INVESTIGATIONS OF HISTONE H1 GENE-SPECIFIC PROMOTER ELEMENTS

Kym Nicholas Duncliffe, B.Sc. (Hons.)

A thesis submitted for the degree of

Doctor of Philosophy

in

The University of Adelaide

(Faculty of Science)

Department of Biochemistry,
University of Adelaide,
Adelaide, South Australia

June, 1994

Awarded 1995
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>vii</td>
</tr>
<tr>
<td>Statement</td>
<td>ix</td>
</tr>
<tr>
<td>Note</td>
<td>x</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xi</td>
</tr>
<tr>
<td>Dedication</td>
<td>xiii</td>
</tr>
</tbody>
</table>

## CHAPTER ONE: INTRODUCTION

1.1 General Introduction                           | 1    
1.2 The Regulation of Eukaryotic Class II Gene Expression | 1    
1.2.1 Class II Gene Structure                        | 2    
1.2.2 Transcriptional Regulation                     | 2    
1.2.2.1 TATA-box                                     | 2    
1.2.2.2 General Promoter Elements                    | 3    
1.2.2.3 Specific Promoter Elements                    | 4    
1.2.2.4 Enhancers                                    | 4    
1.2.2.5 Locus Control Regions                        | 5    
1.2.3 Post-Transcriptional Regulation                | 6    
1.2.3.1 RNA Splicing                                 | 6    
1.2.3.2 3' RNA Processing                            | 7    
1.2.3.3 RNA Stability                                | 7    
1.2.3.4 mRNA Transport                               | 7    
1.2.3.5 Translation                                  | 8    
1.3 Histones                                        | 9    
1.3.1 H1                                            | 9    
1.3.2 The Core Histones                              | 10   
1.4 Histones and Chromatin Structure                 | 10   
1.4.1 The Nucleosome                                 | 11   
1.4.1.1 Structure of the Nucleosome Core Particle    | 11   
1.4.1.2 Attachment of H1 to the Nucleosome           | 12   
1.4.1.3 Role of the core histone tails               | 12   
1.4.2 Higher Order Chromatin Structure               | 13   
1.4.3 Non histone chromosomal proteins               | 14   
1.5 Relationship between Histones, Chromatin Structure, and Transcription | 14   
1.5.1 Histone Subtypes and Variants                  | 15   
1.5.1.1 H1 variants                                  | 16   
1.5.1.2 Core histone variants                        | 18   
1.5.2 Histone Modifications                          | 19   
1.5.2.1 Acetylation                                  | 20   
1.5.2.2 Phosphorylation                              | 21   
1.6 Histone Gene Structure                           | 21   
1.6.1 Structure of the Major Histone Genes           | 23   
1.6.2 Structure of variant histone genes             | 24   
1.7 The Organisation of Histone Genes                | 25   
1.7.1 Plants                                        | 26   
1.7.2 Fungi                                         | 27   
1.7.3 Caenorhabditis elegans                         | 28   
1.7.4 Drosophila                                     | 29   
1.7.5 Sea Urchin                                     | 31   


1.7.6 Trout
1.7.7 Amphibia
1.7.8 Chicken
1.7.9 Mammals

1.8 Regulation of Histone Gene Expression: Lower Eukaryotes
1.9 Regulation of Histone Gene Expression: Higher Eukaryotes
1.9.1 Replication Independent Expression: Embryonic Uncoupling
1.9.2 Replication Independent Expression: Variant Histone Genes
1.9.3 Replication Dependent Expression: Posttranscriptional Regulation
1.9.4 Replication Dependent Expression: Transcriptional Regulation

1.10 The H1 Histone Genes of Chicken

CHAPTER TWO: MATERIALS AND METHODS

2.1 Abbreviations
2.2 Materials
2.2.1 Chemicals, Reagents and Kits
2.2.2 Enzymes
2.2.3 Isotopically Labelled Compounds
2.2.4 Bacterial Strains and Cell Lines
2.2.4 (a) E. coli Strains
2.2.4 (b) Tissue Culture Cell Lines
2.2.5 Media and Buffers
2.2.5 (a) Media
2.2.5 (b) Buffers
2.2.6 Clones and Vectors
2.2.6 (a) Cloned DNA Sequences
2.2.6 (b) Cloning Vectors
2.2.6 (c) Reporter Vectors
2.2.7 Molecular Weight Markers
2.2.7 (a) DNA Markers
2.2.7 (b) Protein Markers
2.2.8 Oligonucleotides
2.3 Methods
2.3.1 Isolation of Plasmid DNA
2.3.1 (a) General Procedure
2.3.1 (b) Miniscreen Procedure
2.3.1 (i) Alkaline extraction
2.3.1 (ii) STET preps
2.3.1 (c) Large Scale Purification on a CsCl Gradient
2.3.2 Restriction Enzyme Digestion and Analysis of DNA
2.3.2 (a) Restriction Enzyme Digestion
2.3.2 (b) Agarose Gel Electrophoresis
2.3.2 (c) Polyacrylamide Gel Electrophoresis
2.3.2 (d) Denaturing polyacrylamide gels
2.3.3 Subcloning of DNA fragments into plasmid vectors
2.3.3 (a) Isolation of DNA from agarose gels
2.3.3 (b) Isolation of DNA from low gelling temperature agarose gels
2.3.3 (c) Preparation of Vector
2.3.3 (d) Ligation into plasmid vectors
2.3.3 (e) Transformation of recombinants into bacteria
2.3.3 (f) Plasmid miniscreen procedure
2.3.4 Purification of oligonucleotides
2.3.5 Labelling of DNA fragments
  2.3.5 (a) Kinasing of synthetic oligonucleotides
  2.3.5 (b) Endfilling of restriction fragments
2.3.6 Manipulation and Harvesting of tissue culture cells
2.3.7 Nuclear extract preparation
  2.3.7 (a) Tissue culture cells
    (i) From Strauss and Varshavsky (1984)
    (ii) From Lee et al (1988)
    (iii) From Dignam et al (1983)
    (iv) From Dalton (1987)
  2.3.7 (b) Small scale nuclear extract preparation from chicken embryos
  2.3.7 (c) Large scale nuclear extract preparation from chicken embryos
  2.3.7 (d) Protein concentration determination
2.3.8 Detection of DNA binding factors in nuclear extracts
  2.3.8 (a) Probe preparation
  2.3.8 (b) Mobility shift assay
  2.3.8 (c) Polyacrylamide gel electrophoresis of mobility shift assays
2.3.9 Chromatographic enrichment of nuclear extract components
  2.3.9 (a) DEAE- / Heparin-Sepharose chromatography
  2.3.9 (b) Large scale concentration of partially purified extracts
  2.3.9 (c) Sephacryl S-300 gel filtration chromatography
  2.3.9 (d) DNA-affinity microprecipitation
    (i) Preparation of DNA Ligand
    (ii) DNA affinity microprecipitation assay
    (iii) DNA affinity column chromatography
2.3.10 SDS Polyacrylamide gel electrophoresis of proteins
  2.3.10 (a) Gel electrophoresis
  2.3.10 (b) Silver staining of protein gels
  2.3.10 (c) Coomassie staining of protein gels
2.3.11 Electroblotting of protein to PVDF membranes
2.3.12 Association of DNA-binding activity with a protein
  2.3.12 (a) South-Western transfer
  2.3.12 (b) Renaturation of binding activity from SDS polyacrylamide gels
    (i) Guanidine-HCl method
    (ii) Saturated urea method
2.3.13 Transient transfection of tissue culture cells
2.3.14 CAT assays
2.3.15 Double Stranded Dideoxy Sequencing of DNA
  2.3.15 (a) Preparation of Template
  2.3.15 (b) Hybridization
  2.3.15 (c) Extension and termination
2.3.16 Polymerase Chain Reaction (PCR)
2.3.17 Autoradiography
2.3.18 Laser Densitometry
2.3.19 Phosphorimaging
2.3.20 Containment facilities and animal ethics
CHAPTER THREE: TOWARD PURIFICATION OF AC-BOX-BINDING FACTORS IN CHICKEN

3.1 Introduction 91
3.2 Characterisation of AC-box Binding Activity 92
3.2.1 Source of AC-box Binding Factors 92
3.2.2 Choice and Preparation of Probe 92
3.2.3 Mobility Shift Assays of Chicken Embryo Extracts with the 01-40mer Probe 93
3.2.4 Analysis of AEV ts34 Nuclear Extracts with the 01-40mer Probe 93
3.2.5 Small Scale DEAE- and Heparin-Sepharose Chromatography of ts34 Crude Nuclear Extracts 94
3.2.6 Attempts to Optimise Recovery of 40-MS-b 94
3.2.7 Differences Between H1-SF and 40-MS-e 95
3.2.8 Mobility Shift Assays with the 014mer Probe 96
3.3 Large Scale Preparation and Chromatography of Chicken Embryo Extracts 98
3.3.1 Heparin-Sepharose Chromatography 98
3.3.2 Ultrafiltration of Heparin-Sepharose Extracts 99
3.4 DNA Affinity Chromatography of Partially Purified Extracts 99
3.4.1 Design of a Suitable Oligonucleotide 99
3.4.2 Preparation of the DNA Affinity Ligand 99
3.4.3 DNA Affinity Microprecipitation 100
3.4.4 Optimisation of Microprecipitation Method 100
3.4.5 Scale-up of Microprecipitation Method 102
3.5 Possible Co-purification of H1-Transcription Factors with 14-MS-b 103
3.6 SDS Polyacrylamide Gel Electrophoresis of DNA Affinity Purified Extracts 105
3.7 Direct Amino Acid Sequencing of a DNA Affinity-Purified Protein 105
3.8 2nd-Round DNA Affinity Microprecipitation 106
3.9 Alternative Purification Methods 107
3.9.1 Sephacryl S-300 Gel Filtration Chromatography 107
3.9.2 DNA Affinity Column Chromatography 108
3.10 Renaturation of 14-MS-b 109
3.10.1 Southwestern Analysis 109
3.10.2 Renaturation of Proteins from Gel Slices 109
3.10.3 Analysis of ts34 Extracts with the 01-14mer Probe 110
3.11 Identification and Renaturation of an 01-40mer Binding Factor in DNA Affinity-Purified Extracts 111
3.11.1 Characterisation of the 40-MS-I Band Shift 111
3.11.2 Southwestern Analysis with the 01-40mer Probe 111
3.11.3 Renaturation of 40-MS-I 112
3.12 Use of Intermediate Length Oligonucleotide Probes 113
3.12.1 Design of 25 bp Oligonucleotides 113
3.12.2 Mobility Shift Assays of Chicken Embryo Extracts with the 25mers 113
3.12.3 DNA Affinity Purification with 25mers 114
3.12.4 Characterisation of a Subtype-Specific AC-box-Binding Factor in ts34 Extracts 114
3.12.5 Mobility Shift Assays of Various Cell Lines with the 25mer Probes 115
3.13 Discussion 116

CHAPTER FOUR: FUNCTIONAL ANALYSES OF THE AC-BOX AND A RELATED H1-SPECIFIC PROMOTER ELEMENT

4.1 Introduction 122
4.2 Effect of the AC-box in a non-S-phase regulated promoter 123
4.2.1 Choice of a suitable reporter system 123
4.2.2 Construction of AC-box/HSVtk promoter fusions in pTKGH 124
4.2.3 Analysis of AC-box/HSVtk promoter activity in p(AC)TKGH series 125
4.2.4 Construction of the p(AC)TKCAT series 126
4.2.5 Analysis of AC-box/HSVtk promoter activity in p(AC)TKCAT.N series 127
4.3 Investigation of a Distal Element in the H1 Histone Promoter
4.3.1 Comparison of H1 Histone Promoter Sequences
4.3.2 Sequencing of the Chicken 02 H1 Histone Promoter
4.3.3 Construction of H1 Promoter / CAT Fusions
  4.3.3 (a) Generation of p(02)TKCAT vectors
  4.3.3 (b) Generation of p(02)HCAT vectors
4.3.4 Effect of the TG-box on H1 Promoter Activity
4.3.5 Effect of an 02 H1 Promoter Fragment on Expression of the HSVtk Promoter
4.4 Discussion

CHAPTER FIVE: FINAL DISCUSSION

BIBLIOGRAPHY
The work presented in this thesis reports investigations of the histone H1 gene-specific promoter element, 5' AAACACA 3', also referred to as the "H1-box" or "AC-box" (Coles and Wells, 1985). The first approach was to isolate and purify a protein factor, H1-SF, from 9 day chicken embryos. This protein had been previously identified in nuclear extracts from AEV ts34 chicken erythroblasts, and had been demonstrated to interact specifically with the AC-box (Dalton and Wells, 1988a).

In mobility-shift assays using a 40 bp oligonucleotide probe (containing the AC-box) derived from the promoter of the chicken histone H1 gene, 01 H1, a candidate AC-box-binding factor, 40-MS-e, was identified in crude nuclear extracts from ts34 cells. Further characterisation revealed anomalies between some properties of this factor and those of H1-SF. In addition, a specific AC-box-binding factor could not be identified in crude chicken embryo extracts. However, a second AC-box-binding factor, 14-MS-b, was identified in chicken embryo extracts using a 14 bp oligonucleotide probe containing only the AC-box.

14-MS-b was purified from chicken embryo extracts on a large scale, using Heparin Sepharose-, Sephacryl S-300-, and finally DNA-affinity chromatography using concatemers of the 14-mer as the ligand. Analysis of DNA affinity-purified extracts on SDS-polyacrylamide gels revealed two enriched proteins of ~86 kD and ~70 kD. To determine whether either of these proteins was 14-MS-b, the MW of 14-MS-b was estimated by renaturation from SDS-polyacrylamide gels. Surprisingly, the renatured 14-MS-b was no greater than 30 kD: much smaller than either of the enriched proteins, or the estimated mass of H1-SF (90 kD - S. Dalton, unpublished results). Thus it was unlikely that 14-MS-b was analogous to H1-SF.

Another factor, 40-MS-1, was identified in mobility shift assays of DNA affinity-purified extracts using the 40 bp oligonucleotide probe. This factor was also renatured from SDS-polyacrylamide gels, but it too was much smaller (42-46 kD) than the estimated molecular mass of H1-SF. Further investigation of the 86 kD and 70 kD proteins indicated that they were non specific DNA binding proteins.
A final series of mobility-shift assays was conducted using four 25 bp oligonucleotides carrying either wild type or mutated AC-boxes derived from the promoters of the chicken histone 01 H1 and 02 H1 genes. In crude nuclear extracts from ts34 cells, a factor (25-MS-1) which specifically bound the wild type 02 H1 oligonucleotide was identified. 25-MS-1 was also detected in extracts from three other chicken cell lines. However, this factor was not detected in 9 day chicken embryo extracts. In addition, no AC-box-specific binding activity analogous to 25-MS-1 was detected with the wild type 01 H1 oligonucleotide.

It had been shown that deletion or mutation of the AC-box both greatly decreased the steady state levels of H1-mRNA and eliminated cell cycle regulation of H1 gene transcription in stably transfected HeLa cell lines (Dalton and Wells, 1988a). To examine the effect of the AC-box on a heterologous promoter, a series of constructs were made in which tandem repeats of the 25 bp AC-box oligonucleotides were cloned immediately 5' to the Herpes Simplex Virus thymidine kinase (HSVtk) promoter of the expression vector pBLCAT.2 (Luckow and Schütz, 1987). It was found that these oligonucleotides, regardless of copy number or orientation, had no significant effect on transcription from the HSVtk promoter in transiently transfected ts34 cells.

Concurrent with the transfection experiments, a computer alignment of the H1 promoters from several species revealed a distal promoter element which was highly conserved in sequence and position. This element (TG-box) was a near perfect inverted repeat of the AC-box, and the spacing between these elements was also well conserved. CAT reporter vectors were constructed which contained fragments of the 02 H1 and 01 H1 promoters. Deletion of a 29 bp region which included the TG-box resulted in a 50% decrease in expression of the 02 H1 promoter in transiently transfected ts34 cells. However, an H1 promoter fragment containing the TG- and AC-boxes, but no proximal H1 promoter elements, had no significant effect on expression of the HSVtk promoter. These data are discussed in relation to the possible role of the TG- and AC-boxes in the regulation of histone H1 gene expression.
STATEMENT

The work presented in this thesis was carried out between March 1988 and December 1993 in the Department of Biochemistry at the University of Adelaide. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying.

Signed:

(Kym Duncliffe)

Date: 8/7/94
NOTE

This thesis was created using various Macintosh computers. The text was typed using Microsoft Word (versions 4.0 and 5.1a) software packages. Macdraw Pro and CA-Cricket Graph III programs were used to create line diagrams and graphs, respectively. Autoradiographs and silver-stained proteins gels presented in Chapter 3 were scanned on a Macintosh Color OneScanner and figures created with the aid of Ofoto and Adobe Photoshop software packages. The brightness of some scans were adjusted to ease visibility of particular bands; otherwise, the figures appear exactly as scanned. In Figures 3.5 and 3.9 (a), lanes that have been spliced together were originally from separate regions of the same gel. Scanned figures were printed on a Hewlett Packard LaserJet 4 printer (300 dpi): remaining figures and text were printed on a Macintosh Personal LaserWriter.
ACKNOWLEDGMENTS

There are many to thank.

Professors W.H. Elliot and G.E. Rogers for their permission to undertake this project in the Department of Biochemistry.

Dr. J.R.E. Wells for allowing me to work in his laboratory, for his supervision, and for his ever enthusiastic support throughout the years.

Dr. B.K. May for his critical evaluation of this thesis.

Denise Turner (Protein Chemistry Unit, this Department), Dr. Jim McInnes, and Sue Rippon (IMVS), for their technical assistance with protein sequencing, photobiotinylation, and hGH assays, respectively.

Everyone in the Office and Central Services for their wisecracks (and assistance).

Brian Denton (workshop) for fixing things.

Jacquie Beale, Darren Miller, and Jenni Bennett for their invaluable help in tissue culture.

All past and present members of Labs 20, 20A, 109, 110 for their help, humour and friendship. There are far too many of you to mention, but what the hell....

The early years: Sally Aplin, Lesley Crocker, Steve Dalton, Gary Hime, Rob King, Paul Krieg, Peter McKinnon, Sarah Newlands, Janelle Rizzon (nee Kamin), Allan Robins, Linda Tabe, Kathryn Tonissen, Angela van Daal, Bill Wells, and Jenny Whiting.

The later years: Christiane Becker, Juliana Beltrame, Tim Blake, Michael Clarkson, Lesley Crocker (definitely worth a second mention), Renate Faast, Andrew Lonie, Ian Lyons, Lena Rondahl, Tom Schulz, Rick Tearle, Varaporn Thonglairoam, and Peter Wigley.

Special thanks must go to Blair Hopwood, fellow H1-factor-purifier, not only for his considerable experimental input, but for his calm, "I'm so nice you can't help but like me" manner in the face of impending doom. And of course to Duncan Sparrow, cohabiter of Lab 20 and uncanny lookalike, for his wit, conversation, musical taste, and for knowing just the right time to suggest a trip to the bar.

I wish also to acknowledge other members of the Department of Biochemistry who became good friends; including Nick Coates, Tim Cox, Nick Kallincos, Rebecca Keough,
Julie Lake (alias "Mrs Peabody"), Michael Reed, and Tim Sadlon. Thankyou all for helping to make it bearable.

My "Uncles"; Rick Tearle, Pete Wigley ("Master"), and Ian Lyons, for their guidance, and for critically reviewing this thesis.

I would also like to thank my parents and my sisters for their continued support, especially in the early years of my project.

Dora, thanks for staying with me through it all. And thankyou Alexander for helping me to understand what was important. My love to you both.

During the course of this work I was supported by a University of Adelaide Scholarship for Postgraduate Research, and countless bottles of red.
DEDICATION

This thesis is dedicated to the memory of Julian Richard Este Wells (1936-1993).
INVESTIGATIONS OF
HISTONE H1 GENE-SPECIFIC
PROMOTER ELEMENTS
CHAPTER ONE
INTRODUCTION
1.1 General Introduction

All living organisms possess the remarkable ability to create new organisms with the same characteristics as the original. This is because within each cell nucleus, these characteristics, as well as all the functions that allow the organism to survive, are encoded within the DNA sequence in the form of genes.

The number of genes in the eukaryote genome has been estimated to range in the tens of thousands (Lewin, 1975). In any particular cell, only a subset of these genes are expressed, and this subset differs depending upon the cell type. The cell draws from a wide range of mechanisms to ensure the correct temporal and spatial expression of these genes. A major focus of investigation over the last twenty years has been directed toward the regulation of gene expression. The next section describes some of the means by which genes are regulated.

1.2 The Regulation of Eukaryotic Class II Gene Expression

Unlike the prokaryotic genes which are transcribed by one type of RNA polymerase, eukaryotic genes are grouped into three classes which are transcribed by distinct RNA polymerases. The Class I genes, transcribed by RNA polymerase I, encode large ribosomal RNAs (rRNAs), which form the bulk of cellular transcripts. RNA polymerase III-transcribed genes encode a variety of different RNAs including transfer RNAs (tRNAs), 5S RNAs, U-snRNAs, and 7SL RNAs. The structure, organisation and regulation of these genes are reviewed elsewhere (Reeder, 1990; Murphy et al, 1989; Sawadogo and Sentenac, 1990). This section will focus on the Class II genes, which essentially encode the messenger RNAs (mRNAs) that are in turn translated into proteins.
1.2.1 Class II Gene Structure

The transcription unit of Class II genes is defined by a coding region which ultimately encodes the protein product, and flanking regions situated immediately 5' and 3' to the coding region. Beyond (and within) the flanking regions are sequences involved with the control of mRNA production. The coding region is almost invariably separated into sections (exons) by intervening sequences (introns). The exons often correlate with separate functional domains of the protein (Gilbert, 1973).

The primary RNA transcript undergoes a number of processing steps before it is exported from the nucleus as mature mRNA. These include the addition of a 7-methyl-G cap to the 5' terminus of the transcript (Shatkin, 1976). 3' to the translated region, the primary transcript is cleaved at a conserved polyadenylation site, and a number of adenylic acid residues are added to the 3' terminus to form the poly(A) tail (Wickens, 1990). The introns are also spliced out of the mRNA after transcription (Sharp, 1987). These processes represent potential control points for the production of the mature mRNA, thus regulating gene expression post-transcriptionally. However, it is generally accepted that the primary mechanism by which gene expression is regulated is through the initiation of transcription (reviewed, Darnell, 1982; Latchman, 1990).

1.2.2 Transcriptional Regulation

In most cases, transcription is primarily regulated by the interaction of sequence-specific DNA binding factors with DNA sequence elements (promoter elements) located 5' to the transcribed region (Dynan, 1989). Depending on the function of these factors and their cognate binding sites, these interactions either potentiate or repress the initiation of transcription by the RNA polymerase II (pol II) transcription complex. Since genes were first cloned, much research has been dedicated to the characterisation of promoter elements and transcription factors, some of which are outlined below.

1.2.2 (a) TATA-box

Located ~30 bp upstream from the transcription initiation site ("cap" site) of the vast majority of Class II genes is an AT-rich element known as the TATA-box (Corden et al,
This element ensures correct initiation of transcription. The binding of TFIID to this element is one of the initial steps in promoter activation, and is a prerequisite for the assembly of the transcription complex at the promoter. Yeast TFIID is a single protein that can complement for partially purified mammalian TFIID in a reconstituted system (Buratowski et al., 1988; Cavallini et al., 1988). However, TFIID of higher organisms is a multisubunit complex consisting of a TATA-binding protein (TBP) and several TBP-associated factors (TAFs). Surprisingly, TFIID appears to be a universal transcription factor, with demonstrated roles in transcription from TATA-less pol II-transcribed promoters, as well as pol I- and pol III-transcribed promoters (reviewed, White and Jackson, 1992). The TBP gene has been isolated from a number of species (Horikoshi et al., 1989; Peterson et al., 1990; Hoffman et al., 1990; Hoey et al., 1990).

Class II genes which do not contain a TATA-box include some constitutively expressed "housekeeping" genes, such as those encoding dihydrofolate reductase (DHFR - Nunberg et al., 1980; Chen et al., 1984) and hypoxanthine phosphoribosyl transferase (HPRT - Melton et al., 1986). The promoters of these genes are characterised by the occurrence of several GC-rich elements and/or direct repeats. Another class of TATA-less promoter without GC-rich sequences is represented by the Ultrabithorax gene of Drosophila (Biggin and Tjian, 1988).

1.2.2 (b) General Promoter Elements

The promoters of pol II-transcribed genes tend to be modular in nature, containing a series of short, discrete sequence elements which contribute in varying degrees to the overall regulation of gene expression. Perhaps the most common promoter element (after the TATA-box) is a sequence 5' CCAAT 3', known as the CCAAT-box, which is usually located ~70-90 bp upstream of the cap site (Kadonaga et al., 1986). This element is either important or essential for the transcription of several genes including herpes simplex virus thymidine kinase (HSVtk - McKnight and Kingsbury, 1982; Jones et al., 1985), β-globin (Grosveld et al., 1982), and heat shock factor (Bienz and Pelham, 1986). Several different transcription factors have been demonstrated to specifically bind the CCAAT-box. Among these are the NF-1/CTF and CP families (Mermod et al., 1989; Chodosh et al., 1988).
Chapter One: Introduction

Another promoter element common to many pol II-transcribed genes is a GC-rich sequence, with the consensus 5' G/T/G/GAAGCGG/T/A/G/A/G/T 3'. This "GC-box" can be found in various copy numbers and locations with respect to the cap site. Some of the pioneering studies of transcription-activating proteins came with the discovery of the Sp1 transcription factor (Dynan and Tjian, 1983a), which was shown to specifically bind to tandemly repeated GC-boxes in the SV40 promoter (Dynan and Tjian, 1983b; Gidoni et al, 1984). Furthermore, a functional link was established between Sp1 binding and transcriptional activation in a number of genes (Gidoni et al, 1985). Sp1 has been subsequently purified and characterised (Briggs et al, 1986).

1.2.2 (c) Specific Promoter Elements

The presence of gene-specific promoter elements enables the expression of a gene to be restricted to a certain population of cells, either at a particular stage of development or differentiation, or of a certain cell type. Often, the transcription factors which bind these elements cooperate with other "general" transcription factors to modulate promoter activity. These elements are obviously too numerous to mention, but compilations of these elements and the various transcription factors which bind them are available (Wingender, 1988; Faisst and Meyer, 1992).

1.2.2 (d) Enhancers

Transcriptional activity can also be regulated by regions of DNA known as enhancers (Benoist and Chambon, 1981), which have different properties to promoters. The most striking difference is that enhancers can function independently of position. Thus, transcription can be influenced by an enhancer either upstream or downstream of the gene, and/or over several kb from the promoter. Enhancers can also function in either orientation with respect to the promoter. Structurally, enhancers can range in size from several bp to several hundred bp, and may contain one or more copies of up to several different transcription factor binding sites. Interestingly, these sites are often found in promoters or other enhancers, indicating that promoters and enhancers function by similar mechanisms (Dynan, 1989).
Chapter One: Introduction

Three categories of enhancers are known (Maniatis et al, 1987). They include the constitutive enhancers, such as the SV40 enhancer (Jones et al, 1988), which operate in all cell types. Inducible enhancers modulate gene activity in response to environmental stimuli, such as the heat shock element (Bienz and Pelham, 1986). Temporal and tissue-specific enhancers function only in certain cell types and/or at certain stages of development or differentiation (Gillies et al, 1983).

1.2.2 (e) Locus Control Regions

Different forms of β-thalassemia have been described where the β- and δ-globin genes were deleted and the remaining γ-globin genes were either expressed at high levels, or repressed completely (see Orkin and Kazaian, 1984). This led to the proposal that the deletions caused alterations to the chromatin structure of the β-globin locus which affected gene expression. A region of nuclease hypersensitive domains ~50 kb 5' to the β-globin gene, known as the Locus Control Region (LCR) was first described by Grosveld et al (1987). When linked to either the β-globin gene or a heterologous gene, the LCR was able to confer high level, gene-copy-number-dependent expression of these genes in transgenic mice and erythroid tissue culture cells, irrespective of the site of integration.

The ability for the human LCR to function in mice illustrated the conservation of LCR function between species. Inspection of the LCR domains revealed binding sites for several general and erythroid-specific transcription factors (Talbot et al, 1990). Although enhancers also contain several transcription factor binding sites, LCRs are distinct from enhancers as the former generally only function when stably integrated into the chromosome. This observation supported the notion that the LCR conferred a transcriptionally competent chromatin conformation to the β-globin locus (Orkin, 1990). It was later proposed that the LCR was involved with interactions that led to the maintenance of a nucleosome-free region across the β-globin promoter (Felsenfeld, 1992).
1.2.3 Post-Transcriptional Regulation

1.2.3 (a) RNA Splicing

As previously mentioned, the production of a protein can also be regulated post-transcriptionally. One major control point for this form of regulation occurs when intronic sequences are spliced from the primary transcript during RNA processing. The intron/exon junctions are characterised by conserved splice sites at which the primary transcript is broken and, after intron removal, rejoined to form the mature transcript. Some of these splice sites may not be utilised during processing, resulting in transcripts which contain intronic sequences, or from which exons have been removed. Alternative splicing can lead to the production of transcripts with different 5' ends, different 3' ends, or different internal sequences (Breitbart et al, 1987; Latchman, 1990).

Many primary transcripts are spliced alternatively in different tissues, or at different developmental stages. An example of the former case occurs with the calcitonin gene (Rosenfeld et al, 1984; Leff et al, 1988). Calcitonin is a 32 amino acid-peptide produced in the thyroid gland, whereas calcitonin-gene-related-peptide (CGRP) is a 36 amino acid-peptide found in specific neurons of the brain and peripheral nervous system. These two peptides, which have different functions, differ in their carboxy-terminal regions due to the tissue-specific alternative splicing of the primary transcript. In somatic cells of Drosophila, the P-element transposase mRNA contains an intron which prohibits its export to the cytoplasm. Only in germ cells (where P-element transposition occurs) is the transcript correctly spliced to produce the transposase (Bingham et al, 1988).

Alternatively spliced transcripts can also occur in the same cell type, such as in the production of membrane-bound and secreted forms of the immunoglobulin heavy chain by B-cells (Peterson and Perry, 1986).

Some forms of alternative splicing are not regulated, but have arisen as the result of mutations which inactivate a splice site, as with mutant globin genes found in some thalassemias (Orkin and Kazazian, 1984).
Chapter One: Introduction

1.2.3 (b) 3' RNA Processing

Some transcripts are alternatively processed to generate mRNAs with different 3' ends. Such alternative transcripts are produced by some histone genes, and the regulation of these transcripts differ markedly. 3' processing is one of the major regulatory mechanisms of histone gene expression, and will be discussed in more detail (1.9.3 (b)).

1.2.3 (c) RNA Stability

The amount of protein produced is also controlled by the rate of turnover of its mRNA. This stability (or instability) may be intrinsic to the mRNA structure, or it may be influenced by some regulatory signal (Brawerman, 1987; Raghow, 1987). The half life of the transcript for the milk protein, casein, increases 40-fold in response to prolactin (Guyette et al, 1979). Similarly, heat shock leads to a 10-fold increase in the half life of the mRNA for the heat shock protein, hsp70 (Theodorakis and Morimoto, 1987). In contrast, an increase in the intracellular levels of iron results in the rapid degradation of the transferrin receptor mRNA. The 3' UTR of this transcript contains several copies of an iron responsive element (IRE) (Casey et al 1988), previously identified in ferritin mRNAs (see 1.2.3 (e)). This element is capable of forming a stem loop structure, and is bound by a regulatory protein in the absence of iron. Addition of iron results in the dissociation of the IRE-binding protein (IRE-BP) and destabilisation of the transcript (Müllner and Kühn, 1988; Müllner et al, 1989).

It should be noted that both mRNA processing and stability play crucial regulatory roles in the expression of the histone gene family (1.9.3).

1.2.3 (d) mRNA Transport

The rate of export of the mature mRNA from the nucleus to the translation machinery in the cytoplasm represents another control point in gene expression. The first example of this form of control came from studies of HIV, where the Rev protein was shown to increase the nuclear export of viral RNAs (Malim et al, 1989).
1.2.3 (e) Translation

Another control point of gene expression occurs at the level of translation. This can be achieved by regulating the amount of transcript available to the translation machinery of the cell. The pronucleus of the Strongylocentrotus purpuratus unfertilised egg contains vast stores of maternally-derived mRNAs that are not translated until fertilisation and the subsequent stages of development. In the early cleavages of sea urchin, the production of histones is dependent upon the rate of recruitment of histone mRNAs to the polysomes (Mauron et al, 1982). After the 16-cell stage, the rapid increase in early histone synthesis is partially due to an increase in the proportion of polysomal histone mRNA (Goustin, 1981).

As with regulation of mRNA stability, sequences within the transcript can affect the rate of translational initiation. Remarkably, the same type of IRE in the 3'UTR of the transferrin receptor mRNA is also found in the 5' UTR of the ferritin mRNA (Aziz and Munro, 1987; reviewed, Klausner et al, 1993). As with the transferrin receptor mRNA, this sequence is bound by the IRE-BP in the absence of iron but dissociates in the presence of iron (Rouault et al, 1988). However, this dissociation leads to increased translation of the ferritin mRNA rather than its destabilisation. This led to the proposal that the opposite effects of the IRE upon translation are determined by the position of the IRE in the transcript. This was confirmed by Casey et al (1988) when they transferred the IRE from the 3' UTR of the transferrin receptor mRNA to the 5' UTR of an unrelated mRNA. The translation of the resultant mRNA was enhanced in the presence of iron.

The purpose of the previous section was to illustrate some ways in which gene expression is controlled. This thesis deals with the regulation of the histone H1 gene; in particular, the role of cis-acting DNA sequences in the transcriptional control of this gene. The regulation of any gene cannot be fully understood without a complete knowledge of the gene product and its function. In the following sections, the histone proteins and their role in chromatin structure are discussed, as well as the organisation and expression of histone genes.


Chapter One: Introduction

1.3 Histones

The histones were first described (and named) 110 years ago by Albrecht Kossel (1884). They are a family of proteins characterised by relatively small size and high basicity. Along with non-histone proteins, they form a complex with DNA known as chromatin, which constitutes the eukaryote chromosome. As implied by their fundamental role in the nucleus, the histones are highly conserved in amino acid sequence and almost ubiquitous throughout eukaryotes. The only possible exception are the dinoflagellates, which do not appear to contain histones (Herzog and Soyer, 1981).

There are five major classes of histones: the linker histone H1 and the core histones H2A, H2B, H3 and H4. "Core" and "linker" refer to the arrangement of these histones within chromatin (see 1.4.1). An early indication that the synthesis of these proteins was strictly regulated came with the discovery that they occurred in stoichiometric quantities in several different cell types, this being equal amounts of the core histones and half the amount of H1. In addition, the mass ratio of histone: DNA was always 1:1, again regardless of species or cell type (Bonner et al, 1968).

1.3.1 H1

Usually about 210-220 amino acids in length, H1 is the largest of the histones. It is also the most heterogeneous, varying in amino acid sequence between species and between subtypes of the same species.

The H1 protein is composed of three domains (Bustin and Cole, 1970; Von Holt et al, 1979; Hartman et al, 1977). The N-terminal 40 amino acids constitute the first domain, which has a relatively high proportion of lysine and proline. The central domain of ~80 amino acids contains a high level of hydrophobic residues. This region is known as the "globular domain" because at physiological pH and ionic strength it forms a secondary structure which is characteristic of globular proteins (Bradbury et al, 1975). The globular domain is the most highly conserved region of H1. The remainder of the molecule (~95 residues) makes up the C-terminal domain, which is highly basic. As well as a high incidence of lysine (~40%), this domain contains a high number of alanine and proline
residues. The N- and C-terminal domains contribute to the heterogeneity of H1, varying considerably in sequence and length. The N-terminal and C-terminal tails have been shown to interact with the linker DNA (see Grunstein, 1990).

1.3.2 The Core Histones

The H2A histone is somewhat smaller than H2B, these being ~130 and 150 amino acids in length, respectively. The H3 and H4 histones are ~135 and 100 amino acids in length, respectively. The amino acid sequences of the latter two histones are among the most highly conserved in nature: the divergence of the H4 proteins of cow and pea by only two residues (DeLange et al, 1969) is often cited.

Each of the core histones consist of two major domains (Isenberg, 1979; Von Holt et al, 1979; Sperling and Wachtel, 1981). The N-terminal third to half of these proteins is rich in arginine and lysine, is hydrophilic, and exhibits little secondary structure in solution. Most of the variability exhibited by the core histones arises from amino acid changes in this region. Except for a short stretch of hydrophilic residues at the C-terminus, the remainder of the core histone molecules is less basic, contains a high number of hydrophobic residues, and has considerable \( \alpha \)-helical content. This hydrophobic domain is responsible for interactions with other core histones in the octamer (see 1.4.1).

1.4 Histones and Chromatin Structure

The average human diploid cell contains \( 6 \times 10^9 \) bp, or about 6 pg, of DNA. If linked into a continuous chain, this DNA would stretch for over 1.7 metres (Lewin, 1980). This must be contained within a nucleus with a diameter of about 6 \( \mu \)m. To reside within this volume, and to be accessible to nuclear processes such as replication, recombination and transcription, the DNA must be highly compacted. To achieve such a conformation, the DNA is combined with histones and certain other nuclear proteins into a nucleoprotein complex known as chromatin (reviewed, McGhee and Felsenfeld, 1980). The chromatin is uniformly folded into a hierarchical series, or "orders", of structures. In each order of
chromatin structure, the DNA is folded with increasing levels of compaction. The mitotic chromosome represents the maximum level of DNA compaction (in the order of 10^4-fold - Du Praw, 1973), and the highest order of chromatin structure.

1.4.1 The Nucleosome

1.4.1.1 Structure of the Nucleosome Core Particle

Several lines of evidence indicated that chromatin was composed of a repeating unit of a histone-DNA complex which became known as the nucleosome (Olins and Olins, 1974; Oudet et al, 1975; reviewed, Kornberg, 1977). The nucleosome core particle appeared to consist of two molecules of each core histone, which included a tetramer of H3 and H4 histones and two heterotypic H2A-H2B dimers (Kornberg and Thomas, 1974; Thomas and Kornberg, 1975a; 1975b; Oudet et al, 1975). Supportive evidence for this "octamer" structure came from studies of the binding affinities between the core histones. In particular, the high affinity between the hydrophobic domains of H3 and H4 stabilise the H3_2H4_2 tetramer (Böhm et al, 1977), while strong interactions between H2A and H2B lead to the formation of heterotypic dimers (Moss et al, 1976).

Electron microscopy, X-ray crystallography, and neutron diffraction studies (Finch et al, 1977; 1981; Klug et al, 1980; Bentley et al, 1981) have revealed that the nucleosome core is a rough disc with a diameter of 11 nm and a height of 5.7 nm, and is slightly wedge-shaped. The 146 bp nucleosomal DNA is wound ~1.75 times around the curved surface of the disc, at ~80 bp per turn. A dyad axis of symmetry passes through the centre of the structure, parallel with the long axis of the disc. The arrangement of the core histones in the octamer, originally deduced from chemical cross-linking studies (Thomas and Kornberg, 1975b; Mirzabekov et al, 1978), agreed with X-ray diffraction analysis at 7Å resolution (Richmond et al, 1984). With further X-ray crystallographic studies at 3Å resolution (Arents et al, 1991) the histone octamer was found to have a tripartite structure in which the centrally located (H3-H4)_2 tetramer was flanked by the two H2A-H2B dimers. The overall shape and dimensions of the octamer were similar to previous reports. However, the central regions of the individual histones do not consist of single globular domains as previously assumed, but
Chapter One: Introduction

of several elongated domains. In each histone dimer, the domains of each histone are extensively interdigitated with those of the other.

It is possible to isolate individual nucleosomes by treatment of chromatin with micrococcal nuclease (Kornberg, 1977; Thomas, 1983). Nucleosomes prepared in this manner consisted of a repeat length of DNA (166-241 bp depending on the source of chromatin), one core histone octamer, and one molecule of H1. Further digestion of the nucleosomal DNA initially trimmed the repeat length to 166 bp without changing the histone composition: this structure is known as the "chromatosome" (Simpson, 1978). Continued nuclease treatment released H1 and reduced the nucleosomal DNA length by another 20 bp, leaving the nucleosome core particle (Noll and Kornberg, 1977). DNase I treatment of nucleosomes nicked the nucleosomal DNA along its entire length, indicating that the DNA was wound around the exterior of the octamer (Lutter, 1978).

1.4.1 (b) Attachment of H1 to the Nucleosome

Digestion of H1 with trypsin removes the basic N-terminal and C-terminal tails of the histone. The remaining globular domain, when supplied to H1-depleted chromatin, was sufficient to protect the terminal 10 bp sections of chromatosomal DNA from nuclease digestion in the same manner as the complete histone. This indicated that the globular domain of H1 was the region which interacted with the nucleosomal DNA at the point where it entered and exited the nucleosome (Allan et al, 1980). This was supported by cross-linking studies of the chromatosome (Belyavsky et al, 1980), and electron micrographs of chromatin at various ionic strengths (Thoma et al, 1979) which suggested that H1 contacts the same face of the nucleosome as where the nucleosomal DNA enters and exits the structure. In this way, H1 would seal two complete turns of DNA within the nucleosome.

1.4.1 (c) Role of the core histone tails

The precise role of the core histone N-terminal domains in chromatin structure has proven difficult to establish. Treatment of chromatin with trypsin can remove 20-30 residues of the N-terminal "tails" with little effect on the nucleosome core (Whitlock and Stein, 1978; Ausio et al, 1989). This could be an indication that the main function of the N-terminal
domains involves interaction with DNA, for which there is some evidence (Hill and Thomas, 1990). These domains could also be involved with the formation of higher order chromatin structure: reconstituted polynucleosomes of "tailless" core histones remain unfolded in conditions that favour formation of higher order structures (Allan et al, 1982). There is also strong evidence that acetylation of these domains has a particular role in transcription (see 1.5.2).

1.4.2 Higher Order Chromatin Structure

The nucleosome is the basic repeat unit of the "nucleosome filament", a fibre of uniform thickness about 10 nm in diameter (Ris and Kubai, 1970). Under certain conditions, this can be transformed into the famous "beads on a string" structure (Olins and Olins, 1974), interpreted as nucleosome "beads" attached at intervals to the DNA. The nucleosome filament is folded to create the 30 nm fibre (Felsenfeld and McGhee, 1986; Butler, 1988), which, along with more compacted structures, are referred to as higher order chromatin structures. The structure of the 30 nm fibre has not been conclusively resolved, but several lines of evidence including electron microscopy (Thoma et al, 1979) and X-ray diffraction studies (Sperling and Klug, 1977) support the Solenoid Model (Finch and Klug, 1976). In this model, the nucleosome filament is folded into a solenoid structure with about six radially arranged nucleosomes per turn. The plane of each nucleosome disc is roughly (but not precisely) parallel with the axis of the fibre. The dyad axes of symmetry of the nucleosomes are at right angles to the fibre axis, with the H1 molecules orientated towards the centre of the structure.

There is strong evidence that H1 is essential for the formation of the 30 nm fibre. Addition of H1 molecules to H1-depleted chromatin changed the structure from a 10 nm fibre to a 30 nm fibre. Regardless of the ionic strength, H1-depleted chromatin was unable to form a stable 30 nm fibre until supplied with H1 (Finch and Klug, 1976; Thoma et al, 1979).

