UNIVERSITY OF ADELAIDE
6.8.93
LIBRARY

TOWARDS CHARACTERISATION OF HISTONE H1 GENE TRANSCRIPTION FACTORS

A thesis submitted to the University of Adelaide
for the degree of Doctor of Philosophy

by

Blair Hopwood, B.Sc. (Hons.)

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

February, 1993

Awarded 1993

CONTENTS

STATEMENT

The work reported in this thesis was carried out from March 1988 to September 1992 in the Department of Biochemistry at the University of Adelaide. This thesis contains no material that has been accepted for the award of any other degree or diploma by any university. To the best of my knowledge it contains no material that has been previously published by any other person, except where due reference has been made in the text. I consent to this thesis being made available for photocopying and loan.

(Blair Hopwood)

SUMMARY

The aim of work presented in this thesis was to isolate and characterise a chicken *trans*-acting factor (H1-SF), reported to bind an histone H1-specific promoter element called the H1 box. Previous work had indicated that the interaction between H1-SF and the H1 box, modulated cell cycle control of histone H1 gene transcription (Dalton and Wells, 1988*a*).

(1) Attempts were made to identify and confirm H1-SF binding activity using the gel retardation assay. Double stranded oligonucleotide probes were used, which contained either the H1 box (40 mer) or a disrupted version of the H1 box (Δ40 mer). For reasons still unclear, an H1 box specific binding activity (H1-SF) was not able to be detected, although a binding activity (H1-F40) was identified that interacted with both the 40 mer and Δ40 mer.

(2) Subsequently, another set of double stranded oligonucleotide probes containing an intact H1 box (14 mer) and disrupted H1 box (Δ14 mer) were designed. An H1-SF candidate (H1-F14), which bound the 14 mer but not the Δ14 mer was identified by gel retardation assay. By these binding criteria and other competitor studies H1-F14 was believed to be H1 box specific and represent H1-SF.

(3) Extensive oligoscreening of a ten day chicken embryo λgt11 cDNA expression library for H1-F14 was carried out. However no H1-F14 clones were isolated using this method.

(4) A large scale protein column chromatography purification protocol for H1-F14 was undertaken. This involved preparation of crude nuclear extract from 600 dozen nine day chicken embryos, partial purification of H1-F14 by heparin-Sepharose column chromatography, membrane filtration concentration, and Sephacryl S-300 gel filtration column chromatography. Affinity chromatography purification of H1-F14 was to be performed next, to purify H1-F14 to homogenity. However before affinity chromatography was carried out, experiments were completed that demonstrated that H1-F14 actually represented a non-specific binding protein. H1-F14 was abandoned at this point as an H1-SF candidate.

(5) It was reasoned that sequences outside of the H1 box, absent from the 14 mer, may be important for H1-SF binding. A further set of double stranded oligonucleotide probes containing an intact H1 box (25 mer) and a disrupted H1 box (Δ25 mer) were designed with this in mind. An H1-SF candidate (H1-F25) was identified by gel retardation assay using these probes. Like H1-F40 which bound the 40 mer and Δ40 mer probes (see (1)), H1-F25 bound both the 25 mer and Δ25 mer probes. However a more rigorous testing of H1-F25 binding specificity, by modification interference analysis, demonstrated that sequences outside of the H1 box along with those contained within it, were involved in binding. As a result H1-F25 was not dismissed as a H1-SF candidate and was pursued further.

(6) A modified oligoscreening method (Kalionis and O'Farrell, 1993), different in a number of important aspects to the method applied in (3), was used to screen a ten day chicken embryo λgt11 cDNA expression library for H1-F25. Two candidate H1-F25 clones were isolated (λ4.1 and λ8.1). These clones were subsequently found, by DNA sequence analysis, to code for novel zinc finger DNA binding proteins with 12 and 5 zinc fingers respectively (nucleic acid and protein databases searched; July, 1992). However λ4.1 (Z12) and λ8.1 (Z5) both appear to be incomplete cDNA clones.

(7) The Z12 and Z5 clones were, after difficulty trying to express them from λ lysogens, subcloned into the pGEX-1 expression vector and sucessfully expressed in *E. coli*. Extracts were prepared and Z12 and Z5 fusion proteins shown to bind 25 mer probe in the gel retardation assay, in a manner analogous to H1-SF25.

(8) Further screening of a ten day chicken embryo λgt11 cDNA expression library was undertaken to isolate the complete cDNA clones for Z12 and Z5. Another Z12 cDNA clone λ16.1 (Z13) was isolated and found to contain additional 5' coding sequence in addition to an extra zinc finger located near the 3' end of the clone. The original Z12 clone would appear to be an incorrectly spliced cDNA copy.

(9) Screening of a λCharon 4a chicken genomic library was carried out. Genomic clones for both Z12 and Z5 (λG7.1 and λG2.1) were isolated. Preliminary characterisation of Z12 and Z5 gene structures was commenced.

ACKNOWLEDGEMENTS

I would like to thank Professor George E. Rogers for the opportunity of undertaking this work in the Department of Biochemistry, Dr. Julian R.E Wells for his supervision throughout the course of the project and the many members of the Department of Biochemistry. In particular, the members of Labs 108, 109, 110, 20, 20A and G02 for providing such a friendly, supportive and stimulating working environment - also for just making lab life fun. I will miss you all.

I am particularly grateful to Rick Tearle, Peter Wigley, Kym Duncliffe, Lena Rondahl, Tom Schulz, Rob King, Varaporn Thonglairoam and Allan Robins for their helpful discussions. I also wish to thank Julian for critically reading the text, Rick for help with database searches, Lesley Crocker for advice on photography, Bill Kalionis for advice on oligoscreening, Stan Bastiras for help with Superose-12 HPLC, Leanne Coles for advice on modification interference analysis and Briony Forbes, Louise O'Keefe and Rob for their help with proof-reading the text.

I wish to thank my family for their endless support and encouragement. I am indebted to my father for the time he found to critically read the text.

And finally, a big thankyou, Louise, for putting up with me.

During the course of this work I was supported by a Commonwealth Postgraduate Research Award.

CHAPTER 1

INTRODUCTION

UNIVERSITY OF ADELAIDE LIBRARY

1.1 INTRODUCTION

Development and differentiation in eukaryotic organisms appears to be controlled largely at the level of transcription by regulatory proteins that bind to specific sites in the vicinity of a gene. The work presented in this thesis will focus on a regulatory factor reported to be involved in the transcriptional control of the chicken histone H1 genes during the cell cycle. This introduction will attempt to put the experimental work presented in this thesis into context by reviewing the two main areas directly relevant to this work. These are the control of eukaryotic gene expression with particular emphasis on transcriptional control, and the organisation, structure and expression of histone genes including a detailed description of the expression of the histone H1 genes.

1.2 REGULATION OF GENE EXPRESSION

The regulation of eukaryotic gene expression is controlled at a number of points along the pathway from gene to gene product. Transcriptional regulation, RNA processing, RNA transport, RNA stability, translational regulation, and post-translational regulation all represent stages at which gene expression can be controlled. Other less common regulatory mechanisms include gene amplification, DNA rearrangement and DNA transposition. However it is generally recognised that the most frequent mode of eukaryotic gene control lies at the level of transcription (reviewed, Darnell, 1982). Before dealing with transcriptional control the other regulatory mechanisms mentioned above will be briefly reviewed below.

1.2.1 RNA processing

A number of processing steps occur to primary RNA transcripts in eukaryotes before mature mRNAs are finally produced. These processing steps include the addition of a 7-methyl-G cap which protects the mRNA from nuclease digestion, and serves as an initiation signal for translation (reviewed, Shatkin, 1976), the addition of adenylic acid residues to create a 3' poly A tail which stimulates translation, and stabilises the

mRNA (reviewed, Wickens, 1990), and finally the splicing out of intronic sequences to produce the mature mRNA (reviewed, Sharp, 1987).

Alternative splicing, in which introns are spliced out of the primary transcript in different combinations, represents a variation for manipulating gene expression. The process can be regulated, often in a developmental or tissue-specific context, to generate different proteins from a single gene. Alternative splicing can produce protein isoforms which are modified to suit different biological situations (e.g. several muscle proteins; review, Leff *et al.*, 1986), or produce two proteins in different tissues that have totally different functions (e.g. calcitonin and calcitonin-gene-related peptide; Leff *et al.*, 1987), or produce selective activation of a protein in certain tissues by generating an active protein by one splicing event, and a non-functional protein by another (e.g. the *Sxl* gene involved in sex determination in *Drosophila*; review, Baker, 1989). Another type of control at the RNA processing level involves alternative sites of transcript cleavage to form the 3' end. This can lead to generation of proteins with different C-terminal sequences (e.g. membrane-bound and secreted forms of immunoglobulin; Peterson and Perry, 1986). Also alternative primary transcripts can be produced by initiation of transcription from different promoter elements of a gene. These transcripts are then processed differentially in such a way that differential splicing is controlled by the presence or absence of the 5' most exon in the primary transcript (e.g. mouse α-amylase gene, salivary gland and liver specific transcripts, Young *et al.*, 1981).

Although most eukaryotic mRNAs have a poly(A) tail added during processing, most histone mRNAs in higher eukaryotes are not polyadenylated. Histone mRNAs instead are processed in such a way that they contain a characteristic stem-loop structure at their 3' end (reviewed, Schümperli, 1988). This novel processing event is involved in regulation of histone gene expression (see 1.10).

Two other novel forms of RNA processing have recently been described. These include *trans*-splicing which involves the covalent linkage of RNA segments from two different transcriptional units (reviewed, Laird, 1989) and RNA editing, in which in one case, in trypanosomes, non-genomically encoded uridine residues are post-transcriptionally inserted into the RNA to form a functional transcript (reviewed, Benne, 1990), and in another example, co- or post-transcriptional substitution of a single nucleotide in rabbit apolipoprotein B mRNA produces two different gene products (Powell *et al.*, 1987).

1.2.2 mRNA stability

Because the stability of an mRNA will influence the time available for translation of the mRNA and therefore the amount of protein produced, the regulation of mRNA stability represents another level at which gene expression can be controlled. Individual mRNAs differ widely with respect to stability, but mRNAs encoding proteins whose levels change rapidly in a cell are often unstable, whereas others such as β-globin mRNA appear to be very stable. Sequences contained in the mRNAs themselves are involved in regulating stability. These regions are often located in the 3' untranslated region of the mRNA and include potential stem loop structures that appear to confer resistance to 3' exonuclease attack (reviewed, Brawerman, 1987) and conversely long A-U rich sequences, believed to promote mRNA decay (Shaw and Kamen, 1986). Although a large number of stability determinants seem to reside in the 3' untranslated region, other regions of the mRNA have also been shown to mediate stability. One example involves the auto-regulation of β-tubulin mRNA where a 13 bp sequence at the 5' end of the β-tubulin mRNA appears to be responsible for destabilisation of the mRNA (Yen *et al.*, 1988).

Stability can be altered in response to physiological signals. For example, the half-life of mRNA for the heat shock protein, hsp70, increases approximately 10-fold after heat shock in HeLa cells (Theodorakis and Morimoto, 1987). Whereas, in the case of mRNA for the transferrin receptor, an increase in intracellular levels of iron causes a

decrease in the stability of transferrin receptor mRNA. This is brought about through a sequence, located in the 3' untranslated region of the transferrin receptor mRNA, termed an iron responsive element that forms a stem loop structure and which binds a protein in the absence of iron. The addition of iron to the system causes the protein to dissociate, thus destabilising the mRNA (Müllner *et al.*, 1989).

The half-life of histone mRNAs are found to fluctuate during the cell cycle (reviewed, Osley, 1991). These fluctuations in histone mRNA stability appear to be tightly coupled to DNA replication. The contribution of regulation of mRNA stability to histone gene expression is discussed in more detail in 1.10.

1.2.3 Translational control

The final stage in the expression of a gene involves the translation of mRNA into a protein product. Examples of translational control, in eukaryotes, mediated via modifications to the translational apparatus affecting the efficiency of translation and by sequences in the mRNAs themselves have been found. For instance, the enhanced translation of ferritin mRNA in response to iron, is mediated by an iron responsive element, similar to that found in the transferrin receptor mRNA (see 1.2.2), but situated in the 5' untranslated region of the ferritin mRNA (Aziz and Munro, 1987). This 5' iron responsive element similarly binds a protein in the absence of iron, but in this case, the RNA-protein complex represses initiation of translation of the mRNA. The addition of iron to the system causes the protein to dissociate from the mRNA, allowing translation (Müllner *et al.*, 1989).

Together with control of mRNA stability, translational control can be seen as supplementing transcriptional regulation in cases where rapid cellular responses are required. Post-translational modification of proteins appears to be a further means of producing a rapid and effective response.

1.2.4 Post-translational regulation

Post-translational modifications, often involve chemical modification of the protein product. Examples of this form of control include phosphorylation (reviewed, Hunter and Karin, 1992), acetylation (e.g. histones; reviewed, Wu *et al.*, 1986), and glycosylation (reviewed, Paulson, 1989).

1.2.5 Less common regulatory mechanisms

DNA rearrangement (e.g. yeast mating type; reviewed, Nasmyth, 1982), gene amplification (reviewed, Kafatos *et al.*, 1985), DNA transposition (e.g. trypanosome variant surface glycoproteins; reviewed, Van der Ploeg, 1987), and RNA transport (e.g. Rev protein of human immunodeficiency virus type 1; Chang and Sharp, 1989), all represent less frequently used forms of gene regulatory mechanisms.

1.3 REGULATION OF TRANSCRIPTION

Both frequency and specificity of initiation are of prime importance in the control of transcription (review, Darnell, 1982). A key feature of transcriptional regulation is the ability to selectively express certain genes. This facet of transcriptional initiation is particularly relevant in the expression of tissue specific, developmentally regulated and inducible genes which are selectively activated in certain cell types or in response to certain physiological stimuli. Regulatory regions of eukaryotic genes and the DNA elements they contain have been identified as being critical for accurate and efficient initiation of transcription. As a result, in the last few years, a major area of research has been involved with the identification of DNA elements involved in the regulation of transcription. Important regulatory sequences have been identified, for example, through their conservation, by sequence comparisons of the promoter regions, of genes with similar expression patterns, and between homolgous genes from different species. The functional significance of specific DNA elements has often also been demonstrated by the use of *in vitro* and *in vivo* gene expression systems, usually in combination with systematic deletion analysis and/or mutagenesis of the sequences. In most cases the DNA elements have been shown to

specifically bind regulatory proteins often called, *trans*-acting factors, of which a relatively large number have now been identified. Exactly how the interactions between the different DNA elements and *trans*-acting factors regulate gene transcription is now beginning to be elucidated.

The next three sections present a brief summary of the nature and properties of some of the classes of these elements and proteins, and how they may cooperate to regulate transcription. However it is appropriate that a brief summary of prokaryotic transcriptional control be presented before eukaryotic transcriptional control is examined, as the characteristics of prokaryotic *cis*-sequences and their interaction with the transcription machinery bear considerable similarity to their eukaryotic counterparts.

1.3.1 Prokaryotic transcriptional control

Prokaryotic promoters contain both specific binding sites for RNA polymerase and regulatory sequences which bind regulatory factors that modulate transcriptional initiation. The two most highly conserved promoter elements are the Pribnow box (found at -10 with respect to the transcription initiation site at +1), and the -35 region. These are involved in binding RNA polymerase, where it is thought that RNA polymerase initially binds at the -35 region and then extends its contacts over the -10 region to cover approximately 60 bp of DNA. However RNA polymerase requires the presence of the factor σ^{70}, an additional subunit to the core enzyme, to be able to bind to the promoter region and initiate transcription. Interestingly, the consensus sequence for the Pribnow box, TATAAT, is very similar to the sequence for the 'TATA box' for eukaryotic RNA polymerase II genes (see 1.5.1). The A/T rich sequence may be able to be 'melted' more easily by the RNA polymerase enzyme, to initiate transcription, and so has been conserved between prokaryotes and eukaryotes (review, Hoopes and McClure, 1987).

Regulatory factors, repressors or activators, both influence the ability of RNA polymerases to initiate transcription. Generally, repressors bind to sequences within, or close to the promoter, and sterically prevent RNA polymerase binding. Activators generally bind just upstream of the -35 region, and are thought to cause activation directly by protein-protein interaction with RNA polymerase (review, Hoopes and McClure, 1987). Although most regulatory sequences are located close to the transcription initiation site, a growing number of sequences have been identified which can influence transcription when located a considerable distance from the promoter (reviewed, Gralla, 1989). These may act by a similar mechanism to that of eukaryotic enhancer sequences (see 1.5.2 (ii)).

Interestingly another type of promoter structure, although less common than the promoters using the -35 and -10 basal elements, has been identified that contains different basal elements located at -12 and -24 and also has the enhancer elements mentioned above. These promoters appear to be predominantly regulated by enhancer driven activation. Recognition of the -12 and -24 elements requires the RNA polymerase to associate with a different sigma subunit, factor σ^{54}. But this is not sufficient to activate transcription. In fact prior to activation the core polymerase plus factor σ^{54} are bound to the basal elements but the complex is unable to transcribe until the enhancer element bound by its activator is triggered to loop out and interact with the poised polymerase complex and initiate transcription (review, Gralla, 1991). This mechanism closely resembles one of the mechanisms proposed to be responsible for RNA polymerase II transcription activation.

1.3.2 Eukaryotic transcriptional control

The control of eukaryotic transcription is similar to that in prokaryotes but more complex. Eukaryotic transcription involves three different RNA polymerase molecules which transcribe discrete classes of genes. RNA polymerase I transcribes genes encoding large ribosomal RNAs (class I genes), RNA polymerase II transcribes genes that encode the wide variety of mRNAs that will be translated into protein and

some small stable nuclear RNAs (class II genes), and RNA polymerase III transcribes genes that encode tRNAs, 5S RNA, and RNAs involved in mRNA processing (U6 RNA) and protein transport (7SL RNA) (class III genes). These classes of genes also differ in their organisation and in the proteins that regulate their transcription (reviewed, Sawadogo and Sentenac, 1990). However there are similarities between the three classes of genes. The next section will briefly discuss eukaryotic class I and III genes, and the subsequent section will deal with the control of transcription of class II genes.

1.4 EUKARYOTIC CLASS I AND III GENES

1.4.1 Class I genes

Ribosomal RNA genes are generally arranged in clusters of tandem repeats in vertebrates, and possibly in many other eukaryotes as well, with the exception of yeast, and in addition there is also a common arrangement of associated regulatory elements (reviewed, Reeder, 1990). Each tandem repeat is arranged into two transcriptional units. One transcriptional unit contains the entire ribosomal RNA precursor (containing the 18S, 5.8S, and 28S rRNAs). The second transcriptional unit lies in the intergenic spacer between rRNA genes and contains a series of enhancer-like elements, and several transcription initiation sites. It has been suggested that the second transcription unit influences control of initiation of the first transcript, possibly by behaving as a 'sink' for RNA polymerse I or essential, but limiting, transcription factors thus providing these for the rRNA transcriptional unit promoter immediately downstream (Mitchelson and Moss, 1987).

There is evidence for at least three *trans*-acting factors involved in RNA polymerase I transcription initiation. These include a factor termed UBF (Bell *et al.*, 1988) that has homologues in all species examined so far and is able to bind to both the intergenic enhancer-like elements, as well as within the rRNA promoter. UBF is possibly involved in enhancer function and may also be involved in the RNA polymerase I initiation complex along with a second type of factor (termed SLI in humans) which

is highly species-specific (Learned *et al.*, 1985). Another factor, recently described, that associates with RNA polymerase I is believed to mediate polymerase recognition of the initiation complex containing UBF and SL1 (Mahajan and Thompson, 1990). The activity of this factor has been found to be modified in response to cell growth rate and influence the ability of RNA polymerase I to initiate rRNA gene transcription.

1.4.2 Class III genes

There are a number of different types of class III genes that have been identified, based on the type of promoter structure that they contain. These include genes with internal control regions (i.e. minimal promoter sequences located within the transcription unit), genes with internal control regions and 5' flanking control regions, and genes with only upstream promoters (reviewed, Murphy *et al.*, 1989).

The group of class III genes with internal control regions encode tRNAs, 5S RNA, and some small viral RNAs. Three promoter elements have been identified within the internal control regions of these genes, designated A, B and C boxes. The tRNA and the small viral RNA genes contain copies of both the A and B boxes, while the 5S RNA genes contain a copy of A and C boxes. Transcription factors TFIIIB, and TFIIIC, are required for the transcription of both sets of genes, with the 5S RNA genes requiring an additional factor, TFIIIA. TFIIIC binds to both the box A, and box B sequences, and TFIIIA binds to box C sequences. TFIIIB then binds to these factors, and in turn promotes the association of RNA polymerase III with the promoter (review, Palmer and Folk, 1990).

The 7SL RNA gene is an example of a class III gene with both an internal control region and additional important 5' flanking sequences located upstream of the transcriptional start site (Ullu and Weiner, 1985). These 5' sequences are required for efficient transcription of the 7SL RNA gene. However with other class III genes, the

5' sequences can be essential for transcription or act to exert either a positive or negative modulating effect, or be responsible for tissue-specific expression.

A quite different type of class III gene possess an upstream promoter only (reviewed, Murphy *et al.*, 1989). Examples include the genes that code for U6 RNA and 7SK RNA, a nuclear RNA of unknown function. These class III genes do not contain either box B, or box C, and, even though they both contain a box A sequence, it has been found non-essential for U6 RNA or 7SK RNA gene transcription, however the presence of the box A element may augment promoter efficiency (Carbon *et al.*, 1987). The promoter structure of these genes resembles class II genes (see 1.5), in that they contain enhancers, proximal promoter elements, and a TATA box. However, it has been confirmed that they are transcribed by RNA polymerase III (review, Murphy *et al.*, 1989). The mechanism by which RNA polymerase selection is achieved in these cases is unknown, however it has recently been determined that a factor with similar or identical DNA binding specificity to TFIID, an RNA polymerase II factor (see 1.5.1 (ii)), is required for U6 RNA gene transcription (Margottin *et al.*, 1991). This suggests a close evolutionary link between RNA polymerase II and III general transcription factors and their associated transcription machinery.

1.5 EUKARYOTIC CLASS II GENES

Transcription of class II genes culminates in the generation of transcripts ultimately translated into cellular protein. Class II genes have been the focus of most investigations into the processes and regulation of gene transcription. As a result a large number of different types of DNA sequences involved in class II gene transcription have been identified. The effects of these sequences are generally mediated through the action of proteins that bind to the DNA elements in a sequence-specific fashion. This section discusses the main classes of DNA sequence elements involved in controlling class II gene transcription and some of the proteins

that bind to these elements. Section 1.6 follows up in greater detail some of the more important functional characteristics of eukaryotic transcription factors.

1.5.1 Initiation elements and associated factors

(i) Specification of transcription start site

The majority of class II genes possess a sequence element called the TATA box (Corden *et al.*, 1980). This sequence functions primarily to ensure that transcripts are initiated accurately and serves as a general assembly point for the transcription machinery. The TATA box is located 25 - 30 bp upstream from the cap site in higher eukaryotes, and 40-120 bp upstream in yeast (review, Struhl, 1989).

However there are class II genes that do not contain TATA boxes. The promoters of these genes can be divided into two groups: one the G/C-rich promoters usually containing several Sp1 binding sites (Dynan and Tjian, 1983), and occasionally a series of direct repeats (Melton *et al.*, 1986), and second a group that includes the remaining promoters which have no TATA box and are not G/C-rich. G/C-rich promoters are found primarily in house-keeping genes (e.g.hypoxanthine phosphoribosyl transferase (Melton *et al.*, 1986) and dihydrofolate reductase (Nunberg *et al.*, 1980; Chen *et al.*, 1984), which are expressed constituitively, usually at a low level, in all cell types. Transcription from this type of gene often initiates at multiple sites. In contrast, many of the second group of non-G/C-rich promoters drive genes that are selectively expressed during differentiation or development. Examples include the *Drosophila* homeotic gene *Ultrabithorax* (Biggin and Tijan, 1988) and mouse terminal deoxynucleotidyltransferase gene involved in immunodifferentiation (Landau *et al.*, 1984). It is likely that the unique structure of these promoters may direct transcription of a select set of genes requiring strict expression during differentiation and development. Smale and Baltimore (1989) have identified an initiator element, Inr, a 17 bp element that includes the initiation start site, in these genes that directs faithful, single site, but low level, transcription

initiation. Interestingly, sequence homology to the Inr element is also found in many TATA box containing genes.

(ii) TFIID and other transcription initiation complex factors

TFIID is a general transcription factor that binds specifically to TATA box sequences. and may be regarded as a commitment factor whose binding to the TATA box is the prerequisite for assembly of the basal transcription aparatus. Initiation of transcription however requires other general transcription factors (TFIIA, TFIIB, TFIIE and TFIIF) and RNA polymerase II to then bind in a defined order to build a competent transcription complex. This series of events has been followed, by the increasing size of the protein complex associated with the DNA, by footprinting studies (Buratowski *et al.*, 1989). Assembly of the initiation complex begins with stable binding of TFIID to the TATA element in a reaction that may be facilitated by TFIIA. Next, TFIIB becomes involved and is believed to act as a bridge between promoter-bound TFIID and RNA polymerase II. TFIIE is then thought to bind the complex. An ATPase function has been suggested for TFIIE as ATP hydrolysis is required for transcription initiation. Another component TFIIF has also been identified and is thought to have some role in encouraging specific RNA polymerase binding to promoter sites (reviewed, Roeder, 1991). TFIIF may also remain associated with RNA polymerase II and influence elongation along with a number of other factors, TFIIS and TFIIX (Bengal *et al.*, 1991). TFIID remains bound to the TATA box following initiation (perhaps blocking nucleosome formation at the promoter), and may catalyse multiple rounds of transcription (Workman and Roeder, 1987). TFIID may be a direct target for activators, but it may also interact with activators indirectly (review, Lewin, 1990; review, Ptashne and Gann, 1990; also see 1.6.2 (i) and (ii)).

TFIID, in higher organisms, has recently been shown to consist of a TATA binding protein (TBP) and a number of associated protein factors, TAFs (reviewed, White and Jackson, 1992). TBP now appears to be a shared general transcription factor

involved in class I, II,and III gene transcription initiation (Cormack and Struhl, 1992; Schultz *et al.*, 1992). It seems likely that it is the TAFs that may confer class specificity upon the different RNA polymerase/general transcription factor complexes. For instance SLI, a component of the RNA polymerase I initiation complex, contains TBP and three TAFs which appear to be different from those in TFIID (see 1.4.1). And as described in 1.4.2, a class III gene coding for U6 RNA requires a TFIID-like factor for transcription, which has been shown to include TBP, although the specifc interacting factors in this case have not been precisely defined.

1.5.2 Regulatory elements directing transcription

A number of different regulatory elements/regions have been identified that are responsible for the correct regulation of gene transcription in addition to the TATA box or alternative initiation sequence determinants. These have been categorised to include promoter elements, enhancer elements and also, recently, another type of region called a locus control region.

(i) Promoter elements

DNA sequences located immediately upstream (typically within a few hundred bp) of the transcription start site are termed promoter elements and are involved in modulating the efficiency and specificity of transcription initiation. Collectively these elements, each represented by discrete functional modules of 7 - 20 bp of DNA, containing one or more binding sites for transcriptional activator or repressor proteins, make up the promoter. The spacing between these elements and their orientation within a promoter seem to be relatively flexible. Elements can be inverted or the spacing between individual elements varied to a degree while still maintaining substantial promoter activity (reviewed, Dynan, 1989). However, as demonstrated in the case of the herpes simplex virus (HSV) thymidine kinase gene, as the promoter elements are moved further apart promoter activity declines (McKnight, 1982). Where promoter elements were separated by odd multiples of half a DNA turn this was found to be more detrimental to transcriptional activity

than insertion of even multiples (Takahashi *et al.*, 1986), suggesting that the proteins interacting with the elements still required relatively precise alignment on the DNA helix to function.

Some promoter elements, such as G/C and CCAAT boxes, which display the above characteristics, are common to many genes transcribed by RNA polymerase II and as such seem to represent general promoter elements. The G/C box, often present in multiple copies in a promoter, is usually located 40 - 100 bp upstream of the start site, while the CCAAT box is typically located 70 - 90 bp upstream of the transcription start site (reviewed, Kadonaga *et al.*, 1986). The G/C box is recognised by the Sp1 transcription factor, and in the case of a number of genes the Sp1-G/C box interaction has been shown to be necessary to activate transcription e.g. human metallothionein IA and IIA genes, HSV thymidine kinase gene and the mouse dihydrofolate reductase gene. Although the G/C box is essential for binding Sp1, G/C boxes within different promoters bind with varying affinities, so the flanking sequences may be influencing binding and thus the efficiency of Sp1 activation of the different promoters (review, Kadonaga *et al.*, 1986). It is not known how Sp1 activates transcription. However recent work suggests that Sp1 may interact with TFIID, although indirectly through a coactivator, to activate transcription (Pugh and Tijan, 1990; also see 1.6.2 (ii)). The CCAAT consensus sequence is critical for the transcription of numerous genes e.g. the β-globin gene (Grosveld *et al.*, 1982), HSV thymidine kinase gene (Jones *et al.*, 1985), and the human heat shock gene (Bienz and Pelham, 1986). However, unlike the G/C box which only seems to interact with the single factor, Sp1, a number of different factors have now been found to recognise the CCAAT consensus including the NF-1/CTF family (Mermod *et al.*, 1989) and the CP family (Chodosh *et al.*, 1988). It remains to be seen whether each of these factors recognises a particular variant of the consensus sequence or whether binding is controlled by sequences flanking the CCAAT box. Thus the CCAAT box is perhaps not as general a promoter element as first thought, with the various factors perhaps providing, or at least contributing, to gene specific activation rather than just acting

in a general stimulatory capacity (e.g. CCAAT box-containing element involved in regulation of histone H1 gene transcription; see 1.11.2).

A large number of less common elements have been implicated in dictating more specific/individualized transcriptional regulation, such as in response to heat shock, to hormones and growth factors, or in the cases of tissue-specific or developmentally restricted regulation of transcription of particular genes. These elements can be binding sites for activators or repressors (see 1.6.3) or both. It is with these more specific promoter elements, often found upstream of the more general (proximal) promoter elements, that the distinction between promoter and enhancers elements (see below) begins to become blurred. A number of these specific promoter/enhancer elements, the genes they are associated with, and the factors that bind them are described in Table 1.1.

(ii) Enhancer elements

Eukaryotic transcription can also be greatly influenced by additional control regions known as enhancers. Often located kilobases either upstream or downstream of the transcription start site, enhancers can interact with a promoter, in an orientation-independent manner, to regulate transcription. Enhancers were originally identified as viral activator sequences (Benoist and Chambon, 1981), but analogous regions were soon found associated with cellular genes. These included constituitive enhancers, active in all cell types of a given organism (e.g. the simian virus (SV40) enhancer; Jones *et al.*, 1988), inducible enhancers that respond to changes in the external environment of a cell to activate (or repress) a gene (e.g. the heat shock element; Bienz and Pelham, 1986), and also, temporal and tissue-specific enhancers that are active only at specific times during development, or only in specific cell types (e.g. lymphoid specific immunoglobulin enhancer; Gillies *et al.*, 1983).

Enhancers, although more compactly organised, appear, like promoters to be composed of multiple individual elements, each of which binds one or more

Table 1.1 Eukaryotic promoter/enhancer elements

Presented in the table are various promoter/enhancer elements, the genes they are associated with and the transcription factors that bind to them. Note: N, represents any nucleotide.

References: (1) review, Faisst and Meyer, 1992; (2) Boyle *et al.*, 1991; (3) Auwerx and Sassone-Corsi, 1991; (4) review, Biggin and Tijan, 1989; (5) Winslow *et al.*, 1989; (6) Williams and Tijan, 1991; (7) Friedman and McKnight, 1990; (8) review, Yamamoto, 1985; (9) Sorger, 1991; (10) Murre *et al.*, 1989; (11) Sturm *et al.*, 1988*a*; (12) Carbon *et al.*, 1987; (13) O'Hare and Goding, 1988; (14) Clerc *et al.*, 1988; (15) Norman *et al.*, 1988; (16) Treisman, 1985.

