol precipitated, and resuspended in ddH2O. A sample of the recovered DNA was run on an agarose gel to determine the approximate concentration relative to the 1Kb band of the size marker.

**2.3.2.9 Radioactive Labelling of DNA Probes**

Double stranded DNA was labelled by random priming (Feinberg & Vogelstein, 1983) using the Klenow fragment of DNA polymerase I. ~50ng of purified probe DNA was diluted with ddH2O to a volume of 35μl and boiled for 5 min. After chilling on ice, 10μl Oligo Labelling Buffer (OLB), 2μl BSA (10mg/ml), 1μl Klenow (1U/μl) and 20μCi [α-32P]-dCTP were added. The reaction was incubated at 37°C for 1-3 hr and applied to a Sephadex G-50 column to remove unincorporated nucleotides. Columns were prepared by pipetting Sephadex G-50 rehydrated in 3x SSC into a 1ml syringe barrel plugged with polymer wool. Subsequent to centrifugation at 250g for 5 min, the labelling reaction was applied to the column along with 100μl 3x SSC and centrifuged as before. The specific activity of the radiolabelled probe was estimated by Cerenkov counting.

**2.3.2.10 Hybridisation**

Filters were generally pre-hybridised for at least 1 hr in ~20ml of prehybridisation solution in a hybridisation bottle (Hybaid) at 65°C. DNA probes were boiled prior to hybridisation. Boiled probes were added to the hybridisation solution to an activity of about 10^6 cpm/ml. Hybridisation occurred overnight in a rotisserie at 65°C.

Subsequent to hybridisation, filters were washed at 65°C at a suitable stringency. The first wash was generally in 3x SSC/ 0.1% SDS for ~30 min, followed by washes at
decreasing concentrations of SSC until the background counts registering on a Geiger counter were 2-5 counts per second (cps). Exposure to X-ray film was for an appropriate length of time with intensifying screens at -70°C.

Southern Blots were then stripped for reuse by the addition of boiling 0.1% SDS.

### 2.3.2.11 Multiple Tissue Northern (MTN) Blot Hybridisation

The MTN blot was prehybridised with ExpressHyb (prewarmed to 68°C) at 68°C for ~30 min. A denatured radiolabelled probe was added to ~5ml fresh hybridisation solution (at ~ 1-2×10⁷ cpm/ml) and hybridised with the MTN blot for 1 hr at 68°C. The blot was then rinsed several times in wash solution 1 at room temperature for 30-40 min, replacing the wash solution several times, followed by washing in wash solution 2 at 50°C for 40 min. Excess wash solution was removed and the blot exposed to X-ray film at -70°C for an appropriate length of time.

### 2.3.3 Subcloning DNA

#### 2.3.3.1 Preparation of Vectors and Inserts

Vectors and inserts were digested with the appropriate enzyme and checked on an agarose gel. The insert DNA was then purified from the gel using LMP agarose, by extracting twice with an equal volume of phenol, once with phenol:chloroform (1:1), once with chloroform and then precipitated with ethanol. The vectors were precipitated and phosphatased by incubating in 100μl volumes containing 1x CiP buffer and 4U phosphatase for 1 hr at 37°C followed by 55°C for 30 min. The vector DNA was then extracted once with phenol:chloroform (1:1) and ethanol precipitated. An aliquot of both vector and insert were run on an agarose gel to determine relative amounts compared to the 1Kb band of the size marker.

#### 2.3.3.2 Ligation of DNA into Vectors

Ligations were generally carried out in 20μl volumes containing 1x ligation buffer, 25ng phosphatased vector, 10-100ng insert DNA and 1U T4 DNA ligase and incubated at 16°C overnight. The ligation reaction was then inactivated by heating at 65°C for 15 min followed by ethanol/3M NaAc precipitation. The subsequent DNA pellet was
resuspended in 10μl ddH₂O. Ligations of PCR products into pCRScript were performed according to the kit instructions.

2.3.3.3 Transformation of Bacteria with Plasmid DNA
Electroporation (Shigekawa & Dower, 1988): 1-2μl of ligation reaction was mixed with 50μl of electrocompetent cells in an eppendorf tube, and then placed in a chilled electrocuvette. The cells were transformed using 1.8KV electrical shock and were then diluted in 450μl SOC and placed in a 37°C orbital shaker for 1 hr. The transformed cells were then plated out on LB agar plate containing the appropriate antibiotic.

Chemical Heatshock (Dagert & Ehrlich, 1974): 1-2μl of ligation reaction was mixed with 100μl of chemical competent cells in an eppendorf tube and incubated on ice for 30 min. The cells were then transformed by heat shock at 42°C for 45 sec and then chilled on ice for a further 2 min. The cells were then diluted in 500μl SOC and placed in a 37°C orbital shaker for 1 hr. The transformed cells were plated out on LB agar plate containing the appropriate antibiotic.

2.3.3.4 Preparation of Electrocompetent Bacterial Cells
A single, fresh colony of cells was used to innoculate a 5ml overnight LB culture and grown overnight. 1ml of this culture was used to innoculate 500ml LB, and was grown at 37°C until A₆₀₀ was in the range 0.45 - 0.55. The culture was then placed on ice for 20 min. After harvesting the cells at 4°C, 3,000g for 5min, the pellet was resuspended in 500ml of 10% glycerol and left for 20 min on ice. After spinning again, the pellet was resuspended in 50ml 10% glycerol and left for a further 20 min on ice. After a final spin, the pellet was resuspended in 2ml 10% glycerol, aliquoted into 40μl aliquots and stored at -70°C.

2.3.3.5 Preparation of Chemically Competent Bacterial Cells
A single, fresh colony of cells was used to innoculate a 5ml overnight LB culture and grown overnight. 1ml of this culture was used to innoculate 400ml LB, and was grown at 37°C until A₆₀₀ was in the range 0.37-0.38. The culture was aliquoted into 8x50ml prechilled falcon tubes and left on ice for 10 min. After spinning for 5 min (3,000g, 4°C),
each pellet was resuspended in 10ml ice-cold CaCl₂ solution. After spinning again, each pellet was resuspended in 10ml cold CaCl₂ solution and left on ice for 30 min. The cells were spun, and finally each pellet resuspended in 1.5ml of cold CaCl₂ solution. The cells were dispensed into 100μl aliquots and stored at -70°C.

2.3.4 Sequencing

For the majority of this project sequencing was performed manually, according to the method of Sanger et al. (1977), however, for the sequencing of certain clones an ABI 377 Automated DNA sequencer (Perkin Elmer) was used.

2.3.4.1 Manual Sequencing

Preparation of Double Stranded Template

Plasmid DNA was prepared as described previously. Approximately 1-5μg of DNA was made up to 20μl with ddH₂O and denatured by the addition of 2μl 2M NaOH, 2mM EDTA, and left for 15 min at room temperature. The DNA was then precipitated by the addition of 7μl ddH₂O, 3μl 3M NaOAc pH 5.3 and 75μl ethanol and left on ice for 15 min. The recovered pellet was washed in 70% ethanol and resuspended in 7μl ddH₂O.

Sequencing Reactions

Sequencing was performed using a Sequenase™ Version 2.0 Sequencing Kit. Annealing reactions were set up by adding 2μl 5x Sequenase buffer and 1μl primer (1pmol/μl) to the 7μl denatured template DNA. The reactions were heated to 65°C for 2 min and then allowed to cool to room temperature. Meanwhile, 2.5μl of each termination mix (dideoxy-nucleotides) were aliquotted into separate wells in a MicroSample plate. A bulk reaction mix was made for several reactions and 5.5μl, comprising the following, was added to each annealed primer-template and incubated at RT for 5 min; 2μl diluted labelling mix (see below), 1μl 0.1M DTT, 0.5μl [α³⁵S]-dATP, 0.5μl DMSO, 2μl Sequenase (diluted 1/8 with Sequenase dilution buffer). The microsample plate was heated to 37°C, and 3.5μl of each reaction was added to each termination mix. After 5 min, 4μl of Stop Solution was added to each termination reaction.
Polyacrylamide Gel Electrophoresis

Sequencing plates were cleaned with detergent and wiped with 70% ethanol. The tagged plate was siliconised by coating with dimethyldichlorosilane solution. The plates were sandwiched together separated by 0.4mm spacers and clamped along the edge. Acrylamide gels were generally 6% and made according to manufacturers guidelines. For 60ml gels, 40ml Sequagel diluent, 14.4ml concentrate, 6ml 10x TBE, 510μl 10% APS and 30μl TEMED were mixed and poured between the plates with a syringe. Sequences were denatured for electrophoresis by placing the microsample plate on a boiling block for 2 min. 3μl of each sequencing reaction were loaded onto the gel and the samples were run at 60W for 2-4 hr depending on the required length of gel run, with 1x TBE buffer as the electrophoresis buffer.

Autoradiography

The tagged plate was prised away from the gel and a sheet of 3MM chromatography paper was laid on top of the gel. The gel was lifted of the remaining plate, covered with clingfilm and left to dry for 1 hr at 80°C. The gel was then exposed overnight to Kodak BioMax film at room temperature.

2.3.4.2 Automated Sequencing

Double stranded DNA sequencing was performed using the ABI Prism Dye Terminator Cycle Sequencing ready reaction kit using AmpliTaq DNA polymerase-FS (Perkin Elmer). 1μg template DNA, resuspended in ddH₂O was placed in an Omnigene PCR tube along with ~5pmol primer and 8μl dye terminator reaction mix (total volume is 20μl). Cycle sequencing was performed in a Geneamp 2400 cycle sequencer using the heated lid function. For any primer the cycle sequence was performed with 20 cycles of the following; 94°C denaturation (20 sec), 53°C annealing (20 sec), 60°C extension (2 min). Following this, the DNA is precipitated at room temperature using 2μl 3M NaAc and 50μl ethanol, washed with 70% ethanol, and resuspended in 4μl of ABI loading buffer. The samples were then boiled and electrophoresed through a 10% polyacrylamide gel in TBE. Sequences obtained were then analysed using the computer software provided with the ABI 377.
2.3.5 Polymerase Chain Reaction

2.3.5.1 General PCR

DNA was amplified by PCR using some modifications to the standard conditions as described by Saiki et al. (1985, 1988). Reactions were carried out in 1x PCR buffer (enzyme dependant), 25mM each dNTP, approximately 50pmol each primer and 1U polymerase. Between 50-100ng of DNA was used as template. PCR reactions were the product was to be used for cloning and subsequent expression studies were generally amplified using Pfu high fidelity Polymerase, while for other applications Bioline Taq Polymerase was used, with each having its own specific 1x PCR buffer. PCR was generally performed in a Hybaid Omnigene with conditions set out as below:

1 Cycle :
Denaturation – 94-96°C, 5 min

20-30 Cycles :
Denaturation - 94°C, 1 min
Annealing - X°C (see below), 30 sec
Elongation - 72°C, 1 min

1 Cycle :
End Filling - 72°C, 10 min

The specific annealing temperature \( T_A - X°C \) for each primer pair was determined by taking the lower \( T_M \) of the primers and subtracting 5°C.

2.3.5.2 RT-PCR and Template Production for RACE

RT-PCR was performed according to the Marathon cDNA Amplification kit instructions. ~1µg of Human Placental Poly-A RNA was combined with 1µl of cDNA synthesis primer (10µM) in a total volume of 5µl and heated at 70°C for 2 min then cooled on ice. To this, 2µl 5x First-strand Buffer, 1µl dNTP Mix (10mM), 1µl ddH2O and 1µl AMV Reverse Transcriptase 20U/µl) were added. The sample was then incubated at 42°C for 1 hr. Second Strand synthesis was performed by taking the 10µl First Strand reaction and adding to this 48.4µl ddH2O, 15µl 5x Second Strand buffer, 1.6µl dNTP Mix (10mM) and 4µl 20x Second Strand Enzyme cocktail. The mix was incubated at 16°C for 1½ hr and then 2µl of T4 DNA Polymerase was added and the mix incubated for a further 45 min at 16°C. Following incubation, 4µl EDTA/Glycogen mix was added to terminate second strand synthesis. 100µl phenol: chloroform: isoamyl
alcohol (25:24:1) was added, the sample vortexed and centrifuged at 14,000g for 10 min. The top aqueous layer was removed to a new eppendorf and the extraction repeated. Again the top layer was removed and ½ volume of 4M Ammonium Acetate and 2½ volumes of room temperature 95% ethanol and the sample vortexed. The sample was spun immediately for 20 min at 14,000g, the pellet washed in 70% ethanol and respun for 10 min. The pellet was then air-fried and resuspended in 10μl ddH₂O.

Adaptors were ligated to the cDNA by taking 5μl ds cDNA and adding 2μl Marathon cDNA Adaptor (10μM), 2μl 5x DNA Ligation Buffer and 1μl T4 DNA Ligase (1U/μl) and incubating the sample overnight at 16°C. The ligase was then heat inactivated by incubating at 70°C for 5 min.

RACE PCR was carried out as for normal PCR using 1μl of cDNA-Adaptor product (cDNA-Adaptor ligation product diluted 1:25) as template for each 50μl reaction.

2.3.6 Protein Analysis

2.3.6.1 Protein Precipitation

Ice cold t richloroacetic acid (TCA) was added to the protein samples to a final concentration of 10%. The samples were incubated on ice for 30 min or more. Proteins were pelleted by centrifugation at >10,000g for 20 min at 4°C. The pellet was washed twice with ice-cold acetone followed by once with 70% ethanol then vacuum dried. The pellet was resuspended in 1xTE.

2.3.6.2 Measurement of Protein Concentration

Protein concentrations were determined using the method of Bradford (1976). A standard curve is generated using 0-40μl of 0.5mg/ml BSA with the sample being made up to 100μl with 0.15M NaCl. 1ml of Bradford reagent was added to the samples and vortexed. After being allowed to stand for 2 min, the OD₅₉₅ was measured. 1-2μl of test sample was analysed as above and the protein concentration determined from the standard curve.

2.3.6.3 SDS-Polyacrylamide Gel Electrophoresis

SDS-PAGE was adapted from the method of Laemmli (1970). Plates were cleaned with detergent and wiped with 70% ethanol. The plates were sandwiched together separated
by 0.75mm or 1mm spacers and clamped along the edge. Separating acrylamide gels were generally in the range of 8 - 12.5%, and were made according to the table below. Once poured, a layer of ddH$_2$O-saturated Isobutanol was placed on top to prevent air from blocking the polymerisation process. Once set, the Isobutanol layer was removed, and the stacking layer, consisting of 0.65ml Protogel (30% (w/v) Acrylamide, 0.8% w/v Bisacrylamide) solution, 1.25ml Tris-HCl/SDS (pH 6.8), 3.05ml H$_2$O, 25μl 10% APS and 5μl TEMED was poured. Following the addition of loading buffer, the protein samples were denatured for electrophoresis by placing the samples in a boiling block for 2 min. Approximately 30μl of each sample were loaded onto the gel. Gels were generally run for 2 hr keeping the voltage constant at 100V.

<table>
<thead>
<tr>
<th>Stock Solution (ml)</th>
<th>8</th>
<th>10</th>
<th>12.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protogel (30% (w/v) Acrylamide, 0.8% w/v Bisacrylamide)</td>
<td>6</td>
<td>7.5</td>
<td>9.375</td>
</tr>
<tr>
<td>4x Tris HCl/SDS (pH 8.8)</td>
<td>5.625</td>
<td>5.625</td>
<td>5.625</td>
</tr>
<tr>
<td>ddH$_2$O</td>
<td>10.875</td>
<td>9.375</td>
<td>7.5</td>
</tr>
<tr>
<td>10% Ammonium persulfate</td>
<td>0.075</td>
<td>0.075</td>
<td>0.075</td>
</tr>
<tr>
<td>TEMED</td>
<td>0.015</td>
<td>0.015</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Table 2 - Recipes for SDS-PAGE separating gels

2.3.6.4 Visualisation of Proteins

The front plate was prised away from the gel, and the gel gently placed into Staining solution for ~1 hr followed by destaining in Destain solution for ~ 2 hr, changing the Destain solution periodically. Following destaining, the gel was dried by placing on a
sheet of 3MM chromatography paper, covered with clingfilm and left to dry for 2 hr at 65°C.

2.3.6.5 Autoradiography

If necessary, the gel was exposed overnight, or longer, to Kodak BioMax-MR film at room temperature.

2.3.7 Immunochemistry

Protocols were generally carried out as described in *Harlow & Lane* (1988).

2.3.7.1 Immunoblotting

Immunoblotting was performed according to the method of *Towbin et al.* (1979). SDS acrylamide gels were blotted using a Biorad tank blotting system. The gel was pre-equilibrated for ~20 min in transfer buffer, while a PVDF membrane pre-wetted in methanol, or a Hybond C membrane pre-wetted in ddH₂O, was also pre-equilibrated in transfer buffer. The gel was laid on a piece of 3MM filter paper, and the uncovered side was overlaid with the membrane, which in turn was overlaid with another sheet of filter paper. The filter papers containing the gel and membrane were then "sandwiched" between two sponge pads and placed in a plastic support. The entire assembly was placed in the tank containing transfer buffer with the membrane positioned on the anode side of the gel. Transfer was achieved by applying a voltage of 100V for ~2 hr.

Following transfer the apparatus was disassembled and the membrane stored moist at 4°C until immunodetection. The proteins could be detected on the membrane by placing the membrane in Ponceau S solution for 5 min, followed by destaining for 2 min in 1x PBS.

2.3.7.2 Western Analysis

Prior to immunodetection, the membrane was washed 1× in PBS. The membrane was then blocked normally overnight in 10ml of the appropriate blocking buffer - normal or special - at room temperature rocking gently. Following blocking, the membrane was incubated with the primary antibody (diluted with blocking buffer) for 2-3 hr at RT. The membrane was then washed 3× for 5 min each with 10ml blocking buffer followed by
incubation with a species-specific alkaline phosphatase-conjugated secondary antibody (diluted 1:1000 in blocking buffer) for 45 min at RT. The membrane was then washed 3× with blocking buffer and twice with 10ml 1x wash buffer for 5 min each. Following incubation with the CDP-Star reagent (diluted 1:250 in 5mls Assay buffer) for 5 min at RT, the membrane was drained of excess developing solution, covered in cling film and exposed to X-Omat blue X-ray film. An initial exposure of 30 sec was used to determine further exposures.

2.3.7.3 Coupling Antibodies to Protein-A-Sepharose Beads

Approximately 100μl of Protein-A-sepharose beads were incubated with 1ml of antibody serum for 1 hr at RT with gentle rocking. The beads were washed twice with 10× volume of 0.2M sodium borate (pH 9.0), pulse-spinning in between to recover the beads. Crosslinking was achieved by resuspending the beads in 10× volume of sodium borate (pH 9.0) containing 20mM dimethylpimelimidate (DMP) and rotating for 30 min at RT. The crosslinking was stopped by washing the beads once with 0.2M ethanolamine (pH 8.0) and then rotating for 2 hr at RT in 10× volume of 0.2M ethanolamine. The beads were then recovered and stored at 4°C in PBS containing 0.1% azide. Prior to use the beads were washed 2× with 100mM glycine (pH 3.0).

2.3.7.4 Immunoprecipitation

To the 6-well dishes, 400μl of lysis buffer was added per dish and left on ice for 10 min. The cells were recovered by scraping and the supernatants passed 3 times through a 21 gauge needle. The lysates were centrifuged at 12,000g for 10 min at 4°C and the lysates recovered. 20μl of antibody-coupled Protein-A beads were washed 2x with lysis buffer and resuspended in 3× volume of lysis buffer. 5μl of Normal Rabbit Serum (NRS) was added and rolled for ~30 min at RT to preclear the lysate. The beads were washed 3× with lysis buffer, resuspended in 20μl lysis buffer and added to the cell lysates. The beads were rolled for 2-4 hr at 4°C, then recovered, washed 5× with lysis buffer then loaded on a SDS polyacrylamide gel, alongside a sample from the supernatant, and electrophoresed. The gel was immunoblotted then probed with the antibody of interest.
2.3.8 In vitro Protein Analysis

2.3.8.1 GST-fusion Protein Purification

Using a fresh bacterial colony, a 5ml LB+ampicillin culture was grown overnight at 37°C. 1ml of this culture was diluted 1:100 into 100ml LB + ampicillin and incubated for 2 hr at 37°C. 100μl IPTG (0.1M) was added and the culture grown for a further 4 hr at 37°C to allow induction of GST expression. The cells were pelleted and washed with 6ml 1x PBS. A 5μl aliquot was removed for SDS-PAGE analysis. After pelleting again, the cells were resuspended in 6ml STE+T containing lysozyme (500μg/ml) and left on ice for 20 min.

To the cells 30μl 1M DTT was added. Following sonication for 1 min at full power the samples were centrifuged for 10 min (10,000g, 4°C) to remove cell debris. To the supernatant, ~100μl Glutathione agarose beads (50% slurry in PBS) was added. After rolling for 1-2 hr the beads were recovered by pulse-spinning. The beads were washed 5x with PBS and stored in 1ml PBS at 4°C. A 50μl aliquot was removed for SDS-PAGE analysis.

If the GST fusion protein was thought to be insoluble, induction was carried out at room temperature or 30°C instead of at 37°C. Alternatively, different detergents were used to solubilise inclusion bodies instead of Triton X-100 including sarcosine, Tween-20 or SDS. The volumes were kept constant throughout for each detergent tested.

2.3.8.2 RNA Synthesis for In vitro Protein Synthesis

RNA was made from DNA which had previously been linearised by restriction digestion followed by treatment with Proteinase-K, phenol-chloroform (1:1) extraction and NaAc/Ethanol precipitation. Approximately 0.5μg resuspended DNA was combined with 5μl rNTP nucleotide mix (10mM each) in 1x transcription buffer containing 10mM DTT. 1U of RNAsin (RNAase inhibitor) and 1U of RNA polymerase (T3, T7 or SP6 depending on which construct was used) was added and the volume made up to 50μl.

The reaction was incubated at 37°C for 2 hr. Upon completion, the reaction was stopped by precipitation with 20μl 5M LiCl, 190μl ddH2O and 600μl Ethanol, leaving at -20°C overnight followed by centrifugation at 12,000g for 20 min. The pellet was resuspended and a fraction run out on a 1% agarose gel to estimate concentration.
2.3.8.3 In vitro Protein Synthesis

$^{35}$S-Methionine-labelled proteins were made using the In Vitro Express™ translation kit from Stratagene according to the manufacturers instructions. A reaction mixture containing 1-3µl of denatured mRNA, 2µl $^{35}$S-methionine and RNase-free ddH₂O to a total of 5µl was prepared, added to 20µl of thawed lysate and incubated in a 30°C waterbath for 1 hr. Following incubation, an aliquot was analysed by adding loading buffer, and subjecting the proteins to SDS-PAGE followed by exposure to Biomax-MR autoradiographic film.

2.3.8.4 TNT (Quick) Coupled Transcription/Translation Protein Synthesis

Approximately 1µg of supercoiled DNA was combined with 25µl TNT rabbit reticulocyte lysate, 2µl Reaction buffer, and 1µl amino acid mixture (minus methionine). 3µl $^{35}$S methionine, 1U of RNAsin (RNAase inhibitor) and 1U of RNA polymerase (T3, T7 or SP6 depending on which construct was used) was added and the volume made up to 50µl with ddH₂O. The reaction was incubated for 1-2 hr at 30°C.

For the Quick-coupled system, 1µg DNA was combined with 2µl $^{35}$S-methionine and ddH₂O up to 10µl. To this 40µl TNT T7 master mix was added and the reaction incubated for 60-90 min at 30°C. In both cases, a 2µl aliquot was analysed by SDS-PAGE followed by exposure to Biomax-MR autoradiographic film.

2.3.8.5 In vitro GST-affinity Capture Assay

An aliquot of glutathione-agarose-protein beads (containing ~0.5µg protein) was washed 1x with PBS, then 2x with TNT buffer and made up to 50µl. 1-2µl of in vitro translated protein was added and incubated for 1 hr at RT. The beads were washed 5x with TNT buffer, loading buffer was added, and the samples boiled for 3 min. Following electrophoresis, fixing, staining and destaining, gels were exposed overnight or longer to Biomax-MR autoradiography film.

2.3.8.6 Nuclear/Cytoplasmic Preparations

Cells were washed in PBS, scraped and collected. After spinning, the cells were resuspended in Buffer A (~1x10⁷cells/ml). An equal volume of Buffer B was added and the samples kept on ice for 10 min. The samples were then centrifuged at 2000g. The
supernatant was recovered and re-spun at 5000g for 5 min. The supernatant was again recovered and kept as the cytoplasmic fraction. The pellet from the original sample was resuspended in 100μl Buffer A and layered onto a 1.2ml sucrose cushion consisting of 0.8M sucrose in Buffer A. The sample was centrifuged for 10 min to pellet nuclei again at 5000g. The supernatant was aspirated off and the nuclei were washed in 1ml Buffer A. The integrity of the nuclei was checked by staining with Trypan Blue stain.

2.3.9 Yeast Analysis

Yeast were transformed using a method adapted from the LiTE protocol of Ito et al. (1983). The use of LiSORB was a modification made by Stephen Elledge (personal communication).

2.3.9.1 Yeast Cell Preparation

A 50ml overnight culture of yeast was inoculated from a single fresh colony and grown at 30°C in YPAD or minimal media (if selecting for a plasmid). The following day the culture was diluted into YPAD such that in 2-3 generations (4-5 hours) growth at 30°C the A600 of the culture is between 0.5 - 0.8. Once attained, the cells were harvested (5 min, 3,000g at 4°C), washed once in sterile ddH2O and finally resuspended in 50ml LiTE or LiSORB. After incubation without shaking for 30 min, the cells were harvested once more and resuspended in 625μl LiTE or LISORB. Cells were kept on ice until further use.

2.3.9.2 Small Scale Yeast Transformation

The carrier DNA mix was prepared by adding 800μl LiTE or LiSORB to 300μl boiled, ice-cooled, sheared salmon sperm DNA (10mg/ml). To 50μl of the carrier mix 1-2μl of miniprep/maxiprep DNA was added. For a negative control carrier mix alone was used. 40μl of prepared cells were added and the cell-DNA mixture was incubated at 30°C for 30 min. 500μl LiTEPEG solution was then added, and the mixture incubated for a further 30 min at 30°C. The cells were heat shocked at 42°C for 7 min then recovered by washing 2-3x with 1x TE, resuspended in 200μl 1x TE, and then plated on selective media.
2.3.9.3 Library Scale Yeast Transformation

The carrier DNA mix was prepared by adding 800μl LiTE or LiSORB to 300μl boiled, ice-cooled, sheared salmon sperm DNA (20mg/ml). To the carrier mix ~20μg of library maxiprep DNA was added. For a negative control carrier mix alone was used. The prepared cells were added and the cell-DNA mixture was incubated at 30°C for 30 min. 900μl LiTEPEG solution was then added for every 100μl cell-DNA mix (e.g. 18ml LiTEPEG for 2ml of cell-DNA mix), and the mixture incubated for a further 30 min at 30°C. Cells were heat shocked at 42°C for 7 min and were recovered by harvesting, resuspending in 100ml MM -L, -W, +H and shaking for 2-3 hr at 30°C. After harvesting and resuspending in 10ml MM -L, -W, -H the cells were plated on selective media. The transformation was either carried out in bulk, or in smaller aliquots, scaling the quantities down accordingly.

2.3.9.4 Colony β-Gal Assay

β-Gal assay buffer was prepared by adding 137μl β-mercaptoethanol and 1.25ml X-Gal (40mg/ml in DMSO) to 50ml Z-buffer. A circle of 3MM filter paper was placed in a petri dish and Assay buffer was added (6ml for 150mm plates and 2ml for 90mm plates) ensuring the paper wick uniformly soaked up the buffer. Colonies were lifted onto Hybond-N+ nylon filters ensuring that the filters and plates were orientated with respect to each other. The filters were immersed in liquid nitrogen for 30 sec then allowed to thaw. The filter was then placed colony side up onto the wet filter paper ensuring there were no trapped air bubbles. The filters were then incubated at 30°C for 5-6 hr or until a blue coloration had appeared.