The proposed packaging of the 30 nm fibre into chromosomes is described by the Scaffold / Radial Loop Model (Marsden and Laemmli, 1979; Paulson, 1988). This model
Chapter One: Introduction

proposes that the 30 nm fibre, in combination with non histone proteins, is folded into loops at least 50-100 kb in length. The ends of these loops are anchored together at a "scaffold" structure which forms the axis of the chromatid. The loops are arranged radially around the central scaffold, much like the arrangement of the nucleosomes in the Solenoid Model. There is some supporting evidence for this model, such as the electron micrographs of histone-depleted metaphase chromosomes (Paulson and Laemmli, 1977).

1.4.3 Non histone chromosomal proteins

In addition to the histones, many other families of proteins are associated with chromatin. These include the HMG proteins (Johns, 1982) and scaffold proteins (Paulson, 1988). These proteins will not be discussed in this thesis, except to mention that they have been implicated in many nuclear processes ranging from the integrity of chromosome structure to roles in transcription (Stein et al, 1983; Adolph, 1988).

1.5 Relationship between Histones, Chromatin Structure, and Transcription

In a multicellular eukaryote, a minority of genes are expressed in any particular cell type. It would therefore be energetically favourable to utilise a mechanism which generally repressed transcription, while the expressed genes were activated by specific transcription factors. It is widely accepted that such a mechanism is provided by the chromatin structure itself (reviewed, Grunstein, 1990). The presence of nucleosomes at a promoter has been shown to block transcription initiation (reviewed, Kornberg and Lorch, 1992). In addition, the presence of H1 in chromatin has been associated with a repressed transcriptional state (reviewed, Zlatanova, 1990).

Alterations in chromatin structure have often been correlated with changes in transcriptional activity. Chromatin in which genes are actively transcribed is postulated to have a relaxed or "open" conformation, presumably to facilitate access of the transcription apparatus to the DNA (Reeves, 1988). One common observation is that regions of
chromosomal DNA in the vicinity of transcribed genes have an increased sensitivity to digestion by DNase I. These "DNase hypersensitive sites" are associated with a particular transcriptional state of the gene (reviewed, Elgin, 1988; Gross and Garrard, 1988). Changes in the sensitivity of these sites to DNase I digestion can be correlated with changes in the expression of the gene in response to certain stimuli, as well as with the differential expression of the gene in different cell types, stages of the cell cycle, or stages of development (Lauchman, 1990). An example of the latter situation occurs with the early histone genes of the sea urchin *S. purpuratus*. These genes are expressed at high levels during early development of the sea urchin embryo but are not expressed in the adult (see 1.9.1 (b)). The transition of the early genes from high transcriptional activity to dormancy is accompanied by a change in the chromatin state in that region of the genome (Spinelli et al, 1982; Bryan et al, 1983). During transcription, the chromatin containing the early histone genes does not have a periodic nucleosomal structure and DNase I hypersensitive sites are present in the 5' flanking regions. With the shut-down of transcription, regular nucleosomal packaging is resumed and the hypersensitive sites are lost (Fronk et al, 1990).

In many cases, alterations in chromatin structure have been attributed to the histones themselves. The following sections illustrate the means by which histones can lead to changes in chromatin structure, some of which coincide with changes in nuclear functions such as transcription.

### 1.5.1 Histone Subtypes and Variants

In most eukaryotes, histone proteins of a particular class are represented by at least two nonallelic isoforms, more commonly referred to as subtypes. The variation in amino acid sequence between subtypes is usually minor. The highest variation occurs with H1: the H1 subtypes of calf thymus differ by 15-20% (Rall and Cole, 1971).

The relative proportions of different subtypes vary in different tissues. The subtype complement has been observed to change during development (Brandt et al, 1979) and differentiation (Seyedin and Kistler, 1979; Lennox and Cohen, 1984; Zweidler, 1984; Meistrich et al, 1985). Although no obvious function may be attributed to any one subtype, this variation suggests that the subtypes are functionally distinct. Indeed, different H1
subtypes were shown to possess differing abilities to condense chromatin (Huang and Cole, 1984).

Most histone classes (with the exception of H4) contain members, known as variants, that are considerably different in amino acid sequence from the major histone subtypes. The expression of these variants often differs from that of the other members. They might only be expressed in certain cell types, or associated with particular chromatin states. They are often referred to as replacement histones because they accumulate in the chromatin of non-dividing cells by replacement of the major histone subtypes. Some of the well characterised histone variants are described below.

1.5.1 (a) H1 variants

A linker histone originally isolated from chicken red blood cells (Neelin and Butler, 1961) was unusual in that it was only found in the erythroid lineage of species that possessed nucleated erythrocytes (Neelin et al, 1964; Edwards and Hnilica, 1968). Histone H5 was considerably different in amino acid sequence from H1, although the two histones were similar structurally (Aviles et al, 1978; Briand et al, 1980).

H1 is partially replaced by H5 in the chromatin of maturing erythroid nuclei (Weintraub, 1978). There is a correlation between this replacement and the chromatin condensation that accompanies erythroid maturation. It has been postulated that H5 plays a role in this condensation, which may in turn lead to the suppression of transcription and replication (Harvey and Wells, 1984). This appears to be the case, as the binding affinity of H5 to chromatin is higher than that of H1 (Kumar and Walker, 1980; Thoma et al, 1983). H5 was also found to have a preference for the formation of higher order chromatin structures, whereas H1 did not (Thomas and Rees, 1983). Remarkably, the expression of chicken H5 in a rat sarcoma cell line resulted in the inhibition of DNA replication (Sun et al, 1989). The high concentration of arginine residues in the C-terminal tail may contribute to this higher compaction of H5-containing chromatin (Doenecke and Tönjes, 1986).

Panyim and Chalkley (1969) described an electrophoretic variant of H1 histone that was enriched in terminally differentiated tissue but absent from proliferating tissue or cells.
Chapter One: Introduction

This protein, later known as H1^0, was found to partially replace H1 in chromatin during terminal differentiation or in cells where the rate of cell division was low. H1^0 has been reported in several mammalian species and in Xenopus (Pehrson and Cole, 1981; 1980; Shimoda et al, 1981; Smith et al, 1984). As with the major H1 proteins, H1^0 consists of a central globular domain flanked by N- and C-terminal tails, but shares higher amino acid homology with H5 than with the major H1 consensus (Doenecke and Tönjes, 1986). This is reflected in immunological studies, where H5 and H1^0 were shown to share antigenic epitopes (Smith et al, 1984; Mendelson et al, 1986). In contrast to H5, H1^0 is found in a wide variety of cell types. It possibly has an analogous function to H5 in non-erythroid cells.

The occurrence of H1^0 in chromatin has been linked with an altered chromatin state. H1^0 was associated with chromatin in which the DNA had a higher resistance to nucleases than the DNA in H1-containing chromatin (Gorka and Lawrence, 1979). It also appeared to be enriched in chromatin where a gene formerly expressed in embryonic tissue was repressed in the adult (Roche et al, 1985). H1^0 has been found in the chromatin of tissues that require hormones for both their function and maintenance. In glands deprived of the maintenance hormone, the level of H1^0 in the chromatin decreases. This loss of H1^0 from chromatin was also observed in regenerating rat liver (Gjerset et al, 1982). It was concluded that the accumulation of H1^0 in the chromatin of terminally differentiated cells played a role in the maintenance of the differentiated state, possibly through the alteration of chromatin structure.

The adult testis of mammals contains high levels of an H1 variant not found in somatic tissues (Branson et al, 1975; Shires et al, 1975; Seyedin and Kistler, 1980). During mammalian spermatogenesis, somatic H1 in the chromatin of spermatocyte nuclei is replaced with the testis-specific variant, H1t. The replacement occurs during the pachytene phase of meiosis (Bucci et al, 1982). H1t is significantly divergent from the major H1 consensus (Cole et al, 1984). This probably reflects a specific function associated with the changes in chromatin structure that occur in meiosis and subsequent steps of spermatogenesis (Meistrich et al, 1985). Testis-specific histone variants of H2A, H2B and H3 have also been
identified (Shires et al., 1975; 1976; Trostle-Weige et al., 1982; Trostle-Weige et al., 1984). In addition, sperm or germ line-specific histones or histone mRNAs have been identified in sea urchin and *Caenorhabditis elegans* (Strickland et al., 1977; Sanicola et al., 1990).

### 1.5.1 (b) Core histone variants

Four H2A subtypes have been characterised in mammals: H2A.1, H2A.2, and the variants H2A.X and H2A.Z. The latter variant is highly divergent from the major H2A subtypes, and constitutes ~5% of the H2A component of chromatin (West and Bonner, 1980). In contrast to the linker histone variants, H2A.Z appears to be expressed in all tissues and at all stages of development (Ernst et al., 1987). The chicken analogue of H2A.Z was first identified by Harvey et al. (1983) with the characterisation of a chicken cDNA clone which hybridised weakly to an H2A-specific probe. This clone, named H2A.F, shared only 60% identity in amino acid sequence with a major H2A. Most of the differences between these two proteins occurred in the N- and C-terminal tails. This histone variant has been highly conserved throughout evolution, and homologous genes or cDNAs have been characterised in *Tetrahymena* (hv1), *Drosophila* (H2AvD), *S. purpuratus*, and mammals (White et al., 1988; van Daal et al., 1988; Ernst et al., 1987; Hatch and Bonner, 1988).

H2A.F/Z type histones seem to be associated with transcriptionally active chromatin. hv1 is localized exclusively in the transcriptionally active macronucleus of *Tetrahymena*; the micronucleus is transcriptionally silent (Allis et al., 1980). Immunofluorescence studies with an hv1-specific antibody labelled discrete regions in the nucleoli of mammalian cells, proposed to be the chromatin associated with the highly active ribosomal RNA genes (Allis et al., 1982). Similar studies with hv1 antiserum specifically labelled a subset of developmentally active loci in *Drosophila* polytene chromosomes (Fry et al., 1987). H2A.F/Z has also been detected either exclusively or at elevated levels in transcriptionally active chromatin fractions of mouse and chicken (Gabrielli et al., 1981; Ridsdale and Davies, 1987). The absence of functional H2AvD in *Drosophila* is lethal (van Daal and Elgin, 1992), indicating that, at least in this species, this histone variant has an essential function.

Another human H2A variant, H2A.X, is almost identical to major H2A except for a 13 amino acid extension in the C-terminal tail. Sequences similar to this extension have been
found in similar regions of H2As from yeast, *Tetrahymena*, and *Drosophila*. H2A.X has also been proposed to have some function in actively transcribed chromatin (Mannironi *et al.*, 1989).

The H3.3 variant histone is not significantly divergent in amino acid sequence from major H3 subtypes (Zweidler, 1980). However, both the gene and mRNA structure of H3.3, as well as its mode of expression, are typical of variant histones (Engel *et al.*, 1982; Wells and Kedes, 1985). Like H2AF/Z, H3.3 appears to be expressed in all tissues. H3.3 is a minor variant of embryos and rapidly dividing adult tissues in chicken, but accumulates in the chromatin of non-dividing adult tissues to become the predominant isoform of its class (Urban and Zweidler, 1983). In quiescent IMR-90 human fibroblasts, the only detectable H3 histone synthesis is that of H3.3 (Wu *et al.*, 1984). The mRNA levels of this variant remain constant during differentiation of MEL cells, resulting in a net increase in its accumulation in the chromatin (Grove and Zweidler, 1984; Brown *et al.*, 1985).

### 1.5.2 Histone Modifications

Histones are invariably subject to at least one of several different types of post-translational modifications, some of which are associated with alterations in chromatin structure (Wu, R.S. *et al.*, 1986; Matthews, 1988; Grunstein, 1990). These modifications most commonly occur at lysine residues within the basic "tail" domains. They include acetylation, phosphorylation (both O-phosphorylation of serine and threonine and N-phosphorylation of lysine and histidine), adenosine diphosphate (ADP) ribosylation, and ubiquitination. The extent of these "reversible" modifications can differ depending on the cell type, metabolic state, stage of development, or phase of the cell cycle (Lennox and Cohen, 1984; 1988; Matthews, 1988). Irreversible modifications such as acetylation of N-terminal residues and methylation are also common but little is known about their function. The most commonly studied modifications, acetylation and phosphorylation, are discussed below.
1.5.2 (a) Acetylation

Although acetylation of core histones is widespread, H1 acetylation (except at the N-terminus) has not been observed. Several correlations between acetylation state of H3 and H4 histones and transcriptional activity have been reported. The levels of hyperacetylated H4 histones in yeast chromatin was found to be ~4-5-fold higher than in higher eukaryotes, and a much higher proportion of the yeast genome is known to be transcriptionally active (40% in yeast compared to ~5% in higher eukaryotes; Hereford and Robash, 1977). All four core histones in the transcriptionally active macronucleus of Tetrahymena were found to be acetylated to varying extents. However, the histones in the transcriptionally inert micronucleus exhibited little, if any, acetylation (Vavra et al, 1982). In the slime mould Physarum, H3 histones from nucleosomes in the transcribed regions of the ribosomal genes were labelled by sulfhydryl-specific reagents, whereas H3 histones in the non-transcribed spacer were not labelled. This correlated with the observation that the chromatin structure in the transcribed regions was more relaxed than in non-transcribed regions, thereby exposing the sulfhydryl group of H3 to the reagent (Prior et al, 1983). On the basis of this finding, Allegra et al (1987) used a mercury-affinity column to selectively purify "unfolded" nucleosomes from rat liver nuclei. The DNA associated with the unfolded nucleosomes in the bound fraction were enriched with sequences actively transcribed in hepatocytes, and the histones in these nucleosomes were hyperacetylated. In contrast, a gene not transcribed in liver was enriched in the run-off fraction (containing "compact" nucleosomes), and the histones in this fraction were not hyperacetylated.

There also appears to be a requirement for acetylated histones during the assembly of chromatin, which occurs concomitantly with DNA replication. Within minutes of their synthesis, H4 histones were assembled into newly replicated chromatin (Worcel et al, 1978). Even before their deposition into chromatin, these histones were acetylated (and phosphorylated). The pattern of acetylation of newly synthesised H4 in duck erythroid cells was different from its acetylation pattern after deposition into chromatin (Ruiz-Carrillo et al, 1975; Jackson et al, 1976). This has also been observed in Physarum (Waterborg and Matthews, 1984). During embryogenesis, in which the rate of chromatin assembly is obviously increased, the requirement for acetylated histones is also implicated. Diacetylated
H4 is stored in *Xenopus* oocytes, and is a major or predominant form of H4 during early embryogenesis of *Drosophila* and sea urchin. As embryogenesis proceeds and the rate of cell doubling decreases, the amount of diacetylated H4 incorporated into chromatin also decreases (Woodland, 1979; Giancotti *et al.*, 1984; Chambers and Shaw, 1984).

Another process which seems to involve histone acetylation is spermatogenesis. Accompanying the compaction of sperm nuclei that occurs during spermatogenesis, in many species the histones are replaced by small, arginine-rich proteins known as protamines. Prior to replacement, the core histones, particularly H4, become hyperacetylated. The proportion of hyperacetylated H4 has been shown to increase dramatically during spermatogenesis in the maturing spermatids of trout and rat (Christensen and Dixon, 1982; Grimes and Henderson, 1984). This hyperacetylation is associated with relaxation in chromatin structure (Christensen *et al.*, 1984; Grimes and Smart, 1985).

It has been postulated that by neutralising the positive charge of lysine residues in the N-terminal tails of core histones, acetylation could lower the interaction of these histones with DNA (Ruiz-Carrillo *et al.*, 1975). This would in turn lead to a relaxation of chromatin structure and facilitate such processes as transcription, replication, or histone replacement. Indeed, the ability of N-terminal H4 peptides to bind DNA was abolished by acetylation (Cary *et al.*, 1982). In studies of higher order chromatin structure, McGhee *et al* (1983) found that histone hyperacetylation did not prevent formation of the chromatin solenoid. However, there was some evidence that the higher order structures composed of hyperacetylated histones were less condensed than those in which the histones were not hyperacetylated.

1.5.2 (b) Phosphorylation

An understanding of the role of phosphorylation has been complicated by the fact that the target histones, in particular the linker histones, occur as multiple forms (or subtypes) which have different numbers and positions of phosphorylation sites. In spite of this, it has been well established that an increase in the phosphorylation state of histone H1 is associated with cell division (Ord and Stocken, 1968; Balhorn *et al.*, 1972; Lennox and Cohen, 1984). In particular, there is a strong correlation between H1-phosphorylation and
chromatin condensation (Bradbury et al., 1973; Karnik, 1983). No such correlation can be found for H2A: 15% of this histone isolated from whole chromatin was found to be phosphorylated at all stages of the cell cycle (Gurley et al., 1973; Prentice et al., 1982).

It was hypothesised that phosphorylation of H1 was a key element in the initiation of chromatin condensation and subsequently mitosis (Bradbury et al., 1974). Supportive (though circumstantial) evidence for this proposal was gained with the study of temperature-sensitive mutants which displayed a link between phosphorylation and chromatin condensation. Matsumoto et al. (1980) isolated a temperature-sensitive line derived from a mouse mammary carcinoma, which arrested at the G2-phase when incubated at the restrictive temperature. Electron microscopy revealed that chromatin condensation did not occur in the arrested cells. There was a significant decrease in the phosphorylation rate of H1 in these cells compared to cells incubated at the permissive temperature. The converse effect was observed in another temperature-sensitive mutant, this time derived from a hamster kidney line (Ajiro et al., 1983). Histones from cells grown at the permissive temperature were only highly phosphorylated in mitosis. Cells grown at the restrictive temperature exhibited premature chromatin condensation (PCC), in which the chromatin attains a structure normally restricted to mitosis. In addition, H1 and H3 histones became superphosphorylated: the appearance of superphosphorylated histones in these cells correlated with the onset of PCC as observed by microscopy. Both superphosphorylation and PCC were inhibited by cycloheximide.

In studies of histone phosphorylation in dividing CHO cells, Gurley and coworkers (1978) defined two states of H1 phosphorylation that occurred at different stages of the cell cycle. During G1-phase, about half of the H1 histones were moderately phosphorylated (containing 1-3 phosphate groups per molecule) while the remainder were unphosphorylated. Cells in this stage also contained little or no phosphorylated H3. As cells progressed through S- and G2-phases, the level of histone phosphorylation increased until it reached a maximum in mitosis, where all H1 molecules contained up to 6 phosphate moieties (depending on the H1 subtype). Phosphorylation of H3 also increased dramatically as cells entered mitosis. This "mitosis state" of phosphorylation coincided with the highest state of chromatin condensation. At the end of mitosis, the phosphorylation state of H1 and
Chapter One: Introduction

H3 rapidly returned to G1-phase-levels. In these and other studies (Ajiro et al, 1981a; 1981b; Lennox et al, 1982), it was demonstrated that different histone subtypes had different patterns of phosphorylation, suggesting that the subtypes were functionally distinct. It was also evident that distinct phosphorylation sites on the same histone molecule were phosphorylated at different stages of the cell cycle, and for different durations. This suggested that the phosphorylations themselves served different functions. This variability at two levels (histone subtype and site of modification) illustrates the potential for histones to confer a wide range of chromatin states associated with different cellular activities.

1.6 Histone Gene Structure

1.6.1 Structure of the Major Histone Genes

One of the most striking features of histone genes is the absence of introns (possibly because events such as splicing would impede the high level of histone production required in embryogenesis). Such a superfamily of intronless RNA polymerase II-transcribed genes is reminiscent of the RNA polymerase I- and III-transcribed genes.

As well as the absence of introns, another feature that distinguishes histone genes from most other RNA polymerase II-transcribed genes (and resembles RNA polymerase I- and III-transcribed genes) is that the mature histone mRNAs are not polyadenylated. Instead of polyadenylation signals, the 3' flanking regions of histone genes are characterised by a highly conserved bipartite DNA sequence (reviewed in Hentschel and Birnstiel, 1981; Maxson et al, 1983a; Osley, 1991). Located within 50 bp 3' to the stop codon is an inverted repeat with the consensus sequence 5'C/AAGC/T/CTTTTCAG/A/GCCACCA 3' (Wells, 1986), which also forms the 3' terminus of most mature histone mRNAs. ~15 bp downstream from this sequence is a purine-rich element, 5' CAAGAAAGA 3' in sea urchin and 5' PuAAAGACT 3' in vertebrates. These elements are conserved throughout many species, and are essential for the correct processing of histone pre-mRNAs (reviewed in Birnstiel et al, 1985; Osley, 1991). They are also involved with the post-transcriptional regulation of histone mRNA levels (see 1.9.3).
Chapter One: Introduction

The 5' untranslated histone mRNA leader sequences are relatively short (~50 bp), and the 5' terminus of histone mRNAs are located within a cap site with the consensus 5' PyCATTCPu 3'.

Generally, the sequences required for optimal transcription of histone genes are located within ~200 bp upstream from the mRNA cap site (Heindl et al, 1988). Apart from a universally conserved TATA box located ~30 bp upstream from the cap site, the promoter architecture of histone genes varies widely across species and histone classes. In virtually all species examined, at least one histone-specific promoter element is conserved between all five classes. In plants, fission and budding yeasts, C. elegans and Drosophila, such histone-specific sequences are the only identified promoter elements to date (1.7.1 - 1.7.4). However, the histone promoters of higher eukaryotes are more complex, containing several general, histone-specific, and class-specific elements (reviewed in Osley, 1991). These elements, and their involvement with transcriptional regulation, will be discussed in more detail (1.9.4).

In species with repeated units of histone genes (see 1.7), an interesting feature of the spacer regions between some genes is the presence of di- and trinucleotide repeats (Hentschel and Birnstiel, 1981; Maxson et al, 1983a). Such repeats have been found in Drosophila (Goldberg, 1979), sea urchin (Schaffner et al, 1978; Sures et al, 1978) and Xenopus laevis (Perry et al, 1985). The number of repeats ranges from 10 consecutive copies of the dinucleotide 5' CT 3' in Drosophila to 28 repeats of 5' CA 3' in Xenopus. It has been postulated that these simple sequence repeats may be sites for recombinational events (Kedes, 1979).

1.6.2 Structure of variant histone genes

Apart from the obvious differences in the nucleotide sequence of the coding region, the genes encoding variant histones can be very dissimilar to those of the major histone classes. They often contain one or more introns, as is the case for H2A.F/Z and the H3.3 variants (Dalton et al, 1989; Engel et al, 1982; Wells and Kedes, 1985). In addition, they
generally do not contain the histone-specific 3' stem loop sequence, and their transcripts are polyadenylated (Krieg et al, 1982; Wells and Kedes, 1985; Kress et al, 1986; White et al, 1988; Dalton et al, 1989). The 5' and 3' untranslated sequences are often considerably longer than those of the major histone genes. Their promoter sequences are also usually highly divergent from those of the major histone genes, although they may share several promoter elements. In addition, they are not regulated in the same manner as the major histone genes (see 1.9.2).

The testis specific histone genes are a unique group of variant histone genes in that they contain the 3' stem loop and purine-rich sequences of replication dependent histone genes, and like these genes their transcripts are not polyadenylated (Cole et al, 1986; Kim et al, 1987; Drabent et al, 1991). The promoters of these genes appear to contain all of the conserved elements found in the major histone genes of the same class. Despite this, they are regulated in a tissue-specific, replication independent manner (see 1.9.2). A cis-acting region unique to this group of genes might account for this regulation; however, such an element has yet to be discovered.

The organisation of the variant histone genes is also different from that of the major genes in two ways. They exist as single copy genes, and they are not clustered with major genes (Krieg et al, 1983; van Daal et al, 1988).

1.7 The Organisation of Histone Genes

Throughout eukaryote species, two trends have emerged regarding histone gene organisation. The first is that almost invariably, each histone class is represented by more than one gene. Depending on the species, the number of histone genes of one class may range from a few to hundreds of copies per haploid genome. There appears to be a correlation between the number of histone genes in a given organism and the rate of cell or nuclear division during embryogenesis and / or genome size (Heintz et al, 1981; Maxson et al, 1983c). The second trend is that the histone genes are usually clustered (for reviews, see Hentschel and Birnstiel, 1981; Maxson et al, 1983a). Like histone gene copy number, the
mode of clustering varies greatly across species. In some organisms, identical, tandemly repeated clusters containing one histone gene from each of the five major classes may be located at a single locus. In other cases, clusters with different numbers and combinations of histone genes are located at several dispersed sites throughout the genome.

To illustrate the high diversity in both the number and clustering of histone genes, this topic will be discussed in relation to several species in which these genes have been studied in some detail.

1.7.1 Plats

Relatively little is known about the organisation and structure of plant histone genes although some fundamental differences between these and animal histone genes are already apparent. The genomes of higher plants appear to contain similar copy numbers of histone genes to those of higher vertebrates. The estimated copy number of the H3 gene ranges from 160 in the tetraploid genome of alfalfa (Wu et al., 1988) and 70 in the diploid genome of maize (Chaubet et al., 1986) to as low as 10 in the diploid genome of Arabidopsis thaliana (Chaboute et al., 1987). The plant genes do not appear to be tightly clustered (Chaubet et al., 1987; Chaboute et al., 1988).

In sharp contrast to the animal histone genes, plant genes do not contain the 3' stem loop sequence, and the transcripts are invariably polyadenylated. The 3' untranslated regions of these transcripts are usually much longer than those of their animal counterparts. A conserved 3' sequence element, 5' AATGAAA 3', has been found ~20 bp upstream from the 3' ends of alfalfa H3 cDNAs (Wu et al., 1989). Related sequences have been located in analogous positions of the H3 and H4 genes of maize (Chaubet et al., 1988) and Arabidopsis (Chaboute et al., 1988). It is possible that this element is the functional equivalent of the polyadenylation sequence (5' AATAAA 3') which is absent from these genes.

As with most eukaryotic genes, the promoters of plant histone genes contain a TATA box about 30 bp upstream from the cap site. Although these promoters do not appear to contain histone class-specific elements, at least three elements common to all or most of the histone gene classes have been noted. The more proximal element is an octamer, 5' GATCCGCG 3', located between -100 and -160. This element, or its reverse
Chapter One: Introduction

complement, has been identified in the promoters of all plant histone genes sequenced to date (Chaubet et al, 1986). The second conserved element is a nonameric sequence, 5' CATCCAACG 3', found in most of the plant histone promoters that have been sequenced as well as the promoters of several other plant genes (Kawata et al, 1988). It is usually located between 130 and 180 bp upstream from the cap site. The most distal element is a hexamer, 5' ACGTCA 3', located upstream of many plant histone genes between -200 and -250. This same element is also found in the promoters of several different animal histone genes (Mikami et al, 1987). Note also that part of the octamer sequence, 5'GATCC 3', is identical to a pentameric element first noted in the histone promoters of several species of sea urchin (Hentschel and Birnstiel, 1981).

1.7.2 Fungi

The eight histone genes per haploid genome of the budding yeast, Saccharomyces cerevisiae, are organised into four unlinked, divergently transcribed gene-pairs: two of which are H2A-H2B and the other two H3-H4 (Smith, 1984). The coding regions of each gene-pair are separated by ~600-800 bp. The two H3 histone genes encode identical proteins, as do the H4 genes. The H2A and H2B proteins differ by 2 and 4 amino acid substitutions, respectively. These H2A and H2B subtypes are redundant. Homozygous diploid and haploid cells in which either (but not both) of the H2B genes were mutated to produce non functional protein were viable (Rykowski et al, 1981). Similar results were obtained with mutants of either of the H2A genes, as well as a double mutant containing one functional H2A and one functional H2B gene (Kolodrubetz et al, 1982).

To date, no S. cerevisiae H1 histone protein nor gene has been identified, and several lines of evidence indicate that yeast does not contain H1 (Grunstein, 1990). The chromatin structure of S. cerevisiae appears to be fundamentally different to that of higher eukaryotes. In particular, features of chromatin that are believed to be at least partially due to H1 (eg; 30 nm fibre) are absent or almost completely absent in this species (Lohr and Hereford, 1979; Rattner et al, 1982; Szent-Gyorgyi and Isenberg, 1983).

In contrast to most animal histone genes, the 3' flanking regions of S. cerevisiae histone genes do not contain the conserved stem loop sequence, and their transcripts are
polyadenylated (check Fahrner et al, 1980). This has also been observed in Neurospora crassa (Woudt et al, 1983) and Tetrahymena (Bannon et al, 1983; Horowitz et al, 1987). It would seem that, in addition to plants, polyadenylation of histone transcripts is prevalent in lower eukaryotes. A conserved sequence in the 3' flanking regions (5'TAGNaTATGN17TT 3') has been proposed to confer correct 3' processing to yeast mRNAs (Zaret and Sherman, 1982).

The promoters of the yeast histone genes contain common sequence elements that are required for expression. As well as a TATA-box upstream of each coding sequence, two or three copies of a 16 bp sequence, 5' GCGAAAAANTNNGAAC 3', can be found in the intergenic spacer of each of the four histone loci (Osley et al, 1986; Freeman et al, 1992).

A similar arrangement of histone genes is found in the fission yeast, Schizosaccharomyces pombe. However, the H2B gene appears to be present as a single copy which is associated with one H2A gene; the second H2A gene is unpaired. S. pombe has three H3-H4 gene-pairs, all of which are divergently transcribed (Choe et al, 1985; Matsumoto and Yanagida, 1985). A single copy of a 17 bp element with the consensus 5' ATCAC/AAACCCTAACCCT 3' is located in the spacer between each of the divergently transcribed genes and upstream of the isolated H2A gene. This element is possibly analogous to the conserved promoter element of S. cerevisiae.

The histone gene complement of Aspergillus nidulans is highly unusual in that all genes contain at least one intron. Aspergillus has two H4 genes and a single copy of the other core histone genes. As with the other yeasts, the genes are organised into H2A-H2B and H3-H4 gene-pairs and one unpaired H4 gene (Ehinger et al, 1990). In contrast to the other yeasts, the histone genes of Aspergillus do not appear to share a common promoter element. The genes, however, are cell cycle-regulated, although one of the H4 genes appears to be expressed in a partially replication dependent manner (Osmani et al, 1987; Ehinger et al, 1990).

1.7.3 Caenorhabditis elegans

There are fundamental differences in both structure and arrangement between the H1 and core histone genes of the nematode C. elegans. The core histones are found clustered at
about 11 different sites throughout the genome (Roberts et al, 1987). Four of these sites contain at least one copy of each of the core histone genes, with H2A-H2B and H3-H4 gene pairs divergently transcribed. In the 5' flanking regions of these genes, the TATA box is the only promoter element common to histone genes of higher eukaryotes. A few bp upstream from this element, a conserved 20 bp sequence (5' CTCCNCTNCCACAN 3') is unique to the C. elegans core histone genes so far examined (Roberts et al, 1989). A second conserved element, 5' CTGCGGGGACACATNT 3', is also found in the intergenic spacer between each gene-pair. The function of these elements is unknown. The 3' stem loop sequence is conserved in these genes, although there is no obvious homology to the purine-rich sequence found in most histone genes.

There are about five H1 histone genes dispersed throughout the genome of C. elegans, none of which are associated with the core histone gene clusters. One of these genes has been isolated and sequenced (Sanicola et al, 1990), revealing two highly unusual features. The first is that the gene contains an intron within the amino-terminal third of the coding sequence. Introns have also been reported within H1 histone genes of Tetrahymena (Wu, M. et al, 1986) and Arabidopsis (Gantt and Lenvik, 1991).

The second unusual feature of the C. elegans H1 gene was the absence of the stem loop sequence found in the core histone 3' untranslated regions. However, a polyadenylation signal was present 3' to the H1 coding region, and the H1 transcripts from this gene were indeed polyadenylated (Sanicola et al, 1990). It is not known whether this unusual histone gene is representative of the remaining H1 genes of C. elegans. If so, then C. elegans is unique in that the regulation of one class of histone genes (H1) appears to be quite different from the other classes.

1.7.4 Drosophila

The five major histone genes of Drosophila melanogaster are organised into a unit which is reiterated about 110 times per haploid genome (see Kedes, 1979). These units are clustered at a single chromosomal locus (Pardue et al, 1977). Two forms of the unit exist; a 5.0 kb (L) unit and a 4.8 kb (S) unit, which occur in a ratio of ~3:1. They differ by a 240 bp tRNA-derived insertion element within the H1-H3 spacer of the L unit (Lifton et al, 1978;
Matsuo and Yamazaki, 1989). Both the L and S units are grouped largely with members of their own type. Although the units are tandemly repeated, the locus is interspersed at several sites with regions of non-histone-coding DNA (Saigo et al., 1980).

Within each unit, the H2A and H2B genes are divergently transcribed, with only ~200 bp separating the 5' ends of the coding regions. This is also the configuration of the H3 and H4 gene-pair. A similar organisation of histone genes is found in Drosophila hydei (Kremer and Hennig, 1990). In addition, D. hydei has a similar number of histone genes to D. melanogaster, which are also found at a single locus. However, in Drosophila virilis, the histone genes are clustered at two loci (Anderson and Lengyel, 1983) and there appears to be a unit that lacks the H1 gene (Domier et al., 1986).

The entire histone repeat unit of another insect, Chironomus thummi thummi, has also been sequenced (Hankeln and Schmidt, 1991). The arrangement of the histone genes within the unit is similar to that of Drosophila except that the orientation of the H1 gene is reversed. In contrast to Drosophila, at least five loci of histone genes are found in the Ch. th. thummi genome, all of which are located on a single chromosome arm. Three length variants of the histone repeat units exist, none of which correspond to the predicted size of the basic repeat unit (5.2 kb). One of these variants (6.2 kb) contains a 1048 bp transposable element (TFB1) within the H1-H3 spacer (Hankeln and Schmidt, 1990).

With the exception of the TATA box, none of the promoter elements of vertebrate histone genes are located in the Drosophila histone promoters. A conserved sequence (5' AGTGAAA 3') is found in the 5' untranslated regions of the 5 histone genes. Another sequence, 5' CCCTC^{T/G} 3', which was first identified upstream of all Drosophila hydei histone genes (Kremer and Hennig, 1990), is only found upstream of some D. melanogaster and Ch. th. thummi histone genes.

The 3' flanking regions of the Drosophila histone genes contain the conserved stem loop and purine-rich sequences common to replication dependent histone genes. The Drosophila histone mRNAs are not polyadenylated (cited by Lifton et al., 1978). However, the genes also possess a polyadenylation signal (5' AATAAA 3') downstream of these conserved sequences. The relevance of the polyadenylation sequences are unknown, but it
has been proposed that these genes may be trancribed to produce both poly(A)− and poly(A)+ histone mRNAs (Matsuo and Yamazaki, 1989).

1.7.5 Sea Urchin

The genome of the sea urchin contains both tandem arrays of the five major histone genes as well as dispersed histone genes ranging from isolated, single copy genes to small clusters. The tandemly arranged histone genes are known as the early genes and encode the α-histone variants. Depending on the species, the repeat unit of the early genes is ~6.3 - 7.2 kb in length and is reiterated ~300-500 times (Maxson et al, 1983a). These highly repeated units allowed the sea urchin histone genes to be among the first eukaryotic genes isolated by molecular cloning techniques (Kedes et al, 1975). Each unit consists of one copy of each histone gene in the same orientation. The dispersed (or late) histone genes consist of 2-12 members per haploid genome (Childs et al, 1982; Maxson et al, 1983b; Knowles and Childs, 1986; Lai and Childs, 1988). These genes encode the β- and γ-histone variants, which differ in both size and sequence from the α-histone variants. The arrangement of the late histone genes is also different from the early genes. Divergently transcribed H2A / H2B (Maxson et al, 1983b) and H3 / H4 (Childs et al, 1982) gene pairs have been isolated; these gene pairs are not adjacent in the early histone gene repeat unit.

As well as the differences mentioned above, there is variation in the temporal expression of the early and late sets of histone genes during the development of the sea urchin embryo (see 1.9.1 (b)).

As mentioned above, the α-histone variants differ in amino acid sequence from the β- and γ-histone variants from the same class. The nucleotide sequences of the early and late H3 coding regions of *Lytechinus pictus* are 19% divergent (Roberts et al, 1984). It has been estimated that the early and late genes evolved from an ancestral sequence some 700 million years ago. This corresponds to the Precambrian period, prior to the bifurcation of the echinoderms from the line leading to vertebrate phyla (Maxson et al, 1983c). It should therefore be possible to find descendants of these ancient early and late genes in other lineages.
The 5' and 3' flanking regions of the early and late histone genes are more diverse than the coding regions. However, they are punctuated with common promoter elements nearing 100% homology. The 3' flanking regions of early and late genes contain the histone-specific stem loop sequence and purine rich sequence (1.6.1). As with lower eukaryotes, the 5' flanking regions of both the early and late genes contain sequences common to most, if not all five histone classes. In addition, histone genes of a particular class share one or more promoter elements specific to that class, as observed throughout the higher eukaryotes. These elements, and the factors which interact with them, will be discussed in more detail (1.9.4).

**1.7.6 Trout**

The rainbow trout, *Salmo gairdnerii*, represents the most primitive vertebrate to date in which histone gene organisation has been studied (Connor *et al.*, 1984). As with simpler eukaryotes and amphibians, one copy of each of the five major histone genes are grouped into a repeat unit. This unit (10.2 kb) is reiterated ~145 times in the haploid genome. As with the sea urchin repeat units, all five genes are in the same orientation; however, the trout genes occur in a unique arrangement. There is no evidence that these units are tandemly repeated.

**1.7.7 Amphibia**

Amphibians have among the largest genomes of any known organism, as well as the highest copy number of histone genes. In the axolotl, *Ambystoma mexicanum*, histone gene reiteration frequency has been reported to exceed 2000 copies per haploid genome (Hilder *et al.*, 1981). Of the urodeles, only in the American spotted newt *Notophthalmus viridescens* has histone gene organisation been studied in detail. One histone gene from each of the five major classes are clustered into a 9 kb unit that is repeated 600-800 times per haploid genome (Stephenson *et al.*, 1981a; 1981b). The genes are transcribed in the same direction with the exception of the H2B gene (associated with H2A in a transcriptionally divergent configuration). These clusters are not tandemly repeated but interspersed with variable lengths (50 kb or more) of satellite DNA.
Chapter One: Introduction

Much investigation of amphibian histone genes has been carried out in the two frog species, *Xenopus borealis* and *Xenopus laevis*. Estimates of histone gene copy number per haploid genome have ranged from 20 in *X. laevis* (Jacob et al., 1976) to 90 in both species (Turner and Woodland, 1983). A more recent estimate in *X. laevis* puts the number of copies at about 50 (Perry et al., 1985). The organisation of histone genes in *Xenopus* is somewhat complicated. Although they are clustered, there is considerable heterogeneity of the histone genes between clusters, from the configuration of the genes to their nucleotide sequence (Zernik et al., 1980; van Dongen et al., 1981; Old et al., 1982). However, it has been established that a 16 kb reiterated "major" cluster, containing one histone gene from each class in a set configuration, accounts for 70% of the histone complement of *X. borealis* (Turner and Woodland, 1983). It is not known if this cluster is repeated in tandem. In *X. laevis*, the heterogeneity appears even more pronounced. Three variant H1 genes were isolated, and each was associated with a different cluster of core histone genes (Destree et al., 1984). Two of these clusters have been cloned and sequenced (Perry et al., 1985). The major cluster (~15 kb) contains one histone gene from each class and appears to be tandemly repeated. This is also the case for the other cluster except that the gene order (and nucleotide sequence) is different.

1.7.8 Chicken

Possibly the entire histone gene complement of the chicken genome has been mapped (D'Andrea et al., 1985). In total, over 40 histone genes, including 8-10 copies of each of the core histone genes and 6 H1 genes, are grouped into ~10 clusters. The combination of histone genes in each cluster appears to be random, and there is no evidence for a repeating unit of genes as found in lower eukaryotes. Although no two clusters are identical, preferred associations are apparent, such as diverently transcribed H2A/H2B gene pairs often closely linked to an H1 gene. In addition, H3 genes are often found near H4 genes. The clusters rarely (if ever) contain histone genes from each class, but often contain two members of one class and none from another. In fact, some clusters seem to have arisen from inverted duplications of smaller clusters (Wang et al., 1985).
1.7.9 Mammals

Like those of the chicken genome, the murine (Sittman et al, 1981) and human (Heintz et al, 1981; Sierra et al, 1982) histone genes are arranged in dispersed, heterogeneous clusters. The mouse genes are found in two clusters on chromosomes 3 and 13, with most genes apparently on the latter chromosome (Marzluff and Graves, 1984). The copy number of mouse histone genes has been estimated at about 10-20 (Jacob, 1976). An interesting feature of the mouse histone genes is that single copy genes on chromosome 3 are expressed at extremely high levels, accounting for 30-40% of the total histone mRNAs of their respective classes (Graves et al, 1985). The high expression of the chromosome 3 genes has been attributed to both increased transcription and efficient 3’ processing of the histone transcripts (Levine et al, 1988; Liu et al, 1989).

The human histone genes have been mapped to chromosomes 1, 6, and 12 (Triputti et al, 1986). Isolation of genomic clones containing human histone gene clusters revealed no repeating arrangement of the genes. It was proposed that the lack of histone gene organisation in human and other higher eukaryotes reflected the loss of selection pressure for high rates of histone synthesis during embryogenesis; species with extremely rapid cell cycle rates at this developmental stage retained high numbers of tandemly reiterated genes (Heintz et al, 1981).

1.8 Regulation of Histone Gene Expression: Lower Eukaryotes

In general, the histone genes of lower eukaryotes are replication dependent, and the major regulatory mechanisms act at the transcriptional level. The control of lower eukaryote histone gene expression has been most extensively studied in Saccharomyces cerevisiae (reviewed, Breeden, 1988; Osley, 1991).

Osley et al (1986) made LacZ fusions with the H2A and H2B genes of the TRTI locus and studied the effects of various promoter deletions in the intergenic spacer. This defined a 250-300 bp region that, when deleted, resulted in almost complete loss of transcriptional activity from both genes in exponentially growing cells. This region
Chapter One: Introduction

contained three of the 16 bp repeats conserved in the 5' flanking regions of all four yeast histone gene-pairs (see 1.7.2). These sequence elements also play a crucial role in the cell cycle regulation of these genes. Not only were three tandem copies of the 16 bp repeat able to function as a UAS, but they were sufficient to confer periodic regulation of transcription to a normally constitutive promoter (Osley et al, 1986). The same properties were demonstrated for the 16 bp repeats of the yeast H3 gene (Freeman et al, 1992).

In the same series of experiments mentioned above, a 70 bp region in the vicinity of (but not including) the 16 bp repeats appeared to negatively regulate transcription. Deletion of this "CCR" (cell cycle regulation) region led to elevated levels of histone-LacZ transcripts in exponentially growing cells. The same elevated levels were observed in cells arrested in G1 at a point where histone genes are not normally expressed. It was subsequently shown that the histone-LacZ fusions were constitutively expressed throughout the cell cycle. This implied that the function of the CCR was to repress histone gene transcription outside the normal G1/S period of histone synthesis. This was reinforced by the finding that the constitutively expressed CYCI-LacZ fusion gene was periodically expressed during the cell cycle when the CCR was placed upstream of the CYCI UAS (Osley et al, 1986). Similar negative regulator regions have been found in the intergenic spacers of both H3-H4 loci. However, although deletion of the negative element (NRS) of one of these loci resulted in increased transcription, the cell cycle periodicity of histone expression was unaffected (Freeman et al, 1992).