Promoter/enhancer element	*Trans*-acting factor	Examples of element-containing genes Tissue specificity of *trans*-acting factor Other features, (Ref.)
$TGA^{G}/_{C}T^{C}/_{A}A^{A}$	AP1 (Homodimer of Jun or heterodimer of Jun and Fos protein families	Various viral and cellular genes e.g. SV40 T antigen, collagenase and metallothionein IIA genes. Ubiquitous. AP1 induced by eg. TPA, growth factors, serum and cAMP. (1, 2, 3)
TAATAATAATAATAA	Antp (Antennapedia)	*Drosophila Antennapedia* and *Ultrabithorax* genes. Involved in developmental process of segmentation. (4, 5)
$CCC^{A}/_{C}N^{G}/_{C}{}^{G}/_{C}{}^{G}/_{C}$	AP2	Several viral and cellular genes e.g. mouse major histocompatibility complex (MHC) genes. Most abundant in early neural crest cells and involved in their differentiation and development. (1, 6)
CCAAT $TGTGG^{T}/_{A}{}^{T}/_{A}{}^{T}/_{A}{}^{G}$	C/EBP (CCAAT/enhancer binding protein)	Serum albumin and phosphoenolpyruvate carboxykinase (PEPCK) genes. Liver specific. (1, 7)
$^{T}/_{A}GATA^{G}/_{A}$	GATA-1	Globin genes. Erythroid specifc; required for erythroid differentiation. (1)
AGAACANNNTGTTCT (Glucocorticoid responsive element; GRE)	GR (Glutocorticoid receptor)	Human metallothionein IIA and growth hormone genes. GR induced by glucocorticoid. (1, 8)
$(NGAAN)_3$ (Heat shock element; HSE)	HSF (Heat shock factor)	e.g. hsp70 (heat shock protein) gene. Ubiquitous. HSF Induced by heat shock. (1, 9)
CAACTGAC	MyoD	Muscle creatine kinase (MCK) and myosin light chain genes. Muscle specific; proliferating myoblasts and differentiated myotubules. (1, 10)
ATGCAAAT	Oct1	Histone H2B, U2 snRNA and HSV immediate early genes. Also functions as a DNA replication element in the adenovirus origin of replication. Ubiquitous. (1, 11, 12, 13)
ATGCAAAT	Oct2	β-globin and immunoglobulin heavy and light chain genes. Lymphoid specific. (1, 14)
GGATGTCCATATTAGGACATCT (Serum response element; SRE)	SRF (Serum response factor)	c-*Fos* and insulin genes. Ubiquitous. SRF induced by serum. (1, 15, 16)

transcriptional activating (or repressing) proteins (reviewed, Dynan, 1989). Since deletion of one element can be compensated by the duplication of another, these elements appear to be functionally interchangeable. Individual elements from an enhancer can also be found in other enhancers and promoters. Furthermore it has been shown that a promoter element can be synthetically multimerised and gain the properties of an enhancer. Thus promoters and enhancers may be homologous entities, and the transcriptional activator proteins that bind to these sequences, interact with the transcriptional machinery by a fundamentally similar mechanism (review, Dynan, 1989).

Although substantial progress has been made in elucidating the structure and function of many different transcription factors that bind promoter and enhancer regions, the mechanisms by which these complexes interact with each other and the initiation complex to bring about transcription are largely unknown. Although a number of models have been proposed to explain how enhancers/promoters function the two most likely models appear to be the scanning model and the looping model (reviewed, Müller and Schaffner, 1990). For the scanning model, enhancer/upstream promoter elements are recognised by RNA polymerase II, or other transcription factors, which bind to the DNA in a sequence-independent fashion, and then slide along the DNA in either direction until proximal promoter elements are reached to then help form a transcription initiation complex. The looping model involves looping out the DNA that separates the enhancer/upstream promoter elements and proximal promoter so as to bring these regions together and permit the proteins, bound to the respective DNA elements, to physically interact with each other and/or interact directly with the initiation complex and effect transcription. Experimental evidence in favour of this model has been that enhancers can act in *trans*, that is via linked circles of DNA (Dunaway and Dröge, 1989). Müller *et al.* (1989) have shown that an enhancer can still stimulate transcription *in vitro* even when attached to a promoter via a protein bridge which would be expected to prevent the functioning of a scanning type mechanism. In

addition, homologous paired chromosomes in *Drosophila*, with an enhancer on one DNA molecule, and a promoter on the other, are still able to activate transcription albeit at a reduced level (Geyer *et al.*, 1990). However, these results although consistent with the looping model do not strictly exclude possible scanning mechanisms, and as such leave some doubt as to the exact mechanism of enhancer action.

(iii) Locus control regions

In addition to the promoter and enhancer elements described above, which are responsible for the correct regulation of gene transcription, there is another recently described type of region, termed a locus control region (LCR), that has been shown to be involved in the regulation of transcription (reviewed, Orkin, 1990). This type of control region was first described in the β-globin gene family (Grosveld *et al.*, 1987), but similar regions have recently been described associated with other genes such as the human CD2 gene (Greaves *et al.*, 1989). In the case of the β-globin gene family, the LCR, is located approximately 50 kb upstream, and consists of a number of DNase I hypersensitive regions (see 1.2.4), which when linked to the β-globin gene, or heterologous genes and introduced into mice, confer high-level, gene-copy-number-dependent expression on these genes, independent of the site of integration in the host genome (Grosveld *et al.*, 1987). Furthermore, deletion of the LCR located approximately 30 kb upstream of the α-globin gene family, as occurs in some rare forms of thalassemia, inactivates the entire α-globin gene complex (review, Orkin, 1990).

The remarkable and unusual conservation, between species, of long stretches of sequence in the noncoding DNase I hypersensitivity domains that make up the β-globin LCR support a claim that the LCR constitutes a new type of regulatory element (review, Orkin, 1990). Although the actual mechanism by which the LCR functions is unknown, it has been proposed that the LCR region works by initially organising the entire β-globin gene locus into a transcriptionally competent (DNase I

sensitive) domain (see 1.7.2 (v)). This permits regulatory factors to gain access to individual genes in the gene family, and then, additionally serve as an enhancer of β-globin gene transcription. GATA (bound by GATA-1; an erythroid specific DNA-binding protein), AP-1 (in this case proposed to bind the erythroid factor NF-E2), and CACCC binding sites have also been identified in the LCR (review, Orkin, 1990). The role of the LCR as a chromatin 'opener' also suggests the involvement of unique protein factors, perhaps mediating nuclear matrix attachment.

1.6 EUKARYOTIC TRANSCRIPTION FACTORS

The successful isolation of genes encoding eukaryotic transcription factors, has lead to an explosion of information on the structure and function of these factors. It has been found that a number of independent domains are necessary for transcription factors to function. These include domains necessary for DNA binding and for protein-protein interaction. This section describes some of the different domain types involved in both of these processes, and briefly reviews how the transcription factors are themselves regulated.

1.6.1 DNA binding domains of transcription factors

(i) Zinc finger

One of the first transcription factors to be purified and cloned was TFIIIA, a protein required for the transcription of 5S ribosomal RNA genes by RNA polymerase III. The DNA binding region of TFIIIA was found to involve an unusual sequence motif comprising 30 amino acids repeated consecutively nine times. This motif was defined by two invariant pairs of cysteines and histidines as well as a number of other conserved amino acids residues. Furthermore, TFIIIA was found to be complexed with zinc, and that zinc was required for DNA binding. Consistent with these findings it was proposed that the cysteine and histidine residues served as a tetrahedral coordination site for a single zinc atom, and that the amino acids between these coordinating residues looped out to form a finger-like stucture, hence the name 'zinc finger' (reviewed, Klug and Rhodes, 1987). The fingers interact with the DNA

by making direct contact with the major groove of the DNA. These structural predictions for the zinc finger (Miller *et al.*, 1985) have since been confirmed by three-dimensional NMR (Lee *et al*, 1989) and x-ray crystallography (Pavletich and Pabo, 1991) studies. The zinc finger, which forms a compact globular domain, is found to contain an antiparallel β ribbon and an α helix. The two cysteines, located near the turn in the β ribbon, and the two histidines, in the carboxyl terminal portion of the α helix, are coordinated by a central zinc atom (Figures 1.1A and 1.1B). DNA binding seems to be through residues located in the amino terminal portion of the α helix.

Many examples of transcription factors containing zinc fingers, from a wide variety of eukaryotic organisms and also involved in a wide variety of regulatory roles, have been identified. Examples include: the general transcription factor Sp1 (Kadonaga *et al.*,1987), *Krüpple* (Rosenberg *et al.*, 1986) and *hunchback* (Tautz *et al.*, 1987) gene products involved in *Drosophila* development, the yeast SWI5 protein involved in mating type switching (Nasmyth *et al.*, 1987) and the MBP-2 protein involved in regulating the human major histocompatibility complex class I genes (van't Veer *et al.*, 1992). There is considerable variation in the number and organisation of the zinc fingers in different transcription factors. From two up to ten or more fingers in the one protein have been reported (Hartshorne *et al.*, 1986; Page *et al.*, 1987). While it seems the most common finger arrangement involves fingers organised into a single cluster of tandem repeats other structural patterns including individual fingers dispersed throughout the protein (e.g. the *Drosophila Suvar(3)7* gene product; Reuter *et al.*, 1990) or groups of fingers organised as 'hands' (e.g. the *Xenopus Xfin* gene product; Ruiz i Altaba *et al.*, 1987) have also been observed in other regulatory proteins. Another interesting observation is that zinc fingers can exhibit systematic sequence alterations in every other finger element (e.g. alteration between three and four inter-histidine residues; Page *et al.*, 1987). In this case the repeat unit would appear to be a pair of mutually distinct fingers and it has been suggested that proteins with this type of finger arrangement actually bind along one face of the DNA double helix, with alternating fingers making inequivalent DNA contacts

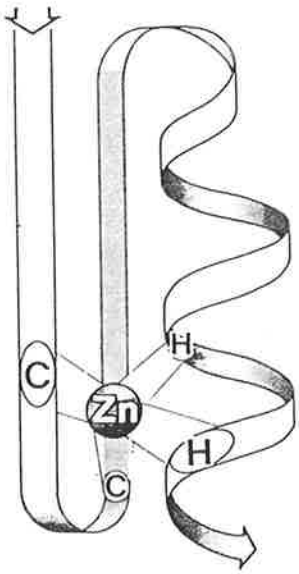
Figure 1.1 Structures of transcription factor DNA binding domains

A Schematic representation of a zinc finger (figure adapted from review, Evans and Hollenberg, 1988). NMR (nuclear magnetic resonance) studies have shown the zinc finger stucture to consist of an antiparallel β ribbon and an α helix. The two cysteines, located either side of the turn in the β ribbon, and the two histidines in the carboxyl terminal portion of the α helix are coordinated by a central zinc atom (Lee *et al.*, 1989).

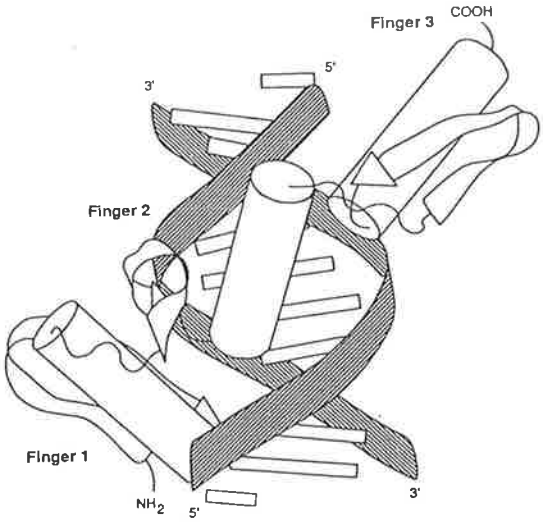
B Diagram showing the relation of a set of contiguous zinc fingers with respect to each other and with respect to the DNA (figure adapted from Pavletich and Pabo, 1991). The contiguous zinc fingers follow the major groove of the DNA with the amino terminal portion of the α helix of each zinc finger making base pair contacts. This structure was generated from x-ray crystallography analysis of the interaction between a peptide containing the 3 zinc finger-DNA binding domain of the mouse immediate early protein Zif268 and an oligonucleotide containing the Zif268 binding site (Pavletich and Pabo, 1991).

C Diagram showing the zinc binding motif of the glucocorticoid receptor (figure adapted from Luisi *et al.*, 1991). Actually depicted are two DNA binding regions of a dimerised complex of receptors bound to DNA. Each receptor consists of two zinc binding modules folded together to form a unified globular domain. Each zinc binding module coordinates two zinc atoms (represented by black discs). This structure was generated from x-ray crystallography analysis of the interaction between a peptide containing the DNA binding domain of the rat glucocorticoid receptor and an oligonucleotide containing the appropriate GRE (glucorticoid reponse element) (Luisi *et al.*, 1991)

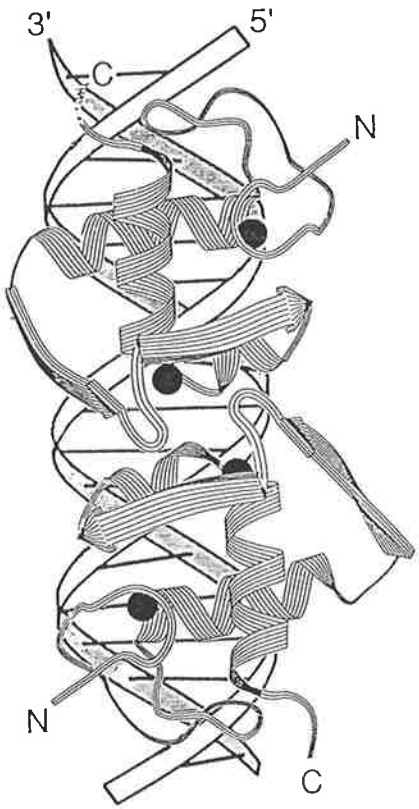
A



B



C



(Fairall *et al.*, 1986; Weiss *et al.*, 1990). All of the above suggest the versatility of the zinc finger as a DNA binding unit, that readily adapts to suit the DNA binding requirements of different regulatory proteins.

A similar, but distinct, zinc binding motif is found in the DNA binding regions of the steroid hormone receptor family (reviewed, Beato, 1989). Representative of this class of motif is the recently described DNA binding structure of the glucocorticoid receptor (Luisi *et al.*, 1991). The motif consists of two zinc binding modules, each containing two pairs of cysteine residues coordinating the zinc atom, instead of the cysteine and histidine pairs described above. The two zinc binding modules fold together to form the single structural DNA binding domain in these receptors (Figure 1.1C). In agreement with a single copy of the motif per receptor, the cognate binding sites are relatively short and involve dyad symmetry, consistent with receptor dimerisation for DNA binding (review, Schwabe and Rhodes, 1991).

The zinc binding structure present in the yeast transcription factor GAL4 represents a third motif type (Marmorstein *et al.*, 1992). The motif, incoporates a zinc coordination scheme similar to that of the metallothionein group of proteins. GAL4 binds two zinc atoms through six cysteine residues in the motif, with the zinc atoms sharing two of the cysteine ligands (Figure 1.2A). Despite all three motifs described above binding zinc, each motif has an obviously different structure. This suggests, along with the disparities in primary amino acid sequence, that the three types of motif are not evolutionarily related.

(ii) Homeodomain

The homeobox was first identified as a region of homology common to a number of homeotic genes of Drosophila, for example *Antennapaedia, Ultrabithorax* and *engrailed* (reviewed, Levine and Hoey, 1988). The 60 amino acid homeodomain was subsequently shown to be critical for the DNA binding ability of these proteins. The homeodomain appears to have been extensively conserved during evolution as

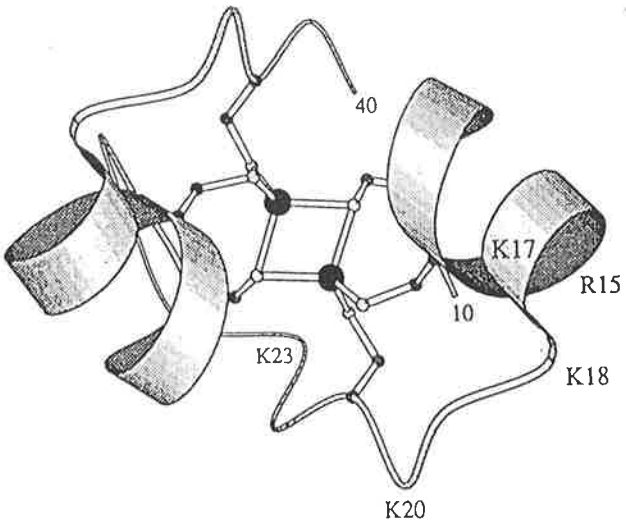
Figure 1.2 Structures of transcription factor DNA binding domains

A Diagram showing the GAL4 DNA-binding domain (figure adapted from Kraulis *et al.*, 1992). In this type of zinc binding motif two zinc atoms (represented by the large black spheres) are coordinated via six cysteine residues (positions 11, 14, 21, 28, 31 and 38 within the GAL4 DNA binding domain; not actually numbered in the diagram but represented by small black spheres included in the polypeptide backbone).

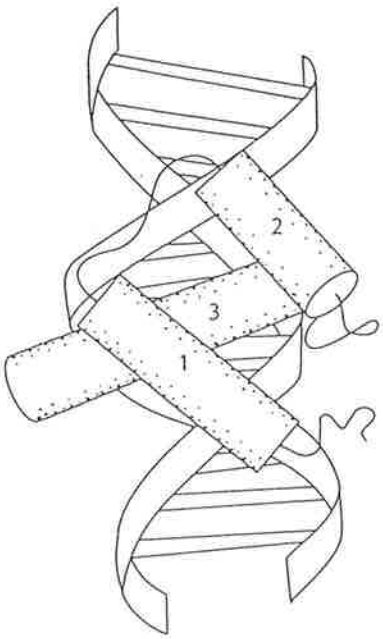
B Schematic representation of the interaction of the homeodomain region with a DNA site (figure adapted from review, Carey, 1991). The 'recognition helix' (helix 3) binds in the major groove of the DNA. However, although helix 3 and helix 2 match closely the helix-turn-helix conformation found in the bacterial repressors, the positioning of the helices with respect to the major groove and the amino acid positions that are involved in DNA binding are very different. Furthermore another helix (helix 1) in the homeodomain is found to make additional base pair contacts in the minor groove of the DNA (review, Gehring *et al.*, 1990).

C Diagram showing the dimerised structure of the basic region-leucine zipper motif bound to DNA (figure adapted from review, Lamb and McKnight, 1991). The white rectangles represent the α helices of the leucine zippers dimerised to form a coiled coil structure. The interaction of the leucine zippers positions the basic region domains (represented by the shaded rectangles), also predicted to form α helical structures, such that they can recognise and bind a dyad symetric DNA binding site (review, Lamb and McKnight, 1991).

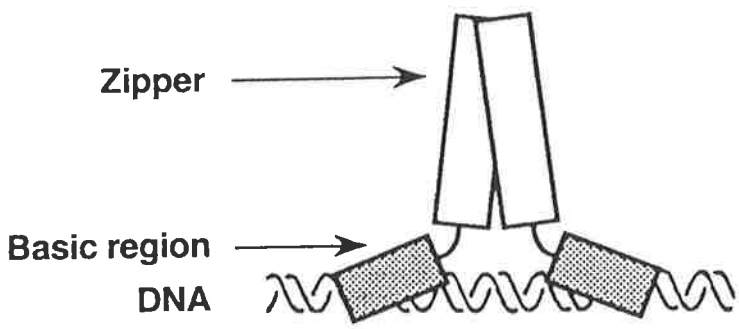
A



B



C



similar homeodomains have since been found in regulatory factors in many different eukaryotic organisms from yeast (e.g. mating type proteins MAT a1 and MAT α2; Shepherd *et al.*, 1984) to humans (Levine *et al.*, 1984).

Structural predictions of the homeodomain indicated that it could form a helix-turn-helix motif similar to the well characterised helix-turn-helix motif found in prokaryotic repressor and activator proteins (reviewed, Harrison and Aggarwal, 1990). The determination of the three dimensional structure of the Antennapaedia homeodomain has confirmed that the homeodomain indeed does form a helix-turn-helix motif (Otting *et al.*, 1990). In the case of the bacterial repressors one of the helices (the 'recognition helix') directly contact bases in the major groove of the DNA, while the other helix lies across the major groove, and makes some non-specific contacts to the DNA. These proteins bind as dimers to DNA sequences with dyad symmetry. However, although a similar 'recognition helix' role for one of the helices has been established, homeodomain proteins bind as monomers. This, together with findings that contacts outside the helix-turn-helix motif contribute to DNA binding, suggests that the homeodomain mode of DNA binding may be quite different from that of the prokaryotic helix-turn-helix motif (review, Gehring *et al.*, 1990) (Figure 1.2B).

More recently, another class of regulatory proteins has been identified in which the homeobox forms part of a larger conserved domain, referred to as the POU domain, that also includes another 75 - 80 amino acid POU specific region. Proteins included in this group are the pituitary specific protein Pit-1 (Ingraham *et al.*, 1988), the octamer binding factors Oct-1, a ubiquitous transcription factor (Sturm *et al.*, 1988*a*) and Oct-2, a tissue (lymphoid) specific transcription factor (Clerc *et al.*, 1988), and the product of the nematode cell lineage control gene, *unc-86* (Finney *et al.*, 1988). The relative contributions of the homeodomain and POU-specific domain to DNA binding, vary between the different proteins (Theill *et al.*, 1989; Sturm and Herr, 1988).

Other additional conserved sequence elements identified, with the homeobox, include the PRD repeat (Frigerio *et al.*, 1986) and the paired box (Bopp *et al.*, 1986) found in several *Drosophila* homeobox genes. Perhaps these domains, along with the POU specific region are involved in providing additional specificity to DNA binding or may participate in protein-protein interactions.

(iii) Leucine zipper

Another motif, the leucine zipper, is found in several transcription factors, including the liver specific transcription factor C/EBP (which binds both CCAAT recognition and enhancer core sequences), the yeast factor GCN4, and the c-Jun, c-Fos and c-Myc oncoproteins (reviewed, Busch and Sassone-Corsi, 1990). This motif consists of an α helical structure containing four or five leucine residues, spaced exactly seven residues apart, such that the leucines occur every two turns and occur on the same side of the helix. However, rather than acting directly as a DNA binding motif, this structure probably facilitates the dimerisation of two leucine-containing helices from different molecules via a coiled coil structure. In agreement with this model, GCN4, C/EBP, and c-Jun all bind to DNA as dimers, and c-Fos and c-Jun can form a heterodimer. Dimerisation results in the generation of the correct protein structure for DNA binding through an adjacent domain, rich in basic amino acids, and predicted to form an α-helical structure (Figure 1.2C). It is has been proposed that the basic region-leucine zipper (bZIP) proteins when bound to DNA take up a Y shaped conformation with the dimeric leucine zipper forming the base of the Y and the helical basic regions forming the arms ('scissor grip' model; Vinson *et al.*, 1989). However, detailed structural confirmation of this class of DNA-binding motif has yet to be carried out.

Dimerisation may represent a common property of sequence specific DNA binding proteins (see also 1.6.1 (ii) and (iv)), that reflects the increased efficiency of binding that it offers the proteins (review, Schleif, 1988). The requirement for dimer formation of these bZIP proteins, prior to DNA binding, also introduces another

potential level of regulation to gene transcription, since different combinations of protein homo- and heterodimers can influence DNA binding specificity. Additionally, where there are multiple related proteins with similar dimerisation and DNA binding activities, but with different abilities to influence transcription, (e.g. c-Jun and c-Fos family members; Kouzarides and Ziff, 1989), the formation of different heterodimers, as well as homodimers has the potential to effect differential expression of the target gene(s). Furthermore, the use of homo- and heterodimerisation dramatically expands the regulatory potential of a system while maintaining an economy of components.

(iv) Helix-loop-helix

Although originally identified in leucine zipper containing proteins, the basic DNA binding domain has also been found in several other transcription factors including the immunoglobulin enhancer binding proteins E12 and E47, the muscle regulatory protein MyoD, c-Myc, and the *Drosophila daughterless* gene product (Murre *et al.*,1989). The basic domain in these proteins is associated with a region that can form a helix-loop-helix structure in which two amphipathic helices are separated by a non-helical loop. This class of protein can also form homo- and heterodimers, and it is the helix-loop-helix structure that is believed to mediate protein dimerisation and facilitate DNA binding by the adjacent basic domain. Consistent with this, introduced mutations in the basic region eliminate DNA binding but not dimerisation (Lassar *et al.*, 1989). Thus, the mechanisms of DNA binding of this class of protein may be similar to that of the leucine zipper proteins described above, but they may dimerise via a different structure.

Interestingly, the c-Myc protein contains both a leucine zipper and helix-loop-helix motif along with the basic domain (review, Kato and Dang, 1992), and the Oct-2 factor contains all three structures in addition to the POU domain (Clerc *et al.*, 1988; reviewed, Busch and Sassone-Corsi, 1990). It would appear that these motifs are not

mutally exclusive, and therefore suggest the existence of sub-families of factors containing combinations of these motifs.

(v) Other DNA binding domains

A number of other DNA binding regions have recently been identified that do not appear to be related to any of the motifs described above. Several examples of these are described below.

A novel and highly conserved 85 amino acid DNA binding domain, the ETS domain (reviewed, Karim *et al.*, 1990), has been identified in the *ets* gene family of proteins (comprising the Ets-1, Ets-2, Elk-1, Elk-2 proteins and the *ets* related gene products, Erg, *Drosophila* E74 and murine PU.1). While the mechanism by which the ETS domain binds to DNA is unknown, it has been suggested that three highly conserved tryptophan residues repeated every 18 amino acids in the ETS domain may play an important role in the structure and function of the domain (a similar arrangement of tryptophan is seen in the Myb proteins; Biedenkapp *et al.*, 1988).

The DNA binding/dimerisation motif found in the heat shock transcription factor (HSF) also appears to be novel (Sorger and Nelson, 1989). It involves an amphipathic α helix containing hydrophobic amino acids every 7 residues, which directs HSF to form a trimeric coiled coil structure. A second adjacent α helix is proposed to stabilize this interaction, but it is not clear how HSF actually binds DNA.

The MADS box represents a further DNA binding/dimerisation region that does not appear to be related to any previously identified DNA binding motif. It has been found in a number of different transcription factors, from which it derives its name, including the yeast mating type specific transcription factor MCM-1 (Passmore *et al.*, 1988), the products of two plant homeotic genes *agamous* (Yanofsky *et al.*, 1990) and *deficiens* (Sommer *et al.*, 1990), and the human serum response factor SRF (Norman *et al.*, 1988).

Regions responsible for the DNA binding ability of proteins such as the CCAAT box binding protein CTF/NF1 (Mermod *et al*, 1989, and the transcription factor AP2 (Williams *et al.*, 1988) also seem to be different to any of the DNA binding motifs previously described.

1.6.2 Transciptional activation domains

Following specific DNA binding, a transcription factor must interact with other factors or the RNA polymerase to be able to activate (or repress) transcription. A number of different transcriptional activation domains in transcription factors have now been identified. Some of these are discussed in this section, while several examples of transcriptional repression are dealt with in the following section.

(i) Acidic domain

The acidic activation domain was initially identified after activation regions of a number of transcription factors were compared and shown that they did not share amino acid homology but did contain a similarly high proportion of acidic amino acids (e.g. the yeast proteins GAL4 and GCN4; Hope and Struhl, 1986; Ma and Ptashne, 1987*a*). Additionally, Ma and Ptashne (1987*b*) found that transcriptional activation could be achieved by fusing the GAL4 DNA binding domain to random polypeptides encoded by fragments of *E. coli* DNA. Many examples of the activating peptides were found to contain an excess of acidic amino acids but also no obvious sequence homology. Interestingly, the strength of activation of these fragments seem to roughly correlate to the abundance of negative charge (this was also found to be the case with various mutants of the GAL4 activating domain; Gill and Ptashne, 1987). As the correlation between charge and activation strength was not always observed, some other structural aspect, in addition to the amount of negative charge, was also important for activating capability. In this respect, many of the *E. coli*-encoded activating peptides had the acidic amino acids arranged in such a way that they formed an amphipathic α-helix with all negative charges displayed along one surface of the helix. The importance of this arrangement was demonstrated by

Giniger and Ptashne (1987) showing that a peptide which could form an acidic amphipathic helix was able to activate transcription when linked to the DNA binding domain of GAL4, but the random distribution of the same negatively charged amino acids on the helix were unable to do so.

The acidic domain appears to be able to carry out its activating function from various positions in the protein with respect to the DNA binding region, and even to activate in heterologous systems (Brent and Ptashne, 1985; Hope and Struhl, 1986). This is consistent with the activation domain representing an independent functional unit, and also, either having a relatively relaxed interactive specificity or being able to interact with a conserved target protein. Perhaps, the acidic activation domains activate transcription by direct interaction with the other proteins of the transcriptional initiation complex, specifically either TFIID (Stringer *et al.*, 1990), or TFIIB (Lin and Green, 1991). Alternatively, it has been suggested that these domains act via another type of factor called a co-activator (reviewed, Lewin, 1990; reviewed, Ptashne and Gann, 1990). Another example of a transcription factor containing an acidic activation domain is the HSV protein, VP16, which possesses an intensely acidic domain in line with its potent activating ability (Triezenberg *et al.*, 1988). However, VP16 does not bind to DNA directly but forms complexes with other cellular factors that do bind DNA to activate transcription (e.g. VP16 interaction with Oct1; Stern *et al.*, 1989)

(ii) Non-acidic domains

Although, the acidic activation domain appears to be common among many transcription factors from a wide variety of eukaryoyic organisms, other activation domains have also been identified. One of these activation domains, rich in glutamine, was initially found in the mammalian transcription factor Sp1 (Courey and Tjian, 1988). Similar regions have also been described in *Antennapaedia, Ultrabithorax, zeste* gene products (Courey and Tjian, 1988), Oct1 (Sturm *et al.*, 1988*a*), Oct2 (Clerc *et al.*, 1988), AP2 (Williams *et al.*, 1988) and SRF (Norman *et al.*, 1988).

Characteristically this domain contains very few charged amino acids, but as its name suggests has a high proportion (20 - 30%) of glutamine residues. However, only in the cases of Sp1 (Courey and Tjian, 1988) and Oct2 has the correlation between activation activity, and the glutamine-rich domain been demonstrated (Müller-Immerglück *et al.*, 1990). Another type of non-acidic activation domain, originally identified in CTF/NF1 (Mermod *et al.*, 1989), is a domain rich in proline (typically 20 - 30% proline). Transcription factors also found to contain regions rich in proline residues, include AP-2 (Williams *et al.*, 1988), Oct2 (Clerc *et al.*, 1988) and SRF (Norman *et al.*, 1988). Other types of activation domain have been reported that contain no preponderance of a particular amino acid, nor a particular net charge (e.g. in C/EBP; Friedman and McKnight, 1990).

Interestingly like the acidic domain, both the glutamine- and proline-rich domains, also demonstrate the ability to activate when in various positions in the protein with respect to the DNA binding region, and to activate heterologous systems. For instance the glutamine-rich sequence from Antennapaedia can partially substitute for the transcription activation domains of Sp1 (reviewed, Mitchell and Tjian, 1989), and the CTF-1 proline rich domain can activate heterologous promoters (e.g. SV40; Mermod *et al.*, 1989). Thus, these domains may operate by a similar mechanism to acidic domains, although probably by interacting with other components of the initiation complex or accessory factors.

1.6.3 Transcriptional repression

To date, although the majority of transcription factors identified act as activators of transcription, a number of cases have now been described in which a transcription factor inhibits transcription. Considerably less is known about the regulatory domains involved and the mechanism by which DNA binding proteins repress transcription. However, several examples of negative transcriptional control have been observed, and these can be grouped into three general mechanistic classes.

These include inhibition of DNA binding, blocking of activation, and 'silencing' or direct repression (reviewed, Renkawitz, 1990).

Inhibition of DNA binding involves a repressor protein binding to a promoter or enhancer region of a gene to influence either the binding of a positive factor or the binding of the transcription initiation complex. This type of transcriptional repression in eukaryotes is related to the most common mechanism of prokaryotic transcriptional repression (see 1.3). An example of inhibition of DNA binding causing transcriptional repression is seen in the β-interferon promoter where the binding of two positively acting factors is necessary for gene activation. Another factor, a repressor, is found to act negatively by binding to this region of DNA and preventing the positively acting factors from binding. However in response to viral infection, the repressor is inactivated, allowing the positively acting factors to bind and transcription to proceed (Goodbourn *et al.*, 1986; Zinn and Maniatis, 1986).