2.3.9.5 Liquid β-Gal Assay

1ml of log phase cells were harvested after measuring OD_{600} by centrifugation. The media was aspirated and the cells washed in 1ml Z-buffer. The cells were then resuspended in 150μl Z-buffer containing β-mercaptoethanol (27μl per 10ml). 50μl chloroform and 20μl 0.1% SDS was added and the sample vortexed. 700μl of pre-warmed ONPG (1mg/ml in Z buffer + βME) and the samples were incubated at 30°C. The reactions were stopped by adding 0.5ml 1M NaCO₃. The samples were pelleted and
the OD_{420} was measured. The relative intensities of different samples was calculated and compared according to the equation:

\[
\text{Miller Units} = \frac{(A_{420} \times 1000)}{(A_{600} \times \text{Time (minutes)})}
\]

2.3.9.6 Recovery of Activation Domain Plasmids from Yeast

A single transformant colony was used to inoculate 5ml of selection media and grown overnight at 30°C. 1.5ml of the culture was placed in an eppendorf, spun for 5 sec at 12,000g in a microfuge to pellet the cells, and the supernatant discarded. The pellet was resuspended in 200μl of yeast lysis buffer. 200μl of phenol:chloroform (1:1) and 0.3g of acid-washed glass beads were added and the eppendorf vortexed vigorously for 2 min. The sample was then spun for 5 min at 12,000g, and the supernatant transferred to a clean eppendorf. The DNA was ethanol precipitated and resuspended in 20μl TE. The DNA was then used to transform HB101 bacterial cells and plated on minimal media plates selecting for leucine.

2.3.10 Cell Culture Techniques

2.3.10.1 Maintenance of Cells

All cells were cultured at 37°C in a 5% CO₂ atmosphere with the media used dependent on the cell line and were generally expanded 1 to 4.

U2OS and HEK-293 attached cells were grown in DMEM + GLUTAMAXII medium, supplemented with 10% Foetal Bovine Serum (FBS), 100IU/ml penicillin and 100UG/ml streptomycin. As they were attached, trypsin was used to detach the cells from the flask.

2.3.10.2 Freezing and Thawing of Cells

Cells were harvested (after detachment in the case of U2OS and HEK-293 cells) by centrifugation at 1000g for 5 min at RT. The cell pellet was resuspended at a density of 10^6 cells/ml in medium containing 20% FBS and 10% DMSO and aliquotted into cryotubes. The cells were then frozen at a rate of approximately 1°C/min at -70°C then transferred and stored in liquid nitrogen.
Cells were thawed by incubating the frozen cryotube at 37°C until defrosted, then the contents were added to 20ml media and incubated overnight. The media was then replaced and cells maintained as described above.

2.3.10.3 Transfection of Cells
Lipofectin Method: ~ 2.5x10^5 cells were seeded in a 6-well tissue dish and grown in DMEM + GLUTAMAXII overnight until 50-70% confluent. The cells were then washed once with PBS and 2ml OPTIMEM was added. Meanwhile, 0-1µg of DNA was precipitated and resuspended in 100µl serum-free OPTIMEM and in a separate eppendorf 0-15µl Lipofectin reagent was added to 100µl OPTIMEM. Both were allowed to stand at RT for 30-45 min. The two solutions were combined and were incubated at RT for 10-15 min. 0.8ml OPTIMEM (containing 4mM Glutamine) was then added to the DNA mix and used to overlay the cells. After incubation overnight, the media was replaced with normal media containing serum and allowed to grow for a further 24hr before harvesting.

Effectene Method: ~ 2.5x10^5 cells were seeded in a 6-well tissue dish and grown in DMEM + GLUTAMAXII overnight until 50-70% confluent. The cells were then washed once with PBS and 1.6ml of DMEM + GLUTAMAXII was added. In an eppendorf, 0-1.5µg DNA was added and the volume made up to 100µl with DNA condenser buffer EC. Enhancer was then added in a 8:1 ratio to the DNA (8µl per µg DNA) and the sample vortexed. Effectene reagent was then added in a ratio between 10:1 and 25:1 (Effectene: DNA) and vortexed. After standing for 5 min, 0.6ml DMEM + GLUTAMAXII was added and the sample overlayed onto the cells. After incubation overnight, the media was replaced with normal media containing serum and allowed to grow for a further 24 hr before harvesting.

2.3.10.4 β-Gal assay of Transfected Cells
Cells transfected with pCMV-βgal were harvested in 0.8 ml 1x reporter lysis buffer. After incubation on ice for 10 min, the cells were scraped and collected in an eppendorf. The lysates were prepared by centrifugation of the samples and collecting the resultant supernatants. Samples were assayed by taking a known volume of lysate and making this
up to 150μl with 1x reporter lysis buffer. 150μl of 2xβgal assay buffer was added, the samples vortexed and incubated at 37°C for 20 min – 3 hr. The reactions were stopped by the addition of 0.5ml 1M Na₂CO₃ and the OD₄₂₀ was measured.

2.3.11 Database Analysis

2.3.11.1 Computer Analysis

All computer programs for both DNA and protein sequence analysis can be accessed at the Human Genome Mapping Project (HGMP) WWW Home Page – http://www.hgmp.mrc.ac.uk/.
Chapter 3

Results

3.1 Work Preceding the BMI-1 Project

The main focus of ongoing work within the laboratory relates to a congenital syndrome, DiGeorge syndrome (DGS). DiGeorge syndrome is a developmental field defect which can involve abnormalities of the cardiac outflow tract, absent or hypoplastic parathyroid glands, absent or hypoplastic thymus gland and facial dysmorphism. DGS usually occurs sporadically, but it may be inherited in a dominant fashion with monosomy for 22q11-pter or interstitial deletions of 22q11 having been described (for review see Demczuk & Aurias, 1995). Within this locus several candidate genes have been mapped, one of which is HIRA.

HIRA encodes a WD40 repeat protein similar to yeast Hir1p and Hir2p (Lamour et al., 1995) and the yeast chromatin assembly factor (CAF) p60 subunit (Gutjahr et al., 1995). Hir1p and Hir2p act as transcriptional co-repressors that regulate cell cycle-dependent histone gene transcription (Sherwood et al., 1993), possibly by remodelling local chromatin structure (Kaufman et al., 1995). The human HIRA protein lacks an obvious DNA binding domain but is predicted to be a transcriptional regulator. Therefore, the identification of transcription factors with which HIRA interacts might indicate which developmental pathways require HIRA. From there, it may be possible to determine whether such pathways are important in the aetiology of DiGeorge syndrome.

In order to detect HIRA-interacting proteins, I performed several library screens using a GAL4-based yeast two-hybrid system using several different fragments of HIRA fused to a GAL4 DNA-binding domain vector as bait with only very limited success. More often than not, several potential positives were identified on the basis of the auxotrophic selection marker. However, on further analysis, these could not be confirmed using the colorimetric selection marker or retransformation experiments, and upon sequencing such positives were often found to encode globin gene family members (both sense and anti-sense strands). Thus with doubts over the suitability of the system as a method of detecting interactors of HIRA, and in order to provide the
possibility of a wider range of corroborative techniques, it was decided that an entirely different bait, a protein with a role in transcriptional regulation and with predicted interactors should be used, hopefully to act as a ‘positive control’ for the system. For this reason, BMI-1 was chosen. In addition, at the outset of the BMI-1 project (April 1996), BMI-1 was also a candidate gene for DGS due to its location at 10p13, at, or close to, the DGSRII locus (Alkema et al., 1993, Shapira et al., 1994).

3.2 Preliminary Experiments

3.2.1 Planning

As mentioned briefly in the Introduction, several interactors of both Psc and its mammalian murine homologue Bmi-1 have already been identified. It should, however, be noted that at the commencement of this project, whilst work was ongoing in other laboratories, no interactors had as yet been found. Subsequent work in these labs, using a GAL4-based yeast two-hybrid system similar to the one used in this project, led to these interactors of Bmi-1 being identified.

In order to investigate whether any regions within the human BMI-1 protein are responsible for mediating protein-protein interactions, it was necessary to use the full length BMI-1, a truncated fragment of BMI-1 or both. Conveniently, an EcoRI site is located approximately a third of the way into Bmi-1 (nucleotide 885 of BMI-1, Genbank Accession No. L13689) which, whilst encoding the majority of the protein, omits the N-terminal RING zinc finger. In addition this EcoRI site is in the same Open Reading Frame (ORF) as the Multiple Cloning Site (MCS) EcoRI site in the majority of GAL4-based DNA-binding domain vectors. At the time this project started the RING finger was thought to be a DNA-binding motif (Lovering et al., 1993) and as such it could interfere with subsequent yeast two-hybrid analysis. If this were the case then it would be advisable to remove the RING finger. The putative helix-turn-helix motif (Murre et al., 1989) is found in a number of proteins that are thought to act as transcription factors and is thought to mediate protein dimerisation. Most proteins that contain this motif also have a basic region of ~15 amino acids adjacent to the HTH domain and specifically binds DNA, but the homo- or heterodimerisation mediated by the HTH motif is independent of the DNA-binding ability. BMI-1 does not have an obvious basic region.
either side of the helix-turn-helix motif, and for this reason the motif is not thought to be involved in DNA binding.

![Diagram of the human BMI-1 protein illustrating the EcoRI fragment used in yeast two-hybrid analysis.](image)

**Figure 10** – *Representation of the human BMI-1 protein illustrating the EcoRI fragment used in yeast two-hybrid analysis.*

### 3.2.2 Subcloning

A 2330bp EcoRI fragment corresponding to the C-ter 191 amino acids of human BMI-1 was prepared from the BMI-1 (K562) cDNA cloned into pSP72. The fragment was subsequently cloned into the EcoRI site of the GAL4 DNA-binding domain expression vector, pGBT9. Restriction fragment analysis and dideoxy-sequencing was used to confirm that the fragment was inserted in the correct orientation and that the ORF was contiguous between the GAL4 DNA-binding domain and the BMI-1 fragment. It was then noted that expression from pGBT9 is relatively weak, as judged by western analysis, compared to another GAL4 expression vector, pAS2-1 (*Clontech techniques, 1995*), and so the same EcoRI fragment was cloned into pAS2-1 and likewise sequenced and restriction-map-confirmed.

### 3.2.3 Autoactivation Analysis

Once it was shown that the bait construct was correct, the bait was tested for autoactivation potential by transforming a suitable yeast strain and observing yeast growth on different selective media. It is sometimes found that proteins or fragments of proteins can autoactivate a reporter gene in the absence of the GAL4 activation domain,
when fused to the GAL4-BD. If this proves problematic, then two strategies can be employed: the first being to choose a different region of the protein that does not autoactivate the reporter in question; the second being to use reagents which inhibit the activity of the reporter gene product.

For this study the pAS2-1-tBMI-1 (pAS2-1-truncated BMI-1 fusion construct) was transformed into the yeast strain Y190. Y190 is a modified yeast strain, which is deleted for, or contains mutated non-functional genes for, the GAL4, LEU2, TRP1 and HIS3 genes, amongst other genes. The deleted GAL4 gene allows the use of the GAL4 based two-hybrid vectors, while the mutated biosynthetic markers allow for the selection of the DNA-BD vector (which contains a wild type TRP1 gene), and the DNA-AD vector (which contains a wild type LEU2 gene). The HIS3 gene as well as an engineered lacZ gene are under the control of the GAL(UAS) and thus the GAL4 gene product.

Using a basic Lithium Acetate (LiTE) transformation protocol (Materials and Methods Sections 2.3.9.1 and 2.3.9.2), Y190 was transformed with either pAS2-1-tBMI-1 or pAS2-1 and plated on minimal media, minus tryptophan, both with and without histidine. Growth was observed on plates containing histidine (-W, +H) (Figure 11A), as expected, since only the presence of the DNA-BD vector is being selected for, and not activation of the HIS3 gene. Unfortunately, very slow background growth was observed on the plates lacking histidine (-W, -H) (Figure 11B). Whether this can be attributed to ‘leaky’ HIS3 expression or a slight level of autoactivation of the histidine reporter gene by the Gal4BD from both the pAS2-1 vector and the pAS2-1-tBMI-1 construct is unknown. This growth was prevented through the addition of 3-Aminotriazole (3-AT) at a concentration of 50mM to plates lacking histidine (-W, -H, +50mM 3-AT) (Figure 11C). 3-AT is a structural analogue of histidine, and is able to act as a competitive inhibitor of histidine synthesis, reducing the level of histidine within the cell to a minimum. However, through the use of 3-AT, the sensitivity for detecting interactions is compromised, thus reducing the likelihood of detecting weak, but real, interactions.

It should also be noted at this point, that the bait protein constitutes a novel fusion whose properties within the cell may not exactly parallel those of the original unfused protein of interest.
Table 3 - Summary of yeast two-hybrid autoactivation analysis. pAS2-1 and pAS2-1-tBMI-1 were tested in Y190 in the presence and absence of histidine and 50mM 3-aminotriazole.

<table>
<thead>
<tr>
<th></th>
<th>Growth on plates +His</th>
<th>Growth on plates -His</th>
<th>Growth on plates -His +3AT(50mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAS2-1</td>
<td>✓ ✓ ✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>pAS2-1-tBMI-1</td>
<td>✓ ✓ ✓</td>
<td>✓</td>
<td>×</td>
</tr>
</tbody>
</table>
Figure 11- *Photograph of yeast (Y190) autoactivation analysis*. Y190 were transformed with either pAS2-1, or pAS2-1-tBMI-1, and plated on minimal media either: (A) containing Histidine; (B) lacking Histidine; or (C) lacking histidine in the presence of 50mM 3-Aminotriazole (3-AT).
3.3 Yeast Two Hybrid Interaction Assays

3.3.1 Strategic Considerations

The pAS2-l-tBMI-l construct was then used as bait to perform three separate screens against different libraries in order to detect possible interaction partners. In each case, in an effort to maximise the transformation efficiency, a two-step strategy was adopted with the bait plasmid being transformed first and yeast selected on minimal media plates minus Tryptophan. Subsequently, yeast containing the bait were transformed against the library plasmids. It has been previously documented that this method should greatly increase the transformation efficiency, rather than the less efficient, conventional, cotransformation method of introducing both plasmids into the yeast simultaneously. Both of the libraries used in this project were obtained from other laboratories where such two-hybrid analyses are routine (see Materials and Methods Section 2.1.10).

3.3.2 Library Screen 1

The truncated BMI-1 fusion construct was first used as bait in a small-scale library screen against a human pre-B-cell library cloned into the GAL4-activation domain vector pACT. It was thought that with BMI-1 having a role in hematopoiesis that this library would represent a suitable source to identify a subset of potential interactors. Using a basic Lithium Acetate (LiTE) protocol (Methods Section 2.3.9.1 and 2.3.9.3), Y190-pAS2-l-tBMI-l was transformed with ~20μg of library DNA, and a small aliquot of the transformation mixture was plated on plates minus tryptophan (-W), minus leucine (-L), plus histidine (+H) plus 3AT, in order to estimate the transformation efficiency (i.e. the number of yeast containing both plasmids).

An overall number of $6 \times 10^4$ cotransformants was observed (which equates to an efficiency of $3 \times 10^3$ cotransformants / μg DNA)

The remainder of the transformation was plated on media (-W, -L, -H, +3-AT) and allowed to grow. However, even after 5 days no colonies were apparent on these selective media plates.

The efficiency observed is fairly low. In contrast to E. coli, the maximum efficiency of transformation in S. cerevisiae is $\sim 10^4$ to $10^5$ transformants/μg DNA. In addition, because GAL4 is an important transcriptional regulator in yeast, it has the disadvantage
that experiments must be performed in \textit{GAL4\textsuperscript{-}} yeast strains to avoid background due to activation of the reporter by endogenous \textit{GAL4}. Such \textit{GAL4\textsuperscript{-}} strains are frequently less healthy and more difficult to transform than wild-type strains resulting in lower transformation efficiencies.

\textbf{3.3.3 Library Screen 2}

With the first library screen showing a relatively low transformation efficiency, efforts were made to try and increase the efficiency. The construct was used against the same B-cell library. As before, \(\sim 20\mu\text{g}\) library DNA was used to transform Y190-pAS2-1-tBMI-1, this time using a modified Lithium Acetate method (LiSORB = LiTE + 1M Sorbitol) (\textit{Materials and Methods Sections 2.3.9.1 and 2.3.9.3}). In addition, instead of performing the transformation in one large aliquot, the sample was divided up into 10 smaller aliquots, each transformed with 2\(\mu\text{g}\) of library DNA and an aliquot of the transformation mixture was plated on plates minus tryptophan (-W), minus leucine (-L), plus histidine (+H) plus 3AT, in order to estimate the transformation efficiency. The aliquots were pooled following transformation and the whole batch allowed to recover prior to plating.

An overall number of \(\sim 5 \times 10^5\) cotransformants was observed (an efficiency of \(2.5 \times 10^4\) cotransformants / \(\mu\text{g}\) DNA) This represents almost a ten-fold improvement on the previous transformation but still remains towards the low end of expected transformation efficiencies.

The remainder of the transformation was plated on media, (-W,-L,-H, +3AT) and allowed to grow for 3-4 days to select for potential interaction positive clones.

After 4 days growth, 85 colonies had appeared that were \textit{HIS3} positive, i.e. activated the histidine reporter gene. Of these 85, 68 activated the second marker, the \textit{lacZ} reporter gene, as detected through a filter \(\beta\)-galactosidase assay (\textit{Methods Section 2.3.9.4}). As positive controls for the \textit{LacZ} assay, Y190 containing either the pCL1 (containing the full-length wild-type \textit{GAL4}) or the combination of pTD1 (expressing a SV40 large T-antigen/\textit{GAL4-AD} hybrid) and pVA3 (expressing a murine p53/\textit{GAL4-DNA-BD} hybrid) was grown on the appropriate media and treated as for the test colonies.
These 68 positives were then restreaked on selection plates containing cycloheximide to test for the ability of the pACT insert-containing plasmids to autologously activate the histidine reporter. The pAS2-1 vector contains a gene, CYH2, which confers sensitivity to the antibiotic cycloheximide. By plating on cycloheximide, in the presence of tryptophan, colonies which no longer contain the pAS2-1-BMI-1 fusion vector but still retain the pACT library plasmid are selected. Any of these colonies that retain the ability to activate the histidine reporter gene in the absence of the pAS2-1-BMI-1 fusion vector are categorised as false positives. Thus only those colonies that don’t activate the HIS3 reporter are of interest for further analysis. Following this selection, 25 of the original 85 positives remained. These were subsequently recovered (Methods Section 2.3.9.6) in the HB101 bacterial strain (which is unable to synthesise Leucine, but can be complemented by the Leucine gene contained on pACT). Each yeast positive recovered was thus found to contain single inserts. Each yeast positive was retransformed into yeast, along with the pAS2-1-BMI-1 fusion vector to confirm the interaction. 14 positives survived this retransformation selection procedure.

These 14 clones were sequenced, and the DNA information obtained was used to search a variety of DNA and protein databases. The following matches were revealed:

Clone 4 - RNA Polymerase II subunit hRPB17
Clone 25 - H.sapiens partial cDNA sequence A7C08
Clone 35 - Human lymphocyte clathrin light chain A
Clone 42 - Human clone 5’ similar to coronin
Clone 54 - Human MCM6 minichromosome maintenance protein
Clones M2/M17 - Human clone 5’ similar to coronin
Clones 11, 16, 32, 39, 48, 53 and 55 - Murine X16 / Human Srp20 splicing factor

Examples of the BLAST output for selected clones are shown in Appendix 4. Before a more detailed examination of these clones is described, the third library screen will be outlined.

3.3.4 Library Screen 3

A second screen was carried out using the same pAS2-1-tBMI-1 bait construct, this time screening against a 9.5/10.5 dpc mouse embryo library cloned into the vector pVP16. The murine homologue of BMI-1 is expressed at this stage of development and
one could speculate that any candidates for interaction would be similarly expressed. pVP16 is another Activation Domain vector that, instead of using the GAL4 protein, uses the activation domain of the Herpes simplex viral protein 16, and is used in an analogous manner to pACT. The transformation was carried out as for the second library screen, using the improved LiSORB protocol in several smaller aliquots (Materials and Methods Sections 2.3.9.1 and 2.3.9.3), again using 20μg library DNA.

The rationale behind using a library from a different species lies in the fact that gene sequence and function are often conserved between species. This conservation is seen especially between human BMI-1 and mouse Bmi-1 with 97% (317/326 residues) identity existing at the protein level (Figure 12). Any positives which are identified in both species are more likely to be better candidates representing a real interaction than positives isolated from only one species. As the majority of false positives tend toward the random, the likelihood of the same false positive occurring in both species is predicted to be small.

Figure 12 - BLAST comparison of murine Bmi-1 against human BMI-1 at the amino-acid level. BLAST analysis was performed using the default page-settings for matrix BLOSUM62 at WWW reference (http://www.ncbi.nlm.nih.gov/gorf/bl2.html)
An overall number of \( \sim 1 \times 10^6 \) cotransformants was observed (representing an efficiency of \( \sim 5 \times 10^4 \) cotransformants / \( \mu \)g DNA). This, again, represents a significant improvement in the transformation efficiency with the observed efficiency in the middle of the expected range of \( 10^4 - 10^5 \) cotransformants/\( \mu \)g DNA.

After selecting for clones growing on selective media (-W,-L,-H, +3AT) for \( \sim 4 \) days, 98 clones were identified that were positive for the HIS3 selection marker. Of these, 74 were also positive for the LacZ reporter gene.

Instead of using cycloheximide selection and retransformation as previously, the 74 LacZ-positive clones were recovered in HB101. Using vector-specific flanking primers, the inserts were PCR amplified, subjected to agarose gel electrophoresis, and Southern blotted. It was observed that most colonies were found to contain single clones, however, on occasion some colonies were found to contain multiple clones, perhaps only one of which might encode a potentially interacting protein. DNA corresponding to positives identified from the second screen was also included in the Southern analysis. The filter of blotted PCR products was then probed with a pool of radiolabelled probes corresponding to 5 clones isolated from the second screen (one from each main category - 4, 16, 25, 54 and M2). Following autoradiography, the pooled probes were found to hybridise to two clones from the set of mouse embryo library positives, along with all 5 positive controls (Figure 13).
Figure 13 – Analysis of (mouse embryo) library screen 3 positives. The library screen used pAS2-1-tBMI-1 as bait in Y190. (A) - Agarose gel electrophoresis of third screen positives; (B) - Corresponding Southern Blot of third screen positives probed with BMI-1 positive human clones from the second screen. (DNA Sizes are indicated).
Upon sequence analysis and database searching against Genbank, one of the two clones (v45), revealed homology to a murine gene, mis5. mis5 is the murine homologue of the human protein MCM6 (also referred to as p105MCM). The second clone (v67), revealed homology to the previously identified EST of unknown function, A7CO8.

Several of the other clones isolated from the mouse library were sequenced and while some appeared to represent unknown ESTs, e.g. v86 = EST 437172, several others were known proteins, e.g. v15 – mitochondrial processing peptidase α-subunit precursor, v36 – heat shock transcription factor 2, and v6 - β-tubulin. Examples of the BLAST output of selected clones are shown in Appendix 4.

### 3.3.5 Summary of Library Transformations

<table>
<thead>
<tr>
<th>Library Screen</th>
<th>B-cell I</th>
<th>B-Cell II</th>
<th>Mouse embryo</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method</strong></td>
<td>LiTE</td>
<td>LiSORB</td>
<td>LiSORB</td>
</tr>
<tr>
<td>No. of Co-transformants</td>
<td>$6 \times 10^4$</td>
<td>$5 \times 10^5$</td>
<td>$1 \times 10^6$</td>
</tr>
<tr>
<td>No. of Histidine positives</td>
<td>0</td>
<td>85</td>
<td>98</td>
</tr>
<tr>
<td>No. of Histidine and LacZ positives</td>
<td>0</td>
<td>68</td>
<td>74</td>
</tr>
<tr>
<td>No. of positives after Cycloheximide selection and retransformation</td>
<td>0</td>
<td>14</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

**Table 4** – *Summary of yeast two-hybrid library screen transformation results*. pAS2-1-tBMI-1 was used as bait to screen three libraries in the yeast strain, Y190. n.d. – not determined. Cycloheximide selection was not performed for the third (mouse embryo) library screen.
### 3.3.6 Summary of Clones Obtained from Library Screens

<table>
<thead>
<tr>
<th>Library</th>
<th>pre-B-Cell in pACT</th>
<th>Mouse 9.5/10.5dpc embryo in pVP16</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>i) Nucleic acid-associated proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA Pol II subunit</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Splicing factor SRp20</td>
<td>11, 16, 32, 39, 48, 53, 55</td>
<td></td>
</tr>
<tr>
<td>MCM6 / mis5</td>
<td>54</td>
<td>v45</td>
</tr>
<tr>
<td><strong>ii) Structural proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clathrin light chain A</td>
<td>35</td>
<td>v54, v64</td>
</tr>
<tr>
<td>β-tubulin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronin</td>
<td>42, M2, M17</td>
<td></td>
</tr>
<tr>
<td><strong>iii) Unknown ESTs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A7CO8</td>
<td>25</td>
<td>v67</td>
</tr>
<tr>
<td>EST 438606</td>
<td></td>
<td>v2, v31, v63</td>
</tr>
<tr>
<td>EST 437172</td>
<td></td>
<td>v66, v86</td>
</tr>
<tr>
<td><strong>iv) Other proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleic acid-associated</td>
<td></td>
<td>All other clones</td>
</tr>
<tr>
<td>Structural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal transduction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5 – *Summary of yeast two-hybrid library screen interactors.* Interactors were identified using pAS2-1-tBMI-1 as bait to screen three libraries in the yeast strain, Y190.
3.4 Analysis of Results

3.4.1 False Positives

Most of the clones derived from the screens, with the overwhelming majority of them coming from the third library screen, were deemed to be false positives. Some false positives can be identified using basic biological principles. For example, within the cell, two proteins will not meaningfully interact with each other if their normal domains of activity reside in separate compartments, and likewise, from a developmental point of view, two proteins will not interact if they are not expressed at the same time and stage of development.

An example of this is the Human lymphocyte clathrin light chain A. Clathrin is the major protein found in intracellular vesicles and plays a major role in endo- and exocytosis. By being expressed in the two hybrid system, the clathrin protein will have been channelled into the nucleus (either via the Nuclear Localisation Signal in the GAL4 AD or through the effects of mass action) where BMI-1 is predominantly found, and not into the cytosolic compartment where it is normally found. From this point of view, it would appear that the tBMI-1-clathrin interaction seen in Y190 is a spurious result. Consequently, the priority of further characterising these clones is substantially lower than that of other clones encoding more plausible BMI-1 interactors.

A similar rationale could be argued for Coronin. Coronin is an actin-binding protein, originally identified in Dictyostelium discoidium (de Hostos et al., 1993). The protein sequence shows a homology to the β-subunits of trimeric G-proteins, leading to the hypothesis that coronin functions by regulating the transmission of signals from chemoattractant receptors via G proteins to the actin skeleton. In phagocytes, coronin tends to accumulate around the phagocytic vacuole and is associated with the oxidase specific protein p40phox (Grogan et al., 1997). Thus from a functional point of view it would appear that coronin is likely to be similar to clathrin, in that as a result of NLS targeting, it is expressed ectopically, resulting in a spurious interaction with BMI-1.

A useful WWW reference page exists that provides a regularly updated list of commonly observed false positives seen in the yeast two-hybrid system.

(Web Page: http://www.fccc.edu/research/labs/golemis/intro.html)
Using this page, several of the clones identified through the BMI-1 screen were also found to occur on this database. For example, various heat shock-associated proteins and cytoskeletal proteins commonly identified as false positives were seen with the mouse embryo screen, e.g. v36 - heat shock transcription factor 2, and v6 - β-tubulin. A suggestion to explain how the heat shock proteins might occur so frequently lies in their function. Frequently, cells protect themselves from stress through the actions of the heat shock proteins primarily because hsp interfere with the uncontrolled protein unfolding that occurs under stress. Their role as chaperones is to ensure that proteins fold correctly and as such use non-specific protein interactions to effect this. Such interactions, whilst biologically functional, are not necessarily significant in the context of searching for protein interactors for a given bait.

The use of the VP16 activation domain may also explain the relatively high number of false positives identified. The VP16 activation domain is relatively stronger than the GAL4 activation domain, and therefore, there is a greater likelihood that weak interactions, such as those associated with false positives, would be detected.