Yeast histone mRNAs are also subject to post-transcriptional mechanisms of cell cycle regulation, predominantly at the level of mRNA turnover. This mechanism appears to be replication independent (Lycan et al, 1987). A 360 bp region of the H2B1 gene including the carboxy-terminal 17 amino acids and the entire 3' untranslated region was sufficient to confer post-transcriptional cell cycle regulation to a neo gene (Xu et al, 1990). Levels of neo-H2B fusion mRNA increased 12-fold during S-phase in synchronised yeast cells and returned to basal levels at the end of S-phase. Deletion of the 54 bp coding region from this fusion did not affect the accumulation of the fusion transcript, but curiously the timing of the accumulation was considerably delayed (early M phase).
Chapter One: Introduction

The rate of histone mRNA turnover in yeast is also affected by histone gene copy number. Doubling the number of H2B1 genes resulted in a 2-fold decrease in the half-life of the H2B1 transcript (Osley and Hereford, 1981). Again, this effect was mediated by sequences in the 3' region of the gene (see Osley, 1991).

1.9 Regulation of Histone Gene Expression: Higher Eukaryotes

The histone genes fall into two broad classes based upon how they are regulated. Most histone genes are replication dependent, meaning that their expression is closely correlated, or coupled, to DNA synthesis. It has been known for some time that histone synthesis occurred in the same phase of the cell cycle as DNA synthesis (Robbins and Borun, 1967). The expression of replication dependent histone genes is controlled at both transcriptional and post-transcriptional levels (reviewed, Osley, 1991). At the onset of S-phase, transcription of these histone genes increases three to five fold higher than non S-phase rates (DeLisle et al, 1983; Heintz et al, 1983; Plumb et al, 1983; Baumbach et al, 1987). In addition, both efficiency of 3'-mRNA processing and mRNA stability increase markedly at this time, which in combination with the increased transcription rate, accounts for the 15-fold to 50-fold increase in steady state mRNA levels (DeLisle et al, 1983; Heintz et al, 1983).

By the end of S-phase the mRNA concentration rapidly returns to non S-phase levels. If DNA synthesis is inhibited during S-phase, a similar drop in histone transcript levels is observed. Removal of the DNA synthesis block leads to a reaccumulation of histone mRNA (Heintz et al, 1983). Additionally, the decrease in transcript levels in the presence of DNA synthesis inhibitors can be prevented by cycloheximide, implying that the degradation of these transcripts requires protein synthesis (DeLisle et al, 1983; Baumbach et al, 1987).

There was some disagreement as to whether the rate of transcription decreased (DeLisle et al, 1983; Baumbach et al, 1987) or remained unchanged (Heintz et al, 1983;
Sittman *et al.*, 1983) in the presence of DNA synthesis inhibitors. However, histone mRNA levels decreased dramatically under these conditions and this was at least partly due to a considerable decrease in the half-life of the transcripts. Thus it appears that a major contribution to the regulation of replication dependent histone mRNA levels during the cell cycle is from post-transcriptional mechanisms.

The expression of the other class of histone genes, which includes most (if not all) of the variant histone genes, is replication independent. It had been known for some time that production of the variant histones was not coupled to DNA synthesis (Appels and Wells, 1972). These genes are usually expressed at a constant, basal level throughout the cell cycle (Wu and Bonner, 1981; Chabanas *et al.*, 1983; Harris *et al.*, 1991). Dalton and coworkers (1986; 1989) have demonstrated that both the steady state levels and the transcription rate of H5 and H2A.F mRNAs remained unaltered throughout S-phase. In contrast to the replication dependent histone genes, the expression of the H10, H3.3, and testis specific histone genes are unaffected by DNA synthesis inhibitors, as measured by mRNA or protein levels (Broke *et al.*, 1980; Zlatanova, 1980; Sittman *et al.*, 1983; Brown *et al.*, 1985; Chiu and Irvine, 1985). In fact, some variant histones were found to be induced by DNA synthesis inhibitors (Pehrson and Cole, 1980).

In some species, the requirement for histones can be so great that even the expression of replication dependent histone genes becomes temporarily uncoupled from DNA synthesis. Before the discussion of variant histone gene expression, the phenomenon of embryonic uncoupling is presented.

### 1.9.1 Replication Independent Expression: Embryonic Uncoupling

In organisms where the rate of cell or nuclear division during embryogenesis is extremely rapid, the demand for histones is obviously high. This demand can be met by several means, such as increasing the rate of histone mRNA transcription, or utilisation of stored mRNAs that were produced in the absence of DNA replication. In other words, the production of histones is not coupled to DNA synthesis. Embryonic uncoupling has been
Chapter One: Introduction

described in *Drosophila*, sea urchin and *Xenopus* (reviewed by Woodland, 1980; Hentschel and Birnstiel, 1981; Maxson et al, 1983a; 1983c).

1.9.1 (a) *Drosophila*

In the first 90 minutes following fertilisation, 9 synchronous nuclear divisions take place in the *Drosophila* egg (Zalokar and Erk, 1976). No gene activity has been detected during this rapid developmental phase, thus the embryo relies upon maternal stores of histone mRNA (which is the most abundant mRNA in the mature oocyte; Anderson and Lengyel, 1984) for chromatin assembly (Zalokar, 1976). Anderson and Lengyel (1980) have demonstrated that embryonic histone mRNA synthesis begins ~90 minutes after oviposition; the histone genes are therefore among the first to be expressed in the *Drosophila* embryo. Throughout the subsequent stages of blastoderm formation and gastrulation, the rate of histone mRNA accumulation in the cytoplasm increases to a peak level that coincides with the maximum rate of DNA synthesis. This is accompanied by an increase in the proportion of histone mRNAs that are associated with the polysomes, indicating that these mRNAs are under translational control (Anderson and Lengyel, 1984). Following gastrulation (~6 hours after oviposition), there is a sharp decline in histone mRNA levels, brought about by a decrease in both mRNA synthesis and stability.

1.9.1 (b) Sea urchin

The early and late histone gene families of sea urchins are differentially expressed during oogenesis and embryonic development. The pronucleus of the sea urchin egg contains large stores of maternally-derived histone mRNA transcribed almost entirely from the early genes (Mauron et al., 1982; DeLeon et al., 1983). These transcripts were known to be produced in the late stages of oocyte maturation (Mauron et al., 1982; Angerer et al., 1984): therefore, they were transcribed up to several months after the cessation of DNA synthesis. In the first six hours after fertilisation (early cleavage), the zygote undergoes 4 synchronous cleavages. Transcription of early histone genes during this period accounted for only 15% of the total histone mRNA; the 48-fold increase in histone protein synthesis during early cleavage was due to the recruitment of maternally-derived histone mRNA.
Chapter One: Introduction

(Goustin, 1981). In addition to this form of embryonic uncoupling, histone synthesis was detected in the G2 phase of these early cleavages (Arceci et al., 1977).

During the period of rapid asynchronous cleavage (6-14 hours post-fertilisation), in which S-phase is as short as 10 minutes, the embryo grows from 16 to ~300 cells. The accumulation of early histone mRNAs also increases dramatically and peaks at ~10-12 hours post-fertilisation (early blastula). At this stage, the levels of core histone mRNA have increased by ~10-fold over the levels in the unfertilised egg. This is due to an increased rate of histone mRNA synthesis (Maxson and Wilt, 1981; Mauron et al., 1982; Weinberg et al., 1983) and an increase in the proportion of histone transcripts sequestered to the polysomes (Goustin, 1981). Histone synthesis also increases by ~8-fold to account for 25-40% of total protein synthesis (Kedes et al., 1969; Seale and Aronson, 1973). From 12 hours post-fertilisation (~200 cells), there is a rapid decrease in early histone mRNA levels. As with Drosophila, this decline is caused by decreases in both mRNA synthesis and in mRNA stability (Mauron et al., 1982; Maxson and Wilt, 1981; 1982; Weinberg et al., 1983).

Although transcripts from the late histone genes are also stored in the unfertilised egg, late histone mRNAs are ~100-fold less abundant than early histone mRNAs (Knowles and Childs, 1984). Using probes derived from early and late H2B coding regions that specifically hybridised their respective transcripts, Maxson et al. (1983b) were able to show that the levels of early and late histone mRNAs of S. purpuratus differed throughout embryogenesis. The embryonic expression of the late genes was activated at about the same time as the early genes. In contrast to the rapid increase in early histone mRNAs up to early blastula, the levels of late histone mRNAs remained relatively low until mid-to-late blastula (14-16 hours post-fertilisation), whereupon there was a rapid (15-fold) increase in the rate of late transcript accumulation. In studies of another sea urchin species, Lytechinus pictus, a 164-fold increase in the level of late transcripts between the egg and gastrula (48 hour) stage was observed. It has since been shown that the maximum transcript levels of different late histone subtypes occur at different stages of development (Mohun et al., 1985; Kemler and Busslinger, 1986). The increase in late histone mRNA levels is primarily due to increased transcription (Ito et al., 1988). By the late blastula stage, expression of the early histone
Chapter One: Introduction

genes ceases for the remainder of the sea urchin life cycle, while the late histone genes continue to be expressed in the adult (Kemler and Busslinger, 1986).

1.9.1 (c) *Xenopus laevis*

The *Xenopus* genome has far fewer histone genes than those of sea urchin or *Drosophila*, and yet the time of DNA replication during early development is roughly the same (S-phase of 12 minutes; Graham and Morgan, 1966) as in these other species. In addition, the genome of *Xenopus* is over 5 times larger than the sea urchin genome and 20 times larger than the *Drosophila* genome. The demand for histones is therefore met with maternally-derived stores of histone mRNA and protein (Woodland, 1980). These vast stores (containing sufficient core histones for 20,000 diploid nuclei) are accumulated throughout oogenesis within the germinal vesicle. Interestingly, a proportion of the histone mRNA is polyadenylated (Levenson and Marcu, 1976). Hormonally-induced maturation of oocytes is accompanied by a 50-fold increase in the rate of core histone synthesis. This is regulated by posttranscriptional events, as the rate is unaffected by physical or chemical enucleation. In the mature egg, polyadenylated histone transcripts are no longer detectable, despite there being no decrease in the total histone mRNA.

Although there are at least two distinct types of histone gene clusters in *X. laevis*, there is no differential transcription from different clusters as observed in sea urchin. Transcripts derived from both clusters are present throughout oogenesis and embryogenesis (Perry *et al.*, 1986). The transcription of these genes is in *X. laevis* oocytes is replication independent, while the same genes follow a replication dependent mode of expression in somatic cells (Old *et al.*, 1985).

Embryonic histones are not detected until the 1000-2000 cell blastula stage, and do not become predominant until early gastrula (Woodland *et al.*, 1979). Thus in *Xenopus*, uncoupling of histone and DNA synthesis in oocytes is essential for producing sufficient histones for the rapid cleavages in early development that could not be provided simply by S-phase transcription of a relatively low number of histone genes. The rate of cell division during embryogenesis of higher eukaryotes such as mammals is considerably lower, and
presumably coupled histone and DNA synthesis can account for the histone requirement (see van Dongen et al, 1984).

1.9.2 Replication Independent Expression: Variant Histone Genes

As already mentioned, both the promoters and 3' flanking sequences of variant histone genes are different from those of the major histone genes. The absence of the 3' palindromic stem loop and purine-rich sequences from most of these genes signifies that they lack an important mechanism of S-phase regulation (see 1.9.3). In addition, variant histone promoters often lack putative S-phase regulatory elements that are present in the replication dependent genes (1.9.4).

H5 is a prime example of such a variant histone gene. The H5 promoter contains none of the conserved H1-specific promoter elements. The promoter has been localised to a 115 bp region immediately upstream of the cap site (Rousseau et al, 1989). An upstream negative element (UNE) was uncovered between -115 and -90: deletion of this region caused a slight increase in H5 transcription. The UNE was also capable of decreasing transcription from the HSVtk promoter, and this effect was not lineage-specific. Two elements which positively affected H5 transcription were also identified. The more distal was a GC-box located at about -80. The second element, ~20 bp in length and known as the upstream positive element (UPE), was positioned a few bp upstream of the TATA-box. Two interesting features of the UPE were that it contained homology to the H4 subtype-specific element (see 1.9.4 (d)), and that it was highly conserved in sequence and position in two H10 promoters (Rousseau et al, 1989). Other potential promoter elements include an inverted GC-box and several CACCC elements (Schüle et al, 1988).

Trainor et al (1987) examined the expression of a cloned H5 gene in erythroid and non-erythroid cell lines. Using an enhancer trap system, they identified a fragment 3' to the H5 coding region that displayed erythroid-specific enhancer-like activity. This was confirmed with the demonstration that this fragment could potentiate transcription from an SV40 early promoter in either orientation, and either 5' or 3' to the reporter gene. This enhancer contained a 34 bp sequence with striking homology to a sequence in the 3' erythroid-specific enhancer of the chicken β-globin gene (Choi and Engel, 1986). Thus, the
Chapter One: Introduction

tissue-specific expression of the H5 gene appears to be regulated by a similar mechanism to other erythroid-specific, non histone genes.

Surprisingly, several 5' and 3' sequence elements with demonstrated roles in the replication dependent expression of the major histone genes are also found associated with some variant histone genes. As well as the ubiquitous TATA-box, the mouse H10 promoter contains GC-boxes and the H1-specific AC-box, all of which are found at similar locations in replication dependent H1 genes (Breuer et al, 1989). In addition, several elements from other histone and non-histone promoters were noted (Steuer et al, 1992). These included HNF-4-like- and Oct-1-like-binding sites, and a H4 subtype-specific element in a similar location to where it is found in the H4 promoter (and in the H5 promoter: see above). In contrast to the major H1 genes, the H10 gene lacks a conserved CCAAT-box. In addition, the human H10 gene does not contain the 3' stem loop and purine rich sequences (Doenecke and Tönjes, 1986).

H10 expression can be induced in several cells lines by various treatments, most of which also induce the cell line to differentiate (Keppel et al, 1977; Pehrson and Cole, 1980; Alonso et al, 1988; Kress et al, 1986). One of the few genes transcribed soon after retinoic acid (RA)-induced differentiation of F9 teratocarcinoma cells is H10 (Alonso et al, 1988). F9 cells were transiently transfected with constructs bearing various H10 promoter deletions. It was demonstrated that a region between -600 and -400 was necessary for both efficient transcription in undifferentiated F9 cells and for the RA-induced increase in transcription (Breuer et al, 1989). Recently, further analysis of this region defined an 80 bp fragment between -531 and -451 that led to the same decrease in transcription when deleted. This region was also reported to enhance transcription from a heterologous promoter (Steuer et al, 1992). Several internal deletions between this region and the cap site also reduced transcription considerably, as did mutations in either the AC-box, Sp1- or H4 subtype-specific elements. It was concluded that the distal 80 bp region cooperated with elements in the proximal promoter to effect efficient basal expression of the H10 promoter (Steuer et al, 1992).

Another recent finding is that the H10 mRNA appears to accumulate in a cell cycle-specific manner. In murine erythroleukemia (MEL) cells, the levels of H10 mRNA were
higher in S-phase than in the G1- or G2-phases, as was shown for H3 mRNA. When these cells were induced to differentiate, the accumulation of H10 mRNA was greatest when induction occurred during S-phase (Grunwald et al, 1991). The same group detected a proliferation-associated increase in H10 mRNA in rat liver cells following partial hepatectomy, despite the decrease in H10 protein in these cells (see 1.5.1 (a)). This indicated that the events controlling H10 protein and transcript levels in regenerating rat liver were uncoupled (Khochbin et al, 1991).

The testis-specific histone genes share both 5' and 3' sequence elements in common with the replication dependent genes of the same class. Like H10, the rat and human H1t promoters share several promoter elements in common with the major H1 (Cole et al, 1986; Drabent et al, 1991). The rat H1t promoter was able to direct S-phase-specific regulation of a CAT reporter gene in somatic cells (Kremer and Kistler, 1992). Similar results were obtained with a testis-specific H2B (TH2B) promoter (Hwang and Chae, 1989). Furthermore, the S-phase regulatory region of the TH2B gene was bidirectional, and could also direct S-phase-specific transcription from the TH2A promoter in somatic cells (Huh et al, 1991). These findings are surprising given the tissue-specific, replication independent regulation of these genes. The mechanism(s) by which this regulation is achieved is unknown.

In addition to these replication independent histone genes, a class of partially replication dependent histone genes has been described (Zweidler, 1984). In terminally differentiating tissues, the levels of partially replication dependent histone mRNAs were increased at the beginning of S-phase but not completely repressed at the end of S-phase. Such a pattern of expression was demonstrated for the mouse H3.2 gene (Brown et al, 1985). An explanation for this phenomenon was observed in a study of the human H2A.X gene (Mannironi et al, 1989). The 3' untranslated region (UTR) of this gene was found to contain both the histone specific stem loop and purine-rich sequences, as well as a polyadenylation signal ~1 kb downstream. This gene was transcribed to produce two alternative mRNAs: a poly(A)+ transcript in which the stem loop sequence resided at the 3'
terminus, and a poly(A)^+ transcript which had a much longer 3' UTR. The poly(A)^- transcript was regulated in the same manner as a replication dependent histone mRNA, while the level of polyadenylated mRNA was neither increased as cells entered S-phase, nor affected by DNA synthesis inhibitors. As with replication independent histones, partially replication dependent histones may have a specific function in the chromatin of terminally differentiating cells, and the transcription of polyadenylated mRNAs from these genes provides a means of accumulating these histones in the chromatin even in the absence of DNA synthesis.

Alternatively processed, independently regulated histone transcripts have also been observed for an H1 gene from mouse, and two chicken H1 genes (Yang et al, 1987; Kirsh et al, 1989). A human H2B cDNA with a similar 3' UTR architecture has also been reported (Collart et al, 1991). A probe derived from the 3' UTR detected a long, polyadenylated mRNA which was regulated in a similar manner to replication independent histone transcripts. It has not been established whether the gene which produced this transcript also produced a replication dependent poly(A)^+ transcript.

The existence of alternatively processed mRNAs with different modes of regulation illustrates the link between S-phase regulation and 3' mRNA processing (see below). By producing transcripts which are not rapidly degraded in the absence of DNA synthesis, variants such as H5, H1^0, H2A.F/Z and H3.3 are able to be expressed in non-dividing cell types.

### 1.9.3 Replication Dependent Expression: Posttranscriptional Regulation

#### 1.9.3 (a) Histone Transcript Turnover

In higher eukaryotes, the post transcriptional regulation of replication dependent histone mRNAs is effected by two main mechanisms (Schümperli, 1988; Osley, 1991). One mechanism affects the stability of the mature mRNA. If DNA synthesis is blocked during S-phase, the half lives of replication dependent histone mRNAs decrease dramatically, from
Chapter One: Introduction

-3-4 fold up to 20-fold (DeLisle et al, 1983; Sittman et al, 1983; Heintz et al, 1983; Heintz and Roeder, 1984; Baumbach et al, 1987). In studying histone transcript stability during the undisturbed CHO cell cycle, Harris et al (1991) found that transcripts isolated from G1-phase cells had similar half lives to those from S-phase cells. However, they concluded that the 35-fold decrease in transcript levels at the end of S-phase was primarily due to altered mRNA stability.

This selective degradation of histone mRNAs is conferred by the stem loop sequence at the 3' terminus. Deletion of the stem loop sequence and addition of polyadenylation signals to histone mRNAs led to an increase in their half-lives when DNA synthesis was inhibited (Levine et al, 1987). Conversely, chimeric α-globin transcripts ending in the histone 3' stem loop structure were regulated in a replication dependent manner (Pandey and Marzluff, 1987). The position of the stem loop was also shown to be an important factor in histone transcript stability. It appeared that the stem loop must coincide with the 3' terminus of the transcript to confer replication dependent regulation. If the 3' terminus was moved downstream by the addition of polyadenylation sequences 3' to the stem loop sequences, the stem loop did not destabilise histone mRNAs when DNA synthesis was blocked. This stabilisation of histone transcripts has been produced artificially (Altermen et al, 1985; Chodchoy et al, 1987; Levine et al, 1987) and also occurs in nature (see 1.9.2). In addition, Graves et al, (1987) demonstrated that the stem loop sequence must lie within ~300 bp of the translational termination codon to effect destabilisation.

1.9.3 (b) 3' Processing of Histone Primary Transcripts

The same stem loop structure required for the cell cycle regulated degradation of histone mRNAs is also a key element in the 3' processing of these transcripts (Birchmeier et al, 1982). In addition to the stem loop sequence, the purine rich sequence that lies ~15 bp downstream in the histone pre-mRNA is essential for correct processing (Birchmeier et al, 1983; 1984; Georgiev and Birnstiel, 1985; Mowry et al, 1989). A factor isolated from sea urchins which enhanced 3' processing of exogenous histone transcripts in frog oocytes (Birchmeier et al, 1984; Galli et al, 1983) was named U7-snRNP (Strub et al, 1984). The role of the U7-snRNA in 3' processing was confirmed when coexpression of the sea urchin
Chapter One: Introduction

H3 and U7-RNA genes in Xenopus oocytes led to correct processing of the H3 mRNA (Strub and Birnstiel, 1986). The U7-snRNA contains a complementary sequence to the purine rich element and evidence suggested that the two sequences must hybridise to promote correct 3' processing (Georgiev and Birnstiel, 1985; Schaufele et al, 1986; Cotten et al, 1988). Other factors implicated in 3' processing of histone pre-mRNAs include a stem loop-binding factor and a heat-labile factor which does not interact directly with the stem loop structure (see Schümperli, 1988; Osley, 1991).

The 3' mRNA processing event is also regulated, and is believed to act at the G1/S-phase border, which is about the same time as the activation of histone gene transcription (1.9.4). This was demonstrated by Schümperli and co-workers with a temperature sensitive mouse mastocytoma cell cycle mutant which could be arrested in G1-phase (Lüscher et al, 1985). At the restrictive temperature, mature H4 mRNA levels decreased by over 120-fold, but the transcription rate only decreased by 3-fold. This suggested that the histone genes were transcribed but inefficiently processed (Lüscher et al, 1985). In addition, H4 precursor mRNA, barely detectable in exponentially dividing cells, accumulated in G1-arrested cells. This was due to a deficiency in these cells of the previously mentioned heat-labile component (Lüscher and Schümperli, 1987). When fused to the SV40 early promoter, the 3' half of a mouse H4 gene was sufficient to direct correct 3' processing of the resultant transcript at the permissive temperature (Lüscher et al, 1985; Stauber et al, 1986). However, inhibition of DNA replication in these cells did not result in degradation of the fusion transcripts, confirming that the down-regulation of these transcripts in G1-arrested cells was due to a deficiency in 3' processing (Stauber and Schümperli, 1988).

Thus the mechanisms regulating 3' RNA processing and stability of replication dependent histone mRNAs, each acting at distinct stages of the cell cycle (the former in G1 and the latter at the end of S-phase), ensure that the cytoplasmic accumulation of mature histone transcripts is carefully restricted to the period in which DNA replication takes place.
1.9.4 Replication Dependent Expression: Transcriptional Regulation

Extensive research has been undertaken to define the promoter regions that give rise to the S-phase-specific increase in histone gene transcription. Comparison of histone promoter sequences has identified several motifs which are conserved in one or more histone classes from many species (Figure 1.1). Functional assays of histone promoter activity are often used to assess the effect of mutation of the element. Alternatively, the element can be placed within a heterologous promoter; this is often adopted to examine whether the element is sufficient to direct an S-phase-specific mode of expression. In another line of investigation, transcription factors which interact with functionally important elements are identified and purified, with the ultimate aim of cloning their genes.

These techniques have led to an improved, but still incomplete understanding of the S-phase-specific transcription of histone genes. The promoters of these genes are typical of the modular architecture of pol II-transcribed genes, containing several distinct sequence elements with additive contributions to the overall promoter activity (see Osley, 1991). Some S-phase-regulatory promoter elements have been identified or at least implicated. Contrary to post-transcriptional regulation, a universal mechanism for the coordinate S-phase regulation of histone gene transcription in higher eukaryotes has yet to be discovered. Surprisingly, each histone class appears to employ distinct means of S-phase regulation via class-specific promoter elements. To illustrate this point, the transcriptional regulation of each histone gene class will be described separately.

1.9.4 (a) H2A

The transcriptional regulation of the H2A gene is possibly the least understood. Consistent with their usual location in most histone genes, TATA- and CCAAT-boxes are present in the H2A promoters throughout higher eukaryotes. Between these two elements in sea urchin H2A promoters is a pentameric histone-specific element, 5' GATCC 3'. This element is found at the same position in a number of sea urchin histone genes from different classes (Hentschel and Birnstiel, 1981; Wells, 1986). However, it is not as highly conserved in vertebrates. Interestingly, the CCAAT-boxes of two late H2A genes (also from sea urchin) were found to be in the reverse orientation (Kemler and Busslinger, 1986).
Inverted CCAAT-boxes at similar locations have been found in several H2A promoters from different species (Perry et al, 1985). In some species such as chicken where the H2A promoter is closely linked to a divergently orientated H2B promoter, these inverted CCAAT-boxes may be involved with the expression of the latter promoter (Sturm et al, 1988).

A 26 bp region (-112 to -138) was demonstrated to be essential for maximal transcription of a sea urchin (Psammechinus miliaris) H2A gene injected into frog oocytes (Grosschedl and Birnstiel, 1980; 1982). This region, also known as a "modulator" was the first such region which, in either orientation, possessed enhancer activity (Grosschedl and Birnstiel, 1980). The modulator from another sea urchin species, Paracentrotus lividus, was shown to enhance transcription from a herpes simplex virus thymidine kinase (HSVtk) promoter, but not in the reverse orientation (Palla et al, 1989). This H2A-specific element is conserved in several species (Perry et al, 1985).

The means by which H2A transcription is S-phase-regulated is unknown. As already stated, H2A genes are often associated with H2B genes as divergently transcribed gene-pairs. If the promoters of these genes overlap, they may utilise common S-phase-regulatory elements (see below). However, this cannot be the case for many H2A genes which do not exist in this configuration.

1.9.4 (b) H2B

Mutational analyses of a human H2B promoter identified five elements (including the TATA- and CCAAT-boxes) which were required for maximal transcription (Sive et al, 1986). Deletion of a group of short direct repeats between -115 and -100 caused an ~2-fold decrease in expression as measured by in vitro transcription assays, as did mutation of a hexameric element, 5' GACTTC 3', located at -68. This hexamer has been found in the promoters of several human histone genes (Zhong et al, 1983). The other element identified as functionally important was a H2B class-specific element with the consensus 5' CCTTATTTGCATAAG 3', located a few bp upstream from the TATA-box. Unlike the histone-specific hexamer, this element could function independently of more distal upstream elements. Known as the H2B-box, this element was originally identified as a highly conserved element of H2B promoters throughout higher eukaryotes (Harvey et al, 1982).
Chapter One: Introduction

Figure 1.1 Topology of higher eukaryote histone promoters

Systematic diagram of the promoter architecture of the five major histone gene classes of higher eukaryotes. The numbers on the scale represent the 5' distance, in bp, from the cap site (+1). Boxes represent the position of sequence elements which are conserved across species and/or have been demonstrated to play a regulatory role in transcription from these promoters, and are referred to in the text. The grey box in the H1 histone promoter denotes the position of the CCAAT-box within the 18 bp H1-specific promoter element.
-200

-100

+1

ATG

H2A

modulator

CCAAT

GATCC

TATA

ATG

H2B

CCAAT

H2B-box

TATA

ATG

H3

CCAAT

CCAAT

GGTCC

TATA

ATG

H4

GC-box

CAAT

GGTCC

TATA

ATG

H1

Hexamer

AC-box

GC-box

CCAT

TATA
Deletion or mutation of the H2B-box also led to significant decreases in the levels of both H2A and H2B mRNAs from a divergently transcribed chicken H2A/H2B gene pair (Sturm et al, 1988).

A transcription factor purified from HeLa nuclear extracts, OTF-1, was shown to specifically activate expression from a truncated human H2B promoter which contained only the TATA- and H2B-boxes (Fletcher et al, 1987). This activation was abolished by mutation of the H2B-box. The first indication that the H2B-box was involved with the S-phase activation of H2B transcription came from in vitro transcription assays of H2B promoter constructs. Mutation of the H2B-box significantly reduced transcription in assays supplemented with extracts prepared from S-phase cells, but not with extracts prepared from G2-phase cells (Fletcher et al, 1987). In transient transfection experiments, mutation of the H2B-box in a H2B promoter/CAT fusion eliminated the S-phase-specific induction of transcription (La Bella et al, 1988). Although deletion of the CCAAT-box and all sequences upstream reduced overall promoter activity, this promoter (which still contained the H2B-box) was induced in S-phase to the same extent as the parental promoter.

When expressed in somatic cells, the testis-specific H2B (TH2B) promoter can direct S-phase-specific expression of cloned genes. Both the H2B-box (conserved in the TH2B promoter) and a CCAAT-box were required for this effect (Hwang and Chae, 1989).

Finally, Ito et al (1989) detected a factor in Chinese hamster fibroblasts which specifically bound the H2B-box in vitro. The DNA binding activity of this factor increased as the cells entered S-phase and decreased at the end of S-phase. This was in contrast to studies by La Bella and Heintz (1991), where it was shown that the DNA binding activity of OTF-1 remained constant throughout the cell cycle in WI38 and HeLa cell lines.

Interestingly, the core octanucleotide of the H2B-box, 5′ ATTTGCGAT 3′ (octamer element), or its reverse complement, are found in the promoters of a variety of otherwise unrelated genes. These include U-snRNA genes, immunoglobulin genes, viral genes, and even RNA polymerase III-transcribed genes. Correspondingly, OTF-1 has since been shown to have identity with NFIII and oct 1; these proteins were previously identified as binding octamer elements in the adenovirus origin of replication and the SV40 enhancer, respectively (see Heintz, 1991).
1.9.4 (c) H3

As with H2A, the pentameric GATCC element resides between the TATA- and CCAAT-boxes of sea urchin H3 promoters (Hentschel and Birnstiel, 1981), but it is not highly conserved in vertebrates. Peculiar to vertebrate H3 promoters, a second copy of the CCAAT-box is located ~20-40 bp upstream from the first. Although some H3 promoters contain putative Sp1 and AP1 binding sites, there appears to be no other promoter elements, H3-specific or otherwise, that are stringently conserved throughout vertebrate species.

Transcriptional control of H3 gene expression has been studied in a number of species. A 1 kb fragment containing the hamster H3.2 promoter was able to direct S-phase-specific expression of a neomycin resistance gene in vivo (Artishevsky et al, 1985). 5'-deletion analysis of this promoter revealed that removal of a 19 bp region between -200 and -180 diminished transcriptional activation in S-phase. This region contained the sequence 5' CGAGTCA 3', which is a near perfect AP1 consensus site. Deletion of a further 13 bp region to -167 decreased transcription by 10-fold and abolished S-phase-specific activation (Artishevsky et al, 1987). This second deleted region contained the sequence 5' GGCTGG 3', several copies of which were tandemly repeated at about -400. The AP1 site was shown to interact with two 42-45 kD proteins purified from hamster cells (Sharma et al, 1989). These proteins were antigenically related to the Jun oncoprotein family; c-Jun is a known component of AP1 (Johnson and McKnight, 1989). The presence of an AP1 site suggested a link between the activation of H3 gene transcription and control of cell proliferation.

A divergently transcribed H3 gene-pair from chicken has also been examined by deletional analysis. Although no cell cycle studies were done, the essential elements for transcriptional activity in transient assays were shown to be the proximal CCAAT-box of the H3-II promoter and an Sp1 binding site in the H3-III promoter (Takami and Nakayama, 1992). Only the H3-II promoter was reported to contain an AP1 site, located adjacent and 3' to the TATA-box. It was also found that the transcriptional activity of the unaltered H3-III promoter was ~20-fold higher than that of the H3-II promoter. It is perhaps puzzling that different elements were shown to be crucial to the expression of the hamster and chicken H3 genes.
Chapter One: Introduction

Using an in vivo analytical approach, Pauli et al (1989) mapped two sites (I and II) of protein-DNA interaction in a human H3 gene. Site I contained a sequence (5' GACGTCAG 3') which is similar to the AP1 site of the hamster H3 promoter. The distal CCAAT-box was centrally located in site II. Because these interactions persisted throughout the cell cycle, it was proposed that they were functionally related to the basal expression of this gene.

1.9.4 (d) H4

Unlike the other histone genes, the H4 promoter does not contain a conserved CCAAT-box, although the sequence 5' CAAT 3' is often found in H4 promoters from several species. At about -40, the sequence 5' GGTCC 3' is conserved throughout vertebrate species and to a lesser degree in sea urchin. This forms part of a H4 subtype-specific element (Wells, 1986). Potential Sp1 and NFkB binding sites are located further upstream at about -80. These elements were contained within a region (-110 to -70) shown to contribute to the maximal transcriptional activity in deletion analysis of the human H4a promoter (Hanly et al, 1985). The ability of this distal region to potentiate transcription was increased severalfold when the in vitro transcription assays were supplied with nuclear extracts prepared from S-phase cells. Interestingly, the conserved GGTCC element did not appear to be involved in this S-phase-specific stimulation of transcription.

Two factors were identified, H4TF-1 and H4TF-2, which interacted specifically with the C-rich element (an inverted repeat of the GC-box) in the distal region and the H4 subtype-specific element, respectively (Dailey et al, 1986). These factors were subsequently purified (Dailey et al, 1988). UV cross-linking experiments demonstrated that H4TF-1 consisted of two polypeptides of 110 and 105 kD. This was similar to the MW of the two protein species which constituted purified Sp1 (Briggs et al, 1986). In addition, the DNA binding activities of both factors were sensitive to chelating agents. However, competition experiments indicated that the DNA binding properties of H4TF-1 were distinct from those of Sp1. Purified H4TF-2 was a 65 kD protein which, like H4TF-1, required Zn\(^{2+}\) to bind DNA. Both purified factors were shown to stimulate transcription from templates containing
the wild type H4 promoter in vitro. The ability of each factor to stimulate transcription was abolished by deletion or mutation of the respective binding site (Dailey et al., 1988).

**In vivo** footprinting analysis and native genomic blotting of the human FO108 H4 promoter also identified two sites of protein-DNA interactions. Site I mapped between 124 and 85 bp upstream from the cap site, and site II mapped between 64 and 35 bp upstream from the cap site (Pauli et al., 1987; 1988). Both of these sites were occupied throughout the cell cycle in proliferating cells (Pauli et al., 1987). Three factors, HiNF-A, HiNF-C and HiNF-D, were shown to bind sequences within these sites in vitro. HiNF-C bound the consensus Sp1 site in the proximal half of site I, HiNF-A bound immediately upstream in the distal half of site I, and HiNF-D bound the H4 subtype-specific element within site II. As indicated in the *in vivo* experiments, these factors were bound to their cognate sites throughout the cell cycle (van Wijnen et al., 1987; 1989).

Site II appeared to have little effect on transcription from the H4 promoter, at least in proliferating cells. Although deletion of site I produced a 3-5 fold decrease in transcription, deletions introduced into site II had no significant effect (van Wijnen et al., 1989). However, when HL-60 cells were induced to differentiate, the shutdown of proliferation was accompanied by a loss of the site II protein-DNA interaction as well as sharp decreases in both HiNF-D DNA binding activity and transcription of H4 templates *in vitro* (Stein et al., 1989; van Wijnen et al., 1989).

Although HiNF-C and HiNF-D appear to be analogous to H4TF-1 and H4TF-2 in their binding specificities, the findings reached by the two groups varied. Unlike the conclusions of Dailey et al. (1988), van Wijnen and coworkers proposed that HiNF-C was similar or identical to Sp1. They also proposed that HiNF-D and H4TF-2 were distinct as they had different biochemical properties (van Wijnen et al., 1989).

An added complication was discovered in studies of primary rat calvarial osteoblasts and WI-38 human foetal lung fibroblasts. In contrast to tumor-derived cells, the DNA binding activity of HiNF-D was S-phase-regulated in extracts from both untransformed cell types (Holthuis et al., 1990). However, La Bella and Heintz (1991) established that the DNA binding activities of OTF-1, H4TF-1 and H4TF-2 were constant throughout the cell cycles of both WI-38 and HeLa cell lines. These inconsistencies were
addressed again by van Wijnen et al. (1992) when they demonstrated that site II was in fact a binding domain for at least three distinct factors: HiNF-D, HiNF-M and HiNF-P. The same region from the H4a gene (bound by H4TF-2) was bound by HiNF-P but not by HiNF-D nor HiNF-M. In combination with other similarities between HiNF-P and H4TF-2, this result suggested that these were analogous factors. Nonetheless, it would seem that the regulation of H4 gene transcription is complex.

1.9.4 (e) H1

As with the core histone promoters, the promoter of the H1 gene is made up of a series of conserved sequence elements, most of which contribute to the maximum transcriptional activity. H1 actually contains two class-specific elements. The CCAAT-box, located only ~25 bp upstream of the TATA-box, lies within an 18 bp H1-specific sequence, consensus 5'GCACCAATCACA/GCGCGC 3'. The more distal H1-specific element, 5' AAACACA 3', also known as the "H1-box" or "AC-box" (Coles and Wells, 1985), is usually found between 100 and 120 bp upstream from the cap site. The sequence of this element is highly conserved, and it is ubiquitous in the H1 promoters of sea urchins and vertebrates. Between the two H1-specific elements is a G-rich element, 5' GGGCGG 3' which has homology to the core Sp1 binding site. Finally, ~40-50 bp upstream from the AC-box is a histone-specific hexamer with identical or related homology to the sequence 5' GACTTC 3', and which is proposed to be analogous to the sea urchin pentameric element, 5' GATCC 3' (see Osley, 1991).

Initial studies of H1 transcription were carried out using a chicken gene, 01 H1, of which deletion mutants of the H1 promoter were assayed by micro-injection into Xenopus oocytes (Younghusband et al, 1986). Interestingly, deletion of the AC-box had no effect on transcription from the H1 promoter, while deletion of the G-rich box caused a 10-fold decrease in H1 mRNA levels. Similar results were obtained when these mutants were transiently transfected into HeLa cells. However, Dalton and Wells (1988a) demonstrated that the AC-box was essential for maximal transcription of the same H1 gene in stably transformed HeLa cells. Deletion or mutation of this element led to a dramatic (15-30-fold)
decrease in steady state H1 transcript levels. In addition, the S-phase-specific increase in H1 mRNA was abolished in synchronised cells.

Mobility shift assays of crude nuclear extracts prepared from either HeLa cells or ts34 cells (an AEV-transformed chicken erythroid cell line) demonstrated the presence of a trans-acting factor (H1-SF) which specifically bound the AC-box sequence (Dalton and Wells, 1988a). The DNA binding activity of this factor in synchronised ts34 cells increased as cells entered S-phase and decreased at the end of S-phase (Dalton and Wells, 1988b). Other experiments demonstrated that H1-SF was a protein of 90 kD under non-denaturing conditions (Dalton, 1987; Dalton and Wells, 1988b).

Using gel mobility shift assays, van Wijnen et al (1988a) detected two major DNA-protein complexes, A and B, which bound a promoter fragment from a human H1 histone gene, FNC16. Complex A had similar properties to a previously characterised factor, HiNF-A. It also bound a 53 bp fragment which contained only the AC-box and a region of short AT-rich repeats. Because HiNF-A had been previously shown to bind similar repeat sequences in the H3 and H4 promoters, it was concluded that this factor also bound these sequences in the H1 promoter. DNase I footprinting and modification interference analysis revealed that complex B, or HiNF-B, primarily bound the CCAAT-box, but also protected the region between the TATA-box and the CCAAT-box. This indicated that HiNF-B was distinct from previously characterised CCAAT-box-binding factors. The presence of the HiNF-A and HiNF-B DNA binding activities in extracts prepared from both S-phase and non S-phase HeLa cells was demonstrated by mobility shift assays. This led to the suggestion that these factors might be involved with the basal, cell cycle independent transcription of the H1 gene (van Wijnen et al, 1988b).

During purification of HiNF-B, partial loss in DNA binding activity coincided with the appearance of another DNA binding activity (complex B'). The DNA binding sites of these two factors were found to overlap. When separate fractions containing the two factors were combined, there was a dramatic increase in the intensity of complex B. These results indicated that HiNF-B was composed of at least two heterologous subunits (van Wijnen et al, 1988a; 1988b).
Mutational analysis of the promoter from another human H1 gene, Hh9, revealed that at least three conserved sequence elements contributed to overall transcriptional activity (Gallinari et al., 1989). In transient transfections of 293 cells, deletion of the AC-box produced a 50% decrease in transcription, while progressive deletions of the G-rich-box and CCAAT-box further reduced transcription (the latter deletion essentially abolished promoter activity). A point mutation in the CCAAT-box reduced the activity of the parental promoter fragment by 50%. The effect of the CCAAT-box mutation and the AC-box deletion on transcription throughout the cell cycle was investigated. The transcriptional activities of the wild type H1 promoter and the two mutant promoters were monitored by in vitro transcription assays supplemented with extracts prepared from HeLa cell populations at various stages of the cell cycle (La Bella et al., 1989). All three promoters were transcribed with equal efficiency in extracts from G1- and G2-phase cells, indicating that neither the CCAAT-box nor the AC-box were necessary for transcription in these phases. In S-phase extracts, transcription from all three promoters increased significantly, but the wild type H1 promoter was transcribed more efficiently than either of the mutant promoters. This indicated that both the H1 CCAAT-box and AC-box were involved with S-phase regulation of H1 transcription. Because each mutation led to a similar decrease in S-phase induction, it was suggested that the contribution of each element to S-phase regulation were approximately equal. The effect of the double mutation on S-phase activation from a single H1 promoter was not investigated.

Two factors, H1TF-1 and H1TF-2, were identified as specifically binding to AC-box and CCAAT-box sequences, respectively (Gallinari et al., 1989). H1TF-2 was subsequently purified and shown to be a 47 kD protein that was distinct from the NFI/CTF CCAAT-binding factor. Purified H1TF-2 stimulated transcription from the H1 promoter in vitro, while transcription from an H1 promoter with a mutated CCAAT-box was unaffected. In contrast to HiNF-B, the H1-CCAAT-box binding activity of H1TF-2 was 2-5-fold higher in extracts prepared from HeLa cells in S-phase than in G1-phase extracts (La Bella et al., 1989). This suggested that the S-phase specific stimulation of H1 promoter activity by H1TF-2 was at least partly due to an increase in its DNA binding activity. The AC-box binding activity of H1TF-1 was the same in both S-phase and G1-phase extracts. If H1TF-1
and H1-SF are analogous factors, then this result is contradictory to the result of Dalton and Wells (1988b).