Repression by blocking of activation, involves a repressor protein interfering with a DNA bound activator protein and preventing it from interacting with the transcriptional initiation complex. This interference may take the form of a negative factor binding to the DNA between the activator protein and the initiation complex, or a protein binding directly to the activation domain of the positive factor (quenching), or to an intermediary protein (squelching) (review, Renkawitz, 1990). An example of the first form of blocking of activation, described above, is found to occur in the *c-myc* gene. The negative factor Myc-PRF binds a repressor element next to the binding site for the positive acting Myc-CF1 and prevents this factor activating transcription (Kakkis *et al.*, 1989). The phenomenon of quenching has been demonstrated with the negatively acting yeast protein GAL80 preventing gene activation by the GAL4 protein, by binding to it and masking the activation domain of GAL4. However in response to the presence of galactose, GAL80 dissociates allowing GAL4 to activate transcription of the genes required for galactose metabolism (Johnston *et al.*, 1987; Ma and Ptashne, 1988). A possible example of

squelching may be used by the adenovirus E1a protein. E1a is able to a repress transcription of genes controlled by certain viral and cellular enhancer elements. A mechanism has been suggested where E1a does not bind to all these different factors, but instead binds to and inactivates an intermediary factor that would normally transmit the positive signal from the different activating factors to the initiation complex (Lillie *et al.*, 1986).

In all of the above cases the negative factor exerts its inhibiting effect by neutralising the action of a positively acting factor by preventing either its DNA binding or its activation of transcription. However, some factors may have an inherently negative effect and may directly inhibit transcription via a discrete domain analogous to the activation domains. Silencing may represent such a mechanism, where a factor binds to a defined DNA element (see 1.5.2 (i)) and directly represses/inactivates the transcription initiation complex. The first described example of a silencer was in the yeast mating type locus (Brand *et al.*, 1985). Another example of silencing has been observed in the chicken lysozyme gene, where a silencer element has been identified approximately 2 kb upstream of the start site. This silencer has been found to be able to function individually in repressing transcription from a minimal TATA promoter, suggesting direct interaction with the initiation complex may be involved in repression (Baniahmad *et al.*, 1990). However, the mechanism by which silencing may be taking place in both of these cases is unclear.

A further point is that some regulatory proteins can function as both an activator and repressor. For example the homeobox protein, Ultrabithorax, has been shown to possess dual function, suppressing *Antennapedia* promoter activity while activating the *Ultrabithorax* promoter (Krasnow *et al.*, 1989). The mechanisms causing the differences in Ultrabithorax regulatory activity are unknown. However it is clear from this and other examples, that there is some confusion in the interpretation of what originally was thought to represent two separate classes of proteins. One explanation is that some eukaryotic transcription factors may possess discrete

activating and repressing domains. The relative contributions of the domains will depend on protein conformation, which could be influenced by, for example, slightly different DNA binding sites for the protein and/or interactions with other proteins (review, Levine and Manley, 1989).

1.6.4 Control of transcription factor activity

What controls the controllers? It is important that transcription factors are themselves correctly regulated so as to be able to appropriately activate/repress their respective target genes. This is achieved in different cases either by controlling the synthesis of the transcription factor so that it is made only when necessary or by regulating the activity of the protein so that pre-existing protein becomes activated when required (review, Mitchell and Tijan, 1989).

Transcription factors are transcriptionally regulated, e.g. restricted expression of Oct -2 (Müller *et al.*, 1988) and C/EBP (Xanthopoulos *et al.*, 1989) in specific cell types, and negative autoregulation of c-Fos (Sassone-Corsi *et al.*, 1988). Post-transcriptional regulation mechanisims are also used to regulate the synthesis of transcription factors, as regulation at the transcriptional level only sets the problem of transcription factor regulation one stage further back, requiring mechanisms to control the transcription of the gene encoding the transcription factor.

Regulation of transcription factor activity by post-translational mechanisms also appears to be very common. This is not surprising considering the obvious advantages of this level of regulation in terms of producing a rapid and effective response as well as avoiding the complications mentioned above. Forms of post-translational regulation of transcription factor activity include ligand binding (e.g. steroid hormone receptor family; reviewed, Yamamoto, 1985), sequestration (e.g. NFκ-B activated after release from cytoplasmic complex with inhibitory protein IκB; Ghosh and Baltimore, 1990), and protein modification (e.g. phosphorylation of CREB; Gonzalez and Montminy, 1989).

1.7 CHROMATIN

In eukaryotes, transcription takes place in an environment in which DNA is packaged into chromatin. Although the rate and specificity of eukaryotic gene transcription are regulated primarily via specific binding of *trans*-acting factors to *cis*-acting regulatory elements, the differential availability of the DNA to the transcription apparatus, because of its association with histone proteins, represents another overall level of control of gene expression. This section discusses the structure of chromatin, and its influence upon transcription.

1.7.1 Chromatin structure

In eukaryotic cells, the DNA is complexed with specific nuclear proteins to form a structure called chromatin. The basic unit of chromatin is the nucleosome (reviewed, Morse and Simpson, 1988), in which approximately 200 bp of DNA is associated with a histone protein core, consisting of two molecules each of the four core histones, H2A, H2B, H3, and H4 (see 1.8 for more detailed description of histone proteins). Approximately 146 bp of DNA is wrapped around the core histone octamer, while the remainder of the DNA serves to link one nucleosome to another. A single molecule of a linker histone (generally H1) is found associated with the exterior of the nucleosome core complex and with the linker DNA. The organisation of DNA into nucleosomes generates a 10 nm beaded fibre structure, but this represents only the first level of DNA packaging. This structure is packed further, with cross-linking of the H1 linker histones, to form a 30 nm coiled fibre structure (reviewed, Felsenfeld and McGhee, 1986). The 30 nm fibre is present in interphase nuclei as looped domains, 35 - 85 kb in length, which when tightly packed during cell division, form mitotic chromosomes.

1.7.2 Active chromatin

There is growing evidence that genes being transcribed still have a nucleosomal organisation (reviewed, Grunstein,1990; see van Holde *et al.* (1992) for a detailed model of transcription through nucleosomes). However the composition and

structure of the chromatin associated with these genes appears to be different from that found with non-expressed genes. The altered chromatin structure associated with expressing (and potentially expressible) genes can be detected by a characteristically heightened sensitivity to DNase I digestion, which is consistent with a more open/transcriptionally accessible form of chromatin. Such an active chromatin structure is believed to be generated in a number of different ways, and some of these are discussed below.

(i) Involvement of histone H1

Most eukaryotic genes are inactive in most tissues, and become active only in one tissue or in response to a particular signal. Therefore it makes sense to have a system in which the gene is constitutively inactive in most tissues without a specific repressor being required. Higher-order chromatin would appear to function as this general repressor of inappropriate gene expression, and it is likely that histone H1 plays a key role in the repression, as its involved in stabilising the higher order chromatin structure (specifically the 30 nm fibre) in the higher eukaryotes (reviewed, Zlatanova, 1990). Genes that are not being expressed appear to be organised into the typical H1-containing nucleosomal arrays of the 30 nm fibre. Furthermore, selective removal of histone H1 from chromatin can cause specific genes to be transcriptionally activatable, while addition of the histone H1 can lead to a corresponding repression of *in vitro* transcription (Wolffe, 1989).

In contrast, the conformation of transcriptionally active chromatin appears to be considerably more extended, approaching the structure of the 10 nm beaded fibre. Consistent with this more open structure, recent studies, involving UV- or chemical-crosslinking of histones to DNA, have shown active chromatin to be partially depleted in histone H1. The intermittent absence of histone H1 from nucleosomes along the DNA perhaps being enough to break down the cooperative interactions that lead to the stable formation of the 30 nm fibre (Kamakaka and Thomas, 1990). In a similar crosslinking experiment histone H1 was found bound to active chromatin

but in a manner that did not allow it to generate higher order structure (Nacheva *et al.*, 1989). In this case histone H1 was no longer associated with active chromatin through its central globular domain but only through its basic carboxyl- and amino-terminal tails.

Chemical modification could be responsible for altering the binding characteristics of the histone H1. For instance phosphorylation of histone H1 has been associated with chromatin decondensation (review, Churchill and Travers, 1991). Linker histone variants (see 1.8) may also influence the stability of higher order chromatin structure, along with certain non-histone proteins, such as the high mobility group proteins HMG 14 and 17, which are found preferentially associated with active chromatin (review, Zlatanova, 1990).

(ii) Altered nucleosome structure

The actual core nucleosome structure of transcriptionally active DNA may differ from the nucleosomes of inactive DNA. It is thought that the histones on active DNA have a more open form called a 'half-nucleosome' (Ryoji and Worcel, 1985). In this form, the DNA is more loosely associated with the histones, and is thus more easily transcribed. These type of structures may be generated in a number of ways, including histone modification (e.g. ubiquination of histone H2A or acetylation of H3 and H4; both forms of modification reducing the positive charge of the histones and probably reducing their ability to interact with the DNA as tightly), a deficiency of histones H2A and H2B, the presence of core histone variants (see 1.8), or interactions with HMG proteins.

(iii) Nucleosome position

The position of the nucleosomes along the DNA may also influence the transcriptional state of a gene. Nucleosomes are frequently positioned at preferred, or fixed positions along specific DNA sequences (reviewed, Grunstein, 1990), and this positioning may affect the accessibility of the DNA to *trans*-acting factors.

(iv) DNA structure

The physical structure of the DNA may also influence the ability of the DNA to be transcribed (review, Grunstein, 1990). For example, altered structures such as Z-DNA, in which the DNA helix coils in a left-handed rather than the conventional right-handed manner, have been associated with transcriptional activity. Z-DNA regions appear to be incompatible with nucleosome formation, and as such may be involved in generating DNase I hypersensitive sites (see below).

(v) DNase I hypersensitive sites

In addition to the general DNase I sensitivity of active chromatin, actively transcribed genes may have DNase I hypersensitive sites. These sites often are present within the transcriptional control regions of a gene, and it is thought that they are due to a displacement of the nucleosomes by the binding of *trans*-acting factors to their DNA-recognition site and/or the formation of Z-DNA which may facilitate nucleosome displacement and subsequent entry of the relevant transcription factors. The combination of the binding of these factors, and the exclusion of nucleosomes from the DNA is thought to facilitate transcription initiation (review, Svaren and Chalkley, 1990).

(vi) Association with nuclear matrix

Another way in which gene transcription appears to be regulated is by the preferential association of active genes with the nuclear matrix (e.g. β-globin gene sequences in erythrocytes; Hentzen *et al.*, 1984). This association of active genes with the matrix takes the form of active genes being located in looped domains of chromatin (see 1.7.1) that are anchored at their bases to the nuclear matrix via specific sequences called scaffold attached regions. LCR regions (see 1.5.2 (iii)) and enhancers are also thought to be organised at the bases of these attached loops of chromatin, and as it is probable that RNA polymerase II is associated with the nuclear matrix (Jackson and Cook, 1985). It has been suggested that the LCR regions and enhancer elements when complexed with the appropriate *trans*-acting factors

provide a strong attraction for RNA polymerase II to initiate transcription. Additionally the points of attachment to the matrix are close to sites for topoisomerase II interaction with the DNA. Therefore it is possible the level of supercoiling and the consequent transcriptional activity of the looped domains could also be regulated from the attachment points. Another interesting point is that histone H1 has been found to be able to preferentially bind the short A/T-rich scaffold attached regions (review, Grunstein, 1990). Thus scaffold attached regions in conjunction with histone H1 may determine the transcriptional potential of the looped domains.

(vii) Chemical modification of DNA

Chemical modification of the DNA also may effect whether or not it is transcribed. For instance methylation of DNA has been linked with a repressed state of chromatin while in a number of cases demethylation has been associated with gene expression. One specific DNA target for methylation is the cytosine residue in the dinucleotide CpG. Under-methylation of such cytosine residues in the dinucleotide CpG, often repeated many times in regions called CpG islands in the 5' regulatory regions of a number of genes, correlates with active transcription of those genes (Bird, 1986). However methylation may just serve to mark the transcriptional state of chromatin and function to help the cell stably maintain and propagate its chromatin structure.

Critical to cells is this ability to be able to faithfully maintain and propagate their individual and specialised chromatin activity states. How this achieved is not well understood, but a number of mechanisms may be involved in this memory process including the methylation state of the DNA and for instance the selective deposition of transcription factors onto the DNA before nucleosome assembly, during DNA replication (reviewed, Svaren and Chalkley, 1990).

1.8 HISTONE PROTEINS

Histones are small, highly basic proteins, that are the principal structural proteins of chromatin (see Section 1.7 for description of chromatin structure). The histone protein family consists of five major classes of histone proteins. The core histones (H2A, H2B, H3 and H4) which are ubiquitous to all eukaryotic organisms and highly conserved throughout evolution, and the H1 or linker histones, which show appreciably more variation between different tissues and species (and which are apparently absent from yeast). Despite the conservation of histone protein sequences in evolution, non-allelic primary sequence variants or subtypes exist for all five histone types. These variants display distinct patterns of expression in a tissue specific manner, during differentiation and development, and throughout the cell cycle (Zweidler, 1984). Furthermore each histone variant or subtype is subject to post-translational modification, including: acetylation, phosphorylation, ADP-ribosylation, methylation and ubiquitination (reviewed, Wu *et al.*, 1986). These modifications appear to further increase the heterogenity and functional diversity of the histone protein family.

H4 appears to be the most highly conserved of the five classes of histones and generally has no variants, *Tetrahymena* being an exception, with H4 variants associated with the biologically distinct micro and macro nuclei of the organism (Bannon *et al.*, 1984). H3 is also highly conserved but a number of variants exsist. For example tissue specific and developmental H3 variants have been described in mouse (review, Wu *et al.*, 1986). H2B has diverged more during evolution than H3 or H4, while H2A shows the most divergence of the core histones. One notable H2A variant is the extreme variant, H2A.Z found in mammals (West and Bonner, 1980), which interestingly appears to be enriched in active chromatin. Related H2A variant proteins have been described in chicken (H2A.F; Harvey *et al.*, 1983) and *Drosophila* (van Daal *et al.*, 1988). These proteins are related to a *Tetrahymena* histone protein variant (hv1), found exclusively in the transcriptionally active macronucleus (Allis *et al.*, 1980; Allis *et al.*, 1986). Therefore, it is conceivable that these related variants may

play some fundemental role in chromosomal organisation of transcriptionally active DNA by altering the properties of nucleosome core particles.

The H1 or linker histones are the most heterogeneous of the five classes of histone proteins. Variants including, histone H5 (Neelin *et al.*, 1964), an avian linker histone variant, found only in nucleated erythrocytes, and histone H1° (Panyim and Chalkley, 1969), a mammalian linker histone variant, structurally similar to H5 (Smith *et al.*, 1980) found in differentiated non-dividing cells but not restricted to just one cell type (Gjerset *et al.*, 1982). Histone H5 gradually replace histone H1 during erythroid cell maturation (although fully mature erythrocytes still contain significant amounts of H1) (Weintraub, 1978). It is thought that H5 contributes to the marked chromatin condensation and transcriptional silencing observed in terminally differentiated erythroid cells. Other less variant histone H1 subtypes also exist. For instance vertebrate species generally have between four and six different histone H1 subtypes. For example six H1 subtypes have been characterised in chicken (Shannon and Wells, 1987) and five H1 subtypes, in addition to H1t, a testis specific H1 variant, in mouse (Lennox and Cohen, 1984). Although the functional significance of the H1 subtypes is not completely understood, it would appear that the different H1 subtypes are functionally distinct. The relative amounts of the different H1 subtypes are found to vary between different tissues in an organism, with evidence for nonrandom distribution of the subtypes between active and inactive chromatin (Lennox and Cohen, 1984) and differing abilities to to condense chromatin reported (Huang and Cole, 1984).

1.9 ORGANISATION OF HISTONE GENES

Histone gene copy number and organisation varies considerably between organisms (reviewed, Hentschel and Birnstiel, 1981; reviewed, Maxson *et al.*, 1983; reviewed, Old and Woodland, 1984; reviewed, Stein *et al.*, 1984). It has been shown that the genes for the five major classes of histones are usually clustered but their arrangement varies widely between different organisms. Organisation ranges from

highly regular tandemly repeated structures, including one copy of each gene per repeat of each of the five major histone types such as in *Drosophila* (reviewed, Lifton *et al.*, 1977), and in the case of the 'early' histone genes of sea urchin (reviewed, Kedes, 1979), to randomly arranged dispersed clusters of genes such as in chicken (Engel and Dodgson, 1981; Harvey *et al.*, 1981), mouse (Sittman *et al.*, 1981) and human (Heintz *et al.*, 1981). Where the histone gene copy number is high (up to several hundred copies), the genes are typically organised into the tandemly repeated structures. Reduced gene number correlates with a breakdown of such tandemly repeated structures. This is most evident in higher eukaryotic organisms such as chicken, mouse and human but also seen in lower eukaryotic organisms such as *Tetrahymena* (Bannon and Gorovsky, 1984) and yeast which has only two copies of each of the four core histones genes (reviewed, Maxson *et al.*, 1983). In organisms which have a low histone gene copy number the histone genes are not completely disorganised. In the chicken for example (probably all or most of the histone gene complement of the chicken has been isolated and characterised; D'Andrea *et al.*, 1985), genes for each of the five major histone classes are represented 6-10 times, and although no long range order is apparent there are some preferred associations, for example, pairs of H2A/H2B genes and H3/H4 genes which are often divergently transcribed.

Two key structural characteristics of most of the genes coding for the five major classes of histones (in the higher eukaryotes) are that they do not contain any introns and are typically transcribed into non-polyadenylated mRNAs, that instead terminate with a highly conserved stem-loop structure (see 1.10).

Genes encoding variant histones have also been characterised. In general, the genes are not linked to the other histone genes and are present in low copy number. In the chicken, four such genes have been isolated. One of these codes for the H5 linker histone (Krieg *et al.*, 1983), another codes for the histone variant, H2A.F (Harvey *et al.*, 1983) and the other two genes code for the H3 variant, H3.3 (Brush *et al.*, 1985).

These variant genes differ from other histone genes in that they are all transcribed to produce polyadenylated mRNAs, and the H3.3 and H2A.F gene sequences contain introns (Dalton *et al.*, 1986; Dalton *et al.*, 1989; Brush *et al.*, 1985).

The structural features of the genes encoding the different variant histones possibly reveal something of their origins. For example the fact that the H5 gene lacks introns and retains residual homology to the H1 genes (Coles and Wells, 1985) suggests that it may have evolved from a pre-existing H1 gene which became isolated from the rest of the histone cluster, perhaps by some unequal crossing-over event. However the structures of the core histone variant genes described above suggest a different origin. These genes, particularly H2A.F, exhibit a high degree of cross-species sequence conservation and are interupted by introns (Dalton *et al.*, 1989; van Daal *et al.*, 1988). This suggests an ancient evolutionary origin for these genes, where they may have originated from intron containing ancestral genes that subsequently diverged to give rise to both the precursors of the current core histone variants and after losing their introns and being duplicated, the genes coding for the five major classes of histone genes (review, Maxson *et al.*, 1983; Mannironi *et al.*, 1989).

1.10 REGULATION OF HISTONE GENE EXPRESSION

Histone genes have also been divided into three distinct classes based on the degree of linkage of their expression with DNA replication and the cell cycle (Zweidler, 1984). The replication-dependant histone genes represent the predominant class (their gene products constituting the most abundant histones of normal dividing cells) with their expression tightly linked to DNA replication, i.e. their expression is induced at the beginning of DNA replication (S-phase), and repressed at the end of S-phase.

A second class of histone genes, the replication-independent variants, also known as the replacement variants, are expressed throughout the cell cycle, i.e. their expression is completely uncoupled from DNA replication. The pattern of expression of this

group of genes may involve low level constitutive expression, or expression in a developmental or tissue-specific manner. Representative of this class is the histone H5 gene discussed above (see 1.8 and 1.9). The expression of these genes presumably represents a way to remodel chromatin structure and effect specific gene expression and subsequently developmental or tissue specific processes independently of DNA replication.

The existence of a third class of histone genes, consisting of partially replication-dependent genes, was originally based on the distinct pattern of expression, in regenerating liver, of two mouse core histone gene subtypes H2A.1 and H2B.2 (Zweidler, 1984). However other histone genes have now been described which have similar expression characteristics (see 1.10.3). The expression of the partially replication-dependent histone genes is characterised by a relaxed link with DNA synthesis where these genes are induced at the onset of DNA synthesis but are not completely repressed after cessation of DNA synthesis.

The regulation of expression of these three different classes of histone genes are further described in this section, while in the next section, transcriptional control of histone gene expression is discussed in more detail with particular emphasis on the transcriptional regulation of the histone H1 genes.

1.10.1 Replication-dependent histone gene expression

Expression of the replication-dependent histone genes involves multiple levels of control. These include transcriptional and post-transcriptional (pre-mRNA processing and mRNA stability) control mechanisms which combine to produce the observed 10 to 50-fold increase in the levels of these histone mRNAs during S-phase of the cell cycle. However the extent to which these different levels of control contribute to histone gene expression varies between the different histone genes in this class and between different organisms. Importantly though, the relative ratios of the five histone types seem to be maintained throughout the cell cycle as well as the

overall levels of histones as a function of DNA replication, indicating coordinate control of all five major histone types and a tight coupling of their expression to DNA replication (review, Osley, 1991). How this coordinate control and the coupling to DNA replication is achieved is not known at present.

At the beginning of S-phase, replication-dependent histone gene transcription is stimulated three to five-fold (Heintz *et al.*, 1983; DeLisle *et al.*, 1983; Alterman *et al.*, 1984) producing primary transcripts which include a highly conserved palindromic sequence followed by a conserved stretch of purine residues near the 3' termini (review, Schümperli, 1988; also see Figure 1.4). These regions of message (the palindromic sequence immediately preceding the cleavage site and the purine-rich sequence just downstream from the cleavage site) interact with U7 snRNA and other factors, to generate the mature histone mRNA 3' end, consisting of a characteristic stem-loop stucture (Georgiev and Birnstiel, 1985). The efficiency of this 3' processing step is enhanced in S-phase by virtue of an unmasking of the relevant 5' U7 RNA sequences, which are then available to pair with of the purine-rich element of the newly synthesized histone pre-mRNA (Hoffmann and Birnstiel, 1990). However, the 3' processing pathway is believed to contribute more to regulation of histone gene expression during transitions between states of cell proliferation and growth arrest rather than controlling histone mRNA abundance throughout the normal cell cycle (review, Schümperli, 1988).

A further regulatory step occurs at the level of the stability of the mature, non-polyadenylated messages. The half-life of histone mRNA in S-phase is increased approximately three-fold relative to their half-life in other stages of the cell cycle. Following S-phase, or in the presence of inhibitors of DNA synthesis, histone mRNA turnover increases by a mechanism which is linked to translation of the mRNA (Levine *et al.*, 1987). It has been demonstrated with chimeric mRNA that the histone-specific stem-loop structure is necessary and sufficient to confer this level of

regulation (Pandey and Marzluff, 1987). However the 5' leader of histone mRNAs has also been implicated to be important for transcript stability (Morris *et al.*, 1986).

Although decreased stability of histone mRNA appears to be a major effector of histone gene expression at the end of S-phase, the down regulation of histone gene transcription at the end of S-phase also contributes to the decline in histone mRNA for the remainder of the cell cycle.

Generally, all three levels of control mentioned above operate in higher eukaryotes, but interestingly, this is not the case in some lower eukaryotes (e.g. fungi and ciliates). For instance, in these lower eukaryotes, the mRNAs from the histone genes are polyadenylated and as a result are not subject to control by the same pre-mRNA processing and mRNA stability pathways. The genes are still cell cycle regulated but transcriptional control is more prominant in the overall control of the genes. Control of histone gene expression at the level of translation has also been reported to occur in the lower eukaryote, *Physarum*. However in general, translational control does not play a significant role in regulation of histone genes in other eukaryotes (review, Osley, 1991).

Another interesting level of control of histone gene expression is seen in some lower eukaryotes including the sea urchin where developmental switching between 'early' and 'late' sets of histone genes takes place (reviewed, Maxson *et al.*, 1983). The sea urchin 'early' histone genes as described above are organised in tandemly repeating units (400 copies/haploid genome) while 'late' genes are represented at 10-20 copies/haploid genome. The switch in expression from 'early' to 'late' histone gene expression corresponds to the change from rapid cell division (where there is a need for large amounts of histone proteins in a short period of time which can be generated from the large numbers of early genes) to the onset of cellular differentiation and morphogenesis. This differential expression of distinct sets of histone genes during development does not seem to occur in higher eukaryotes.

However, qualitative tissue specific/developmental variant histone gene expression takes place in both lower and higher eukaryotes (see 1.10.2 and 1.10.3).

What signals/controls are actually involved in coupling replication-dependent histone gene expression with DNA synthesis and the cell cycle? Replication-dependent histone gene expression appears to be temporally, but not necessarily functionally, coupled to DNA replication. For instance, although histone gene transcription is down-regulated and their mRNAs destabilized in the presence of some DNA synthesis inhibitors (notably inhibitors which interfere with deoxynucleotide metabolism, e.g. hydroxyurea) there are other DNA synthesis inhibitors (that do not affect deoxynucleotide metabolism, e.g. novobiocin) that have no detectable effect on either control of histone gene transcription or transcript stability (Graves and Marzluff, 1984). Therefore a functional link between deoxynucleotide levels and coupling of DNA replication to histone mRNA levels has been postulated (Marzluff and Graves, 1984). How this process may operate and also whether it is relevant to regulation of histone gene expression during the normal uninterrupted cell cycle is unknown.

A further observation is that protein synthesis inhibitors can counteract the destabilization of histone mRNAs (Stimac *et al.*, 1984; Sive *et al.*, 1984). This suggests that a short-lived protein may be involved in the destabilization process; perhaps a regulatory protein involved in coupling deoxynucleotide metabolism and histone mRNA stability or a protein directly involved in histone mRNA degradation (Marzluff and Graves, 1984). It has also been proposed that histone biosynthesis is subject to autoregulation whereby levels of histone proteins regulate their own synthesis, perhaps by controlling levels of their corresponding mRNAs (Butler and Mueller, 1973; Stein and Stein, 1984; Sariban *et al.*, 1985). In support of this model, Peltz and Ross (1987), using an *in vitro* system, observed specifically upon addition of histone proteins, increased histone mRNA degradation. It is possible that both a link with deoxynucleotide metabolism and feedback control by histone proteins

could operate to regulate histone mRNA metabolism (review, Marzluff and Pandey, 1988).

1.10.2 Replication-independent histone gene expression

The replication-independent histone genes, as mentioned above, are expressed constitutively throughout the cell cycle and are believed to have specialized functions. Consistent with their constitutive expression, mRNAs for replication-independent variants remain stable when DNA synthesis is inhibited by drug treatment (Sittman *et al.*, 1983). These genes encode polyadenylated mRNAs that lack the 3' terminal stem-loop structure, required for the post-transcriptional regulation of the replication-dependent histone genes. Additionally these genes may lack the appropriate 5' promoter sequences involved in conferring transcriptional control on the replication-dependent histone genes (e.g. the histone H5 gene lacks the H1 box, a promoter element believed to be involved in the cell cycle control of the histone H1 genes in chicken; Dalton *et al.*, 1986) (see 1.11). These differences between the two classes of histone genes would appear to be the responsible for the uncoupling of expression of the replication-independent histone genes and DNA replication in the cell cycle.

1.10.3 Partially replication-dependent histone gene expression

Expression of the partially replication-dependent histone genes is characterised by a relaxed link with DNA synthesis where these genes are induced at the onset of DNA synthesis but are not completely repressed after cessation of DNA synthesis. Histone genes that have been identified as belonging to this class include: mouse variant histone H1 (Cheng *et al.*, 1989) and H3 genes (Brown *et al.*, 1985), the mammalian variant histone H2A.X gene (Mannironi *et al.*, 1989), and two chicken histone H1 genes (Kirsh *et al.*, 1989). These genes have been reported to produce both relatively stable polyadenylated mRNA (characteristic of the constitutively expressed replication-independent class of genes) and short-lived non-polyadenylated mRNA (characteristic of the main class of replication-dependent histone genes). Unlike the

replication-independent genes these genes contain the conserved 3' palindrome and purine box elements involved in producing the replication-dependent/poly(A)$^-$ transcripts (these sequences are found in their mature polyadenylated mRNAs). The two types of messages produced by these genes are therefore the result of alternative processing events, and by virtue of their different 3' termini, their abundance in the cell can be regulated independently.

It would appear that the ability to produce relatively stable polyadenylated mRNA, able to escape the replication-dependent histone gene specific degradative process, allows the partially replication-dependent histone genes to continue expressing their histone products when DNA synthesis ceases. This may allow these genes to increase the ratio of their histone products to that of other histone subtypes of the same class in nondividing cells/differentiated tissue. An example of this is seen in the case of the partially replication-dependent chicken histone H1 genes, H1.01 and H1.10. These genes, which have been suggested to possess increased capacities to condense chromatin (Coles *et al.*, 1987), encode the histone protein subtypes H1c' and H1c respectively and have been found to be particularly prevalent in nondividing differentiated tissues (H1c' and/or H1c may serve a similar function to mammalian histone H1°; Winter *et al.*, 1985). However, both the H1.01 and H1.10 genes have 5' promoter and 3' processing sequence elements characteristic of replication-dependent histone genes (Coles *et al.*, 1987). Furthermore in the case of the H1.01 gene it has been shown to be cell cycle regulated in dividing cells (Dalton and Wells, 1988*a*; see 1.11.2). Consistent with the above findings, Kirsh *et al.* (1989) have shown that the H1.01 gene is expressed predominantly as a polyadenylated mRNA in tissues with few dividing cells whereas the replication-dependent/poly(A)$^-$ mRNA species is the major transcript generated from the H1.01 gene in tissues containing large numbers of dividing cells.

The signals and mechanisms involved in bringing about the differential expression of this class of histone genes are unknown. Although, both Kirsh *et al.* (1989) and

Cheng *et al.* (1989) suggest that since the alternative RNA processing pathways (involved in the production of the poly(A)$^-$ and poly(A)$^+$ mRNAs) may be in competition a reduction in the efficiency of 3' stem-loop formation in nondividing cells (perhaps via a cell specific differentiation signal) would lead to a decrease in production of the poly(A)$^-$ mRNA and an increase in the poly(A)$^+$ mRNA.

1.11 TRANSCRIPTIONAL REGULATION OF HISTONE GENES

As mentioned above, a significant level of control of histone gene expression occurs at the level of transcription. *Cis*-acting regulatory elements (and *trans*-acting regulatory factors that interact with these elements) have been identified that are critical to the transcriptional regulation of histone genes during the cell cycle. These elements (and their cognate transcription factors) along with more general promoter elements also present in the promoters of the histone genes are described below. A more in depth discussion of the transcriptional control of histone H1 genes follows.

1.11.1 Histone gene promoter elements and transcription factors

(i) General promoter elements

The promoter architecture of histone genes is similar to promoters of other RNA polymerase II transcribed genes, being modular in nature and containing a number of discrete independently functioning elements that contribute to transcriptional activity (Wells and McBride, 1989; review, Osley, 1991). Typically histone gene promoters have, approximately 20 - 30 bp upstream from the transcription initiation site, a conserved TATA box element, which is characteristic to most RNA polymerase II transcribed genes (see 1.5.1 (i)). Between 40 - 50 bp further upstream, many histone genes also have a CCAAT box element (see 1.5.2 (i)). Additionally some histone genes, usually within 100 bp of the TATA box, possess a G/C rich element(s) (see 1.5.2 (i)). However, as the majority of these elements are common to a wide range of genes transcribed by RNA polymerase II (see 1.5.2 (i)), it is likely they play a general role in histone gene transcription and are not involved

specifically in selective promoter recognition associated with developmental, tissue specific or coordinated expression of histone genes during the cell cycle (although in the case of a CCAAT box-containing element in the histone H1 genes this may not be the case; see 1.11.2).

(ii) Specific promoter elements

Histone gene-specific regulatory elements have been identified from the extensive sequence analysis and comparisons of the promoter regions of the different histone genes. The functional significance of some of these specific elements has been demonstrated by the use of *in vitro* and *in vivo* gene expression systems, usually in combination with deletion and/or mutagenesis analysis of the sequences. Additionally, *in vitro* and *in vivo* techniques for detecting specific protein-DNA interactions have been applied to identify factors that interact with some of these specific regulatory elements.