3.4.2 SRp20

One intriguing set of positives identified is the SRp20 splicing factor. Seven SRp20 encoding clones were identified, several differing at their 5' end, as shown in Figure 14. SR920 is a member of the highly conserved SR family of splicing regulators originally identified through a shared phospho-epitope recognised by the monoclonal antibody mAB104 (Roth et al., 1990, 1991). The SR proteins constitute a large family of nuclear phosphoproteins required for constitutive pre-mRNA splicing (Ge & Manley, 1990, Krainer et al., 1990, Fu & Maniatis, 1992, Zahler et al., 1992). Whilst their tissue distribution is ubiquitous, cell type variations in relative SR concentrations do occur, giving rise to the idea that these factors have a concentration-dependent effects on alternative splicing regulation.

3.4.3 MCM6

One protein fragment, identified from two different library screens from two different species encoded MCM6, a member of the minichromosome maintenance family of proteins. The alignment between the two clones, clone 54 and v45, and the human and
murine sequence is shown in Figure 15. The MCM proteins have been postulated as being replication licensing factors that are involved in ensuring that DNA replication occurs only once per cell cycle (for a review, see Kearsey et al., 1996). Further information regarding MCM6 and its possible role will be discussed later.

Upon examination of the two clones, 54 and v45 it was found that although the 5' ends of the two clones are not identical, the protein-coding reading frame associated with the both sequences is the same, and the regions that are associated with a potential BMI-1 binding activity are virtually identical, encompassing the C-Ter ~80 amino acids (Figure 15).

3.4.4 A7C08
A7C08 is an EST for which there is very little information. The DNA sequence, upon database searching, bears no significant homology to any known genes, and similarly, at the protein level when screened against the Swissprot and NBRF protein databases, there is no homology to any known proteins. What information there is about this clone, and subsequent characterisation will be discussed in Results Section 3.11.

3.4.5 Other ESTs
Within the database, there are several thousand ESTs (Expressed Sequence Tags) for which there is varying amounts of information available. In some cases they might correspond to DNA sequences of known genes, or fragments of genes, but in other cases they might represent completely novel coding regions.

Whilst some of the ESTs potentially encode proteins that represent real interactors of bait proteins, others are present that might encode false positives similar to those described in Results Section 3.4.1. For example, cytoskeletal proteins that are ectopically expressed and can interact in a similar manner to coronin or clathrin. Without detailed analysis of each individual EST such information cannot be obtained.
DNA sequence | Linker sequence |
---|---|
Clone | 53 GGC CAC GAA GGC CGG AGG AAA GCG GGA AGA CTC ATC GGA GCG TGT GGA |
| 11 GGC CAC GAA GGC C | GA AGA CTC ATC GGA GCG TGT GGA |
| 16 GGC CAC GAA GGC C | GA AGA CTC ATC GGA GCG TGT GGA |
| 39 GGC CAC GAA GGC C | GA AGA CTC ATC GGA GCG TGT GGA |
| 48 GGC CAC GAA GGC C | GA CTC ATC GGA GCG TGT GGA |
| 32 GGC CAC GAA GGC C | GT GGA |
| 55 GGC CAC GAA GGC C | GT GGA |
Protein Sequence | G H E G R K A G R L I G A C G |
| (R) |
| 53 TTT GAG CCG CCG CAT TTT TTA ACC CTA GAT CTC GAA ATG CAT CGT GAT |
| 11 TTT GAG CCG CCG CAT TTT TTA ACC CTA GAT CTC GAA ATG CAT CGT GAT |
| 16 TTT GAG CCG CCG CAT TTT TTA ACC CTA GAT CTC GAA ATG CAT CGT GAT |
| 39 TTT GAG CCG CCG CAT TTT TTA ACC CTA GAT CTC GAA ATG CAT CGT GAT |
| 48 TTT GAG CCG CCG CAT TTT TTA ACC CTA GAT CTC GAA ATG CAT CGT GAT |
| 32 TTT GAG CCG CCG CAT TTT TTA ACC CTA GAT CTC GAA ATG CAT CGT GAT |
| 55 TTT GAG CCG CCG CAT TTT TTA ACC CTA GAT CTC GAA ATG CAT CGT GAT |
| F E P P H F L T L D L E M H R D |
| Initiating Methionine |

Figure 14 – *Alignment of the 5' ends of the different SRp20-encoding fragments*. Each clone was independently identified from the B-cell library through a yeast two-hybrid screen with pAS2-1-tBMI-1 as bait.
Figure 15 – Comparison of DNA and protein sequences of the 3' region of human MCM6 (d84557), murine mis5 (d86726), clone 54 and v45. Vector linker sequences shown in blue (pACT), and Red (pVPI6).
3.5 Ranking of Clones in Order of Priority for Further Characterisation

As reported by Alkema et al. (1997a), murine Bmi-1 was reported to interact with another PcG gene product, murine polyhomeotic (Mph) as part of a large mammalian multiprotein complex. However, using the truncated BMI-1 clone, neither Mph nor either of its two known human homologues HPH1 or HPH2 (Gunster et al., 1997), depending on the species from which the libraries were derived, nor indeed any other Pc-G members, were identified as potential interactors.

However, several potential non-Pc-G interactors were identified from the above analysis. Of particular interest was the independent occurrence of the MCM6 fragment and the EST A7CO8 in two separate screens. Thus it was decided that efforts to characterise the interactions should concentrate mainly on the two positives 25, and 54 (cDNA A7CO8, and C-ter MCM6 respectively) since, being isolated from two separate libraries, they may be more likely to represent real interactions. In addition, clone 16 (SRp20) was also analysed further, in yeast at least, due to its level of representation in the B-cell library positives.

3.6 Further Yeast Analysis

Once positives were identified, it was necessary to retransform the bait and preys into yeast in order to test whether the interactions were i) reproducible and ii) whether both bait and prey were essential for reporter activation (as should be the case for a real interaction). In order to do this, as well as testing the bait against each of the isolated clones, the pAS2-1-tBMI-1 construct was transformed against pACT or pVP16 (not containing an insert), and conversely, the isolated clones in pACT or pVP16 were transformed against pAS2-1 (lacking the tBMI-1 fragment). These combinations were tested in three different yeast strains.

pAS2-1-tBMI-1 vs. pACT or pACT-16/25/54
pAS2-1 vs. pACT or pACT-16/25/54
3.6.1 Y190

Y190 was the strain used for the initial library screen. Thus, it was essential to confirm the interactions in the same strain. The above vector combinations were transformed into Y190 and plated on media (-W, -L, +/-H, +50mM 3-AT). Colonies which grew were restreaked and tested by filter assay for β-galactosidase activity. Table 5 below and Figure 16 show the results of this transformation experiment.

It can be seen from Figure 16 that both the bait and the preys together are necessary to activate transcription both of the reporters in this system.

3.6.2 Y166

Y166 is similar strain to Y190 in that it has similar HIS3 and LacZ reporter constructs. It differs however, in that the endogenous URA3 gene is partially deleted, and instead a full-length URA3 gene is present under the control of the GAL1 promoter so that it acts as an additional auxotrophic reporter construct. The same combinations as above were tested and were this time plated on media (-W, -L, +/-Ura, +/-H, +50mM 3-AT). The results obtained were very similar to those from Y190 with all three clones activating the reporters only in the presence of the pAS2-1-tBMI-1 construct.
Table 6 – *Summary of yeast two-hybrid retransformation experiments*. pAS2-1 and pAS2-1-tBMI-1 were tested against pACT-16/25/54 and pACT alone. Yeast (Y190) were plated on (-W, -L, +H, +50mM 3-AT) and (-W, -L, -H, +50mM 3-AT). Similar results were obtained in both Y190 and Y166.
Figure 16 - Photograph illustrating retransformation results in the yeast strain Y190. pAS2-1 and pAS2-1-tBMI-1 were tested against pACT-16/25/54 and pACT alone. The combinations were plated on: (A) –W, -L; and (B) –W, -L, -H +50mM 3AT. (C) β-Galactosidase filter assay of plate (B). (D) orientation of transformation combinations.
3.6.3 PJ69-4A

PJ69-4A is an improved yeast strain for use in two-hybrid experiments that has been designed to eliminate the selection of false positives (James et al., 1996). The reporters found in the majority of yeast strains, e.g. Y190 and Y166, are under the control of a single GAL4-responsive promoter, e.g. the GAL1 UAS for both Y190 and Y166. This can result in an increase in the background of false positives since some false positives can be promoter specific. PJ69-4A is designed so that by using a different GAL promoter for each reporter, the region of similarity between the promoters is limited to the degenerate GAL4 binding sites (GAL1→HIS3, GAL2→ADE2, GAL7→LacZ). Thus promoter specific false positives can be virtually eliminated. In addition, there are three reporters HIS3, ADE2 and LacZ, each of which is very tightly regulated whilst retaining sensitivity.

As the strain was new to our laboratory, nutritional growth requirements were not precisely known. To determine these requirements, the strain was restreaked on a variety of minimal media plates, each selectively lacking an amino acid, or adenine or uracil. Comparing the different plates revealed that, apart from the obvious selection requirements placed on the strain by the GAL4-responsive reporters, the only difference between PJ69-4A and Y190 was the requirement of PJ69-4A for methionine which was suggested from the genotype. Thus, all minimal media plates used for two-hybrid selection were supplemented with methionine.

The same vector combinations as above were tested and were this time plated on media (-W, -L, -Ade, +/-H, +5mM 3-AT (as described in James et al., 1996) – as opposed to 50mM 3-AT for Y190). The results obtained were very similar to those from Y190 with all three pACT-derived clones activating the reporters only in the presence of the pAS2-1-tBMI-1 construct. The results of this series of transformations are shown in Figure 17.
Figure 17 - Photographs illustrating retransformation results in the yeast strain PJ69-4A. pAS2-1 and pAS2-1-tBMI-1 were tested against pACT-16/25/54 and pACT alone. The combinations were plated on: (A) -W, -L; and (B) -W, -L, -Ade, -H +5mM 3AT. (C) Orientation of transformation combinations.
A disadvantage with this strain, however, is that the \textit{GAL7} promoter controlling the \textit{LacZ} reporter is inducible by permeabilization with liquid nitrogen so a filter-based $\beta$-Galactosidase assay could not be carried out. However this could be overcome by using a chloroform-based liquid $\beta$-Galactosidase assay. Colonies grown on $-W$, $-L$, $+H$ were grown in liquid media and assayed as described in \textbf{Material and Methods Section 2.3.9.5}.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure18.png}
\caption{Liquid $\beta$-Galactosidase assay results for the yeast strain, PJ69-4A. Assayed combinations are from Figure 17A, and were performed as described in Materials and Methods Section 2.3.9.5.}
\end{figure}
3.7 tBMI-1 and Mph1

3.7.1 Comparison of Truncated BMI-1 and Full-Length Bmi-1 Interactions

To test whether the truncated BMI-1 fragment is capable of binding Mph1, and conversely, whether the full length Bmi-1 is capable of binding 16, 25 or 54, Maarten van Lohuizen kindly sent yeast GAL4-based two-hybrid vectors corresponding to the GAL4-DNA-BD fused to the full length murine Bmi-1 (fBmi-1), in the vector pPC97, and the GAL4-AD fused to a C-ter portion of Mph1, in the vector pPC67, that had previously been used successfully in two-hybrid screens (Alkema et al., 1997a). The nutritional markers on these vectors were however different from the yeast two-hybrid vectors used in the library screens. In the system used in this project pAS2-1-tBMI-1 is selected by tryptophan and pACT/pVP16 by leucine. However, the van Lohuizen vectors were designed the other way around so that their GAL4-BD vector was selected by leucine and the GAL4-AD by tryptophan. This problem was overcome by swapping the selection markers of the van Lohuizen vectors so that they could be used in this study. The ‘new’ constructs were tested by plating Y190 transformed with either of the vectors on the appropriate media.

Satisfied that the derived constructs grew on their respective media, various combinations of vectors were tested in both Y190 and PJ69-4A. Both pAS2-1-tBMI-1 and the re-engineered full-length Bmi-1 were tested against pACT, 16, 25, 54 and the re-engineered Mph1 construct, and were plated on -W, -L, +/−H, +50mM 3-AT for Y190 and -W, -L, +/−Ade, +/−H, +5mM 3-AT for PJ69-4A.

Results obtained were similar for both Y190 and PJ69-4A and are described below (see Figures 19 and 20):

i) As expected, yeast containing both fBmi-1 and Mph1 grew on media selecting for the auxotrophic markers (Adenine and/or Histidine).

ii) As previously demonstrated, pAS2-1-tBMI-1 grew on marker-selective media with 16, 25 and 54 but not with pACT alone.

iii) Growth on selective media was NOT observed for combinations of fBmi-1 and either of 16, 25, 54 or pACT. Growth on media +His, + Ade confirmed the presence of both two-hybrid plasmids.

iv) Similarly, growth was not observed for the combination of pAS2-1 and Mph1.
v) Growth on selective media was observed for the combination of tBMI-1 and Mph1.

One noticeable observation was that yeast containing both the full-length Bmi-1 and Mph1 vectors grew more quickly than other transformations. For example, growth on -W, -L, -H was seen after 2-3 days for the full-length Bmi-1/Mph1 combination in Y190, whereas growth was only visible after ~4 days for the tBMI-1 construct along with 16, 25 or 54 when plated on similar media.

In addition, for combinations of pAS2-1-tBMI-1 with either 16, 25, 54, or Mph1, there was an obvious phenotype difference on the basis of colour. In this case, yeast containing the combinations of pAS2-1-tBMI-1 with either 16, 25, 54 showed a slight pink coloration compared to yeast containing fBmi-1 and Mph1 which remained white. A pink colour in yeast often reflects a defect in adenine metabolism with the darker pink the colour, the more severe the defect. It would seem that the tBMI-1 25/54/Mph1 combinations are not activating the Adenine reporter as strongly as the fBmi-1-Mph1 combinations. The colonies were not, however, dark pink or red, indicating some residual adenine metabolism was present.

For PJ69-4A, liquid β-Gal assays were carried out as described previously. Colonies grown on -W, -L, +H were grown in liquid media and assayed as described in Material and Methods Section 2.3.9.5. The liquid β-Gal assay results (Figure 20) tend to confirm the growth-based results with again, the fBmi-1-Mph1 vector combination showing the highest activity.

The protein expression levels of the GAL4-fusion constructs were not determined and so whether the different growth rates reflects the presence of stronger promoters on the fBmi-1 and Mph1 constructs driving stronger expression, or a stronger interaction between tBMI-1 and Mph1 compared to tBMI-1 and 25/54/Mph1 remains unknown.
Figure 19 - Photographs illustrating retransformation results comparing fBmi-1 and tBMI-1 interactions in the yeast strain PJ69-4A. pAS2-1-tBMI-1 and fBmi-1-pPC97 were tested against pACT-16/25/54, pACT alone and Mph1(C-Ter)-pPC67 and plated on: (A) -W, -L; and (B) -W, -L, -Ade, -H +5mM 3AT; (C) Orientation of transformation combinations.
Figure 20 - Liquid β-Galactosidase assay results comparing fBmi-1 and tBMI-1 interactions for the yeast strain, PJ69-4A. Assayed combinations are from Figure 19A, and were performed as described in Materials and Methods Section 2.3.9.5.
3.8 Further Characterisation of Protein Interactions Between BMI-1 and the Potential Interactors

As can be seen from the two hybrid screens carried out thus far (both as part of this project and from other laboratories), one common feature of the yeast two hybrid system is the recurrence of false positives. Thus, any potential real interactions that are identified using this system need to be distinguished from these false positives by being characterised and corroborated through other independent means. Two approaches were adopted, each which its limitations and applications, and each will be described as appropriate.

3.9 GST Affinity Capture Assays

3.9.1 The Underlying Principle

In this method, a protein of interest is purified as a GST fusion protein using glutathione agarose beads and is then used as a ‘bait’ to test for binding to a known or suspected ‘test’ protein. This protein is normally labelled by in vitro translation using radiolabelled methionine ($^{35}$S), leucine ($^3$H) or cysteine ($^{35}$S). Alternatively, the test protein can be labelled with $^{32}$P in vitro by incorporating a five-amino acid recognition motif recognised by a specific kinase. Beads with bound GST-fusion protein are incubated with test protein in the presence of a competitor protein to block non-specific interactions. Bound proteins are then washed, and the capture of test protein is determined by SDS-PAGE and autoradiography (see Figure 21).
Figure 21 – The principle underlying the GST-affinity capture method for determining protein-protein interactions. Glutathione agarose beads with bound GST-fusion protein (the ‘bait’) is incubated with radiolabelled ‘test’ protein in the presence of competitor protein to block non-specific interactions. Bound beads are then washed repeatedly and the capture of test protein by the GST-fusion protein is determined by SDS-PAGE and autoradiography.

3.9.2 Strategy I

The initial strategy adopted was to use a GST-BMI-1 fusion protein as a bait to affinity capture either of the three proteins that work was focussing on – the MCM6 C-ter fragment (clone 54), the unknown cDNA A7CO8 (clone 25) and the splicing factor SRp20 (clone 16).

Two GST-BMI-1 constructs were made. The first corresponded to the EcoRI truncated fragment used in the initial two hybrid screen cloned into the EcoRI site of the vector pGEX-4T-1, while the second was a full length BMI-1 fragment (generated through PCR with BMI-1-specific ORF-specific primers with SalI linkers (Materials and Methods Section 2.2.9)) cloned into the SalI site of the vector pGEX-4T-1. The linking fragments of the pGEX vector and of the BMI-1 sequence generated through
PCR were checked by sequencing and restriction mapping and the ORF was maintained.

Initial small scale (5ml) expression experiments were carried out to test the constructs in the bacterial strain BL21 (see Materials and Methods Section 2.3.8.1). Purification with glutathione-agarose beads was performed with aliquots of retained protein analysed by SDS-PAGE. The results of a typical experiment are shown in Figure 22A.

In each case, the pGEX-4T-1 construct alone, upon induction by IPTG, produced GST protein which was able to be successfully captured on glutathione-agarose. However, neither of the two different BMI-1 fusion constructs appeared to produce a band that would suggest that the fusion protein is being synthesised in a soluble form, but the absence of a band of a similar size to GST alone would suggest that a protein is being synthesised (Figure 22A). One problem that can occur with this procedure, especially with larger fusion proteins (e.g. >50KDa), is that of insolubility. Often fusion proteins form insoluble inclusion bodies which can only be resolubilised through the use of detergents and denaturing agents.

Two methods were employed to try to increase the solubility of the fusion proteins as described in Materials and Methods Section 2.8.3.1 (Frangioni & Neel, 1993). 

i) Induction of GST synthesis was carried out at lower temperatures (at room temperature and at 30°C), with no apparent success. Lowering the temperature of induction is reported to increase the solubility of insoluble proteins.

ii) A range of detergents were used, to try and solubilise the inclusion bodies. Using sarcosine, Triton X-100, Tween-20, or SDS the proteins remained insoluble and thus were not captured on glutathione agarose beads.

With hindsight, the expression of both of the proteins could have been detected via the use of anti-GST monoclonal antibodies on western blots of cell lysates pre- and post induction with IPTG, although at the time these experiments were being performed appropriate reagents were not available within the laboratory.

3.9.3 Strategy II

With the first strategy being unsuccessful, it was decided to attempt a different approach. In this case, GST fusion proteins of the three principle proteins of interest
from the yeast two-hybrid library screens – encoded by clones 16, 25 and 54 - were to be used as bait in an attempt to capture \textit{in vitro} translated BMI-1.

The relevant inserts in the Activation Domain vector (pACT) were PCR amplified using pACT- specific primers (\textbf{Materials and Methods Section 2.2.9}) and the PCR products were subcloned into the vector pCRScript. The large MCS provided by this vector facilitated subsequent insert-shuttling steps into a variety of different vectors. \textit{EcoRI} fragments containing the inserts were prepared and cloned into the \textit{EcoRI} site of the vector pGEX-3X, with the fragments checked by sequencing and restriction mapping to ensure that the orientation and ORF were maintained.

Again, small-scale cultures were set up and using a mild lysis buffer (STE+T) (\textbf{Materials and Methods Section 2.8.3.1}) protein bands of approximately the expected size were detected by SDS-PAGE for all of the fusion proteins and GST alone with the exception of GST-16 (i.e. GST alone, GST-25 and GST-54) (\textbf{Figure 22B}). Induction of expression at 30°C as opposed to 37°C improved the yield of the three soluble proteins, but despite the use of different lysis buffers and temperature inductions the GST-16 remained insoluble.
Figure 22 - SDS-PAGE analysis for GST-agarose purification of different pGEX-BMI-1 constructs: (A) - GST purification of GST, GST-tBMI-1 and GST-fBMI-1; (B) - GST-purification of GST, GST-25 and GST-54. (P - Pre-induction with IPTG, I - Post-induction with IPTG, B - Bead capture using glutathione agarose) ((A) - 10% Acrylamide stacking gel, (B) - 12% Acrylamide separating gel, Marker Sizes in KDa).
3.9.4 In vitro Translation of BMI-1

As part of the affinity capture assay, the GST-X construct is used as bait to capture a test protein, in this case BMI-1, normally in the form of a radiolabelled protein. Two methods were applied to obtain radiolabelled protein.

3.9.4.1 Translation from Prepared mRNA Templates

The first approach was to transcribe RNA from a suitable template using the SP6 bacteriophage RNA polymerase (Materials and Methods Section 2.3.8.2), and then translate the RNA using a rabbit reticulocyte translation system allowing the incorporation of $^{35}$S-methionine (Materials and Methods Section 2.3.8.3). RNA and labelled protein from two different linearised constructs was prepared:

a) A fusion construct consisting of the BMI-1 ORF fused to the C-ter of the Herpes Simplex VP16 activation domain and;
b) The VP16 activation domain alone.

The RNA produced was of good quality, as judged by agarose gel electrophoresis, and did not appear to be significantly degraded. SDS-PAGE analysis, however, revealed that the proteins produced consisted of numerous bands (Figure 23A) – too many to allow confident assignments of bands of interest, and thus these preparations were of little use in the affinity capture assay.

3.9.4.2 Coupled Transcription-Translation

As an alternative to the two-step process, a one-step coupled system became commercially available and was used with the two constructs as described above (Materials and Methods Section 2.3.8.4). With the SP6 coupled in vitro transcription-translation system, the translated protein was of higher quality than previously and allowed the bands of interest to be identified over other background bands also present, although the yield was lower than expected (Figure 23B).

In addition, both full-length BMI-1 and the truncated EcoRI BMI-1 fragments were cloned into a HIS-tag vector pET28-a which has been optimised for high quality transcription and translation. Using this vector, and the T7 Quick-coupled™ in vitro
transcription-translation system, high quality translation products were obtained, with very little background incorporation (Figure 23C).

Figure 23 – SDS-PAGE comparison of different methods of synthesising radiolabelled BMI-1. (A) – in vitro transcription from DNA templates (pVP16 and pVP16-BMI-1) followed by translation. (B) - coupled transcription-translation from pVP16 and pVP16-BMI-1. (C) - Quick-coupled transcription-translation from full-length BMI-1 and tBMI-1 in pET28-a. (12% Acrylamide separating gel, Marker Sizes in KDa).
3.9.5 GST-Affinity Capture Assay

With GST-fusion proteins available corresponding to clones both 25 and 54, and in vitro translated BMI-1 also available, the next step was to perform the affinity capture assay (Materials and Methods Section 2.3.8.5).

Approximately equal amounts (~0.5 μg) of the GST proteins (as determined by Coomassie staining intensity) attached to glutathione-agarose beads were used as baits to capture in vitro translated BMI-1 in TNT buffer following incubation with BSA to prevent non-specific binding. Beads were then washed several times to remove any unbound radiolabelled proteins, and analysed by SDS-PAGE and autoradiography.

3.9.5.1 His-tag Derived Constructs

With the full-length BMI-1 construct as the test protein, both of the GST-fusion proteins successfully captured BMI-1 (Figure 24B), whilst GST protein alone showed minimal binding. Similar results were also seen using the truncated BMI-1 fragment (Figure 24C) with both of the GST-fusion proteins successfully capturing tBMI-1, and GST protein alone demonstrating minimal binding.

3.9.5.2 VP16 Derived Constructs

Using the VP16-based constructs, similar results were obtained as for the His-tag derived constructs. Using the VP16-full-length BMI-1 construct as the test protein, both of the GST-fusion proteins successfully captured BMI-1 (Figure 25B), whilst GST protein alone showed minimal binding. As a negative control the VP16 protein was tested, with the observation that none of the bait proteins significantly pulled down VP16 alone (Figure 25A).

One significant observation was that using all three BMI-1-derived constructs, it was apparent that the captured protein signal was stronger for GST-25 when compared to GST-54. As the input of radiolabelled protein was identical for each set of experiments, one interpretation is that GST-25 had a higher affinity for BMI-1 than GST-54. This is also reminiscent of the stronger β-Galactosidase reporter activity seen for clone 25 in the yeast two-hybrid system over clone 54.
Figure 24 – SDS-PAGE analysis for the GST-Affinity capture of $^{35}$S-methionine labelled BMI-1 by GST fusion proteins. (A) – Coomassie staining of bait proteins, GST, GST-25 and GST-54. (B) – Affinity capture assay using His-tag fBMI-1 as test protein. (C) - Affinity capture assay using His-tag tBMI-1 as test protein. (12% Acrylamide separating gel, Marker Sizes in KDa).
Figure 25 – SDS-PAGE analysis for the GST-Affinity capture of $^{35}$S-methionine labelled VP16 and VP16-BMI-1 by GST fusion proteins. (A) Affinity capture using VP16 as test protein; and (B) Affinity capture using VP16-BMI-1 as test protein. Both proteins were tested against GST alone, GST-25 or GST-54 fusion proteins. (12% Acrylamide separating gel, Marker sizes shown in KDa).
Table 7 – *Summary of GST-affinity capture assay results*. GST alone, GST-25 and GST-54 were used to capture $^{35}$S-methionine-labelled VP16, VP16-BMI-1, His-tag fBMI-1 and His-tag tBMI-1.

<table>
<thead>
<tr>
<th></th>
<th>GST alone</th>
<th>GST-25</th>
<th>GST-54</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP16</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>VP16-BMI-1</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>His-tag fBMI-1</td>
<td>×</td>
<td>✓✓</td>
<td>✓✓</td>
</tr>
<tr>
<td>His-tag tBMI-1</td>
<td>×</td>
<td>✓✓</td>
<td>✓✓</td>
</tr>
</tbody>
</table>

$x$ - No interaction observed; ✓ - moderate interaction observed; ✓✓ - strong interaction observed

As can be seen from these results, summarised in Table 7, there does appear to be a significant interaction between both GST-25 and GST-54 with BMI-1. This interaction is seen to occur with both full-length BMI-1 and the truncated BMI-1 fragment. The truncated BMI-1 interaction corroborates the interaction seen in the yeast two-hybrid experiment. However, the interaction detected between the GST-fusion proteins and the full-length BMI-1 (biologically more relevant) was not replicated in the yeast two-hybrid assays.
3.10 Immunoprecipitation Studies of BMI-1 and Potential Interactors

3.10.1 Introduction

The underlying principle of immunoprecipitation is that within the cell, proteins are often found in large multimeric complexes which, depending on many factors, can or cannot be isolated. Using simple immunoprecipitation, specific antibodies can be used to study target antigens for example, individual members of these complexes. However, a more useful application is that of co-immunoprecipitation, providing antibodies are available, to study protein partners and interactions that occur within these complexes. A summary of the co-immunoprecipitation strategy is shown in Figure 26.

---

Figure 26 - The principle underlying the (co-)immunoprecipitation method for analysing protein-protein interactions.
A target-protein-specific antibody is used to immunoprecipitate (Figure 26, step i) the target protein of interest from a complex protein lysate. Depending upon several parameters, for example the salt concentration in the lysis buffer, the target protein may be associated with several other interacting proteins. At the low concentrations of antigen used most antibodies are non-precipitating, and so a sandwich reagent is used to isolate the antibody-antigen complex (step ii). Two of the most common reagents are Staphylococcus protein-A or Streptococcus protein-G coupled to sepharose. Following several washes to remove non-specific proteins (step iii), the complexes are subjected to SDS-PAGE (step iv) and western blotting (step v). Finally, an antibody either directed to the target protein used initially, or another protein of interest is used to detect its presence on the western blot.