1.10 The H1 Histone Genes of Chicken

Six H1 protein subtypes (a, a', b, c, c', and d) have been described in chicken (Shannon and Wells, 1987), and the genes encoding all six subtypes have been cloned and sequenced (Sugarman et al, 1983; Coles and Wells, 1985; Coles et al, 1987). The subtype encoded by each gene has also been determined. The relative proportion of H1 subtypes varies in different tissues. In AEV ts34 erythroblasts, both the steady state levels and the fold-increase upon entry into S-phase varied for each of the six H1 mRNAs (Tabe, King, Wells, unpublished results). The gene for the H1-d histone, 02 H1, was shown to be embryonically expressed, at least in red blood cells (Sugarman et al, 1983). In erythrocytes, H1-a (11L.H1 gene) is the most abundant subtype, whereas in other differentiated tissues H1-c (10 H1 gene) is the predominant subtype. Winter et al (1985a; 1985b) found that as chicken myoblasts differentiated into nondividing myotubes, the relative levels of H1-a, -b, and -d decreased considerably, whereas the relative level of H1-c increased. Furthermore, synthesis of H1-c was uncoupled from DNA replication. This uncoupling was demonstrated to occur at the level of mRNA accumulation. H1-c mRNA levels remained unchanged during myogenesis, and in myoblasts where DNA synthesis was inhibited. Because of its prevalence in differentiated tissue, H1-c was proposed to be analogous to mammalian H10 (Winter et al, 1985a).

The H1-c fraction has since been resolved into two subtypes, H1-c and H1-c', which are encoded by the 10 H1 and 01 H1 genes, respectively. The proportion of H1-c and H1-c' in erythrocytes is approximately equal (Shannon and Wells, 1987). In contrast to previous findings, these genes were shown to produce both the expected poly(A)- mRNA as well as unusually long, polyadenylated transcripts in non dividing tissue (Kirsh et al, 1989). The polyadenylated transcript levels were constant throughout G0- and S-phases, and the half life of these transcripts were longer than non polyadenylated transcripts from the same
genes. These genes were therefore classed as partially replication dependent variants. Interestingly, the 01 H1 mRNA has been demonstrated to exhibit S-phase transcriptional regulation in HeLa cells (Dalton and Wells, 1988a).

As illustrated above, the control of histone gene expression is complex: even genes from the same class in the same species appear to be regulated differently. In an effort to gain a better understanding of the transcriptional regulation of chicken H1 histone gene expression, the role of conserved promoter elements in these genes was investigated. The results of this study are presented in this thesis.
CHAPTER TWO

MATERIALS AND METHODS
Chapter Two: Materials and Methods

2.1 Abbreviations

Abbreviations are as described in 'Instructions to authors', Biochem. J., 273, 1-19, 1991.

In addition:

\[ A_{\text{xxx}} \]: absorbance at \( \text{xxx} \) nm
\[ \text{APS} \]: ammonium persulphate
\[ \text{BCIG} \]: 5-bromo-4-chloro-3-indolyl-\( \beta \)-D-galactoside
\[ \text{bisacrylamide} \]: N,N'-methylene-bisacrylamide
\[ \text{CAPS} \]: 3-(cyclohexylamino)-1-propanesulphonic acid
\[ \text{CsCl} \]: cesium chloride
\[ \text{DEPC} \]: diethylpyrocarbonate
\[ \text{DMEM} \]: Dulbecco's Modified Eagle's Medium
\[ \text{DMF} \]: dimethylformamide
\[ \text{DMSO} \]: dimethyl sulphoxide
\[ \text{DTT} \]: dithiothreitol
\[ \text{FCS} \]: foetal calf serum
\[ \text{IPTG} \]: isopropyl-\( \beta \)-D-thio-galactopyranoside
\[ \text{kD} \]: kilodaltons
\[ \text{NP-40} \]: Nonidet P-40
\[ \text{PBS} \]: phosphate buffered saline
\[ \text{PEG} \]: polyethylene glycol
\[ \text{PMSF} \]: phenylmethylsulfonyl fluoride
\[ \text{poly(dIdC)} \]: polydeoxyinosinic-deoxycytidylic acid
\[ \text{PVDF} \]: polyvinylene difluoride
\[ \text{RNase} \]: ribonuclease
\[ \text{SDS} \]: sodium dodecyl sulphate
\[ \text{rpm} \]: revolutions per minute
\[ \text{RSP} \]: reverse sequencing primer
\[ \text{TEMED} \]: N,N,N',N'-tetramethylethylenediamine
Chapter Two: Materials and Methods

TLC  thin layer chromatography
USP:  universal sequencing primer
UV:  ultraviolet

2.2 Materials

2.2.1 Chemicals, Reagents and Kits

All chemicals were of analytical reagent grade, or the highest available purity. Most chemicals and materials were obtained from a range of suppliers, the major source of the more important chemicals and reagents are listed below.

Acrylamide:  Sigma; Bio-Rad
Agarose; ampicillin; ATP:  Sigma
BCIG; bromophenol blue; BSA:  Sigma
Bisacrylamide:  Bio-Rad
Chicken serum:  Flow Laboratories
CsCl:  Boehringer Mannheim
DMEM:  Gibco
dNTPs; DTT:  Sigma
Ethidium bromide:  Sigma
F12:  Gibco
Foetal Calf Serum:  Commonwealth Serum Laboratories
Geneclean kit:  Bio-101
Glass beads (212-300 microns):  Sigma
Heparin-Sepharose CL-6B:  Pharmacia
Immobilon-P (PVDF):  Millipore
IPTG:  Sigma
Low gelling temperature agarose:  BRL; Sigma
Mixed bed resin AG 501-X8 (D):  Bio-rad
Chapter Two: Materials and Methods

Nitrocellulose: Schleicher and Schuell
Nonidet P-40: BDH Ltd.
Photobiotin: Bresatec
PMSF: Sigma; Boehringer Mannheim
Poly(dIdC): Sigma
Salmon sperm DNA; SDS: Sigma
Sepharose CL-6B: Pharmacia
Streptavidin Agarose: BRL
Sequencing kit: Bresatec
TEMED: Bio-rad
Urea (ultra-pure): Merck
Xylene cyanol: Sigma
Zetaprobe: Bio-rad

2.2.2 Enzymes

Enzymes were obtained from the following sources:

Calf intestinal phosphatase: Boehringer Mannheim

E.coli DNA-polymerase I,

Klenow fragment: Bresatec

Lysozyme: Sigma

Pfu DNA polymerase: Stratagene

Proteinase K: Boehringer Mannheim

Restriction endonucleases: Boehringer Mannheim; New England Biolabs;

Pharmacia; Bresatec

Ribonuclease A: Sigma

Sequenase: USB

T4 DNA ligase: Bresatec

T4 polynucleotide kinase: Boehringer Mannheim

T7 RNA polymerase: Bresatec

Trypsin: Sigma
2.2.3 Isotopically Labelled Compounds

\( \alpha-^{32}\text{P-dNTPs} \) (3000 Ci/m mole): Bresatec
\( \gamma-^{32}\text{P-ATP} \) (4000 Ci/m mole): Bresatec
\( \alpha-^{32}\text{P-UTP} \) (3000 Ci/m mole): Bresatec
\( \alpha-^{35}\text{S-dATP} \) (1500 Ci/m mole): Bresatec
\( ^{3}\text{H-Thymidine} \) (78 Ci/m mole): ICN Biomedicals, Inc.
\( ^{14}\text{C-Chloramphenicol} \) (57 mCi/mmol): Amersham

2.2.4 Bacterial Strains and Cell Lines

2.2.4 (a) \textit{E. coli} Strains

BB4: \( \text{sup}F58, \text{sup}E44, \text{hsd}R514, \text{gal}K2, \text{gal}T22, \text{trp}R55, \text{met}B1, \text{ton}A, \)
\( \Delta\text{lac}U169, \text{F'}[\text{pro}AB^+, \text{lac}I^\text{A}, \text{lac}Z\Delta M15, \text{Tn}10(\text{tet})] \) (Bullock et al., 1987)

DH5a: \( \text{sup}E44, \Delta\text{lac}U169 (\phi80 \text{lacZ\Delta M15}), \text{hsd}R17, \text{rec}A1, \text{end}A1, \text{gyr}A96 \)
\( \text{thi}-1, \text{rel}A1 \) (Sambrook et al., 1989; gift from Dr. H. Richardson, Department of Biochemistry, University of Adelaide)

ED8799: \( \text{hsd}S, \text{met}B7, \text{sup}E, (\text{gln}V)44, \text{sup}F, (\text{tyr}T)58, \Delta(lacZ)M15, \text{r}_{\text{k}}^{-}, \text{m}_{\text{k}}^{-} \)
(gift from Dr. S. Clarke, Biotechnology Australia)

JM101: \( \Delta(lac-pro \text{ AB}), \text{sup}E, \text{thi} , \text{F'}[\text{lac}I^\text{A} lacZ\Delta M15, \text{tra}D36, \text{pro}AB^+] \)
(Messing, 1979)

MC1061: \( \text{ara}D139, \Delta(\text{ara, leu })7697, \Delta\text{lac}X74, \text{gal}U , \text{gal}K , \text{hsr}^{-}, \text{hsm}^{+}, \text{str}A \)
(Casadaban and Cohen, 1980)

2.2.4 (b) Tissue Culture Cell Lines

CHO-K1: Chinese hamster ovary cell line (ATCC CCL 61)
CEF-38: Chicken finite fibroblast line
HD2 (ts34): AEV-transformed chicken erythroid cell line (Beug \textit{et al}, 1982)
HD11: MC29-transformed chicken macrophage cell line (Beug \textit{et al}, 1979)
Chapter Two: Materials and Methods

HeLa: Human cervix epitheloid carcinoma cell line (ATCC No. CCL 2)
LMH: Chicken hepatoma cell line (J. Virology (1990) 64, 3249-3258)

### 2.2.5 Media and Buffers

#### 2.2.5 (a) Media

All buffers and media were prepared with distilled and deionised water and sterilised by autoclaving, except heat labile reagents, which were filter sterilised.

All bacteria, except *E. coli* JM101, were grown in L-broth or on L-agar plates.

(*E. coli* JM101 was grown in minimal medium, 2 x YT broth and on minimal plus glucose plates.)

- **L-broth:** 1% (w/v) bacto-tryptone / 0.5% (w/v) bacto-yeast extract / 1% (w/v) NaCl, pH 7.0
- **L-agar plates:** 1.5% (w/v) bacto-agar in L-broth
- **Minimal medium:** 2.1% (w/v) K2HPO4 / 0.9% (w/v) KH2PO4 / 0.2% (w/v) (NH4)2SO4 / 0.1% (w/v) tri-sodium citrate
- **2 x YT broth:** 1.6% (w/v) bacto-tryptone / 1% (w/v) bacto-yeast extract / 0.5% (w/v) NaCl, pH 7.0

Ampicillin was added where required for selection at a concentration of 50 or 100 mg/ml.

Tissue culture cells were maintained at 37°C in an atmosphere of 5% CO2.

CHO-K1 cells were grown in Ham's F12 medium supplemented with 20 mM HEPES, 50 μM β-mercaptoethanol, sodium bicarbonate (2.4 g / litre), gentamycin (50 mg / litre), and 10% (v/v) FCS.

HeLa cells were grown in DMEM medium supplemented with 20 mM HEPES, 50 μM β-mercaptoethanol, sodium bicarbonate (2.4 g / litre), gentamycin (50 mg / litre), and 10% (v/v) FCS.
ts34, HD11, CEF-38, and LMH cells were maintained in the same medium, except that 2% (v/v) chicken serum was added.

2.2.5 (b) Buffers

Commonly used buffers were:

- **HeBS**: 20 mM HEPES pH 7.05 / 137 mM NaCl / 5 mM KCl
- **PBS**: 7.5 mM Na₂HPO₄ / 2.5 mM NaH₂PO₄·2H₂O / 145 mM NaCl
- **TBE**: 50 mM Tris-borate pH 8.3 / 1 mM EDTA
- **TE**: 10 mM Tris-HCl pH 7.4 / 1 mM EDTA
- **TAE**: 40 mM Tris-acetate pH 8.2 / 1 mM EDTA
- **Tris-Glycine buffer**: 40 mM Tris-HCl / 0.6 M Glycine pH 8.5
- **Urea load buffer**: 4 M Urea / 50% (w/v) sucrose / 50 mM EDTA / 0.1% (w/v) Bromophenol blue
- **Acrylamide load buffer**: 50% (w/v) sucrose / 5 mM EDTA / 50 mM Tris-HCl pH 7.4 / 0.1% (w/v) Bromophenol blue / 0.1% (w/v) Xylene cyanol
- **Tris Glycine load buffer**: 120 mM Tris-Glycine pH 8.5 / 20% (v/v) glycerol / 0.1% (w/v) Bromophenol blue
- **SDS load buffer**: 10% (v/v) glycerol / 375 mM Tris-HCl pH 8.8 / 5% (w/v) SDS / 0.1% (w/v) Bromophenol blue / 5% (v/v) β-mercaptoethanol
- **Formamide load buffer**: 80% (v/v) formamide / 250 mM EDTA / 0.1% (w/v) Bromophenol blue / 0.1% (w/v) Xylene cyanol
- **TM Buffer**: 50 mM Tris-HCl pH 7.9 / 12.5 mM MgCl₂ / 1 mM EDTA / 20% (v/v) glycerol / 1 mM DTT / 0.5 mM PMSF

2.2.6 Clones and Vectors

2.2.6 (a) Cloned DNA Sequences

- **pgKNeopA**: gift from Varaporn Thongolaim
- **pCH4.7E**: gift from Leeanne Coles

63
2.2.6 (b) Cloning Vectors

pBluescript SK+: gift from Dr. A. Robins
pUC19: gift from Dr. A. Robins (Yanisch-Perron et al., 1985)

2.2.6 (c) Reporter Vectors

pTKGH: gift from Dr. A. Robins
pBLCAT.2; pBLCAT.3: gift from Dr. C. Harn
pRSVCAT: gift from Dr. C. Ham

2.2.7 Molecular Weight Markers

2.2.7 (a) DNA Markers

DMW-S1: Eco RI digested SPP1 phage (380 bp - 7.84 kb): Bresatec
DMW-P1: Hpa II digested pUC19 (26 - 502 bp): Bresatec

2.2.7 (b) Protein Markers

MW - SDS - 200: high molecular weight (29000 - 205000): Sigma

2.2.8 Oligonucleotides

All oligonucleotides were synthesized by Bresatec using an Applied Biosystems Model 380B DNA synthesizer. The sequence of the oligonucleotides used in this work are as follows:

H5E-t: 5' dTCGAGAGCCAGGAGGAGGAGGACTCCTCCTTGTCCATAGGAGTGAGGCACACCCG 3'
H5E-b: 5' dGATCCGGGCTGTCCCTCACTGACAAGGACTGAGTCCCTCTCCTGCTCCTCGGTCC 3'
01-40-t: 5' dAATTCCTTTGTAGTCCAAAGAAACACAAATCGAGCACAG 3'
01-40-b: 5' dGATCCCTGTGCATTCTTGTGTTTCCTTGGACTAAACAAAG 3'
Δ01-40-t: 5' dATTCTTTTGTAGTCCAAAGGACTACGAATCGAGCACAG 3'
01-40-b: 5’ dGATCCCTGTGCTGATTCGATCTCCTTTGGGACTAACAAAAAG 3’
01-14-t: 5’ dAAGAAACACAAAGCT 3’
01-14-b: 5’ dGTGTGTTTCTTATGCT 3’
01-25-t: 5’ dGTCCAAAGAAACACAAATCGAGCAC 3’
01-25-b: 5’ dGTTCCTGATTGTGTTGTTTCTTGGGAC 3’
Δ01-25-t: 5’ dGTCCAAAGATCTCTCTCAATCGAGC 3’
Δ01-25-b: 5’ dGTTCCTGATTGAGATTTGGGAC 3’
02-25-t: 5’ dGCTTGTAAAGACACACAAATAGGC 3’
02-25-b: 5’ dGCTTATTTTCTTCTTCAACAGGC 3’
Δ02-25-t: 5’ dGCTTGTAAAGATCTCTCTCAATAGGC 3’
Δ02-25-b: 5’ dGCTTATTTTCTTAACTACAGGC 3’
01-G-rich-t: 5’ dTGCAAGGGGGGGGGTCCTAGGC 3’
01-G-rich-b: 5’ dGTTCATTTTCTTCTTGGTAAAAGGC 3’
01-CCAAT-t: 5’ dCAACCGAATACACCGGCG 3’
01-CCAAT-b: 5’ dCGGAGATGATTTGGCCGTTG 3’
Drosophila HSE-t: 5’ dGTCCAGCGGCGCCCTCGAATGTTCTAGGAAAAG 3’
Drosophila HSE-b: 5’ dGTCCCTTTTCTAGAACATTCGAGGCGCGTCG 3’
5’-01L: 5’ dGGTGGAGATTCAGCTGAGGCG 3’
5’-01S: 5’ dGCTTCATTCCCTCTCTCCTCAAATCA 3’
3’-01: 5’ dGAGATGTCGACGGGTGCTCGATTTTG 3’
5’-02L: 5’ dGGTGGAGATTCAGCTGAGGCG 3’
5’-02S: 5’ dGCTTCATTCCCTCTCTCCTCCTAAATCG 3’
3’-02: 5’ dGAGATGTCGACTCCCGCTCTTTTG 3’
RSP: 5’ dAACAGCATTGACCAGT 3’
USP: 5’ dGTAAACGACGGGAGT 3’
Sequencing primer (-40): 5’ dGTTTCCAGGATCAGGAC 3’
T3 primer: 5’ dATTAACCCCTCCTAAAG 3’
T7 primer: 5’ dATACGAGCTATATAG 3’
2.3 Methods

2.3.1 Isolation of Plasmid DNA

2.3.1 (a) General Procedure

50 ml of L-broth with the appropriate antibiotic was inoculated with either a single bacterial colony or an overnight culture and grown overnight at 37°C with aeration. The cells were harvested by centrifugation at 5000 rpm (HB-4 rotor) for 5 minutes and resuspended in 2 ml of lysozyme buffer (15% (w/v) sucrose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA, and 4 mg/ml lysozyme). The suspension was kept on ice while 4 ml 0.2 M NaOH / 1% (w/v) SDS was added, followed by 2.5 ml 3 M sodium acetate pH 4.6. The lysate was kept on ice for 30 minutes with occasional gentle mixing. Chromosomal DNA and cellular debris were removed by centrifugation at 10000 rpm (HB-4 rotor) for 15 minutes. The supernatant was retained, and the plasmid DNA precipitated by adding 0.6 volumes of ice cold isopropanol. After a 15 minute incubation on ice, the sample was centrifuged at 10000 rpm (HB-4 rotor) for 10 minutes.

The precipitate was dissolved in 2 ml water and mixed with 4 ml of 5 M LiCl. The mixture was incubated on ice for 15 minutes, followed by centrifugation for 5 minutes at 12000 rpm to pellet the RNA. The supernatant was transferred to a fresh 50 ml oakridge tube. Plasmid DNA was precipitated by the addition of 2.5 volumes of ethanol followed by incubation for 30 minutes on ice. After centrifugation at 12000 rpm for 5 minutes, the DNA pellet was resuspended in 300 μl of TE buffer and transferred to an eppendorf tube. 2 μl of RNAse A (10 mg/ml) was added and the sample was incubated at 37°C for 30 minutes. 8 μl of 10% (w/v) SDS and 2.5 μl of Proteinase K (20 mg/ml) were then added and the sample incubated for a further 15 minutes at 37°C. The sample was extracted twice with an equal volume of a 1:1 phenol : chloroform mixture, and then once with an equal volume of chloroform. The aqueous phase was mixed with 0.25 volumes of 7 M ammonium acetate and 2.5 volumes of ethanol. The DNA was precipitated, washed once with 70% (v/v) ethanol, and resuspended in an appropriate volume of water.
2.3.1 (b) Miniscreen Procedure

(i) Alkaline extraction

A single bacterial colony was picked into 2 ml of L-broth plus the appropriate antibiotic and grown at 37°C with aeration for 6 hours or overnight. 1.5 ml of the culture was transferred to an eppendorf tube and the cells were harvested by centrifugation for 1 minute in an eppendorf microfuge. The cellular pellet was resuspended in 100 μl of lysozyme buffer (2.3.1 (a)). The suspension was kept on ice while 200 μl of 0.2 M NaOH / 1% (w/v) SDS was added, followed by 125 μl of 3M sodium acetate pH 4.6. After mixing gently, the suspension was kept on ice for 5-10 minutes. Chromosomal DNA and cellular debris were precipitated by centrifugation for 5 minutes in an eppendorf microfuge. The supernatant was transferred to a fresh eppendorf tube, and 2 μl of 10 mg/ml RNAse A was added. After incubation at 37°C for 20 minutes, the sample was extracted twice with phenol : chloroform. Plasmid DNA was precipitated by the addition of 2.5 volumes of ethanol, followed by incubation at -20°C for at least one hour. After centrifugation for 10 minutes at 4°C, the precipitated DNA was rinsed in 70% ethanol and briefly air-dried before resuspension in 20 μl of water.

(ii) STET preps

Alternatively, the cellular pellet from 1.5 ml of overnight culture was resuspended in 200 μl of STET buffer (8% (w/v) sucrose, 5% (v/v) Triton X-100, 50 mM Tris-HCl pH 8.0, and 50 mM EDTA). 20 μl of 5 mg/ml lysozyme was added and the mixture was heated to 100°C for 45 seconds. The cell debris was precipitated by centrifugation for 15 minutes in a microfuge and removed from the sample using a sterile toothpick. The supernatant was centrifuged for 5 minutes to remove any remaining debris. Plasmid DNA was precipitated by the addition of 200 μl of isopropanol. After 10 minutes on ice, the sample was centrifuged for 10 minutes and the DNA was washed twice in 70% ethanol. The precipitated DNA was briefly air-dried before resuspension in 20 μl of water.
2.3.1 (c) Large Scale Purification on a CsCl Gradient

A single bacterial colony was used to inoculate 100 ml of L broth with the appropriate antibiotic and the culture was grown at 37°C with aeration overnight. Plasmid DNA to be purified on CsCl gradients was prepared as in (2.3.1 (a)) up to and including the isopropanol precipitation. The DNA pellet was resuspended in 1.4 ml of water, and added to 1.5 g CsCl in a 10 ml polycarbonate tube. After the CsCl had dissolved, 120 μl of 10 mg/ml EtBr was added and the mixture was transferred in a Beckman 2.2 ml heat sealable tube. The sample was then centrifuged at 80000 rpm for 16 hours in a Beckman TL-100 ultracentrifuge.

The lower band in the CsCl gradient containing supercoiled plasmid DNA was removed from the tube using a 26-gauge needle and syringe. To remove EtBr, the DNA solution was extracted with an equal volume of water-saturated butan-1-ol. This step was repeated until the aqueous phase was colourless, at which stage it was made up to twice the original volume with water. Plasmid DNA was precipitated with 2.5 volumes of ethanol, pelleted by centrifugation and rinsed twice in 70% ethanol. The precipitate from the 70% ethanol wash was resuspended in an appropriate volume of water, and stored at -20°C.

The amount of DNA recovered was quantitated by measuring the \(A_{260}\) on a Shimadzu UV-160A Spectrophotometer. This was confirmed by electrophoresis of a DNA sample alongside a known amount of control DNA in an agarose gel.

2.3.2 Restriction Enzyme Digestion and Analysis of DNA

2.3.2 (a) Restriction Enzyme Digestion

Restriction endonuclease digestions were performed using the conditions for each enzyme as recommended by the supplier. Analytical digests were in 20 μl reactions containing 200-500 ng of DNA, and a five-fold excess of enzyme for 1 hour. Preparative digests were in 50-100 μl containing 10-20 μg DNA.

Analytical digests were stopped by adding a quarter volume of urea loading buffer and loaded onto either an agarose or polyacrylamide gel. Alternatively, the reaction was extracted once with an equal volume of phenol : chloroform. DNA was ethanol precipitated
from the aqueous phase by adjusting to 0.3 M sodium acetate, pH 4.6, and adding 2.5 volumes (preparative digests) or 4 volumes (analytical digests) of cold, nuclease-free ethanol. After incubation at -20°C for at least 30 minutes, the DNA was pelleted by centrifugation for 10 minutes in an Eppendorf bench centrifuge. The DNA pellet was washed with 70% (v/v) (preparative digests) or 80% (v/v) (analytical digests) nuclease-free ethanol, and briefly air-dried before being resuspended in an appropriate volume of water.

2.3.2 (b) Agarose Gel Electrophoresis

Agarose was dissolved in 1xTAE buffer to the concentration required (0.7% - 3.0%), and poured onto 7.5 cm x 5 cm microscope slides, to be used as horizontal gels. The gels were submerged in 1xTAE.

DNA samples in a volume of ~5 µl were mixed with 2 µl of urea loading buffer and loaded into the wells. A current of 80-110 mA was applied. After electrophoresis, the DNA was visualised by staining in 0.02% ethidium bromide and exposure to UV light.

2.3.2 (c) Polyacrylamide Gel Electrophoresis

Electrophoresis of DNA species of less than 500 bp in length was carried out on vertical 14 cm x 14 cm x 0.5 mm slab gels containing 8-20 % (w/v) acrylamide : bisacrylamide (40:1), which had been deionised with mixed bed resin.

The acrylamide was polymerised in 1xTBE buffer by the addition of 0.1% (w/v) APS and 0.1% (v/v) TEMED. The gel was pre-electrophoresed for 10-45 minutes at 400 V prior to loading. Samples were mixed with a quarter volume of acrylamide load buffer before loading into the wells.

All non denaturing polyacrylamide gels were electrophoresed at 200-500 V until the bromophenol blue had migrated the desired distance. DNA labelled with a radioactive isotope was visualised by autoradiography (2.3.17). Unlabelled DNA was visualised by "UV shadowing": the gel was placed over a TLC plate and illuminated from above with UV light. The position of DNA in the gel was indicated by a shadow cast onto the TLC plate directly underneath the DNA.
2.3.2 (d) Denaturing polyacrylamide gels

For DNA sequencing (2.3.15), 6-10% (w/v) polyacrylamide / 8 M Urea gels containing 1xTBE buffer were required. 420 x 350 x 0.25 mm slab gels were usually cast. The gels were electrophoresed in 1xTBE buffer at 2000 V or 45 W. Pre-electrophoresis was maintained for 45 minutes before loading.

After electrophoresis, gels were fixed with 200 ml of 10% (v/v) acetic acid, and washed with 4 litres of 20% (v/v) aqueous ethanol. The fixed gels were dried in a 100°C oven for 30 minutes and allowed to cool to room temperature before autoradiography (2.3.17). Alternatively, the gels were transferred to a sheet of Whatman 3MM, overlayed with a sheet of plastic wrap, and dried at 65°C for 30 minutes in a Hoefer Scientific Instruments Drygel Sr. Slab gel drier (SE 1160).

2.3.3 Subcloning of DNA fragments into plasmid vectors

2.3.3 (a) Isolation of DNA from agarose gels

(i) Geneclean

DNA fragments >500 bp were isolated from agarose gels using the protocols and solutions provided by Bio-101, and based on the method of Vogelstein and Gillespie (1979). The agarose containing the required fragment was placed in an eppendorf tube, 2-3 volumes of 6 M sodium iodide solution added, and the agarose dissolved by incubating at 50°C for 5 minutes. A 5 µl aliquot of Glassmilk solution (containing a silica matrix suspension) was added and incubated at room temperature for at least 5 minutes to allow the DNA to bind. The DNA / Glassmilk complex was pelleted by centrifugation for 5 seconds, and the supernatant discarded. The pellet was washed three times with 500 µl of a wash buffer containing NaCl and ethanol to remove all traces of sodium iodide.

The final pellet was resuspended in 5-10 µl of water, and incubated at 50°C for 5 minutes. After centrifugation for 20 seconds, the supernatant containing the DNA was transferred to another tube, and the elution step repeated with another 5-10 µl of water. The two eluates were pooled and centrifuged for 20 seconds to remove any traces of Glassmilk.
(ii) Spin-columns

DNA fragments smaller than ~500 bp were recovered from agarose gel slices by centrifugation. The bottom of a 0.5 ml PCR eppendorf tube was pierced with a 21-gauge needle. The hole was covered with ~20 µl of acid-washed glass beads (212-300 microns) and the agarose slice containing the fragment to be purified was placed on the beads. The PCR tube was then placed inside a 1.5 ml eppendorf tube and the assembly was centrifuged for 5 minutes at 6000 rpm in an Eppendorf 5415 C bench centrifuge. The PCR tube and its contents were discarded and the eluate from the gel slice, which had collected in the 1.5 ml eppendorf tube, was extracted once with phenol : chloroform. The aqueous phase containing the DNA was adjusted to 0.3 M sodium acetate, pH 4.6, and 4 volumes of ethanol was added. After incubation at -20°C, the DNA was pelleted by centrifugation for 10 minutes in an eppendorf microfuge and rinsed in 80% ethanol. The DNA pellet was briefly air-dried before resuspension in an appropriate volume of water.

2.3.3 (b) Isolation of DNA from low gelling temperature agarose gels

The smallest slice of agarose possible, containing the DNA fragment, was placed in an eppendorf tube with an equal volume of TE buffer, and was melted at 65°C. The mixture was extracted twice with phenol, and once with phenol : chloroform before the DNA was ethanol precipitated. After rinsing in 70-80% ethanol, the fragment was resuspended in an appropriate volume of water.

2.3.3 (c) Preparation of Vector

Vector DNA was linearised with the required restriction enzyme(s), extracted once with phenol : chloroform, ethanol precipitated, and resuspended in 10 µl of water. If necessary, 1 µl of 10x dephosphorylation buffer and 1 unit of calf intestinal phosphatase was added. For sticky ends, the reaction was incubated for 30 minutes at 37°C, while for blunt ends, a 15-minute incubation at 37°C was followed by the addition of fresh enzyme and incubation at 55°C for a further 45 minutes. The reaction was extracted once with phenol : chloroform, ethanol precipitated, and resuspended in 10 µl of water. The linearised,
dephosphorylated vector was purified from uncut vector by electrophoresis on an agarose gel as described above.

2.3.3 (d) Ligation into plasmid vectors

20-50 ng of purified vector DNA was combined with a 1-3 fold molar excess of purified insert DNA in a 20 µl reaction volume containing 50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 1 mM DTT, 0.5 mM ATP, and 0.1 units (sticky ends) to 0.5 units (blunt ends) of T4 DNA ligase. Alternatively, 2 µl of 10xOne-Phor-All buffer (Pharmacia) was included instead of the ligation buffer. Ligations were usually performed at room temperature for 1 hour.

When the DNA fragment had incompatible protruding 5' or 3' termini, it was treated with the Klenow fragment of DNA polymerase I to fill-in the single-stranded ends. This was carried out prior to purification of the fragment by agarose gel electrophoresis in a 20 µl reaction volume containing 10 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM MgCl₂, 10 mM DTT, 0.1 mM of each dNTP, and 5-10 units of *E.coli* DNA-polymerase I, Klenow fragment.

2.3.3 (e) Transformation of recombinants into bacteria

0.5 ml of an overnight culture of *E.coli* MC1061, *E.coli* DH5α, or *E.coli* ED8799, was diluted 100-fold in L-broth, and grown with aeration at 37°C to an A₆₀₀ of 0.6-0.8. The cells were pelleted by centrifugation (HB-4 rotor, 5000 rpm for 5 minutes), washed in 1/2 volume of ice-cold 0.1 M MgCl₂, and centrifuged as before. The pellet was resuspended in 1/20th volume of ice-cold 0.1 M CaCl₂, and left on ice for at least 1 hour before use.

Usually 1/10th of the ligation reaction was combined with 200 µl of competent cells and left on ice for 30 minutes. The transformation mix was then heat shocked at 42°C for 2 minutes, and returned to ice for 30 minutes. After slowly warming to room temperature, 0.5 ml of L-broth was added, and the cells incubated at 37°C for 20 minutes. The transformation mix was transferred to an eppendorf tube and briefly centrifuged to pellet the cells. Most of the supernatant was removed and the cells resuspended in the final few µl

72
before spreading them directly onto an L-agar plate containing the appropriate antibiotic(s). The plate was incubated at 37°C overnight.

When colour selection was required with *E.coli* strain ED8799, 60 μl of BCIG (20 mg/ml in DMF), and 60 μl of IPTG (10 mg/ml) were spread with the cells directly onto the L-agar plate.

2.3.3 (f) Plasmid miniscreen procedure

Transformant colonies were each picked into 2 ml of L-broth with the appropriate antibiotic and grown at 37°C overnight with continual aeration. Plasmid DNA was recovered by either the alkaline extraction or the STET procedure (2.3.1 (b) (i)). The DNA prepared in this way was pure enough to be cut with restriction enzymes.

2.3.4 Purification of oligonucleotides

10 μg of oligonucleotide was mixed with acrylamide loading buffer and loaded into a 1 cm well of a 20% (w/v) polyacrylamide gel. After electrophoresis, the oligonucleotide was visualized by UV-shadowing (2.3.2 (c)) and excised from the gel. The gel slice containing the oligonucleotide was placed into a 1.5 ml eppendorf tube and 400 μl of TE buffer was added. The oligonucleotide was eluted from the gel slice by incubation at 37°C overnight.

To remove acrylamide fragments, the eluate was filtered through a Sartorius Ministart NML 0.2μm filter. The filtrate was adjusted to 0.3 M sodium acetate, pH 4.6, and the oligonucleotide was precipitated by the addition of 4 volumes of ethanol. The precipitate was rinsed in 80% ethanol and briefly air dried before resuspension in 20 μl of water. 1 μl of the sample was quantitated by measuring the A260 on a Shimadzu UV-160A spectrophotometer.

2.3.5 Labelling of DNA fragments

2.3.5 (a) Kinasing of synthetic oligonucleotides

Oligonucleotides were 5' end-labelled with T4 polynucleotide kinase and γ-32P-ATP. Normally 100 ng of oligonucleotide was kinased in a 10 μl reaction containing 60 mM
Chapter Two: Materials and Methods

Tris-HCl pH 7.5, 10 mM MgCl₂, 10 mM DTT, 50 μCi of γ-³²P-ATP, and 3 units of enzyme. The reaction was incubated at 37°C for 1 hour. 2 μl of acrylamide loading buffer was added and the mixture loaded onto a 20% (w/v) polyacrylamide gel for purification. The labelled oligonucleotide was visualized by autoradiography (2.3.17), and excised from the gel. The gel slice containing labelled oligonucleotide was placed into a 1.5 ml eppendorf tube and 400 μl of TE buffer was added. The oligonucleotide was eluted from the gel slice by incubation at 37°C overnight.

2.3.5 (b) Endfilling of restriction fragments

500 ng of DNA fragments with 5' overhangs was incubated in a 10 μl reaction mixture containing 10 mM Tris-HCl pH 7.4, 10 mM MgCl₂, 30 mM NaCl, 100 mM of each dNTP (excluding the appropriate labelled dNTP(s)), and 50 μCi of the appropriate α-³²P-dNTP (dATP, dCTP, or both). One unit of E.coli DNA polymerase I, Klenow fragment was added to the reaction, and the mixture incubated at 37°C for 5 minutes. The reaction was terminated by phenol: chloroform extraction. Labelled DNA was ethanol precipitated and resuspended in 10 μl of water.

2.3.6 Manipulation and Harvesting of tissue culture cells

Approximately 16 hours prior to experimental manipulation, cells were seeded in fresh media at ~5 x 10⁵ cells per ml for suspension cultures, and 30-40% confluence for attached cells in order to ensure that they were highly mitotic, and in an exponential phase of growth.

Suspension cultures were harvested by centrifugation (Hettich Universal centrifuge, 1500 rpm, 5 minutes) and the cell pellet was washed once by resuspension in ~100 volumes of PBS. The cells were then pelleted and resuspended in the appropriate buffer for manipulation.

Attached cell lines were collected by initially washing in 15-20 ml PBS. After decanting, cells were detached by a brief treatment with 2 ml of trypsin (1 mg/ml dissolved in 1x versene solution). The detached cells were resuspended in at least an equal volume of culture medium, and harvested by two sequential steps involving centrifugation, decanting.
2.3.7 Nuclear extract preparation

2.3.7 (a) Tissue culture cells

Nuclear extracts were prepared from tissue culture cells using a variety of methods, which are described below:

(i) From Strauss and Varshavsky (1984)

Approximately $10^9$ cells were harvested, washed twice in PBS (2.3.6), and pelleted by centrifugation (JA-10 rotor, 3000 rpm for five minutes). All subsequent procedures were performed at 4°C to minimise proteolytic degradation. The pellet was resuspended in ~3 volumes of 0.25 M sucrose in Buffer A (60 mM KCl, 15 mM NaCl, 0.25 mM MgCl$_2$, 0.5 mM Na-EGTA, 0.15 mM spermidine, 0.5 mM spermine, 15 mM Tris-HCl, pH 7.4, 14 mM β-mercaptoethanol, and 0.5 mM PMSF), and homogenised with fifteen strokes in a Dounce homogeniser. The cell lysate was diluted 3-fold with 2.0 M sucrose in Buffer A, and layered onto a 9 ml cushion of 1.7 M sucrose in Buffer A. The nuclei were pelleted by centrifugation (HB-4 rotor, 12000 rpm for 45 minutes) through the sucrose cushion. The nuclear pellet was resuspended in 20 ml of 0.25 M sucrose in Buffer A and centrifuged at 7000 rpm for 5 minutes. At this stage, the pellet could be snap frozen in liquid nitrogen; otherwise it was resuspended to a final DNA concentration of 0.5 mg/ml in 0.4 M NaCl, 5 mM EDTA, 10 mM HEPES-NaOH, pH 7.5, 14 mM β-mercaptoethanol, and 0.5 mM PMSF, and incubated on ice for 1 hour, with occasional gentle mixing. The suspension was transferred to eppendorf tubes, and centrifuged for 10 minutes at 4°C. Glycerol was added to the supernatant to a final concentration of 20% (v/v). Aliquots of the crude nuclear extract were snap frozen in liquid nitrogen and stored at -80°C.

(ii) From Lee et al (1988)

Between $10^7$ and $10^8$ cells were harvested and washed in PBS (2.3.6), and then pelleted by centrifugation at 1000 rpm for 5 minutes (Hettich Universal centrifuge). All
subsequent procedures were performed at 4°C to minimise proteolytic degradation. The cell pellet was resuspended in 0.5 ml of Buffer A (10 mM HEPES-NaOH, pH 7.9, 10 mM KCl, 1.5 mM MgCl₂, 1 mM DTT, and 0.5 mM PMSF) and incubated on ice for 15 minutes. The suspension was slowly drawn into a 1 ml syringe, and expelled rapidly through a 25-gauge needle. This step was repeated 12 times; light microscopy revealed that this was sufficient to lyse >95% of the cells. The lysate was centrifuged for 20 seconds in an eppendorf microfuge, and the nuclear pellet was resuspended in 100 μl of Buffer C (20 mM HEPES-NaOH, pH 7.9, 420 mM NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 25% (v/v) glycerol, 1 mM DTT, and 0.5 mM PMSF). The suspension was kept on ice for 30 minutes with occasional mixing.

The suspension was centrifuged for 5 minutes at 4°C, and the supernatant was aliquoted into eppendorf tubes. The aliquots of crude nuclear extract were snap frozen in liquid nitrogen and stored at -80°C.

(iii) From Dignam et al (1983)

Cells were harvested and washed twice in PBS (2.3.6), and then pelleted by centrifugation (Beckman J-21B Centrifuge, JA-20 rotor, 3000 rpm for 5 minutes). The cell pellet was resuspended in 5 volumes of Buffer A (2.3.7 (a) (ii)) and kept on ice for 10 minutes. The cells were pelleted by centrifugation as before and resuspended in 2 volumes of Buffer A, whereupon they were lysed by 10 strokes of a Dounce homogeniser. The lysate was centrifuged as before and the nuclear pellet was resuspended in 3 ml/10⁹ cells of Buffer C (2.3.7 (a) (ii)). The nuclei were lysed by 10 strokes of a Dounce homogeniser and the nuclear lysate was kept on ice for 1 hour with occasional mixing. Insoluble nuclear material was precipitated by centrifugation for 10 minutes at 4°C in an eppendorf microfuge and the supernatant (crude nuclear extract) was dispensed into aliquots which were snap frozen in liquid nitrogen and stored at -80°C.

(iv) From Dalton (1987)

Cells were harvested and washed twice in PBS (2.3.6), and then pelleted by centrifugation (Beckman J-21B Centrifuge, JA-20 rotor, 3000 rpm for 5 minutes). The
pellet was resuspended in lysis buffer (10 mM HEPES-NaOH, pH 8.0, 50 mM NaCl, 0.5 M sucrose, 0.1 mM EDTA, 0.5% (v/v) Triton X-100, 5 mM MgCl₂, and 1 mM DTT), at ~4x10⁷ cells/ml. The cells were lysed with several strokes of a Dounce homogenizer. Lysed cells were pelleted by centrifugation as before, and resuspended at ~5x10⁷ cells/ml in lysis buffer. The lysate was adjusted to 5 mM spermidine and 0.5 M NaCl, and kept on ice for 1 hour with occasional mixing. The suspension was dispensed into eppendorf tubes and centrifuged for 10 minutes at 4°C. The supernatant was mixed with ammonium sulphate (0.33 g/ml) at 4°C for 1 hour, and precipitated proteins were collected by centrifugation (14000 rpm, 20 minutes). The precipitate was resuspended in 1 ml of 0.1 M KCl in Tris-Magnesium (TM) buffer (2.2.5 (b)). The crude extract was snap frozen in aliquots and stored at -80°C. Alternatively, it was dialysed extensively against 0.1 M KCl in TM buffer at 4°C prior to storage.

2.3.7 (b) Small scale nuclear extract preparation from chicken embryos

Preparation of chicken embryo nuclei was based on the method of Panyim et al (1971). A dozen embryos were removed from their eggs, and placed in 30 ml of ice-cold 0.25 M sucrose in Grinding Buffer (10 mM Tris-HCl, pH 8.0, 10 mM MgCl₂, 50 mM NaHSO₃, 1 mM DTT, and 0.5 mM PMSF). All subsequent procedures were performed at 4°C to minimise proteolytic degradation of the extract. Embryos were homogenised for 1 minute in a Sorvall Omni-Mixer. The homogenate was filtered through four layers of muslin, and then a single layer of Miracloth to remove large particulate matter. Cells were pelleted by centrifugation for 10 minutes at 2600 rpm (Sorvall HB-4 rotor). The cell pellet was washed twice by resuspension in 15 ml of 0.25 M sucrose, 0.2% (v/v) Triton X-100 in Grinding Buffer, homogenisation for 15 seconds in a Sorvall Omni-Mixer, and centrifugation as before. The washed pellet was resuspended in 10 volumes of Grinding Buffer containing 2.2 M sucrose, 0.2% (v/v) Triton X-100, and cells were lysed with three strokes in a Dounce homogeniser. The cell lysate was made up to 29.4 ml and layered over a 12.6 ml cushion of Grinding Buffer containing 2.4 M sucrose, 0.2% (v/v) Triton X-100, in a Ti-70 heat sealable tube. The gradient was centrifuged in a Beckman L8-70 Ultracentrifuge (35000 rpm for 3 hours at 5°C). The nuclear pellet was washed with 10
Chapter Two: Materials and Methods

volumes of Buffer A (2.3.7 (a) (i)), and pelleted by centrifugation (Sorvall RC-5B centrifuge, SS-34 rotor, 7000 rpm for 5 minutes).

Nuclear extract was prepared from the purified nuclei as described above (2.3.7 (a) (i)).