Most of the histone gene-specific regulatory elements identified so far have been located within a couple of hundred base pairs of the site for transcription initiation. An element, in *S. cerevisiae*, has been found associated with each of the 4 histone gene loci (no histone H1 or equivalent has been identified in yeast; see 1.8). This element has been shown to be involved in cell cycle expression of the histone genes and represents a strong candidate for coordinating expression of these genes (Osley *et al.*, 1986). A second cell cycle regulatory element, found at all but one of the 4 histone gene loci, appears to act as a negative regulator of *S. cerevisiae* histone gene expression. In higher eukaryotes, no histone gene-specific elements involved in cell cycle regulation have been found that are common to all the members of the histone gene family. A hexameric sequence (GACTTC) has been identified that appears to be present in the promoters of all the histone genes. However, this element seems to be involved solely in helping to generate maximal levels of histone gene transcription (review, Osley, 1991). Instead in higher eukaryotes, each of the five

classes of histone genes seem to possess different class-specific element(s) critical for their transcriptional control during the cell cycle (see Table 1.2).

Distinct cell cycle regulatory elements, and corresponding factors that interact with these elements, have been identified for the histone H1 (see 1.11.2) and H2B genes (Harvey *et al.*, 1982; Sive *et al.*, 1986; Fletcher *et al.*, 1987; La Bella *et al.*, 1988). In the case of the H2B genes a specific element the, H2B box, has been identified as being essential for cell cycle regulation. The H2B box element contains an octamer sequence (see Table 1.2) found in the promoters of a number of other unrelated genes (see Table 1.1), and, like these other genes, has been shown to bind the Oct1 transcription factor (Sive and Roeder, 1986; Fletcher *et al.*, 1987). The mechanism of how the H2B box/Oct1 complex specifically effects the cell cycle regulation of the H2B genes is not known, although specific post-translational modification of Oct1 or the existence of additional regulatory protein(s) that form a complex with Oct1 when bound to the H2B box may be involved (reviewed, Heintz, 1991; reviewed, Osley, 1991)

Histone H4 gene specific regulatory elements (and interacting transcription factors) have also been found (Hanly *et al.*, 1985; Dailey *et al.*, 1986; Dailey *et al.*, 1988; van Wijnen *et al.*, 1989). However, although these elements have not yet directly been shown to be necessary for cell cycle regulation of the H4 genes, their similar positioning in the H4 gene promoters, compared to the positioning of the H1 and H2B gene specific elements in their respective promoters, suggests the H4 specific elements, paticularly the proximal element bound by H4TF2, may indeed have cell cycle regulatory roles (see below). Little is known of the promoter elements and corresponding factors that regulate transcription of histone H2A and H3 genes in the cell cycle, although a 32 bp region in the promoter of an H3 gene in hamster has been found to be required for accurate S-phase regulation of this H3 gene (Artishevsky *et al.*, 1987). Further studies by Sharma *et al.* (1989) have resulted in the identification of a close match to the AP1 consensus sequence within this 32 bp region, and the

Table 1.2 Histone gene class-specific promoter elements

Presented in the table are various histone gene class-specific elements and the transcription factors that bind to them (figure adapted from review, Heintz, 1991). The H1 box, H1-CCAAT box, and H2B box have been shown to be involved in cell cycle regulation of their respective genes (see 1.11.1 (ii) and 1.11.2). The H4 specific proximal promoter element (bound by H4TF2), yet to be demonstrated to be involved in cell cycle regulation, represents a strong candidate for carrying out such a fuction (see 1.11.1 (ii)). The H3 sequence tabulated, represents a 32 bp region shown to be involved in cell regulation of the hamster H3.2 gene. However, as yet, a more defined H3 class specific element along the lines of those identified in the H1, H2B and H4 genes has not been identified. Likewise, in the case of the H2A genes no class specific cell cycle regulatory element(s) have been identified.

CCAAT, octamer and AP1 consensus elements (found in other unrelated genes; see Table 1.1), contained within the H1-CCAAT box, H2B box and the H3 32 bp sequence respectively, have been underlined. Note: N, represents any nucleotide.

References: (1) review, Heintz, 199; (2) Coles *et al.*, 1987; (3) van Wijnen *et al.*, 1988; (4) Harvey *et al.*, 1982; (5) La Bella *et al.*, 1988; (6) review, Osley, 1991; (7) Sharma *et al.*, 1989; (8) Dailey *et al.*, 1988

Histone gene class	Histone gene class-specific element	*Trans*-acting factor	Species included (Ref.)
H1	H1-CCAAT box: GCACCAATCAC$^{A}/_{C}$GCGCG$^{G}/_{C}$	H1TF2 HiNFB	Human, chicken (1, 2, 3)
	H1 box: AA$^{G}/_{C}$AAACACA	H1TF1 H1-SF	
H2A	Unknown		
H2B	H2B box: $^{C}/_{C}$NATTTGCATA$^{A}/_{C}$	Oct1	Human, chicken (1, 4, 5)
H3	TGGCGAGTCAGCCAGCCGCGGGGCTGGACAAG	Jun-related protein(s)	Hamster (6, 7)
H4	$^{T}/_{C}{}^{C}/_{T}$TTCAGGTT$^{C}/_{T}$TCA$^{G}/_{A}{}^{A}/_{T}$NNGGTCCG	H4TF2	Human (1, 6, 8)
	CCCCTCCCCC	H4TF1	

preliminary characterisation of two Jun-like proteins that bind this site/region. In higher eukaryotes, many H2A and H2B genes, and also H3 and H4 genes, are closely linked, often forming divergent transcription units. Interestingly, there is some evidence that elements identified as being involved with the regulation of the H2B and H4 genes may also influence H2A and H3 gene regulation respectively (Sturm *et al.*, 1988*b*; van Wijnen *et al.*, 1991). The lack of readily identifiable specific regulatory elements for H2A and H3 genes may imply that more gene copy specific elements are involved in the transcriptional regulation of these genes which await characterisation (review, Schümperli, 1988). The promoter sequences of the H2A and H3 genes have not been as extensively characterised as those for the H1, H2B, and H4 genes.

The histone class-specific elements, that have been identified, are highly conserved in sequence and position within a specific class of histone genes, and are often also conserved between genes of the same class from different species. Additionally, and consistent with a general histone gene promoter structure, the different histone class-specific elements are located similar distances upstream in their respective promoters from the transcription start site, with their positions relative to the other general promoter elements also conserved. For instance, the proximal promoter specific elements in the histone H1, H2B and H4 genes are similarly positioned with respect to the TATA boxes in each of their promoters, to suggest the presence of important functional interactions between the regulatory factors involved (H1TF2, Oct1 and H4TF2 factors respectively; see Table 1.2) and the TATA box binding factor, TFIID (review, Heintz, 1991).

Clearly, the sequences that regulate transcription have diverged among the five classes of histone genes in higher eukaryotes, while in yeast, a more simply co-ordinated regulatory alternative prevails. The reason, for the evolution of this increased regulatory complexity in higher eukaryotes, may stem from the proportionally increased demands on the regulatory system to ensure appropriate

histone gene expression in the large variety of cell types in the biologically more complex higher eukaryotic organisms. Additionally, Heintz (1991) suggests that the ability to independently regulate synthesis of each of the different histone types may accomodate the apparent evolutionary instability of histone gene copy numbers within the histone gene family.

The maintainance of balanced histone synthesis appears to be critical to cell viability (Meeks-Wagner and Hartwell, 1986; review, Osley; 1991). Thus even though the five major classes of histone genes, in higher eukaryotes, regulate their transcription during the cell cycle via different elements, they must also, perhaps like the more directly coordinated yeast histone gene complement, be coordinately controlled so as to ensure balanced histone synthesis. Additional upstream regulatory signals must be operating to coordinate histone gene transcription and link histone gene expression with the cell cycle.

1.11.2 Regulation of H1 histone gene transcription

Several distinct regulatory elements have been found in the promoters of the histone H1 genes examined to date. Apart from canonical TATA box and G/C rich elements, two different histone H1 specific regulatory elements have been identified and shown to be involved in the cell cycle regulation of H1 gene transcription (see Table 1.2 and below). These are a proximal promoter specific element, the H1-CCAAT box, which has a core CCAAT motif but involves a more extended consensus sequence (Gallinari *et al.*, 1989), and a distal promoter specific element, the H1 box, a highly conserved element that bears no sequence similarity to other known regulatory elements (Coles and Wells, 1985). The H1 box element has been found in the promoters of all histone H1 genes analysed from higher eukaryotes (including the full complement of 6 chicken histone H1 genes). The H1-CCAAT box element also seems to be highly conserved in the histone H1 genes of higher eukaryotes. The position of these two elements in the promoters of a number of histone H1 genes,

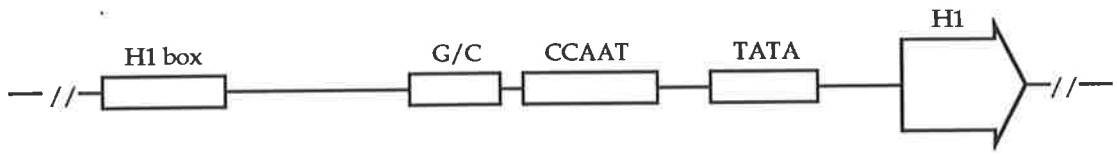
along with the positions of the TATA box and G/C rich elements, are displayed in Figure 1.3.

The importance of each of the H1 box, G/C rich and H1-CCAAT box elements to maximal H1 promoter activity have been demonstrated both *in vivo* and in an *in vitro* transcription assay with a series of 5' deletions of a human histone H1 promoter (Gallinari *et al.*, 1989). Deletion of the H1 box resulted in an approximate 50% decrease in transcription, the additional loss of the G/C rich Sp1 binding site resulted in a further decrease in transcription to 15% of wild type promoter activity while removal of the H1-CCAAT box reduced transcription to an undetectable level. A point mutated H1-CCAAT box alone resulted in an approximate 50% decrease in promoter activity.

Cell cycle regulatory roles for the H1 box and for the H1-CCAAT box have also been demonstrated using similar experimental approaches to those described above. Dalton and Wells (1988*a*), examining transcription from H1 constructs transfected into HeLa cells, reported that deletion or base substitution of the H1 box in the promoter of the chicken histone H1.01 gene (see Figure 1.4) significantly reduced H1 mRNA levels in randomly growing cells and eliminated cell cycle control of H1 transcription in aphidicolin synchronized cells. While La Bella *et al.* (1989) using extracts generated from centrifugally elutriated S-phase cells and an *in vitro* transcription system to assay the extract's effect on transcription from appropriately mutated human histone H1 promoter-containing constructs, obtained similar results to Dalton and Wells (1988*a*), demonstrating a role for the H1 box in cell cycle control of H1 transcription. Additionally, La Bella *et al.* (1989) were able to show that the H1-CCAAT box element also exhibited a major cell cycle regulatory effect on H1 transcription.

Figure 1.3 Histone H1 gene promoter structure

The promoter architecture of the histone H1 genes from chicken and a number of other organisms are shown (figure adapted from Coles *et al.*, 1987; Dalton and Wells, 1988*a*). The conserved elements from the promoters of the six chicken H1 genes (Coles *et al.*, 1987), human Hh9 H1 gene (Gallinari *et al.*, 1989), mouse H1 gene (Yang *et al.*, 1987), Rat H1t gene (Cole *et al.*, 1986), Xenopus XLHW8 H1 gene (Turner *et al.*, 1983) and trout H1 gene (Mezquita *et al.*, 1985) have been aligned for comparison. Distances between these elements are in base pairs and numbering of the elements is with respect to the respective cap sites. Gaps (-) were introduced into the chicken H1.03 gene for alignment.



		H1 box	G/C	CCAAT	TATA
		-118	-74	-55	-32
Chicken	H1.01	AAGAAACACAAA...32...GCGGGGCGGGCT..7..GCACCAATCA..15..CTAAAAATA			
	H1.02	-120 AAGAAACACAAA...32...GCGGGGCGGGCT..7..GCACCAATCA..15..CTATAAATA			
	H1.10	-118 AAGAAACACAGA...32...GTTAGGCGGGCT..7..GCACCAATCA..15..CTATAAATT			
	H1 11L	-105 AAGAAACACAAG...27...GCGGGGAGGGCT..7..GCACCAATCA..15..CTATAAATA			
	H1 11R	-113 AAGAAACACAAG...31...GCGGGGCGGGGC..5..GCACCAATCA..14..CTATAAATG			
	H1.03	-89 AAGAAACACAAC...17...GCGGGGCGA----------ACCAATCA..15..CTATAAAGG			
Human	Hh9 H1	-100 AACAAACACAAA...11...CAAGGGCGGGGC.13..GCACCAATCA..14..CTATATAAG			
Mouse	H1	-112 AACAAACACAAG.....45.....GCACCAATCA..16..CTATATATA			
Rat	H1t	-102 AAGAAACACAAA...11...CAGGGGCGGGGG.14..GCACCAATCA..14..CTATATAAG			
Xenopus	XLHW8 H1	-153 TAAAAACACAGA...36...AATGGGCGGGGT..8..CAACCAATGA..45..ATATAAGGA			
Trout	H1	-113 AGAAAACACAAG...33...GCTGGGTGGCTT.....24.....GTTTAAGGC			

Figure 1.4 Chicken histone H1.01 gene sequence

The DNA sequence of the chicken histone H1.01 gene is shown (figure adapted from Coles and Wells, 1985). The protein coding region is presented as triplets and includes initiator and terminator codons with the predicted amino acid sequence shown above the DNA sequence. In the non-coding regions, 5' and 3' conserved elements are underlined. The conserved palindromic and purine-rich 3' mRNA processing elements are identified by the bracketed numbers 1 and 2 respectively. Bases representing the start and end of the H1.01 mRNA are marked with +1 and the symbol, #, respectively.

-600
TCGGGCATTAACGAATTGTTTACAGC

TCTATAATAAGTGCAATAGGAGGTGGTATGAGACACTATTTCAATAGAGTTTAGCGTTTTAGGAGGTAAAACACATGGC

-500
CAAAGGAACCGGTTCTTAATCCATATTCAGAGCTGAGAGGGGAGTTTGTGTTACTTCCTCCTCTTCAAAATCAATTTAA

-400
ACTGTCAAAATAGCTTTAAATCGTCAGATTTCGATTTTAGTCGCAAGAAAACCCTCTAAAGAACAACATACAGTGGTCG

-300
CTGTAACATTTCTGCCGAAAGAAAATTTTTAATTTGCAATGAAAAACAAAAGTTTCTCTCTTCAGAGGTCTCTCCAGCA

GCAGCACACGGGATTTATCGCCTCTCCTTTAACTCAGGACGCGTGTCTGCGGCTGGAAACTCTCCCGAACGCAAGTACC

-200
TGCTCTTCTCTTCCCTCACCGAGGAGACGGGGCGATTTGGTGGCAGAAATTCCGAGGAAAATACACTTTTGTTAGTCCA

-100
AAGAAACACAAATCGAGCACACCGAAGGGCTCCCCGGCCGTGCAGCGGGGCGGGCTTAGCAACGCACCAATCACCGCGC

+1
GGCTCCTCTCTAAAAATACGAGCATCTGACCCGCGCCAGCCCAATTGTGTTCGCCTGCTCCGCAGAGGACTGCGCCGCG

met ser glu thr ala pro ala ala ala pro asp ala pro ala pro gly ala lys ala ala
ATG TCC GAG ACC GCT CCC GCC GCC GCC CCC GAT GCG CCC GCG CCC GGC GCC AAG GCC GCC

100
ala lys lys pro lys lys ala ala gly gly ala lys ala arg lys pro ala gly pro ser
GCC AAG AAG CCG AAG AAG GCG GCG GGC GGC GCC AAA GCC CGC AAG CCC GCG GGC CCC AGC

200
val thr glu leu ile thr lys ala val ser ala ser lys glu arg lys gly leu ser leu
GTC ACC GAG CTG ATC ACC AAG GCC GTG TCC GCC TCC AAG GAG CGC AAG GGG CTC TCC CTC

ala ala leu lys lys ala leu ala ala gly gly tyr asp val glu lys asn asn ser arg
GCC GCG CTC AAG AAG GCG CTG GCC GCC GGC GGC TAC GAC GTG GAG AAG AAC AAC AGC CGC

300
ile lys leu gly leu lys ser leu val ser lys gly thr leu val gln thr lys gly thr
ATC AAG CTG GGG CTC AAG AGC CTC GTC AGC AAG GGC ACC CTG GTG CAG ACC AAG GGC ACC

gly ala ser gly ser phe arg leu asn lys lys pro gly glu val lys glu lys ala pro
GGC GCC TCG GGC TCC TTT CGG CTC AAC AAG AAG CCG GGT GAG GTG AAG GAG AAG GCT CCG

400
arg lys arg ala thr ala ala lys pro lys lys pro ala ala lys lys pro ala ala ala
AGG AAG CGA GCG ACT GCT GCC AAG CCC AAG AAG CCG GCG GCC AAG AAG CCT GCG GCT GCT

500
ala lys lys pro lys lys ala ala ala val lys lys ser pro lys lys ala lys lys pro
GCC AAG AAG CCC AAG AAG GCG GCG GCG GTG AAG AAG AGC CCC AAG AAA GCC AAG AAG CCG

ala ala ala ala thr lys lys ala ala lys ser pro lys lys ala ala lys ala gly arg
GCG GCT GCC GCC ACC AAG AAG GCG GCC AAG AGC CCC AAG AAG GCC GCC AAG GCT GGC CGC

600
pro lys lys ala ala lys ser pro ala lys ala lys ala val lys pro lys ala ala lys
CCC AAG AAG GCC GCC AAG AGC CCG GCC AAG GCA AAG GCG GTG AAG CCC AAG GCT GCC AAG

*pro lys ala thr lys pro lys ala ala lys ala lys lys thr ala ala lys lys lys ****
CCC AAG GCG ACC AAA CCC AAG GCG GCC AAG GCG AAG AAG ACG GCA GCC AAG AAG AAG TAA

700 #
GTTATCCCAGAAGAGTCCTGCTCTACCTATTTTGATATCCAACGGCTCTTTTAAGAGCCACCCACACTTTCCCTAAAGG

800 (1) (2)
AGCTGAGGCACCGAGGTCGTCAGAAACTTCCAGCACGGAGGCAGCAATTCGTAAGTCGTCAGAGGTCAATTGCCTTTTC

900
CCCTCCGATTACCGAAACCTAACGAGCACGGTTGAACGCGGCGGCTTTAGGGAAGTGTAGACTTTGTATCTTTTGCCGA

1000
GTAATTGGTTTGACTACCGTGAAGAAACGTTTTGTAATGATTTGATAAAAATCGGGTGACACTTTTTTTAAGAATATAT

TTTGTAACAGAAGTAATGGATTTCCCAGGCGCAAGCTACTACTGAGCCATGTCTAACGTGTTGTGTTGTTCCTCTTTAA

1100
GGTGTCTCCTTAAATGCTTTTGTGTATTAGGGGAAGACGGGAGATTTTTCTTACTGACGCGGTAACAGCCCCGAGCTCT

1200
CCCATCTCTTTTGTTCCCGCTGAGACAGAACAGCGGCTTCTGCTGTTGGAAAAGCCCGCCCTGGCCGAGGATTGGCCAC

1300
GAGGAGCCCGGCCCGCTGCCCCTTCCCCCTCCCACCGCAGTCCCCGCCTTGGGCCCGGCGCTTTTGGCCGCGTTGAAGA

1400
AAGGAACAGGCGTGGGGGAAAGGAGGGGGGAGGGGGGCGGGCGGTGACGGGACCTCCCGGAACGCATTGGTTTCCTTTC

GTACGATATACGAACTGAAAGTGTAACGGCGCGTCCCGGGAGAAACTTCTTTTGGGAGAACGCTTTGGGCACGACTTTG

1500
TTAACGGAAGCATGGAAAGCGTTGCTATTATTACCCACCAAATAATACTGATAATAAATATGAGAAAAAAAAAAGAAAA

AAAAAAAGCACGGC

The functional significance of the H1 and H1-CCAAT boxes to histone H1 transcriptional regulation is further supported by the identification of potential regulatory proteins that specifically interact with each of these elements.

A 47 kDa factor, H1TF2, that specifically interacts with the H1-CCAAT box element has been extensively purified. Considering its distinct molecular size and binding characteristics compared to other previously identified CCAAT box binding factors, H1TF2 would appear to be a histone H1 specific CCAAT-binding activity. Furthermore, consistent with a role for this factor in histone H1 cell cycle expression, H1TF2 DNA binding activity has been shown to increase during S-phase (La Bella *et al.*, 1989). However van Wijnen *et al* (1988) have identified a multiprotein complex, HiNFB, which also binds the H1-CCAAT box. It is unclear if HiNFB and H1TF2 are related. However, H1TF2 is unable to activate transcription in an *in vitro* transcription assay. This suggests that H1TF2 requires additional protein component(s) to be fully functional, and may indeed form part of a multiprotein complex. Perhaps, H1TF2 represents the DNA binding portion of the HiNFB complex.

Two DNA binding proteins, H1-SF in chicken (Dalton and Wells, 1988*a*) and H1TF1 in mammals (Gallinari *et al.*, 1989), both approximately 90 kDa, have been identified as specifically binding the H1 box element. These factors have not been characterised as well as the H1TF2 factor, and there is some discrepancy in the DNA binding activities of these two seemingly homologous factors. For example, while H1-SF DNA binding activity has been reported to increase approximately 12-fold in S-phase (Dalton and Wells 1988*b*), H1TF1 DNA binding activity was found to remain constant during the cell cycle (La Bella *et al.*, 1989). However, on further analysis, H1TF1 was found not to bind all of the histone H1 boxes tested, suggesting that H1TF1 and H1-SF are different protein factors (unpublished work by N. Segil and N. Heintz; cited Heintz, 1991).

1.12 THESIS AIMS

The primary aim of work presented in this thesis was to isolate a chicken *trans*-acting factor (H1-SF), reported to bind a histone H1 specific promoter element called the H1 box. Previous work had indicated that the interaction between H1-SF and the H1 box modulated cell cycle control of histone H1 gene transcription (Dalton and Wells, 1988*a*).

Following the isolation of H1-SF it was hoped to begin studies of its biological characterisation, to allow more precise definition of the role of this factor in regulation of histone H1 gene expression. It was also hoped to isolate the H1-SF gene with the aim of being able to examine the regulation of H1-SF expression, and hopefully gain, insight into further upstream regulatory mechanisms involved with coordinating histone gene expression.

CHAPTER 2

DETECTION OF HI-SF CANDIDATES: H1-F40 AND H1-F14

2.1 INTRODUCTION

Sequence analysis of the full complement of 6 chicken histone H1 genes, previously isolated and sequenced in this laboratory (Coles *et al.*, 1987), revealed a number of conserved promoter elements (see Figure 1.3). Of particular interest was a conserved sequence (AAGAAACACA), called the H1 box, located at approximately -100 with respect to the site of transcription initiation. Further studies conducted in this laboratory indicated that the H1 box was important in controlling H1 gene expression during the cell cycle (Dalton and Wells, 1988*a*). Furthermore, double stranded oligonucleotides containing the H1 box sequence were used with the gel retardation assay technique to detect a DNA binding activity (H1-SF) that interacted specifically with the H1 box (Dalton and Wells, 1988*a*).

Work described in this chapter is concerned with the further investigation of this H1-SF/H1 box interaction, using the gel retardation assay technique. Once a specific H1-SF/H1 box interaction could be demonstrated, attempts would be made to purify H1-SF (dealt with in Chapters 3 and 4).

2.2 STUDY OF H1-SF/H1 BOX INTERACTION WITH 40 MER AND Δ40 MER

The experiments described below were designed to confirm the specific H1-SF/H1 box interaction initially identified by Dalton (1987). In order to detect this specific DNA-protein interaction the gel retardation assay technique was employed. The gel retardation or gel shift assay (Fried and Crothers, 1981; Garner and Revzin, 1981) works on the basis that stable DNA-protein complexes migrate more slowly than free DNA when electrophoresed on non-denaturing gels. The assay is simple, quick and sensitive. Protein extract is incubated with radiolabelled binding site DNA. After the binding reaction is complete the sample is separated on a non-denaturing polyacrylamide gel and visualised by autoradiography. The gel retardation assay can be used both qualitatively, to identify and characterise a DNA binding protein, and/or quantitatively, to help measure binding affinity and relative concentration of

a protein. The gel retardation assay conditions used in the bulk of this work are described in 7.3.12 (i) unless otherwise specified.

2.2.1 Preparation of 40 mer and Δ40 mer oligonucleotide probes

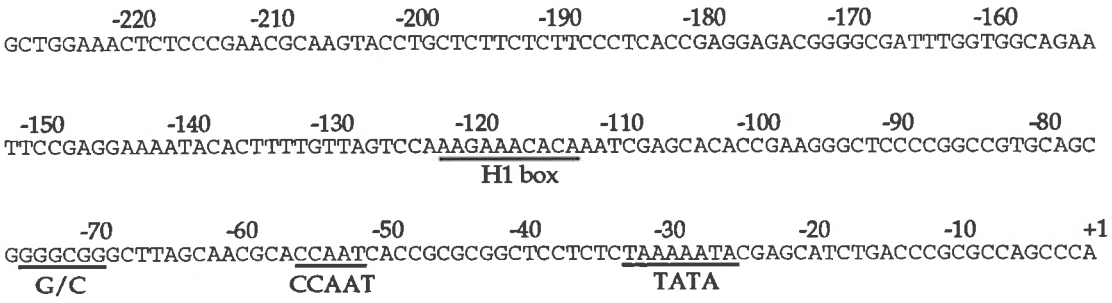
The studies conducted by Dalton and Wells (1988*a*), demonstrating the functional importance of the H1 box were all carried out using the chicken histone H1.01 gene (Figure 1.4) and H1 box-containing oligonucleotide probes of corresponding H1.01 sequence to detect H1-SF activity. Thus the same two double stranded oligonucleotide probes, one containing an intact H1 box (40 mer) and the other a disrupted version of the H1 box (Δ40 mer), were used in this study. The sequences of the 40 mer and Δ40 mer probes are shown in Figure 2.1 along with the sequence of the proximal promoter region of the histone H1.01 gene. A double stranded oligonucleotide probe which contains a histone H5 specific enhancer sequence, was also used in the study as a control probe for the gel retardation assay (Sparrow, 1991; see 7.2.8 for sequences of the oligonucleotides used to form the double stranded H5 control probe). Gel retardation probes were prepared as described in 7.3.12 (i).

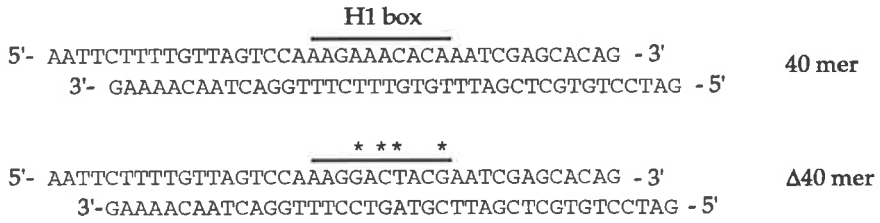
2.2.2 Preparation of nuclear extracts

H1-SF activity had previously been detected in crude nuclear extract prepared from LSCC HD2 tissue culture cells (Dalton, 1987), a chicken erythroid cell line transformed with the avian erythroblastosis virus (Beug *et al.*, 1982). Thus LSCC HD2 nuclear extracts were also used in this study, but in addition nuclear extracts were prepared from nine day chicken embryos, as they represented an inexpensive and readily available source of dividing cells likely to contain the required factor. Several different methods for preparing nuclear extract from nine day chicken embryos were evaluated. Ultimately the method of choice was a hybrid procedure derived from methods described by Panyim *et al.* (1971) and Lee *et al.* (1988). Extracts from both LSCC HD2 cells and nine day chicken embryos were prepared as described in 7.3.7 (i) and (ii).

Figure 2.1 40 mer and Δ40 mer gel retardation probes

The proximal promoter region of the chicken histone H1.01 gene is shown (see Figure 1.4 for further chicken histone H1.01 5' noncoding sequence). Conserved promoter elements are highlighted and numbering is with respect to the cap site. The sequences of both the 40 mer and Δ40 mer gel retardation probes are also shown. Base changes made to the Δ40 mer probe disrupting the H1 box are designated with asterisks. Additionally, both 40 mer and Δ40 mer contain *Eco*RI and *Bam*HI overhangs.


-220 -210 -200 -190 -180 -170 -160
GCTGGAAACTCTCCCGAACGCAAGTACCTGCTCTTCTCTTCCCTCACCGAGGAGACGGGGCGATTTGGTGGCAGAA
-150 -140 -130 -120 -110 -100 -90 -80
TTCCGAGGAAAATACACTTTTGTTAGTCCAAAGAAACACAAATCGAGCACACCGAAGGGCTCCCCGGCCGTGCAGC
H1 box
-70 -60 -50 -40 -30 -20 -10 +1
GGGGCGGGCTTAGCAACGCACCAATCACCGCGCGGCTCCTCTCTAAAAATACGAGCATCTGACCCGCGCCAGCCCA
G/C
CCAAT
TATA


H1 box
5'- AATTCTTTTGTTAGTCCAAAGAAACACAAATCGAGCACAG - 3'
3'- GAAAACAATCAGGTTTCTTTGTGTTTAGCTCGTGTCCTAG - 5'
40 mer
* ** *
5'- AATTCTTTTGTTAGTCCAAAGGACTACGAATCGAGCACAG - 3'
3'-GAAAACAATCAGGTTTCCTGATGCTTAGCTCGTGTCCTAG - 5'
Δ40 mer

2.2.3 Detection of control H5 enhancer binding protein

As a positive control, labelled H5 probe was incubated with LSCC HD2 and nine day chicken embryo (9DCE) crude nuclear extracts with, or without, unlabelled heterologous competitor DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by non-denaturing polyacrylamide gel electrophoresis (7.3.12 (i)). The assay detected two retarded DNA-protein complexes, resulting from the binding of proteins present in the crude extract to the labelled probe (see Figure 2.2A). The specificity of the protein-DNA complexes formed was tested by incubation of labelled DNA with crude extract in the presence of unlabelled heterologous competitor DNA. As had previously been shown (Sparrow, 1991), addition of an excess of unlabelled heterologous DNA (poly(dI-dC)) competed out the upper band, representing non-specific protein interaction(s) with the probe, but did not compete out the lower band which represented the specific H5 enhancer binding protein complex. Both extracts examined contained H5 enhancer binding protein activity, indicating that procedures to prepare nuclear extracts were satisfactory.