3.10.2 Strategies for Co-immunoprecipitation

Two different approaches were used in this project.
- The first was a direct approach, using antibodies raised against BMI-1 and full-length MCM6 for co-immunoprecipitation studies. As 25 represents a potentially novel protein, no antibodies were immediately available.
- The second approach was slightly more indirect. As only small quantities of the α-BMI-1 antibody were available, epitope-tagged fusion protein constructs were generated for BMI-1, full-length MCM6 and clone 25 (using the reading frame predicted from the yeast two-hybrid vector). These were cloned in eukaryotic expression vectors under the control of the constitutive CMV (Cytomegalovirus) promoter, and upon transfection into mammalian cells were tested for expression of the proteins of interest. Commercial antibodies are available against the epitope tags (e.g. the HA and FLAG tags) and can be used to probe protein functions in a similar fashion to antibodies raised against native proteins.

3.10.3 Characterisation of Antibodies

3.10.3.1 Preparation of Antibodies

Since only limited quantities of antibodies directed against Bmi-1 and MCM6 were available from other laboratories, it was decided to raise antibodies against protein-specific peptide sequences. For this, a company based in the United States (Research Genetics) was used. Before any co-immunoprecipitation experiments can be carried out,
one has to be confident that the antibodies to be used are specific for the proteins of interest. One of the main problems with anti-peptide antibodies is that whereas the antisera raised this way will often recognise the peptide to which they were designed, the antisera may not recognise the native protein. Assays that need or benefit from anti-native protein antibodies, such as immunoprecipitation, will only succeed when the peptide sequence on the surface of the peptide-carrier conjugate reflects the structure of the native molecule.

Peptides were designed and linked to either Keyhole-limpet Hemocyanin (KLH) or MAP to act as a carrier and were used to immunise two rabbits for each peptide. Bleeds were taken at 4, 8 and 10 weeks and serum prepared from each of these bleeds were analyzed together with a sample taken from the rabbit prior to inoculation (preimmune).

Two peptides were designed to raise antibodies against the Human BMI-1 protein and one against the Human MCM6 protein. The sequence for the MCM6 sequence was modified from a peptide sequence previously published (Holthoff et al., 1998).

With two rabbits immunised for each peptide, a total of 6 sets of sera were to be analysed. This was done in two stages. Initially, total cell lysates, from either 293 or U2OS cells, were subjected to SDS-PAGE, western blotted (Materials and Methods Section 2.3.7.1 and 2.3.7.2) and the resultant filters cut into strips. The serum from the rabbits was then used to detect the relevant proteins on these blots. To test whether any bands detected correlated to the correct proteins (at least, by the size of the bands detected), the two previously characterised antibodies, F6 anti-Bmi-1 (Alkema et al., 1997a) and anti-MCM6 (Materials and Methods Section 2.1.5) were used as positive controls. Antibodies detecting a protein of the correct size were then used to detect a sample of the in vitro translated protein.
3.10.3.2 Initial Characterisation of BMI-1 Antibodies

BMI-1 Nter (HRTTRIKITELNPH)g-MAP
Rabbit No. 43582 & 43583

BMI-1 Cter QSSFANRPRKSSVNGSSA (KLH)
Rabbit No. 43584 & 43585

Figure 27 – Simple representation of BMI-1 with the peptides to which antibodies were raised shown.

Preimmune, 4 and 8 week sera were tested on 293 total cell lysates prepared in RIPA buffer.

43582 - Both the 4 and 8 week sera detected one band of approximately 60KDa which was also detected in the preimmune serum.

43583 - Both the 4 and 8 week sera detected a band of approximately 30 KDa which was not detected by the preimmune serum.

43584 - Both the 4 and 8 week sera detected a similar band of ~30KDa as 43583 and in addition a larger band of approximately 60KDa, neither of which were detected by the preimmune serum.

43585 - The 4 week serum detected several bands and varying sizes (between 30 and 80KDa) several of which were also detected using the preimmune serum.
When 43583 and 43584 were compared with the F6 α-BMI-1 control antibody, neither of the polyclonal antibodies appeared to detect a band that correlated with the band detected by F6 (Figure 28). This suggests that the polyclonal antibodies are not detecting BMI-1, and consequently are of little use.

Figure 28 – Western analysis of peptide-raised polyclonal antibodies (43583 and 43584) and the characterised F6 αBMI-1 polyclonal antibody. All were tested on 293 total cell lysates and detected with AP-linked anti-rabbit IgG. (12% Acrylamide separating gel, Marker sizes shown in KDa).
3.10.3.3 Initial Characterisation of MCM6 Antibodies

**MCM6 antibody**  PGAGSQHLEVRDEVAEKC

**Rabbit No.**  83244 & 83245

![Diagram of MCM6 domain](image)

**PGAGSQHLEVRDEVAEKC**

Figure 29 – *Simple representation of MCM6 with the peptides to which antibodies were raised shown.*

Preimmune and 8 week sera were tested against western blots of U2OS total cell lysates prepared in RIPA buffer and mouse 9.5/10.5 dpc embryo total lysates also in RIPA (Materials and Methods Section 2.3.7.1 and 2.3.7.2).

**83244** - A band of approximately 100KDa was detected by 83244 in U2OS extracts, along with a slightly smaller band in the mouse embryo extracts, indicating some cross reactivity.

**83245** - This antibody failed to detect any significant bands in either U2OS extract or mouse embryo extract.
The band of ~100KDa detected by 83244, when compared against the control MCM6 antibody, is of a similar size band, suggesting that the antibodies are recognising the same protein (Figure 30).

Figure 30 – Western analysis of peptide-raised polyclonal antibodies (83244 and 83245) and the characterised αMCM6 polyclonal antibody. All were tested on U2OS (U) total cell lysates and mouse 9.5/10.5 dpc embryo lysates (M) and detected with AP-linked anti-rabbit IgG. (12% Acrylamide separating gel, Marker sizes shown in KDa).
3.10.3.4 Further Characterisation of MCM6 Antibodies

The next step in the characterization of 83244 as a possible αMCM6 antibody was to test it against \textit{in vitro} preparations of MCM6. MCM6 was \textit{in vitro} translated using a rabbit reticulocyte lysate T7 Quik-coupled transcription-translation kit using ‘cold’ methionine. A negative control transcription-translation reaction was also set up. Aliquots of the lysates were then subjected to SDS-PAGE and western blotted. 83244 was then used to probe the western blot. As a positive control, MCM6 was \textit{in vitro} translated using $^{35}$S-Methionine and an aliquot was run on the same gel.

As Figure 31 shows, a band of approximately 100KDa was detected by 83244 in MCM6 cold translations and 293 cell extracts that was not detected in the negative control translations. In addition, this band is of a similar size to that in the lane corresponding to $^{35}$S-Methionine-labelled MCM6, suggesting that the band detected by 83244 is indeed MCM6.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure31.png}
\caption{Western blot of 83244 anti-MCM6 antibody tested against \textit{in vitro} translated MCM6. (A) - Western blot of ‘cold’ \textit{in vitro} translated MCM6 and 293 total cell lysate probed with 83244 and detected with AP-linked anti-rabbit IgG. (B) - $^{35}$S-Methionine labelled MCM6. (10\% Acrylamide separating gel, Marker sizes shown in KDa).}
\end{figure}
With 83244 appearing to detect the correct protein, as a control, the peptide used to immunise the rabbit was used to block the antibody, to determine whether or not binding to the MCM6 protein was specific and could be abolished.

83244 was incubated either with, or without an excess of blocking peptide prior to being used on western blots of U2OS and 293 cell lysates. Upon detection with anti-rabbit AP-linked antibody it was observed that addition of the peptide prevented detection of the MCM6 band in both extracts. As expected, the MCM6 was detected in the control lanes without blocking peptide (Figure 32).

![Figure 32](image)

**Figure 32** - *Western analysis showing the effect of blocking 83244 with peptide on the detection of MCM6*. 83244 was detected with AP-linked anti-rabbit IgG. (10% Acrylamide separating gel, Marker sizes shown in KDa).
It was then determined whether MCM6 could be immunoprecipitated by 83244. The antibody was incubated with total cell lysates (in RIPA buffer – optimised for immunoprecipitation) from U2OS and 293 cell lines, both in the presence and absence of blocking peptide. The immunoprecipitation was performed as described in Materials and Methods Section 2.3.7.4 and the samples subjected to western blotting. As a positive control, a sample of the lysate was also included. When probed with 83244, no immunoprecipitation was observed from either cell line. MCM6 was detected in the cell lysate as expected. This suggests that the antibody, whilst being suitable for western blot analysis, is not suitable for immunoprecipitation.

3.10.4 Transfection Studies

3.10.4.1 Optimisation of Transfection Efficiency

With a suitable antibody available only for MCM6 western analysis, constructs were generated to allow epitope tagging of the relevant proteins. However, prior to testing the constructs generated in cells, conditions for transfection were optimised. As transfection efficiency can vary widely between different cell types, and even within the same type dependent on passage number, the conditions for each cell line need to be determined empirically. For this, a plasmid constitutively expressing the LacZ gene product (pCMV-β-galactosidase) was used as a reporter. Two methods of transfection were used, the first, a liposomal lipid based system, Lipofectin, available from Stratagene, and the second, a non-liposome lipid based system, Effectene, available from Qiagen.

U2OS cells were transfected with the pCMV-β-gal construct using either of the two transfection methods as described in Materials and Methods Section 2.3.10.3. The absolute amounts of DNA were varied along with the DNA : lipid ratios. Cell lysates were prepared and β-galactosidase activity was assayed using a simple colorimetric assay (Materials and Methods Section 2.3.10.4), and a basic linear relationship between transfection efficiency and β-galactosidase activity was assumed.
As can be seen from Figure 33, the Effectene reagent transfected U2OS cells with a far higher efficiency than the lipofectin reagent. Using 1µg of reporter DNA, with an Effectene:DNA ratio of 10:1, there is an approximately 10-fold greater reporter activity than the highest Lipofectin results for the same quantity of DNA. Thus, it was decided for future transfections to use 1µg of total DNA with an Effectene:DNA ratio of 10:1.

**Figure 33 – Comparison of different lipid-based methods of transfecting U2OS cells with pCMV-β-Gal.** β-Gal reporter activity was measured as described in Materials and Methods Section 2.3.10.4.
3.10.4.2 Analysis of HA-BMI-1 Expression Constructs

A eukaryotic construct expressing HA-tagged BMI-1 was kindly provided by Maarten van Lohuizen. The construct was tested by transfecting into U2OS cells and probing western blots of resultant lysates with HA antibody. In addition, the HA antibody was used to immunoprecipitate the HA-BMI-1 and western blots again probed with HA antibody.

Western analysis of the HA-BMI-1 transfected lysates revealed a single band of the correct size that was absent in mock transfected lysates. Similarly, using an anti-HA monoclonal antibody, the HA-BMI-1 protein was successfully immunoprecipitated.

3.10.4.3 Analysis of FLAG-25/MCM6 Expression Constructs

Eukaryotic constructs expressing FLAG-tagged clone 25 and MCM6 were generated by cloning either the PCR product corresponding to the full-length MCM6 ORF (as opposed to the C-ter fragment encoded by clone 54) or a Xhol fragment corresponding to clone 25 into the FLAG-epitope expression vector, pCMV-Tag1. The construct was tested by transfecting into U2OS cells and probing western blots of resultant lysates with aFLAG antibody.

The constructs were checked by in vitro translation using a coupled transcription-translation system (Materials and Methods 2.3.8.4) and bands corresponding to the expected size were detected for both constructs, indicating that the inserts had been cloned as intended. However, initial results based on western analysis on cell extracts failed to detect significant expression of either FLAG-25 or full-length MCM6. Anti-FLAG antibodies from different companies were obtained and tested on the extracts again, without significant expression detected.

It was then decided to shuttle the FLAG-25/MCM6 fragments from pCMV-Tag1 into another eukaryotic expression vector, pcDNA3. Once more the constructs were checked by in vitro translation with the expected results. The pcDNA3 constructs were then transfected side-by-side along with the pCMV-Tag1-based constructs into U2OS, and western blots of the lysates again probed with αFLAG antibody. This time, there was significant expression from the pcDNA3-derived constructs (Figure 34B), but again, very little detectable expression from the pCMV-Tag1-derived constructs (Figure 34A).

As both sets of constructs expressed in vitro, suggesting that the constructs were cloned correctly and that the phage promoter (either T3 or T7) was functional, and both
were essentially similar vectors with the same promoter driving eukaryotic expression, one possibility was that the CMV promoter of the pCMV-Tag1 constructs was defective.

Figure 34 – Western analysis of U2OS cells transfected with FLAG-25/MCM6 Expression Constructs. U2OS cells were Effectene-transfected with 1μg of either: (A)- pCMVTag1-25 or pCMVTag1-MCM6; or (B)- pcDNA3-25 or pcDNA3-MCM6. Filters were probed with αFLAG and detected with AP-linked anti-mouse IgG. (10% Acrylamide separating gel, Marker sizes shown in KDa).
3.10.5 Co-immunoprecipitation Experiments

A series of co-immunoprecipitation experiments were conducted to test whether BMI-1 co-precipitated with either 25 or MCM6. As a positive control, a myc-tagged RING1 construct (a known BMI-1 interactor (Satijn et al., 1997)) was used in parallel experiments.

U2OS cells were co-transfected with the HA-BMI-1 expression vector and either the FLAG-25/MCM6 or myc-RING1 expression vector. Lysates were prepared in ELB and rat anti-HA antibody was used as the primary antibody to immunoprecipitate HA-BMI-1 (Materials and Methods Section 2.3.7.4). As a control for the HA-BMI-1-myc-RING1 immunoprecipitation, HA-specific peptide was used to block the interaction between the anti-HA antibody and HA-BMI-1 to indicate the specific nature of the immunoprecipitation. Following immunoprecipitation, the samples were western blotted and were probed with either anti-FLAG or anti-myc antibody, depending on which construct was used. As a positive control for construct expression, an aliquot of each transfected cell lysate was included in the western analysis.

In each series of co-transfections, neither FLAG-25 nor FLAG-MCM6 could be seen to co-immunoprecipitate with HA-tagged BMI-1 using the anti-HA antibody (Figure 36). In contrast, myc-RING1 could be successfully co-immunoprecipitated with HA-BMI-1 using the anti-HA antibody in a reaction which could be blocked by the addition of the HA-peptide (Figure 35) indicating that the technique was working appropriately.

Two possibilities could be that, either the proteins of interest have been dissociated from a complex or that somehow, the proteins have not been completely extracted from the cell lysate. For example, in the case of the MCM6 protein, it may be that with such a high number of molecules present within the cell (~10^5-10^6), extraction may have released the majority of the MCM6, accounting for its detection in the cell lysate, but a significant, and possibly functionally relevant, portion may still be associated with DNA or other proteins.

To test this hypothesis, one method is to alter the stringency of protein extraction – the most obvious way through changing the NaCl concentration. Similar co-transfections were again set up as previously, this time harvesting the cells in ELB with differing NaCl concentrations (100mM and 400mM as opposed to the normal 250mM) and
immunoprecipitating as previously. Again, no co-immunoprecipitation was seen with HA-BMI-1.

One other possibility is that the presence of the epitope prevents incorporation of either of the proteins into the correct complex. As the HA-tagged BMI-1 had been successfully used for co-immunoprecipitation experiments previously, it was thought that the HA-epitope was not interfering in its function. Thus it was decided to investigate the effect of the FLAG epitope in relation to the MCM6 protein. A simple transfection only with the HA-BMI-1 construct was set up, and lysates collected. The 83244 anti-MCM6 antibody was then used to detect MCM6 instead of the FLAG-tagged version used previously. As seen previously, while the endogenous MCM6 was detected in the lysate, no interaction was detected between the HA-tagged BMI-1 and MCM6.
Figure 35 – Western blot of positive control co-immunoprecipitation analysis of U2OS cells transfected with HA-BMI-1 and myc-RING1. Rat α-HA was used as the primary antibody: (A) – probed with mouse α-HA monoclonal antibody; (B) – probed with mouse α-myc monoclonal antibody. Both blots were detected with AP-linked anti-mouse IgG. One set of co-immunoprecipitations was blocked with the HA peptide to demonstrate specificity. (12% Acrylamide separating gel, Marker sizes in KDa).
Figure 36 – *Western blot of test co-immunoprecipitation analysis of U2OS cells transfected with HA-BMI-1 and either FLAG-25 or FLAG-MCM6*. Rat α-HA was used as the primary antibody: (A) – probed with mouse α-HA monoclonal antibody; (B) – probed with mouse α-FLAG monoclonal antibody. Both blots were detected with AP-linked anti-mouse IgG. (12% Acrylamide separating gel, Marker sizes in KDa).
3.11 Analysis of Potential BMI-1 Interactor Clone 25

3.11.1 DNA Sequence Analysis

The partial sequence of clone 25 from the region immediately adjacent to the two-hybrid vector (pACT or pVP16) cloning site failed to give any clue as to the identity or possible function of the protein encoded by this cDNA, and so further characterisation of the clone was carried out. Initially, the entire clone 25 insert was sequenced (either from the original recombinant itself, or from a derivative of the clone in the vector pCRScript) and used to assemble a composite sequence. A program called NIX (Nucleotide Identify X) was used to analyse this sequence. NIX is a WWW tool used to view the results of running several DNA analysis programs on a DNA sequence of interest including: GRAIL, Fex, Hexon, MZEF, Genemark, Genefinder, FGene, BLAST (against many databases, including dbEST, GENBANK, EMBL, SWISSPROT, and NBRF) Polyah, RepeatMasker, and tRNAscan. It should be noted that not all of the programs ran were actually relevant to this sequence, but were performed as an integral part of the programme (for example Genefinder which searches for potential genes within genomic DNA). The default settings for the programmes are too numerous to mention here, but information can be found on the NIX Introduction WWW page –

http://menu.hgmp.mrc.ac.uk/Nix/Help/Mainhelp.html

The Results of the NIX search with BMI+25 are shown in Figure 37.

BLAST searching against the dbEST database provided several ESTs which were found to contain significant sequence identity to clone 25 (Figure 38). No known genes from any species showed any significant homology to the cDNA sequence. From analysis of these GENBANK ESTs, it was found that several of them appeared to be part of a Tentative Human Consensus Sequence (THC), 175068 (Figure 39).

3.11.2 Definition of a Tentative Human Consensus Sequence (From the TIGR Database)


THCs are created by assembling ESTs into virtual transcripts. In some cases, THCs contain full or partial cDNA sequences obtained by classical methods. THCs contain information on the source library and abundance of ESTs, and in many cases represent full-length transcripts. Alternative splice forms are built into separate THCs.
Sequence data contained within THC 175068 was obtained from the TIGR database. When compared against 25, the THC175068 represents the antisense strand.

No opposite end information could be found regarding any of the ESTs contained within the THC (with the exception of ESTs yc20a06 and HSBA7C08 – both highlighted in Figure 39), and so the THC could not be extended, either 5-prime or 3-prime of the known sequence by database searching. Similarly, information regarding the size of ESTs (when available) failed to extend the THC.

When the relative lengths of the THC and 25 were compared, it was evident that the THC was entirely contained within clone 25, and so ultimately the THC proved to be of little use apart from serving as a useful sequence comparison against the clone 25. When the two sequences were compared, they were found to be identical with the exception of a 1 base frame shift. In the clone 25 DNA sequence there is an additional A residue that would be located between bases 689 and 690 of the THC. As this would naturally be expected to have a fundamental effect on any Open Reading Frames (ORF) present, this region of 25 was repeatedly sequenced, both from the pACT-25 yeast two-hybrid vector itself and pCRscript containing clone 25 as mentioned in Section). In each case the A residue was present. In addition, out of the two ESTs which provide THC sequence data for the region of interest (11-EST186923 and 13-EST114163), EST186923 also contains the A residue. The sequence quality of the other EST is lower, leaving the possibility that the omission of the A could be as a result of an error in the sequencing of EST114163.
Figure 37 – **NIX output for BMI-1 yeast two-hybrid positive clone 25.** The colours do not have any specific meaning except that programs with similar purposes have generally been grouped together and given the same colour. The colours each have three shades to indicate the quality or confidence of the prediction. The stronger (more intense) the shade the better the prediction. The allocation of the quality of the results is made to one of three strengths: 'excellent', 'good' or 'marginal' and is based on a subjective interpretation by the program of what the scores of the various programs mean.
Figure 38 – BLAST output for BMI-1 yeast two-hybrid positive clone 25. Clone 25 DNA sequence was screened against the dbEST database. The main parameter of interest, the Expect value cutoff is set to 0.1.
Selected ESTs from THC175068

<table>
<thead>
<tr>
<th>#</th>
<th>EST id</th>
<th>GB#</th>
<th>ATCC#</th>
<th>left</th>
<th>right</th>
<th>library</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fa e48h01.s1</td>
<td>AA598461</td>
<td>1 224</td>
<td></td>
<td></td>
<td>prostate NCI_CGAP_Pr6, DKRizman</td>
</tr>
<tr>
<td>2</td>
<td>D ng78f09.s1</td>
<td>AA492214</td>
<td>1 378</td>
<td>prostate NCI_CGAP_Pr6, DKRizman</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>F yc20a06.s1</td>
<td>T57131</td>
<td>321035</td>
<td>1 166</td>
<td>Lung, Stratagene (#937210)</td>
<td></td>
</tr>
<tr>
<td>3b</td>
<td>F yc20a06.s1</td>
<td>T57062</td>
<td>321035</td>
<td>38 147</td>
<td>Lung, Stratagene (#937210)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>F yz35g07.s1</td>
<td>N67557</td>
<td>542197</td>
<td>1 381</td>
<td>Fetal cochlea Morton</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>F zk77b01.r1</td>
<td>AA046335</td>
<td>1 373</td>
<td>pregnant uterus NbHPU, Soares</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>F zp57c10.s1</td>
<td>AA182774</td>
<td>782714</td>
<td>1 451</td>
<td>Endothelial cell, Stratagene (#937223)</td>
<td></td>
</tr>
<tr>
<td>7a</td>
<td>L HSBA7C082</td>
<td>Z25235</td>
<td>2 325</td>
<td>skeletal muscle, Stratagene (#936215)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7b</td>
<td>L HSBA7C081</td>
<td>Z28858</td>
<td>378 662</td>
<td>skeletal muscle, Stratagene (#936215)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>A EST186923</td>
<td>AA314882</td>
<td>113251</td>
<td>245 716</td>
<td>HCC cell line (matastasis to liver in mouse) II</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>D PMY2187</td>
<td>AA471214</td>
<td>250 572</td>
<td>promyeloblast KG1a cell line</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>A EST114163</td>
<td>AA298547</td>
<td>177619</td>
<td>413 735</td>
<td>HSC172 cells I</td>
<td></td>
</tr>
</tbody>
</table>

Figure 39 – THC175068 (from the TIGR database). The entries highlighted in bold indicate ESTs with information relating to both 3’ and 5’ ends.
3.11.3 Northern Analysis

The XhoI fragment (excised from the two-hybrid vector, pACT, containing clone 25) was used as a probe on a Clontech Multiple Tissue Northern (MTN) to estimate transcript size and tissue distribution. Using the hybridisation protocol as per manufacturer’s recommendations (Methods Section 2.3.2.11) several transcripts of different sizes were detected upon autoradiography. The high molecular weight transcript ([Figure 40 – transcript 1]) could possibly be carry over of ribosomal RNA since it is migrating at approximately the same position as 28S rRNA would be expected to. However, the fact that the other ribosomal RNA band (approximately 1.8Kb) is not detectable and that the RNAs used for the blot are predominantly polyA RNA makes it unlikely that this band is ribosomal in origin. Alternatively, it might represent an alternative transcript of clone 25. The low molecular weight transcript ([Figure 40 – transcript 3]) could represent a degradation product of mRNA since the band is more of a smear than either of the other two bands seen. However, the electrophoretic conditions used for the MTN Blots have been optimised to give the best resolution of transcripts in the range of 1.0-4.0 Kb, and so the smaller band could represent a real spliced transcript of clone 25 that was not resolved sufficiently by electrophoresis.

The third band ([Figure 40 – transcript 2]) was approximately 1Kb in size, and does not appear to be a degradation product, and at the same time does not migrate with any of the ribosomal RNAs. Thus, this transcript might also represent a transcript encoding clone 25.

The second point to note is the wide variety of tissues in which the transcript is expressed. All three different sized transcripts are seen in a range of tissues indicating a fairly ubiquitous expression pattern. This is also reflected in the wide range of tissues from which ESTs used to derive the THC were obtained. There are apparent differences in the level of expression, as judged by comparison with the β-actin probe, with expression higher in heart and skeletal muscle than in, for example, lung tissue.
Figure 40 – Northern blot analysis of clone 25. Upper panel - Northern blot (Clontech MTN) probed with 25 (Xhol fragment) according to manufacturer’s instructions (Methods Section 2.3.2.11). Transcripts 1, 2 and 3 are indicated. Lower panel – Control hybridisation using β-actin. (Marker sizes in Kb)

As the northern blot suggests the possibility of a ~4Kb transcript containing clone 25 sequence, strategies were employed in an attempt to extend clone 25 (which is only approximately 800bp). The first strategy was a straightforward library screen (Materials and Methods Sections 2.3.1.4, 2.3.2.6, and 2.3.2.7). The Xhol fragment from the yeast two-hybrid vector, pACT, containing clone 25, was used to screen approximately $5 \times 10^5$ clones from a Clontech human fetal brain 5' STRETCH PLUS cDNA library.
This library was chosen due to the range in size of inserts (from 0.6 to 4.5Kb). In addition, this library had been used in other studies to isolate a 10Kb brain-specific transcript encoding Doublecortin, a gene mutated in human X-linked lissencephaly (Gleeson et al., 1998), and thus was thought a suitable library from which to obtain long transcripts. Several of the ESTs identified from BLAST searches with the DNA sequence of clone 25 were derived from human fetal brain or total human fetus indicating the source of the library was also appropriate.

From the library screen, 7 primary positives were identified, of which 4 remained following secondary and tertiary screens. Following rescue from phage, the inserts were subcloned into pBluescript and sequenced (the positives are referred to as HFB (Human Fetal Brain) 1, HFB2, HFB5 and HFB6).

The four clones rescued ranged in size from 0.7-1.0 Kb in size, and therefore do not correspond to the potential 4 Kb transcript. Sequence analysis revealed that the clones all had similar 3-prime ends which matched the 3-prime end of clone 25, but each clone differed at the 5-prime end. The 3' of HFB1 is similar to the 3' of clone 25, however, the 5' region bears no similarity at all. Database screening with this 5' region revealed no similarity to any known genes or ESTs from any species. Whether it represents a splice variant of clone 25, an unspliced human cDNA or a coligation product remains to be determined.

The remaining clones all extended upstream of the 5-prime end of clone 25, though in the case of HFB2 and HFB5, only by 90 and 60 bp respectively. The remaining cDNA, HFB6, extended approximately 240bp upstream of the 5-prime end of 25, taking the total transcript length to about 1Kb (Figure 41), which correlates with the transcript (2) seen on the northern blot (Figure 40). Of note from sequencing HFB2, -5, and -6 was that the extra A residue, lacking from the THC175068 but present in 25, is also present in each of these cDNAs.
Figure 41 - Alignment of 5-prime ends of Human Fetal Brain library clones (HFB2, -5, and -6), THC175068 and clone 25. The potential initiation methionine and in-frame upstream STOP codons are in red. The extra A residue is in green.
3.11.4 Other Strategies to Obtain cDNAs Corresponding to the Long Transcript Detected by Clone 25

With none of the cDNAs identified from the library screening extending clone 25 significantly, other means were employed in an attempt to find a cDNA corresponding to the 4Kb transcript predicted by northern analysis. Primary amongst these were 3-prime and 5-prime RACE (Rapid Amplification of cDNA Ends). 3-prime RACE was performed because it was thought that clone 25 might extend further 3' than the known sequence. Within the DNA sequence for clone 25, there is no classic polyadenylation signal which is characteristic of most mRNAs, which could suggest that the true 3' end is not that of clone 25 and is in fact further downstream. 5-prime RACE was performed to try and obtain the upstream sequence which is normally underrepresented by normal methods of cDNA library synthesis. cDNA was prepared from human placental mRNA and adaptors were ligated to each end as per manufacturer's instructions (Methods Section 2.3.5.2). 3' and 5' RACE PCR was performed using nested gene-specific primers and primers designed against the adaptors.

3-Prime RACE

3-prime RACE was performed using nested internal 25-specific primers (3' RACE1 and 3' RACE2) and the primers AP1 and AP2 (see Methods Section 2.2.9), with the human placental cDNA as template. A product was obtained for first round (using 3' RACE1 and AP1) PCR which, when used as template for a second round of nested-PCR (using 3' RACE2 and AP2), resulted in a slightly smaller product being obtained. Both of these fragments were cloned and sequenced. It was found that the fragments corresponded to sequences between the relevant primer (3' RACE1 or-2) and the 3-prime end of 25. Thus it was apparent that 3' RACE had not extended downstream of the known clone 25 sequence.