2.3.7 (c) Large scale nuclear extract preparation from chicken embryos

Crude nuclear extract was prepared from large numbers of chicken embryos using a combination of the modified methods of Panyim et al (1971) and Lee et al (1988). 15-20 dozen chicken embryos were removed from their eggs, and placed in 500 ml of ice-cold Grinding Buffer (2.3.7 (b)). All subsequent procedures were performed at 4°C to minimise proteolytic degradation of the extract. Embryos were homogenised and filtered as before (2.3.7 (b)). Cells were pelleted by centrifugation (Beckman JA-10 rotor, 5000 rpm for 10 minutes). The cell pellet was washed twice by resuspension in 500 ml of Grinding Buffer containing 0.25 M sucrose, 0.2% (v/v) Triton X-100, homogenisation for 15 seconds, and centrifugation as before. The pellet was resuspended in an equal volume of Buffer A (2.3.7 (a) (ii)) and incubated on ice for 15 minutes. Cells were lysed by 5 passes through a 23 gauge needle, and the lysate was centrifuged at 8000 rpm for 10 minutes (Sorvall SS-34 rotor). The pellet was resuspended in an equal volume of Buffer C (2.3.7 (a) (ii)) and incubated on ice for 1 hour, with occasional gentle mixing. The suspension was centrifuged (Sorvall SS-34 rotor, 8000 rpm for 10 minutes), and the supernatant snap frozen in liquid nitrogen, and stored at -80°C.

2.3.7 (d) Protein concentration determination

Protein concentration of nuclear extracts was routinely determined by the method of Bradford (1976), using BSA as a standard.
2.3.8 Detection of DNA binding factors in nuclear extracts

2.3.8 (a) Probe preparation

For gel retardations involving oligonucleotide probes, one of the two complementary oligonucleotides containing the binding site of interest was kinased with $\gamma^{32}$P-ATP (2.3.5 (a)). ~1.5 ng of $32$P-labelled oligonucleotide was combined with 300 ng of unlabelled, complementary oligonucleotide in a 15 μl annealing reaction containing 200 mM NaCl. The mixture was heated to 100°C for 45 seconds, then allowed to cool to room temperature slowly by incubation as follows; 5 minutes at 85°C, 5 minutes at 70°C, 5 minutes at 55°C, 5 minutes at 37°C, and finally 5 minutes at room temperature.

2.3.8 (b) Mobility shift assay

DNA binding activity was assayed essentially as described (Schneider et al, 1986). Protein extract (usually 5 μg if crude extract) was combined with competitor DNA and 0.2 ng of annealed, $32$P-labelled probe in a 10-20 μl reaction volume containing 25 mM HEPES-NaOH, pH 7.6, 1 mM EDTA, 5 mM DTT, 100 mM NaCl, and 20% (v/v) glycerol. The contribution of NaCl and glycerol from the extract was included in the total concentration of the reaction. Extract was always added last to the reaction. After incubation at room temperature for 20 minutes, 5 μl of Tris-Glycine loading buffer was added to the reaction, which was then electrophoresed on a 10% non-denaturing Tris-Glycine polyacrylamide gel. Band shifts were visualised by autoradiography (2.3.17).

2.3.8 (c) Polyacrylamide gel electrophoresis of mobility shift assays

The products of a mobility shift assay were resolved on a 10% (w/v) polyacrylamide gel containing 40 mM Tris-glycine, pH 8.5. The gels used were 14 x 14 x 0.05 cm, and were pre-electrophoresed for 45 minutes at 4°C before use. The gels were run in 40 mM Tris-glycine buffer at 250-450 V at 4°C. When the dye had run the required distance, the gel was autoradiographed (2.3.17) at -80°C with a tungsten intensifying screen.
Chapter Two: Materials and Methods

2.3.9 Chromatographic enrichment of nuclear extract components

2.3.9 (a) DEAE- / Heparin-Sepharose chromatography

DEAE-Sepharose beads or Heparin-Sepharose CL-6B beads were rehydrated according to the manufacturer's protocol, packed into either a 10 ml column (pilot scale) or a 100 ml column (preparative scale), and allowed to settle under gravity. The column was then equilibrated with 3-5 column-volumes of 0.1 M KCl in TM Buffer. Crude nuclear extract was diluted to a final KCl concentration of 0.1 M with TM Buffer, and loaded onto the column. To remove unbound proteins, the column was washed with 3-5 column-volumes of 0.1 M KCl in TM Buffer. Bound proteins were eluted from the column by a series of washes with TM Buffer containing increasing KCl concentrations. Usually, 0.2 M, 0.3 M, and 0.4 M KCl washes were employed, followed with a 1.0 M wash to strip the column. Extract and salt washes were delivered to the column via a Gilson Minipuls 2 peristaltic pump, at a flow rate of 1 ml/minute. A Pharmacia Frac-100 fraction collector was employed to collect 5 ml samples from each salt wash, which were snap frozen in liquid nitrogen and stored at -80°C.

2.3.9 (b) Large scale concentration of partially purified extracts (from Hopwood, 1993)

Ultrafiltration of partially purified chicken embryo extracts was performed using an Amicon ultrafiltration stirred cell, model 8200 (200 ml capacity), under the specifications of the manufacturer. YM10 membranes were used, which have a 10 000 MW rating.

2.3.9 (c) Sephacryl S-300 gel filtration chromatography (from Hopwood, 1993)

Sephacryl S-300 beads were rehydrated according to the manufacturer's protocol, and packed into a 400 ml column 800 mm x 25 mm (for preparative scale chromatography), and allowed to settle under gravity. The column was then equilibrated with TM buffer plus 0.3 M NaCl. Heparin-Sepharose purified extract was diluted to a final NaCl concentration of 0.3 M with TM buffer, and loaded onto the column. TM buffer plus 0.3 M NaCl (2-3 column volumes) was continuously pumped onto the column at a flow rate of 0.5
ml/minute, and fractions collected from the point after the extract was applied to the column. Fractions were snap frozen in liquid nitrogen, and stored at -80°C.

2.3.9 (d) DNA-affinity microprecipitation (Franza et al, 1987)

(i) Preparation of DNA Ligand

10 μg each of the complementary oligonucleotides containing the binding site of interest were kinased separately in 20 μl reactions containing 60 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 10 mM DTT, 5 mM rATP, and 6 units of T4 polynucleotide kinase. The reactions were incubated at 37°C for 1 hour, after which they were pooled and combined with ~3 ng of one of the oligonucleotides that had been end-labelled with ³²P (2.3.5 (a)).

The combined reactions were incubated at 70°C for 10 minutes, then allowed to cool slowly to room temperature. A 50 μl cocktail containing 100 mM Tris-HCl pH 7.5, 20 mM MgCl₂, 2 mM DTT, and 1 mM rATP was added, and the annealed oligonucleotides were concatenated by the addition of 15 units of T4 DNA ligase. After 3 hours at room temperature, the reaction was extracted once with phenol : chloroform. The concatenated DNA was ethanol precipitated and resuspended in 10 μl of 0.1 mM EDTA at a DNA concentration of 1 μg/μl. A small sample of the reaction was electrophoresed on an 8% non denaturing polyacrylamide gel. The extent of concatenation was examined by autoradiography of the gel.

The concatenated DNA was biotinylated using a Photobiotin kit (Bresatec), following the supplier's protocol.

(ii) DNA affinity microprecipitation assay

Heparin sepharose fractionated extract was diluted to 0.15 M NaCl with Buffer B (50 mM Tris-HCl, pH 8.0, 0.01% (v/v) NP-40, 20% (v/v) glycerol, 1.5 mM MgCl₂, 1 mM DTT, and 0.5 mM PMSF). The extract was then centrifuged for 2 minutes to remove insoluble material. 200 μl of diluted, centrifuged extract was combined in an eppendorf tube with ~500 ng of biotinylated, concatenated oligonucleotide in the presence or absence of poly(dIdC). The sample was mixed thoroughly by attaching the tube to a rotating wheel for
30 minutes at room temperature. The extract was centrifuged briefly (eppendorf, 2 minutes), and the supernatant added to 50 µl (packed volume) of streptavidin-agarose beads which had been pre-washed with 0.15 M NaCl in Buffer B. Extract and beads were mixed for a further 30 minutes. The beads were precipitated (eppendorf, 15 seconds), and washed three times with 0.15 M NaCl in Buffer B. Bound proteins were eluted by mixing the beads with 50 µl of 1.0 M NaCl in Buffer B for 30 minutes. The beads were precipitated and the supernatant snap frozen in liquid nitrogen, and stored at -80°C.

An adaptation of this procedure involved attachment of the ligand to the beads prior to treatment of the extract. 0.5-1.0 µg of biotinylated, concatenated oligonucleotide in 200 µl of 0.15 M NaCl in Buffer B was mixed with 50 µl (packed volume) of pre-washed streptavidin-agarose beads for 3 hours at room temperature. The beads were precipitated and washed twice with 0.15 M NaCl in Buffer B. The extent of attachment of DNA to the beads was monitored by scintillation counting. Generally, >90% of the added counts was retained in the beads. 200 µl of diluted extract was then added to the beads, with mixing for 3 hours at 4°C. The rest of the procedure was continued as before.

In addition, it was found that the DNA-agarose beads could be regenerated after each experiment by three washes with 2.5 M NaCl in Buffer B followed by 3 washes with 0.15 M NaCl in Buffer B. In this manner, the matrix could be used over 10 times.

(iii) DNA affinity column chromatography

A sawn-off pasteur pipette was packed with 0.25 ml of DNA-agarose slurry and equilibrated with 0.15 M NaCl in Buffer B (2.3.9 (b) (ii)). 20 mg of partially purified extract was diluted to 0.15 M NaCl with Buffer B and loaded onto the column. The flow-through was delivered back to the column by means of a peristaltic pump. The extract was recirculated through the column for three hours at 4°C. After the flow-through was collected, the column was washed with 20 column-volumes of 0.15 M NaCl in Buffer B. Bound proteins were eluted with 0.75 M NaCl in Buffer B; about 10 x 100 µl fractions were collected. The fractions were snap frozen in liquid nitrogen and stored at -80°C.
Chapter Two: Materials and Methods

2.3.10 SDS Polyacrylamide gel electrophoresis of proteins

2.3.10 (a) Gel electrophoresis

Protein was electrophoresed on vertical 14 x 14 x 0.05 cm SDS polyacrylamide gels. The gel mix contained 10% (w/v) acrylamide (39:1 acrylamide : bisacrylamide), 2.5% (w/v) SDS, 375 mM Tris-HCl pH 8.8, 0.2% (w/v) ammonium persulphate, and 0.1% (v/v) TEMED. The gels also contained a 2 cm 4% (w/v) acrylamide stacking gel containing 125 mM Tris-HCl pH 6.8, 2.5% SDS, 0.2% ammonium persulphate, and 0.1% TEMED. Protein samples were treated prior to loading by addition of an equal volume of SDS loading buffer, and incubation at 70°C for 10 minutes.

Gel tanks contained approximately 1 litre running buffer (50 mM Tris-HCl pH 8.3, 380 mM glycine, 0.1% (w/v) SDS), and gels were run at 20 mA until the dye reached the bottom of the gel.

When protein samples were to be renatured from SDS polyacrylamide gels, or electroblotted to PVDF membranes for direct amino acid sequencing or Southwestern blotting, the gels were made entirely from Biorad reagents and allowed to polymerise overnight. In addition, the tank buffer in the top reservoir contained 1 mM thioglycollic acid, and gels were pre-electrophoresed for 30 minutes before loading. The loading buffer was made using Biorad reagents and recrystallised SDS (gift from Briony Forbes). These measures were to decrease the possibility of free radicals from the gel reacting with the N-terminus of the peptide of interest, which could result in N-terminal blocking or destruction of the DNA binding activity of the peptide.

2.3.10 (b) Silver staining of protein gels

When low levels of protein were present in the polyacrylamide gels (<200 ng), the protein was stained with silver by a modification of the simplified method described by Heuseshoven and Dernick (1985).

Gels were fixed overnight in 500 ml of 30% (v/v) ethanol, 10% (v/v) acetic acid. The gel was washed thoroughly with distilled water to attain a near neutral pH, and sensitized prior to staining in 200 ml of Farmer’s Reducer (0.15% (w/v) potassium
Chapter Two: Materials and Methods

ferricyanide, 0.3% (w/v) sodium thiosulphate, 0.05% (w/v) sodium carbonate) for 30 seconds.

The Farmer's Reducer was removed by washing with two changes of distilled water, for 10 minutes with each change. The gels were stained in 200 ml 0.1% (w/v) silver nitrate for 30 minutes, and the stain developed in 2.5% (w/v) sodium carbonate, 0.05% (v/v) formaldehyde (the formaldehyde was added just prior to developing). The developer was changed twice, when it became yellow-brown (after 10-60 seconds). When the protein bands had developed sufficiently, the reaction was stopped with 1% (v/v) acetic acid. The gel was washed with two changes of distilled water, for 10 minutes with each change, before it was dried between two sheets of cellophane on a Hoefer Scientific Instruments slab gel drier.

2.3.10 (c) Coomassie staining of protein gels

For large amounts of protein, the protein in the gel was visualized by staining with 0.1% (w/v) Coomassie blue in 50% (v/v) methanol, 10% acetic acid overnight at room temperature. The gel was destained in several changes of 5% (v/v) methanol, 10% (v/v) acetic acid by diffusion at room temperature. The gel was then dried as described above (2.3.10 (b)).

2.3.11 Electroblotting of protein to PVDF membranes

Protein was transferred to PVDF membrane (Immobilon) by the method of Matsudaira (1988), using a Hoefer Scientific Instruments transphor electrophoresis unit. Briefly, protein to be transferred was fractionated on a 10% SDS polyacrylamide gel (2.3.10 (a)). The gel was transferred to a PVDF membrane in 10 mM CAPS, 10% (v/v) HPLC-grade methanol at 400 mA for 30 minutes.

The membrane was washed for 5 minutes in distilled water, then stained in 0.5% (w/v) Coomassie blue, 50% (v/v) methanol, 5% (v/v) acetic acid for 5 minutes. The membrane was then destained in 50% (v/v) methanol, 10% (v/v) acetic acid for 10 minutes. Whilst still moist, the stained protein band was cut out, air dried, and stored at -20°C.
2.3.12 Association of DNA-binding activity with a protein

2.3.12 (a) South-Western transfer

300 μg of Heparin-Sepharose-purified chicken embro extract (2.3.9 (a)) was fractionated on a 10% SDS polyacrylamide gel and electroblotted to PVDF membrane as described (2.3.11). A track containing MW-SDS-200 markers (2.2.7 (b)) was run in a track adjacent to the extract. After the electroblot, the section of the membrane to which the markers had been transferred was stained with coomassie blue. All washing and blocking steps were performed at 4°C on a rocking platform. The section of the membrane containing the extract was washed with 1% (v/v) Triton X-100 in Buffer S (150 mM NaCl, 10 mM Tris-HCl, pH 7.4, 0.5 mg/ml sodium azide, 0.5 mM PMSF, and 1 mM DTT) for 30 minutes. The membrane was then blocked with 1% (w/v) BSA in Buffer S for 3 hours, and washed in 0.1% (v/v) Tween-20 in Buffer S for 10 minutes. The washing buffer was changed and the membrane was probed overnight with 1 μg of double-stranded 32P labelled oligonucleotide plus 2 μg poly(dIdC).

After hybridisation, the membrane was washed twice in 0.1% (v/v) Tween-20 in Buffer S for 45 minutes, air dried, and autoradiographed (2.3.17). The size of the hybridizing band was estimated by comparison with the coomassie stained section of the filter.

2.3.12 (b) Renaturation of binding activity from SDS polyacrylamide gels

Prior to renaturation of a DNA binding protein from SDS gels, the possibility of renaturing this protein from a sample of denatured extract was examined. 100 μg of Heparin-Sepharose-purified extract was heated to 70°C in 2.5% SDS / 700 mM β-mercaptoethanol for 10 minutes. The denatured extract was precipitated with 4 volumes of acetone, washed once in 80% acetone, and resuspended in 5 μl of 6 M guanidine-HCl. After incubation at 37°C for 30 minutes, the sample was diluted to 4 M guanidine-HCl in a final volume of 50 μl. This was dialysed extensively against 0.3 M TM buffer at 4°C overnight. The dialysed extract was centrifuged briefly to remove insoluble material, and the supernatant was analysed by mobility shift assay.
Chapter Two: Materials and Methods

Renaturation from SDS polyacrylamide gels was performed by one of two methods, each of which were based on a combination of techniques described by Briggs et al (1986) and Baeuerle and Baltimore (1988).

(i) Guanidine-HCl method

Protein extract (approximately 100 μg) was separated on a 10% SDS polyacrylamide gel. A parallel track containing protein molecular weight markers was separated from the sample track, and silver stained (2.3.10 (b)). The sample track was cut into sections of different MW ranges. Each section was placed in a separate eppendorf tube, ground, and 300 μl of elution buffer (50 mM Tris-HCl, pH 7.9, 0.1% (w/v) SDS, 0.2 mM EDTA, 10% glycerol, 1 mM DTT, and 0.5 mM PMSF) was added. The samples were shaken vigorously overnight at 4°C. Eluates were removed from ground gel slices and centrifuged for 10 minutes at 4°C to precipitate the remaining acrylamide. The supernatants were transferred to fresh eppendorf tubes, and proteins were precipitated with the addition of 4 volumes of HPLC-grade acetone. Precipitates were washed once in 80% acetone, resuspended in 2.5 μl of 6 M guanidine-HCl and incubated at 37°C for 10 minutes. 125 μl of Buffer B (2.3.9 (d) (ii)) was added, and the proteins allowed to renature overnight at 4°C. Renatured proteins were analysed by mobility shift assay (2.3.8 (b)).

(ii) Saturated urea method

This method is similar to (2.3.12 (b) (i)) except that the ground gel slices were eluted into 300 μl of 0.1% (v/v) Tween-20 in Buffer S (2.3.12 (a)). The acrylamide was removed by centrifugation, and proteins were precipitated with the addition of 4 volumes of HPLC-grade acetone. Precipitates were washed once in cold 100% HPLC-grade methanol, dried in vacuo, and resuspended in 1 μl of saturated urea. 50 μl of Buffer S was added to each precipitate and the proteins were allowed to renature overnight at 4°C.

2.3.13 Transient transfection of tissue culture cells

Cells were harvested (2.3.6) and washed twice in ice cold PBS. The cell pellet was resuspended in cold PBS and adjusted if necessary to 1x10^7 cells/ml (except for CHO-K1
cells, which were adjusted to 6.25x10^6 cells/ml. DNA to be transfected was mixed with 0.5 ml of cells (0.8 ml for CHO-K1 cells) in an electroporation cuvette. After incubation at 4°C for 10 minutes, the cells were electroporated at the appropriate voltage and capacitance. The transfected cells were left at room temperature for 10 minutes before plating in a 10 cm petri dish with 10 ml of the appropriate culture medium. The cells were left to recover for 48-72 hours before harvesting.

Transfection of HeLa cells was carried out as above except that HeBS supplemented with 6 mM glucose was used in place of PBS.

### 2.3.14 CAT assays

Extract preparation and CAT assays were performed essentially as described (Gorman et al, 1982). AEV-ts34 cells were harvested and washed once with PBS (2.3.6), and then pelleted. Cell pellets were resuspended in 50 µl of 250 mM Tris-HCl, pH 7.5, and cells were lysed by three rounds of freeze/thawing. The lysate was centrifuged for 10 minutes in an Eppendorf microfuge at room temperature, and the supernatant was incubated at 68°C for 10 minutes. The heat-treated extracts were centrifuged as before and the protein concentration of the supernatants were estimated by the method of Bradford (1976). Equal amounts of protein were added to each CAT assay.

Assays mixtures usually contained 10 µl of protein extract in 570 mM Tris-HCl, pH 7.5, 2 mM DTT, 0.5 mM Acetyl CoA (Sigma) and 6 nCi of D-threo-[dichloroacetyl-1-^14^C]chloramphenicol in a final volume of 150 µl. Reactions were allowed to proceed at 37°C for 2 hours. The reactions were then extracted with 1 ml of cold ethyl acetate, and the organic phase was evaporated in a speedivac. The chloramphenicol was resuspended in 20 µl of ethyl acetate, spotted onto silica gel thin layer plates, and run in chloroform : methanol (95:5, ascending). Plates were dried and exposed in a phosphorimaging cassette, and the samples were visualized on a PhosphorImager (2.3.19).
Chapter Two: Materials and Methods

2.3.15 Double Stranded Dideoxy Sequencing of DNA

Although there are considerable differences, the principle of this method is based on that described by Sanger et al (1977).

2.3.15 (a) Preparation of Template

2-5 ng of purified plasmid DNA (2.3.1 (a), (c)) or half of the DNA from a miniprep (2.3.1 (b)) was adjusted to 12 µl with water and mixed with 1.3 µl of 10 mg/ml RNase A. After a 15 minute incubation at 37°C, 3.3 µl of 1 M NaOH, 1 mM EDTA was added and the mixture incubated at 37°C for a further 15 minutes. During this procedure, spin columns were prepared (2.3.3 (a) (ii)). After the addition of acid-washed glass beads, the PCR tube was filled with CL-6B Sepharose (Pharmacia). The tube was then placed inside a 1.5 ml eppendorf tube and the storage buffer was removed by centrifugation (5 minutes at 6000 rpm in an Eppendorf 5415 C centrifuge). The 1.5 ml eppendorf tube was replaced with a new tube. The denatured DNA was layered onto the Sepharose and the assembly centrifuged as before. Usually ~15 µl of DNA was recovered.

2.3.15 (b) Hybridization

5 ng of the appropriate sequencing primer was annealed to 7 µl of denatured template in a 10 µl volume containing 40 mM Tris-HCl, pH 7.5, 50 mM NaCl, and 20 mM MgCl₂. The annealing reaction was heated to 65°C for 3 minutes, and allowed to cool to room temperature.

2.3.15 (c) Extension and termination

4.5 µl of a cocktail containing 0.6 µM dCTP, dGTP, and TTP, 20 mM DTT, and either α-³²P- or α-³⁵S-dATP (5 µCi) was added to each annealing reaction. 1 µl (2 units) of Sequenase (USB) was then added and the reactions incubated at 37°C for 3-5 minutes. 2.5 µl of the appropriate termination mixes; A, C, G, and T (Superbase sequencing kit - Bresatec) were placed into separate eppendorf tubes and pre-warmed to 37°C.
Chapter Two: Materials and Methods

3.5 μl aliquots from the extension reaction were dispensed into each termination mix. After a 3-5 minute incubation at 37°C, 4 μl of formamide loading buffer was added to each tube. The reactions were boiled for 2 minutes and loaded onto a 6% denaturing polyacrylamide gel. Electrophoresis was performed as described (2.3.2 (d)).

2.3.16 Polymerase Chain Reaction (PCR)

10 ng of target DNA was incubated in a 20 μl reaction containing tris, mg, 0.2 mM dNTPs, ~40 ng of each primer and 2.5 U of pfu polymerase. Reactions were placed in 0.5 ml eppendorf tubes and overlayed with mineral oil. The PCR included 16 cycles, each consisting of sequential denaturation (95°C, 1 minute), annealing (55°C, 2 minutes), and extension (72°C, 3 minutes) steps. In the first cycle, the denaturation step was for 2 minutes and the annealing step was at 52°C, while in the last cycle, the extension step was for 9 minutes. Cycles were executed using a PTC-100 Programmable Thermal Controller (M.J. Research, Inc.).

2.3.17 Autoradiography

32P-labelled DNA, which had been electrophoresed on polyacrylamide gels, was visualised by autoradiography. Gels were either covered by a thin sheet of plastic wrap before exposure, or dried onto a piece of Whatman 3MM paper using a Hoefer Scientific Instruments slab gel drier. A sheet of Fuji X-ray film or Kodak X-OMAT AR-50 film was placed over the gel enclosed in an Ilford autoradiography cassette, and exposed at room temperature for the required length of time. For detection of low levels of radioactivity, autoradiography was carried out in a cassette with a tungsten intensifying screen, at -80°C. After exposure, the X-ray film was developed, fixed, washed, and dried automatically.
2.3.18 Laser Densitometry

Protein gels and autoradiographs of mobility shift assays were quantitated on a Molecular Dynamics 300A Computing Densitometer.

2.3.19 Phosphorimaging

Genescreen membranes containing slot blot hybridisations or TLC plates containing CAT assay products were sealed in plastic bags and placed in a phosphorimager cassette for the required time. The phosphor screen was then scanned and the results quantitated using a Molecular Dynamics PhosphorImager.

2.3.20 Containment facilities and animal ethics

All manipulations involving recombinant DNA were carried out in accordance with the regulations and approval of the Australian Academy of Science Committee on Recombinant DNA and the University Council of the University of Adelaide.

All manipulations involving animals were carried out with the approval of the University of Adelaide Animal Ethics Committee.
CHAPTER THREE
TOWARD PURIFICATION OF
AC-BOX-BINDING FACTORS
FROM CHICKEN
Chapter Three: Toward Purification of AC-box-Binding Factors from Chicken

3.1 Introduction

Experiments conducted in this laboratory suggested that the H1 gene-specific promoter element, 5' AAACACA 3' (also known as the "H1-box" or "AC-box"), had an important role in the transcriptional regulation of the H1 gene: particularly the S-phase-specific increase in transcription. Deletion of this element, or altering a few bases of its sequence, led to a 15- to 30-fold decrease in steady state H1 transcript levels in exponentially dividing cells. In addition, these same mutations eliminated the S-phase associated increase of H1 transcription in synchronised cells (Dalton and Wells, 1988a).

Mobility shift assays using a 40 bp oligonucleotide probe derived from the promoter of the chicken H1 histone gene, 01 H1, identified a specific AC-box-binding activity in nuclear extracts from AEV ts34 chicken erythroblasts. This activity fluctuated in an S-phase-specific manner in parallel with the levels of H1 transcript (Dalton and Wells, 1988b). Other evidence indicated that the factor responsible for this activity, known as H1-SF, was a double stranded DNA binding protein with a MW of approximately 90 kD (Dalton, 1987).

The initial aim of my project was to characterise H1-SF, with the prospect that this would ultimately lead to a better understanding of the control of H1 histone gene expression. The first step toward characterisation of H1-SF was to purify a sufficient amount of the protein with which to generate amino acid sequence data: this information could be used for the production of antibodies against H1-SF, and/or in the design of oligonucleotides to screen libraries for the H1-SF gene. At least 30 pmol of homogeneous protein (~3 µg of a 90 kD protein) was required for direct N-terminal amino acid sequence analysis (B. Forbes, pers. comm.). This amount would require processing of several hundred mg of crude nuclear extract. Therefore, I set out to purify H1-SF on a large scale.

In this chapter, the identification of candidate AC-box binding factors from chicken embryo extracts is described. These factors were partially purified using several different chromatography steps and renatured from SDS polyacrylamide gels. It was concluded that none of the candidate factors were analogous to H1-SF. Finally, the characterisation of a DNA binding factor that appeared to be specific for the AC-box from a particular H1 gene subtype is presented.
Chapter Three: Toward Purification of AC-box-Binding Factors from Chicken

3.2 Characterisation of AC-box Binding Activity

3.2.1 Source of AC-box Binding Factors

To facilitate the purification of HI-SF, 9 day chicken embryos were chosen as the starting material. If HI-SF is involved with the positive activation of H1 histone genes in S-phase (as previous evidence suggested), then it should be expressed at near optimal levels in rapidly dividing cells, of which chicken embryos were an obvious source. Furthermore, compared to tissue culture cells, chicken embryos were an inexpensive means of accumulating a large amount of crude nuclear extract.

3.2.2 Choice and Preparation of Probe

Early mobility shift assays were carried out with the same 40 bp oligonucleotide probe used in the initial characterisation of HI-SF, from now on referred to as the 01-40mer. The AC-box was centrally located within this probe, which spanned the promoter of the 01 H1 gene from -135 to -101 (Figure 3.1). A second oligonucleotide pair, Δ01-40mer, was identical to the 01-40mer except for 4 base substitutions within the AC-box. To facilitate ligation of these oligonucleotides into plasmids (if required), Eco RI and Bam HI restriction sites were engineered at the 5' and 3' termini, respectively.

During the original characterisation of HI-SF, both strands of the 01-40mer were end labelled with $^{32}$P and annealed in stoichiometric quantities for use as a probe in mobility shift assays (Dalton, 1987). The annealed product was purified from single stranded oligonucleotides on a non denaturing polyacrylamide gel and eluted for use in the assay. However, it was found that a significant proportion of the probe (~10-20%) prepared in this manner denatured during the course of the assay (data not shown). This could have led to ambiguous results from band shifts produced by single stranded DNA binding proteins. Therefore to minimize the level of single-stranded probe, only one strand of the oligonucleotide pair (usually 01-40-t, see 2.2.8) was end-labelled with $^{32}$P before annealing with a 200-fold excess of unlabelled complementary oligonucleotide.
Chapter Three: Toward Purification of AC-box-binding Factors from Chicken

Figure 3.1 Oligonucleotides used in this study

The DNA sequence of the 01 H1 and 02 H1 promoters are shown from -143 to -34, and from -145 to -36, respectively. The core AC-box-, G-rich- and CCAAT-elements are boxed. Also illustrated are the sequences of the oligonucleotides described in this thesis, presented directly above or below the region of the promoters from which they were derived. Only the mutated AC-boxes of the Δ01-40mer, Δ01-25mer, and Δ02-25mer are shown, as the remainder of these oligonucleotides are identical to the corresponding wild type oligonucleotides. The mutated base pairs within these elements are asterisked.
3.2.3 Mobility Shift Assays of Chicken Embryo Extracts with the 01-40mer Probe

Before large scale preparation of chicken embryo extracts, it was important to confirm the presence of a factor equivalent to H1-SF. Crude nuclear extracts from 9 day chicken embryos were prepared on a pilot scale using a modification of the procedure described by Panyim et al. (1971) (2.3.7 (b)). In a typical mobility shift assay, 5 µg of extract was incubated with 100 pg (~3.8 fmol) of 01-40mer probe in the presence of 100 ng of poly(dIdC). Protein-DNA complexes were separated on a 10% non-denaturing Tris-glycine polyacrylamide gel (2.3.8 (c)) and visualized by autoradiography.

Mobility shift assays of chicken embryo crude extract with the 01-40mer probe produced a number of band shifts. These band shifts were competed out by increasing amounts of non-specific competitor DNA (Figure 3.2 (a)). Thus, no AC-box-specific binding activity was detected in chicken embryo crude extract. To determine whether or not this was a reflection of the experimental procedure, it was necessary to prepare and assay ts34 crude nuclear extracts in order to reproduce the H1-SF band shift previously demonstrated in this laboratory.

3.2.4 Analysis of AEV ts34 Nuclear Extracts with the 01-40mer Probe

On advice, (S. Dalton, pers. comm.) the initial method of ts34 crude nuclear extract preparation was based on the method of Strauss and Varshavsky (1984; see 2.3.7 (a) (i)). Extracts prepared using this method were analysed in mobility shift assays with the 01-40mer probe. In the presence of a 500-fold excess (by mass) of poly(dIdC), a number of band shifts were detected, designated 40-MS (mobility shift) -a, -b, -c, -d, -e, and -f (Figure 3.2 (b)). In contrast, no more than three band shifts were detected in similar assays previously conducted in this laboratory (Dalton, 1987). Based on their mobility relative to that of free probe, the H1-SF band shift appeared to be equivalent to the 40-MS-e band shift (data not shown). In addition, although some of the band shifts were not competed out by a 1000-fold excess of non-specific DNA, 40-MS-e was the only band shift that was not competed out by a 10000-fold excess of poly(dIdC) (Figure 3.3 (a)). This represented a highly specific factor (protein)-DNA interaction. 40-MS-e was therefore a strong candidate
Chapter Three: Toward Purification of AC-box-Binding Factors from Chicken

for H1-SF. A band shift of identical mobility and intensity to the 40-MS-e band shift was produced when the 01-40-b strand of the duplex probe was labelled (data not shown).

Consistent with the hypothesis that 40-MS-e was an AC-box-specific double stranded DNA binding protein, the 40-MS-e band shift was competed out with excess unlabelled 01-40mer, but not with excess unlabelled Δ01-40mer (Figure 3.3 (b)).

3.2.5 Small Scale DEAE- and Heparin-Sepharose Chromatography of ts34 Crude Nuclear Extracts

Further evidence that H1-SF and 40-MS-e were the same factor was obtained through partial purification of 40-MS-e by DEAE- and Heparin-Sepharose chromatography (2.3.9 (a)). 40 mg of ts34 crude nuclear extract was diluted to 0.1 M NaCl and loaded onto a 10 ml DEAE-Sepharose column that was connected in series to a 10 ml Heparin-Sepharose column. The flow-through and KCl washes were each collected as 8x5 ml fractions. Selected fractions from each salt wash were analysed by mobility shift assay. All detectable 40-MS-e was found exclusively in the flow-through from the DEAE column (data not shown). ~75% of the 40-MS-e that had eluted from the Heparin-Sepharose column was detected in fractions from the 0.3 M KCl wash (Figure 3.4). This was the same salt concentration at which H1-SF eluted from Heparin-Sepharose (Dalton, 1987). In addition, the specific activity (Units / μg protein) of 40-MS-e was highest in the 0.3 M KCl fractions (Table 3.1). Although a small amount of 40-MS-e was detected in other salt washes, this was most probably because the column had been slightly overloaded.

3.2.6 Attempts to Optimise Recovery of 40-MS-e

Long exposure times (several days with fast film) were required for detection of the 40-MS-e band shift, whereas the H1-SF band shift was detectable after an overnight exposure (S. Dalton, pers. comm.). It was possible that the intensities of the band shifts were affected by the conditions of the assay, and that a set of conditions might be found where the interaction with the probe was increased. Except for increasing NaCl concentration - which generally decreased binding to the probe - altering pH, Mg2+, EDTA
Figure 3.2 (a) Mobility shift assays of chicken embryo crude extract with the 01-40mer probe

9-day chicken embryo crude nuclear extract (~5 μg) was incubated with 100 pg of 01-40mer probe and increasing amounts of either poly(dIdC) or salmon sperm DNA. The products of the reaction were then separated on a 10% non denaturing Tris-glycine polyacrylamide gel. The fold-excess (by mass) of non specific competitor is shown beneath the corresponding lane. The position of unbound probe is indicated.

Figure 3.2 (b) Mobility shift assays of AEV ts34 crude nuclear extract with the 01-40mer probe

100 pg of 01-40mer probe in the presence of 50 ng (500-fold excess by mass) of poly(dIdC) was incubated with either no extract (lane 1) or 5 μg of ts34 crude nuclear extract (lane 2). The position of the 40-MS-a, -b, -c, -d, -e, and -f band shifts are indicated.
Figure 3.3 The 40-MS-e band shift indicates a highly specific interaction with the 01-40mer

(a) Mobility shift assays of ts34 crude nuclear extract were carried out with increasing amounts of non specific competitor as in Figure 3.2 (a). The type and fold-excess of competitor used in each assay is displayed below the corresponding lane. The 40-MS-e band shift is shown.

(b) 5 μg of Heparin-Sepharose-purified ts34 extract (3.2.5) was incubated with 01-40mer probe and increasing amounts of unlabelled 01-40mer or unlabelled Δ01-40mer. The fold-excess of competitor oligonucleotide is shown below the corresponding lane.
Figure 3.4 Partial purification of 40-MS-e by Heparin-Sepharose chromatography

ts34 crude nuclear extract was purified through DEAE- and Heparin-Sepharose columns as described in the text (3.2.5). 3 µl of selected fractions were analysed by mobility shift assays with the 01-40mer probe. The number of the fraction from each salt wash is indicated below the corresponding lane. The "C" lanes contain a control assay of 5 µg of ts34 crude nuclear extract. The position of the 40-MS-e band shift is indicated with an arrow.
Table 3.1 Specific activity of 40-MS-e in fractions from the small scale Heparin-Sepharose column

The 40-MS-e band shifts produced in mobility shift assays of fractions from the small scale Heparin-Sepharose column (Figure 3.4) were analysed by laser densitometry. For each fraction in which the 40-MS-e band shift was detected, the units of 40-MS-e binding activity, μg of total protein in the assay, and specific activity of the fraction are presented. One unit of 40-MS-e binding activity is defined as the amount of 40-MS-e required to produce the 40-MS-e band shift with an intensity equal to that produced by 5 μg of ts34 crude nuclear extract. As shown in the table, the specific activity of the fractions from the 0.3 M KCl wash were considerably higher (up to ~7-fold) than that of either the crude extract or the other salt wash fractions. NR = not recorded.
<table>
<thead>
<tr>
<th>Fraction no.</th>
<th>Flow-through 8</th>
<th>0.1 M KCl 1</th>
<th>0.2 M KCl 2</th>
<th>0.3 M KCl 3</th>
<th>0.3 M KCl 4</th>
<th>0.3 M KCl 5</th>
<th>0.3 M KCl 6</th>
<th>0.3 M KCl 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Units (U)</td>
<td>0.12</td>
<td>0.13</td>
<td>0.07</td>
<td>0.07</td>
<td>0.10</td>
<td>0.32</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td>Protein in assay (µg)</td>
<td>0.87</td>
<td>1.00</td>
<td>0.45</td>
<td>NR</td>
<td>0.26</td>
<td>0.44</td>
<td>0.26</td>
<td>0.16</td>
</tr>
<tr>
<td>Specific Activity (U/µg)</td>
<td>0.14</td>
<td>0.13</td>
<td>0.16</td>
<td>0.09</td>
<td>-</td>
<td>0.38</td>
<td>0.73</td>
<td>0.88</td>
</tr>
</tbody>
</table>
and DTT concentrations in the mobility shift assays had no significant effect on the intensities of these band shifts (data not shown).

Another reason for the apparently low abundance of 40-MS-e could be attributed to the method by which the extracts were prepared. Therefore, the method used in the original characterisation of H1-SF was adopted (2.3.7 (a) (iv)). Crude nuclear extract from ts34 cells prepared in this manner was assayed with the 01-40mer probe in the presence of various non-specific competitors (Figure 3.5). As well as the major band shifts observed with extracts prepared by the initial method (2.3.7 (a) (i)), the extracts prepared by this method produced an additional band shift with a lower mobility than the 40-MS-e band shift. However, although not competed out by excess poly(dIdC), this band shift was eliminated by addition of 1000-fold excess E. coli or salmon sperm DNA. Thus, extract prepared by this method produced no new band shifts which could be attributed to specific interaction of a factor with the 01-40mer. In addition, mobility shift assays conducted on a nuclear extract prepared using another well documented method (2.3.7 (a) (iii)) failed to increase the intensities of any of the band shifts, nor detected any band shifts that had not already been identified (data not shown).

3.2.7 Differences Between H1-SF and 40-MS-e

Further characterisation of 40-MS-e revealed some anomalies with H1-SF. Firstly, Heparin-Sepharose-purified ts34 extracts incubated with the 01-40mer probe in the presence of excess unlabelled 01-40-t (2.2.8) competed out the 40-MS-e band shift (Figure 3.6 (a)). In such assays, a significant level of single stranded probe was observed. This suggested that the excess unlabelled 01-40-t had displaced the labelled 01-40-t from the duplex probe. Thus it was not certain whether the disappearance of the 40-MS-e band shift was due to competition by unlabelled single stranded or double stranded oligonucleotide.

To address this question, labelled 01-40-t was incubated with ts34 extracts in a mobility shift assay. This produced a band shift with a slightly higher mobility than the 40-MS-e band shift (Figure 3.6 (b)). A similar band shift was produced with labelled 01-40-b (data not shown). The band shift produced with 01-40-t was at least 10 times more intense than that produced with double stranded probe. One explanation for these results
Chapter Three: Toward Purification of AC-box-Binding Factors from Chicken

was that 40-MS-e was in fact a single stranded DNA binding protein, and that the 40-MS-e band shift was produced by labelled 01-40-t present at minimal levels with the 01-40mer probe. Nevertheless, the 40-MS-e band shift was not competed out by an excess of unlabelled heterologous single stranded oligonucleotide (Figure 3.6 (a)), implying that even if 40-MS-e was a single stranded DNA binding protein, it was still specific for the AC-box oligonucleotide.

Another inconsistency was that a band shift of identical mobility to that of 40-MS-e was observed when the extract was assayed with labelled 01-40mer (Figure 3.6 (c)). This band shift (Δ40-MS-e) was also ~3- to 5-fold more intense than the 40-MS-e band shift. Because the specific activities of the 01-40mer and Δ01-40mer probes were similar, this implied that the factor responsible for these band shifts (if it was the same factor) had a higher affinity for the mutant oligonucleotide than for the wild type oligonucleotide. However, this did not agree with the previous finding that the 40-MS-e band shift was not competed out by unlabelled Δ01-40mer (Figure 3.3 (b)).

The conclusion reached in this study was that a factor which specifically bound the 01-40mer was detected in ts34 extracts, although it was possibly a single stranded DNA binding factor. However, because the results with the 01-40mer and Δ01-40mer probes did not agree with results previously obtained in this laboratory, it was concluded that this factor was not H1-SF.

3.2.8 Mobility Shift Assays with the 14mer Probe

One reason that an AC-box-specific factor was not readily detectable in chicken embryo crude nuclear extracts may have been that it was present in very low abundance. One or more rounds of purification may have been necessary before such a factor could be assayed successfully. In addition, a shorter oligonucleotide probe might also facilitate the detection of this factor. Gel mobility shift assays of chicken embryo and ts34 crude nuclear extracts using the 01-40mer probe detected several band shifts arising from non-specific DNA binding proteins. It was possible that an oligonucleotide with a less complex sequence would bind fewer proteins than the 01-40mer. A 14 bp oligonucleotide that was originally designed as a ligand for DNA affinity chromatography (see 3.4.1) contained only the
Figure 3.5 Mobility shift assays of ts34 crude extract prepared using the original method

5 µg of extract prepared by the method used in the original characterisation of H1-SF (2.3.7 (a) (iv)) was incubated with 01-40mer probe in the presence of 1000-fold excess (by mass) of either poly(dIdC) (lane 2), E. coli DNA (lane 3), or salmon sperm DNA (lane 4). Lane 1 contains a control assay of ts34 extract prepared by the usual method (2.3.7 (a) (i)). The position of a unique band shift in the extract prepared by the original method is indicated with an arrow, as is the position of the 40-MS-e band shift.
Figure 3.6 Anomalies between the 40-MS-e and H1-SF band shifts

(a) Heparin-Sepharose-purified ts34 extract was incubated with 01-40mer probe in the presence of 500-fold excess (by mass) of either poly(dIdC) (lane 1), unlabelled 01-40-t (lane 2), or unlabelled H5E-t (lane 3), which spans the 3' enhancer of the chicken H5 gene (Sparrow, 1991). The 40-MS-e band shift is indicated with an arrow.

(b) The same extract as in (a) was incubated with either double stranded (ds) 01-40mer probe (lane 1) or single stranded (ss) 01-40-t probe (lane 2). The 40-MS-e band shift is shown.

(c) ts34 crude nuclear extract was incubated with either wild type ("w.t.") 01-40mer probe (lane 1) or mutated ("mut.") Δ01-40mer probe (lane 2) in the presence of 1000-fold excess (by mass) of poly(dIdC). Indicated with arrows are the 40-MS-e and Δ40-MS-e band shifts.
chicken AC-box consensus flanked by cohesive ends (Figure 3.1) This oligonucleotide (01-14mer) was used as a probe in mobility shift assays of chicken embryo crude nuclear extracts.