2.2.4 Search for H1-SF candidate with 40 mer

(i) LSCC HD2 extract

Labelled 40 mer probe or Δ40 mer probe was incubated with LSCC HD2 crude nuclear extract, and the reactions analysed by gel retardation assay (7.3.12 (i)). The assay detected several retarded DNA-protein complexes, resulting from the binding of proteins present in the nuclear extract to the labelled 40 mer probe (see Figure 2.2B). The specificity of DNA-protein complexes formed was tested by incubation of labelled DNA with nuclear extract in the presence of unlabelled heterologous or homologous competitor DNA. Addition of an excess of unlabelled heterologous DNA competed out all except one band which was thought to represent H1-SF. However the same band was also seen with the Δ40 mer probe, suggesting that the binding activity identified in the LSCC HD2 nuclear extract (designated H1-F40) was not interacting specifically with the H1 box and was not the H1-SF activity identified by Dalton (1987). However there did seem to be a degree of sequence specificity, as

Figure 2.2

A H5 control gel retardation

Labelled H5 control probe (100 pg; see 2.2.1) was incubated with nuclear extract (1 μl; 10 μg/μl) with or without unlabelled heterologous competitor DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a non-denaturing 10% polyacrylamide gel (7.3.12 (i)). Non-specific (NS) and specific (S; H5) complexes formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain H5 probe. Lanes are as follows;

1; H5 probe alone

2; LSCC HD2 nuclear extract

3; LSCC HD2 nuclear extract plus 100-fold excess of poly(dI-dC)

4; Nine day chicken embryo (9DCE) nuclear extract

5; 9DCE nuclear extract plus 100-fold excess of poly(dI-dC)

B 40 mer and Δ40 mer/LSCC HD2 extract gel retardations

Labelled 40 mer or Δ40 mer probe (100 pg) was incubated with LSCC HD2 nuclear extract (1 μl; 10 μg/μl) with or without various competitor DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a non-denaturing polyacrylamide gel (7.3.12 (i)). Non-specific (NS) and candidate specific (S; H1-F40) complexes formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe were detected by autoradiography and are indicated. Lanes 1 - 5 contain 40 mer probe while lanes 6 - 8 contain Δ40 mer probe. Lanes are as follows;

1; 40 mer probe alone

2; LSCC HD2 nuclear extract

3; LSCC HD2 nuclear extract plus 100-fold excess of poly(dI-dC)

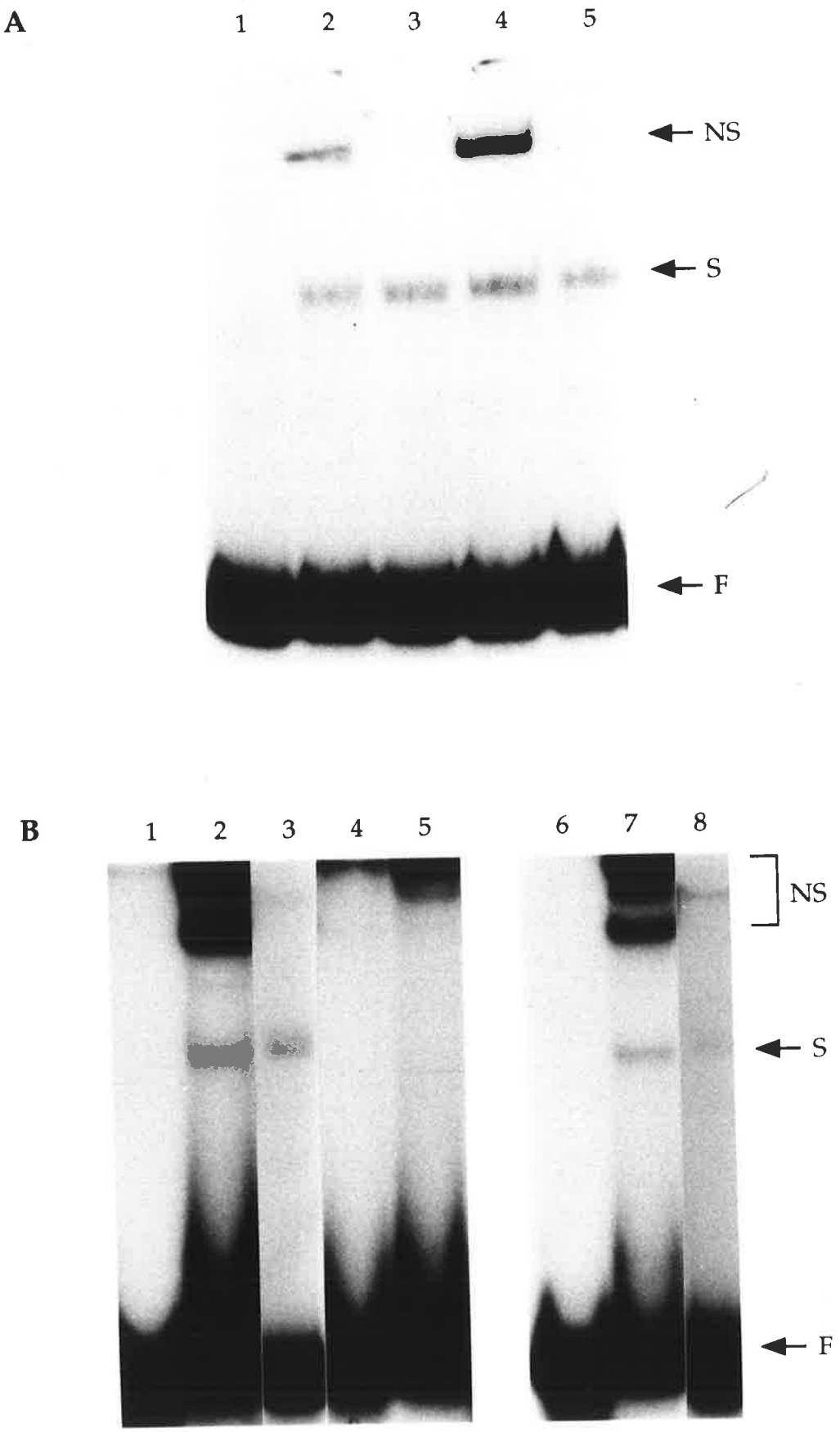
4; LSCC HD2 nuclear extract plus 10-fold excess of unlabelled 40 mer

5; LSCC HD2 nuclear extract plus 10-fold excess of unlabelled Δ40 mer

6; Δ40 mer probe alone

7; LSCC HD2 nuclear extract

8; LSCC HD2 nuclear extract plus 100-fold excess of poly(dI-dC)


A
1
2
3
4
5
NS
S
F
B
1
2
3
4
5
6
7
8
NS
S
F

the complex was not competed out using a 100-fold excess of heterologous DNA (poly(dI-dC)), but was competed out with only a 10-fold excess of homologous DNA (40 mer or Δ40 mer). Similar experiments employing salmon sperm DNA as the heterologous DNA competitior were also performed giving similar results to when poly(dI-dC) was included (data not shown).

(ii) Nine day chicken embryo (9DCE) extract

Labelled 40 mer probe or Δ40 mer probe was incubated with 9DCE crude nuclear extract and the reactions analysed by gel retardation assay (7.3.12 (i)). A slightly modified gel retardation reaction buffer was used for these experiments in an attempt to reproduce the gel retardation conditions of Dalton (1987). Binding conditions with the modified buffer were as follows: 25 mM Tris-HCl pH 7.5, 1 mM EDTA, 5 mM DTT, 150 mM NaCl, 20 mM $MgCl_2$, 0.5 mM PMSF and 10% (v/v) glycerol. Both the modified buffer and the standard buffer (7.3.12 (i)) were recommended by S. Dalton (personal communication).

The assay detected several retarded DNA-protein complexes, resulting from the binding of proteins present in the nuclear extract to the labelled probes (see Figure 2.3A). The specificity of DNA-protein complexes formed was again tested by incubation of labelled DNA with nuclear extract in the presence of unlabelled heterologous or homologous competitor DNA. Addition of an excess of unlabelled heterologous DNA (poly(dI-dC)) competed out all except two bands. An extra band was now detected in both the LSCC HD2 control and the 9DCE nuclear extracts with the 40 mer probe. The upper band of the two, appeared to be equivalent to the single band (representing H1-F40) produced when LSCC HD2 nuclear extract was incubated in the standard gel retardation reaction buffer (2.2.4 (i)), and the lower band only detected in this set of experiments with the modified gel retardation reaction buffer. However the same two bands were also seen when extract (9DCE or LSCC HD2) was incubated with the Δ 40 mer probe (Figure 2.3B). Furthermore both bands were, like the single band generated in 2.2.4 (i) gel retardations, competed out

Figure 2.3 40 mer and Δ40 mer/9DCE extract gel retardations

Labelled 40 mer or Δ40 mer probe (100 pg) was incubated with either LSCC HD2 or 9DCE nuclear extract (1 μl; 10 μg/μl), in modified gel retardation buffer, with or without various competitor DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a non-denaturing 10% polyacrylamide gel (7.3.12 (i)). Non-specific (NS) and candidate specific (S; H1-F40) complexes formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe were detected by autoradiography and are indicated.

A Lanes 1 - 6 contain 40 mer probe. Lanes are as follows;

1; 40 mer probe alone

2; LSCC HD2 nuclear extract control

3; 9DCE nuclear extract

4; 9DCE nuclear extract plus 100-fold excess of poly (dI-dC)

5; 9DCE nuclear extract plus 10-fold excess of unlabelled 40 mer

6; 9DCE nuclear extract plus 10-fold excess of unlabelled Δ40 mer

B Lanes 1 - 8 contain Δ40 mer probe. Lanes are as follows;

1; Δ40 mer probe alone

2; LSCC HD2 nuclear extract control

3; 9DCE nuclear extract

4; 9DCE nuclear extract plus 100-fold excess of poly (dI-dC)

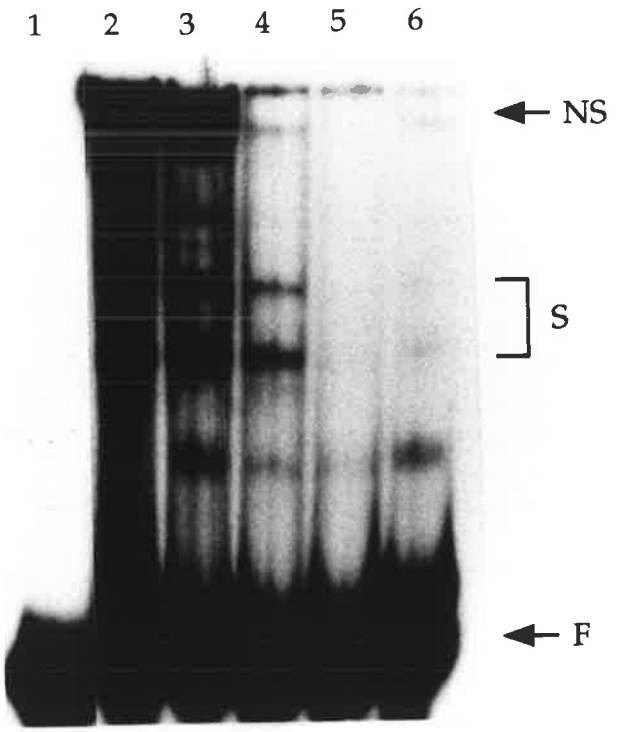
5; 9DCE nuclear extract plus 2-fold excess of unlabelled 40 mer

6; 9DCE nuclear extract plus 10-fold excess of unlabelled 40 mer

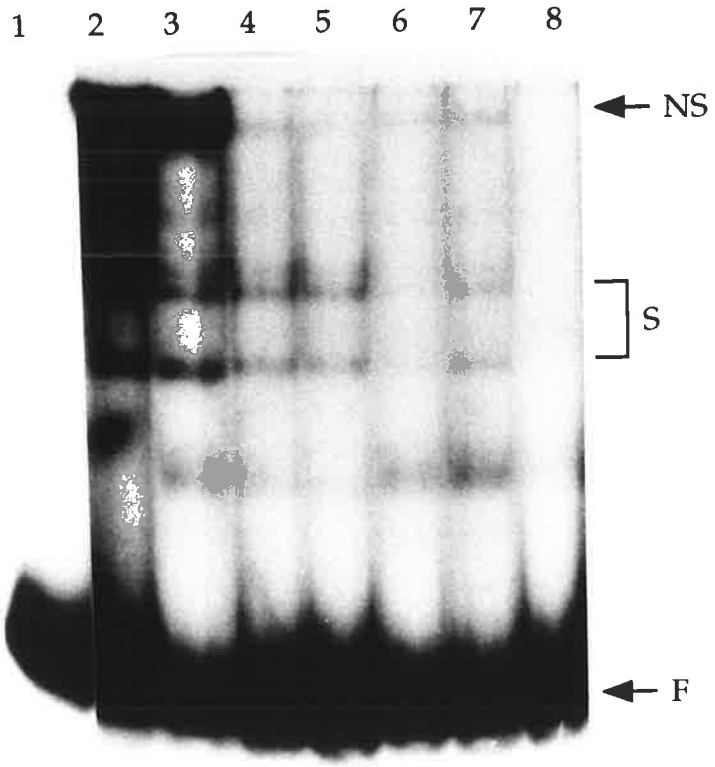
7; 9DCE nuclear extract plus 2-fold excess of unlabelled Δ40 mer

8; 9DCE nuclear extract plus 10-fold excess of unlabelled Δ40 mer

A



B



with an excess of unlabelled homologous DNA (40 mer or Δ40 mer) (Figures 2.3A and 2.3B). Again it appeared that the binding activity identified in the 9DCE nuclear extract was not H1 box specific and was not the H1-SF activity identified by Dalton (1987).

Both sets of gel retardation experiments presented in this section were internally consistent (i.e. similar band shift patterns were generated with both 40 mer and Δ40 mer probes and both LSCC HD2 and 9DCE nuclear extracts), the only difference being the appearance of a second band with the change of gel retardation reaction buffer. However this data was inconsistent with previous results obtained by Dalton (1987).

2.2.5 Replacement 40 mer and Δ40 mer

The 40 mer and Δ40 mer oligonucleotides were synthesised again to rule out the possibility of a mistake in preparing the oligonucleotides. The replacement 40 mer probe and Δ40 mer probes were prepared as described in Section 7.3.12 (i). Replacement 40 mer probe or Δ40 mer probe was then incubated with 9DCE nuclear extract and the reactions analysed by gel retardation assay (7.3.12 (i)).

Several retarded DNA-protein complexes were detected, resulting from the binding of proteins present in the crude extract to the labelled probes (Figure 2.4A). The specificity of DNA-protein complexes formed was again tested by incubation of labelled DNA with nuclear extract in the presence of unlabelled heterologous or homologous competitor DNA. However the profiles of these gel retardations reactions were essentially the same as those seen in the first set of experiments carried out using the original 40 mer and Δ40 mer probes and standard reaction buffer. The addition of an excess of unlabelled heterologous DNA (poly(dI-dC)) removed all bands, with the exception of one (H1-F40), but this band was also seen with the Δ40 mer probe (Figure 2.4A). Likewise addition of an excess of unlabelled homologous DNA (40 mer or Δ40 mer) competed out the band of interest (Figure

Figure 2.4 Replacement 40 mer or Δ40 mer gel retardations

Labelled replacement 40 mer or Δ40 mer probe (100 pg) was incubated with 9DCE nuclear extract (1 µl; 10 µg/µl) with or without various competitor DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 10% non-denaturing polyacrylamide gel (7.3.12 (i)). Non-specific (NS) and candidate specific (S; H1-F40) complexes formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe were detected by autoradiography and are indicated.

A Lanes 1 - 3 contain 40 mer probe and lanes 4 - 6 Δ40 mer probe. Lanes are as follows;

1; 40 mer probe alone

2; 9DCE nuclear extract

3; 9DCE nuclear extract plus 100-fold excess of poly (dI-dC)

4; Δ40 mer probe alone

5; 9DCE nuclear extract

6; 9DCE nuclear extract plus 100-fold excess of poly (dI-dC)

B Lanes 1 - 5 contain 40 mer probe and lanes 6 - 10 contain Δ40 mer probe. All lanes contain 9DCE nuclear extract. Lanes are as follows;

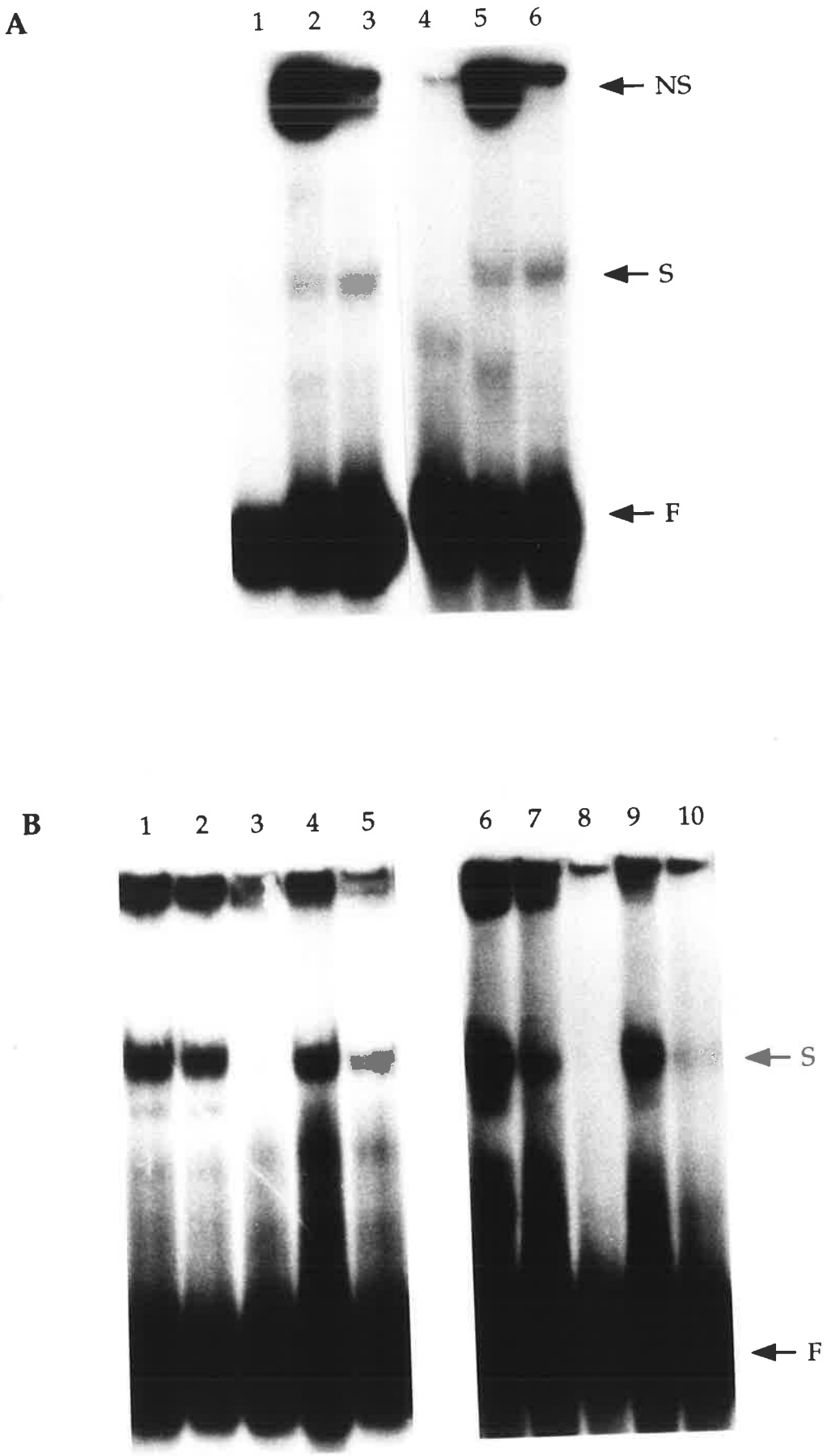
1 and 6; 100-fold excess of poly(dI-dC)

2 and 7; 2-fold excess of unlabelled 40 mer

3 and 8; 10-fold excess of unlabelled 40 mer

4 and 9; 2-fold excess of unlabelled Δ40 mer

5 and 10; 10-fold excess of unlabelled Δ40 mer


A
1 2 3 4 5 6
NS
S
F
B
1 2 3 4 5 6 7 8 9 10
S
F

2.4A and 2.4B). However the Δ40 mer appeared not to be as effective a competitior as the 40 mer in competing out both H1-F40 and the equivalent complex generated with the Δ40 mer probe.

2.2.6 Dideoxy sequencing of replacement 40mer and Δ40 mer

The replacement 40 mer and Δ40 mer double stranded oligonucleotides were cloned into *Bam*HI/*Eco*RI cut pBluescript SK+ (7.3.3) and sequenced (Sanger *et al.*, 1977; 7.3.21) to rule out any chance of the oligonucleotides having been made incorrectly. However both probes were found to be of the correct sequence (data not shown).

2.2.7 Troubleshooting: continued search for H1-SF candidate with 40 mer

Since the replacement 40 mer and Δ40 mer were shown to contain the correct sequences, it was reasoned that other parameters of the system, overlooked to this point in time, must be responsible for the failure to detect H1-SF. Therefore a number of parameters involved with the preparation of the crude protein extract and the gel retardation assay were systematically varied to try to recreate the conditions that permitted the detection of H1-SF by Dalton (1987).

(i) Nuclear extracts

In the process of developing an efficient method to prepare crude nuclear extract from 9 day chicken embryos, a number of extracts were prepared by slightly different versions of the Panyim *et al.* (1971) and Lee *et al.* (1988) hybrid method (7.3.7 (ii)) and also by the Lee *et al.* (1988) method alone. Furthermore several separate batches of crude nuclear extract were prepared using the final optimised hybrid method (7.3.7 (ii)). However none of these extracts formed any other retarded complexes with the 40 mer in the gel retardation assay (apart from H1-F40), that by appropriate competitior studies resembled the H1-SF binding activity described by Dalton (1987).

During the initial investigations for H1-SF (2.2.2 and 2.2.4 (i)), several separate preparations of crude nuclear extract were made from LSCC HD2 cells, by the method of Strauss and Varshavsky (1984) (7.3.7 (i)). Dalton (1987) reported using this method to prepare LSCC HD2 crude nuclear extract that contained H1-SF binding activity. However none of these extracts behaved any differently or formed any other retarded complexes with the 40 mer (apart from H1-F40), that by appropriate competitior studies resembled the H1-SF binding activity described by Dalton (1987).

Additional proteinase inhibitors (antipain, chymostatin, leupeptin and pepstatin A; 1 μg/ml each) were included in the buffers used in the methods for both 9DCE and LSCC HD2 extract preparation, to help prevent possible protein proteolysis, but with no obvious influence (data not shown). Increased concentrations of DTT (25 mM) and EDTA (25 mM) were trialled in the buffers used to prepare the 9DCE extract, to help protect against possible protein oxidation and further protease action respectively. However these modifications were also unsuccessful in producing extract, that by gel retardation, contained detectable H1-SF binding activity (data not shown). Extract was also assayed directly after being prepared, instead of being snap frozen (stored at -80°) and thawed before being assayed as had previously been the case. This was done as it was thought that H1-SF activity may have been lost because of excessive or inapproprate freeze-thawing of the extracts. However, no additional retarded complexes (other than H1-F40) were produced from extract treated in this manner (data not shown).

(ii) Gel retardation assay

The concentration of NaCl (0 - 0.5 M) or $MgCl_2$ (0 - 50 mM) in the gel retardation reaction (7.3.12 (i)) was varied to try to detect H1-SF binding activity (Figures 2.5A and 2.5B). However no additional retarded complexes (other than H1-SF40) were observed over the concentration range of NaCl investigated. Maximum H1-SF40 binding was observed at a concentration of 0.2 - 0.3 M NaCl. When $MgCl_2$ was

Figure 2.5 Troubleshooting gel retardations

Labelled 40 mer probe (100 pg) was incubated with 9DCE nuclear extract (1 µl; 10 µg/µl) in reactions with varying concentrations of NaCl or $MgCl_2$ (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 10% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F40) complex and free (F) unbound probe were detected by autoradiography and are indicated.

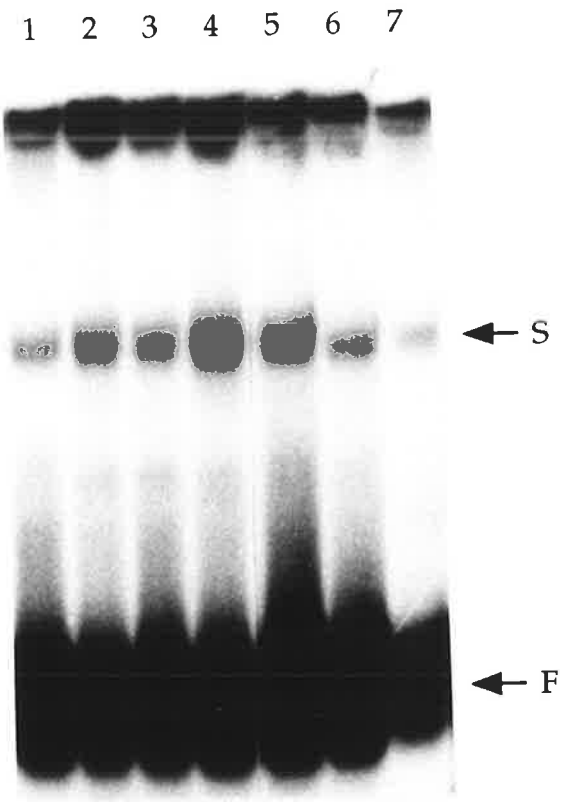
A NaCl concentration in gel retardation buffer

All lanes contain 40 mer probe and 9DCE nuclear extract. Lanes are as follows;
1, NaCl omitted from the gel retardation reaction.
2 - 6; 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 M NaCl respectively in the gel retardation reactions.

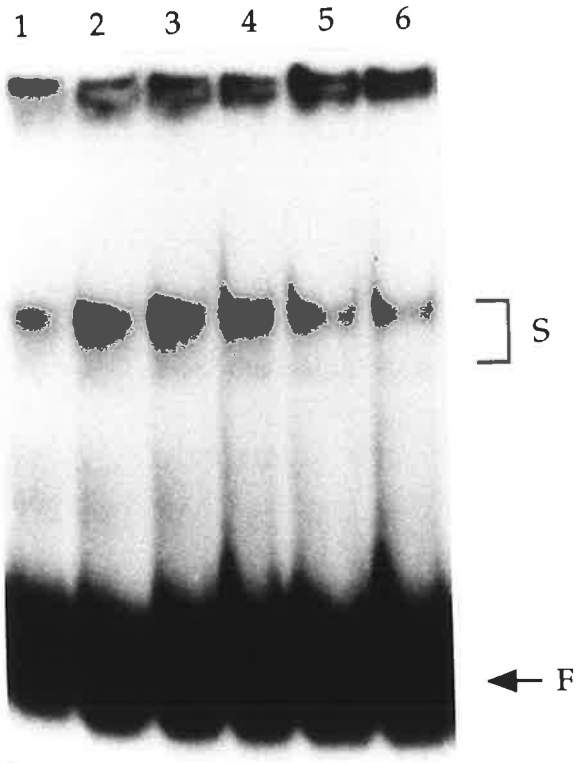
B $MgCl_2$ concentration in gel retardation buffer

All lanes contain 40 mer probe and 9DCE nuclear extract. Lanes are as follows;
1 - 6; 5, 10, 20, 30, 40, 50 mM $MgCl_2$ respectively in the gel retardation reactions.

A



B



included in the reaction buffer, an additional retarded complex to H1-F40 was faintly detectable. This may correspond to the complex (lower band) that was detected, when the modified gel retardation reaction buffer, which also contained $MgCl_2$, was used earlier to detect H1-F40 (2.2.4 (ii); see Figure 2.3A). Maximum H1-F40 binding in this latest case was observed at a concentration of 20 mM $MgCl_2$. The extra complex (lower band) may represent another DNA binding protein that requires Mg^{2+} for optimal binding activity or perhaps represents a breakdown product of H1-F40. The addition of $MgCl_2$ to the reaction buffer may promote increased protease activity in the nuclear extract, resulting in proteolytic degradation of H1-F40 during the incubation reaction.

The temperature and reaction time of the gel retardation reaction (i.e. incubation reaction of the labelled DNA and protein extract) were also varied. Incubation temperatures between 4°C and 37°C were tested, along with reaction times of between 15 minutes and 3 hours. None of these variations resulted in the detection of any additional retarded complexes, although incubation at 37°C was found to significantly reduce H1-F40 binding (data not shown).

To evaluate the possibility that H1-F40 was being irreversibly disassociated from the 40 mer probe in the time between the addition of loading buffer to the incubation reaction mixture and running an aliquot of the reaction mixture on the gel, a number of different loading buffers were tested. The gel running buffer (40 mM Tris-glycine pH 8.5 or 1 x TBE, pH 8.3), voltage, and temperature (4°C or room temperature) at which the gels were run were also varied. However, no additional retarded complexes were generated by any of these variations (data not shown). Although it was noted that H1-F40 binding activity was significantly decreased when the gel was run at room temperature instead of at 4°C. This was probably caused by the gel run at room temperature, running much warmer and as a result compromising H1-F40 activity.

Additionally, 0.01% or 0.1 % (v/v) NP-40 (a nonionic detergent) or BSA (1 mg/ml) were included in the gel retardation reaction to minimize protein loss by adsorption to the plastic or glass surfaces. The inclusion of either of these reagents did not lead to the formation of any additional retarded complexes (Data not shown).

In summary, none of the parameter changes described above resulted in the detection of any additional retarded complexes. Therefore from the combined results of this troubleshooting section, the earlier gel retardation data presented in 2.2.4 - 2.2.6, and parallel work done by K. Duncliffe who likewise was consistently only able to detect H1-F40 binding activity with the 40 mer/Δ40 mer probes using independently prepared extracts (this laboratory; personal communication), did not reproduce or explain the previously reported results of Dalton and Wells (1988*a*).

2.3 STUDY OF H1-SF/H1 BOX INTERACTION WITH 14 MER AND Δ14 MER

With the failure to detect a distinct H1-SF candidate with the 40 mer probe, a second set of probes was designed. It had been suggested to prevent as many non-specific interactions from occurring as possible, that double stranded oligonucleotide probes be made as small as possible while still including the relevant DNA recognition sequence (R. Tijan, Department of Molecular and Cell Biology, University of California; communication with J.R.E. Wells). Using these guidelines, two probes, one containing an intact H1 box (14 mer) and the other a disrupted version of the H1 box (Δ14 mer), were designed (see Figure 2.6A).

2.3.1 Search for H1-SF candidate with 14 mer

The 14 mer probe was prepared as described in 7.3.12 (i) and then incubated with 9DCE nuclear extract with, or without, unlabelled heterologous DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by non-denaturing polyacrylamide gel electrophoresis (7.3.12 (i)). However it was not possible to detect any retarded bands in the gel retardation because the 14 mer probe was being broken

Figure 2.6

A 14 mer and Δ14 mer gel retardation probes

The proximal promoter region of the chicken histone H1.01 gene is shown (see Figure 1.4 for further chicken histone H1.01 5' noncoding sequence). Conserved promoter elements are highlighted and numbering is with respect to the cap site. The sequences of both the 14 mer and Δ14 mer gel retardation probes are also shown. Base changes made to the Δ14 mer probe disrupting the H1 box are designated with asterisks. Additionally, both 14 mer and Δ14 mer contain *Sac*I overhangs at both ends.

B 14 mer/9DCE extract gel retardation

Labelled 14 mer probe (100 pg) was incubated with 9DCE nuclear extract (1 μl; 10 μg/μl) with or without heterologous competitor DNA (7.3.12 (i)). Samples were electrophoresed on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). Free (F) unbound probe and free single strand (F(SS)) probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe. Lanes are as follows;

1; 14 mer probe alone

2; 9DCE nuclear extract

3; 9DCE nuclear extract plus 100-fold excess of poly (dI-dC)

C 14 mer/heparin-Sepharose purified 9DCE extract gel retardation

Labelled 14 mer probe (100 pg) was incubated with each of the fractions collected from a small scale heparin-Sepharose column purification of 9DCE nuclear extract (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel electrophoresis (7.3.12 (i)). A candidate specific (S; H1-F14) complex and free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe, while 10 μl each of the 0.1 (flow-through), 0.2, 0.3, 0.4 M NaCl fractions and 5 μl of the 1.0 M NaCl fraction from the column were analysed. Lanes are as follows;

1; 0.1 M NaCl (flow-through) fraction

2 - 6; 0.2, 0.3, 0.4 and 1.0 M NaCl fractions

A

-220 -210 -200 -190 -180 -170 -160
GCTGGAAACTCTCCCGAACGCAAGTACCTGCTCTTCTCTTCCCTCACCGAGGAGACGGGGCGATTTGGTGGCAGAA

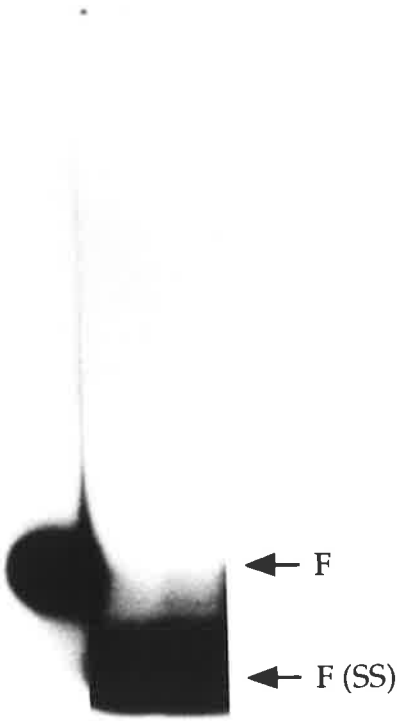
-150 -140 -130 -120 -110 -100 -90 -80
TTCCGAGGAAAATACACTTTTGTTAGTCCAAAGAAACACAAATCGAGCACACCGAAGGGCTCCCCGGCCGTGCAGC
H1 box

-70 -60 -50 -40 -30 -20 -10 +1
GGGGCGGGCTTAGCAACGCACCAATCACCGCGCGGCTCCTCTCTAAAAATACGAGCATCTGACCCGCGCCAGCCCA
G/C CCAAT TATA

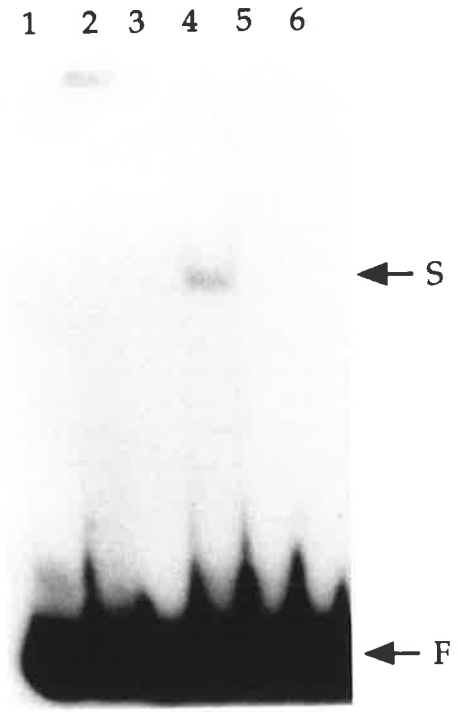
H1 box
5'- AAGAAACACAAGCT -3' **14 mer**
3'- TCGATTCTTTGTGT -5'

* ** *
5'- AAGGACTACGAGCT -3' **Δ14 mer**
3'- TCGATTCCTGATGC -5'

B



C



down (actually being 'melted' into its two single stranded constituents), presumably by a component of the crude nuclear extract (Figure 2.6B).