5-Prime RACE

5-prime RACE was performed using nested internal 25-specific primers (5' RACE1 and 5'RACE2) and the primers AP1 and AP2 (see Methods Section 2.2.9), again with the human placental cDNA as template. First round PCR (using 5' RACE1 and AP1) resulted in a smeared product. Using this as template, second round PCR (using 5' RACE2 and AP2) was performed using various combinations of primers (e.g. 5'
RACE2 and AP2, 5' RACE1 and AP2, and using 5' RACE2 and AP1). In each case no PCR product was obtained.

At this time, of the three transcripts detected from northern analysis, only one (*transcript 2, Figure 40*) can be adequately accounted for, with the cDNA encoded by HFB6 seeming a suitable candidate for this transcript. Both the low Molecular Weight transcript (*transcript 3, Figure 40*), and the ~4Kb transcript (*transcript 1, Figure 40*) remain unaccounted for. Whether they actually represent alternative splice variants of clone 25 or are related sequences remains to be determined.

### 3.11.5 Chromosomal Location

One important piece of information which can be used to determine whether a gene represents a candidate for a particular disease is its genetic location. Within the human genome, many markers exist which have been physically mapped to a distinct region of a chromosome, so-called Sequence Tag Sites or STSs. These STSs are used as the basis for electronic PCR. The virtual PCR products generated thus are screened against DNA databases resulting in ESTs or genes containing this PCR sequence being identified. These 'hits' are then further screened against DNA databases, ultimately with a cluster of transcripts that maps to the STS in question being assembled.


Searching of the UniGene database with ESTs contained within THC175068 identified one such STS. This marker, stSG2963, has been mapped to the long arm of chromosome X between the genetic markers DXS1059 and DXS1047. Originally, using electronic PCR, this STS would have been screened against DNA databases, revealing two ESTs (*AA314882* and *Z28858*). These ESTs would then have been linked to other overlapping ESTs and the resultant cluster generated.

Unfortunately, in this case, knowing the genetic locus is not particularly informative. Only a few disease loci have been characterised within this region, with the majority of those diseases that have been fine-mapped to this regions already having candidate genes assigned, e.g. the Doublecortin gene (*Gleeson et al., 1998*) mutated in human X-linked lissencephaly has been mapped to Xq22.3-q23.
Figure 42 - Summary of DNA sequences relating to BMI-1 yeast two-hybrid clone 25. (ESTs from THC175068, Other ESTs, 3’RACE products, HFB cDNA clones)
3.11.6 Protein Analysis of Clone 25

3.11.6.1 Open Reading Frames

As nucleotide similarity searches found no significant matches with clone 25, the predicted ORF was used to screen protein sequence databases instead. The predicted ORF was derived from the junction sequence from the two-hybrid vector, pACT, for which the ORF of the $GAL4$ Activation Domain is known. The predicted protein sequence, shown in **Figure 43**, encodes an ORF that extends for approximately 210 amino acids. Using an ORF prediction program, this reading frame was shown to be the only significant reading frame present within the DNA sequence (**Figure 44**).

```
1  RGHEGRGTM A AAADERSPED GEDEEEEQL VLVELSGIID SDFLSKCENK
51  CKVLGIDTER PILQVDSCVF AGYEYDTLTG CVIFEENVEH ADTEGNKTV
101  LKYKCHTMKK LSMTRTLLTE KKEGEENIGG VEWLQIKDND FSYPNMCIN
151  FLHENEDEEV VASAPDKSLE LEEEIQMNDD SSNLSCEQEK PMHLEIEdSG
201  PLIDIPSETE GSVFMETQML P*KSLLDEMF LIITCQELFR VVT*K*LLCK
251  KKKAFVAS
```

**Figure 43** - Predicted protein sequence of BMI-1 yeast two-hybrid positive 25. The pACT vector linker sequence is in red. The potential initiation methionine and in-frame STOP codon are underlined.

---

```
Frame from to Length
+3  9..650  642
-1  2..367  366
+2  296..400  105
+2  2..103  102
```

**Figure 44** - Predicted potential Open Reading Frames encoded by BMI-1 yeast two-hybrid positive 25. The Open Reading Frame derived from the yeast two-hybrid vector, pACT, is indicated →. The Program used is at WWW Reference: http://www.ncbi.nlm.nih.gov/gorf/gorf.html

143
Within the predicted ORF lies a Methionine in a favourable context for initiation as defined by the Kozak consensus sequence. For efficient translation in vertebrates, both the ATG initiation codon and the sequences flanking the initiation codon have been shown to be required to direct the position of initiation. Using a series of preproinsulin mutant mRNAs, Kozak derived a consensus sequence of GCC(A or G)CCATGG to signal the initiation site (Kozak, 1987, Kozak, 1996, and Iida & Masuda, 1996).

When the consensus is compared to the potential site in the unknown cDNA, there are only two minor differences with two 5’Cs being replaced by Gs.

```
Kozak Consensus       GCC (A/G) CCATGG
clone25 sequence      GGG--A--CCATGG
```

One other piece of supporting evidence is that upon examination of the protein coding sequence of HFB6, assuming the frame dictated by clone 25 and pACT, two upstream in-frame STOP codons are located approximately 30bp and 75bp from the potential initiation methionine, as well as STOP codons existing in both of the other translational frames.

To test the functionality of the Kozak consensus an *in vitro* translation experiment was performed. The *XhoI* fragment from the pACT-25 two-hybrid vector was cloned into pBluescript and used as a template for *in vitro* translation using $^{35}$S-methionine and a T7 coupled transcription-translation kit (Methods Section 2.3.8.4). Two proteins were visible upon autoradiography, one of approximately 21-22 KDa and the other of approximately 37-38 KDa (Figure 46A). These two proteins were also present following translation of the 25+ clones obtained from the human fetal brain cDNA library. To determine which protein was the expected protein, the FLAG-25 construct in pcDNA3 (from Results Section 3.10.4.3) was *in vitro* translated. Two proteins were again produced, with the lower band of the same mass as the lower band of 25 in pBluescript. The upper band was of a slightly increased mass which would correlate with the addition of the FLAG epitope tag. In addition, an immunoprecipitation experiment was performed using the anti-FLAG antibody, and this resulted in only the larger protein being captured (Figure 46B).

This series of *in vitro* translation experiments appear to conflict with the protein sequence data which predicts a protein of approximately 25KDa. This mass discrepancy is also apparent when comparing the GST-25 fusion protein with GST alone (from
Results Section 3.9.3). The difference in mass between GST alone and GST-25 would also suggest a protein of greater than 25KDa (Figure 45). The occurrence of this mass discrepancy in both a eukaryotic-based system (reticulocyte lysate) and a prokaryotic-based system would seem to argue against it being due to a post-translational modification. Whether the smaller protein represents a degradation product of the protein encoded by 25, or is due to initiation at an alternative methionine remains unknown. The fact that the size of the smaller protein remains unaffected by the addition of the FLAG epitope might argue for an internal initiation methionine being present. Other ATG codons are present in the clone 25 DNA sequence, though none are in as favourable a context as that described previously.

Other examples of proteins which do not migrate as expected are provided by the Polycomb-group proteins themselves. For example, BMI-1 protein, migrates with a mobility of ~45KDa as judged by in vitro translation, somewhat larger than the 37KDa predicted from the 326 amino acid ORF (Alkema et al., 1993).

Figure 45 – Coomassie stained SDS-PAGE gel of GST and GST-25. The gel illustrates the size difference between GST alone and the fusion protein GST-25. (12% Acrylamide separating gel, Marker sizes in KDa).
Figure 46 – SDS-PAGE analysis of in vitro translation and immunoprecipitation of clone 25 constructs. (A) – translation of 35S-methionine 25 from pCRScript. (B) – translation of 35S-methionine FLAG-tagged 25 and immunoprecipitation of FLAG-25 using anti-FLAG monoclonal antibody. (12% Acrylamide separating gel, Marker sizes in KDa).

3.11.6.2 Motif and Pattern Analysis

WWW protein motif databases searched using the predicted protein sequence included: Pfam (A collection of protein families and domains); ProDom (A collection of homologous domains); Motifs (a collection of protein motifs, including many of those in ProDom and Pfam); SAPS (Statistical Analysis of Protein Sequences to evaluate a range of protein sequence properties); and PEST (Used to identify proteins with short half-lives).
Physical analysis of the protein sequence revealed that the protein is slightly unusual in its composition, having a relatively high proportion of negatively charged amino acids (aspartic acid and glutamic acid) (Figure 47), helping to give the protein an overall negative charge at pH7. This is reflected in the predicted Isoelectric point (pI) of 4.1. This might be one explanation to account for the aberrant mobility seen with SDS-PAGE analysis.

A: 8 (3.8%); C: 7 (3.3%); D+: 16 (7.5%); E++: 38 (17.8%); F: 6 (2.8%
G: 11 (5.2%); H: 4 (1.9%); I: 12 (5.6%); K: 13 (6.1%); L: 19 (8.9%)
M: 8 (3.8%); N: 11 (5.2%); P: 8 (3.8%); Q: 6 (2.8%); R: 4 (1.9%)
S: 15 (7.0%); T: 11 (5.2%); V: 12 (5.6%); W: 1 (0.5%); Y: 3 (1.4%)

**Figure 47 – Amino acid composition for the predicted protein encoded by clone 25.**

In terms of functional analysis, very little in the way of identifiable motifs was revealed with the exception of several potential phosphorylation and glycosylation sites for example, Protein Kinase C phosphorylation sites and Casein kinase II phosphorylation sites (as shown in Appendix 5B). PEST (Rodgers et al., 1986) analysis to determine proteins with short intracellular half-lives revealed two potential PEST sequences within the protein sequence (see Appendix 5A).

### 3.11.6.3 Nuclear or Cytoplasmic Localisation of Clone 25

The function of BMI-1 requires it to be located in the nucleus, and so it seems reasonable to suggest that any potential interactors would also be located in the nucleus. Upon sequence examination of the protein sequence encoded by 25, one notable feature was the absence of a Nuclear Localisation Signal. To determine whether 25 could be targeted to the nucleus, a crude method was employed to determine its intracellular distribution. U2OS cells were transiently transfected using the pcDNA3-FLAG-25 (from Results Section 3.10.4.3), along with a known nuclear protein to act as a positive control for nuclear targeting (the transcriptional regulator Pax3, containing a N-ter FLAG epitope in pcDNA3). Crude nuclear and cytoplasmic preparations were made as in Methods Section 2.3.8.6 and western blots of these extracts were prepared. The blots
were then probed with the anti-FLAG monoclonal antibody. The results suggested that the protein encoded by 25 is a predominantly cytoplasmic protein with little, if any, detectable expression in the nucleus (Figure 48). The FLAG-Pax3 protein was detected in the nuclear extract as expected. One should, however, be cautious in this interpretation and assume that overexpression of the proteins in question has not altered their subcellular distribution, for example by overwhelming a post-translational pathway necessary for sub-cellular localisation, resulting in some protein not being modified.

![Figure 48](image-url)  

**Figure 48** – *Crude intracellular localisation of FLAG-25 and FLAG-Pax3 in transfected U2OS cells.* The blots were probed with anti-FLAG monoclonal antibody and detected with AP-linked anti-mouse IgG. (C) – Cytoplasmic extract; (N) – Nuclear extract. (12% Acrylamide separating gel, Marker sizes in KDa).
Chapter 4
Discussion

The discussion will be split into several sections: 4.1) brief summary of results; 4.2) the validity of the positives as real interactors of BMI-1; 4.3) the role of MCM proteins in DNA replication; 4.4) the methods used to investigate protein-protein interactions; 4.5) future work, and 4.6) conclusions and hindsight.

4.1 Brief Summary of Results

For this study, a truncated BMI-1 fragment was chosen as bait for use in a GAL4-based yeast two-hybrid system. Following testing for the potential to autoactivate the HIS3 reporter, three library screens were performed. The initial lack of interaction detection in the first library screen was thought to be due to a low transformation efficiency. This was subsequently overcome by using a modified transformation protocol, and the two ensuing library screens (against B-cell and mouse embryo libraries) demonstrated significantly improved transformation efficiencies. From the second screen, several positives were identified and subsequent use of cycloheximide selection and retransformation procedures allowed some of these positives to be eliminated from further analysis. From the mouse embryo library screen many potential positives were identified – too many to permit easy characterisation, and it was thus decided to screen a Southern blot of PCR-amplified products of all of the mouse embryo-derived positives with a pool of radiolabelled probes identified from the second screen. This way, it was hoped that any positives conserved between species could be easily identified for subsequent characterisation. Using this approach, two positives (clones 25 and 54) were identified and subsequently sequenced, clone 25 (EST A7CO8) and clone 54 (C-ter fragment of a member of the minichromosome maintenance family of proteins, MCM6). These two positives, along with a clone (16) encoding the splicing factor SRp20 (chosen for its multiple occurrence in the second library screen) were then further characterised. Retransformation experiments in different yeast strains, Y190, Y166 and PJ69-4A confirmed the initial yeast two-hybrid results and the inclusion of ‘empty’ vectors as
appropriate controls served to show the specificity of the interactions. The bait (tBMI-1) was also tested against Mph1, and tBmi-1 was tested against the three positives in a series of yeast two-hybrid experiments. The interaction observed between the positives and tBMI-1 was not observed when tBmi-1 was used as bait. In addition, the magnitude of the β-Gal reporter activation between tBmi-1 and Mph1 was significantly larger than that observed between tBMI-1 and the three positives tested.

To provide further corroborative evidence for the interactions, GST-fusion constructs were generated for each of the three clones (16, 25 and 54) and were expressed in bacteria. Only two of the fusion proteins (GST-25 and GST-54) were successfully purified. These two proteins were then used as bait to successfully capture in vitro translated 35S-methionine-labelled full-length BMI-1 and a truncated BMI-1 fragment corresponding to that used in the initial two-hybrid screens. GST protein alone failed to significantly pull down BMI-1.

In an attempt to corroborate the interaction in mammalian cells, a series of co-immunoprecipitation experiments were performed. Initially, anti-BMI-1 and anti-MCM6 polyclonal antibodies were raised and characterised. While the anti-BMI-1 antibodies failed to detect a band corresponding to BMI-1 on a western blot, the anti-MCM6 antibody detected a specific band corresponding to MCM6. However, further investigation of this antibody suggested that while the antibody was suited for western analysis it could not be used for immunoprecipitation of MCM6. A different approach was attempted with FLAG epitope-tagged constructs being generated for clone 25 and the full-length MCM6. Following optimisation of transfection conditions in U2OS cells, a series of co-immunoprecipitation experiments were performed with the FLAG-tagged constructs cotransfected into U2OS cells along with HA-tagged BMI-1. Thus far, no interactions have been detected in the co-immunoprecipitation studies between BMI-1 and the FLAG-tagged proteins encoded by clone 25 and MCM6. In contrast, an interaction was observed between BMI-1 and a known interactor of BMI-1, RING1.

Further examination of clone 25 (EST A7CO8) revealed little homology to known genes at the DNA level. Using clone 25 as a probe for northern analysis, three transcripts were identified. However, cDNA library screening and RACE analysis identified just one of the transcripts. The identity of the other two transcripts remains unknown. Within the DNA sequence there is a favourable Kozak translation initiation signal and the predicted
protein sequence from this initiating Methionine reveals a 213 amino acid protein with no homology to known sequences and little in the way of identifiable motifs. The apparent mass of 37-38 KDa, as suggested from in vitro translation analysis, is somewhat larger than the predicted mass of ~25 KDa. Nuclear vs. cytoplasmic distribution analysis using a FLAG-tagged 25 construct suggests that the protein is predominantly cytoplasmic in distribution.

4.2 Validity of the Positives as Real Interactors of BMI-1

This section will focus on known interactors of BMI-1, the interactors that have been identified in this study and those interactors that may have been missed.

4.2.1 Known Bmi-1 Interacting Proteins

Following the initial identification of Bmi-1, more recent work has focussed on identifying other components of Polycomb-group complexes that interact with Bmi-1. Using Bmi-1 as a bait, yeast two hybrid analysis identified Mph1, (or rae28 as it is also known) as an interacting protein (Alkema et al., 1997a), and coimmunoprecipitation experiments suggest that Bmi-1 and Mph1 constitute a large multimeric protein complex together with M33 and MPc2 (Alkema et al., 1997b) and other proteins. The Mph1 gene encodes a protein sharing several characteristic motifs and highly homologous regions with the Drosophila gene polyhomeotic (ph); a single zinc finger, a glutamine-rich region and two highly homologous regions, one consisting of 28 amino acids (Homology Domain I) and the other, a 66 amino acid SEP domain (yeast sterility-, Ets related-, Pc-G proteins) (Homology Domain II). Similarly, using a human BMI-1 as bait against a human library, two proteins, HPH1 and HPH2 were isolated (Gunster et al., 1997). These two proteins have little sequence homology with each other, except in two highly conserved domains which they also share with Drosophila Ph (Homology Dominains I and II). The presence of two proteins with homology to Ph again points to the possibility of there being different mammalian multimeric Pc-G complexes.

Interestingly, Mph1−/− (Takahara et al., 1997) mice show a phenocopy which, although is not identical, is reminiscent of the phenotypes seen in the humans with congenital disorders derived from neural crest defects such as the CATCH-22 (Cardiac
abnormalities, Abnormal facies, Thymic hypoplasia, Cleft lip, Hypocalcemia associated with hemizygosity for chromosome 22q11) group of syndromes which encompasses DiGeorge syndrome and velo-cardio-facial syndrome (reviewed in Scambler, 1993a, 1993b). The homozygous mice show skeletal posterior transformations and multiple developmental defects in the eyes, hard palate, parathyroid glands, thymus, heart and spleen. In humans, the CATCH-22 syndrome is autosomal dominant and is frequently associated with microdeletions within chromosome 22q11. However, a second DiGeorge syndrome locus was identified which mapped to 10p13, though has since been shown to exclude the BMI-1 gene (Daw et al., 1996) by deletion mapping. Nevertheless, the remarkable similarity of the phenotypes between the rae28-/- homozygotes and the CATCH-22 syndrome may indicate the presence of a defect in a common pathway of neural crest cells.

Extensive characterisation of interactions between BMI-1 and other Polycomb-Group proteins has recently been performed by Satijn et al. (1999), in particular in relation to interactions between BMI-1 and RING1A and RING1B (also known as dinG). Comparing RING1A and dinG there are three regions that demonstrate extensive homology, a 150 amino acid N-ter region (almost 100% identical) and two C-ter regions (~70% identical), and interaction analysis suggests that these domains are capable of mediating similar interactions. There is, however, a central region which exhibits very little homology between the two proteins, and this region is thought to mediate differential interactions, suggesting that RING1A and dinG are able to interact with different proteins, which could provide specificity for the formation of different Pc-G protein complexes (Figure 49).
**Figure 49 – Model of human Pc-G multimeric protein complexes.** RING1A and RING1B possess different protein binding properties interactions that permit the formation of different Pc-G complexes, with different target specificities.

### 4.2.2 fBmi-1 and Mphl Interactions

One noticeable feature of the three library screens carried out in this project was the lack of identification of any Pc-G proteins as candidate interactors. The N-ter RING finger has been shown to be necessary, though in some cases not sufficient, for mediating protein-protein interactions, for example, heterodimerisation of Bmi-1 to dinG/RING1B requires intact RING finger structures on both proteins. The RING finger of Bmi-1 is also essential for its oncogenic potential in Eμ-myc transgenic mice, though the C-terminal Pro-Ser region is not required (Hemenway et al., 1998). Thus, using the tBMI-1 clone lacking the RING finger, any interactions that might occur as a result of the RING finger would be unlikely to be detected. However, Alkema et al. (1997a) domain-mapped the region of murine Bmi-1 involved in Mphl binding and found it excluded the RING finger.

As can be seen from **Figure 50**, the fragment of BMI-1 used as bait in the two-hybrid screens overlaps substantially with the Mphl binding domain of Bmi-1. Alkema et al. also showed, using two-hybrid analysis, that a region within the Mphl binding fragment, but not contained within the two-hybrid fragment (Region A indicated), is,
along with the rest of the Mphl binding fragment, necessary to mediate the interaction. Thus from their results it would appear that the tBMI-1 fragment should not bind Mphl.

The interaction between BMI-1 and Mphl may have remained undetected even if a larger bait fragment encompassing the whole Mphl interacting region had been used in the library screens, because of the average length of cDNAs in the Activation Domain vector libraries. As both libraries were oligo-dT primed, there is an inherent bias towards the 3’prime end of mRNAs. This, coupled with the fact that the libraries were size selected resulted, in the case of the mouse embryo library, in relatively small inserts being subcloned. For example, as demonstrated by Alkema et al. (1997a), the region of Mphl responsible for Bmi-1 binding extends from amino acids 720-1012, almost 300 amino acids, corresponding to ~900bp of DNA sequence. As can be seen from Figure 13A (Results Section 3.3.4), the average size of inserts in the mouse embryo library in pVP16 is ~500bp, much smaller than 900bp required, and so it is unlikely that any Mphl clone present in the library would contain a sufficient Bmi-1
binding domain, with the consequence that such interactions could have remained undetected.

Even though the fragment used should not bind *Mphl*, as Figure 20 (Results Section 3.7) shows, there is some *LacZ* reporter activity between the pAS2-1-tBMI-1 and *Mphl*. Although this level is low alongside the full-length *Bmi-1* in pPC97 and *Mphl* combination, it does appear significant when compared with the pAS2-1 and *Mphl* combination. Whether this difference between pPC97-t*Bmi-1* and pAS2-1-tBMI-1 with respect to *Mphl* can be attributed to differences in the relative promoter strengths of the GAL4 BD vectors and stability of the plasmids, or to the abolition of the *Mphl* binding site has not been determined. It seems unlikely that the difference can be attributed to the use of clones originating from different species due to the high level of conservation between murine *Bmi-1* and human BMI-1. If, however, the interaction between pAS2-1-tBMI-1 and *Mphl* is not significant, and the *LacZ* result is as a result of some background binding activity, the comparable strengths of the interaction between pAS2-1-tBMI-1 and pACT-16, -25, and -54 would seem to suggest that, they too, are as a result of a background binding capability. If this were the case, then perhaps a stronger interaction may be detected using a different fragment of BMI-1.

The other main difference observed between the tBMI-1 and t*Bmi-1* clones was in their relative interactions with pACT-16, pACT25 and pACT-54. In contrast to the scenario with *Mphl*, the full-length *Bmi-1* failed to detect an interaction with either of these three fusion proteins whereas the truncated BMI-1 suggested a slight, though significant, interaction.

One possibility is that by removing the N-ter region of the BMI-1 protein the tBMI-1 fragment is able to adopt a conformation which permits binding of these proteins but which bears no significance to the ‘normal’ conformation adopted by BMI-1 *in vivo*. However, the GST-affinity capture data from Results Section 3.9 would seem to suggest that the removal of the N-Ter region of BMI-1 has no significant effect on the ability to be captured by either GST-25 or -54, and is suggestive of a similar conformation being adopted by both the t*Bmi-1* and tBMI-1 proteins. If this is the case, then an alternative explanation may be that the presence of the GAL4 DNA-Binding Domain causes an alteration of the structure of the full-length *Bmi-1*, through the RING
finger, which somehow prevents it from interacting with either of the two proteins. Using the tBMI-1 fragment, the RING finger-mediated change may not occur, allowing the interactions to be detected. However, the fact that the GAL4BD-tBmi-1 can still interact with Mph1 would imply that the structure of the central Mph1-interacting region is unaltered. As the domain responsible for Mph1 interaction has been only approximately identified, and not fine-mapped, it may be that the area responsible for Mph1 interaction is very small and so less affected by significant global structural alterations.

One further alternative is that a post-translational modification, that is necessary for the interaction to occur between full-length BMI-1 and the C-ter fragment of MCM6 or clone 25, fails to occur in yeast, but does so in rabbit reticulocyte lysates. This modification, e.g. phosphorylation, could alter the RING finger from conformation I (Figure 51B) which prevents MCM6 binding to full-length BMI-1, to conformation II (Figure 51C) which permits MCM6 binding to full-length BMI-1. The absence of the RING finger from the original bait construct could alleviate the necessity for the post-translational modification and allow the interaction to be detected in yeast.
Figure 51 – Different conformations of BMI-1 could effect different protein interactions. (A) In yeast, the truncated BMI-1 is capable of binding to the C-Ter fragment of MCM6. (B) The presence of the RING finger somehow prevents the C-Ter MCM6 fragment binding. (C) In mammalian cells, a post-translational modification could alter the BMI-1 structure to ‘open’ the RING finger domain, allowing the C-Ter fragment of MCM6 to bind.

4.2.3 The Role of Electrostatic Interactions?

One possibility of explaining the interactions could be as a result of the physical properties of the proteins. When comparing the relative pI’s of each of the proteins in question it is interesting to see that while both of the Polycomb proteins, BMI-1 and the C-Ter Mphl are relatively basic (pI’s of 8.90 and 8.91 respectively), the two proteins encoded by clones 25 and 54 are relatively acidic (pI’s of 4.13 and 4.55 respectively). The BMI-1 interactions with the clone 25/54 proteins could thus be mediated as a result of electrostatic attraction between the proteins. The interaction between Bmi-1 and Mphl is seen despite the electrostatic repulsion that would be expected to occur between proteins of similar, basic pI’s. Whether the difference between the pI’s of
BMI-1 and clones 25 and 54 is responsible for mediating interactions is difficult to determine, though it could be postulated as one mechanism by which the detection of false positives could occur.

4.2.4 Domain Swap Experiments

One method of corroborating an interaction is to switch the GAL4 domains so that the bait protein is fused with the activation domain and the library-derived protein is fused to the DNA binding domain. Observed interaction between reciprocal hybrids is a strong indication of true physical association. The domain swap experiment was not carried out in this study due to incompatibilities between the GAL4 DNA binding domain vector (pAS2-1) and the GAL4 activation domain (pACT) multiple cloning sites. Even if it had been performed, a non-detection of an interaction may not necessarily reflect a false-positive. An example of relevance is demonstrated by BMI-1 itself. Satijn & Otte (1999) found that BMI-1 self-associated on the basis of a series of yeast two-hybrid experiments. The full-length BMI-1 fused to the GAL4 DNA binding domain was found to interact with an N-Ter region (amino acids 1-136) fused to the GAL4 activation domain. However, the same N-Ter region was not found to interact with full-length BMI-1 when the GAL4 DNA binding domain and activation domain were swapped (Figure 52), presumably due to the fusion proteins having different properties (e.g. conformation and stability).

![Diagram of GAL4 Domain Swap experiments altering two-hybrid BMI-1 self-association](image)

Figure 52 – GAL4 Domain Swap experiments alter two-hybrid BMI-1 self-association, as determined by yeast two-hybrid experiments. (Taken from Satijn & Otte, 1999.)
4.2.5 Correlation Between Interaction Strength and Biological Significance

One important underlying principle in the interpretation of these results is the assumption that the degree of interaction as determined by two-hybrid approaches reflects the degree of interaction as determined by biochemical techniques. Estojak et al. (1995) determined that while the strength of interaction as determined by the two-hybrid approach in general (i.e. the combination of auxotrophic and colorimetric reporters) correlates with that determined in vitro, there was no single reporter for which the amount of gene expression linearly reflected affinity measured in vitro. Such assumptions of result-interaction correlation should always, however, be treated with a certain degree of caution since many other factors can affect the two-hybrid result, including the expression, stability, nuclear localisation, 3-D structure of a fusion protein, the fact that discrete domains might interact more strongly than the corresponding full-length protein and the fact that the interaction is occurring in a biologically unrelated species (Vidal & Legrain, 1999). In addition, it should also be emphasised that the magnitude of the two-hybrid response does not necessarily correlate with the biological significance of the interaction.