The 01-14mer probe gave rise to three band shifts, referred to as 14-MS-a, -b, and -c, when incubated with chicken embryo crude extracts (Figure 3.7). As well as binding fewer proteins, another advantage of the 01-14mer probe was that the exposure time required to view the band shifts (overnight) was much shorter than for the 01-40mer probe (up to several days). The 14-MS-a doublet had a low mobility, and was therefore possibly caused by binding of high MW aggregates to the probe. The 14-MS-c band shift was unusual because it had a slightly higher mobility than that of free probe; one explanation for this result was that 14-MS-c had a high negative charge. 14-MS-c also appeared to be a highly abundant single stranded DNA binding protein. Incubation of the same amount of extract with only the labelled strand of the 01-14mer probe, 01-14-t (2.2.8), resulted in the disappearance of free single stranded probe and a single intense band shift at the identical position to the 14-MS-c band shift (data not shown). In contrast, the 14-MS-b band shift was absent in the same assay, suggesting that 14-MS-b was a double stranded DNA binding factor. Therefore, the most likely candidate for an AC-box-specific factor analogous to H1-SF was 14-MS-b.

The 01-14mer probe denatured readily under the conditions of the standard mobility shift assay. Relatively high NaCl concentrations (300 mM) and 1000-fold excess of unlabelled complementary strand over the labelled strand were required to retain double-stranded probe in the assay. Even under these conditions, only ~50% of the probe remained annealed (data not shown).

The 14-MS-b band shift was competed out by excess unlabelled non-specific competitors. Poly(dIdC) was more effective at competing out 14-MS-b than salmon sperm DNA (see Figure 3.14 (a)). However, the 14-MS-b band shift was produced by an AC-box-specific interaction because it was not detected with an 01-14mer probe in which the same base changes were made to the AC-box as with the Δ01-40mer (Hopwood, 1993). In addition, 14-MS-b displayed sequence specificity when unlabelled heterologous oligonucleotides were used as competitors (see Figure 3.9 (b)). 14-MS-b was therefore a
candidate for H1-SF, and the purification of this factor from chicken embryo extracts was pursued on a large scale. However, competitor DNA such as poly(dIdC) was no longer included in mobility shift assays in which the 01-14mer was used as a probe.

### 3.3 Large Scale Preparation and Chromatography of Chicken Embryo Extracts

500 dozen 9 day chicken embryos were processed in batches of 100 dozen, using a modified version of the method of Lee et al (1988) (2.3.7 (a) (ii)). This produced ~1 L of 10 mg/ml crude protein extract, in 0.4 M NaCl. Extracts were diluted to 0.2 M NaCl in preparation for loading onto a Heparin-Sepharose column. Dialysis was avoided as there was evidence that this resulted in some loss of DNA-binding activity (data not shown).

#### 3.3.1 Heparin-Sepharose Chromatography

In a typical experiment, ~1 g of diluted crude protein extract was loaded onto a 100 ml Heparin-Sepharose column equilibrated with 0.2 M TM Buffer. Chromatography was carried out essentially as described (2.3.9 (a)). Fractions from each salt wash were analysed in mobility shift assays with the 01-14mer probe. The 14-MS-b activity eluted in the 0.3 M KCl wash, and the fractions containing this activity were pooled. After five column runs, ~1.5 L of <1 mg/ml Heparin-Sepharose extract in 0.3 M KCl was obtained.

N.B. Large-scale Heparin-Sepharose chromatography was carried out with the assistance of Duncan Sparrow and Blair Hopwood.

#### 3.3.2 Ultrafiltration of Heparin-Sepharose Extracts

Because the sample loaded on a gel filtration column could not exceed 5% of the column volume, it was necessary to reduce the volume of the Heparin-Sepharose extract. The pooled fractions containing the 14-MS-b activity were concentrated as described
Figure 3.7 Detection of DNA binding activities in chicken embryo crude extract using the 01-14mer probe

01-14mer probe was incubated with chicken embryo crude extract and the products electrophoresed as with mobility shift assays using the 01-40mer probe. The 14-MS-a (a doublet), -b, and -c band shifts, as well as unbound probe, are shown. Lane 1: probe alone. Lane 2: probe + extract.
Chapter Three: Toward Purification of AC-box-Binding Factors from Chicken

(2.3.9 (b)). The Heparin-Sepharose extract was reduced to 150 ml, with a protein concentration of ~8 mg/ml.

N.B. Ultrafiltration was carried out exclusively by Blair Hopwood.

3.4 DNA Affinity Chromatography of Partially Purified Extracts

3.4.1 Design of a Suitable Oligonucleotide

It was generally accepted that when purifying proteins by DNA affinity chromatography, the oligonucleotide containing the binding site should be as short as possible, and concatenated to produce the ligand (R. Tjian, pers. comm.). To this end, a 14 bp oligonucleotide (01-14mer) was designed containing only the chicken AC-box consensus, 5' AAGAAACACA 3', flanked by cohesive 5' and 3' ends (Figure 3.1). As it had already been demonstrated by mobility shift assays that this probe was bound by a sequence-specific factor (14-MS-b), it was most likely an appropriate ligand for DNA affinity chromatography.

Apart from its length, another reason for not selecting the 01-40mer as a DNA affinity ligand was that concatenation of this oligonucleotide was very inefficient: after a typical self-ligation reaction, >50% of the 01-40mer remained as a monomer (data not shown).

3.4.2 Preparation of the DNA Affinity Ligand

01-14-t (2.2.8) was kinased with a trace of γ-32P-ATP, annealed, and self-ligated (2.3.9 (d) (i)). A sample of the reaction was electrophoresed on an 8% non-denaturing polyacrylamide gel. Autoradiography of the gel revealed that the oligonucleotides had ligated to an average length of ~400 bp (data not shown). These concatemers were subsequently photobiotinylated using a Bresatec Photobiotin kit under the directions of the manufacturer.
3.4.3 DNA Affinity Microprecipitation

The initial method of DNA affinity purification was an adaptation of the microprecipitation technique (2.3.9 (d) (ii)). In early experiments, 100-500 μg of concentrated, Heparin-Sepharose-purified extract was mixed with 500 ng of concatenated, photobiotinylated DNA (referred to as DNA ligand) in an eppendorf tube for 30 minutes before the addition of 50 μl of streptavidin agarose beads. After further mixing, the supernatant was removed and the beads were washed extensively with 0.15 M NaCl. Bound proteins were eluted with 1.0 M NaCl. Because the 14-MS-b band shift was readily competed out by non specific competitors in mobility shift assays (see 3.2.8), such competitors were not included in the binding reaction in an effort to improve the yield of DNA affinity-purified 14-MS-b.

Samples of untreated extract, supernatant, 0.15 M NaCl-wash, and eluate were analysed in gel mobility shift assays with the 01-14mer probe. A typical result is shown in Figure 3.8 (a). The first observation was that there was a high loss of 14-MS-b activity (~84%) during the course of the microprecipitation. Although the eluate contained a proportion of the recovered 14-MS-b (~8%), it appeared that the majority of recovered 14-MS-b was retained in the supernatant from the microprecipitation assay (~70%).

Another point of interest was that the eluate contained no trace of 14-MS-c. This is in agreement with the finding that the 14-MS-c band shift was produced exclusively by an interaction with single stranded 01-14-t probe (3.2.8), none of which would be present in the DNA-agarose matrix.

In light of the low recovery of 14-MS-b in the eluate from the microprecipitation, it was necessary to establish conditions in which this recovery was maximised.

3.4.4 Optimisation of Microprecipitation Method

Before embarking on large scale microprecipitation of 14-MS-b, several modifications were made in order to improve the yield of this factor in the eluate.

Because the DNA ligand was radioactively labelled, the proportion which had attached to the streptavidin agarose beads during the microprecipitation procedure could be monitored. Radioactivity in samples of supernatant and beads from several
microprecipitation experiments were measured by scintillation counting. It was estimated that up to 60% of the DNA ligand was retained in the supernatant (data not shown), implying that not all of the DNA had attached to the streptavidin agarose matrix. This was most likely a major reason for the low yield of 01-14-b. One explanation for this finding was that some of the DNA, particularly <200 bp, had escaped photobiotinylation (J. McInnes, pers. comm.). In addition, the proportion of DNA ligand attached to the beads might be increased by extending the mixing time.

To address these problems, the DNA ligand was mixed with the streptavidin agarose beads and unbound DNA was removed prior to the addition of extract. Typically, 500 ng of DNA ligand was mixed with 50 µl of streptavidin agarose beads in an eppendorf tube for three hours at room temperature. After washing unbound DNA from the matrix, it was estimated by scintillation counting that 70 to >90% of the DNA remained bound (data not shown). The DNA-agarose beads were then mixed with ~750 µg of partially purified extract, and the rest of the procedure was identical to the initial microprecipitation technique.

Several ways in which the recovery of 14-MS-b might be further improved were investigated. Firstly, samples of partially purified extract were mixed with DNA-agarose for either 1 hour at room temperature or 3 hours at 4°C. There was at least a 2-fold increase in the level of 14-MS-b in the eluate from the 3 hour mixing (Figure 3.8 (b)). There was no further increase in this recovery after 6 or 12 hours of mixing (data not shown).

In another series of experiments, equal amounts of partially purified extract were incubated with equal amounts of DNA-agarose in either 75 mM, 150 mM, or 300 mM NaCl. The optimal recovery of 14-MS-b in the eluate occurred when the salt concentration of the extract / DNA-agarose mixture was 150 mM (data not shown).

The loss of 14-MS-b activity after one round of microprecipitation was negligible (that is, the total 14-MS-b activity in the supernatant and eluate was approximately equal to the 14-MS-b activity in the starting material). The supernatant from one round of microprecipitation still contained a significant level of 14-MS-b. This supernatant could be treated a second time with DNA-agarose. The amount of 14-MS-b in the eluate from this second microprecipitation (as determined by mobility shift assay) was comparable to the amount obtained from the first microprecipitation, thus increasing the yield of affinity-
purified 14-MS-b by up to 2-fold (Figure 3.8 (c)). The resultant loss of 14-MS-b activity during this second treatment of the supernatant was about 50%. Finally, a third treatment of the same supernatant did not produce significant levels of DNA affinity purified 14-MS-b, and was therefore no longer pursued.

The presence of 14-MS-b in the supernatant indicated that the binding capacity of the DNA-agarose was being exceeded. It was conceivable that the amount of extract mixed with DNA-agarose could be reduced without lowering the amount of 14-MS-b in the eluate. This in turn could reduce the loss of 14-MS-b activity that occurred during microprecipitation. However, when the amount of extract was decreased, the level of 14-MS-b in the eluate also decreased (data not shown).

Using the modified DNA affinity microprecipitation procedure, recovery of 14-MS-b from Heparin-Sepharose-purified extracts was ~40% as estimated by laser densitometric analyses (see Table 3.2).

During the course of these experiments, it became apparent that significant levels of 14-MS-a were also present in the DNA affinity-purified extracts (see Figures 3.8 (b) and (c)). As non specific competitor was not included in the binding reaction (3.4.3), it was possible that any protein capable of binding the 01-14mer (specifically or non specifically) would be purified in this manner. Thus the appearance of the 14-MS-a band shift in the eluates from the microprecipitation experiments was not surprising.

### 3.4.5 Scale-up of Microprecipitation Method

Once the conditions for optimal recovery of 14-MS-b from DNA affinity microprecipitation were established, the procedure was scaled-up. It appeared that simply increasing the volume of extract and DNA-agarose beads in the mixing reaction resulted in decreased yields of DNA binding factor in the eluate (D. Sparrow, pers. comm.). Therefore, a large scale experiment consisted of 20-30 eppendorf tubes, each containing an extract / DNA-agarose mixture equivalent to one pilot scale experiment. Large scale experiments normally involved processing of ~20 mg of Heparin-Sepharose-purified extract at a time.
Chapter Three: Toward Purification of AC-box-binding Factors from Chicken

Figure 3.8 (a) Purification of 14-MS-b by DNA affinity microprecipitation

160 µg of Heparin-Sepharose-purified chicken embryo extract was processed in a DNA affinity microprecipitation experiment as described (2.3.9 (b) (i)). 1/40th, 1/20th, 1/50th, and 1/10th of the Heparin-Sepharose-purified extract (lane 1), supernatant (lane 2), low salt wash (lane 3) and high salt eluate (lane 4), respectively, were incubated with 01-14mer probe in mobility shift assays. Fractions of the total sample incubated in each assay were not the same due to the volume constraints of the assay. The 14-MS-b band shift is indicated.

Figure 3.8 (b) Optimisation of microprecipitation conditions #1: binding reaction time

2x400 µg samples of Heparin-Sepharose-purified chicken embryo extract were mixed with DNA-agarose in two identical microprecipitation experiments, except that one sample was mixed for 1 hour at room temperature and the other for 3 hours at 4°C. 1/100th, 1/10th, 1/50th, and 1/10th of the Heparin-Sepharose-purified extract (lane 2), supernatant (S), low salt wash (W) and high salt eluate (E), respectively, were assayed for the presence of 14-MS-b by incubation with 01-14mer probe. Lane 1 contains a mobility shift assay with no extract.

Figure 3.8 (c) Optimisation of microprecipitation conditions #2: re-treatment of supernatant

The supernatant from a typical microprecipitation experiment was mixed with a fresh sample of DNA-agarose in a second microprecipitation, and the supernatant from this experiment was processed in a third treatment. One tenth of the supernatants (S) and eluates (E) from all three treatments were assayed for the presence of 14-MS-b. Lanes 1 and 2: first treatment. Lanes 3 and 4: second treatment. Lanes 5 and 6: third treatment.
Chapter Three: Toward Purification of AC-box-Binding Factors from Chicken

It was found that the yield of DNA affinity-purified 14-MS-b from large scale microprecipitation of Heparin-Sepharose-purified extract was similar to that obtained in the pilot scale experiments (~40%).

3.5 Possible Co-purification of H1-Transcription Factors with 14-MS-b

It was envisaged that other transcription factors, in particular, the factors which interacted with the G-rich- and CCAAT-elements of the H1 promoter, were most likely to be present in the chicken embryo extracts. In addition, the conserved spacing between these elements and the AC-box, and their proximity, suggested that the factors which bound these elements may interact with each other in vivo. These factors could remain associated in nuclear extracts; therefore purification of the AC-box binding protein could facilitate the isolation of the other transcription factors. Characterisation of these factors would ultimately lead to a more complete understanding of transcriptional control of the H1 gene.

To investigate the possibility of purifying these factors, oligonucleotides containing the G-rich-box and CCAAT-box sequences were synthesized (2.2.8). These oligonucleotides were 20 bp and 19 bp in length, and spanned the 01 H1 promoter from -81 to -62 and from -62 to -44, respectively (Figure 3.1).

The G-rich-box 20mer and CCAAT-box 19mer were end-labelled with \(^{32}\)P and annealed in the same manner as the 01-40mer, and used as probes in mobility shift assays with chicken embryo crude nuclear extracts. Each probe detected a band shift with very similar mobility to that of 14-MS-b (data not shown). This same band shift was also detected with both probes in DNA affinity purified extracts (see Figure 3.9 (a)). These observations were consistent with the hypothesis that the factors which bound these sequences were associated in chicken embryo extracts, and that the G-rich-box- and CCAAT-box-binding factors were being co-purified with the AC-box-binding factor.

To verify this hypothesis, the G-rich-box 20mer or CCAAT-box 19mer probes were incubated with DNA affinity purified extract in the presence of excess unlabelled 01-14mer.
Chapter Three: Toward Purification of AC-box-Binding Factors from Chicken

If 14-MS-b existed as a complex of the AC-box-, G-rich-box- and CCAAT-box-binding factors, then the MW of this complex would be increased with the binding of the unlabelled oligonucleotide. This in turn should cause a decrease in the mobility of the band shift. However, instead of a decreased mobility, the band shift produced by either probe was competed out by unlabelled 01-14mer (Figure 3.9 (a)). In addition, unlabelled G-rich-box 20mer or CCAAT-box 19mer oligonucleotides competed out the 14-MS-b band shift (see Figure 3.9 (b)). In short, the band shift, regardless of the probe used, was competed out by any of the unlabelled oligonucleotides. This implied that 14-MS-b was not a complex of DNA binding factors, but a factor with a single DNA binding domain.

Although 14-MS-b was capable of binding several different DNA sequences, it displayed high specificity for the 01-14mer sequence. Unlabelled 01-14mer was ~20 times more effective at competing out the band shift produced with the G-rich-box 20mer probe than unlabelled homologous oligonucleotide; similar results were obtained with the CCAAT-box 19mer probe (Figure 3.9 (a)). The specificity of 14-MS-b for the 01-14mer was confirmed in a series of competition experiments, where DNA affinity purified eluates were incubated with the 01-14mer probe and increasing amounts of unlabelled homologous or heterologous oligonucleotides (Figure 3.9 (b)). All of the oligonucleotides tested at least partially competed out the 14-MS-b band shift. However, unlabelled 01-14mer was at least 100 times more effective at competing it out than any of the heterologous oligonucleotides.

It was interesting to note that the effects on the 14-MS-a doublet in these experiments were converse to the effects on 14-MS-b. This band shift doublet was unaffected by excess unlabelled 01-14mer, but was decreased considerably by either unlabelled G-rich-box 20mer or CCAAT-box 19mer, thereby demonstrating specificity for the latter two oligonucleotides. It was therefore possible that the factors which bound these sequences in the H1 promoter were being co-purified with 14-MS-b. However, in the course of this work, this was not able to be confirmed.
Figure 3.9 (a) 14-MS-b binds different DNA sequences

DNA affinity-purified chicken embryo extract was incubated with 01-14mer, G-rich-box 20mer, or CCAAT-box 19mer probes in the presence or absence of unlabelled competitors. Lanes 1, 2, and 5: no competitor. Lanes 3 and 6: 100-fold excess (by mass) of unlabelled G-rich-box and CCAAT-box oligonucleotides, respectively. Lanes 4 and 7: 100-fold excess (by mass) of unlabelled 01-14mer.

Figure 3.9 (b) Specificity of 14-MS-b for the 01-14mer

DNA affinity-purified chicken embryo extract was incubated with 01-14mer probe in the presence of various unlabelled oligonucleotide competitors. The numbers below each lane refer to the fold-excess (by mass) of competitor in the assay. Lane 1 contains an assay without extract. Lane 2: no competitor. Lanes 3 and 4: unlabelled 01-14mer. Lanes 5 and 6: unlabelled CCAAT-box 19mer. Lanes 7 and 8: unlabelled G-rich-box 20mer. Lanes 9 and 10: unlabelled Drosophila HSE 32mer (contains a consensus heat shock factor binding site and flanking sequences derived from the D. melanogaster hsp 70 promoter (Pelham, 1982).
Chapter Three: Toward Purification of AC-box-Binding Factors from Chicken

3.6 SDS Polyacrylamide Gel Electrophoresis of DNA Affinity Purified Extracts

To determine whether the DNA affinity-purified extracts contained any visibly enriched proteins, samples of Heparin-Sepharose-purified extract, supernatant, 0.15 M NaCl-wash, and eluate from the microprecipitation procedure was electrophoresed on SDS polyacrylamide gels. Because the protein concentration in the eluates was low (<250 µg/ml), proteins were visualized by silver staining (2.3.10 (b)). This method is reported to be up to 100-1000 times more sensitive than coomassie staining (see Sambrook et al, 1989).

When the protein profiles of supernatant and DNA affinity eluate from the microprecipitation experiments were compared, enriched proteins were found in the eluate (Figure 3.10). In particular, two proteins of 86.0 kD and 69.6 kD were consistently enriched. Laser densitometric analysis of the eluate protein profiles revealed that the 86 kD and 70 kD proteins each constituted between 5 and 10% of the total eluate protein content, respectively (data not shown). The former protein was particularly interesting because it had a similar MW to that reported for H1-SF (Dalton, 1987). However, it must be emphasized that because of the absence of poly(dIdC) in these experiments (3.4.3), it was not certain that these proteins were specifically interacting with the AC-box.

3.7 Direct Amino Acid Sequencing of a DNA Affinity-Purified Protein

In anticipation that the enriched proteins in the DNA affinity eluates were AC-box-specific factors, attempts were made to determine their N-terminal amino acid sequence. The eluate from one large scale DNA affinity microprecipitation of Heparin-Sepharose-purified extract (~250 µg of DNA affinity-purified extract) was TCA-precipitated and electrophoresed on an SDS polyacrylamide gel. After electrophoresis, the gel was electroblotted onto a PVDF filter (2.3.11), which was briefly stained with coomassie blue. The 70 kD protein
could not be sequenced by this method as it co-migrated with another protein. The region of filter containing the 86 kD protein was excised and the protein sequenced directly on an Applied Biosystems 470 A Protein Sequencer. No sequence data were obtained, suggesting that the protein was N-terminally blocked (13 amino acids of N-terminal sequence was obtained from a second protein on the same filter - data not shown).

A second sample of DNA affinity-purified extract was acetone-precipitated, resuspended in SDS loading buffer and electrophoresed on an SDS polyacrylamide gel that had been modified to lower the possibility of N-terminal blocking (R. Simpson, pers. comm.; see 2.3.10). From this sample, 8 residues of N-terminal sequence from the 86 kD protein were obtained (data not shown). Because the amount of protein in the sample was low, most of the amino acid identities were uncertain.

In order to obtain reliable amino acid sequence data, it was necessary to increase the level of the 86 kD protein by up to 10-fold. However, a threefold increase in the amount of DNA affinity purified extract overloaded the SDS polyacrylamide gel (data not shown). Thus to load a sufficient amount of the 86 kD protein for direct amino acid sequencing, it was necessary to apply a second round of DNA affinity purification to the extract.

3.8 2nd-Round DNA Affinity Microprecipitation

~50 µg of eluate from one round of DNA affinity microprecipitation was diluted to 0.15 M NaCl and incubated with 100 µl of DNA-agarose beads. Samples of first round eluate, supernatant, low salt wash, and second round eluate were electrophoresed on an SDS polyacrylamide gel, silver-stained, and analysed by laser densitometry. In the high salt eluate from the second round of DNA affinity purification, the 86 kD and 70 kD proteins had increased to 20% and 40% of the total protein content, respectively (Figure 3.11 (a)). However, there was approximately an 8-fold decrease in the amount of the 86 kD protein after the second round of DNA affinity purification, and the losses of the 70 kD protein appeared to be even higher (see Figure 3.11 (a)). In addition, mobility shift assays detected ~13% recovery of 14-MS-b activity in the 2nd round eluates (Figure 3.11 (b)). It would
Figure 3.10 Analysis of DNA affinity-purified extracts by SDS polyacrylamide gel electrophoresis

1/100th (~5 µg), 1/50th, 1/25th, and 1/2, respectively, of the Heparin-Sepharose-purified chicken embryo extract (H.S.), supernatant (S), low salt wash (W), and eluate (E) from a microprecipitation experiment were separated on a 10% SDS polyacrylamide gel and silver-stained as described (2.3.10). "M" refers to MW markers, the molecular masses of which are shown in kD. The enriched proteins in the eluate (~115.5 kD, 86.0 kD, and 69.6 kD) are indicated with arrows. The largest of these proteins was only observed in this particular experiment. The molecular masses of the 86.0 and 69.6 kD proteins were calculated by averaging the estimated molecular masses of these proteins from three different protein gels.
<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>H.S.</th>
<th>S</th>
<th>W</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>205 kD</td>
<td>←</td>
<td></td>
<td>116 kD</td>
<td></td>
<td>115 kD</td>
</tr>
<tr>
<td>116 kD</td>
<td>←</td>
<td></td>
<td>97 kD</td>
<td></td>
<td>86 kD</td>
</tr>
<tr>
<td>97 kD</td>
<td>←</td>
<td></td>
<td>66 kD</td>
<td>←</td>
<td>70 kD</td>
</tr>
<tr>
<td>66 kD</td>
<td>←</td>
<td></td>
<td>45 kD</td>
<td>←</td>
<td></td>
</tr>
<tr>
<td>45 kD</td>
<td>←</td>
<td></td>
<td>29 kD</td>
<td>←</td>
<td></td>
</tr>
<tr>
<td>29 kD</td>
<td>←</td>
<td></td>
<td></td>
<td>←</td>
<td></td>
</tr>
</tbody>
</table>
Chapter Three: Toward Purification of AC-box-binding Factors from Chicken

Figure 3.11 Second round DNA affinity microprecipitation

(a) A second round of DNA affinity microprecipitation was applied to a sample of DNA affinity-purified chicken embryo extract. Aliquots from the first round extract (E1), supernatant (S2), low salt wash (W2), and second round eluate (E2) (1/10th, 1/25th, 1/50th, and 1/4 of the total fractions, respectively) were loaded onto an SDS polyacrylamide gel which was silver-stained after electrophoresis. The enriched proteins are indicated with arrows, as are the molecular weight markers (M).

(b) 1/15th, 1/30th, 1/50th, and 1/14th of the E1, S2, W2, and E2 fractions, respectively, were incubated with 01-14mer probe in a mobility shift assay. The 14-MS-b band shift is also indicated.
therefore be necessary to process at least several hundred mg of Heparin-Sepharose-purified extract in order to accumulate a sufficient amount of second round DNA affinity-purified material for direct amino acid sequencing of the 86 kD and 70 kD proteins. This was equivalent to about half of the total Heparin-Sepharose-purified extract. As there was no direct evidence linking 14-MS-b with any of the proteins enriched in DNA affinity-purified extracts, two different lines of investigation were pursued. Firstly, alternative means of purifying the 86 kD and 70 kD proteins without the need for two rounds of DNA affinity microprecipitation were investigated. Secondly, experiments were conducted to determine which protein(s) was responsible for the 14-MS-b band shift.

3.9 Alternative Purification Methods

3.9.1 Sephacryl S-300 Gel Filtration Chromatography

In each of six column runs, 20 ml (~160 mg) of concentrated Heparin-Sepharose extract was loaded onto a 400 ml Sephacryl S-300 column (2.3.9 (c)) equilibrated with 0.3 M TM buffer. After loading the extract, 0.3 M TM was pumped onto the column until all the fractions were collected. Fractions containing 14-MS-b activity were pooled and concentrated with the Amicon ultrafiltration cell as before (3.3.2). In total, 180 ml of pooled S-300 fractions were concentrated to ~60 ml of 0.5 mg/ml extract.

The yield and fold-purification of 14-MS-b after this chromatography step are summarised in Table 3.2.

N.B. Sephacryl S-300 chromatography and ultrafiltration were carried out exclusively by Blair Hopwood.

Large scale microprecipitation of ~1 mg batches of Sephacryl S-300-purified extract were performed, and the recovery of 14-MS-b was examined by mobility shift assay. It was found that the yield of 14-MS-b from large scale microprecipitation varied depending on the extract from which it was purified. As mentioned above, the overall yield of 14-MS-b after DNA affinity microprecipitation of Heparin-Sepharose-purified extracts was ~40%.
Chapter Three: Toward Purification of AC-box-Binding Factors from Chicken

However, the loss of 14-MS-b activity that occurred during large scale microprecipitation of Sephacryl S-300-purified extract was considerably greater than that of Heparin-Sepharose-purified extract (data not shown). In addition, the yield of 14-MS-b was much lower (≤17%) in the eluates from DNA affinity microprecipitation of Sephacryl S-300-purified extract than those of Heparin-Sepharose-purified extract (see Table 3.2).

3.9.2 DNA Affinity Column Chromatography

Until now, all DNA affinity experiments involved mixing the extract and the agarose-bound DNA in eppendorf tubes. This procedure was especially labour intensive in large scale experiments. In addition, the multiple centrifugation steps during washing may have dissociated some proteins from the DNA, as well as damaging the agarose beads. In an alternative procedure, a small column was packed with 0.25 ml of the DNA-agarose slurry and ~0.25-0.5 mg of Sephacryl S-300-purified extract was recirculated through the column (2.3.9 (d) (iii)). Samples of the flow-through, low salt wash, and high salt eluate fractions were analysed by mobility shift assay (Figure 3.12 (a)).

In terms of the yield of DNA affinity-purified 14-MS-b, chromatography compared favourably with the microprecipitation procedure. After Sephacryl S-300-purified extract was applied to only one round of DNA affinity chromatography, the yield of DNA affinity-purified 14-MS-b was as high as 33% (Table 3.2). Therefore column chromatography became the method of choice for large scale DNA affinity-purification of 14-MS-b.

Analysis of the flow-through, low salt wash, and eluate from the DNA affinity chromatography column on an SDS polyacrylamide gel also revealed that the 86 kD and 70 kD proteins had become considerably enriched in the eluate (Figure 3.12 (b)).

To further test the performance of the column, a second-round DNA affinity chromatography experiment was conducted. A sample of DNA affinity-purified extract was loaded onto the column and the procedure was as for a first round experiment. There was an 18.5% recovery of 14-MS-b in the eluate from this second round of purification, as determined by densitometric analysis of band shifts (Figure 3.12 (c) and Table 3.2).
Figure 3.12 DNA affinity column chromatography of 14-MS-b

(a) Sephacryl S-300-purified chicken embryo extract was circulated through a DNA affinity column as described (2.3.9 (b) (ii)). Aliquots of untreated extract (S-300), flow-through (S), low salt wash (W) and fractions eluted with high salt (1-10) were assayed with 01-14mer probe. \( \frac{1}{360} \)th of the total S-300 and flow-through samples, \( \frac{1}{500} \)th of the total low salt wash sample, and \( \frac{1}{40} \)th of each of the high salt eluate fractions were assayed. 

p = assay without extract.
Figure 3.12 DNA affinity column chromatography of 14-MS-b

(b) Samples of untreated extract (S-300; \(1/2000^{th}\) of the total sample), flow-through (S; \(1/2000^{th}\) of the total sample), 0.15 M NaCl wash (W; \(1/500^{th}\) of the total sample) and a fraction from the 0.75 M NaCl elution which contained a relatively high level of 14-MS-b (E3; \(1/3^{rd}\) of the total fraction) were electrophoresed on a 10% SDS polyacrylamide gel. The enriched proteins are indicated with arrows.

(c) Second round DNA affinity chromatography. DNA affinity-purified chicken embryo extract was circulated through the DNA affinity column. Aliquots of first round extract (E1; \(1/90^{th}\) of the total sample), flow-through (S; \(1/150^{th}\) of the total sample), low salt wash (W; \(1/300^{th}\) of the total sample) and high salt eluate fractions (1-8; \(1/30^{th}\) of each fraction) were assayed for 01-14mer binding activity. The 14-MS-b band shift is indicated.
b

116 kD
97 kD
66 kD
45 kD

M S-300 S W E3

86 kD
70 kD

14-MS-b

c

14-MS-b

E1 S W 1 2 3 4 5 6 7 8
Chapter Three: Toward Purification of AC-box-binding Factors from Chicken

Table 3.2 Summary of 14-MS-b purification from chicken embryo extracts

The yield, specific activity and fold-purification of 14-MS-b after each chromatographic treatment of chicken embryo extracts are shown. The figures for crude extract, Heparin-Sepharose-, and Sephacryl S-300-purified extract (the latter two obtained after ultrafiltration) are provided courtesy of Hopwood, 1993. Specific activity and % yield are based on laser densitometric analyses of band shift intensities in mobility shift assays of the extracts. For DNA affinity microprecipitation, the figures on the left refer to purification from Heparin-Sepharose-purified extract, while those on the right refer to purification from Sephacryl S-300-purified extract. DNA affinity purification figures are from the results of typical experiments. Note that for some DNA affinity-purified extracts, the protein concentration was below the limit of detection by the Bradford assay. In such cases, the concentration was estimated from silver-stained SDS polyacrylamide gels where extracts were run alongside a known amount of protein MW markers.
<table>
<thead>
<tr>
<th>Procedure</th>
<th>Protein (mg/ml)</th>
<th>Specific Activity (Units/mg)</th>
<th>Fold Purification</th>
<th>% Yield (cumulative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude Nuclear Extract</td>
<td>10</td>
<td>100</td>
<td>-</td>
<td>(100)</td>
</tr>
<tr>
<td>Heparin Sepharose</td>
<td>8</td>
<td>700</td>
<td>7</td>
<td>(87)</td>
</tr>
<tr>
<td>Sephacryl S-300</td>
<td>0.5</td>
<td>12300</td>
<td>18</td>
<td>(49)</td>
</tr>
<tr>
<td>DNA affinity microprecipitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first round</td>
<td>0.2</td>
<td>9250</td>
<td>13</td>
<td>40 (35)</td>
</tr>
<tr>
<td>second round</td>
<td>.016</td>
<td>44130</td>
<td>3.6</td>
<td>16.5 (8)</td>
</tr>
<tr>
<td>DNA affinity chromatography</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first round</td>
<td>.005</td>
<td>164500</td>
<td>~13</td>
<td>33 (16)</td>
</tr>
<tr>
<td>second round</td>
<td></td>
<td></td>
<td></td>
<td>18.5 (3)</td>
</tr>
</tbody>
</table>
3.10 Renaturation of 14-MS-b

3.10.1 Southwestern Analysis

In parallel with the DNA affinity chromatography experiments, attempts were made to identify 14-MS-b. The first technique adopted was Southwestern blotting. DNA affinity-purified extract was electrophoresed on an SDS polyacrylamide gel and electroblotted onto PVDF in the same manner as for direct amino acid sequencing. The filter was blocked, probed and washed as described (2.3.12 (a)), and then visualised by autoradiography.

The first Southwestern blots, in which the filters were probed with the 01-14mer, failed to detect any DNA-binding proteins (data not shown). Analysis of the probe on a non-denaturing polyacrylamide gel revealed it had completely denatured under the conditions of the Southwestern. It was therefore not possible to identify 14-MS-b by Southwestern blotting. Subsequent blots, in which the 01-40mer was used as a probe, are discussed below (3.11.2).

3.10.2 Renaturation of Proteins from Gel Slices

As an alternative to Southwestern blots, attempts were made to renature 14-MS-b from SDS polyacrylamide gels. To determine whether such a renaturation was possible, a sample of Heparin-Sepharose-purified extract was first denatured by heating to 70°C in SDS loading buffer. The denatured extract was acetone-precipitated, resuspended in 6 M guanidine-HCl, and dialysed extensively against 0.3 M TM buffer at 4°C overnight (2.3.12 (b)). Samples of denatured and dialysed extracts were analysed by mobility shift assay. Despite no detectable band shifts in extract treated with SDS, there was a band shift in the renatured sample with identical mobility to that of native 14-MS-b (Figure 3.13 (a)). It was estimated that ~5% of the starting material had been renatured in this manner.

Following the success of the previous experiment, 100 μg of Sephadryl S-300-purified extract was loaded into a 1 cm well of an SDS polyacrylamide gel. The gel had been modified in the same manner as those used in direct amino acid sequencing of proteins electroblotted onto PVDF (2.3.10). After electrophoresis, the lane containing the extract was cut into ~0.5 cm slices. Proteins were eluted from each slice and renatured as described
(2.3.12 (b) (i)). These eluates were each tested in mobility shift assays with the 01-14mer probe.

After renaturation, the eluate from only one of the gel slices produced a band shift with identical mobility to the 14-MS-b band shift. Surprisingly, this eluate corresponded to a region near the bottom of the gel, containing the dye-front (Figure 3.13 (b) and (c)). This indicated that the protein(s) responsible for this band shift was no larger than 29 kD.

This renatured factor displayed the same characteristics as native 14-MS-b. The band shift was competed out by non-specific competitors; as with native 14-MS-b, poly(dIdC) was a more effective competitor than salmon sperm DNA (Figure 3.14 (a)). It was also competed out by various oligonucleotides, but unlabelled 01-14mer was at least 20 times more effective at competing out the band shift than any heterologous oligonucleotide (Figure 3.14 (b)). It also produced the same band shift with G-rich-box and CCAAT-box oligonucleotide probes (data not shown). There was little doubt that this renatured factor was the same protein(s) that had previously been identified as 14-MS-b. Therefore, neither the 86 kD nor the 70 kD proteins, although enriched in DNA affinity purified eluates, were likely to be 14-MS-b.

In order to obtain a more accurate estimate of the MW of 14-MS-b, attempts were made to renature the factor from 15% SDS polyacrylamide gels. However, no 01-14mer binding activity could be detected in the renatured eluates from these gels (data not shown).

### 3.10.3 Analysis of ts34 Extracts with the 01-14mer Probe

If H1-SF was a 90 kD protein, then it should bind the 01-14mer probe to give rise to a band shift with lower mobility than the 14-MS-b band shift. In an effort to confirm this, a mobility shift assay of ts34 crude nuclear extract with the 01-14mer probe was conducted. Surprisingly, despite repeated attempts, no band shift was detected. This result indicated that 14-MS-b was not present in ts34 extract. Together with the finding that 14-MS-b was much smaller than the predicted MW of H1-SF, it seemed unlikely that these factors were analogous.
Figure 3.13 Renaturation of 14-MS-b

(a) A sample of Heparin-Sepharose-purified chicken embryo extract was denatured and renatured as described in the text. Equal amounts of extract before (lane 1) and after (lane 2) denaturation, after resuspension in 6 M Guanidine-HCl (lane 3), and after dialysis (lane 4) were analysed by mobility shift assay. Note that some 14-MS-b activity was restored after resuspension in guanidine.

(b) and (c) Renaturation of 14-MS-b from SDS polyacrylamide gels. 100 µg and 2.5 µg samples of Sephacryl S-300-purified chicken embryo extract were electrophoresed in separate lanes on a 10% SDS polyacrylamide gel. The lane containing the 100 µg sample was then cut into 11 slices which were treated as described in the text. (b) The 2.5 µg sample (S-300) after silver staining. The slices into which the "100 µg" lane had been sectioned were numbered 1-11 as indicated.

(c) The renatured eluates from the gel slices were assayed for 01-14mer binding activity. The number below each lane refers to the number of the gel slice. A sample of Sephacryl S-300-purified chicken embryo extract (C) was also assayed as a positive control.
Figure 3.14 DNA binding properties of the renatured factor are the same as those of native 14-MS-b

The renatured factor was incubated in a series of mobility shift assays with 01-14mer probe in the presence of various non specific (a) and oligonucleotide (b) competitors. The type and fold excess (by mass) of competitor is shown below the corresponding lane. Lanes 1 contain no competitor. Note that poly(dIdC) was >30 times more effective at competing out the band shift than salmon sperm DNA.
3.11 Identification and Renaturation of an 01-40mer Binding Factor in DNA Affinity Purified Extracts

3.11.1 Characterisation of the 40-MS-1 Band Shift

The renaturation of 14-MS-b conclusively demonstrated that this factor was much smaller than the 86 kD and 70 kD proteins enriched in DNA affinity eluates. It was still not known which of these species, if any, were analogous to H1-SF. In an attempt to resolve this, the 01-40mer probe was employed in mobility shift assays of DNA affinity-purified extracts. Although this probe was unsuccessful in detecting a specific AC-box-binding factor in chicken embryo crude nuclear extracts, it was assumed that the relative abundance of such a factor in DNA affinity-purified extracts was sufficiently high to allow detection.

A band shift with a similar mobility to 40-MS-e from ts34 nuclear extracts was detected in DNA affinity-purified extracts from chicken embryos (Figure 3.15 (a)). The factor responsible for this band shift (40-MS-1) appeared to be a double stranded DNA binding protein. Although a highly intense band shift was produced by incubation of labelled 01-40-t or -b with DNA affinity-purified extracts, this band shift had a significantly lower mobility than the 40-MS-1 band shift (Figure 3.15 (b)).

The interaction of 40-MS-1 with the 01-40mer was highly specific. The 40-MS-1 band shift was completely removed by 100-fold excess of unlabelled 01-40mer, but remained unaffected by competition with up to 1000-fold excess of unlabelled heterologous oligonucleotides or by a 10000-fold excess of poly(dIdC) (Figure 3.15 (c)). It was also unaffected by competition with unlabelled 01-14mer, indicating that it was distinct from 14-MS-b. The competition experiments identified two additional band shifts produced by highly specific interactions with the 01-40mer. These band shifts occurred as a low mobility doublet (Figure 3.15 (c)).

3.11.2 Southwestern Analysis with the 01-40mer Probe

In an attempt to determine the MW of 40-MS-1, a Southwestern blot of Sephacryl S-300-purified extract was performed. The blocking, probing and washing steps were carried out as before (3.10.1), except that the 01-40mer was used to probe the PVDF filter.
A number of proteins in the Southwestern blot bound the 01-40mer probe, including the 86 and 70 kD proteins enriched in DNA affinity-purified extracts (Figure 3.16). The most intense labelling was of a doublet which co-migrated with the 45 kD marker. Proteins with molecular masses of ~45 kD were previously detected in DNA affinity-purified extracts which had been separated on SDS polyacrylamide gels and silver-stained (data not shown). Intense labelling also occurred at the ion front, possibly due to binding of the probe to 14-MS-b. From this result alone it was not possible to ascertain which (if any) of the labelled proteins were 40-MS-1.

### 3.11.3 Renaturation of 40-MS-1

The approximate MW of 40-MS-1 was determined by renaturation of this factor from SDS polyacrylamide gels. Sephacryl S-300-purified chicken embryo extract was electrophoresed on an SDS polyacrylamide gel, and eluted proteins from the gel slices were renatured in the same manner as for 14-MS-b. However, mobility shift assays of these renatured eluates were unable to detect a band shift (data not shown).

Using a slightly different renaturation technique (2.3.12 (b) (ii)), a band shift with identical mobility to the 40-MS-1 band shift was detected in the renatured eluate from one of the slices of the SDS gel. The region of the gel from which this slice was obtained corresponded to a range of ~39-46 kD (Figure 3.17). A second experiment narrowed this range to 42-46 kD (data not shown). It was quite possible that renatured 40-MS-1 was one or both of the ~45 kD proteins identified as an intensely labelled doublet by Southwestern blotting (see above). This factor was therefore distinct from 14-MS-b, but still much smaller than H1-SF or the enriched proteins in DNA affinity purified extracts. Thus neither the 01-40mer nor the 01-14mer probes detected a factor analogous to H1-SF in chicken embryo extracts.

In addition to 40-MS-1, a second 01-40mer binding factor was renatured in this fashion. This factor was recovered from the gel slice containing the ion front (Figure 3.17). The mobility of the band shift produced by this factor was higher than that of 40-MS-1. It was therefore possible that this band shift was produced by 01-14-b.
Chapter Three: Toward Purification of AC-box-binding Factors from Chicken

Figure 3.15 (a) Identification of an 01-40mer binding factor in DNA affinity-purified extracts

01-40mer probe was incubated with either 5 μg of ts34 crude nuclear extract (lane 1), 10 μg of Heparin-Sepharose-purified chicken embryo extract (lane 2), or 0.75 μg of DNA affinity-purified chicken embryo extract (lane 3). Assays were in the presence of either 1000-fold (lanes 1 and 2) or 5000-fold (lane 3) excess of poly(dIdC). The position of the 40-MS-e and 40-MS-1 band shifts are shown.