2.3.2 Small scale heparin-Sepharose column chromatography

It was decided to partially purify the 9DCE extract by heparin-Sepharose column chromatography (7.3.11 (i)). It was hoped that this would separate the activity responsible for melting the probe from potential H1-SF binding activity. An aliquot of 9DCE nuclear extract (10 ml; 10 mg/ml) was diluted to 0.1 M NaCl concentration in TM buffer, and passed down a small scale heparin-Sepharose column (10 ml volume). The column was washed with two column volumes of 0.1 M NaCl TM buffer, and fractions of the flow-through material collected. Subsequently, the column was washed with a step gradient of 0.2, 0.3, 0.4, and 1.0 M NaCl TM buffer washes, each of two column volumes, and fractions collected. The fractions from the column were tested in a gel retardation assay with the 14 mer probe (Figure 2.6C). The partial purification of the 9DCE nuclear extract, by heparin-Sepharose chromatography, seemed to overcome the melting problems with the 14 mer probe. A single retarded band was detected in the 0.3 M NaCl fraction from the heparin-Sepharose column.

2.3.3 Characterisation of H1-SF candidate: H1-F14

Labelled 14 mer or Δ14 mer was incubated with 9DCE 0.3 M NaCl heparin-Sepharose (9DCE HS) extract and the reactions analysed by gel retardation assay (7.3.12 (i)).

(i) 14 mer vs Δ14 mer

The H1-SF candidate complex produced with the 14 mer probe and the 9DCE HS extract was not seen when the Δ14 mer was used as the probe (Figure 2.7). This suggested that the binding activity detected with the 14 mer (designated H1-F14) was H1 box specific and may represent the H1-SF activity identified by Dalton (1987).

Figure 2.7 14 mer vs Δ14 mer gel retardation

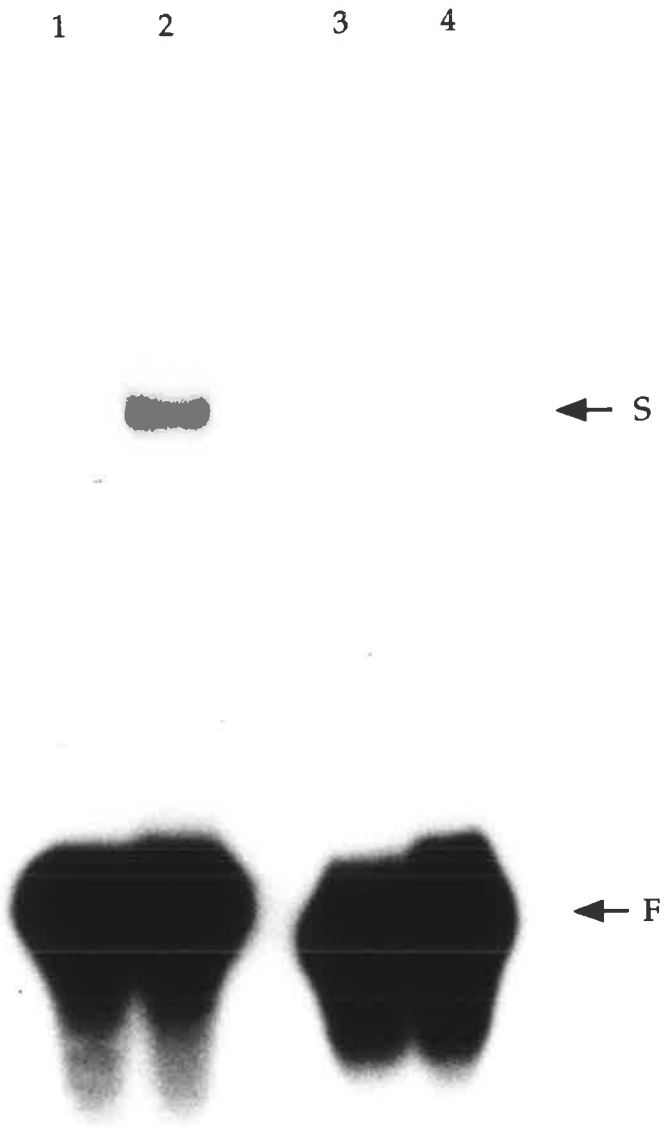
Labelled 14 mer or Δ14 mer probe (100 pg) was incubated with 9DCE 0.3 M NaCl heparin-Sepharose (9DCE HS) extract (10 µl; 1 µg/µl) (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). A candidate specific (S; H1-F14) complex and free (F) unbound probe were detected by autoradiography and are indicated. Lanes are as follows;

1; 14 mer probe alone

2; 14 mer probe and 9DCE HS extract

3; Δ14 mer probe alone

4; Δ14 mer probe and 9DCE HS extract


1
2
3
4
S
F

(ii) Heterologous and homologous competitors

The effects of several different unlabelled heterologous competitor DNAs (poly(dI-dC), salmon sperm DNA, calf thymus DNA and *E. coli* DNA) on H1-F14 binding activity were examined (Figure 2.8A). The various competitor DNAs competed with the H1-F14 complex to different extents. Poly(dI-dC) was found to be the most effective heterologous competitor. In fact, it competed out the H1-F14 complex almost as if it was a nonspecific interaction (< 100-fold excess sufficient to compete out the H1-F14 complex), while the H1-F14 complex behaved more as expected for a specific interaction with the other three heterologous competitors (1,000, 500 and 500-fold excesses required for the salmon sperm, calf thymus and *E. coli* DNA competitors respectively, to compete out the H1-F14 complex). These observations did not necessarily indicate that the binding activity observed was due to a non-specific DNA-binding protein. There are precedents for sequence specific DNA-binding proteins binding with higher avidity to some heterologous competitor DNAs compared to others (Lichtsteiner *et al.*, 1987). The heterologous competitors are quite different in sequence and would presumably have varying abilities to compete with the H1-F14 complex.

The effects of unlabelled homologous competitor DNAs (14 mer or Δ14 mer) on the H1-F14 complex were examined (Figure 2.8B). A 10-fold excess of 14 mer was sufficient to compete out the H1-F14 complex, while a 100-fold excess of Δ14 mer only just began to compete out the H1-F14 complex. This was consistent with the binding activity observed, representing a specific interaction with the H1 box.

(iii) Double stranded vs single stranded binding protein

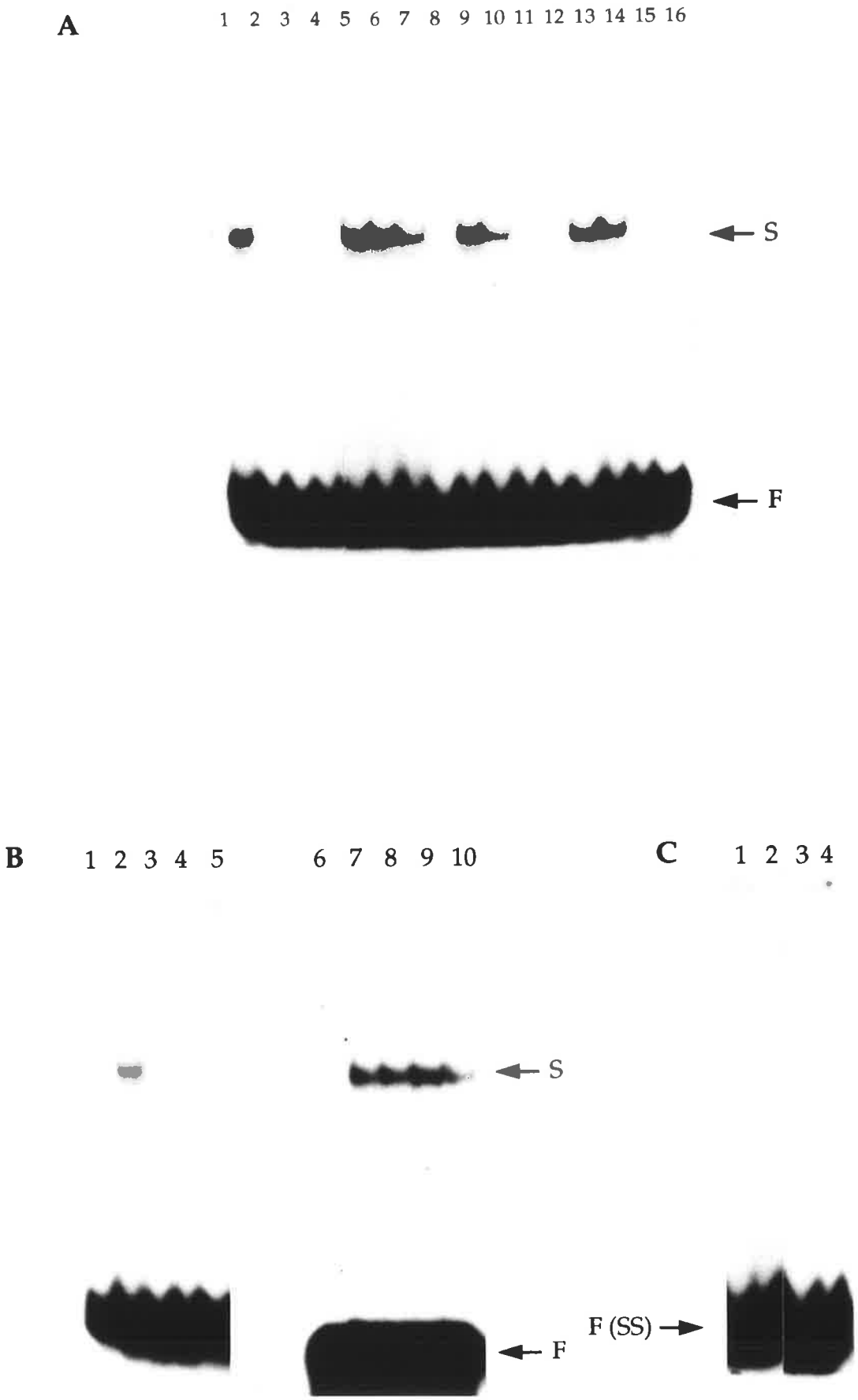
Finally, it was examined whether the H1-F14 binding activity was double stranded H1 box specific, or represented an interaction with a single stranded binding protein. Labelled single stranded 14 mer probes (+/coding strand and -/noncoding strand) were incubated with 9DCE HS extract and the reactions analysed by gel retardation assay (7.3.12 (i)). Neither single stranded 14 mer probes produced retarded

Figure 2.8 Heterologous and homologous competitor analysis of H1-F14
Labelled 14 mer probe (100 pg) was incubated with 9DCE HS extract (10 µl; 1 µg/µl) with or without various competitor DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F14) complex and free (F) unbound probe were detected by autoradiography and are indicated.

A All lanes contain 14 mer probe and 9DCE HS extract. Lanes are as follows;
1 - 4; 10, 100, 500 and 1,000-fold excess poly(dI-dC)
5 - 8; 10, 100, 500 and 1,000-fold excess salmon sperm DNA
9 - 12; 10, 100, 500 and 1,000-fold excess calf thymus DNA
13 - 16; 10, 100, 500 and 1,000-fold excess *E. coli* DNA

B All lanes contain 14 mer probe, while all lanes except 1 and 6 contain 9DCE HS extract. Lanes are as follows;
1; 14 mer probe alone
2 - 5; Competitor omitted, 10, 50 and 100-fold excess unlabelled 14 mer
6; 14 mer probe alone
7 - 10; Competitor omitted, 10, 50 and 100-fold excess unlabelled Δ14 mer

C Labelled single stranded 14 mer probe (+/coding strand or -/noncoding strand; 100 pg) was incubated with 9DCE HS extract (10 µl; 1 µg/µl) (7.3.12 (i)). Samples were electrophoresed on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). Free single strand (F(SS)) probe was detected by autoradiography and is indicated. Lanes are as follows;
1; Single stranded 14 mer probe (+ strand) alone
2; Single stranded 14 mer probe (+ strand) and 9DCE HS extract
3; Single stranded 14 mer probe (- strand) alone
4; Single stranded 14 mer probe (- strand) and 9DCE HS extract


A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
S
F
B
1 2 3 4 5 6 7 8 9 10
S
F
C
1 2 3 4
F (SS)

complexes (Figure 2.8C). This suggested that the H1-SF candidate binding activity was not caused by a single stranded binding protein.

Taken together, the 14 mer/Δ14 mer set of gel retardation results suggested that an H1 box specific binding activity (H1-F14) had been detected with the 14 mer probe in 9DCE HS extract and that H1-F14 binding activity represented the H1-SF activity reported by Dalton and Wells (1988*a*).

2.4 DISCUSSION

This chapter describes the efforts made to detect a chicken *trans*-acting protein factor (H1-SF) that had previously been shown to interact specifically with the H1 box, a highly conserved histone H1 promoter element (Dalton and Wells, 1988*a*).

Two double stranded oligonucleotide probes (designed and used previously to detect H1-SF; Dalton and Wells, 1988*a*), one containing an intact H1 box (40 mer) and the other a disrupted version of the H1 box (Δ40 mer), were used in conjunction with the gel retardation assay technique to investigate the H1-SF/H1 box interaction. Crude nuclear extracts prepared from both, nine day chicken embryos and the chicken erythroid cell line, LSCC HD2, shown previously to contain H1-SF (Dalton, 1987), were assayed for H1-SF activity. Gel retardation analysis of both extracts with the 40 mer and Δ40 mer probes failed to identify an H1 box-specific binding activity (H1-SF). Although a DNA binding activity (designated H1-F40) was identified, in both types of crude nuclear extract, which interacted with both the 40 mer and Δ40 mer probes. H1-F40 binding activity did appear to show some sequence specificity even though it bound both 40 mer and Δ40 mer. The H1-F40 complex was not competed out using a 100-fold excess of heterologous DNA but was competed out with only a 10-fold excess of homologous DNA (40 mer or Δ40 mer) (see Figures 2.2B, 2.3A 2.4A and 2.4B). Furthermore the Δ40 mer appeared not to be as effective a competitor as the 40 mer in competing out both the H1-F40 complex and the equivalent complex generated with the Δ40 mer (see Figure 2.4B). This was

consistent with both the 40 mer and Δ40 mer probes complexing with the same DNA binding protein (H1-F40) which bound the 40 mer with slightly higher affinity than the Δ40 mer. If the 40 mer and Δ40 mer complexes involved two different proteins (that by chance ran at similar positions on the gel) it would be expected that the Δ40 mer competitor would have competed out the Δ40 mer complex more efficiently than the H1-F40 complex.

S. Dalton (personal communication) never observed a complex with the DNA binding characteristics described for H1-F40. H1-SF binding activity had only ever been detected with the 40 mer and never with the Δ40 mer (not even trace or reduced levels of binding). Furthermore, Gallinari *et al.* (1989), using similarly designed probes, containing intact and disrupted copies of the H1 box but corresponding to the sequence of the human histone Hh9 H1 gene, have reported an H1 box specific binding activity (H1TF1) in HeLa cells (see 1.11.2) that presumably by definition only bound the probe containing the intact H1 box. Therefore, considering the consistency of these studies it was reasoned that other parameters of the system, overlooked to this point in time, must be responsible for the failure to detect H1-SF. Other parameters concerned with the preparation of crude nuclear extract and the gel retardation assay were thoroughly investigated (see 2.2.7) with the aim of recreating the conditions that enabled Dalton (1987) to detect H1-SF. Additionally the 40 mer and Δ40 mer oligonucleotides were resynthesized to rule out the possibility of an error in synthesis of the oligonucleotides and sequenced to confirm they had been made correctly (see 2.2.5 and 2.2.6). The oligonucleotides were found to be correct and none of the other changes carried out permitted the detection of a potential H1-SF candidate, or for that matter any additional DNA binding activity, apart from the above mentioned H1-F40. Thus, the inconsistencies between the results described in 2.2, and the results reported by Dalton and Wells (1988*a*) could not be explained. Interestingly, the H1-F40 complex ran at an identical position to the H1-SF complex run under similar gel electrophoresis conditions (comparison to unpublished primary data; S. Dalton).

With the failure to detect a distinct H1-SF candidate with the 40 mer probe, a second set of probes were designed, one containing an intact H1 box (14 mer) and the other a disrupted version of the H1 box (Δ14 mer) (see Figure 2.6A). The rationale for these changes was to prevent as many non-specific interactions from occurring as possible by making the oligonucleotide probes as small as possible while still including the relevant DNA recognition sequence.

A binding activity (H1-F14) was identified in 9DCE nuclear extract, but only after partial purification by heparin-Sepharose chromatography (see 2.3.2). H1-F14 activity was detected exclusively in the 0.3 M NaCl fraction from the heparin-Sepharose purification of 9DCE extract (see Figure 2.6C). Interestingly, H1-SF activity had also previously been found to elute exclusively in the 0.3 M NaCl fraction from a similar heparin-Sepharose column purification (Dalton, 1987). With this in mind, and also in case similar purification had actually been necessary to detect H1-SF with the 40 mer, the heparin-Sepharose fractions were assayed for H1-F40 activity with the 40 mer and Δ40 mer probes. H1-F40 activity was detected predominantly in the 0.2 M NaCl fraction with faintly detectable activity in the 0.3 M NaCl fraction. However the Δ40 mer probe again produced the same complex (H1-F40) and in the same fractions (data not shown).

Importantly, in contrast to H1-F40, H1-F14 was found not to bind the Δ14 mer probe containing the disrupted H1 box (see Figure 2.7). This strongly suggested that the H1-F14 binding activity was H1 box specific and may actually represent the H1-SF activity identified by Dalton (1987). Competition studies of H1-F14 binding activity were carried out with both heterologous competitors (poly(dI-dC), salmon sperm DNA, calf thymus DNA and *E. coli* DNA) and homologous competitor DNAs (14 mer or Δ14 mer). The competitor studies were generally consistent with H1-F14 representing a specific interaction with the H1 box (see Figures 2.8A and 2.8B; also see 2.3.3 (ii) and (iii)). Although, in the case of poly(dI-dC), the H1-F14 complex was competed out rather easily, almost as if it represented a nonspecific interaction (but

still at least 10-fold less efficiently than 14 mer competitor DNA). There are precedents for sequence specific DNA-binding proteins binding with higher avidity to some heterologous competitor DNAs compared to others and binding to these non-specific DNAs with almost comparable avidity to the specific DNA interactions (Lichtsteiner *et al.*, 1987). Furthermore the small size of the 14 mer may artifactually limit the strength of the H1-F14/14 mer interaction, because flanking sequences perhaps normally involved in specific contacts or just general complex stabilization are missing.

Following the demonstration that H1-F14 appears to represent an H1 box specific binding activity (and possibly represents the H1-SF activity detected by Dalton (1987)), the next step in the study of H1-F14 was to isolate larger quantities of the protein, in a sufficiently pure form to allow determination of the protein sequence, with the ultimate aim of isolating the gene coding for H1-F14.

CHAPTER 3

APPROACHES TO THE PURIFICATION OF H1-F14

3.1 INTRODUCTION

Various approaches have been taken to isolate genes coding for DNA binding proteins. The traditional approach has usually involved column chromatography techniques to obtain sufficient quantities of the purified protein to generate amino acid sequence data which in turn allows design of specific oligonucleotide probes, to screen recombinant DNA libraries. In the case of DNA binding proteins, the highly efficient and specific technique of DNA affinity chromatography can be incorporated into the purification scheme (Kadonaga and Tjian, 1986). A second approach, developed recently to isolate DNA binding proteins, involves oligoscreening of cDNA expression libraries with appropriate recognition site probes (Singh *et al.*, 1988; Vinson *et al.*, 1988). This approach avoids the difficult and labour-intensive step of protein purification in the isolation of clones encoding DNA binding proteins.

Work involving both approaches is discussed in this chapter. Extensive oligoscreening of a ten day chicken embryo cDNA λgt11 expression library for the H1-SF candidate, detected by the 14 mer (H1-F14), was carried out and a large scale protein purification protocol for H1-F14 was also undertaken.

3.2 OLIGOSCREENING

Oligoscreening is a recently developed method designed to isolate cDNA clones encoding DNA binding proteins (Singh *et al.*, 1988; Vinson *et al.*, 1988). The method has been successfully used to isolate a large number of different types of DNA binding proteins (reviewed, Singh *et al.*, 1989). Oligoscreening essentially involves probing protein replica filters of an expression library with labelled double stranded recognition site DNA, in a manner analogous to the immunological screening of an expression library (Young and Davis, 1983*b*). The success of the method relies heavily on the functional expression in, *E.coli*, of the DNA binding domain of a protein and just as importantly, strong binding of the DNA binding domain to its DNA recognition site. However there are limitations to the method which include its inability to be used to isolate DNA binding proteins that either require eukaryotic

post-translational modifications or two or more subunits for DNA binding activity. Another problem is that the β-galactosidase fusion may render DNA binding proteins incapable of binding DNA. Despite these limitations, the oligoscreening technique is a very powerful method and as mentioned, can by-pass protein purification of a DNA binding protein for the purpose of isolating its gene.

A ten day chicken embryo cDNA λgt11 expression library was obtained (Clontech library, catalogue number CL1001b; 1.06 x 10^6 independent clones) and oligoscreened with the aim of isolating a clone(s) encoding H1-F14. Oligoscreening was carried out using the Vinson *et al.* (1988) method as described in 7.3.15 (ii). As a control, a CCAAT/enhancer binding protein (C/EBP) λgt11 clone (Landschultz *et al.*, 1988) was obtained from S.McKnight (Howard Hughes Research laboratories, Carnegie Institution of Washington).

3.2.1 Probe preparation

Concatenated probes for both C/EBP (see 7.2.8 for sequences of the oligonucleotides used to form the double stranded C/EBP probe) and H1-F14 oligoscreening were generated by kinasing the C/EBP binding site and 14 mer oligonucleotides respectively with γ^{32}P-ATP, annealing, and ligating to form multimers of each of 5 - 10 copies. The procedure is described in more detail in 7.3.15 (i).

3.2.2 C/EBP control oligoscreening

An aliquot of the C/EBP λgt11 clone was absorbed to *E.coli* strain Y1090, and plated onto a 87 mm L-agar + 50 μg/ml ampicillin plate so as to give approximately 100 C/EBP plaques on the plate. Another plate, that served as a negative control, had an aliquot of ten day chicken embryo cDNA λgt11 library plated on it so as to give approximately 5,000 plaques on the plate. After 4 hours of growth at 42°C, fusion proteins were induced by the addition of IPTG soaked filters for 6 hours (first protein lifts), or 2 hours (second protein lifts). Filters were air-dried at room temperature for 15 minutes before denaturation/renaturation processing (7.3.15 (ii)). Filters were

probed with a concatenated C/EBP probe, washed and autoradiographed (7.3.15 (ii)). The filters lifted from the plate containing λgt11 library plaques only, were negative as anticipated (unless of course, fortuitously, there had been C/EBP clone(s) in the 5,000 ten day chicken embryo cDNA λgt11 library plaques plated), while duplicate C/EBP positives were detected on the filters lifted from the C/EBP plate, suggesting that the oligoscreening technique was working satisfactorily (Figure 3.1).

3.2.3 H1-F14 oligoscreening

The ten day chicken embryo cDNA λgt11 library was chosen because, as described in Chapter 2, nuclear extract from nine day chicken embryos was shown to contain H1-F14 binding activity. To closely follow the Vinson *et al.* (1988) method, oligoscreening for H1-F14 was carried out using their oligoscreening buffer. This was done after it had been shown that a similar level of H1-F14 activity was detected with the Vinson *et al.* (1988) oligoscreening buffer compared to the standard gel retardation buffer (Figure 3.2A). Labelled 14 mer was incubated with 9DCE HS extract, in either standard gel retardation reaction buffer (7.3.12 (i)) or Vinson *et al.* (1988) oligoscreening buffer (7.3.15 (ii)) and the reactions analysed by gel retardation assay (7.3.12 (i)).

H1-F14 oligoscreening was carried out as described for the C/EBP control, except that the ten day chicken embryo cDNA λgt11 library was absorbed to *E.coli* strain Y1090, and plated onto 15 cm L-agar + 50 μg/ml ampicillin plates with approximately 50,000 plaques per plate. Filters were analysed with the concatenated 14 mer probe. The ten day chicken embryo cDNA λgt11 library was screened four times (~ 10^6 phage per screening). A number of screening parameters were varied in an attempt to obtain positives. These included processing the filters with or without the denaturation/renaturation step and also screening with either concatenated 14 mer probe or single copy 14 mer probe. Of the four screening combinations, all were carried out in duplicate except the initial screen (denaturation/renaturation and concatenated probe). In total 31 first round positives were isolated, however only 8

Figure 3.1 C/EBP control oligoscreening

Shown are the results from the C/EBP control oligoscreening. Oligoscreening was carried out according to the method described by Vinson *et al.* (1988) (see 3.3.2 and 7.3.15 (ii)). Filters were probed with concatenated C/EBP probe (see 3.2.1). Filters are as follows;

1 and 2; First and second duplicate filter lifts from the negative control plate containing approximately 5,000 ten day chicken embryo cDNA λgt11 library plaques.

3 and 4; First and second duplicate filter lifts from the positive control plate containing approximately 100 C/EBP λgt11 clone plaques.

1

2

3

4

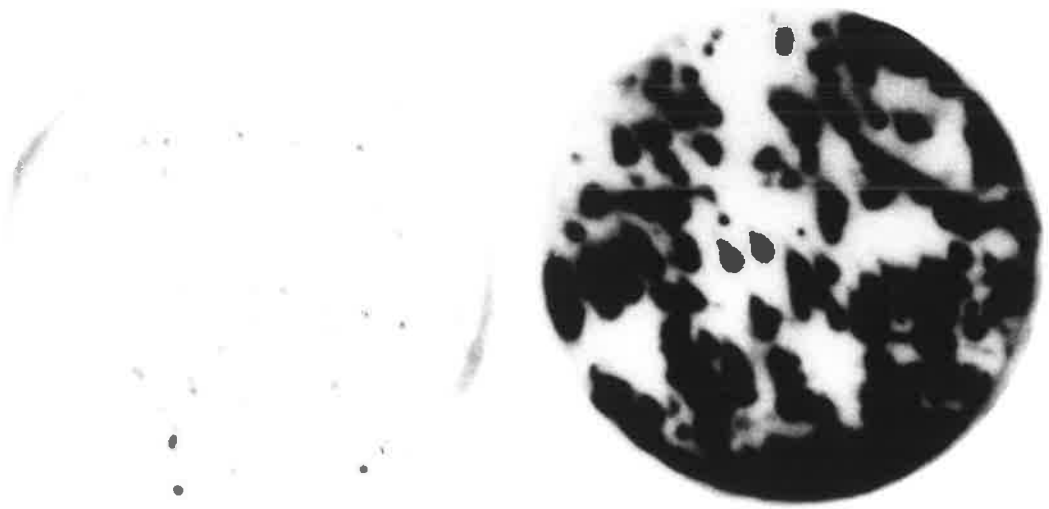


Figure 3.2

A Vinson *et al.* (1988) oligoscreening buffer vs gel retardation buffer

Labelled 14 mer probe (100 pg) was incubated with 9DCE HS extract (10 µl; 1 µg/µl) in either standard gel retardation reaction buffer (7.3.12 (i)) or Vinson *et al.* (1988) oligoscreening buffer (7.3.15 (ii)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F14) complex and free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe. Lanes are as follows;

1; 14 mer probe alone (standard gel retardation reaction buffer)

2; 9DCE HS extract (standard gel retardation reaction buffer)

3; 14 mer probe alone (Vinson *et al.* (1988) oligoscreening buffer)

4; 9DCE HS extract (Vinson *et al.* (1988) oligoscreening buffer)

B Bulk 9DCE extract/14 mer gel retardation

Labelled 14 mer probe (100 pg) was incubated with either 9DCE HS extract (10 µl; 1 µg/µl) or an aliquot from the bulk preparation of 9DCE nuclear extract (1 µl; 10 µg/µl) (see 3.3.2) (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F14) complex and free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe. Lanes are as follows;

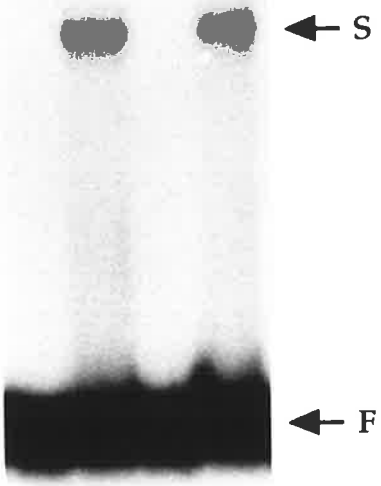
1; 14 mer probe alone

2; 9DCE HS extract control

3; 9DCE nuclear extract (see 3.3.2)

Note: an extra band appears in the 14 mer probe alone track (lane 1) not seen previously with the 14 mer probe. This band appears to be probe specific and not caused by a DNA-protein interaction but the exact origin of the band is unknown. Similar bands are seen in a number of other gel retardation figures in this chapter.

A 1 2 3 4



B 1 2 3

of these were duplicate positives. The remaining 23 single positives were picked because they looked like they were of plaque origin and also because of the large discrepancy in signal strengths between first and second lifts, and therefore the uncertainty over the reproducibility of duplicate signals. All first round positive plaques were picked into 1 ml of PSB, the phage allowed to elute overnight at 4°C, and subjected to a second round screening under the conditions used for their detection in the initial screening. This time, all protein lifts were done in duplicate, however for all first round positives, the second round screenings failed to give a signal on either first or second lift filters. This suggested that these positives were false, and not due to specific DNA-protein interaction. For further discussion of the H1-F14 oligoscreening results see 3.4.

3.3 COLUMN CHROMATOGRAPHY PURIFICATION

This section describes the work carried out to try to isolate H1-F14 by column chromatography. Partial purification of H1-F14 was to be achieved using heparin-Sepharose column chromatography having already been shown, on a pilot scale, to successfully fractionate H1-F14 (see 2.3). It was then envisaged that DNA (concatenated 14 mer) affinity chromatography could be carried out directly on this partially purified extract to purify H1-F14 to homogenity.

DNA affinity chromatography is a highly efficient and specific technique that can be used to isolate DNA binding proteins (Kadonaga and Tjian, 1986). Briefly, the method involves coupling concatenated double stranded oligonucleotides, containing the recognition site for the DNA binding protein of interest, to Sepharose beads to make a chromatography column. Protein extract, that has had non-specific competitor DNA added to it, is then passed over the column. The DNA binding protein of interest can be purified because it will bind with a higher affinity to the DNA recognition sites on the column than to the non-spesific competitor DNA in solution. Alternatively, a method developed by Franza *et al.* (1987), involves carrying out the DNA affinity purification in solution. The protein extract containing

the DNA binding protein of interest is incubated, in the presence of non-specific DNA, with specific recognition site DNA that has been biotinylated. The DNA binding protein/biotinylated DNA complexes are then collected on streptavidin agarose beads.