4.2.6 Clone 16 (SRp20) as a Candidate Interactor of BMI-1

It remains difficult to say whether clone 16 represents a real interactor of BMI-1 simply because time did not permit a more detailed analysis. There are several lines of argument both for and against this being a real interaction. One strong argument against this interaction is based around their respective mechanistic details of action. BMI-1’s known function is that of a transcriptional repressor, whilst SRp20 functions as a member of a splicing complex, which is indicative of a role in transcription. At the nuclear level, initial localisation studies revealed that the SR proteins are localised to subnuclear domains called speckles, distinct from sites of gene transcription. Co-localisation studies of SC35 (a member of the SR family) along with BMI-1 have showed that BMI-1 does not associate with these structures (Alkema et al., 1997a). However, refinements of the initial SR antibody studies showed that these speckles may represent storage sites and that their actual distribution is far more wide-spread, generally being associated with sites of active transcription (Neugebauer & Roth, 1997). Whilst BMI-1
localisation has not been carried out to this resolution, it is still remains difficult to rationalise the two proteins apparent contrasting function. In addition, it has been reported that SRp20 has been identified in a number of different yeast two-hybrid screens as a false positive (Karla Neugebauer, personal communication).

Despite this, one line of evidence supporting the argument for the interaction being real concerns the number of positives identified through the library screen. Seven of the fourteen B-cell positives identified from the first successful library screen were SRp20 and while they all code for the full-length protein, they differed in their 5-prime untranslated sequence (which in the yeast two-hybrid system is translated) (Figure 14). No other clones were identified in such number.

Of interest in this regard, is the idea that BMI-1 might interact with a protein SPOP (Maarten van Lohuizen, personal communication), a novel nuclear speckle-type protein, which itself co-localises with a splicing factor, snRNP B/B (Nagai et al. 1997). In addition, analysis with the PcG proteins Eed and Ezh2 revealed that, in vitro at least, these two proteins were capable of binding RNA, and that RNA altered the interaction between these proteins (Denisenko et al., 1998). It is therefore possible, that BMI-1 might have as yet uncharacterised functions in RNA metabolism.

Alternatively, it may be that SRp20 has an as yet undescribed function in the regulation of transcription by Pc-G proteins. While this seems unlikely, it cannot be ruled out as new functions for previously characterised proteins are constantly being reported. One of the genes predisposing to Wilms’ Tumour, WT1, has been suggested to have dual roles. The deduced protein sequence indicates that it may act as a DNA-binding factor, in that it has four zinc finger domains and a proline/glutamine rich transregulatory domain similar to that found in many known transcriptional regulators (Call et al., 1990). In vitro, WT1 has been shown to bind to a specific DNA sequence similar to that of the transcription factor EGR1 (Rauscher et al., 1990). However, the activity of the protein can be modulated by the inclusion or omission of a tripeptide, lysine-threonine-serine (KTS), between zinc fingers 3 and 4. The inclusion of the KTS tripeptide reduces the affinity of WT1 for DNA and results in the colocalisation of the protein, detected using both mono- or polyclonal antibodies, with components of the pre-mRNA splicing machinery (Larsson et al., 1995).

The weight of evidence observed since the initial isolation of SRp20 suggests that a more detailed analysis of this protein and its possible interaction with BMI-1 would be
beneficial, and this clone has been sent to Maarten van Lohuizen for more extensive characterisation.

4.2.7 Clone 25 (ESTA7CO8) as a Candidate Interactor of BMI-1

Evidence suggesting the polypeptide encoded by clone 25 as a real interactor of BMI-1 is contradictory. Evidence for the interaction occurring is essentially two-fold. Primarily, using the tBMI-1 fragment as bait, its detection in the yeast two-hybrid library screens from two different species could be argued as quite strong evidence. Not only is the gene highly conserved in both species, but also its potential binding function is apparently conserved in that it was detected by the same fragment of the bait protein. Such conservation between species would not be expected to occur by chance. Secondly, the interaction was corroborated *in vitro* using a GST-25 fusion construct. Using this construct interactions were detected between GST-25 and both the truncated BMI-1 and, more importantly, full-length BMI-1, neither of which were detected by GST alone, indicating the necessity of the polypeptide encoded by clone 25 for this affinity capture to occur.

Unfortunately, arguments against the interaction are essentially three-fold. Firstly, the interaction has not yet been successfully detected *in vivo* using epitope-tagged proteins and co-immunoprecipitation techniques. This may be as a consequence of several factors, for example, the lysis buffer used may be of the wrong stringency. However, different stringencies, based on different NaCl concentration were tested, each without apparent success. As a positive control, BMI-1 and RING1 were successfully co-immunoprecipitated. The second argument that the interaction was not confirmed in the yeast two-hybrid system against the full-length BMI-1 is also valid since the full-length BMI-1 construct detected *Mph1*, a characterised interactor of BMI-1.

The third, and possibly most convincing argument against the interaction, relates to the cellular distribution of the protein. With the function of BMI-1 requiring it to be nuclear in distribution, in order for there to be a significant interaction between the two proteins one would expect the protein encoded by 25 to be nuclear as well. Therefore it was with some disappointment that upon determination of its cellular localisation, via crude cytoplasmic vs. nuclear preparations, 25 appears to exist in a predominantly cytoplasmic
form (Figure 48). Perhaps, this was not so surprising, since analysis of the peptide sequence fails to identify a classic Nuclear Localisation Sequence (either monopartite or bipartite), as determined by the PSORT WWW server (WWW page - http://psort.nibb.ac.jp:8800/). In addition, cellular distribution programs within the same server predict the protein as being cytoplasmic, although caution should used with this result as the prediction methods used are based on heuristic probabilities rather than strict, defined patterns. This absence of a NLS alone cannot guarantee non-nuclear entry since there are examples of proteins which are nuclear in distribution which lack a NLS. An example is illustrated by the E2F transcription factors, which bind to DP protein family members to form functional E2F heterodimers. While E2F-1 possesses an intrinsic NLS, E2F-4 and -5 are devoid of such a signal. In the case of these latter two proteins, distinct processes govern their nuclear accumulation whereby the nuclear localisation signal is supplied in trans from either a DP heterodimeric partner, or a physically associated pocket protein (Allen et al., 1997). In addition, the subcellular distribution of some proteins without a NLS can be affected by different stages of the cell cycle, e.g. the MCM proteins as mentioned in Discussion Section 4.3.2.

When looked at objectively, it would seem that the weight of evidence is against 25 being a real interactor of BMI-1. It looks, in all probability likely to be a false positive that, by ectopic expression in the yeast nucleus, was detected in a similar fashion to Clathrin and Coronin in Results Section 3.4.1 - a weak, probably non-specific, interaction was detected in yeast, which, when further analysed by GST capture was greatly enhanced through the laws of mass action. If this is so, it shows the fallibility of both the yeast two-hybrid and the GST-affinity capture assays in investigating protein interactions. However, until a complete cell cycle-dependent localisation study of the polypeptide encoded by 25 has been carried out a question mark will hang over the subcellular localisation of this protein, and the resultant ability to bind to BMI-1.
4.3 MCM Proteins and DNA Replication

Before any connections between MCM6 and BMI-1 can be made, it would be useful to discuss mechanisms of replication and the involvement of the MCM proteins in this process.

4.3.1 DNA Replication

DNA replication is an essential process that has progressively evolved with increased cellular complexity. Replication of the relatively small genome of prokaryotic cells involves the specific interaction of replication proteins with a well-defined DNA sequence which acts as an origin for the initiation of DNA replication. In eukaryotes, the presence of both larger genomes than prokaryotes and multiple chromosomes, results in the synthesis of DNA during the S phase of the cell being achieved through the use of multiple replication origins.

As well as ensuring that DNA is accurately replicated, one crucial feature of the replication process is to ensure that replication of DNA occurs only once during each S phase before the completion of mitosis. Over-replication would be expected to have disastrous consequences for the cell (for example, in terms of gene dosage effects) when the sister chromatids are pulled apart during mitosis.

Within eukaryotes, one of the best-studied systems of DNA replication is that of the budding yeast Saccharomyces cerevisiae. Early studies by Rao & Johnson (1970) provided important evidence into the mechanisms of re-replication prevention. By fusing G1- or G2-phase cells with S-phase cells, they found that only the G1 nucleus and not the G2 nucleus could replicate the DNA. Replication of the G2 nucleus DNA required that the hybrid cell passed through mitosis. This evidence, along with studies on the effect of nuclear permeabilization on DNA replication in Xenopus egg extracts ultimately led to the 'licensing factor' hypothesis (Blow & Laskey, 1988).

In this model, a two-step process has emerged that explains why origins of replication fire only once per cell cycle. In the first step, which occurs during early G1, a pre-replicative complex (pre-RC) is established on chromatin by the sequential binding of ORC, Cdc6p and the MCM proteins. The ORC and Cdc6p proteins are thought to act as loading factors that allow the recruitment the MCM proteins (Donovan et al., 1997). In the second step, which defines the G1-S phase transition, the cyclin-dependent kinases and Cdc7/dbf4 kinase are activated, DNA replication is initiated and Cdc6 and MCMs
are lost from the origin (reviewed by Dutta & Bell, 1997). In addition, S-, and M-phase-specific kinases (possibly cyclin E/cdk2) then prevent the rebinding of MCM proteins to chromatin (Dahman et al., 1995, Piatti et al., 1996, Tanaka et al., 1997), thereby preventing reinitiation until after the next mitosis. The long-known role of the nuclear envelope in regulation of DNA replication (for example in experiments where perturbation of the nuclear envelope disrupts DNA replication) is a crucial factor in this model. The nuclear envelope is thought to create, through selective nuclear transport, an intranuclear environment that is permissive for DNA replication (Leno & Laskey, 1991, Cox, 1992, Walter et al., 1998). According to this model, during interphase, an inhibitor is gradually concentrated within the nucleus. As DNA replication occurs, the inhibitor prevents dissociated MCM rebinding, thus preventing re-replication. Following the nuclear envelope breakdown in mitosis the inhibitor is diluted by the cytoplasm, thus allowing the renewed binding of MCM to chromatin.

Whilst comparatively little is known about replication in higher eukaryotes, certain aspects of this model have been observed, for example the conservation of MCM proteins in a wide range of species, from yeast, to Drosophila, to humans.

4.3.2 MCM Proteins

The products of the \textit{cdc} genes and the products of the minichromosome maintenance (MCM) genes in \textit{Saccharomyces cerevisiae} have provided candidate genes for the initiation licensing factor. Mutations of the \textit{cdc} genes result in the arrest of DNA synthesis at the G1/S transition or at the beginning of S phase, while mutations of the MCM genes result in an inability to maintain the extrachromosomal replication of minichromosomes (~large scale screening for mutants with this particular inability led to their initial discovery (Gibson et al., 1990, Yan et al., 1991, Chen et al., 1992)).

Human cells, like other eukaryotic cells, express as least six MCM proteins conserved in the central region. MCM proteins are abundant nuclear proteins with copy numbers ranging $10^4$-$10^6$ per cell nucleus, and they occur either free in the nucleosol or are bound to nuclear chromatin (Holthoff et al., 1998). MCM proteins are associated with chromatin in a cell-cycle dependent manner, with the fraction of chromatin-bound MCM proteins highest at the end of the G1 phase or the beginning of the S-phase of the cell cycle, but gradually decreasing as S phase proceeds. In post-replicative cells, essentially
all MCM proteins are free in the nucleosol and disperse throughout the cell upon nuclear breakdown in mitosis, but they rapidly reassemble on chromatin sites in nuclei after mitotic telophase (Kubota et al., 1995, Tsuruga et al., 1997). In human cells, MCM4, -6 and -7 proteins form a stable complex, and MCM2 is loosely associated with this complex (Schulte et al., 1996, Ishimi, 1997). MCM3 and -5 proteins also form a stable complex. All six of the MCM proteins contain DNA-dependent ATPase motifs in a central domain that is conserved from yeast to mammals, and these motifs are found in several enzymes that unwind the DNA duplex (Koonin, 1993).

4.3.3 Functions of MCM Proteins

Despite the fact that MCM proteins have been identified in a variety of species, very little is known relating to their role in replication. Since MCM proteins have the domains required for DNA-dependent ATPase activity, and a DNA helicase activity is associated with a complex of MCM4, -6, and -7 proteins (Ishimi, 1997) it may be that they function as the DNA helicase that is required for DNA unwinding during DNA replication. With MCM proteins being thought to act as hexamers, functional similarities have been made between the MCM proteins and the replicative helicases of E.coli (DnaB) and SV-40 (T antigen) which also function as hexamers (Lohman & Bjornson, 1996, West, 1996). One problem that is apparent though, is that the helicase activity associated with in vitro MCM complexes is weak compared to other known replicative helicases, though this may be explained by the absence of protein cofactors that stimulate the activity in vivo. Aparicio et al. (1997) found that while MCM proteins were found to specifically associate with replication specific DNA sequences during G1 phase, their association with specific sequences changed during S-phase, moving from replication sequences to other, non-replication, sequences. In further studies that examined changes in distribution of MCM proteins in a region of chromosome III during a synchronous cell cycle, it was found that their distribution closely paralleled that of DNA polymerase (Pol ε), a known component of the replication fork. Assuming that the results reflect the actual distribution of proteins on chromatin, this suggests that the MCM proteins are part of the replisome (see Figure 53).
Recruitment
(Pre-RC Assembly)

Initiation
(Replicative Complex (RC) assembly)

Elongation
(DNA synthesis)

Figure 53 – A model of DNA replication complexes during the Cell Cycle (as proposed by Aparicio et al., 1997). After binding the origin DNA, the ORC-origin complex is recognised by Cdc6p, which recruits MCMs to the origin forming the preRC. Activation of this complex leads to the formation of unwound DNA at the origin, and the recruitment of the DNA replication enzymes (e.g. DNA Pole) and their accessory proteins to the origin forming the replicative complex (RC). After initiation, MCM proteins and Cdc45p become components of the resulting replicative complex at the replication fork.

An alternative explanation for the apparent movement of the MCM proteins with the replication fork is that MCM proteins are actually distributed relatively uniformly along chromatin but are inaccessible during G1 except at replicator sequences, and that the movement of the replication fork exposes the MCM proteins as it approaches. Of interest is the observation that two proteins, MCM3, and MCM6 may also occur independently on chromatin as monomeric units not tightly linked to other MCM proteins (Holthoff et al., 1998).
In addition, apparent links between MCM proteins and the core tetramer of histones (H3 and H4) (Ishimi et al., 1996) have been used to suggest that the MCM proteins may play a role in DNA replication by remodelling chromatin structures, similar to the SWI/SNF/NURF chromatin remodelling factors (Cote et al., 1994, Kruger et al., 1995), preparing the chromatin for efficient passage of the replication fork. In this regard, DNA-dependent ATPases are also known to function in chromatin remodelling as well as origin unwinding.

4.3.4 Cloning and Characterisation of hsMCM6

hsMCM6 was identified as the human homologue of the mis5+ gene of *S. pombe* (Holthoff et al., 1996, Tsuruga et al., 1997). Using a fragment from the Xenopus mis5 homologue, a full length cDNA was obtained that contained an open reading frame of 3624 bp encoding an 821 amino acid protein with a predicted mass of 93.1 KDa and an apparent electrophoretic mass of 105KDa. Motifs present within the protein include a characteristic DNA-binding zinc finger and a MCM/P1 family motif.

One interesting feature of MCM6 has been observed in Xenopus. Distinct maternal and zygotic genes have been identified for MCM6 (Sible et al. 1998). Specifically, zMCM6 contains an additional 21 amino acids at the carboxy terminal and consensus phosphorylation site for cyclin-cdk complexes that is missing from maternal MCM6. The presence of this phosphorylation site and the role of cyclins in preventing re-initiation suggests that differential regulation of mMCM6 and zMCM6 may occur, especially at the midblastula transition where zMCM6, along with zygotic gene transcription in general, is first seen.

4.3.5 MCM6 as a Candidate Interactor of BMI-1

Similar for-and-against arguments could be postulated for the MCM6 protein as for clone 25 relating to the interaction with BMI-1. The occurrence of a similar MCM C-ter fragment (clone 54 and v45) from libraries from two distinct species is indicative of a conservation of function and so suggests a potential interaction. The interaction between the same C-ter fragment (as a GST fusion protein) with both the truncated BMI-1 and full-length BMI-1 also seems suggestive of a real interaction. However, the lack of
interaction in vivo with either endogenous MCM6 or epitope-tagged MCM6 would indicate otherwise. However, one fundamental difference is observed between the two proteins encoded by clone 25 and MCM6 - that of their respective known cellular distributions. MCM6, as clone 25, lacks an obvious Nuclear Localisation Signal. Despite this, Tsuruga et al. (1997) showed that while present in the cytoplasm continuously, the nuclear distribution of MCM6 is cell-cycle dependent. Furthermore, within the nucleus, the chromatin association is also cell-cycle dependent. Interphase nuclei were found to contain MCM6 with a peak at the G1/S phase boundary. The timing of re-association in the nucleus at telophase was apparently before nuclear membrane formation, which would account for the lack of Nuclear Localisation Signals seen in the protein sequence of MCM6. Thus, the presence of MCM6 in the nucleus would permit an interaction to occur with BMI-1.

One possibility is that the C-ter domain, or fragment, isolated, is capable of interacting with BMI-1, but that either the structure of the domain in isolation bares no resemblance to its structure within the full-length MCM6 protein overall or a post-translational modification of some kind masks or alters the ‘BMI-1’ binding site on MCM6. As such, it would appear to be a false positive.

One other option is that it does represent a biological interaction. If this is the case, what is its possible significance? With no evidence to suggest that BMI-1 is a member of the replication complex, one option may be that the interaction is a transient one that occurs as a result of the passage of the replication fork through chromatin. While studies on Pc-G proteins have focussed on the association of Pc-G with chromatin during mitosis (Buchenau et al., 1998, Saurin et al., 1998), virtually nothing has been researched into the association of Pc-G proteins with chromatin during S-phase, the most logical time for the inheritance of Pc-G mediated repression to occur as discussed in the Introduction Section 1.4.3. A little more work has been done to study the inheritance of chromatin during S-phase, with the current view that the replication machinery moving along the DNA transiently disrupts the chromatin fibre and a random segregation of parental nucleosomes to the daughter strands occurs (reviewed by Gruss & Sogo, 1992). It has been postulated that the nucleosomes dissociate during the passage of replication into H2A/H2B dimers and (H3.H4)$_2$ tetramers. The (H3.H4)$_2$ tetramers are thought to segregate to both of the daughter strands with the addition of the H2A/H2B
dimers completing the assembly of the nucleosome (Jackson, 1987, 1990). With the possibility that the MCM proteins could act as a chromatin remodelling DNA-helicase to unwind the parent strand, it may be that it is the MCMs that are responsible, in part at least, for the disassociation and/or segregation of the nucleosomes. This may explain the interaction seen between certain MCM proteins and Histone H3 (Ishimi et al., 1996).

As with the random segregation of parental nucleosomes to the daughter strands, so too might the Pc-G complexes distribute to the daughter strands in a similar fashion, again potentially mediated by the action of MCM proteins. If this were to be the case, this could provide a mechanism for the inheritance of Pc-G repression. If even a fraction of the Pc-G proteins were distributed in such a fashion, they could potentially act as nucleation markers to facilitate the recruitment of other Pc-G members and the subsequent reforming of repressive Pc-G complexes. While the BMI-1-MCM6 interaction may occur, the functional significance may be secondary and so, in overall terms, this could be viewed as a non-informative biological interaction.

Another possibility is that the replication fork moves along the DNA with the MCM protein complex disrupting the chromatin fibre as described previously. This time, however, the Pc-G proteins, instead of being shuttled to the daughter strands, are simply displaced from the DNA strand entirely by the passage of the MCM complex, and it is another mechanism that is responsible for the epigenetic inheritance of the Pc-G proteins. One possibility could be that histone acetylation plays a role in this inheritance. One line of evidence for this comes from Ekwall et al. (1997). They found that treating S.pombe for five cell doublings with the histone deacetylase inhibitor trichostatin A (TSA) causes centromeric heterochromatin to become hyperacetylated. This hyperacetylated state resulted in derepression of reporter genes at centromeric heterochromatin and defects in chromosome segregation, and, surprisingly, was still observed even after 80-100 generations after the TSA had been removed, suggesting that the state of histone acetylation may be an important factor in determining patterns of heterochromatin inheritance. A connection between this hypothesis and the observation that Xenopus Pc and human Pc2 are capable of binding the C-terminal binding protein (Sewalt et al., 1999) (Introduction Section 1.4.2), which itself can bind the histone deacetylase HDAC1 (Sundqvist et al., 1998), could provide the key to unravel the mystery of Polycomb-Group inheritance.
Were either of these two possibilities to be the case, the transience of the MCM6-BMI-1 interaction could explain why it failed to be detected \textit{in vivo}, through techniques such as co-immunoprecipitation. The advantage of the yeast two-hybrid system over other techniques is that it is capable of detecting weak or transient interactions that other, physical, methods can miss. In addition, if the interaction only occurred during S-phase, the chances of detecting an interaction would be greatly reduced since, for an unsynchronised population of cells the proportion of cells in S phase is relatively small. An alternative explanation regarding the lack of detection \textit{in vivo} could be explained by the relative numbers of protein molecules within the cell. The introduction of the FLAG-MCM6 eukaryotic expression construct into cells would be expected to increase the level of MCM6. However, with such high numbers of endogenous protein present anyway ($10^4$-$10^6$ molecules per cell nucleus), the FLAG-MCM6 contribution could be relatively insignificant and competition between endogenous MCM6 and FLAG-MCM6 could still favour the endogenous MCM6 functioning as the biologically relevant species.

Ongoing work in relation to this interaction has been hampered through the very transience of the potential interaction and of the DNA replication machinery in general. The two-hybrid clone 54, as well as the FLAG-epitope MCM6 have both been sent to \textit{Maarten van Lohuizen} for further analysis.

4.4 Methods Used to Investigate Protein-Protein Interactions

4.4.1 The Yeast Two-Hybrid System

4.4.1.1 Suitability of the System.

At the time this project started it was predicted that mammalian Polycomb-Group proteins might act in large multi-protein complexes similar to the Pc-G protein complexes found in \textit{Drosophila} and subsequent evidence from other laboratories has demonstrated this to be true. Furthermore, a significant proportion of these interactions were originally identified and characterised using yeast-two hybrid methodology (\textit{Alkema et al., 1997a, Gunster et al., 1997, Satijn et al., 1997, van Lohuizen et al., 1998, Sewalt et al., 1998}), thus demonstrating the suitability of the yeast two-hybrid system for investigating protein interactions. Several of these interaction experiments have involved
screens using either the murine Bmi-1 or its human homologue, indicating that this particular protein can be used successfully in such screens. One particular screen of interest, performed by Gunster et al. (1997), utilised the same Binding Domain vector, pAS2-1, and the same yeast strain, Y190, as used in this study. Noticeable differences between their study and this were that Xenopus BMI-1 (XBmi-1) was used instead of human BMI-1, the concentration of 3-AT used was 30mM not 50mM, and the library screened was a human leukocyte library as opposed to a pre-B-cell library or mouse embryo library.

4.4.1.2 Ease of Use?

The yeast two-hybrid system has developed as an immensely useful tool in analysing protein-protein interactions. It is not, however, without its limitations and problems. The majority of system-related problems are associated with the fact that each protein used is different and unique, and so problems encountered are likely to be different for each protein. There is no ‘hard and fast’ set of rules relating to protein structure and function in this regard. For example, simply by examining the BMI-1 protein sequence, it would have been impossible to predict whether or not the Gal4-BD fusion of this protein would have autoactivated the HIS3 reporter gene. In addition, there is no way of predicting whether any given yeast two-hybrid system will work for any protein. One example of this was observed for the HirA transcriptional repressor. Using several different HirA constructs and exhaustive screening with a GAL4-based yeast two-hybrid system, whilst several positives were identified, none appeared significant or reproducible, and invariably only one of the reporters was activated. Using a LexA-based yeast two-hybrid system, however, several positives have been identified, including Histone H3 and Pax3 (Magnaghi et al., 1998), both of which have been corroborated as interactors of HirA using other techniques. Such factors cannot be determined a priori and can only be deduced by empirical means.

As well as protein-related obstacles, there are the problems encountered relating to the maintenance and manipulation of yeast cultures, especially in a laboratory where yeast-based techniques were not routine. Reversion events can occur, and so yeast stocks need to be constantly maintained and checked for genotypic and phenotypic alterations. In terms of yeast manipulation, for applications where a relatively low transformation efficiency is required, e.g. simple cotransformations, experiments are relatively
straightforward. However, for library screening, where a high level of efficiency is essential, care is required for each of the harvesting, heat-shock and recovery stages. Even slight deviations can dramatically reduce the overall efficiency, with, in the case of two-hybrid screens, a concomitant reduction in number of positives identified.

There are, in addition to the methodical problems, some protein-protein interactions which cannot be detected. For example, it may not be possible to detect interactions involving extracellular proteins due to glycosylation events and disulphide bond formation, events generally incompatible with a nuclear-based system. Thus, the system may be of limited use in analysing, for example, receptor-ligand processes.

4.4.1.3 Yeast Strains

As mentioned in the Results Section 3.6, differences can arise between strains, resulting in one strain being more suited for two-hybrid analysis than other strains. This is probably best borne out by differences between Y190 and PJ69-4A. From a functional viewpoint, the main differences relate to the promoters controlling reporter expression, with PJ69-4A being more sensitive and at the same time less prone to false positives through the use of different GAL4 responsive promoters. However, differences exist beyond simply the promoter controlling the activation of individual reporters. Each yeast strain will be slightly different in terms of its functional genotype, or 'background', i.e. the numbers of genes expressed, the time of gene expression and the relative levels of genes expressed. These differences can act to enhance or reduce the ability to detect significant interactions, and also to differentiate these from false positives which may also occur. As can be seen from the library screens carried out in Y190, the occurrence of false positives is relatively common, and can interfere with further analysis of real positives. Providing a cut-off between real and false positives is, to a large extent, arbitrary, and in the case of some of the ESTs identified can prove exceedingly difficult. How many of these false positives would have been screened out by using PJ69-4A remains unknown, but it would seem reasonable to surmise that the proportion of false positives would have been reduced.
4.4.1.4 Binding Domain Vectors

The choice of Binding Domain vector can have a critical effect on the results of a library screen. pAS2-1 was chosen over pGBT9 due to the relatively increased expression of the GAL4-fusion protein seen with the former plasmid (Clontechtechniques, 1995). This is thought to occur due to the increased length of the ADH1 promoter in pAS2-1. However, expression of the GAL4-fusion protein can be affected in other ways. Louvet et al. (1997) analysed pAS2-1 expression, stability and transformation efficiency and compared it with a new DNA-binding vector (pODB8) based on pRS424 and carrying the expression cassette from pAS2-1. They found that the pAS2-1 vector was unstable when compared with pODB8, leading to disadvantages such as weaker protein expression, frequent plasmid rearrangements in yeast and low transformation efficiency in library screening, all of which can reduce the likelihood of detecting real interactions. Despite the disadvantages of this vector, pAS2-1 has been successfully used in conjunction with Y190, the same yeast strain as in this study, to identify the human homologues of Mph1 (HPH1 and HPH2) (Gunster et al., 1997) as interactors of BMI-1.

4.4.1.5 The Use of Full-length vs. Truncated Protein as Bait

One strategic dilemma that can occur is whether to use the full-length protein, or not. Sometimes it may be beneficial, especially for larger proteins, to use a fragment of a protein as opposed to the full-length protein, while in other scenarios, the full length protein may be essential to determine interaction partners. One example, which though not determined using the yeast two-hybrid system, demonstrates differing interactions seen between full-length and truncated c-Myc protein interactions. c-Myc is a sequence-specific DNA binding protein that requires heterodimerisation with Max for its activity. Blackwell et al. (1990) showed that the c-Myc homodimer is not observed in vivo. This is presumably due to steric constraints within the full-length protein, since truncated c-Myc proteins containing just the C-terminal B-HEL-LZ (Basic Helix-loop-helix Leucine Zipper) Motif can dimerise in vitro. Thus the steric influences of the remainder of the protein can affect protein interactions. In addition, the GAL4 binding domain can also potentially alter binding capabilities, through steric constraints.

In the case of whether it would have been better to have used the full-length BMI-1 protein instead of a truncated fragment, it remains difficult to say one way or the other. At the time, the function of the zinc finger (the RING finger) was not known. Early
studies speculated that it could have been a putative DNA binding motif (Lovering et al., 1993), and if this were the case, then it would have been wise to remove it to prevent any interference with the reporter systems used. It has since been shown that this DNA-binding activity should be considered to be more artifactual than of true in vivo relevance (Borden & Freemont, 1996). Instead, the RING finger is thought to be involved in mediating protein–protein interactions (Hemenway et al., 1998). Even so, it is still conceivable that some investigators might decide to remove this domain in order to enrich for other, possibly weaker, interactions.