Figure 3.15 (b) 40-MS-1 does not appear to be a single stranded DNA binding factor

~50-100 ng of DNA affinity-purified chicken embryo extract was incubated with either double stranded or single stranded oligonucleotide probe in the presence of 5000-fold excess of poly(dIdC). Lane 1: labelled 01-40-t probe. Lane 2: labelled 01-40-b probe. Lane 3: 01-40mer probe (labelled 01-40-b strand). Lane 4: 01-40mer probe (labelled 01-40-t strand). The 40-MS-1 band shift is shown. Also indicated is a doublet of band shifts which appear to be produced by highly specific interactions with the 01-40mer. "ss" denotes the position of intense band shifts produced by interactions with the single stranded probes.

Figure 3.15 (c) The 40-MS-1 band shift denotes a highly specific interaction

~50-100 ng of DNA affinity-purified chicken embryo extract was incubated with 01-40mer probe in the presence of various unlabelled competitors. The type of competitor and fold-excess (by mass) are indicated below the corresponding lane. "CCAAT", "G-rich", and "HSE" refer to the CCAAT-box 19mer, G-rich-box 20mer, and Drosophila HSE 32mer, respectively. The 40-MS-1 band shift is indicated with an arrow, as well as the lower mobility doublet. Lane 11 is considered to be an anomalous result.
Figure 3.16 Southwestern analysis of chicken embryo extract with the 01-40mer probe

~200 µg of Sephacryl S-300-purified chicken embryo extract was electrophoresed on a 10% SDS polyacrylamide gel and electroblotted onto an immobilon filter which was blocked, washed, and probed as described (2.3.12 (a)). After visualising the DNA binding proteins by autoradiography, the filter was stained with coomassie blue. Alignment of the stained filter with the autoradiograph allowed the positions of the 86 and 70 kD proteins to be determined (indicated with arrows). In addition to these proteins, an intensely labelled doublet at ~45 kD ("?") and a single protein which co-migrated with the ion-front ("14-MS-b?") are shown. Although not indicated in this figure, a protein which migrated immediately above the 70 kD protein was also labelled.
Figure 3.17 Renaturation of 40-MS-1 from SDS polyacrylamide gels

~160 µg and 2.5 µg samples of Sephacryl S-300-purified chicken embryo extract were electrophoresed in separate lanes on a 10% SDS polyacrylamide gel. The lane containing the 160 µg sample was then cut into 9 slices which were treated as described in the text.

(a) The lane containing the 2.5 µg (S-300) sample after silver staining. The numbering to the right of the lane is as in Figure 3.13 (b).

(b) The renatured eluates from the gel slices were assayed for 01-40mer binding activity. The numbering below each lane is as in Figure 3.13 (c). A sample of Sephacryl S-300-purified chicken embryo extract (C) was also assayed as a positive control.
3.12 Use of Intermediate Length Oligonucleotide Probes

3.12.1 Design of 25 bp Oligonucleotides

It appeared that the difficulties incurred with the 01-14mer and 01-40mer probes might be overcome with a probe of sufficient length to allow strong sequence-specific binding, but which also bound fewer non-specific proteins. To meet these requirements, three 25 bp oligonucleotides were designed with the core AC-box centrally located and 7 or 8 bases of promoter sequence flanking either side (2.2.8). Two of these oligonucleotides, referred to as the 01-25mer and 02-25mer, were derived from congruent regions of the promoters of the chicken histone 01 H1 and 02 H1 genes, respectively. The 01-25mer included the 01 H1 promoter sequence between -126 and -102, while the 02-25mer spanned the promoter of the 02 H1 gene from -130 to -106. The third oligonucleotide, Δ01-25mer, had the same flanking sequences as the 01-25mer, but every base except the central 'C' of the core AC-box was altered (Figure 3.1). The 5' and 3' termini of the oligonucleotides were blunt; restriction sites were not included as they may have created spurious binding sites. GC base pairs were chosen as oligonucleotide termini to aid in preventing the 25mers from denaturing under the conditions of the mobility shift assay.

It should be noted that the bases flanking the core AC-box in the 01- and 02-25mers shared very little homology, so that any common band shift detected by these probes almost certainly represented an interaction with the AC-box. Such a result could be confirmed if the same band shift was not detected with the Δ01-25mer.

3.12.2 Mobility Shift Assays of Chicken Embryo Extracts with the 25mers

The oligonucleotides were end-labelled with $^{32}$P and annealed in the same manner as the 01-40mer. Chicken embryo extracts at various stages of purification were incubated with the probes in mobility shift assays. Several band shifts were detected by all three probes; in fact, the 01-, 02-, and Δ01-25mers all produced near identical band shift patterns (Figure 3.18). There were no band shifts produced with the 01- or 02-25mers that were not also
detected with the Δ01-25mer. Thus, no specific AC-box-binding factor was detected in chicken embryo extracts with the 25mer probes.

3.12.3 DNA Affinity Purification with 25mers

It was still not known whether the 86 kD or 70 kD proteins enriched in DNA affinity purified extracts interacted specifically with the AC-box. In a final attempt to determine whether these proteins were AC-box-specific, three different DNA affinity columns were prepared. Each column contained either the 01-, 02-, and Δ01-25mers as the ligand. These oligonucleotides had been concatenated, photobiotinylated, and attached to steptavidin agarose beads in the same manner as the 01-14mer (2.3.9 (d) (i)). Equal samples of Sephasyl S-300-purified extract were loaded onto each column. High salt eluate fractions from each experiment were electrophoresed on SDS polyacrylamide gels. If the 86 kD or 70 kD proteins were AC-box-specific, then they should not be enriched in the eluate from the Δ01-25mer column.

The enrichment of 86 kD and 70 kD proteins observed in eluates from the 01-14mer affinity purifications was also apparent in the eluate fractions from all three chromatography experiments, including the Δ01-25mer column (Figure 3.19). This implied that these proteins did not specifically interact with the AC-box, since the Δ01-25mer did not contain this element. There were no other enriched proteins, nor were there any significant differences between the protein profiles of the eluates from the three experiments. It was therefore not possible to distinguish a specific AC-box-binding factor in chicken embryo extracts using the 25mers.

3.12.4 Characterisation of a Subtype-Specific AC-box-Binding Factor in ts34 Extracts

As no specific AC-box-binding factor could be identified with the 25mers in chicken embryo extracts, it was decided to assay extracts from various cell lines. Firstly, crude nuclear extracts were prepared from ts34 cells using the modified method of Lee et al (1988) (2.3.7 (a) (ii)). These extracts were incubated with the 25mer probes in mobility shift assays. As with chicken embryo extracts, the band shift patterns were very similar for each
Figure 3.18 Analysis of chicken embryo extracts with 25mer probes

The 01-, Δ01-, and 02-25mers were used as probes in mobility shift assays of chicken embryo extracts at various stages of purification. The probe used in each assay is indicated below the figure. The assays are as follows; Lanes 1: no extract. Lanes 2: 8 μg of crude extract. Lanes 3: 8 μg of crude extract plus 500-fold excess of poly(dIdC). Lanes 4: 0.75 μg of Sephacryl S-300-purified extract plus 500-fold excess of poly(dIdC). Lanes 5: ~25 ng of DNA affinity-purified extract. Lanes 6: ~25 ng of DNA affinity-purified extract plus 500-fold excess of poly(dIdC).
3x175 µg samples of Sephacryl S-300-purified extract were circulated through one of three DNA affinity columns (2.3.9 (d) (iii)) containing concatenated 01-, 02-, or Δ01-25mers as the ligand. Aliquots of untreated extract ("F" lane; 1/700th of the total sample) and 0.75 M NaCl eluate fractions 1-7 (lanes 1-7; 1/3rd of the total fraction) were electrophoresed on 10% SDS polyacrylamide gels and silver-stained. The protein profiles of the eluate fractions from each DNA affinity column were essentially identical; only the untreated extract and 0.75 M NaCl eluate fractions from the Δ01-25mer column are shown. The 86 kD and 70 kD proteins are indicated with arrows. Also indicated are the molecular masses of the proteins in the marker lane ("M").
probe (Figure 3.20 (a)). However, a single unique band shift (02-25-1) was detected with the 02-25mer in both this extract and an older extract prepared by the initial method (2.3.7 (a) (i)). 02-25-1 was not a single stranded DNA binding factor: single stranded 02-25-t probe failed to detect a band shift with a similar mobility to 02-25-1. This factor was also highly specific: the 02-25-1 band shift was competed out with a 100-fold molar excess of unlabelled homologous oligonucleotide, but not competed out with up to 1000-fold molar excess of unlabelled 01-25mer nor Δ01-25mer (Figure 3.20 (b)). It was also unaffected by a 10000-fold excess of poly(dIdC) (see Figure 3.21 (b)).

Although the 02-25-1 band shift represented a highly specific 02-25mer-binding activity, it appeared that the AC-box was not essential for this interaction. This was because the 01-25mer probe did not give rise to a similar band shift in the same extract, even though it also contained an AC-box. To confirm this finding, another 25 bp oligonucleotide, Δ02-25mer, was synthesized with identical sequence to the 02-25mer except that the same changes were made to the AC-box as with the Δ01-25mer (Figure 3.1).

Like the other 25mers, Δ02-25mer probe produced a number of band shifts in mobility shift assays of ts34 crude nuclear extracts. However, none had the same mobility as the 02-25-1 band shift (Figure 3.21 (a)). In addition, unlabelled Δ02-25mer did not compete out the 02-25-1 band shift (Figure 3.21 (c)). Therefore the presence of the AC-box in the 02-25mer was essential for the generation of the 02-25-1 band shift. An attempt to identify the nucleotides bound by 25-MS-1 using methylation interference analysis was unsuccessful (data not shown).

### 3.12.5 Mobility Shift Assays of Various Cell Lines with the 25mer Probes

The biological significance of the 02-25-1 band shift remained unclear. This band shift had not been detected in chicken embryo extracts. To determine whether this activity was a phenomenon of ts34 cells, crude nuclear extracts were prepared from several cell lines. These included three chicken cell lines: HD11 macrophages, CEF-38 fibroblasts (a finite line), and LMH hepatomas. CHO and HeLa cell lines were also analysed.

Although undetectable in CHO or HeLa crude nuclear extracts (data not shown), a band shift with identical mobility to that of 02-25-1 was present in the extracts from each of
the chicken cell lines (Figure 3.22 (a)). As with 02-25-1, this band shift was not detected with the 01-, Δ01-, nor the Δ02-25mer probes. In addition, it was competed out by unlabelled 02-25mer, but not competed out by a 1000-fold excess of heterologous oligonucleotides, nor by up to a 10000-fold excess of poly(dIdC) (Figure 3.22 (b)). These results were identical for all of the chicken cell lines examined.

3.13 Discussion

Previous work in this laboratory had identified a factor, H1-SF, from ts34 nuclear extracts which specifically interacted with the AC-box. H1-SF had been characterised as a double stranded DNA binding protein with a molecular mass of ~90 kD. This chapter describes the purification and characterisation of H1-SF-like factors from chicken embryo extracts.

Mobility shift assays of chicken embryo crude nuclear extracts using the same 01-40mer probe that was employed in the initial characterisation of H1-SF failed to detect the same band shift that was produced with H1-SF. Although a number of band shifts were detected with this probe in chicken embryo crude nuclear extracts, they appeared to be due to non-specific interactions. One reason why the H1-SF band shift was not detected in these extracts may be that H1-SF was a very low abundance protein, and one or more rounds of purification may have been required before it could be detected by mobility shift assay. Others attempting to purify an AC-box-specific factor from human cell lines have estimated that this factor is 20 times less abundant than Sp1 (N. Heintz, pers. comm.).

In an attempt to reproduce the H1-SF band shift, crude nuclear extracts were prepared from AEV ts34 erythroblasts. Mobility shift assays of these extracts using the 01-40mer probe detected a band shift, 40-MS-e, with a similar mobility to the H1-SF band shift. Like H1-SF, 40-MS-e was highly sequence specific. 40-MS-e also eluted in the same salt wash (0.3 M) from a Heparin-Sepharose column as H1-SF.

However, two major discrepancies between this factor and H1-SF became evident. The first was that 40-MS-e appeared not to bind the 01-40mer probe, but instead bound
Figure 3.20 Analysis of ts34 crude nuclear extracts with 25mer probes

(a) 5 µg of extract was incubated with single stranded or double stranded 25mer probes in the presence of 1000-fold excess (by mass) of poly(dIdC). The probes used in each assay are as follows; Lane 1: labelled 01-25-t. Lane 2: labelled 02-25-t. Lane 3: labelled Δ01-25-t. Lane 4: labelled 01-25mer. Lane 5: labelled 02-25mer. Lane 6: labelled Δ01-25mer. The position of 25-MS-1 is indicated with an arrow.

(b) 5 µg of extract was incubated with 02-25mer probe in the presence of 1000-fold excess (by mass) of poly(dIdC) and excess unlabelled oligonucleotide. The fold-excess and type of oligonucleotide competitor is indicated below the corresponding lane. Lane 1 contains poly(dIdC) competitor only. The 25-MS-1 band shift is indicated with an arrow.
Chapter Three: Toward Purification of AC-box-binding Factors from Chicken

Figure 3.21 The AC-box is essential for the 25-MS-1 band shift

(a) 5 μg of ts34 crude nuclear extract was incubated with one of each of the 25mer probes in the presence of 1000-fold excess (by mass) of poly(dIdC). The probe used is indicated below the corresponding lane.

(b) 5 μg of extract was incubated with either 02-25mer (lanes 1 and 2) or Δ02-25mer (lanes 3 and 4) in the presence of increasing amounts of poly(dIdC). The fold-excess of poly(dIdC) is indicated below the appropriate lane.

(c) 5 μg of extract was incubated with the 02-25mer probe in the presence of 1000-fold excess (by mass) of poly(dIdC) and increasing amounts of unlabelled 02-25mer (lanes 2 and 3) or unlabelled Δ02-25mer (lanes 4 and 5). The fold-excess of oligonucleotide competitor is indicated below the appropriate lane. Lane 1 contains poly(dIdC) competitor alone.

These experiments uncovered a low mobility doublet of band shifts (25-MS-2) which may also have been due to AC-box-specific interactions. However, previous experiments (see Figure 3.20 (b)) had indicated that these band shifts were produced by non-specific interactions.
Figure 3.22 25-MS-1 is found in several different chicken cell lines

(a) 5 µg of CEF 38 crude nuclear extract was incubated with one of each of the 25mer probes in the presence of 1000-fold excess (by mass) of poly(dIdC). The probe used is indicated below the corresponding lane. Odd numbered lanes contain ts34 extract (5 µg) as a control; even numbered lanes contain CEF 38 extract. This result was repeatable with the HD11 and LMH extracts (not shown).

(b) 5 µg of HD11 crude nuclear extract was incubated with 02-25mer probe in the presence of various non specific competitors. The type and fold-excess of competitor is indicated below each lane. This result was repeatable with the CEF 38 and LMH extracts (not shown).
single stranded 01-40-t, which was always present at low levels in the mobility shift assays. H1-SF was proposed to bind duplex DNA (S. Dalton, 1987). In light of the original method of probe preparation, in which significant levels of single stranded probe were prone to occur, the possibility that the H1-SF band shift had arisen from an interaction with single stranded DNA cannot be ruled out. Consistent with this hypothesis, the H1-SF band shift was detectable after relatively short exposure times, whereas the appearance of the 40-MS-e band shift, in which the amount of single stranded probe was minimized, generally required much longer exposure times. Based on these observations, H1-SF and 40-MS-e may still have been analogous factors.

Some DNA binding proteins have been reported to preferentially interact with one strand of the DNA binding site. The sterol regulatory element (SRE) is found in the promoters of several sterol-regulated genes, and is essential for sterol-mediated repression of transcription. Two proteins, CNBP (19 kD) and SRE-BF (45-49 kD) each bound the SRE in a sequence- and strand-specific manner (Rajavashisth et al, 1989; Stark et al, 1992). This binding specificity correlated with in vivo function; point mutations which decreased or were predicted to decrease sterol responsiveness in vivo abolished binding to the single stranded SRE.

Another cis-acting element, the oestrogen responsive element (ERE), was also specifically bound on one strand by the oestrogen receptor (Lannigan and Notides, 1989). This finding is surprising considering the ERE is largely palindromic. In another report, purified human oestrogen receptor could not form a complex with a double stranded oligonucleotide containing the ERE until provided with a single stranded binding protein from yeast (Mukherjee and Chambon, 1990).

Other proteins implicated in binding to single stranded DNA promoter elements include MyoD (Santoro et al, 1991) and the yeast α1 and MCM1 proteins (Grayhack, 1992).

The second discrepancy between H1-SF and 40-MS-e was difficult to explain. Incubation of chicken embryo extracts with the Δ01-40mer probe gave rise to a band shift, Δ40-MS-e, with identical mobility to the 40-MS-e band shift. This was highly suggestive that the two band shifts were produced by the same factor. The Δ40-MS-e band shift was
more intense than the 40-MS-e band shift under the same conditions, implying that this factor bound the Δ01-40mer probe more avidly than the 01-40mer probe. In contrast, the 40-MS-e band shift was not competed out by excess unlabelled Δ01-40mer. These conflicting results could be explained if the 01- and Δ01-40mers were bound by distinct factors. The increased intensity of the Δ40-MS-e band shift could therefore be due to a higher affinity of the Δ01-40mer binding factor for its respective probe, and/or that this factor was more abundant than 40-MS-e. However, the added requirement that these factors have identical electrophoretic mobilities makes this a highly unlikely possibility. In addition with the disagreement between these and previous results (Dalton, 1987) regarding the 01- and Δ01-40mers, it was concluded that 40-MS-e was not analogous to H1-SF.

Although an AC-box-specific factor could not be identified with the 01-40mer in chicken embryo crude nuclear extracts, a band shift, 14-MS-b, was detected with a shorter probe (01-14mer) which only contained the core AC-box flanked by cohesive ends. This band shift was produced by a specific interaction with double stranded probe, despite the fact that it was competed out by a relatively low excess of non specific competitors such as poly(dIdC). It was possible that the 01-14mer was too short for even an AC-box-specific binding factor to form a stable interaction.

Although it displayed some specificity for the AC-box sequence, the binding of 14-MS-b to several different sequences was more suggestive of a general DNA binding protein than a sequence-specific transcription factor.

Chicken embryo crude nuclear extracts were prepared and purified on Heparin-Sepharose and Sephacryl S-300 columns (in that order) on a large scale. Fractions containing 14-MS-b were pooled and concentrated. Concatenated 01-14mer was used as a ligand in DNA affinity micro precipitation, and later in DNA affinity chromatography of these partially purified extracts. Mobility shift assays of the eluates from the optimised DNA affinity experiments revealed a 40% recovery of 14-MS-b from Heparin-Sepharose-purified extracts. This was comparable with the literature for one round of purification (Briggs et al, 1986; Mitchell et al, 1987; Fletcher et al, 1987).

Analysis of the DNA affinity purified extracts on SDS polyacrylamide gels revealed two enriched proteins of 86 kD and 70 kD. This enrichment was increased by a second
round of DNA affinity purification. In order to determine whether one or both of these proteins were 14-MS-b, attempts were made to renature this factor from SDS polyacrylamide gels. A renatured factor which displayed all the characteristics of 14-MS-b in mobility shift assays was found to be no larger than 30 kD. This was much smaller than the estimated mass (90 kD) of renatured H1-SF (S. Dalton, pers. comm.). In addition, there is indirect evidence that HITF1, a factor isolated in HeLa nuclear extracts which specifically interacted with the AC-box, was also 90 kD (Gallinari et al, 1989). For these reasons, it was concluded that 14-MS-b was not analogous to H1-SF. This was supported by the finding that no band shifts were detected in mobility shift assays of ts34 crude nuclear extracts with the 01-14mer probe.

Mobility shift assays of the DNA affinity purified extracts with the 01-40mer probe detected a band shift (40-MS-1) with very similar mobility to 40-MS-e. This band shift was produced by a highly specific interaction. Renaturation experiments from SDS gels identified a factor of 42-46 kD that produced a band shift with identical mobility to 40-MS-1. Again, this was considered too small to be equivalent to H1-SF.

In a final effort to determine whether the proteins enriched in DNA affinity purified extracts were AC-box-specific factors, 25 bp oligonucleotides, the 01-25mer and 02-25mer, were designed containing the AC-box and flanking sequences from the 01 H1 and 02 H1 promoters, respectively. Mobility shift assays of chicken embryo extracts with these probes gave rise to several band shifts, but all of these band shifts were also detected with an 01-25mer probe in which the AC-box had been mutated (Δ01-25mer). In addition, partially purified chicken embryo extract was applied to each of three DNA affinity chromatography columns in which the ligand was either the 01-25mer, 02-25mer, or Δ01-25mer in concatenated form. The eluate fractions from each column, when electrophoresed on SDS polyacrylamide gels, were found to be indistinguishable. Thus no AC-box-specific factor was detected with the 25mers either by mobility shift assay or DNA affinity chromatography. As the 86 kD and 70 kD proteins were enriched in the DNA affinity eluates from all three columns, it appeared that these proteins were non specific DNA binding proteins.
Mobility shift assays of ts34 crude nuclear extracts with the 25mer probes revealed a highly specific band shift (02-25-1) unique to the 02-25mer. The absence of the 02-25-1 band shift in similar assays with the 01-25mer probe suggested that this band shift had arisen from an interaction with the sequence flanking the AC-box in the 02-25mer probe (recall that the flanking sequences vary significantly between the 01- and 02-25mers). However, an 02-25mer probe in which the AC-box had been mutated (Δ02-25mer) also failed to produce this band shift. This implied that sequences both within and flanking the AC-box of the 02-25mer were essential for this interaction.

This result was not restricted to ts34 extracts. The same selective interaction with the 02-25mer probe was also observed in extracts from a number of chicken cell lines (macrophage, hepatoma, and fibroblast), suggesting this phenomenon was not cell type specific. Therefore the absence of 02-25-1 from chicken embryo extracts was puzzling. One possible explanation was that the relative abundance of 02-25-1 in chicken embryo extracts was lower than in the chicken cell lines.

If 02-25-1 was a true AC-box binding factor, then the selective interaction with the 02-25mer and not with the 01-25mer may reflect different levels of expression of these genes. It is known that the protein and transcript levels of the H1 subtypes vary in different tissues (see 1.10). Evidence in this laboratory indicated that the transcript levels from the 02 H1 gene were ~5-fold higher than those from the 01 H1 gene in ts34 cells (L. Tabe and R. King, unpublished results). If this variation was due to different rates of transcription initiation, then it was possible that a factor such as 02-25-1 could play a role in the initiation rate through its selective binding to the 02 H1 promoter. However, because this selective binding was identical in several different cell types, it seemed unlikely that 02-25-1 was involved with the variation in the relative amounts of H1 subtypes in different tissues.

The fact that this band shift was not detected in CHO nor HeLa crude nuclear extracts implied that 02-25-1 was an avian-specific factor. Alternatively, an equivalent factor in the mammalian extracts may have diverged to a point where it could not bind the chicken AC-box and flanking sequences. This seemed unlikely for an AC-box-binding factor, as the AC-box is highly conserved throughout higher eukaryotes (Coles and Wells, 1985; Wells,
In addition, H1-SF had been detected in both ts34 and HeLa extracts (Dalton and Wells, 1988a).

In the final analysis, none of the oligonucleotides used in this study conclusively demonstrated the presence of an AC-box-specific binding factor analogous to H1-SF in nuclear extracts. 40-MS-e, detected in mobility shift assays of ts34 extracts with the 01-40mer probe, was rejected as an H1-SF candidate because the results obtained with the Δ01-40mer probe were incompatible with previous results (Dalton, 1987). 40-MS-1, which was detected by the 01-40mer probe upon incubation with DNA affinity-purified extracts, was found to be much smaller than H1-SF. Also found to be significantly smaller than H1-SF was 14-MS-b, which bound the 01-14mer probe to give rise to the 14-MS-b band shift. In addition, no 01-14mer binding activity was detected in ts34 nuclear extracts. Finally, the only factor which specifically bound the AC-box (and flanking sequences) within the 02-25mer probe did not bind another probe (01-25mer) which also contained the AC-box. It should be noted that characterisation of an apparently AC-box-specific factor from human nuclear extracts was not pursued because it did not bind some AC-box-containing oligonucleotides (Heintz, 1991).
CHAPTER FOUR

FUNCTIONAL ANALYSES OF THE

AC-BOX AND A RELATED 

H1-SPECIFIC PROMOTER 

ELEMENT
Chapter Four: Functional Analyses of the AC-box and a Related H1-Specific Promoter Element

4.1 Introduction

In light of the difficulties encountered with the attempts to purify a specific AC-box-binding factor, a new approach toward understanding the role of the AC-box was adopted. As previously mentioned, deletion or mutation of the AC-box sequence had significant to dramatic effects upon H1 gene transcription, including either a decrease or abolition of the S-phase-specific activation of the H1 promoter (Dalton and Wells, 1988a; La Bella et al, 1989). This disruption to normal S-phase regulation may be a direct consequence of the loss of the AC-box sequence, or it may reflect the disruption of a cooperative interaction between the AC-box and another S-phase-regulatory element. It was desirable to know whether the AC-box was sufficient to direct S-phase-specific transcriptional activation. One way to test this would be to place the AC-box within a normally non-S-phase regulated promoter and to determine whether S-phase regulation was subsequently conferred to that promoter. Short, conserved sequence elements from S. cerevisiae histone promoters are sufficient to direct cell cycle regulation of formerly constitutive genes (Osley et al, 1986). Similar results have been obtained with a sequence from the yeast thymidylate synthase promoter, which contained a hexamer (5' ACGCGT 3': "Mlu I motif") conserved in several cell cycle regulated yeast genes (McIntosh et al, 1991). However, a promoter element with this capability has not been described for the histone genes of higher eukaryotes.

This chapter describes the strategies employed in determining the effect of the AC-box, in single and multiple copies, upon the transcription of a heterologous promoter. While these experiments were in progress, a report concerning a distal element in the H10 promoter which produced a dramatic transcriptional effect (Steuer et al, 1992) prompted a comparison between several distal H1 promoter regions. A computer alignment revealed a highly conserved motif ~450-480 bp upstream from the cap site of all the H1 genes analysed. The effect of this motif upon transcription from the chicken 02 H1 promoter was also investigated.
4.2 Effect of the AC-box in a non-S-phase regulated promoter

4.2.1 Choice of a suitable reporter system

The following experiments were designed to answer two questions regarding the AC-box: 1) can the AC-box increase transcription from a heterologous promoter?, and 2) can the AC-box direct S-phase activation of a heterologous promoter? The first question, and possibly the second, could be addressed using transient transfection assays of a reporter construct into tissue culture cells. Because the AC-box sequences were derived from chicken H1 histone genes, it was felt that these experiments should initially be carried out in an actively dividing chicken cell line. For this reason, the chicken macrophage line HD11 was adopted. Because this was an attached cell line, stable transformants could be selected if it was required to establish permanent lines for cell cycle analysis.

The herpes simplex thymidine kinase promoter (HSVtk) was selected to monitor the effect of the AC-box on transcription. There were several reasons for this choice. Firstly, HSVtk is a small, well characterised promoter. The regions important for expression from this promoter have been defined by linker scanning analysis (McKnight and Kingsbury, 1982). Secondly, it is a basally expressed promoter in which even a small enhancement of transcription by a putative regulatory element can be easily detected. Importantly, it is not normally S-phase regulated.

The reporter plasmid used in initial experiments was pTKGH (Selden et al, 1986). This vector contained a 250 bp fragment of the HSVtk promoter spanning bases -196 to +57 (Figure 4.1 (a)). The reporter gene, encoding human growth hormone (hGH), was chosen because when expressed, the hormone was secreted into the medium. This enabled HSVtk promoter activity to be assayed without the need for preparation of cellular extracts. Expression of hGH was not expected to affect the chicken cell line (J. Beale, pers. comm.).

While the AC-box/HSVtk promoter fusions were being constructed, trial transfections of pTKGH were carried out to ascertain that this vector was capable of producing detectable levels of hGH in HD11 cells. Despite repeated transfections by electroporation or calcium phosphate precipitation with either 5, 10, or 20 μg of pTKGH DNA, no hGH was detected. Thus HD11 appeared unsuitable for examination of HSVtk
Figure 4.1 AC-box/pTKGH series

(a) A diagram of the reporter vector pTKGH is shown, indicating the insertion site of the AC-box oligonucleotides. Cloning of the 25mer oligonucleotides was achieved by opening the vector with Hind III, end-filling with Klenow, and adding the appropriate oligonucleotide in a blunt-ended ligation. The resultant constructs are tabulated in (b). 01 and 02 refer to the 01-25mer and 02-25mer (Chapter 3), respectively, and \( n \) refers to the copy number of the particular oligonucleotide in the construct. Also listed are constructs containing \( \Delta 01 \)- or \( \Delta 02 \)-25mers.
(AC-box)$_n$

p(AC)TKGII

p(4C01)rKGH

p(4C02)rKGH

p(4C02)RTKGH

p(4C02)2TKGH

p(4C02)3rKGH

hGH

Eco RI

(b) p(AC)TKGH series

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>p(ΔAC)TKGH series</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(AC01)TKGH</td>
<td>1</td>
<td>p(ΔAC01)TKGH</td>
</tr>
<tr>
<td>p(AC02)TKGH</td>
<td>1</td>
<td>p(ΔAC02)TKGH</td>
</tr>
<tr>
<td>p(AC02)RTKGH</td>
<td>1 (reversed)</td>
<td></td>
</tr>
<tr>
<td>p(AC02)2TKGH</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>p(AC02)3TKGH</td>
<td>3</td>
<td>p(ΔAC02)3TKGH</td>
</tr>
<tr>
<td>p(AC02)6TKGH</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
promoter activity in pTKGH. To verify that experimental technique was sound, CHO-K1 cells were also transfected with pTKGH. Detectable levels of hGH had been previously produced from this line after transient transfection with pTKGH (P. McKinnon, pers. comm.). Culture medium isolated from CHO cells 72 hours after electroporation with pTKGH contained significant levels of hGH (Table 4.1). Therefore, the effect of the AC-box on HSVtk promoter expression was initially studied in CHO cells.

### 4.2.2 Construction of AC-box/HSVtk promoter fusions in pTKGH

To clone one or more copies of the AC-box into pTKGH, the same 25 bp oligonucleotides that were employed in mobility shift assays (Chapter 3) were ligated into the end-filled Hind III site immediately upstream of the HSVtk promoter (Figure 4.1 (a)). Initially, the double stranded oligonucleotides were concatenated and purified from agarose gels before ligating into pTKGH. Although the concatenation was successful, the subsequent ligation into the vector was not (data not shown). Therefore, the unconcatenated oligonucleotides were simply ligated directly to the vector. Four types of recombinants were constructed, containing either the 01-25mer, Δ01-25mer, 02-25mer, or Δ02-25mer oligonucleotides. Because the ligations were blunt-ended, there was no control over the direction (or number) in which the oligonucleotides were cloned. Therefore, the recombinants were sequenced to determine the number and orientation of cloned oligonucleotides.

In this manner, several useful recombinants were obtained for functional analysis. The nomenclature for these recombinants is listed in Figure 4.1 (b). These include four recombinants, each containing one of the 25mers in the 5' to 3' orientation with respect to the AC-box sequence. In the event that a single copy of the AC-box was not sufficient to produce a detectable transcriptional effect, recombinants containing two and three tandem copies of the 02-25mers, in the required orientation, were also selected. The largest useful recombinant isolated contained an insertion of seven 02-25mers: the 5' oligonucleotide was in the reverse orientation with respect to the AC-box, but the remaining six were all in the correct orientation.
4.2.3 Analysis of AC-box/HSVtk promoter activity in p(AC)TKGH series

It was assumed that in a population of randomly dividing cells, about \( \frac{1}{3} \)rd would be in S-phase at any one time. Thus if the transient transfections indicated that the AC-box increased transcription from the HSVtk promoter, this increase may represent a general transcriptional effect or an S-phase-specific activation of transcription. Further investigation with synchronised cell populations would then determine whether either or both of these possibilities were true. On the other hand, if the AC-box produced no effect on transcription in transient assays, then it would almost certainly be insufficient to direct S-phase-specific activation. Therefore transient transfections of unsynchronised cells should provide an early indication of the effect of the AC-box.

Recombinants p(AC02)TKGH, p(AC02)2TKGH, and p(AC02)RTKGH, contained one copy, two copies and one reversed copy of the 02-25mers, respectively. Along with a pTKGH control, 5 µg of each of these constructs were electroporated into CHO cells and AEV ts34 cells. Units of hGH in culture medium 72 hours after transfection are shown in Table 4.2. The most consistent finding was that the highest producers of hGH in both CHO and ts34 cells were those transfected with p(ΔAC02)TKGH (one copy of the AC-box). Cells transfected with pTKGH produced between 46 and 75% of the amount of hGH produced by cells transfected with p(AC02)TKGH. One problem apparent in these experiments was that CHO cells, formerly producing high hGH levels upon transfection, were now producing low levels. Because of this inconsistency, it was decided to use ts34s in all future transfections.

Triplicate samples of ts34 cells were each transfected with 5 µg of either pTKGH or p(AC02)TKGH. This experiment demonstrated that one copy of the 02-25mer, placed directly 5′ to the HSVtk promoter, caused approximately a three-fold increase in the production of hGH (Figure 4.2 (a)).

To confirm that the apparent increase in HSVtk promoter activity in p(AC02)TKGH was caused by the AC-box, a similar experiment was conducted in which triplicate samples of ts34 cells were each transfected with 5 µg of either pTKGH or p(ΔAC02)TKGH. In addition, transfections with 5 µg of p(AC02)RTKGH were carried out to investigate whether any effects of the AC-box upon transcription were orientation-specific.
Table 4.1 hGH production in pTKGH-transfected cells

The table shows the levels of hGH in culture media from HD-11 and CHO-K1 cells 72 hours after transfection with either 0, 5, or 10 µg of pTKGH. Note that for the CHO cells, transfection with 5 µg of pTKGH gave rise to a higher level of hGH than transfection with 10 µg of pTKGH. This result correlates with the observed dose-response of L cells to transfection with increasing amounts of this reporter plasmid (Selden et al, 1986).

Table 4.2 hGH production in cells transfected with p(AC)TKGH constructs

CHO-K1 and ts34 cells were transfected with 5 µg of either pTKGH, p(AC02)TKGH, p(AC02)2TKGH, or p(AC02)RTKGH. Levels of hGH in the culture medium 72 hours after transfection are indicated.
### 4.1

<table>
<thead>
<tr>
<th>Cell line</th>
<th>HD-11</th>
<th>CHO-K1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfected pTKGH (µg)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>hGH (µU/ml)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### 4.2

<table>
<thead>
<tr>
<th>Cell line</th>
<th>CHO-K1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfected construct</td>
<td>no DNA</td>
</tr>
<tr>
<td>hGH (µU/ml)</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cell line</th>
</tr>
</thead>
<tbody>
<tr>
<td>ts34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transfected construct</th>
<th>no DNA</th>
<th>pTKGH</th>
<th>p(AC02)TKGH</th>
<th>p(AC02)RTKGH</th>
<th>p(AC02)2TKGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>hGH (µU/ml)</td>
<td>0</td>
<td>27</td>
<td>59</td>
<td>35</td>
<td>23</td>
</tr>
</tbody>
</table>
Figure 4.2 hGH production in ts34 cells transfected with p(AC)TKGH constructs

(a) ts34 cells were transfected with 5 μg of either pTKGH or p(AC02)TKGH, and the levels of hGH in the culture medium 72 hours after transfection were measured. The mean hGH levels (+ standard deviations) from three experiments are indicated.

(b) A similar series of transfections were carried out as depicted in (a), except that the constructs used were pTKGH, p(AC02)RTKGH and p(ΔAC02)TKGH. Again, the mean hGH levels (+ standard deviations) from three experiments are shown.
Cells transfected with p(AC02)RTKGH produced ~3-fold higher levels of hGH than those transfected with pTKGH (Figure 4.2 (b)). This was about the same increase as that found for p(AC02)TKGH, which suggested that the effect of the AC-box-containing 02-25mer was orientation-independent. However, there was approximately a five-fold increase in the production of hGH from cells transfected with p(ΔAC02)TKGH relative to those transfected with pTKGH (Figure 4.2 (b)). It was possible that the mutated AC-box in p(ΔAC02)TKGH was a fortuitous binding site for a transcriptional activator. Nevertheless, the effect of the AC-box on transcription could not be conclusively defined in these experiments.

4.2.4 Construction of the p(AC)TKCAT series

In order to confirm what was possibly a positive effect on transcription by the AC-box, the AC-box/HSVtk promoter fusions were resected from the pTKGH vectors and cloned into pBLCAT.2 (Luckow and Schütz, 1987). This plasmid contains a CAT reporter gene under the control of the HSVtk promoter. The promoter fragment in pBLCAT.2 is shorter (~150 bp) than that of pTKGH, but it still contains all the elements required for basal level transcription (McKnight and Kingsbury, 1982). Quantitation of CAT activity is a widely accepted method for analysis of transcriptional activity in eukaryote systems, including the study of several histone promoters and promoter elements (Breuer et al, 1989; Hwang and Chae, 1989; Lai et al, 1989).

Prior to cloning the AC-box/HSVtk promoter fusions into pBLCAT.2, a transcription unit containing the neomycin phosphotransferase (neo) coding sequence under the control of phosphoglycerate kinase (pgk) promoter and polyadenylation sequences was cloned into the Sac I site at the 3' end of the CAT gene. This would allow the selection of stable transformants if it was required to establish permanent lines. A recombinant was selected in which the direction of transcription of the neo gene was the same as that of the CAT gene, as it was felt that convergently transcribing promoters may cause some masking of CAT expression. This recombinant was known as pBLCAT.N.

Both pTKGH and pBLCAT.N were derived from the same series of pUC plasmids. Advantage was taken of unique Nde I and Mlu I sites in the plasmid and HSVtk promoter
sequences, respectively, to transfer the AC-box/HSVtk promoter fusions from pTKGH to pBLCAT.N (Figure 4.3).

4.2.5 Analysis of AC-box/HSVtk promoter activity in p(AC)TKCAT.N series

pTKCAT.N, p(AC02)TKCAT.N, p(AC02)6TKCAT.N, and p(ΔAC02)TKCAT.N, which contained the HSVtk promoter alone, one copy and six copies of the 02-25mer, and one copy of the Δ02-25mer, respectively, were transfected into ts34 cells. pTKCAT.N differed from pBLCAT.N in that it contained the larger fragment of the HSVtk promoter derived from pTKGH; the same promoter fragment in the p(AC)TKCAT.N constructs.

The CAT activities of extracts from each transfection are depicted in Figure 4.4 (a). These data indicated that the activity of the HSVtk promoter was unaffected by one copy or six copies of the 02-25mer. A similar experiment indicated that there was no significant difference between the CAT activities of extracts from ts34 cells which were transfected with p(AC02)3TKCAT.N and p(ΔAC02)3TKCAT.N, which contained three copies each of the 02-25mer and Δ02-25mer, respectively (data not shown). To test the possibility that the sequences flanking the AC-box in the 02-25mers were somehow masking AC-box function, the experiment was repeated with constructs that contained one copy each of the 01-25mer and Δ01-25mer (p(AC01)TKCAT.N and p(ΔAC01)TKCAT.N). The sequences flanking the AC-box in these oligonucleotides are considerably different from those in the 02-25mers. Again, there was no significant difference between the CAT activities of extracts from cells transfected with either construct (data not shown). From these experiments it was concluded that the AC-box alone was not sufficient to affect transcription of the HSVtk promoter in transient transfections of ts34 cells.

The % conversion of chloramphenicol from unacytlated to acetylated forms was very low (<1%) in these experiments. It was possible that the strong pgk promoter, driving the neo gene, was interfering with transcription from the moderate HSVtk promoter. This interference may have blocked transcription from the latter promoter, and the low levels of CAT activity could be derived from read-through transcription initiated at the pgk promoter. Thus it was necessary to examine the CAT activity of cells transfected with AC-box/HSVtk.
Figure 4.3 Construction of p(AC)TKCAT series

The strategy for construction of this series is shown. pBLCAT.N and the p(AC)TKGH vectors were each digested with Nde I and Mlu I, and the AC-box-containing Nde I / Mlu I fragments from the p(AC)TKGH vectors were ligated into pBLCAT.N. N.B. The pBLCAT.2 vector shown in this diagram actually included a 1.8 kb fragment which contained the neo coding sequence flanked by the pgk promoter and polyadenylation sequences. This fragment was cloned into the Sac I site at the 3' end of the CAT gene, in the vicinity of the Eco RI site indicated, to create pBLCAT.N. For simplicity, this fragment has been excluded from the diagram.
Figure 4.4 CAT activity in ts34 cells transfected with p(AC)TKCAT.N constructs

(a) ts34 cells were transfected with 15 µg of either pTKCAT.N, p(AC02)TKCAT.N, p(AC02)6TKCAT.N, or p(ΔAC02)TKCAT.N. 72 hours after transfection, protein extracts were prepared from the harvested cells and the extracts were assayed for CAT activity. For each construct, the mean activity (expressed as % conversion from non-acetylated to acetylated 14C-chloramphenicol) from three experiments is shown. Error bars denote the standard deviation in these experiments.

(b) 15 µg of either pBLCAT.2, pTKCAT, p(AC02)3TKCAT, or p(ΔAC02)3TKCAT were transfected into ts34 cells and extracts were prepared 72 hours after transfection. CAT activities in extracts from three such experiments are expressed as in Figure 4.4 (a).
**a**

![Graph](image)

**b**

![Graph](image)
promoter constructs which did not contain the pgk-neo unit. Three such constructs (pTKCAT, p(AC02)3TKCAT, and p(AAC02)3TKCAT) were used to transfect ts34 cells, extracts from which were assayed for CAT activity (Figure 4.4 (b)). The first finding was that the CAT activity of these extracts was about an order of magnitude higher than those prepared from cells transfected with the p(AC)TKCAT.N vectors. This suggested that the pgk promoter had indeed affected the expression of the HSVtk promoter.

Another finding in this experiment was that three tandem copies of the AC-box did not positively affect transcription from the HSVtk promoter. In fact, the CAT activities of extracts from cells transfected with p(AC02)3TKCAT (3 copies of the AC-box 5' to the HSVtk promoter) were ~50% lower than those of extracts from cells transfected with pTKCAT (HSVtk promoter alone). However, further investigation of the possible negative effect of the AC-box on transcription from the HSVtk promoter were postponed in light of a discovery which was likely to have a bearing on this study (see below).

4.3 Investigation of a Distal Element in the H1 Histone Promoter

During the course of the AC-box/HSVtk promoter fusion experiments, a report appeared in the literature concerning an 80 bp region in the distal H10 promoter. Deletion of this region produced a dramatic decrease in transcription (Steuer et al., 1992). A similar effect by a distal region had been implicated in studies of the mouse H1t promoter (Kremer and Kistler, 1992). Although these genes are replication independent, they share several common promoter elements with the replication dependent H1 promoters, including the AC-box. In addition, there is evidence that the H10 and H1t promoters are cell cycle regulated (see 1.9.2). It was therefore interesting to know if such a distal element also existed in the major H1 promoters.
4.3.1 Comparison of H1 Histone Promoter Sequences

The promoters (for which several hundred bp of upstream sequence was known) of major and variant H1 genes from several species were aligned with the aid of the GCG Pileup program (Figure 4.5). Apart from the previously characterised promoter elements, a region of extremely high homology was situated -450 to 480 bp upstream from the cap site in all of the promoters examined. This region contained a 7 bp sequence, 5' TGTGTTA 3', which is a near perfect inverted repeat of the AC-box (the perfect inverted repeat is 5' TGTGTTT 3'); hence, I have named this element the "TG-box".