3.3.1 Source material

The first consideration in the large scale purification of the H1-SF candidate, H1-F14, was the source material. *Trans*-acting factors are usually present in very low quantities, of the order of 0.001% of the total cellular protein (Kadonaga and Tjian, 1986), thus a very large number of cells is required as starting material. For this project it was decided to use nine day chicken embryos as the source material rather than LSCC HD2 tissue culture cells (7.2.4 (iii)), since a large number of chicken embryos was readily available, they were comparatively inexpensive and provided a good yield of the factor. The LSCC HD2 cells, in contrast would take a relatively long time to achieve sufficient yields which in turn would be an expensive excercise.

3.3.2 Nuclear extract preparation

The first step in the large scale purification scheme was to optimise the method of preparation of the crude nuclear extract. A suitable method had already been developed on a small scale for preparing nuclear extract from nine day chicken embryos (7.3.7 (ii)). This method was scaled up with a few minor modifications to the procedure (7.3.7 (iii)). The large scale method yielded a similar amount of total protein relative to H1-F14 binding activity compared to the small scale method. Six hundred dozen nine day chicken embryos, processed in batches of 100 dozen, were used to generate 750 ml of crude nuclear extract containing approximately 7.5 g of protein (the amount of protein determined using the method of Bradford (1976; see 7.3.8). Extract was either snap frozen with liquid nitrogen, and stored at -80°C, or loaded directly onto a heparin-Sepharose column (7.3.11 (i)).

The nine day chicken embryo (9DCE) nuclear extract was assayed for H1-F14 binding activity with labelled 14 mer probe, by gel retardation assay (7.3.12 (i)).

However no retarded bands were detected because the 14 mer probe was 'melted' into single stranded form, presumably by a component in the crude nuclear extract (Figure 3.2B). This was not unexpected, as the problem had already been encountered in earlier characterisation of H1-F14 (see 2.3.1). It was found previously that H1-F14 binding activity was able to be detected after partially purifying the 9DCE nuclear extract by heparin-Sepharose column chromatography (see 2.3.2).

3.3.3 Heparin-Sepharose column chromatography

A large (200 ml) heparin-Sepharose column was used to process the crude nuclear extract prepared above (7.3.11 (i)). The 9DCE nuclear extract was diluted to 0.2 M NaCl, in TM buffer (7.2.5 (ii)), and passed onto the column. The column was washed with 2 - 3 column volumes of 0.2 M NaCl TM buffer, and 45 x 10 ml fractions of the flow-through were collected. Subsequently, the column was washed with a step gradient of 0.3, 0.4, and 1.0 M NaCl TM buffer washes, each of 2 - 3 column volumes, and 45 x 10 ml fractions collected for each. The maximum capacity of the column was approximately 1 - 2 g of nuclear extract protein, making it necessary to run the fractionation process five times. Each set of column fractions were assayed for H1-SF14 binding activity, with 14 mer probe, by gel retardation assay (7.3.12 (i)). The gel retardation analysis of the fractions collected from the first column run are shown in Figure 3.3. H1-F14 binding activity was present in fractions 16 - 45 of the 0.3 M NaCl eluate and 1 - 5 of the 0.4 M NaCl eluate. H1-F14 binding activity was detected, in all five column runs, in fractions from the 0.3 M NaCl and 0.4 M NaCl washes, with the peak of binding activity in the 0.3 M NaCl fractions and a small amount of the binding activity in the first few 0.4 M NaCl fractions. Approximately 1.5 Litres of heparin-Sepharose purified 9DCE (9DCE HS) extract, containing H1-F14 binding activity, was generated by this procedure. The scale up procedure resulted in some of the H1-F14 binding activity eluting from the heparin-Sepharose column at 0.4 M NaCl concentration whereas in the small scale experiments H1-SF14 activity had only been detected in the 0.3 M NaCl fractions. The slight difference in elution characteristics of the factor between the small scale and large scale heparin-

Figure 3.3 Large scale heparin-Sepharose purification of H1-F14

Gel retardation analysis of fractions from large scale heparin-Sepharose column chromatography purification of 9DCE nuclear extract (see 3.3.3). Fractions were eluted from the column with sequential 0.2, 0.3, 0.4 and 1.0 M NaCl TM buffer washes. Labelled 14 mer probe (100 pg) was incubated with each of the fractions collected from the column (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). Non-specific (NS) and candidate specific (S; H1-F14) complexes formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe while 10 µl each of the 0.2, 0.3 and 0.4 M NaCl fractions and 5 µl of the 1.0 M NaCl fraction were analysed. Lanes are as follows;

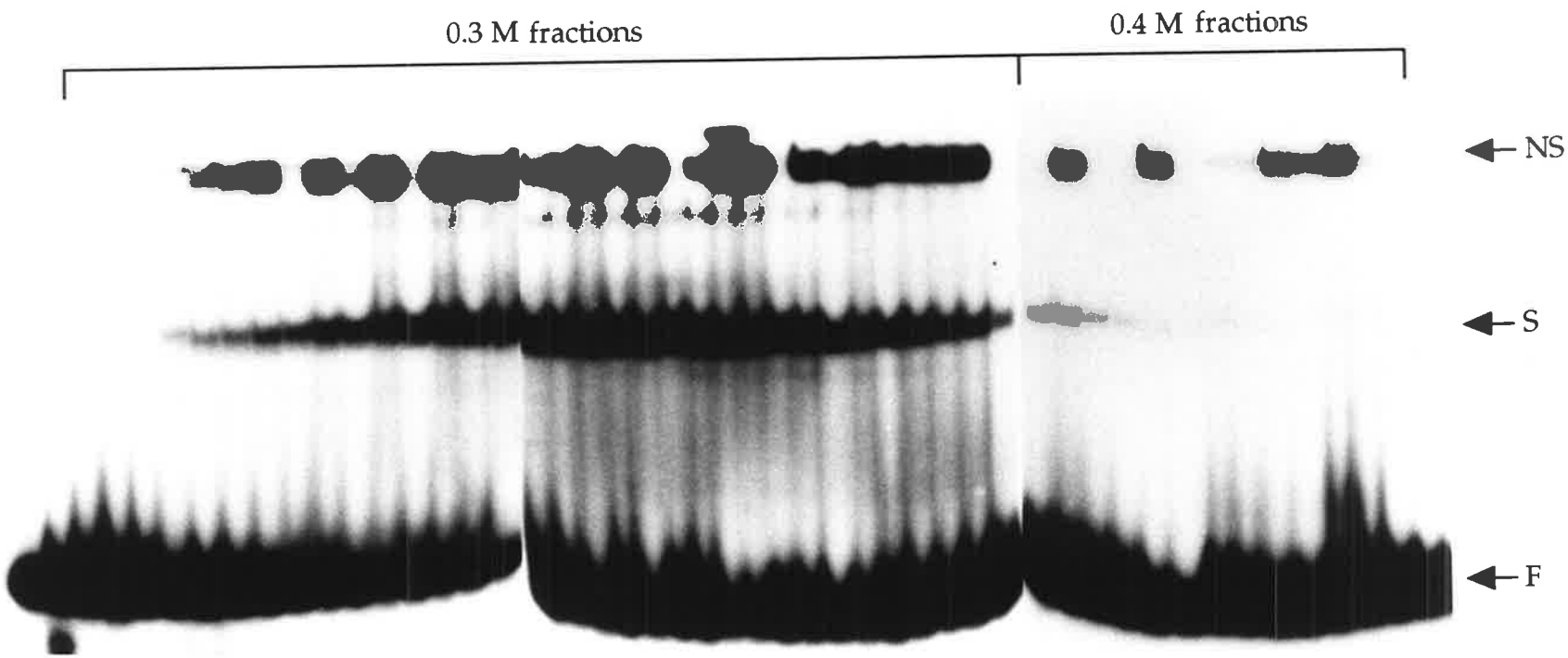
1; 14 mer probe alone

2; 0.2 M NaCl (pooled) fraction

3 - 35; 0.3 M NaCl fractions, 1, 5, 10, 15 and 20 - 45 (bracketed as 0.3 M fractions)

36 - 48; 0.4 M NaCl fractions, 1-5, 10, 15, 20, 25, 30, 35, 40 and 45 (bracketed as 0.4 M fractions)

49; 1.0 M NaCl (pooled) fraction


0.3 M fractions
0.4 M fractions
NS
S
F

Sepharose columns may be due to a difference in the rate of change of salt concentration during the elution procedure, or perhaps due to the fact that the fractions from the large scale heparin-Sepharose column(s) contained relatively more non-specific DNA binding activity. The increased non-specific DNA binding activity observed in these fractions is most likely the result of the bulk nuclear extract prepared by the large scale method containing increased amounts of contaminating proteins (probably because of inherent inefficiencies associated with scale up process) that non-specifically bind the 14 mer probe.

3.3.4 Ammonium sulphate precipitation

It was necessary to reduce the large volume of 9DCE HS extract to make the sample amenable to further purification. In a preliminary experiment, 9DCE HS extract was fractionated into two ammonium sulphate cuts (0 - 40%, 40 - 80% saturated ammonium sulphate; 7.3.9). These samples were assayed for H1-F14 binding activity, with 14 mer probe, by gel retardation (7.3.12 (i)). The H1-F14 binding activity was found to be in the 0 - 40% ammonium sulphate cut. However, a comparison of the activity of the starting material, and the 0 - 40% ammonium sulphate cut material revealed a significantly large (~ 80%) loss of H1-F14 activity (data not shown). Therefore it was decided not to use ammonium sulphate precipitation to concentrate the 9DCE HS extract and this approach was not pursued any further.

3.3.5 Pilot scale membrane filtration

The suitability of a membrane filtration approach to concentrate the 9DCE HS extract was investigated. The 9DCE HS extract (a 150 ml aliquot of the total 1.5 Litre sample) was concentrated using an Amicon ultrafiltration stirred cell (model 8200; 200 ml capacity) and ultrafiltration membrane (YM10; 10 kDa molecular mass cut off rating). The extract was concentrated approximately 10-fold over 4 hours under 400 kpa pressure at 4°C (7.3.10 (i)). The concentrated 9DCE HS extract was assayed for H1-F14 binding activity, with 14 mer probe, by gel retardation (7.3.12 (i)), but

unexpectedly the 14 mer probe was 'melted' again by this extract and H1-F14 binding activity not detected (Figure 3.4A). It was also feared that the H1-F14 activity may have been lost or inactivated during the concentration procedure. These alternatives were investigated further.

3.3.6 Troubleshooting: 14 mer probe melting

A number of gel retardation experiments were carried out to identify the cause of the 14 mer probe melting or possible loss of H1-F14 activity found after membrane filtration concentration of the 9DCE HS extract.

Concentrated 9DCE HS extract (1 µl; 8 µg/µl) was incubated in a reaction containing 9DCE HS extract (10 µl; 1 µg/µl) and 14 mer probe (7.3.12 (i)). Following gel retardation analysis the 14 mer probe was found to be intact and H1-F14 binding activity was detected (Figure 3.4A). This result seemed to suggest that it was not something in the concentrated 9DCE HS extract that was melting the 14 mer probe. However something had changed since the 9DCE HS extract had been concentrated.

It was possible that a component necessary for the H1-F14 binding activity had been lost during concentration. A 10 µl aliquot of the flow-through fraction, kept from the membrane filtration concentration run, was incubated with concentrated 9DCE HS extract (1 µl; 8 µg/µl) and 14 mer probe (note: a 10 µl aliquot of the flow-through was added back to the concentrated extract as the extract had been concentrated approximately 10-fold). The result in this case was that the 14 mer probe remained intact and H1-F14 binding activity was detected (Figure 3.4B). Flow-through (10 µl) by itself incubated with 14 mer probe was unable to produce H1-F14 binding activity (Figure 3.4B). Subsequently TM buffer (no NaCl added) or 0.3 M NaCl TM buffer (10 µl each) were added to reactions containing concentrated 9DCE HS extract (1 µl; 10 µg/µl) and 14 mer probe. In the case of TM buffer, 14 mer probe melting was not improved, however with the addition of 0.3 M NaCl TM buffer the 14 mer probe remained predominantly intact and H1-F14 binding activity was detected (Figure

Figure 3.4 Troubleshooting: 14 mer probe melting

A Labelled 14 mer probe (100 pg) was incubated with either 9DCE HS extract (10 µl; 1 µg/µl) or concentrated 9DCE HS extract (1 µl; 8 µg/µl) (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F14) complex, free (F) unbound probe and free single strand (F(SS)) probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe. Lanes are as follows;

1; 14 mer probe alone

2; 9DCE HS extract control

3; Concentrated 9DCE HS extract (see 3.3.5)

4; 9DCE HS extract plus concentrated 9DCE HS extract (see 3.3.6)

B Labelled 14 mer probe (100 pg) was incubated with either 9DCE HS extract (10 µl; 1 µg/µl), concentrated 9DCE HS extract (1 µl; 8 µg/µl) or flow-through (10 µl) from membrane filtration concentration of the 9DCE HS extract (see 3.3.5) (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F14) complex, free (F) unbound probe and free single strand (F(SS)) probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe. Lanes are as follows;

1; 14 mer probe alone

2; 9DCE HS extract control

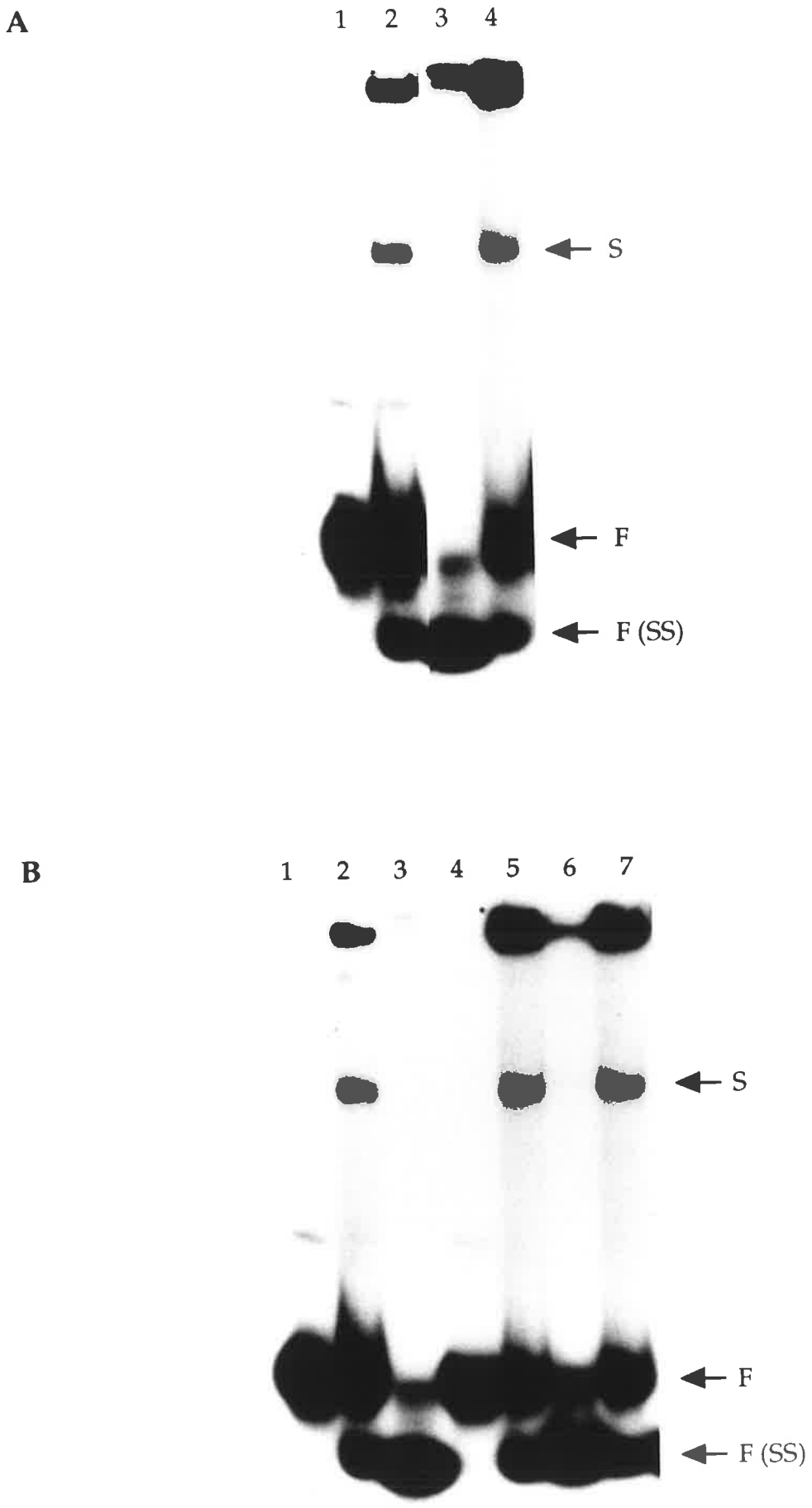
3; Concentrated 9DCE HS extract

4; Flow-through from membrane filtration concentration of 9DCE HS extract

5; Concentrated 9DCE HS extract plus flow-through from membrane filtration concentration of 9DCE HS extract

6; Concentrated 9DCE HS extract plus TM buffer (no NaCl added) (10 µl)

7; Concentrated 9DCE HS extract plus 0.3 M NaCl TM buffer (10 µl)


A
1 2 3 4
S
F
F (SS)
B
1 2 3 4 5 6 7
S
F
F (SS)

3.4B). These experiments suggested that it was not a protein component (neccessary for H1-F14 activity) lost in the flow-through during concentration, but a critical salt concentration, required to maintain 14 mer probe integrity so that the H1-F14 complex could be detected. However this threshold salt concentration was only required in the presence of extract to prevent the 14 mer probe from breaking down. The 14 mer probe, alone, was stable in standard reaction buffer. This helped explain the earlier result, that 14 mer probe remained intact and H1-F14 activity was detectable in 9DCE HS (10 µl; 1 µg/µl) extract spiked with concentrated 9DCE HS (1 µl; 8 µg/µl) extract. In this case the necessary concentration of NaCl (0.3 M) to maintain 14 mer integrity was achieved with the addition of the 10 µl of 9DCE HS extract to the reaction mixture. That is, 10 µl of 9DCE HS extract at 0.3 M NaCl plus the 2 µl of gel retardation buffer (10 x stock at 1.5 M NaCl) routinely added to the incubation reaction (20 µl total volume) gave a final NaCl concentration of 0.3 M in the reaction mixture. Whereas when only 1 µl of concentrated 9DCE HS extract was incubated with the 14 mer probe in gel retardation buffer the reaction mixture was only increased marginally to 165 mM NaCl.

(i) Salt component

The effect of NaCl concentration on the system (14 mer/H1-F14 interaction) was investigated. The concentration of NaCl (0.165 - 0.515 M) was varied in incubation reactions containing concentrated 9DCE HS extract (1 µl; 10 µg/µl) and 14 mer probe. The rections were analysed by gel retardation assay (7.3.12 (i)). A concentration of at least 0.215 M NaCl was necessary to detect H1-F14 activity, while 0.265 M NaCl or more was required to maintain the bulk of the 14 mer in a double stranded form (Figure 3.5A). H1-F14 binding activity was observed even at a concentration of 0.515 M NaCl.

(ii) Protein component

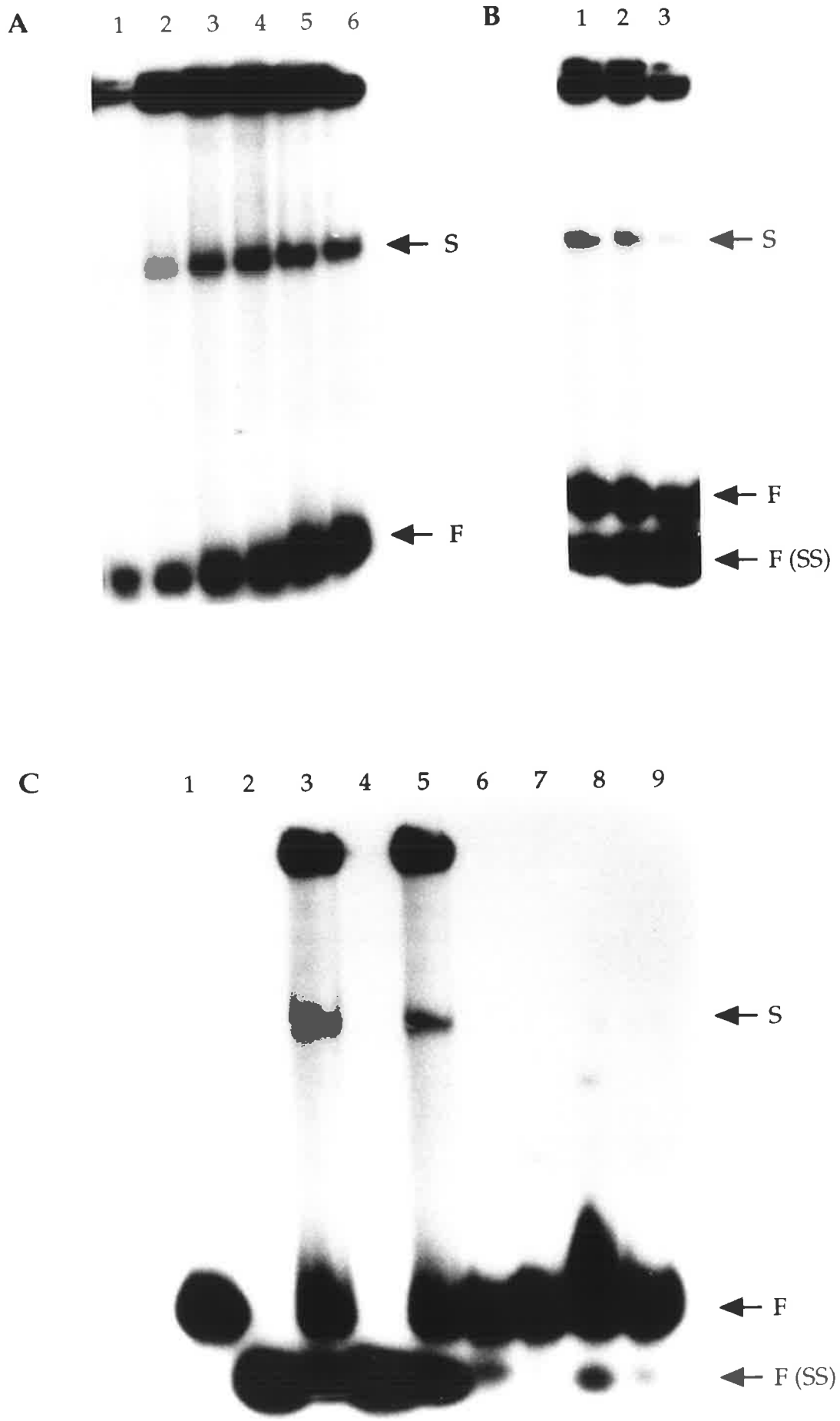
When increasing amounts of concentrated 9DCE HS extract were added to incubation reactions with 14 mer probe, increased 14 mer probe melting and

Figure 3.5 Troubleshooting: 14 mer probe melting

A Labelled 14 mer probe (100 pg) was incubated with concentrated 9DCE HS extract (1 µl; 8 µg/µl) in reactions with varying concentrations of NaCl (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F14) complex and free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe and concentrated 9DCE HS extract. Lanes are as follows;
1 - 6; 0.165, 0.215, 0.265, 0.315, 0.415 and 0.515 M NaCl (final concentration) in the respective reactions.

B Labelled 14 mer probe (100 pg) was incubated with increasing amounts of concentrated 9DCE HS extract (1 µl; 8 µg/µl) (7.3.12 (i)). Incubation reactions were made up to a final concentration of 0.3 M NaCl. DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F14) complex, free (F) unbound probe and free single strand (F(SS)) probe were detected by autoradiography and are indicated. Lanes are as follows;
1 - 3; 14 mer probe and 1, 5 or 10 µl concentrated 9DCE HS extract respectively

C Labelled 14 mer probe (100 pg) was incubated with variously treated concentrated 9DCE HS extract samples (1 µl; 8 µg/µl) (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F14) complex, free (F) unbound probe and free single strand (F(SS)) probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe. Lanes are as follows;
1; 14 mer probe alone
2 and 3; Concentrated 9DCE HS extract -/+ 1 µl of 3.0 M NaCl respectively
4 and 5; Concentrated 9DCE HS extract (incubated for 1 hr at room temperature before incubation with 14 mer probe) -/+ 1 µl of 3.0 M NaCl respectively
6 and 7; Concentrated 9DCE HS extract (incubated with proteinase K for 1 hr at room temperature (see 3.3.6. (ii)) before incubation with 14 mer probe) -/+ 1 µl of 3.0 M NaCl respectively
8 and 9; Concentrated 9DCE HS extract (incubated for 15 minutes at 100°C before incubation with 14 mer) -/+ 1 µl of 3.0 M NaCl respectively


A
1 2 3 4 5 6
S
F
B
1 2 3
S
F
F (SS)
C
1 2 3 4 5 6 7 8 9
S
F
F (SS)

corresponding decreased H1-F14 binding activity was observed (note: NaCl concentration adjusted to 0.3 M in each reaction) (Figure 3.5B). This was investigated further with a set of experiments designed to examine if a protein component was involved in the probe melting problem. Concentrated 9DCE HS extract (20 µl) was either treated with proteinase K (5 µl; 100 µg/ml) for 1 hour at room temperature or heated at 100°C for 15 minutes. Aliquots (1 µl) from each of the treated extracts were then incubated with 14 mer probe and analysed by gel retardation (7.3.12 (i)). It was found that the treated extracts did not cause 14 mer probe melting whereas untreated controls did (Figure 3.5C). Thus it appeared that there was still a heat labile, proteinase K-sensitive component (presumably a protein), associated with the 9DCE HS extract, melting the 14 mer probe, that had not been removed by heparin-Sepharose chromatography. However by adding extra NaCl this effect could be overcome and H1-F14 activity detected. Whether or not the NaCl concentration required to maintain 14 mer probe integrity was actually optimal for H1-F14 binding was not clear.

It was concluded that another step was needed to separate H1-F14 binding activity from the probe melting activity. Otherwise the 9DCE HS extract in its current state of purity may prevent optimal DNA affinity chromatography purification of H1-F14.

3.3.7 Bulk membrane filtration concentration of 9DCE HS extract

The remaining 9DCE HS extract (1.35 Litres) was batch concentrated (150 ml at a time) using the Amicon ultrafiltration stirred cell device as described above (7.3.10 (i)). The 9DCE HS extract was concentrated approximately 10-fold generating 120 ml of concentrated 9DCE HS extract. This extract was shown, by gel retardation assay (7.3.12 (i)), to contain H1-F14 activity if the incubation reaction was carried out at a final concentration of 0.3 M NaCl (data not shown).

3.3.8 DEAE-Sepharose column chromatography

DEAE-Sepharose fractionation of the concentrated 9DCE HS extract was investigated with the hope that it would separate the activity responsible for the 14 mer probe melting from H1-F14 binding activity. An aliquot of concentrated 9DCE HS extract (5 ml; 10 mg/ml) was diluted to 0.1 M NaCl concentration in TM buffer, and passed down a small scale DEAE-Sepharose column (10 ml volume) (7.3.11 (ii). The column was washed with two column volumes of 0.1 M NaCl TM buffer, and the flow-through fractions collected. Subsequently, the column was washed with a step gradient of 0.2 and 0.3 M NaCl TM buffer washes, each of two column volumes, and fractions collected. The fractions from the column were tested by gel retardation assay with the 14 mer probe (7.3.12 (i)). H1-F14 activity was detected exclusively in the 0.1 M NaCl fractions (Figure 3.6A).

The 0.1 M NaCl fractions collected from the DEAE-Sepharose column were pooled, concentrated approximately 10-fold using an Amicon centriprep spin column (7.3.10 (ii)) and assayed for H1-F14 activity by gel retardation analysis with the 14 mer probe (7.3.12 (i)). Probe melting activity still appeared to be co-purifying with H1-F14 activity (Figure 3.6B). Furthermore, H1-F14 activity was significantly diluted during DEAE-Sepharose chromatography and it seemed that little obvious enrichment of H1-F14 activity with respect to non-specific DNA binding activity as judged by gel retardation (Figure 3.6B) and total protein (see Figure 3.10A) was achieved. Therefore DEAE-Sepharose purification of H1-F14 was not pursued any further.

3.3.9 Gel filtration column chromatography

Gel filtration column chromatography was next investigated as an additional purification step. A Sephacryl S-300 column (400 ml) was used to fractionate concentrated 9DCE HS extract (20 ml) (7.3.11 (iii)). Once the sample had been applied to the column 0.3 M NaCl TM buffer was run onto the column and 80 x 5 ml fractions were collected. The fractions were tested by gel retardation analysis for H1-

Figure 3.6 DEAE-Sepharose column chromatography

A Gel retardation analysis of fractions from a trial DEAE-Sepharose column chromatography purification of concentrated 9DCE HS extract. Labelled 14 mer probe (100 pg) was incubated with each of the 0.1, 0.2 and 0.3 M NaCl fractions collected from the column (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F14) complex and free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe while 10 µl each of the 0.1, 0.2 and 0.3 M NaCl fractions were analysed. Lanes are as follows;

1; 14 mer probe alone

2 and 3; 0.1 M NaCl fractions

4 and 5; 0.2 M NaCl fractions

6 and 7; 0.3 M NaCl fractions

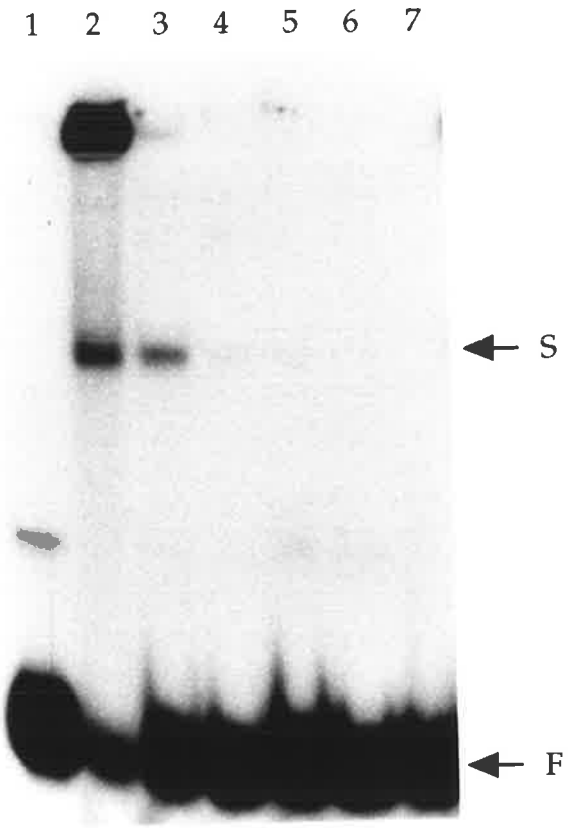
B Labelled 14 mer probe (100 pg) was incubated with either 9DCE HS extract (10 µl; 1 µg/µl) or concentrated 9DCE HS 0.1 M NaCl DEAE-Sepharose (9DCE HSDEAE) extract (1 µl; 6 µg/µl) (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F14) complex and free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe. Lanes are as follows;

1; 14 mer probe alone

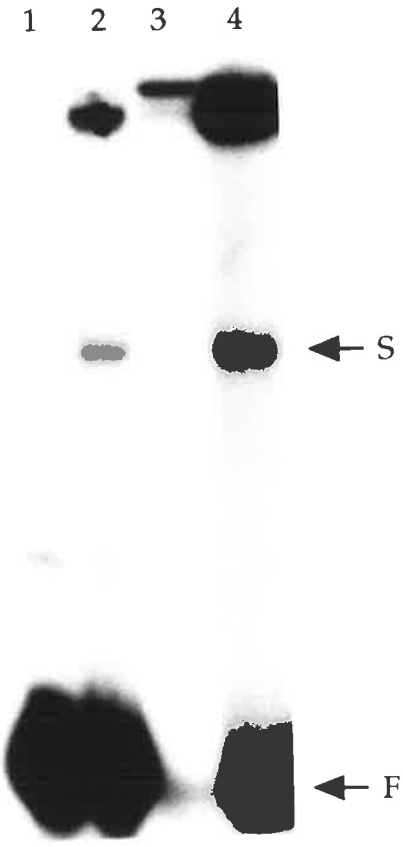
2; 9DCE HS extract control

3 and 4; Concentrated 9DCE HS DEAE extract -/+ 1 µl of 3.0 M NaCl respectively

A



B



F14 binding activity. Peak H1-F14 binding activity was detected, in fractions 61 - 66 (Figure 3.7A).