One other difference relates to the relative strengths of interactions as determined by the LacZ assay. If for example, using the full-length protein, a positive had been isolated, which gave a liquid β-Gal assay result as strong as that of Mph1, were other positives to have been identified that were a lot weaker (for example, compare the relative signals as shown in Figure 20), they may well have been ignored as false positives. In cases such as this, one has to be especially careful not to assume that the strength of interaction is necessarily related to the biological significance of the interaction in question.

4.4.1.6 Maximising Transformation Efficiency

One of the main factors in determining whether or not a two-hybrid screen will be successful is related to the efficiency of transformation. Unlike screening genomic or cDNA libraries using a DNA probe, there are no simple empirical formulae that can be employed in order to predict the number of independent colonies which must be screened. Screening for protein interactions, due to the very nature of protein interactions, is subject to multiple factors such as whether cofactors are necessary to mediate an interaction, or how many domains within a protein are necessary to mediate an interaction. Consequently, perhaps one of the few guidelines that can be employed is obvious – maximise the transformation efficiency to maximise the chance of an interaction taking place. In each case, in an effort to maximise the transformation efficiency, a two-step strategy was adopted with the bait plasmid being transformed first, and yeast containing the bait subsequently transformed against the library plasmids. This two-step method is thought to greatly increase the transformation efficiency, rather than the less efficient, conventional, cotransformation method of introducing both plasmids into the yeast simultaneously. In addition, dividing the transformation into multiple
smaller aliquots and the use of Sorbitol both appear to help increase the overall transformation efficiency. The role of Sorbitol is unknown, though it may be possible that it acts to prevent osmotic shock.

4.4.1.7 The Choice of Libraries

The choice of libraries, B-cell and mouse 9.5/10.5dpc, in this project reflected what is known about BMI-1. BMI-1 is expressed in these tissues / at these stages, and one would hope that some, if not all, interactors of BMI-1 are similarly expressed. In addition, the quality of the libraries can also influence the two-hybrid results. A comprehensive two-hybrid library must be of sufficient complexity not only to represent each gene, but also to provide as many different end-points within each gene as possible. The library must also include fusions that encode the amino terminus of as many proteins as possible, though this region is the most difficult to represent adequately. Other potential problems in library construction relate to the reading frame of the cDNAs of interest. Whilst only one reading frame may be correct, cDNAs may be cloned into a vector in either orientation resulting in six potential reading frames. The use of directional cloning has partially overcome this by ensuring that cDNAs can only be in one orientation, reducing the number of potential reading frames by half to three. The libraries used in this project were obtained from other laboratories where such two-hybrid analyses are routine (see Materials and Methods Section 2.1.10).

4.4.1.8 Specificity vs. Sensitivity

When using the yeast two-hybrid system, a compromise has to be made between the relative needs for specificity and sensitivity. Usually, higher specificity is required because a large number of false positives can obscure the biologically relevant interactions. However, there is a trade-off between specificity and sensitivity. Increasing the specificity of a screen usually leads to a reduction in sensitivity and thus a greater number of false positives.
4.4.2 GST-Affinity Capture

As with any technique, there are advantages and disadvantages to using GST-affinity capture to investigate protein interactions. The first of these is with regard to the purification of the GST-fusion protein itself. Whilst, in general, such fusions are often soluble, there can be a problem of insolubility for larger fusion proteins (as a guideline, fusion proteins >50 KDa). BMI-1 would belong to this category, with the GST-BMI-1 fusion being approximately 75 KDa, thus possibly explaining the problems experienced in obtaining the GST-BMI-1 fusion protein. It was with some consolation to find that the Dutch group led by Maarten van Lohuizen working on BMI-1 have also had problems with the solubility of GST-derived BMI-1 fusion proteins. Purification of the pGEX-Bmi-1 is possible, although the necessary conditions were not provided in the literature (Alkema et al., 1997a).

For GST-fusions of clones 25 and 54 successful purification was achieved, and in both these cases it was interesting to see that the recovery of protein, while not measured in a quantitative manner, appeared to improve with a lower temperature of induction. In addition, one main advantage was that denaturing conditions (for example the use of the detergent, sarcosine) were not required at any stage during purification and consequently, the fusion polypeptides may retain their functional activity and, though not relevant to this project, their functional antigenicity.

The sensitivity of this assay rests in the mass action effects of using high concentrations of bait which drives the detection of even weak interactions. However, mass action effects can also enhance the formation of non-physiological and non-specific interactions, thus creating the potential for artifacts. This is of particular significance in a scenario such as the one used in these experiments where the target protein is synthesised in vitro in relatively small quantities. False positives, which may occur through weak non-specific interactions, may thus be classed as potentially real interactions.

Artifacts may also arise by taking a protein out of its normal context. A protein that normally exists as a complex may, when expressed as a fusion protein, expose surfaces that are not normally exposed. These surfaces may be 'sticky' and lead to a general affinity for other proteins. Thus, because of the potential for artifacts, GST capture alone is not sufficient to establish that an interaction takes place in vivo, but is merely used to support evidence from other independent assay methods.
Despite these disadvantages, the GST-affinity capture technique remains powerful due to its simplicity. It is quicker to perform than two-hybrid interaction assays, and, depending on the ability to quantitate the amount of the bait and test protein, the binding constant and the half-life of the binding reaction can be determined. The simplicity of the method is of particular use when determining regions within a sequence that are sufficient for interaction. Different regions or fragments of both the bait and test proteins can be generated and tested against each other in a binding assay, quickly narrowing down the regions required to mediate interactions. Amongst other advantages are that interaction assays can be performed in a defined environment, in contrast to yeast-based experiments. In addition, step-wise interactions can be investigated, for example, those interactions that depend upon the presence of a third protein or the phosphorylation state of a given protein.

4.4.3 Co-immunoprecipitation Techniques

The environment defined by the mammalian cell is, from a biological perspective, the most representative of an in vivo setting, and so the identification of protein interactions from within these cells is probably one of the best means of establishing an interaction. The correct folding of proteins, the presence of cofactors, along with all the post-translational modifications that occur all contribute to the functional in vivo nature that is so difficult to reconstitute in vitro (n.b. while the yeast two-hybrid system is essentially an in vivo technique, the fact that the cells are yeast and not mammalian opens the possibility of artifactual interactions). Co-immunoprecipitation assays to detect protein-protein interactions probably remain the most highly regarded out of any of the three techniques used in this study to characterise interactions. Because of the high specificity of antibodies, specific antigens that are minor components of complex mixtures can be isolated. This high specificity can, however, work against the investigator and one should also be aware that, as with GST affinity capture, the technique of co-immunoprecipitation is essentially an in vitro technique. For example, different sample buffers used can lyse membranes and dissociate complexes to different extents, and similarly, the pH of buffers used can promote or inhibit protein aggregate formation.

One of the basic requirements for immunological studies is that it is essential to have a good working antibody. Raising antibodies against peptides from specific proteins can be
expensive with no guarantee of success. Peptide design is not an exact science, and while there are certain guidelines which can be helpful (e.g. choosing hydrophilic sequences and those containing prolines rather than hydrophobic sequences \((\text{Hopp & Woods, 1981})\) and the use of C-ter and N-ter sequences), is it by no means an exact art. For example, in this project, of the 6 sets of sera raised against peptides of interest using a combination of these criteria, only one was found to react against the epitope to which it was designed. In this regard, the introduction of epitope-tag eukaryotic expression vectors to which there are well-characterised, commercially available antibodies has greatly simplified experiments, making protein identification, purification and analysis both relatively cheaper, and in some cases, easier. Another application of epitope tags is the investigation of the subcellular localisation of a given protein, and this was of particular relevance in investigating the subcellular localisation of clone 25. However, one should bear in mind the potential problems that can arise with epitope-tagged proteins. Firstly, the epitope could alter the structure of the protein as a whole, or specific domains within the protein, leading to altered function. Second, using constitutively eukaryotic expression vectors, the protein in question may be overexpressed so that its level within the cell is no longer physiological. This may result in aberrant targeting and non-significant complexes forming within the cell. Alternatively, if the protein is part of a multiprotein complex overexpression of one of the constituents would alter the mass-action equilibrium that exists favouring complex formation. The introduction of inducible expression systems should help to alleviate at least some of these problems.

4.5 Future work

As mentioned, the three principle clones of interest, 16 (SRp20), 25 (EST A7CO8) and 54/MCM6 have all been sent to Maarten van Lohuizen for further analysis. There are several techniques available within the laboratory which could be used to investigate the interactions further. One obvious technique to test for the interactions would be to see whether the interactions detected in the yeast two-hybrid system can also be detected in the mammalian two-hybrid system. If corroborated in this way, it would suggest that the interaction is not an artifact seen as a result of expressing the proteins in yeast cells.
Additional techniques such as confocal microscopy could be used to investigate the
subnuclear localisation and co-localisation of BMI-1 and either SRp20 or MCM6 within
the cell nucleus. Antibodies are available against SRp20 as well as other SR member
splicing factors. These resources could allow the investigation of whether the interaction
does occur, and if it does, is it specific to BMI-1 or can BMI-1 bind other splicing
factors. Do other Pc-G members interact with other splicing factors?

Alternatively, FACS (fluorescence activated cell sorting) analysis could be used to
specifically enrich populations of cells at certain points within the cell cycle, e.g. S phase.
This enriched population of cells could then be used for a series of immunoprecipitation
studies. Alternatively, if the antibodies used for immunoprecipitation were thought to
disrupt the interactions, immunoaffinity purification may be more suited for investigating
the interactions. For example if the interaction is enhanced by other proteins within a
complex, the purification of the holocomplex could permit the detection of the
interaction. One example of such a purification procedure was shown for the PCAF
Histone acetylase complex (*Ogryzko et al., 1998*). A cell line was established expressing
FLAG epitope-tagged PCAF, and nuclear extracts from this cell line were subjected to a
series of size chromatography and affinity chromatography (anti-FLAG and anti-PCAF
affinity columns) steps. Proteins contained within the complex were then analysed by
SDS-PAGE. Specific bands from Coomassie stained SDS-PAGE gels were then excised,
digested with trypsin and analysed by mass spectrometry. Using this procedure PCAF
was found to exist in a complex of more than 20 distinct polypeptides with several TAFs
(TATA-binding protein associated factors) being identified as constituents.

The MCM6-BMI-1 link could be very informative, however, a major problem is that
resources relating to DNA replication, and MCM proteins in particular, are less abundant
than for the splicing factors. In addition, the detection of the interaction could prove
difficult due to the very transience of the presumed interaction. One method that may be
of some use is similar in principle to a method used for drug screening, the *Scintillation
Proximity Assay*. This method consists of a free radiolabel (donor) and an acceptor
molecule. If the two are apart from each other the energy emitted by the radiolabel is
absorbed by the medium (*Figure 54A*). If, however, the two are brought into proximity
to each other the energy from the radiolabel can pass to the acceptor, stimulating it to
emit light (*Figure 54B*) which can then be detected.
It may be possible to adapt this process to investigate protein interactions. Through recombinant techniques it might be possible to couple an electron donor to e.g. BMI-1 and an electron acceptor to e.g. MCM6. The acceptor could be chosen so that it could emit light of a certain wavelength which could be detected by a fluorescent microscope with suitable filters. If the two proteins were in proximity with each other, as described in Discussion Section 4.3.5 this could be detected by the emission of light. Furthermore, if expression of these proteins could be inducible, the levels could be adjusted to represent normal physiological protein concentrations. This technique could have the advantage over some other methods of interaction detection in that it could be followed in living cells throughout the cell cycle.

Figure 54 – The principle of the scintillation proximity assay. (A) The donor and acceptor are not in close proximity and energy emitted by the donor is absorbed by the surrounding medium; (B) The donor and acceptor are in close proximity and energy emitted by the donor is transmitted to the acceptor causing the acceptor to emit light.

In terms of the Polycomb-group as a whole, the field is still relatively new and there are many questions about which we know almost nothing about. Why are some Polycomb group members ‘weak’ and others ‘strong’ as determined by the phenotypic defects seen in mutant mice? Does this reflect differing roles within Pc-G complexes, for example would a mouse that is mutant for a Pc-G member thought to be involved in Pc-G complex nucleation exhibit a stronger phenotype than a mouse mutant in a non-nucleation Pc-G member? Or, do the weak and strong phenotypes reflect the action of
Pc-G proteins on different target genes through the formation of different Pc-G complexes?

Whilst work is progressing on investigating the composition of Pc-group complexes, very little is known of their actual function. To investigate the nature of Polycomb-group protein action it would be interesting to investigate whether Pc-G repressed regions are early replicating or, like heterochromatin, are late replicating. What is the nature and significance of a link between the C-Terminal binding protein, the histone deacetylase HDAC1 and Polycomb-group proteins? Would the histone deacetylase inhibitor trichostatin A (TSA) have an affect on Pc-G mediated repression or the epigenetic inheritance of the repression pattern?

Another potentially interesting area is the connection between the Pc-G proteins and the inheritance of kinetochore activity – do the various Pc-G mutant mice, e.g. the Bmi-1-/- and M33-/- mice, as well as exhibiting posterior transformations exhibit defects in chromosomal structure and segregation?

4.6 Conclusions and Hindsight

The main crux of protein-protein interaction determination is the differentiation between: 

i) informative biological interactions (representing real and significant interactions);

ii) non-informative biological interactions (those interactions that occur in, for example, a degradation pathway such as the ubiquitin-dependent proteolysis pathway); and

iii) non-significant, false, interactions (those that do not occur in nature due to expression at different times or in different locations). Whilst methodology and technology has improved (the use of improved strains, more refined reporter systems and the use of inducible libraries have all been utilised in the elimination of false positives with varying degrees of success), as can be seen from the three screens carried out the identification of false interactions remains a problem.

As with almost everything in life, there is a saying that ‘hindsight is a wonderful thing’, and this project is no exception. Perhaps the main problem associated with this project only became apparent relatively late on, with the identification and differentiation of positives. Using the pAS2-1-tBMI-1 construct as bait, both clones 25 and 54 were identified as positives when compared to the appropriate negative control vector.
combinations (pAS2-1-tBMI-1 vs. pACT and pAS2-1 vs. clones 25 and 54). From the three yeast strains tested, auxotrophic marker selection and the LacZ colorimetric assay provided evidence to suggest that the interactions occurred. It remains difficult to determine relative interaction strength simply by using the auxotrophic markers, however, clear growth vs. non-growth was observed which allowed the differentiation between an interaction occurring or not. It was not until the fBmi-1+Mphl combination was tested alongside the tBMI-1+25/54 combinations that the relative scale of 'positiveness' was identified. To adapt an Orwellian phrase, 'all clones were tested positive – some more positive than others'. However, as mentioned earlier, the strength of an interaction as determined by yeast two-hybrid analysis, while perhaps suggestive of an interaction occurring, is by no means a guide to the biological significance of the interaction.

Perhaps using the full-length BMI-1 would have made a difference to the overall outcome, but then again perhaps not. Given the strength of interaction of the fBmi-1-Mphl interaction, it is quite conceivable that lesser, though potentially more interesting, interactions may have been missed. Also, as can be seen from these experiments, using the truncated BMI-1, interactions have been detected which screening with the full-length BMI-1 might have missed. The use of PJ69-4A, had it been available at the time, might have aided positive identification, though such things are purely speculative.

Overall, like any system, the yeast-two hybrid has its advantages and it has its limitations. Without doubt, the fact that it allows the genetic selection of genes encoding potential interacting proteins without the need for such time and labour-intensive biochemical is a major plus point. Advances are continuously being made, which along with a suitable degree of human input, will allow the correct identification and differentiation of real positives from false positives. Considering most of these advances have been made in the last 5 years or so, it is conceivable that technology will have advanced in the next 10 years or so to permit a vastly increased avenue of applications for the system, from functional proteomics, to disease and allele investigation, to drug discovery and beyond.
References


Tie F., Furuyama T., Harte P.J. (1998). The Drosophila Polycomb group proteins esc and E(z) bind directly to each other and co-localise at multiple chromosomal sites. Development 125, 3483-3496.


Warburton P.E., Cooke C.A., Bourassa S., Vafa O., Sullivan B.A., Stetten G.,
Gimelli G., Warburton D., Tyler-Smith C., Sullivan K.F., Poirier G.G., Earnshaw
at the inner kinetochore plate of active centromeres. Curr Biol 7, 901-904.

methylation of genomic sequences in plants. Cell 11, 567-576

Wedeen C., Harding K., Levine M. (1986). Spatial regulation of Antennapedia and
bithorax gene expression by the Polycomb locus in Drosophila. Cell 44, 739-748.

maternal terminal pattern information in Drosophila. Science 248, 495-498.


pumps. Cell 26, 177-180.

kinetochore assembly (review). Cell 93, 313-316.

193, 296-307.

120, 587-590.

Yan H., Gibson S., Tye B.K. (1991). Mcm2 and Mcm3, two proteins important for
ARS activity, are related in structure and function. Genes Dev 5, 944-957.

family of pre-mRNA splicing factors. Genes Dev. 6, 837-847.

Polycomb protein with Antennapedia regulatory sequences in polytene chromosomes of
Drosophila melanogaster. EMBO J. 10, 153-162.
World Wide Web References

Http://www.hgmp.mrc.ac.uk/
The Human Genome Mapping Project home page.

http://menu.hgmp.mrc.ac.uk/Nix/Help/Mainhelp.html
The Human Genome Mapping Project Web page for the NIX (Nucleotide Identity X) multiple database search tool.

The National Centre for Biotechnology Information page containing the Basic Local Alignment Search Tool (BLAST) program for searching nucleic acid and protein databases.

The National Centre for Biotechnology Information page containing the Basic Local Alignment Search Tool (BLAST) program allowing two sequences (DNA or Protein) to be compared against each other.

The National Centre for Biotechnology Information page allowing access to the UniGene database. UniGene is an experimental system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters.

The National Centre for Biotechnology Information page allowing access to the ORF (Open Reading Frame) Finder Program, a graphical-analysis tool which allows the user to find all open reading frames in a given sequence.

http://www.tigr.org/
The Institute for Genomic Research home page.

http://www.tigr.org/tdb/hgi/hgi.html
The Institute for Genomic Research Human Gene Index page.

http://www.tigr.org/tdb/hgi/hgi_info.html
The Institute for Genomic Research Human Gene Index Definitions and Protocols page.

The IMAGE clone database home page.
http://www.fccc.edu/research/labs/golemis/intro.html
Useful page about common false positives generated through the Yeast Two hybrid system.

http://www.atcc.org/
The American Type Culture Collection home page.

http://www.expasy.ch/
The ExPASy molecular biology WWW server of the Swiss Institute of Bioinformatics (SIB), providing several programs and links relating to the analysis of protein sequences and structures.

Compute pI/Mw is a tool which allows the computation of the theoretical pI (isoelectric point) and Mw (molecular weight) for a list of SWISS-PROT and/or TrEMBL entries or for a user entered sequence.

http://www.sacs.ucsf.edu/home/HerskowitzLab/protocols/
Useful page containing several yeast-based protocols.

http://www.sanger.ac.uk/Software/Pfam/
Pfam is a high-quality comprehensive collection of protein domain families.

http://www.toulouse.inra.fr/prodom.html
ProDom is a comprehensive collection of protein families. The novelty of ProDom is that the modular arrangement of proteins have been taken into account and whenever domain boundaries were detected the sequences were cut to produce consistent families of domains.

http://blocks.fhcrc.org/blocks/blocks_search.html
Blocks compares a protein or DNA sequence to a database of protein blocks (blocks are short multiply aligned sequences corresponding to the most highly conserved regions of proteins). The rationale behind searching a database of blocks is that information from multiply aligned sequences is present in a concatenated form, reducing background and increasing sensitivity to distant relationships.

http://www.motif.genome.ad.jp/
Allows the searching Protein and Nucleic Acid Sequence Motifs, including Pfam, ProDom and Blocks, as shown above.

http://psort.nibb.ac.jp:8800/
PSORT is a useful computer program for the prediction of protein localisation sites in cells.
http://www.isrec.isb-sib.ch/software/SAPS_form.html
SAPS - Statistical Analysis of Protein Sequences analyses proteins for statistically significant features like charge-clusters, repeats, hydrophobic regions, compositional domains etc.

http://www.icnet.uk/LRITu/projects/pest/index.html
The PEST search utility identifies possible PEST regions in a submitted probe to identify proteins with short intracellular half-lives.
Appendix 1
Subcloning strategy for BMI-1 constructs

'Full-length BMI-1 clones'

- pGEX-4T1 and pET28a

'Truncated BMI-1 clones'

- pAS2-1, pGEX-4T1 and pET28a

PCR using BMI-1 specific primers with Sall linkers

EcoRI fragment (nucleotides 885-3203)

Full-length BMI-1 (K562) cDNA in pSP72

BMI-1 specific primers for PCR
Appendix 2
Subcloning strategy for pACT (B-cell library)-derived constructs

Inserts in Xhol site of pACT

- GAL4 AD
- BgII
- EcoRI
- BamHI
- Xhol
- Xhol
- BgII

PCR inserts using pACT-specific primers and clone products into pCR-Script

- EcoRI Fragment

pGEX-3X, pCMV-Tag1 and pcDNA3
Appendix 3
The principle vectors used in this study

i) pAS2-1
ii) pACT

MCS sequence for pACT:

MATCHMAKER 5' Insert-screening Amplimer → GAL4 Activation Domain

AAT ACC ACT ACA ATG GAT GAT GTA TAT AAC TAT CTA TTC GAT GAT GAA GAT ACC CCA

CCA AAC CCA AAA AAA GAG ATC TGG AAT TCG GAT CCT CGA GAG ATC TAT GAA TCG

TAG ATA CTG AAA AAC CCC GCA AGT TCA CTT CAA CTG TGC ATC GTG CAC CAT CTC

STOP (ORF 1) STOP (ORF 2) MATCHMAKER 3' Insert-screening Amplimer

GAL4 ad = GAL4 activation domain sequence
P = promoter
T = transcription termination sequence
A = SV40 T-antigen nuclear localization signal

Unique sites are in bold
* cDNAs are inserted at this site.
† The Bgl II sites may be used as a unique site.
iii) pGEX

Thrombin

CTG GTT CCG CGT GGA TCC

CTCG AAA TCC CGG GTG CCA CTC CAG CGG CCG CAT GAC TGA

BamHI EcoRI Smal Sal I XhoI Not I Stop codons

Factor Xa

ATC GAA GGT COT GGG ATC COC GGG AA TCA TOG TOA CTC ACT QAC

BamHI Smal EcoRI

Stop codons

pGEX-4T-1

pGEX-3X

pGEX

≈700 bp
iv) pET28a

The diagram illustrates the pET-28a (+) cloning/expression region with various restriction sites and gene sequences. The text mentions the use of primers for upstream and downstream regions, as well as the T7 promoter and terminator sequences.
v) pCMV-Tag1

The diagram shows the structure of the pCMV-Tag1 mammalian expression vector. The vector contains several elements:

- **MCS1**: Multicloning Site 1
- **Kozak**: Kozak sequence
- **FLAG**: FLAG tag
- **Stop Codon**: Stop codon
- **MCS2**: Multicloning Site 2
- **c-myc**: c-myc tag

The vector is 4309 bp in length. The diagram also includes restriction enzyme recognition sites, such as Bgl II, EcoRI, HindIII, Sal I, Sph I, and Not I, which are crucial for cloning and inserting foreign DNA into the vector.
vi) pcDNA3

There is an ATG upstream of the XbaI site.
Appendix 4
Selected BLAST (Basic Local Alignment and Search Tool) search results of BMI-1 potential interactors identified through yeast two-hybrid analysis

i) BMI-1 POSITIVE 16 (B-CELL LIBRARY) BLAST SEARCH RESULTS vs. GENBANK

Query= bmil6.abi, 526 bases, 9395642A checksum.
(526 letters)
Database: genbank
1,249,146 sequences; 821,455,368 total letters.
Searching..............................................done

Smallest Sum
Sequences producing High-scoring Segment Pairs:
<table>
<thead>
<tr>
<th>High Score</th>
<th>Probability P(N)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB:AA160555 AA160555 zo77d12.r1 Stratagene pancreas (#937...</td>
<td>614</td>
<td>1.0e-159</td>
</tr>
<tr>
<td>GB:H08033 H08033 yl91f01.r1 Homo sapiens cDNA clone 454...</td>
<td>1252</td>
<td>3.8e-149</td>
</tr>
<tr>
<td>GB:AA075682 AA075682 zm75c11.r1 Stratagene neuroepitheliu...</td>
<td>1619</td>
<td>3.8e-145</td>
</tr>
<tr>
<td>GB:HS108K1I Z595896 Human DNA sequence *** SEQUENCING IN P...</td>
<td>1017</td>
<td>1.0e-141</td>
</tr>
<tr>
<td>GB:H17675 H17675 ym35b07.r1 Homo sapiens cDNA clone 501...</td>
<td>1754</td>
<td>1.3e-140</td>
</tr>
<tr>
<td>GB:T65520 T65520 yc73g11.r1 Homo sapiens cDNA clone 216...</td>
<td>1348</td>
<td>6.5e-139</td>
</tr>
<tr>
<td>GB:C06342 C06342 similar to splicing factor SRp20.</td>
<td>1712</td>
<td>5.4e-137</td>
</tr>
<tr>
<td>GB:HUMSRP20 L10838 Homo sapiens SR protein family, pre-mR...</td>
<td>1692</td>
<td>2.0e-135</td>
</tr>
<tr>
<td>GB:W03181 W03181 za54c07.r1 Soares fetal liver spleen 1...</td>
<td>1408</td>
<td>6.5e-131</td>
</tr>
<tr>
<td>GB:AA206244 AA206244 q54d08.r1 Stratagene neuroepitheliu...</td>
<td>987</td>
<td>2.0e-127</td>
</tr>
<tr>
<td>GB:AA073385 AA073385 mm93c11.r1 Stratagene mouse heart (#...</td>
<td>1522</td>
<td>2.4e-121</td>
</tr>
<tr>
<td>GB:AA073385 AA073385 mm93c11.r1 Stratagene mouse heart (#...</td>
<td>1522</td>
<td>2.4e-121</td>
</tr>
</tbody>
</table>

WARNING: Descriptions of 68 database sequences were not reported due to the limiting value of parameter V = 20.