Along with the AC-box, the TG-box was one of the most highly conserved elements among the thirteen H1 promoters analysed (including 02 H1). With two exceptions (discussed below), the sequence of the TG-box in major and variant H1 promoters was invariant throughout higher eukaryotes. Interestingly, this conservation included the 3' "A" of the TG-box; the only nucleotide that did not form part of the inverted repeat. One exception was in the sea urchin (S. purpuratus) late histone H1-β promoter (Lai and Childs, 1988), in which a "T" occupied the 3' position. Remarkably, the AC-box of this promoter had a "T" instead of an "A" at the 5' position, thus preserving, at the same position, the single nucleotide that disrupted the inverted repeat. The second exception was in the H1A promoter from the X. laevis X1h3 histone cluster (Perry et al, 1985). This promoter contained a variant AC-box (5' AAGCACA 3'). Once again, a compensatory change was found in the TG-box sequence (5' TGTGCTA 3') to maintain the inverted repeat.

As well as the sequence of the TG-box and the AC-box, the distance between these two elements was also highly conserved. Of the thirteen H1 promoter sequences examined, eight contained 351-354 bp between the 5' ends of each element. The exceptions were the human and mouse H10 promoters (339 and 336 bp, respectively), the human H1.2 promoter (359 bp), the Xenopus H1A promoter (342 bp) and the sea urchin late H1-β promoter (344 bp).

It is interesting to note that the TG-box of the mouse H10 promoter maps within the 80 bp region essential for basal level expression of this promoter (Steuer et al, 1992).
Figure 4.5 Identification of an H1-specific promoter element

The promoter sequences of several H1 histone genes were aligned using the GCG Pileup program (gap weight 5; gap length weight 0.3). Predicted cap sites are located at the extreme 3' ends of the sequences, and the distances (in bases) upstream from these sites are shown at the right of each row. For simplicity, only highly conserved regions of homology are boxed. The TG-box (see text) is the most distal of these conserved regions. References and/or GenBank accession numbers are listed below the alignment. Also shown are the distances (in bases) between the 5' ends of the TG- and AC-boxes of each promoter.
<table>
<thead>
<tr>
<th>Species</th>
<th>Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicken (02 H1)</td>
<td>-359 AAGGCCCGCG CAGACACGCG TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>Duck (DH81)</td>
<td>-359 AAGGAAAAAA CAGAAAGAG TGCCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>Chicken (01 H1)</td>
<td>-359 AAAACATGC CGAAAGACG TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>Human H1.2</td>
<td>-359 ACCATCCAG GACCTGGCC TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>Human H1.4</td>
<td>-359 ATGCTGGCC TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>Human FNC16</td>
<td>-359 ATGCTGGCC TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>Human H1t</td>
<td>-359 ATGCTGGCC TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>Rat H1t</td>
<td>-359 ATGCTGGCC TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>Mouse H1o</td>
<td>-359 ATGCTGGCC TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>Mouse H1o° 3141</td>
<td>-359 ATGCTGGCC TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>X. laevis H1B</td>
<td>-359 ATGCTGGCC TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>X. laevis H1A</td>
<td>-359 ATGCTGGCC TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
<tr>
<td>Sea urchin (λSpHl-1)</td>
<td>-359 ATGCTGGCC TGTTGCTGGG GAGTGGCCTG CAGACACGCG TGGTCCTGGG CAGACACGCG</td>
</tr>
</tbody>
</table>

**Table Entries:**

- **Chicken (02 H1)**: Accession numbers for chicken, species-specific DNA sequences.
- **Duck (DH81)**: Accession numbers for duck, species-specific DNA sequences.
- **Chicken (01 H1)**: Accession numbers for chicken, species-specific DNA sequences.
- **Human H1.2**: Accession numbers for human, species-specific DNA sequences.
- **Human H1.4**: Accession numbers for human, species-specific DNA sequences.
- **Human FNC16**: Accession numbers for human, species-specific DNA sequences.
- **Human H1t**: Accession numbers for human, species-specific DNA sequences.
- **Rat H1t**: Accession numbers for rat, species-specific DNA sequences.
- **Mouse H1o**: Accession numbers for mouse, species-specific DNA sequences.
- **Mouse H1o° 3141**: Accession numbers for mouse, species-specific DNA sequences.
- **X. laevis H1B**: Accession numbers for xantid, species-specific DNA sequences.
- **X. laevis H1A**: Accession numbers for xantid, species-specific DNA sequences.
- **Sea urchin (λSpHl-1)**: Accession numbers for sea urchin, species-specific DNA sequences.

**Accession Numbers Description:**

- **AAAGCAAGGA**: Accession number for given species-specific DNA sequence.
- **GAGACCGGCG**: Accession number for given species-specific DNA sequence.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Distance (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicken (02 H1)</td>
<td>352</td>
<td>This thesis</td>
</tr>
<tr>
<td>Duck (DH81)</td>
<td>352</td>
<td>(Tönjes and Doenecke, 1987)</td>
</tr>
<tr>
<td>Chicken (01 H1)</td>
<td>352</td>
<td>(Coles and Wells, 1985)</td>
</tr>
<tr>
<td>Human H1.2</td>
<td>359</td>
<td>X57129 (Eick et al, 1989)</td>
</tr>
<tr>
<td>Human H1.4</td>
<td>355</td>
<td>M60748 (Albig et al, 1991)</td>
</tr>
<tr>
<td>Human FNC16</td>
<td>354</td>
<td>(van Wijnen et al, 1988)</td>
</tr>
<tr>
<td>Human H1t</td>
<td>353</td>
<td>M60094 (Drabant et al, 1991)</td>
</tr>
<tr>
<td>Rat H1t</td>
<td>351</td>
<td>M28409 (Wolfe and Grimes, 1986)</td>
</tr>
<tr>
<td>Human H1°</td>
<td>339</td>
<td>M87841 (Bouterfa et al, 1992)</td>
</tr>
<tr>
<td>Mouse H1° 3141</td>
<td>336</td>
<td>(Breuer et al, 1989)</td>
</tr>
<tr>
<td>X. laevis H1B</td>
<td>352</td>
<td>(Perry et al, 1985)</td>
</tr>
<tr>
<td>X. laevis H1A</td>
<td>342</td>
<td>(Perry et al, 1985)</td>
</tr>
<tr>
<td>S. purpuratus (λSplH1-5)</td>
<td>344</td>
<td>(Lai and Childs, 1988)</td>
</tr>
</tbody>
</table>
4.3.2 Sequencing of the Chicken 02 H1 Histone Promoter

The immediate aim was to determine whether the TG-box played a role in the transcription of a major H1 histone promoter. Of the six H1 genes of chicken, only 01 H1 had been sequenced extensively both 5' and 3' to the coding region (Coles and Wells, 1985). However, previous work in this laboratory (L. Tabe and R. King, unpublished) had indicated that the transcript levels of the 02 H1 gene in ts34 cells were much higher in S-phase than those of 01 H1. Other investigations had suggested that 01 H1 was not a fully replication dependent histone gene (Kirsh et al, 1989). The 02 H1 promoter was therefore selected for transcription analysis.

The isolation of the chicken genomic clone λCH02 has been described (Harvey et al, 1981). A 682 bp Kpn I / Sac I fragment was isolated from the subclone pCH4.7E (Coles, 1986). This fragment, which contained the 02 H1 promoter sequence from -771 to -90, was cloned into pBluescript SK+ and sequenced. The remainder of the promoter (-89 to +17) was sequenced in pUC.02-P (see below). The entire sequence of the chicken 02 H1 histone promoter is shown in Figure 4.6.

As expected, a TG-box was found in this promoter at -472, in an essentially identical location to the TG-box of the 01 H1 promoter. A second AC-box sequence at -527 in the 01 H1 promoter was absent from the 02 H1 promoter. The 02 H1 sequence has been included in the computer alignment in Figure 4.5.

A comparison between the promoter sequences of the 02 H1, 01 H1, and a duck H1 histone gene (Figure 4.6) indicated that the 02 H1 promoter was more homologous with the duck promoter (84%) than with the 01 H1 promoter (61%). This finding was in agreement with similar observations made by Tönjes and Doenecke (1987). The extremely high conservation in nucleotide sequence between the 02 H1 and duck promoters suggest that the two genes are homologues. This allowed the tentative placement of the extreme 5' end of these H1 genes at ~-640. Beyond this point, homology between the two genes decreases dramatically.
4.3.3 Construction of H1 Promoter / CAT Fusions

To examine the effect of the TG-box upon transcription, functional 02 H1 promoters which either included or lacked the TG-box were cloned into a CAT reporter plasmid. However, the palindromic relationship and conserved distance between the AC-box and the TG-box implied a relationship between these two elements. Therefore the intermediate products of the cloning strategy were designed with the long term aim of examining whether a promoter fragment containing both the TG- and AC-boxes (and the intervening sequence), but no other proximal H1 promoter sequences, could influence transcription from a heterologous promoter. If a transcriptional effect was observed, then deletion of the TG-box from the promoter fragment would demonstrate whether this element was required for this effect.

4.3.3 (a) Generation of p(02)TKCAT vectors

pCH4.7E was used as the template for the PCR-mediated generation of two overlapping, but slightly different fragments from the same region of the 02 H1 promoter. Three primers, 5'-02L (long), 5'-02S (short), and 3'-02 were designed for the two PCRs (2.2.8). Eco RI and Sal I sites were engineered into the 5' primers and the 3' primer, respectively, to facilitate cloning of the PCR products. In each reaction, either 5'-02-L or 5'-02S were used in combination with 3'-02 to amplify regions of the 02 H1 promoter from -493 to -100 and from -464 to -100, respectively. These two promoter fragments, CH02-L (395 bp) and CH02-S (367 bp), were identical except that CH02-S lacked 29 bp (which included the TG-box) from the 5' end. Both CH02-L and CH02-S contained the AC-box close to the 3' end.

These fragments were digested with Eco RI and Sal I and cloned between the Eco RI and Sal I sites of pUC19, generating pUC.02-L and pUC.02-S, respectively. The fidelity of the PCRs was confirmed by sequencing. The promoter fragments were then removed from pUC19 by digestion with Nde I and Sal I and cloned between the Nde I and Sal I sites of pBLCAT.2. The resultant constructs, p(02L)TKCAT and p(02S)TKCAT, contained the CH02-L and CH02-S promoter fragments, respectively, directly 5' to the HSVtk promoter (Figure 4.7). Also constructed were a similar pair of vectors, p(01L)TKCAT and
Figure 4.6 Comparison of the nucleotide sequence between three avian H1 histone promoters

The alignment of the chicken 02 H1, duck H1, and chicken 01 H1 promoter sequences is shown. As with Figure 4.5, the numbers to the right of each row depict the distance in bp upstream from the predicted cap sites (located at the extreme 3' ends of the sequences). Homologous bases between two or three of the sequences are boxed. Note that the duck promoter shares significantly more homology with the 02 H1 promoter than the 01 H1 promoter.
Chapter Four: Functional Analyses of the AC-box and a Related H1-Specific Promoter Element

Figure 4.7 Construction of H1 / HSV-tk promoter fusions

The plasmid pCH4.7E was used as the template to generate 395 bp (CH02-L) and 367 bp (CH02-S) fragments by PCR. The black regions denote the positions of the TG- and AC-boxes. These fragments were cloned between the Eco RI and Sal I sites of pUC19, creating pUC.02-L and pUC.02-S, respectively. The fidelity of the PCR was confirmed by sequencing. The Nde I / Sal I fragments from pUC.02-L and pUC.02-S were then cloned into the CAT reporter vector pBLCAT.2, creating p(02L)TKCAT and p(02S)TKCAT.
1. PCR
2. Eco RI / Sal I

1. Ligate into pUC19
2. Sequence

1. Nde I / Sal I
2. Ligate into pBLCAT.2

```
1. PCR
2. Eco RI / Sal I

1. Ligate into pUC19
2. Sequence

1. Nde I / Sal I
2. Ligate into pBLCAT.2
```
Figure 4.8. H1 promoter / CAT fusion vectors

(a) A 461 bp Cla I / Sau 3AI fragment of the 02 H1 promoter (-449 to +17) was removed from pUC.02-P and cloned into both p(02L)TKCAT and p(02S)TKCAT between the Cla I and Bgl II sites to create p(02L)HCAT and p(02S)HCAT, respectively.
Figure 4.8. H1 promoter / CAT fusion vectors

(b) A 397 bp Nde I / Cla I fragment from pUC.02-P was cloned into Nde I / Cla I-digested p(02S)HCAT to generate p(02P)HCAT, which contained an additional 151 bp of 02 H1 promoter sequence 5' to the TG-box.
Nde I / Cla I

ligate

Nde I / Cla I

p(02S)HCAT

pUC.02-P

p(02P)HCAT

Eco RI

Nde I

Cla I

Sau 3AI

Eco RI

Nde I

Eco RI

Ecl 136II/Sma I

Cla I

Sau 3AI/Bam HI

Hind III

amp

puc.02-P
Figure 4.8. H1 promoter / CAT fusion vectors

(c) The deletion mutants were constructed by first digesting pUC.02-P with Ava I and treating the opened vector with Exonuclease III (Erase-a-base kit, Promega). The shortened vectors were then recircularized and sequenced. Suitable deletions were removed from pUC.02-P by digestion with Cla I / Sau 3AI and cloned into Cla I / Bgl II-digested p(02L)TKCAT.
p(01S)TKCAT, in which the PCR fragments were derived from the 01 H1 promoter (not shown).

4.3.3 (b) Generation of p(02)HCAT vectors

The final products of the cloning strategy were created by replacing the HSVtk promoter in p(02L)TKCAT and p(02S)TKCAT with a 461 bp fragment containing the proximal region of the 02 H1 promoter (Figure 4.8 (a)). A 637 bp Sma I / Sau 3AI fragment spanning the 02 H1 promoter from -623 to +17 was firstly cloned into pUC19 between the Ecl 136II and Bam HI sites, generating pUC.02-P. The 461 bp Cla I / Sau 3AI fragment was then removed and cloned into Cla I / Bgl II - digested p(02L)TKCAT or p(02S)TKCAT; creating p(02L)HCAT and p(02S)HCAT, respectively.

To ascertain the possible contribution of more distal elements to the expression of the 02 H1 promoter, the 397 bp Nde I / Cla I fragment from pUC.02-P was ligated into Nde I / Cla I - digested p(02S)HCAT (Figure 4.8 (b)). The resultant construct, p(02P)HCAT, contained an additional 151 bp of 02 H1 promoter sequence 5' to the TG-box.

Finally, to investigate the importance of the highly conserved spacing between the AC- and TG-boxes, three deletions (21 bp, 214 bp, and 223 bp) were made between these elements. These deletion mutants were all generated in pUC.02-P. Digestion of this construct with Ava I removed a 21 bp fragment from the 02 H1 promoter between -295 and -275 inclusive. The vector was recircularized and the Cla I / Sau 3AI fragment was cloned into Cla I / Bgl II - digested p(02L)HCAT to create p(02LΔ21)HCAT. Larger deletion mutants were derived from Ava I - digested pUC.02-P using a Promega Erase-a-base kit according to the manufacturer's instructions. Once again, the Cla I / Sau 3AI fragment from the 02 H1 promoter deletions were transferred from pUC.02-P to p(02L)HCAT. In this way, p(02LΔ214)HCAT (a 214 bp deletion from -407 to -194) and p(02LΔ223)HCAT (a 223 bp deletion from -439 to -217) were constructed (Figure 4.8 (c)).

4.3.4 Effect of the TG-box on H1 Promoter Activity

The p(02)HCAT series of vectors were transfected into ts34 cells and the activity of each promoter was monitored by CAT assay. The results of six such experiments are
Chapter Four: Functional Analyses of the AC-box and a Related H1-Specific Promoter Element

depicted in Figure 4.9 (a). The first point of interest was that, in five out of the six experiments, the transcriptional activity of p(02S)HCAT was ~50% lower than that of p(02L)HCAT. The only difference between these constructs was that the former lacked 29 bp of distal 02 H1 promoter sequence which contained the TG-box. This indicated that the deletion of the TG-box was responsible for the decreased activity of p(02S)HCAT, as it was the only sequence in the 29 bp region that is conserved throughout the major and variant H1 genes of several species (Figure 4.5).

The promoter activities of p(02P)HCAT and the deletion mutants were generally lower (up to 50%) than that of p(02L)HCAT. Due to a high variability in the CAT activities of these extracts, no definitive conclusions about the effect of deletions between the AC- and TG-boxes on transcription of the 02 H1 promoter could be reached in these experiments. Similarly, the effect of the additional 150 bp 5' to the TG-box was uncertain.

In a second series of experiments, the CAT activities of extracts from cells transfected with p(02P)HCAT, p(02L)HCAT, and p(02S)HCAT were compared (Figure 4.9 (b)). There was no significant difference between the transcriptional activities of p(02P)HCAT and p(02L)HCAT, indicating that the 150 bp region 5' to the TG-box did not affect transcription of the 02 H1 promoter. In contrast, the transcriptional activity of p(02S)HCAT was significantly (>50%) lower than that of p(02L)HCAT, as had been indicated in the previous experiment (Figure 4.9 (a)). This was confirmed when the same result was obtained with different isolates of these constructs (data not shown). These results indicate that the TG-box is required for maximal expression of the 02 H1 promoter.

4.3.5 Effect of an 02 H1 Promoter Fragment on Expression of the HSVtk Promoter

To examine whether the AC- and TG-boxes, in combination, could produce a transcriptional effect, AEV ts34 cells were transfected with 15 μg of either p(02L)TKCAT, p(02S)TKCAT, or pBLCAT.2, and the activities of these promoter fusions were monitored by CAT assays.

The CAT activities of extracts prepared from cells transfected with p(02L)TKCAT or p(02S)TKCAT were comparable with the pBLCAT.2 control, as shown in Figure 4.9 (c).
Figure 4.9 Deletion of the TG-box affects transcription from the 02 H1 promoter

(a) ts34 cells were transfected with the p(02)HCAT vectors (15 μg DNA/transfection). Extracts from these cells were prepared 72 hours after transfection. The relative CAT activities of extracts from six transfection experiments are shown. p(02LΔ223)HCAT was not included in the first three experiments.
The graph shows the relative CAT activity across different experiments (Expt. 1 to Expt. 6) and treatments. The treatments include:

- p(O2P)CAT
- p(O2L)CAT
- p(O2S)CAT
- p(O2LΔ21)CAT
- p(O2LΔ214)CAT
- p(O2LΔ223)CAT

Each experiment contains data for all treatments, with the y-axis representing the relative CAT activity ranging from 0 to 110.
Figure 4.9 Deletion of the TG-box affects transcription from the 02 H1 promoter

(b) p(02P)HCAT, p(02L)HCAT, and p(02S)HCAT were transfected into ts34 cells as in Figure 4.9 (a), and the resultant extracts were assayed for CAT activity. The values shown are the means (± standard deviation) from five experiments.

(c) 15 µg of p(02L)TKCAT, p(02S)TKCAT, or pBLCAT.2 were transfected into ts34 cells as described in Figure 4.9 (a). The CAT activities of extracts from these transfections are shown. The activities are expressed as the mean (± standard deviation) from three experiments.
**Conversion**

**Relative CAT Activity**

- \( p(02P) \) HCAT
- \( p(02L) \) HCAT
- \( p(02S) \) HCAT

**% conversion**

- \( pBLCAT2 \)
- \( p(02L)JKCAT \)
- \( p(02S)JKCAT \)
Chapter Four: Functional Analyses of the AC-box and a Related H1-Specific Promoter Element

This suggested that neither the CH02-L (+ TG-box) nor CH02-S (- TG-box) promoter fragments had a significant effect on transcription from the HSVtk promoter in transient assays. The same result was obtained with different isolates of these constructs, as well as with the p(01L)TKCAT or p(01S)TKCAT constructs (data not shown). A possible explanation for this result is that these sequences cooperate with additional proximal sequences from the H1 promoter to produce a transcriptional effect.

4.4 Discussion

The experiments in this chapter concern the functional analysis of two H1-specific promoter elements. The first element, the AC-box, was tested for its ability to influence transcription from a heterologous promoter (HSVtk). One to six copies of 25 bp oligonucleotides containing the AC-box were placed directly upstream of the HSVtk promoter in hGH and CAT reporter plasmids. These constructs were transfected into AEV ts34 erythroblasts and the HSVtk promoter activities were compared. There was no significant difference in promoter activity between constructs containing equal numbers of wild type or mutated AC-boxes. This demonstrated that, in transient transfections of ts34 (and CHO) cells, the AC-box did not affect transcription from the HSVtk promoter. This element is certainly required for maximal expression of the H1 promoter (Dalton and Wells, 1988a; Gallinari et al, 1989; La Bella et al, 1989). The results presented in this chapter suggest that the AC-box interacts with another element(s) from the H1 promoter in order to produce a transcriptional effect. This interaction may occur via transcription factors that bind these elements, or through the DNA elements themselves.

An alternative explanation for these results is that the AC-box was unable to influence transcription because of its location. The distance of the AC-box from the cap site of the HSVtk promoter (~200) was at least 40 bp upstream with respect to its usual distance from the cap site in the H1 promoter (~100 to -160). Its well conserved position in the H1 promoter implies that this position has functional significance. Placement of the AC-box at a different position may interfere with this function. In subsequent experiments (see below),
the distance between the AC-box and the HVS\textsubscript{t}k cap site was shorter than the corresponding distance in the \textit{Xenopus} H1 promoters and very similar to the corresponding distance in the sea urchin H1-\textbeta{} promoter (see Figure 4.5). The AC-box still did not affect expression of the HVS\textsubscript{t}k promoter in these experiments, lending support to the notion that other H1 promoter elements cooperate with the AC-box to potentiate transcription.

A third possible explanation for the lack of a transcriptional effect by the AC-box is that this element is not a binding site for a transcription factor. Instead, it may be a cis-acting element involved with maintaining the H1 promoter region in a state (such as a particular chromatin state) that is amenable to transcription. This would explain the decrease in transcriptional activity of H1 promoters where the AC-box has been mutated or deleted (Dalton and Wells, 1988a; Gallinari \textit{et al}, 1989), and the failure of the AC-box to activate transcription in a heterologous promoter that does not normally require the AC-box for expression.

The second element analysed, the TG-box, was uncovered in a computer-aided comparison of both major and variant H1 promoters from several higher eukaryote species. This element is a near perfect inverted repeat of the AC-box. From the available sequence data, it appears that the TG-box resides at \textasciitilde{}470 in every H1 promoter that contains the AC-box. Except for a single base change in the TG-box sequence of two H1 promoters, this element is invariant in thirteen different H1 histone promoters. In the promoters containing the two slightly different TG-boxes, compensatory mutations in the corresponding positions of the AC-boxes are found. Thus, at least in the promoters examined, these elements are more conserved than any other H1 promoter element. Also highly conserved is the distance between the AC-box and the TG-box. The distance between the 5' ends of each element was between 336 and 359 bp depending on the promoter, and in eight of the thirteen promoters this distance was restricted to between 351 and 354 bp.

The effect of the TG-box on the expression of the chicken 02 H1 gene was examined. The most consistent finding was that deletion of a 29 bp region containing the TG-box led to a 50\% decrease in the activity of the 02 H1 promoter in transient assays. This
Chapter Four: Functional Analyses of the AC-box and a Related H1-Specific Promoter Element

decrease is most likely to have resulted from the loss of the TG-box sequence (although not conclusively shown), as it is the only conserved element of this 29 bp region in the thirteen H1 promoters examined.

The decrease in 02 H1 promoter activity upon removal of the TG-box-containing region may represent a general effect on transcription, or it may reflect a loss of the S-phase-specific activation of transcription that is a feature of the major histone genes. The AC-box has been implicated in the S-phase activation of H1 gene transcription (Dalton and Wells, 1988a; La Bella et al, 1989). This in turn suggests that the TG-box may also be involved, as the palindromic relationship and conserved distance between these two elements suggest that their functions are interrelated. In contrast, these elements are also present in the non-cell-cycle-regulated H1t and H10 variant genes, and deletion of H10 sequences which contained the TG-box also produced a decrease in expression (Steuer et al, 1992). However, there is evidence that the H10 and H1t promoters, and promoters from other replication independent histone genes, are activated in S-phase or can direct cell cycle regulation of heterologous genes (Hwang and Chae, 1989; Huh et al, 1991; Grunwald et al, 1991; Khochbin et al, 1991; Kremer and Kistler, 1992).

To determine whether the position of the TG-box with respect to the more proximal H1 promoter elements was also important for maximum expression, 21 bp, 214 bp, and 223 bp deletions were made between the AC- and TG-boxes. Although there was some suggestion that these deletions all led to decreased transcriptional activity, the level of variability in these experiments meant that no firm conclusions could be drawn. More significant results were reported for the H10 gene, where a number of deletions between the AC- and TG-boxes, ranging from 27 to 185 bp, produced a marked decrease in transcription (Steuer et al, 1992). The magnitude of the decrease was similar regardless of the size or position of the deletion. The results of the H10 experiments implied that, rather than removing sequences that were important for transcription, the deletions were disrupting an interaction between the proximal promoter elements and some distal region (possibly the TG-box).

A 395 bp 02 H1 promoter fragment, containing only the TG- and AC-boxes and the intervening sequences, had no significant effect on transcription from the HSVtk promoter
in transiently transfected AEV ts34 cells. As suggested by the AC-box/HSVtk promoter fusions (see above), these elements may interact with other H1 promoter sequences to produce a transcriptional effect. In contrast to indications from previous work carried out in this laboratory (Dalton, 1987), the AC-box did not behave as an enhancer element in these experiments.

Another explanation for this result may have been the location of the AC- and TG-boxes with respect to the HSVtk cap site. In p(02L)TKCAT and p(02S)TKCAT, the 5' end of the AC-box is at -145. This is 25 bp further upstream than its position in the 02 H1 promoter, but is very similar to the location of the AC-box in the H1 promoters of *Xenopus* and sea urchin (Perry *et al*, 1985; Wells, 1986). Although a relatively short distance, this difference may be sufficient to disrupt the normal function of these elements.

Also consistent with these results is the alternative explanation that these elements are not binding sites for transcription factors, but that their function is to maintain the H1 promoter region in a "transcribable" state (see above).

Since the completion of these experiments, several detailed investigations of the regulation of H10 gene transcription in mouse, *Xenopus* and human have appeared in the literature (Breuer *et al*, 1993; Khochbin and Wolffe, 1993; Bouterfa *et al*, 1993). In the mouse H10 promoter, the previously characterised 80 bp fragment essential for basal transcription (Steuer *et al*, 1992) was demonstrated to activate transcription from the HSVtk promoter by 2-3-fold. The TG-box is contained within the 3' region of this fragment. A similar enhancement of transcription was observed when a 33 bp oligonucleotide, also containing the TG-box, was placed upstream of the HSVtk promoter. However, the TG-box was not responsible for this effect, as transcription from HSVtk was not activated by a mutated oligonucleotide in which the TG-box sequence remained intact.

Khochbin and Wolffe (1993) recognised a 44 bp distal region of homology between the mouse and *Xenopus* H10 promoters. This upstream conserved element (UCE) resulted in a ~10-fold decrease in transcription when deleted from the *Xenopus* promoter. The TG-box resides at the extreme 5' end of the UCE. Mutation of the TG-box sequence led to a similar decrease in transcription to that which resulted from deletion of the entire UCE. This
result conclusively demonstrated that the TG-box was essential for maximal activity of the *Xenopus* H10 promoter. Further mutational analysis of this promoter indicated that the UCE coordinated with the AC-box and the "H4 subtype-specific element" (conserved in H10 promoters - see 1.9.2) to mediate both basal and butyrate-inducible expression. However, the UCE did not possess the properties of a transcriptional enhancer.

Only a portion of the UCE, including the TG-box, is conserved in the human H10 promoter (Bouterfa *et al*, 1993). Considerable decreases in transcription were also observed in deletion mutants of the human promoter which lacked the TG-box.

It is interesting to note that the magnitude of the decrease in transcription upon removal of the TG-box is much greater for the H10 genes (10-20-fold) than for the 02 H1 gene (2-fold). From these data, it would appear that the TG-box has a more influential role in the expression of the H10 genes than of the major H1 genes. On the other hand, the large difference in the observed effects may be a consequence of the different systems in which the effect is studied. H10 transcription is most conveniently studied in cell types in which the H10 gene is specifically induced, either in terminally differentiated cells or cells which have been treated with agents that induce differentiation. In such systems, the expression (rate of transcription initiation) of the H10 gene may be significantly greater than the expression of a replication dependent H1 gene in actively dividing cells. The deletion of the TG-box may cause expression to drop to a "basal" level that is relatively similar for both genes, meaning that the decrease in the rate of transcription initiation is greater for the H10 gene. Alternatively, the variation in the magnitude of this decrease in transcription may simply reflect a fundamental difference in the regulation of these genes (recall that the distance between the AC- and TG-boxes of the H10 promoters is about 15 bp shorter than that of the major and the testis-specific H1s; see Figure 4.5). It should also be noted that the 2-fold decrease in 02 H1 expression upon deletion of the TG-box is similar to the magnitude of the decrease in transcription of a human H1 promoter construct upon deletion of the AC-box (Gallinari *et al*, 1989).
CHAPTER FIVE

FINAL DISCUSSION
The motivation behind the work presented in this thesis was to gain a better understanding of the function of an H1 histone promoter-specific element. This element, known as the H1-box or AC-box, was first identified by Coles and Wells (1985) as being conserved in sequence and position in the H1 promoters of several higher eukaryote species. Functional analyses of different H1 genes confirmed that the AC-box was necessary for both maximal transcription as well as the S-phase-specific activation of the H1 promoter (Dalton and Wells, 1988a; Lai et al, 1988; La Bella et al, 1989). Trans-acting factors were subsequently demonstrated to specifically bind this element (Dalton and Wells, 1988a; 1988b; Gallinari et al, 1989). The initial aim of this project was to purify H1-SF, the AC-box-specific binding factor identified and partially characterised in this laboratory (Dalton, 1987). Purification of sufficient quantities of H1-SF for direct amino acid sequencing would enable the possibility of isolating the H1-SF gene. In addition, the production of H1-SF-specific antibodies could provide information about the distribution and activity of this factor. These data would in turn lead to a more complete understanding of the control of H1 gene transcription.

Mobility shift assays of crude nuclear extracts prepared from both chicken embryos and AEV ts34 erythroblasts were conducted, using the same 01-40mer probe used in the initial characterisation of H1-SF (Dalton, 1987). An AC-box-binding factor with the same specificity as H1-SF was not detected, although a band shift (40-MS-e) with a similar mobility to the H1-SF band shift was detected in ts34 extracts. 40-MS-e was similar in other respects to H1-SF. 40-MS-e and H1-SF eluted in the same salt washes from DEAE- and Heparin-Sepharose columns, and 40-MS-e displayed the same high specificity for the 01-40mer probe in the presence of non-specific competitors. In apparent contrast to H1-SF, 40-MS-e bound single stranded 01-40mer probes. Close inspection of some previous results concerning H1-SF (Dalton, 1987) indicate that it also may have bound single stranded probe, though it was still AC-box-specific. Several transcription factors have been demonstrated to bind the single stranded form of their cognate binding site. However, a more serious anomaly between H1-SF and 40-MS-e was indicated when incubation of ts34 extracts with the Δ01-40mer probe gave rise to a band shift with identical mobility to that of
40-MS-e. If 40-MS-e bound the mutated AC-box probe, then this represented an irrefutable difference between this factor and H1-SF, and they must therefore be distinct.

Incubation of a shorter AC-box probe (01-14mer) with chicken embryo crude nuclear extracts gave rise to a candidate AC-box-specific band shift; the factor which produced this band shift was designated 01-14-b. Although easily competed out by non specific competitors such as salmon sperm DNA or poly(dIdC), this band shift displayed higher specificity in the presence of heterologous oligonucleotide competitors. 01-14-b was purified on a large scale from chicken embryos. Crude nuclear extracts were fractionated by Heparin-Sepharose and Sephacryl S-300 chromatography, and fractions containing 01-14-b were purified by DNA affinity microprecipitation / chromatography. Two proteins (86 kD and 70 kD) were consistently enriched in DNA affinity-purified extracts as analysed on SDS polyacrylamide gels. However, renaturation of 01-14-b from SDS gels demonstrated that it was no larger than 30 kD. In addition, size fractionation of this factor on a Superose-12 HPLC column indicated that it was no larger than 40 kD (Hopwood, 1993). Previous experiments had indicated that H1-SF was ~90 kD under non denaturing (Dalton, 1987; Dalton and Wells, 1988b) and denaturing (S. Dalton, pers. comm.) conditions. Gallinari et al (1989) had also partially characterised a 90 kD protein (H1TF1) that specifically bound the AC-box from a human H1 histone promoter. From these data, it was clear that 01-14-b and H1-SF were not analogous factors. With the additional discovery that no 01-14mer binding activity was detected in ts34 extracts, the purification of 01-14-b was not pursued.

Although a specific 01-40mer-binding factor was not detected in chicken embryo crude nuclear extracts, such a factor was detected in DNA affinity-purified extracts. Renaturation of this factor from SDS polyacrylamide gels demonstrated that it too was much smaller (42-46 kD) than H1-SF. Thus further purification of this factor was also not pursued.

In order to overcome inherent problems associated with the 01-14mer and 01-40mer probes, 25 bp oligonucleotide probes derived from two different chicken H1 promoter sequences were designed. However, these probes detected no specific AC-box-binding factor in mobility shift assays of chicken embryo extracts. In addition, use of these 25mers as ligands in DNA affinity chromatography experiments finally determined that the 86 kD
and 70 kD proteins enriched in earlier DNA affinity-purified extracts were not AC-box-specific factors.

Mobility shift assays of ts34 crude nuclear extracts with the 25mer probes detected a factor (25-MS-1) that specifically bound the probe derived from the 02 H1 promoter. Although not detected in chicken embryo extracts, this same factor was detected in several different chicken cell lines. Nucleotides in both the AC-box and the sequences flanking the AC-box were required for the binding of 25-MS-1 to the 02-25mer probe. However, the 25-MS-1 band shift was not detected by the 01-25mer probe, suggesting that it was not an AC-box-specific factor.

One possible flaw with the experimental approach behind attempts to purify H1-SF was the assumption that previous results concerning H1-SF were irrefutable. It should be pointed out that simultaneous attempts to reproduce the H1-SF band shift in this laboratory were also unsuccessful (Hopwood, 1993). In addition, despite detection of a factor (H1TF1) which specifically bound a human AC-box sequence, the subsequent finding that this factor did not interact with some AC-boxes dismissed it as a bona fide AC-box-binding factor (see Heintz, 1991). In an analysis of another human H1 promoter, van Wijnen et al (1988) did not detect an AC-box binding activity, though proteins which interacted with other conserved motifs in the promoter were detected. The reason why previous experiments in this laboratory detected H1-SF and later experiments did not is unknown, although operator technique is a possible factor. As discussed below, subsequent evidence indicated that the AC-box may not be a binding site for a transcription factor.

At this stage, it was decided to investigate the function of the AC-box from a different perspective. To study the possible transcriptional activating ability of the AC-box, oligonucleotides containing AC-box sequences were placed directly upstream of the HSVtk promoter. Transcription from this promoter was unaffected by one, three or six copies of the AC-box. This suggested that the AC-box was insufficient to produce a transcriptional effect on the HSVtk promoter in transient assays.

The discovery of the TG-box could explain much of the difficulties encountered in earlier experiments. This element is almost absolutely conserved in both major and variant H1 promoters throughout the higher eukaryotes from sea urchin to human. The TG-box
bears an interesting relationship to the AC-box. Firstly, it is a near perfect (6 out of 7 bases) inverted repeat of the AC-box, with the 7th position of the TG-box and the 1st position of the AC-box occupied by identical, instead of complementary, nucleotides. Remarkably, in the only two cases where the TG-box sequence has diverged from the consensus, compensatory base changes are found in the corresponding positions of the AC-box to maintain this near perfect inverted repeat structure. Secondly, the distance between these elements (~350 bp) is also highly conserved. These findings are strongly suggestive that the functions of these two elements are interrelated. Thus, the failure to detect a specific AC-box-binding factor in mobility shift assays may be because the full binding site for such a factor consists of both the AC- and TG-boxes. However, the inability of the AC-box to activate transcription from the HSVtk promoter did not appear to be due to the lack of the TG-box. A fragment from the 02 H1 promoter containing both these elements also failed to activate transcription. This could mean that both of these elements potentiate transcription by interaction with other H1 promoter elements, or that they are not bound by transcription factors (discussed below).

The TG-box, like the AC-box, appears to be essential for maximal H1 promoter activity. Deletion of a 29 bp region which included the TG-box produced a 50% decrease in expression of the 02 H1 promoter. To the best of my knowledge, this study represents the first indication of an effect by the TG-box on the transcription of a replication dependent histone H1 gene. Similar results were obtained in studies of the H1\(^{0}\) gene, except that removal or mutation of the TG-box had a much greater effect (10-fold reduction) on transcription (Steuer et al, 1992; Breuer et al, 1993; Khochbin and Wolffe, 1993; Bouferta et al, 1993). This difference may be an effect of the different systems in which the studies were performed, or it may simply reflect a difference in the regulation of the H1\(^{0}\) and 02 H1 genes.

Although the role of the TG-box is unknown, a recent finding might help to elucidate its function. The 5.0 kb repeat unit containing the five *Drosophila melanogaster* histone genes has been sequenced (Matsuo and Yamazaki, 1989). Apart from the TATA box, none of the promoter elements of vertebrate histone genes were located in the *Drosophila*
promoters, suggesting that a different (or at least highly diverged) set of transcription factors regulate *Drosophila* H1 gene expression. However, I have identified previously unreported AC- and TG-boxes in the H1-H3 spacer region, ~600 bp upstream of the H1 coding sequence (see Figure 5.1). These AC- and TG-boxes are identical in sequence to the consensus. They occur as pairs separated by 12 and 9 bp, respectively, and the innermost boxes of this arrangement are 256 bp apart. The association of these elements with the *Drosophila* histone genes suggests that they operate independently of the other H1 promoter elements of sea urchin and vertebrates (and *Drosophila*), with the possible exception of the TATA-box. However, there was no evidence for the cooperation of the HSVtk TATA-box with the AC- and TG-boxes (see above), and their location in the *Drosophila* H1-H3 spacer is several hundred bp upstream from their usual location in the H1 promoters of other higher eukaryotes. These elements could therefore represent a novel mechanism of gene regulation.

Curiously, this same arrangement of AC- and TG-boxes was not found in the histone gene repeat unit of *Drosophila hydei* nor *Chironomus thummi thummi* (Kremer and Hennig, 1990; Hankeln and Schmidt, 1991), although perfect or near perfect matches to both of these elements are present in the repeat units of both species.

The *Drosophila melanogaster* AC- and TG-boxes lie within a scaffold attachment region (SAR - Mirkovitch *et al*, 1984). SARs are believed to be anchorage points to the nuclear "scaffold" or "matrix" and are often found in the vicinity of putative eukaryotic DNA replication origins (Benbow *et al*, 1992). It has been postulated that DNA replication and transcription occurs at the nuclear scaffold (Pardoll *et al*, 1980; Vogelstein *et al*, 1980; Jackson and Cook, 1985). The presence of AC- and TG-boxes within the SAR may be related to the S-phase-specific expression of the *Drosophila* histone genes. These elements may play a role in maintaining the nucleosomal or chromosomal structure of the H1 promoter in a state amenable to transcription during S-phase. Supportive (though circumstantial) evidence for this hypothesis can be found when comparing the effects of deleting these elements in different systems. In permanent cell lines, deletion or mutation of the AC-box (Dalton and Wells, 1988a) led to a much greater decrease in H1 promoter activity than deletion of the TG-box (or the AC-box - Gallinari *et al*, 1989) in transient assays. This is suggestive that the AC- and TG-boxes must reside within a chromosomal
context for a maximal effect upon H1 gene transcription, although they can also affect transcription when in a nucleosomal context.

The arrangement of the AC- and TG-boxes (two halves of an AT-rich short inverted repeat separated by a large distance) is reminiscent of the arrangement of inverted repeats found in prokaryotic and eukaryotic replication origins. Such repeats in the E. coli origin of replication (oriC), known as dnaA boxes (5' TTATC/ACAC/GAA 3') are binding sites for the dnaA initiator protein (Fuller et al., 1984). It has been noted that an element found in many S. cerevisiae autonomously replicating sequences (ARSs); 5' A/TTAAT/CTATATTTA A/T 3', is also similar to the AC-box (Younghusband et al., 1986). However, in contrast to the AC-box, this element has been implicated in transcriptional repression rather than activation (Miller and Nasmyth, 1984; Brand et al., 1987; Fangman and Brewer, 1991). Regions of DNA (orS) from monkey CV-1 cells with similar properties to ARSs are also characterised by AT-rich regions and inverted repeats (Landry and Zannis-Hadjopoulos, 1991; Pearson et al., 1991).

Short, AT-rich inverted repeats have also been proposed to play a role in nucleosome positioning. The highly repetitive α-satellite DNA of the African green monkey genome is a 172 bp repeat unit. Using a band-competition assay, Strauss and Varshavsky (1984) detected and purified a factor (α-protein) that specifically bound this unit. DNase I footprinting identified three binding sites for the α-protein (I, II, and III) within the α-DNA unit. Sites II and III are AT-rich inverted repeats, and site II of one unit lies 145 bp (one nucleosome core DNA length) from site III of the adjacent unit. The authors suggested that through its binding to the regularly spaced AT-rich regions, the α-protein may be a nucleosome positioning protein.

A final interesting observation is the identification of AC- and TG-box sequences in non histone promoters. The promoters of many, if not all, plant seed storage protein genes contain the sequence 5' AACACA 3' (Goldberg, 1986), and TG-boxes can also be found in at least some of these genes (see Dasgupta et al., 1993). In addition, I have identified a TG-box ~200 bp upstream from a previously identified AC-box in the promoter of a sheep keratin gene (Powell et al., 1993). The significance of the association of AC- and TG-boxes
Figure 5.1. TG- and AC-boxes are present in the *D. melanogaster* H1-H3 spacer

The location of two TG-boxes and two AC-boxes in the spacer between the H1 and H3 histone genes of *Drosophila melanogaster* is illustrated. Also shown is the SAR (Mirkovitch *et al.*, 1984). The coding sequences of the H1 and H3 genes are represented by arrows in the 5' to 3' direction.
Chapter Five: Final Discussion

with non histone genes is unknown, but it is suggestive that a common mechanism exists between the regulation of these genes and the H1 histone genes.

In closing, the AC- and TG-boxes may represent a novel transcription-activating mechanism distinct from promoter or enhancer elements, and they may also be involved in the regulation of non histone genes. This work has uncovered what promises to be an exciting future in the study of histone gene expression, and perhaps transcriptional regulation in general.
BIBLIOGRAPHY
Bibliography


Bibliography


Bibliography


Bibliography


Bibliography

Bibliography


Bibliography


Bibliography


Bibliography

Bibliography


Bibliography


Bibliography


Bibliography

Bibliography


Bibliography


Bibliography

Bibliography

Bibliography


Bibliography