>GB:AA160555 AA160555 zo77d12.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone 592919 5' similar to gb:L10838 PRE-MRNA SPlicing FACTOR SRP20 (HUMAN);
Length = 588

Plus Strand HSPs:

Score = 614 (169.7 bits), Expect = 1.0e-159, Sum P(5) = 1.0e-159
Identities = 156/210 (74%), Positives = 156/210 (74%), Strand = Plus / Plus

Query: 300 C C GT GT AAGAGT GGAAC T GT C GAAT GGT GAAAAAAGAAGT AGAAAT C GT GGC C C AC C T C C
Sbjct: 319 C C GT GT AAGAGT GGAAC T GT C GAAT GGT GAAAAAAGAAGT AGAAAT C GT GGC C C AC C T C C

Score = 561 (155.0 bits), Expect = 1.0e-159, Sum P(5) = 1.0e-159
Identities = 119/128 (92%), Positives = 119/128 (92%), Strand = Plus / Plus

Query: 101 T GGAC T GT AAGGT T TAT GT AGGC AAT C T T GGAAAC AAT GGC AAC AAGAC GGAAT T GGAAC
Sbjct: 116 T T GGC T GT AAGGT T TAT GT AGGC AAT C T T GGAAAC AAT GGC AAC AAGAC GGAAT T GGAAC

212
Query: 161 GGGCTTTTGCTACTATGGACCACTCCGAAGTGCTGTTGGTGTAGCTANAAACCCAACCCGGCT 220
Sbjct: 176 GGGCTTTTGCTACTATGGACCACTCCGAAGTGCTGTTGGTGTAGAAACCCAACCCGGC 235

Query: 221 TTGCTTTTT 228
Sbjct: 236 TTTGCTTTT 243
Score = 423 (116.9 bits), Expect = 1.0e-159, Sum P(5) = 1.0e-159
Identities = 87/90 (96%), Positives = 87/90 (96%), Strand = Plus / Plus

Query: 14 AAGGCGGAAGCAGCTCATCCGGCCGACGCTAGCTAGTGGATTTTAGCCGCCGCGAGACTTTTTTTTACCGCTAGATC 73
Sbjct: 30 AAGGCGGAAGCAGCTCATCCGGCCGACGCTAGCTAGTGGATTTTAGCCGCCGCGAGACTTTTTTTTACCGCTAGATC 89
Score = 269 (74.3 bits), Expect = 1.0e-159, Sum P(5) = 1.0e-159
Identities = 57/62 (91%), Positives = 57/62 (91%), Strand = Plus / Plus

Query: 207 AAACCCACCCCGGCTTTTGCTTTTTTGAATTGGAANATCCCAGATGCAGCTAGCTAGTGGATTTTAGCCGCCGCGAGACTTTTTTTTACCGCTAGATC 266
Sbjct: 223 AAACCCACCCCGGCTTTTGCTTTTTTGAATTGGAANATCCCAGATGCAGCTAGCTAGTGGATTTTAGCCGCCGCGAGACTTTTTTTTACCGCTAGATC 282

Query: 267 CC 268
Sbjct: 283 CC 284
Score = 153 (42.3 bits), Expect = 1.0e-159, Sum P(5) = 1.0e-159
Identities = 31/32 (96%), Positives = 31/32 (96%), Strand = Plus / Plus

Query: 267 CCGAGAGCTANATGGAAAGAACACTATGGCTGCT 298
Sbjct: 284 CCGAGAGCTAGATGGAAAGAACACTATGGCTGCT 315
ii) BMI-1 POSITIVE 25 (B-CELL LIBRARY) BLAST SEARCH RESULTS vs. GENBANK

Query = bmi25.abi, 525 bases, F365FD07 checksum.
(525 letters)
Database: genbank
1,249,146 sequences; 821,455,368 total letters.
Searching......................................................done

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Score</th>
<th>P(N)</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB:HSBA7C081 Z28858 H. sapiens partial cDNA sequence; clone A7C08; version 1; strand(-), single read.</td>
<td>1383</td>
<td>1.8e-109</td>
<td>1</td>
</tr>
<tr>
<td>GB:N34591 N34591 xy61d09.r1 Homo sapiens cDNA clone 2...</td>
<td>1332</td>
<td>2.3e-105</td>
<td>1</td>
</tr>
<tr>
<td>GB:N26710 N26710 xy61d09.s1 Homo sapiens cDNA clone 2...</td>
<td>781</td>
<td>2.8e-68</td>
<td>3</td>
</tr>
<tr>
<td>GB:AA105138 AA105138 mp4f10.r1 Barstead MPLRB1 Mus mus...</td>
<td>842</td>
<td>9.8e-63</td>
<td>1</td>
</tr>
<tr>
<td>GB:AA105138 AA105138 mp4f10.r1 Barstead MPLRB1 Mus mus...</td>
<td>842</td>
<td>9.8e-63</td>
<td>1</td>
</tr>
<tr>
<td>GB:AA182774 AA182774 zp57c10.s1 Stratagene endothelial ...</td>
<td>354</td>
<td>1.8e-53</td>
<td>4</td>
</tr>
<tr>
<td>GB:AA046335 AA046335 zk77b01.r1 Soares pregnant uterus</td>
<td>620</td>
<td>3.0e-42</td>
<td>1</td>
</tr>
<tr>
<td>GB:HSBA7C082 Z25235 H. sapiens partial cDNA sequence; cl...</td>
<td>387</td>
<td>3.9e-22</td>
<td>1</td>
</tr>
<tr>
<td>GB:N67557 N67557 yz35g07.s1 Homo sapiens cDNA clone 2...</td>
<td>142</td>
<td>2.2e-07</td>
<td>3</td>
</tr>
<tr>
<td>GB:AA155454 AA155454 mn5hh12.r1 Beddington mouse embryo...</td>
<td>147</td>
<td>0.12</td>
<td>1</td>
</tr>
<tr>
<td>GB:HSBB1CIP4A L14320 Bovine herpesvirus type 1 early-inte...</td>
<td>146</td>
<td>0.16</td>
<td>1</td>
</tr>
<tr>
<td>GB:AA202960 AA202960 LD03373.5prime LD Drosophila Embryo...</td>
<td>142</td>
<td>0.26</td>
<td>1</td>
</tr>
<tr>
<td>GB:HEM19ANON1 L47334 Human DNA from cosmid f23280 from ch...</td>
<td>143</td>
<td>0.27</td>
<td>1</td>
</tr>
</tbody>
</table>

WARNING: Descriptions of 85 database sequences were not reported due to the limiting value of parameter V = 20.

>GB:HSBA7C081 Z28858 H. sapiens partial cDNA sequence; clone A7C08; version 1; strand(-), single read.
Length = 286

Plus Strand HSPs:

Score = 1383 (382.1 bits), Expect = 1.8e-109, P = 1.8e-109
Identities = 279/285 (97%), Positives = 279/285 (97%), Strand = Plus / Plus

Query: 90 GTTGTTCTTGTGGATATTATGAGAAATTATTGATTCAGACTTCTCTCAAAATGTGAATAA 149

Sbjct: 2 GTTGTTCTTGTGGATATTATGAGAAATTATTGATTCAGACTTCTCTCAAAATGTGAATAA 61

Query: 150 TAATGCAAGGTTTGGGCGATTGAGCATGAGAACCCATCTCTGCAAGATGAGCTGTGT 209

Sbjct: 62 TAATGCAAGGTTTGGGCGATTGAGCATGAGAACCCATCTCTGCAAGATGAGCTGTGT 121

Query: 210 CTTTCGCTGGGGAGTNTGGAAGACACTCTAGGGACCTCTGTGTATATTTGGAANAAATGTTGA 269

Sbjct: 122 CTTTCGCTGGGGAGTNTGGAAGACACTCTAGGGACCTCTGTGTATATTTGGAANAAATGTTGA 181

Query: 270 ACATCGCTGATACAGAAGGCAATTAAATAAAAAACTGTGCTAAAAATAATGACGATACATGAA 329

Sbjct: 182 ACATCGCTGATACAGAAGGCAATTAAATAAAAAACTGTGCTAAAAATAATGACGATACATGAA 241

Query: 330 NAAGCTCAGCATGACAAAGAAGAATCTCTCTGAGAACAGAAAANAVAGAAGGAGG 374

Sbjct: 242 GAAGCTCAGCATGACAAAGAAGAATCTCTCTGAGAACAGAAAANAVAGAAGGAGG 286
iii) BMI-1 POSITIVE 54 (B-CELL LIBRARY) BLAST SEARCH RESULTS vs. GENBANK

Query = bmi54.abi, 532 bases, 88B50DCC checksum.
(532 letters)
Database: genbank
1,240,544 sequences; 816,476,863 total letters.
Searching ......................................................done

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Score</th>
<th>Probability</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB:HUMHSMCM6 D84557 Human mRNA for HsMcm6, complete cds.</td>
<td>1484</td>
<td>9.1e-177</td>
<td>3</td>
</tr>
<tr>
<td>GB:HSU46838 U46838 Human p105MCM mRNA, complete cds.</td>
<td>1475</td>
<td>5.0e-176</td>
<td>3</td>
</tr>
<tr>
<td>GB:W58160 W58160 zdl8h05.r1 Soares fetal heart NbHH19...</td>
<td>1866</td>
<td>7.7e-150</td>
<td>1</td>
</tr>
<tr>
<td>GB:N40913 N40913 yw63h07.r1 Homo sapiens cDNA clone 2...</td>
<td>1367</td>
<td>3.4e-128</td>
<td>2</td>
</tr>
<tr>
<td>GB:H67765 H67765 yu54d11.r1 Homo sapiens cDNA clone 2...</td>
<td>1435</td>
<td>5.6e-114</td>
<td>1</td>
</tr>
<tr>
<td>GB:HSMCM6S4 U67284 Human mis5 homolog (MCM6) gene, exon...</td>
<td>806</td>
<td>2.5e-103</td>
<td>5</td>
</tr>
<tr>
<td>GB:W57982 W57982 zdl8h05.s1 Soares fetal heart NbHH19...</td>
<td>617</td>
<td>6.9e-98</td>
<td>4</td>
</tr>
<tr>
<td>GB:D86726 D86726 Mouse mRNA for mMIS5, complete cds.</td>
<td>1105</td>
<td>5.5e-84</td>
<td>1</td>
</tr>
<tr>
<td>GB:RNU17565 RNU17565 Rattus norvegicus intestinal DNA rep...</td>
<td>1075</td>
<td>1.5e-81</td>
<td>1</td>
</tr>
<tr>
<td>GB:AA132164 AA132164 zl38c08.r1 Soares pregnant uterus ...</td>
<td>1026</td>
<td>1.5e-79</td>
<td>1</td>
</tr>
<tr>
<td>GB:AA132079 AA132079 zl38c08.s1 Soares pregnant uterus ...</td>
<td>920</td>
<td>7.9e-70</td>
<td>1</td>
</tr>
<tr>
<td>GB:AA218363 AA218363 mv55f01.r1 Soares mouse 3NME12 5 M...</td>
<td>2374</td>
<td>7.9e-68</td>
<td>2</td>
</tr>
</tbody>
</table>

WARNING: Descriptions of 25 database sequences were not reported due to the limiting value of parameter V = 20.

>GB:HUMHSMCM6 D84557 Human mRNA for HsMcm6, complete cds.
Length = 2927

Plus Strand HSPs:

Score = 1484 (410.1 bits), Expect = 9.1e-177, Sum P(3) = 9.1e-177
Identities = 300/306 (98%), Positives = 300/306 (98%), Strand = Plus / Plus

Query: 18 C C GAAAGGT GGAAGAAGAAGAGGAC GAGT CAGCAT TAAAGAGGAGC GAGC T T GT TAAC T G
Sbjct: 2254 C AGAAAGGT GGAAGAAGAAGAGGAC GAGT C AGC AT T AAAGAGGAGC GAGC T T GT T AAC T G

Query: 78 GT AC T T GAAGGAAAT C GAAT C AGAGAT AGAC T C T GAAGAAGAAC T TAT AAAT NAAAAAAG
Sbjct: 2314 GT AC T T GAAGGAAAT C GAAT C AGAGAT AGAC T C T GAAGAAGAAC T TAT AAAT AAAAAAAG

Query: 138 AAT CAT AGAGAAAGT TAT T CAT C GAC T C AC AC AC TAT GAT CAT GT T C T AAT T GAGC T C AC
Sbjct: 2374 AAT CAT AGAGAAAGT TAT T CAT C GAC T C AC AC AC TAT GAT CAT GT T C T AAT T GAGC T C AC

Query: 198 C C AGGC T GGAT T GAAAGGC T C C AC AGAGGGAAGT GAGAGC TAT GAAGAAGAT C C C T AC T T
Sbjct: 2434 C C AGGC T GGAT T GAAAGGC T C C AC AGAGGGAAGT GAGAGC TAT GAAGAAGAT C C C T AC T T

Query: 258 GGT AGT T AAC C C T AAC T AC T T GC T C GAAGAT T GAGAT AGT GAAAGT AAC T GAC C AGAGC T
Sbjct: 2554 GAGGGA 323

Score = 624 (172.4 bits), Expect = 9.1e-177, Sum P(3) = 9.1e-177
Identities = 128/135 (94%), Positives = 128/135 (94%), Strand = Plus / Plus

Query: 321 GGAACGTGGCCACAGCACCTCGTGGGCCTGGAGCCTGCTGCTCTANGGGACAGAA 380
Sbjct: 2556 GGAACGTGGCCACAGCACCTCGTGGGCCTGGAGCCTGCTGCTCTANGGGACAGAA 2615
Query: 381 NTGTTTCTGGAAAGTATGCTTCCANGATTTTTCANAACAANAAATTGATGATGTG 440
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2616 GTGTTTCTGGAAAGTATGCTTCCANGATTTTTCAGAAAAACAAATTGATGATGTG 2675

Query: 441 CCTATGTGTCNCATT 455
        ||||||||
Sbjct: 2676 CCTATGTGTCACATT 2690

Score = 220 (60.8 bits), Expect = 9.1e-177, Sum P(3) = 9.1e-177
Identities = 48/55 (87%), Positives = 48/55 (87%), Strand = Plus / Plus

Query: 455 TTCATCNACAGTTTTCATNCACAACACAGGCTTCCACTCTCTTTGTGTGTNTTTC 509
        |||||      |||          |                          |||||         |      ||
Sbjct: 2689 TTCATCACAGTTTTCATACCAAACACACAGGCTTCCACTCTCTTTGTGTGTNTTTC 2743
### iv) BMI-1 POSITIVE v15 (9.5/10.5dpc MOUSE EMBRYO LIBRARY) BLAST SEARCH RESULTS vs. GENBANK

**Query**: v15, 120 bases, ABA082F5 checksum.

(Database: genbank)

- Sequences producing High-scoring Segment Pairs:
  - **GB:AA155273** AA155273 mr98d07.r1 Stratagene mouse embryoni... 600 2.4e-42 1
  - **GB:AA155273** AA155273 mr98d07.r1 Stratagene mouse embryoni... 600 2.4e-42 1
  - **GB:RATMPP** M57728 Rat general mitochondrial matrix proce... 564 6.4e-39 1
  - **GB:R21055** R21055 yg52b05.r1 Homo sapiens cDNA clone 361... 528 2.5e-36 1
  - **GB:D50913** D50913 Human mRNA for KIAA0123 gene, partial ... 528 6.6e-36 1
  - **GB:HUMORFB1** D21064 Human mRNA for KIAA0123 gene, partial ... 528 6.6e-36 1
  - **GB:W37985** W37985 zcl31all.r1 Soares parathyroid tumor Nb... 116 0.9993 1

**>GB:AA155273** AA155273 mr98d07.r1 Stratagene mouse embryonic carcinoma (#937317)

Mus musculus cDNA clone 605485 5' similar to SW:MPP1_RAT P20069
MITOCHONDRIAL PROCESSING PEPTIDASE ALPHA SUBUNIT PRECURSOR ;
Length = 466

**Plus Strand HSPs:**

Score = 600 (165.8 bits), Expect = 2.4e-42, P = 2.4e-42

Identities = 120/120 (100%), Positives = 120/120 (100%), Strand = Plus / Plus

Query: 

```
T T GAC AGC AAAGAC GAAAT C C T GC T T AC GC T GGAAAAAC AT GGT GGT AT C T GT GAC T GC C
```

Sbjct: 

```
315 T T GAC AGC AAAGAC GAAAT C C T GC T T AC GC T GGAAAAAC AT GGT GGT AT C T GT GAC T GC C
```

```
61 AGAC C T C AAGAGAC AC C AC CAT GT AT GC T GT GT C T GC T GAC AGC AAAGGC T T GGAC AC T G
```

Sbjct: 

```
375 AGAC C T C AAGAGAC AC C AC CAT GT AT GCTGTGTCTGCT GAC AGC AAAGGC T T GGAC AC T G 434
```

**>GB:AA155273** AA155273 mr98d07.r1 Stratagene mouse embryonic carcinoma (#937317)

Mus musculus cDNA clone 605485 5' similar to SW:MPP1_RAT P20069
MITOCHONDRIAL PROCESSING PEPTIDASE ALPHA SUBUNIT PRECURSOR ;
Length = 466

**Plus Strand HSPs:**

Score = 600 (165.8 bits), Expect = 2.4e-42, P = 2.4e-42

Identities = 120/120 (100%), Positives = 120/120 (100%), Strand = Plus / Plus

Query: 

```
1 TTGACAGCAAAGACAAATCCTGCTTGAGAAAAACATGTGTATCTGTGACTGCC
```

Sbjct: 

```
315 TTGACAGCAAAGACAAATCCTGCTTGAGAAAAACATGTGTATCTGTGACTGCC
```

```
61 AGACCTCAAAGAGACACCACTGATGTGCTGCTGCTGACAGCAAAGGCCTTGGACACTG
```

Sbjct: 

```
375 AGACCTCAAAGAGACACCACTGATGTGCTGCTGCTGACAGCAAAGGCCTTGGACACTG
```

217
v) BMI-1 POSITIVE v36 (9.5/10.5dpc MOUSE EMBRYO LIBRARY) BLAST
SEARCH RESULTS vs. GENBANK

Query = v36, 76 bases, 33F66104 checksum. (76 letters)

Database: genbank
1,352,761 sequences; 920,511,008 total letters.
Searching.................................done

Sequences producing High-scoring Segment Pairs:

<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
<th>Score</th>
<th>Expect</th>
<th>P (N)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB:MMHSF2</td>
<td>X61754 M. musculus mRNA for heat shock transcripton factor 2.</td>
<td>380</td>
<td>1.4e-23</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>GB:HUMHSF2</td>
<td>M65217 Human heat shock factor 2 (HSF2) mRNA, ...</td>
<td>272</td>
<td>1.4e-14</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>GB:R21303</td>
<td>R21303 yg48d10.rl Homo sapiens cDNA clone 3590...</td>
<td>230</td>
<td>3.2e-11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>GB:HS18</td>
<td>282209 Human DNA sequence from clone J428A131.</td>
<td>114</td>
<td>0.14</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>GB:R14019</td>
<td>R14019 yf62a12.rl Homo sapiens cDNA clone 2659...</td>
<td>120</td>
<td>0.49</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>GB:M85446</td>
<td>M85446 EST01962 Homo sapiens cDNA clone HFBCI31.</td>
<td>109</td>
<td>0.9996</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

>GB:MMHSF2 X61754 M. musculus mRNA for heat shock transcription factor 2.
Length = 1972

Plus Strand HSPs:

Score = 380 (105.0 bits), Expect = 1.4e-23, P = 1.4e-23
Identities = 76/76 (100%), Positives = 76/76 (100%), Strand = Plus / Plus

Query: 1 ATAAGGGACTAGAAGCTCAACAGGACAGCTGTTGTTCAACATGTGTAGAAGAGGAAGAAAATCTAAAT C C A 72
Sbjct: 1339 ATAAGGGACTAGAAGCTCAACAGGACAGCTGTTGTTCAACATGTGTAGAAGAGGAAGAAAATCTAAAT C C A 233

>GB:HUMHSF2 M65217 Human heat shock factor 2 (HSF2) mRNA, complete cds.
Length = 2411

Plus Strand HSPs:

Score = 272 (75.2 bits), Expect = 1.4e-14, P = 1.4e-14
Identities = 64/76 (84%), Positives = 64/76 (84%), Strand = Plus / Plus

Query: 1 ATAAGGGACTAGAAGCTCAACAGGACAGCTGTTGTTCAACATGTGTAGAAGAGGAAGAAAATCTAAAT C C A 72
Sbjct: 1326 ATAAGGGACTAGAAGCTCAACAGGACAGCTGTTGTTCAACATGTGTAGAAGAGGAAGAAAATCTAAAT C C A 233

>GB:R21303 R21303 yg48d10.rl Homo sapiens cDNA clone 35903 5' similar to gb:M65217 HEAT SHOCK FACTOR PROTEIN 2 (HUMAN);.
Length = 388

Plus Strand HSPs:

Score = 230 (63.6 bits), Expect = 3.2e-11, P = 3.2e-11
Identities = 54/64 (84%), Positives = 54/64 (84%), Strand = Plus / Plus

Query: 13 AAGCTCAACAGGACAGCTGTTGTTCAACATGTGTAGAAGAGGAAGAAAATCTAAATCCCA 72
Sbjct: 174 AAGCTCAACAGGACAGCTGTTGTTCAACATGTGTAGAAGAGGAAGAAAATCTAAATCCCA 233

Query: 73 AGCC 76
Sbjct: 234 AACC 237
vi) BMI-1 POSITIVE v45 (9.5/10.5dpc MOUSE EMBRYO LIBRARY) BLAST SEARCH RESULTS vs. GENBANK

Query= V45F.seq, 526 bases, BA269BBE checksum.
(526 letters)

Searching......................................................done

Sequences producing High-scoring Segment Pairs:

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
<th>Score</th>
<th>Probability</th>
<th>Sum Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB:D86726</td>
<td>D86726 Mouse mRNA for mMIS5, complete cds.</td>
<td>1134</td>
<td>7.6e-97</td>
<td>2</td>
</tr>
<tr>
<td>GB:AA218363</td>
<td>AA218363 mv55f01.r1 Soares mouse 3NME12 5 M...</td>
<td>1057</td>
<td>2.3e-96</td>
<td>2</td>
</tr>
<tr>
<td>GB:AA048222</td>
<td>AA048222 mj26c03.r1 Soares mouse embryo NbM...</td>
<td>1044</td>
<td>2.6e-95</td>
<td>2</td>
</tr>
<tr>
<td>GB:W84969</td>
<td>W84969 mf42e03.r1 Soares mouse embryo NbM...</td>
<td>915</td>
<td>1.6e-84</td>
<td>2</td>
</tr>
<tr>
<td>GB:W14754</td>
<td>W14754 mb23f12.r1 Soares mouse p3NMF19.5 Mu...</td>
<td>941</td>
<td>1.1e-81</td>
<td>2</td>
</tr>
<tr>
<td>GB:RNU17565</td>
<td>U17565 Rattus norvegicus intestinal DNA rep...</td>
<td>973</td>
<td>4.9e-73</td>
<td>1</td>
</tr>
<tr>
<td>GB:AA209971</td>
<td>AA209971 mu38g08.o1 Soares NbMT Mus muscul...</td>
<td>788</td>
<td>1.9e-71</td>
<td>2</td>
</tr>
<tr>
<td>GB:H67765</td>
<td>H67765 yu54d11.r1 Homo sapiens cDNA clone 2...</td>
<td>857</td>
<td>1.2e-64</td>
<td>1</td>
</tr>
<tr>
<td>GB:HUMH5CM6</td>
<td>HUMH5CM6 Human mRNA for HsMcm6, partial cds.</td>
<td>857</td>
<td>4.4e-62</td>
<td>1</td>
</tr>
<tr>
<td>GB:HUMH5CM6</td>
<td>HUMH5CM6 Human mRNA for HsMcm6, complete cds.</td>
<td>857</td>
<td>4.4e-62</td>
<td>1</td>
</tr>
<tr>
<td>GB:W58160</td>
<td>W58160 zd18h05.r1 Soares fetal heart NbH19...</td>
<td>829</td>
<td>1.4e-61</td>
<td>1</td>
</tr>
<tr>
<td>GB:N40913</td>
<td>N40913 yw63h07.r1 Homo sapiens cDNA clone 2...</td>
<td>832</td>
<td>3.0e-61</td>
<td>1</td>
</tr>
<tr>
<td>GB:HSU46838</td>
<td>U46838 Human pl05MCM mRNA, complete cds.</td>
<td>848</td>
<td>3.2e-61</td>
<td>1</td>
</tr>
</tbody>
</table>

WARNING: Descriptions of 821 database sequences were not reported due to the limiting value of parameter V = 20.

>GB:D86726 D86726 Mouse mRNA for mMIS5, complete cds.
Length = 2901

Plus Strand HSPs:

Score = 1134 (313.3 bits), Expect = 7.6e-97, Sum P(2) = 7.6e-97
Identities = 244/275 (88%), Positives = 244/275 (88%), Strand = Plus / Plus

Query: 71 GTTGCGCTCAAGAGAAGCCGAGTTTGTCAACTGGTNCCTGNNGAAATCGAGTCAGANAT 130
|| || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || || | 219
vii) BMI-1 POSITIVE v67 (9.5/10.5dpc MOUSE EMBRYO LIBRARY) BLAST SEARCH RESULTS vs. GENBANK

Query= v67, 83 bases, 54793916 checksum.
(Database: genbank)
1,352,761 sequences; 920,511,008 total letters.

Sequences producing High-scoring Segment Pairs:

<table>
<thead>
<tr>
<th>Database</th>
<th>Sequence Description</th>
<th>Score</th>
<th>P(N)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB:HSBA7C081 Z28858</td>
<td>H. sapiens partial cDNA sequence; clone A7C08; version 1; strand(-), single read.</td>
<td>262</td>
<td>5.2e-14</td>
<td>1</td>
</tr>
<tr>
<td>GB:SYCSLRG D64005</td>
<td>Synechocystis sp. PCC6803 complete genome, 24/27, 3002966-3138603.</td>
<td>100</td>
<td>0.11</td>
<td>2</td>
</tr>
<tr>
<td>GB:CET13H5 Z66524</td>
<td>Caenorhabditis elegans cosmid T13H5.</td>
<td>112</td>
<td>0.999</td>
<td>1</td>
</tr>
</tbody>
</table>

>GB:HSBA7C081 Z28858 H. sapiens partial cDNA sequence; clone A7C08; version 1; strand(-), single read.
Length = 286

Plus Strand HSPs:
Score = 262 (72.4 bits), Expect = 5.2e-14, P = 5.2e-14
Identities = 58/65 (89%), Positives = 58/65 (89%), Strand = Plus / Plus

Query: 19 GT T T GT T C T AGT AGAAT T GT C AGGAAT CAT T GAT T C AGAC TTTCTCTC C AAAT GT GAAAA

Sbjct: 2 GTTGGTTCTGGTGGAATTATCAGGAATTATTGATTCAGACTTCCTCTCAAAATGTGAAAA

Score = 100 (27.6 bits), Expect = 0.12, Sum P(2) = 0.11
Identities = 24/29 (82%), Positives = 24/29 (82%), Strand = Minus / Plus

Query: 4 0 TGACAATTCCTACTAGAAACTGCTCCT 12

Sbjct: 87903 TGACAATTCCTACTAGAAACTGCTCCT 87931

Score = 80 (22.1 bits), Expect = 0.12, Sum P(2) = 0.11
Identities = 16/16 (100%), Positives = 16/16 (100%), Strand = Minus / Plus

Query: 55 TGAATCAATGATTCCT

Sbjct: 63456 TGAATCAATGATTCCT 63471
Appendix 5
Motifs present within the protein sequence encoded by clone 25

A - PEST Analysis
(Potential PEST sequences, as predicted by the PEST search tool, are shown in red)

Candidate number 1
Probe: (8-38) - RSPEDGEDEEEEQLVVELSGIIDSDFLSK
PEST Score: 8.69

Candidate number 2
Probe: (145-159) - HENEDEEVVASAPDK
PEST Score: 9.25

---

B - Motif Analysis
(The motifs, as predicted by the Motifs database search tool, are shown in red)

i) Motif Name: ASN_GLYCOSYLATION (N-glycosylation site)
Prosite Name: PS00001
Pattern: N-{P}-{ST}-{P}.
Appears: 89..92 NKTV, 171..174 NDSS, 175..178 NLSC
Sequence:

MAAAADERSPDGEDEEEEEQLVVELSGIIDSDFLSKCENKCKVGLIDTERPILQVDSC
VFAGEYEDTLGTCVIFEEVNEHADTEGNKTVLKYKCHTMKGLSMTRLLTEKKEGEENI
GGVEWQLIKDNDFSYRPNNICNFHLHENEDEEVVASAPDKSLEESEEIQMNDDSNLSCEQ
EKPMHLEIEDSGPLIDIPSETEGSVFMETQMLP

---
ii) Motif Name: TYR_PHOSPHO_SITE *(Tyrosine kinase phosphorylation site)*
Prosite Name: PS00007
Pattern: [RK]-x(2,3)-[DE]-x(2,3)-Y.
Appears: 129..135 KDNDFSY
Sequence:

```
MAAAAADERSPEDGEDEEEEQLVVELSGIIDSFLSKCENKCKVGLIDTERPILOVQVDSC
VFAGYEYETLTCVIFEENVEHADTEGNKTVLKYKCHTMKKSMTTRLLEKEEENI
GGVEWLQIKDNDFSYRPNMCNFHNEEVEVASSAPKSLEEEEIQMNDSLSECEQ
EKPMHLEIEDSPGIDLIDIPSETGVSFMETQMLP
```

iii) Motif Name: PKC_PHOSPHO_SITE *(Protein kinase C phosphorylation site)*
Prosite Name: PS00005
Pattern: [ST]-x-[RK].
Appears: 50..52 TER, 99..101 TMK, 111..113 TEK, 134..136 SYR
Sequence:

```
MAAAAADERSPEDGEDEEEEQLVVELSGIIDSFLSKCENKCKVGLIDTERPILOVQVDSC
VFAGYEYETLTCVIFEENVEHADTEGNKTVLKYKCHTMKKSMTTRLLEKEEENI
GGVEWLQIKDNDFSYRPNMCNFHNEEVEVASSAPKSLEEEEIQMNDSLSECEQ
EKPMHLEIEDSPGIDLIDIPSETGVSFMETQMLP
```