Sephacryl S-300 purified fractions containing H1-F14 activity were pooled, concentrated using an Amicon centriprep spin column (7.3.10 (ii)), and then reassayed for H1-F14 activity by gel retardation analysis with 14 mer probe (7.3.12 (i)). A large amount of the non-specific binding activity in the 9DCE HS extract was removed by the Sephacryl S-300 purification (Figure 3.7B). Additionally Sephacryl S-300 chromatography seemed to provide a significant purification of H1-F14 binding activity with respect to total protein (see Figure 3.10A). However it failed to remove the probe melting activity from H1-F14 containing fractions (Figure 3.7B).

Nevertheless it was decided that DNA affinity chromatography should be attempted on the Sephacryl S-300 purified 9DCE HS extract with the hope that H1-F14 activity could still be efficiently purified in spite of the probe melting activity. In the case of DNA affinity chromatography the 14 mer would be concatenated to form a much larger probe, thereby hopefully substantially stabilizing the DNA and avoiding the melting problems that have plagued the gel retardation analysis of H1-F14, with the single copy 14 mer probe.

Therefore, bulk Sephacryl S-300 chromatography purification of the remaining concentrated 9DCE HS extract was carried out. The maximum amount of sample that could be loaded on the Sephacryl S-300 column (400 ml total volume) was approximately 20 ml (~ 5% of the total column volume). This made it necessary to run the fractionation process five times to purify the remaining 100 ml of concentrated 9DCE HS extract (7.3.11 (iii)). Each set of fractions from each run were tested by gel retardation assay, with the 14 mer probe, for H1-F14 binding activity. Approximately 180 ml of Sephacryl S-300 purified 9DCE HS (9DCE HSS-300) extract, containing H1-F14 binding activity, was obtained and pooled.

Figure 3.7 Gel filtration chromatography

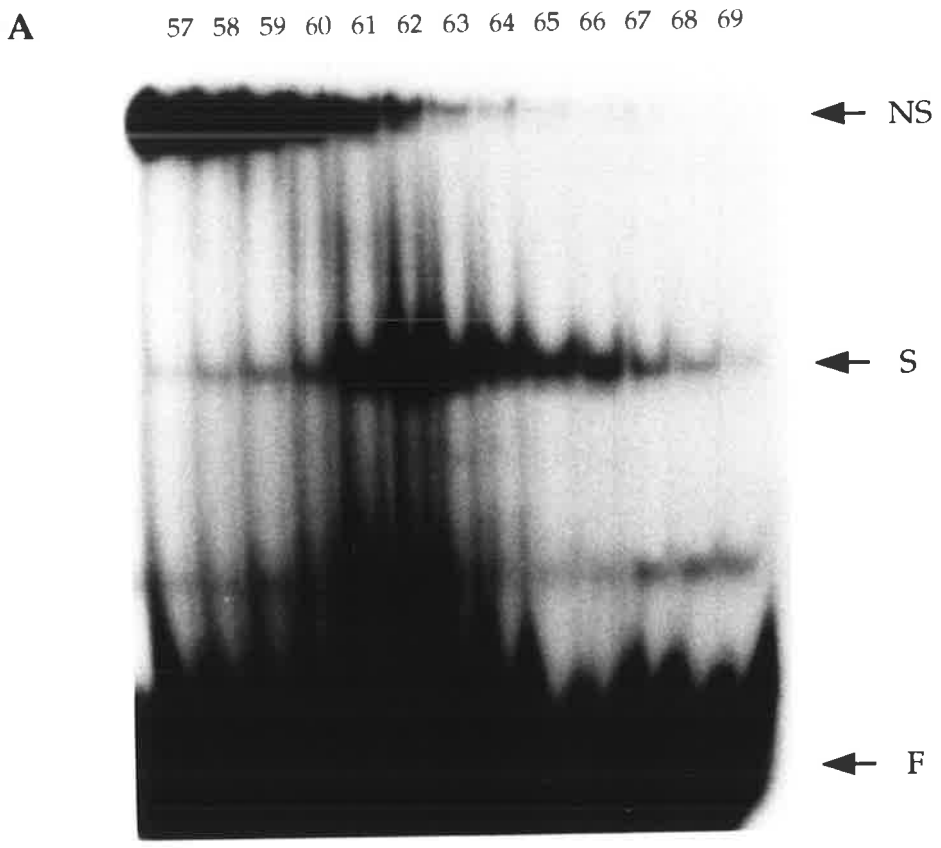
A Gel retardation analysis of fractions from Sephacryl S-300 column chromatography purification of concentrated 9DCE HS extract. Labelled 14 mer probe was incubated with each of the fractions collected from the column (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). Non-specific (NS) and candidate specific (S; H1-F40) complexes formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe (100 pg) while 10 µl each of the 80 x 5 ml fractions were analysed. However only those fractions containing H1-F14 binding activity are shown and the numbering represents the fraction numbers as they were collected from the column.

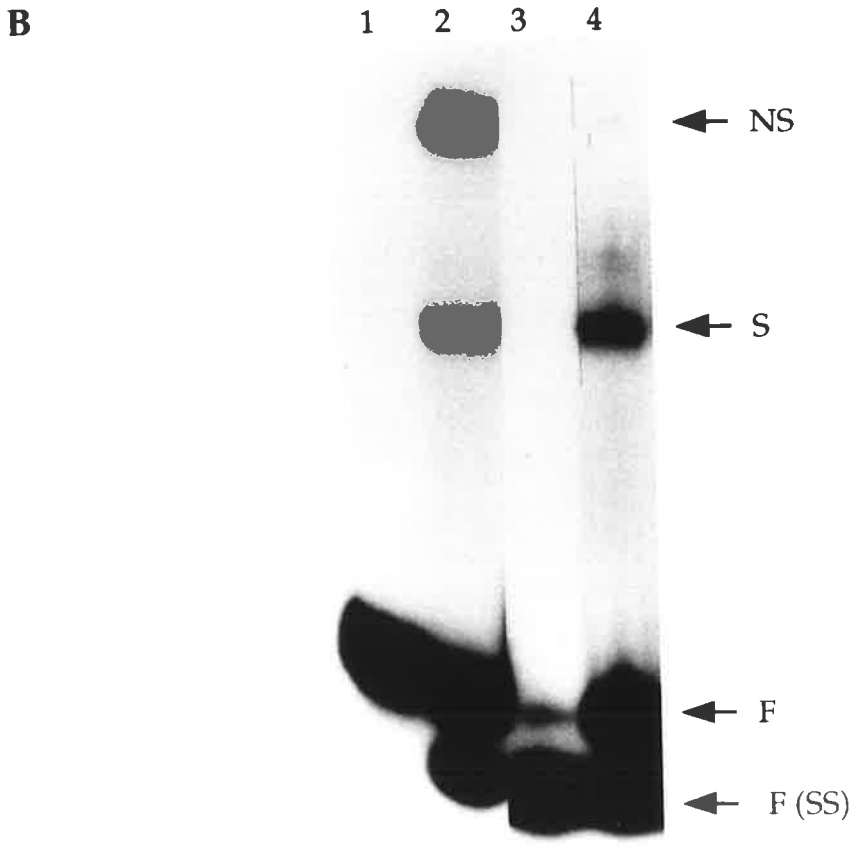
B Labelled 14 mer probe (100 pg) was incubated with either 9DCE HS extract (10 µl; 1 µg/µl) or concentrated Sephacryl S-300 purified 9DCE HS (9DCE HSS-300) extract (1 µl; 0.5 µg/µl) (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). Non-specific (NS) and candidate specific (S; H1-F40) complexes formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe and free single strand (F(SS)) probe were detected by autoradiography and are indicated. All lanes contain 14 mer probe. Lanes are as follows;

1; 14 mer probe alone

2; 9DCE HS extract control

3 and 4; Concentrated 9DCE HSS-300 extract -/+ 1 µl of 3.0 M NaCl respectively


A
57 58 59 60 61 62 63 64 65 66 67 68 69
NS
S
F


B
1 2 3 4
NS
S
F
F (SS)

3.3.10 Bulk membrane filtration concentration of 9DCE HSS-300 extract

The 9DCE HSS-300 extract (180 ml) was concentrated using the Amicon ultrafiltration stirred cell device as described above (7.3.10 (i)) so as to make it amenable to DNA affinity purification. The 9DCE HSS-300 extract was concentrated approximately three-fold generating 60 ml of concentrated 9DCE HSS-300 extract.

3.3.11 H1-SF14 is not H1-SF

At this time, data generated from work described above (3.3.11 (i)), from an experiment done as a consequence of these observations (3.3.11 (ii)), and a set of experiments (3.3.11 (iii)) carried out at the same time by K. Duncliffe (this laboratory; personal communication), were evaluated.

(i) Gel filtration size estimation

H1-F14 activity was present in fractions from the Sephacryl S-300 column chromatography purification that suggested it was much smaller than the H1-SF protein reported by Dalton and Wells (1988*b*) of approximately 90 kDa. Although the Sephacryl S-300 column had not been calibrated with protein size standards, a Superose-12 column (HPLC system) which had been trialled earlier for H1-F14 purification (data not shown) had been calibrated. H1-F14 activity was detected in fractions from the Superose-12 column that suggested that it was no larger than approximately 40 kDa. Furthermore the Sephacryl S-300 fractions containing H1-F14 activity analysed by SDS-polyacrylamide gel electrophoresis were predominantly enriched for protein species less than approximately 50 kDa (see Figure 3.10A).

(ii) Southwestern analysis

The southwestern procedure involves the probing of protein, separated by SDS-polyacrylamide gel electrophoresis and transferred onto a membrane, with labelled double stranded recognition site DNA (Matsudaira, 1987 and Miskimins *et al.*, 1985). The method can be used to gain an idea of the size of a DNA binding protein of interest. Southwestern analysis was carried out on 9DCE nuclear extract (5 µl)

extract to try to determine the size of H1-F14 (7.3.14). A single band that represented a protein of approximately 30 kDa was detected using a concatenated 14 mer probe (7.3.15 (i)) (Figure 3.8).

The results described in sections 3.3.11 (i) and (ii) were consistent with each other and together seemed to suggest that H1-F14 was much too small to be the H1-SF identified by Dalton (1987).

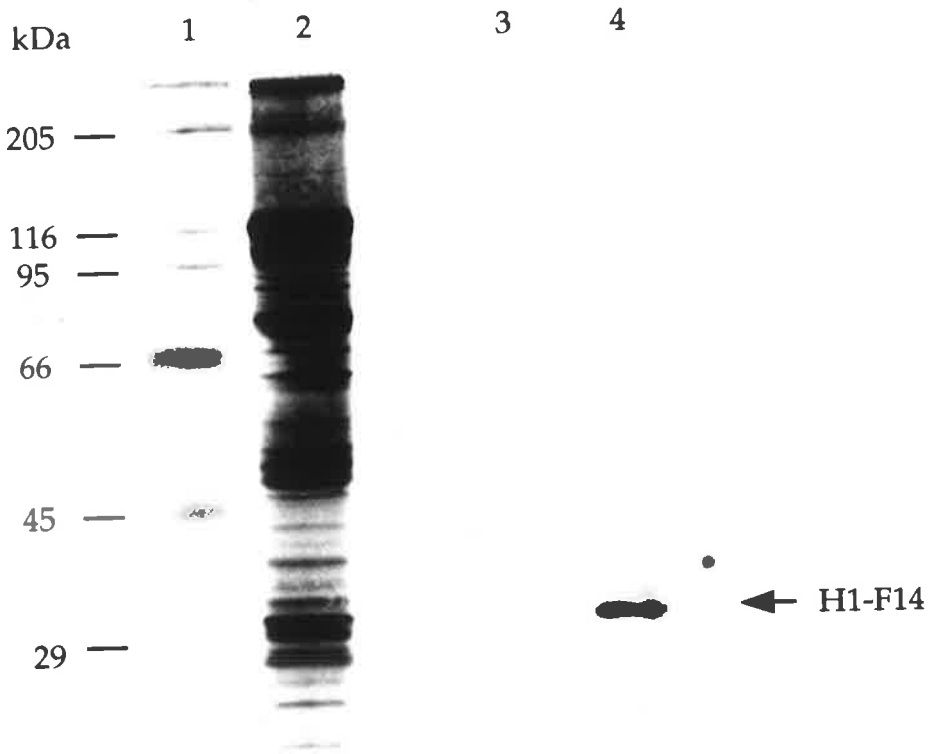
(iii) G/C and CCAAT box gel retardation assays

At this time, K. Duncliffe (personal communication), carried out a set of gel retardation experiments designed to examine the possibility that H1-F40 binding activity as detected with the 40 mer and Δ40 mer probes (see Section 2.2) may actually represent histone H1 specific G/C or CCAAT box binding activity. N. Heintz and co-workers (N. Heintz, Howard Hughes Medical Institute, Rockefeller University; communication with J.R.E. Wells) had recently purified a factor which seemed to be a good candidate for H1-SF. However it was found to be the Sp1 transcription factor. They were able to identify a 'cryptic' Sp1 binding site overlapping the H1 box that presumably resulted in the purification of the Sp1 factor. While J. Stein and co-workers (J. Stein, Department of Cell Biology, University of Massachusetts; communication with J.R.E. Wells) had observed extensive sequence homology between the H1 box (AAGAAACACA) and the histone H1 CCAAT element bound by HiNFB (see 1.11.2). The possibility that H1-F14 binding activity could actually represent one or the other of these binding activities was also examined at the same time. However, it was unlikely that H1-F14 binding activity represented Sp1 because the 14 mer would appear to lack the flanking sequence necessary for this interaction.

K. Duncliffe (personal communication) showed that the G/C and CCAAT box probes (see 7.2.8 for sequences of the oligonucleotides used to form the double stranded G/C and CCAAT box probes) were each able to form a retarded complex

Figure 3.8 H1-F14 Southwestern analysis

Shown is the Southwestern analysis of 9DCE nuclear extract with concatenated 14 mer probe. The 9DCE nuclear extract (1 µl; 10 µg/µl) was subjected to SDS-polyacrylamide gel electrophoresis, transferred to PVDF membrane, probed with concatenated 14 mer probe, in the presence of unlabelled heterologous DNA (salmon sperm DNA), and autoradiographed (see 7.3.14). Duplicate samples of protein molecular size markers and 9DCE nuclear extract were run on the SDS-polyacrylamide gel. One set of markers and nuclear extract were not transferred to the membrane but instead silver stained in the gel (lanes 1 and 2 respectively), while the other set of markers and nuclear extract was transferred to the membrane and probed as described above (lanes 3 and 4 respectively). The sizes of the markers are indicated aswell as the presumed H1-F14 complex.


kDa
1
2
3
4
205
116
95
66
45
29
H1-F14

that by competitior studies appeared to be the same as H1-F14 detected with the 14 mer probe. These gel retardation experiments were repeated and the results are presented below. Probes (14mer, Δ14 mer, G/C, and CCAAT box) were incubated with concentrated 9DCE HSS-300 extract (1 μl; 0.5 μg/μl) and the reactions analysed by gel retardation assay (7.3.12 (i)). The H1-F14 complex produced with the 14 mer probe was not seen with the Δ14 mer probe (also see 2.3.3), but retarded complexes of virtually the same mobility were produced with the G/C and CCAAT box probes (Figure 3.9A).

Such complexes may represent other unrelated DNA binding protein interactions with the G/C and CCAAT box probes, that produce similar sized retardation complexes, rather than H1-F14 binding activity. One point that should be made is that the concentrated 9DCE HSS-300 extract used in these studies was enriched for H1-F14 and more than likely does not contain Sp1 and CCAAT box factors that would normally be expected to preferentially interact with the G/C and CCAAT box probes. Another set of gel retardations were set up to investigate the possibility that the G/C and CCAAT box probe complexes were unrelated DNA binding proteins. In these experiments 14 mer, G/C, and CCAAT box probes were each incubated with 9DCE HSS-300 extract (1 μl; 0.5 μg/μl), along with 14 mer, Δ14 mer, G/C, and CCAAT box competitor DNAs, and the reactions analysed by gel retardation assay (7.3.12 (i)). H1-F14 binding activity was competed out with equivalent amounts of the 14 mer, G/C, and CCAAT box competitors while a similar amount of Δ14 mer, did not compete out the H1-F14 complex (Figure 3.9B). Furthermore the complexes formed by the G/C and CCAAT box probes were also competed out with 14 mer, G/C, and CCAAT box competitors (Figures 3.9C and 3.9D).

These results suggested that the H1-F14 binding activity was not H1 box specific. As the G/C and CCAAT box competitors competed out the H1-F14 complex almost as efficiently as 14 mer competitor, and the G/C and CCAAT box probes also both bound H1-F14 as determined by the ability of the 14 mer to compete out the G/C

Figure 3.9 G/C and CCAAT box gel retardation analysis of H1-F14
Labelled 14 mer, Δ14 mer, G/C or CCAAT box probe (100 pg) was incubated with 9DCE HSS-300 (1 µl; 0.5 µg/µl) extract with or without various competitor DNA (7.3.12 (i)). Incubation reactions were made up to a final concentration of 0.3 M NaCl. DNA-protein complexes were separated from free DNA by electrophoresis on 12% non-denaturing polyacrylamide gels (7.3.12 (i)). The candidate specific (S; H1-F14) complex and free (F) unbound probe were detected by autoradiography and are indicated.

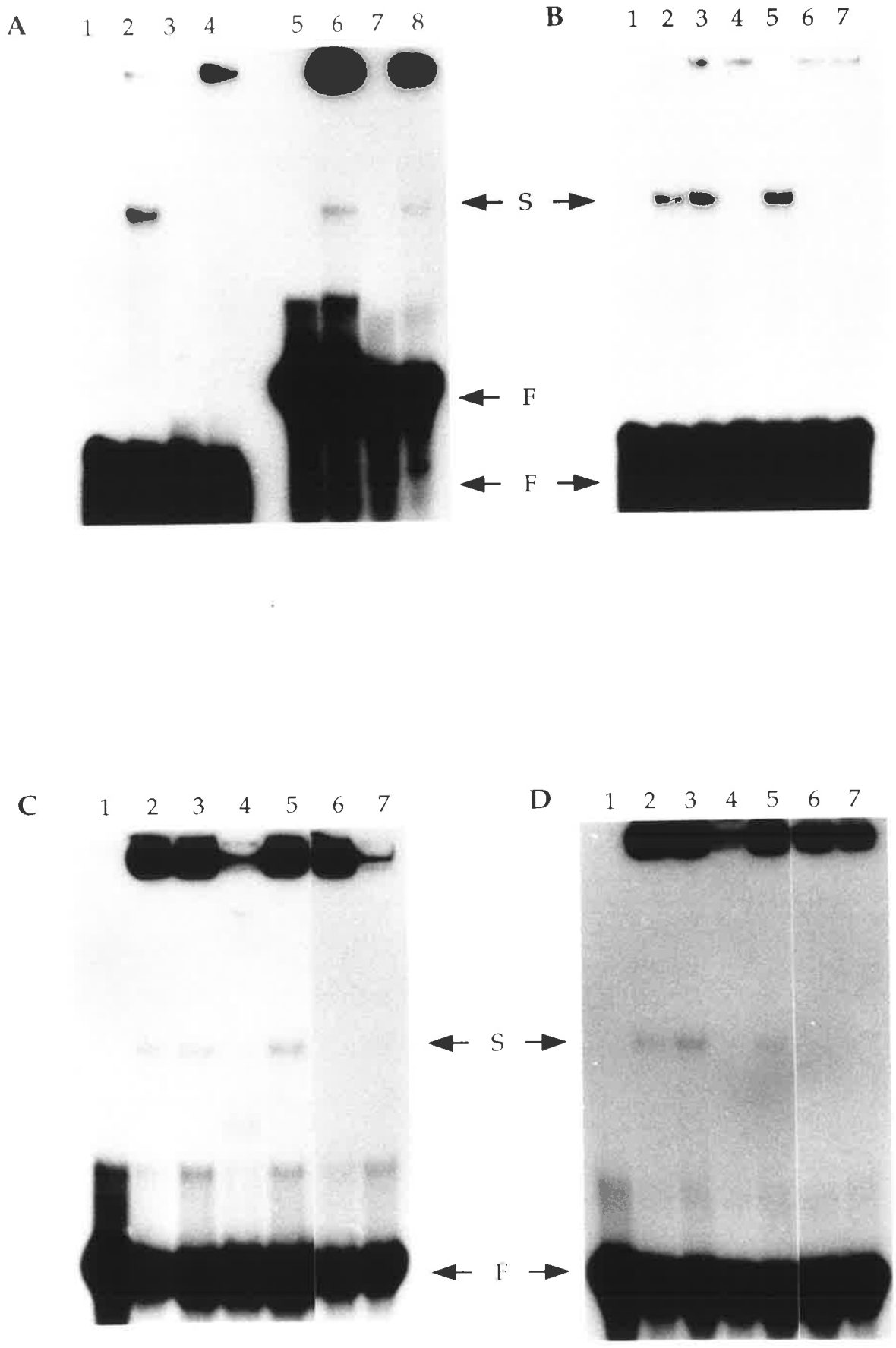
A Lanes are as follows;
1 and 2; 14 mer probe alone, 14 mer probe and 9DCE HSS-300 extract
3 and 4; Δ14 mer probe alone, Δ14 mer probe and 9DCE HSS-300 extract
5 and 6; G/C box probe alone, G/C box probe and 9DCE HSS-300 extract
7 and 8; CCAAT box probe alone, CCAAT box probe and 9DCE HSS-300 extract

B All lanes contain 14 mer probe, while all lanes except lane 1 contain 9DCE HSS-300 extract (1 µl; 0.5 µg/µl). Lanes are as follows;
1; 14 mer probe alone
2; 9DCE HSS-300 extract control
3; 20-fold excess of poly(dI-dC)
4; 20-fold excess of unlabelled 14 mer
5; 20-fold excess of unlabelled Δ14 mer
6; 20-fold excess of unlabelled G/C box competitor
7; 20-fold excess of unlabelled CCAAT box competitor

C Lanes 1 - 7 are the same as described for B above except that G/C box probe is substituted for 14 mer probe.

D Lanes 1 - 7 are the same as described for B above except that CCAAT box probe is substituted for 14 mer probe.

Note, considerably more non-specific DNA binding activity (represented by the complexes at the top of the gel not indicated) was detected with both the G/C and CCAAT box probes compared to the 14 mer and Δ14 mer probes.


A
1 2 3 4 5 6 7 8
B
1 2 3 4 5 6 7
S
F
F
C
1 2 3 4 5 6 7
D
1 2 3 4 5 6 7
S
F

and CCAAT box complexes. Why the H1-F14 activity did not bind the Δ14 mer probe was not clear, but the gel retardation data pointed strongly to H1-F14 activity being a non-specfic DNA binding protein activity and not H1-SF. H1-F14 was abandoned at this point as an H1-SF candidate.

3.4 DISCUSSION

This chapter describes the efforts made to isolate the H1-SF candidate, H1-F14. Extensive oligoscreening of a ten day chicken embryo cDNA λgt11 expression library for H1-F14 clone(s) and a large scale protein purification protocol for H1-F14 were undertaken.

The oligoscreening approach failed to isolate any H1-F14 clone(s) (see 3.2). A number of promising first round positives were detected but none of these came through a second round of screening. This suggested that these positives were false, and not due to specific DNA-protein interaction. A high background signal along with a large discrepancy in signal strengths between first and second lifts (e.g. see C/EBP oligoscreening control; Figure 3.1) made it difficult to identify bona fide positives. The oligoscreening technique had appeared to work satisfactorily with the C/EBP control which suggested that there were other problems involved with the screening process peculiar to H1-F14. For instance H1-F14 clones may have not been represented in the library to start with considering the low abundance of transcription factors in general. Additionally any one of a number of limitations of the oligoscreening method, mentioned in 3.2, may have applied in the case of H1-F14 and prevented its isolation.

With the failure of the oligoscreening technique to isolate potential H1-F14 clones the emphasis was placed on purifying H1-F14 by conventional column chromatography methods. To this end, a large amount of nuclear extract was prepared from nine day chicken embryos (3.3.2) and H1-F14 binding activity partially purified from the nuclear extract by heparin-Sepharose column chromatography. Heparin-Sepharose

chromatography was chosen as the initial column chromatography step because it had previously been reported that most DNA binding proteins bound heparin and large amounts of other material present in the crude extract passed through the column (Sorger *et al.*, 1989). The DNA binding proteins were also generally found to be able to be eluted off the column within narrow salt concentrations thus allowing further purification. Furthermore H1-F14 binding activity had already been successfully fractionated on a small scale by heparin-Sepharose column chromatography (see 2.3.2), and Dalton (1987) reported using heparin-Sepharose chromatography in a pilot scale purification of H1-SF.

Approximately 1.5 Litres of heparin-Sepharose purified 9DCE (9DCE HS) extract, containing H1-F14 binding activity was generated from the fractionation of 750 ml of crude nuclear extract. The bulk 9DCE HS extract was then concentrated by a membrane filtration procedure (3.3.5 and 3.3.7) to reduce the volume of the sample and make it more amenable to further purification. However after concentration of the 9DCE HS extract, H1-F14 binding activity was no longer able to be detected by gel retardation assay. This problem was partially solved and H1-F14 binding activity 'recovered', when it was found that a protein component in the extract was melting the 14 mer probe and in the process preventing the detection of H1-F14. By carrying out gel retardation analysis for H1-F14 in the presense of 0.3 M NaCl the 14 mer probe was stabilised and H1-F14 binding activity detected. H1-F14 binding activity had also not been able to be detected initially in crude nuclear extract (see Figure 3.2B), but on increasing the NaCl concentration to 0.3 M, binding activity was restored (see Figure 3.10A lane 6).

The removal of the probe melting activity from the extract was felt to be important to be able to carry out efficient DNA affinity chromatography purification of H1-F14. DEAE-Sepharose chromatography was carried out in the hope that it would separate the probe melting activity from H1-F14 binding activity, as it had been reported to be very useful for removing contaminating nucleic acids and importantly, phosphatases

from relatively crude protein samples (Briggs *et al.*, 1986). However, DEAE-Sepharose chromatography failed to eliminate the probe melting activity from the fractions containing H1-F14 binding activity (see 3.3.8) and was not pursued further.

Gel filtration chromatography had also been reported to be useful for separating transcription factors from contaminating nucleases (Biggin and Tijan, 1988). Therefore Sephacryl S-300 column chromatography was used to further purify H1-F14 (see 3.3.9). However, while Sephacryl S-300 column chromatography provided substantial purification of H1-F14, it failed to separate the probe melting activity from H1-F14 binding activity. The relative purification of H1-F14 achieved via the heparin-Sepharose and Sephacryl S-300 chromatography steps is presented in Figures 3.10A and 3.10B. The purification scheme achieved a 123-fold purification with a 49% recovery. The probe melting activity and H1-F14 binding activity appeared to have very similar fractionation qualities and were not separated by conventional chromatography. In fact H1-F14 binding and the probe melting activities may reside in the same polypeptide. The binding of H1-F14 to the 14 mer probe may cause the breakdown of the probe by opening up the duplex, which in the case of the small 14 mer probe would lead to its complete disassociation. The 14 mer probe is particularly susceptible to disassociation into its single stranded constituents. Whereas the 40 mer probe when incubated with crude nuclear extract was relatively stable (see 2.2; likewise the even larger H5 50 mer probe was also stable). Therefore, DNA affinity chromatography was to be attempted with the Sephacryl S-300 purified material considering that the concatenated 14 mer to be used for affinity chromatography would probably be equally as resistant to breakdown as the 40 mer and H5 50 mer probes.

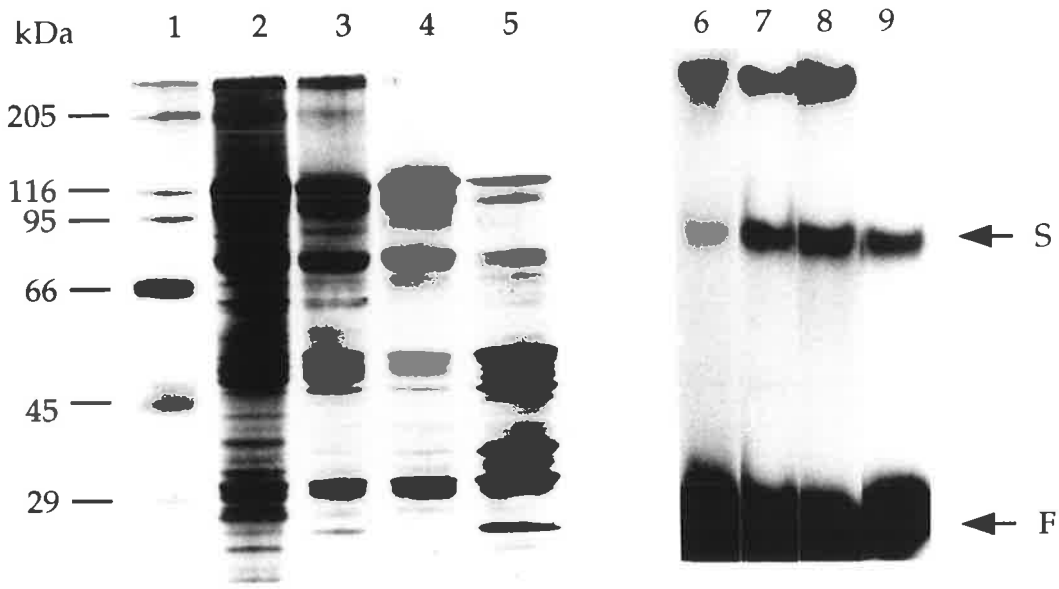
However, at this time, evidence demonstrating that the H1-F14 binding activity represented a non-specific DNA binding activity was obtained. K. Duncliffe (this laboratory; personal communication) showed that histone H1 specific G/C and CCAAT box probes were each able to form a retarded complex that by competitior

Figure 3.10 Summary of H1-F14 partial purification

A SDS-polyacrylamide gel electrophoresis (lanes 1 - 5) and gel retardation analysis (lanes (6 - 9) of protein samples from the various protein purification steps used to partially purify H1-F14. SDS-polyacrylamide gel electrophoresis and silver staining of protein carried out as described in 7.3.13 (i) and (iii). Lanes are as follows;
1; Protein molecular size markers (sizes of markers indicated)
2; 9DCE nuclear extract (1 µl)
3; Concentrated 9DCE HS extract (1 µl)
4; Concentrated 9DCE HS DEAE extract (1 µl)
5; Concentrated 9DCE HSS-300 extract (20 µl)
Gel retardation analysis carried out using 14 mer probe (100 pg) as described in 7.3.12 (i). The candidate specific (S; H1-F14) complex and free (F) unbound probe were detected by autoradiography and are indicated. Lanes are as follows;
6 - 9; Contain same extract fractions as lanes 2 - 5, but only 1 µl of extract in each case.

B Data represented in this table summarises the enrichment of H1-F14 from nuclear extract derived from nine day chicken embryos. The partial purification of H1-F14 from 9DCE nuclear extract involved chromatography using heparin-Sepharose (3.3.3) and then gel filtration through Sephacryl S-300 (3.3.9). Note that 9DCE HS and 9DCE HSS-300 refer to concentrated fractions of 9DCE HS and 9DCE HSS-300 extract respectively (see 3.3.7 and 3.3.10). DEAE-Sepharose chromatography was not included as it was only tested as a potential purification step and eventually not used (3.3.8). The amount of protein present in the nuclear extract (9DCE) and after each purification step is indicated in addition to sample volumes. SDS-polyacrylamide gel electrophoresis was performed on protein samples from each stage of the purification scheme (see A above) and protein measured by the method of Bradford (1976; see 7.3.8). Gel retardation analysis was performed on protein samples from each stage of the purification scheme (see A above). Units of activity are an approximate measure of H1-F14 binding activity present in the respective extracts. The bands representing the H1-F14/14 mer complexes from each of the fractions were excised from the gel (A above; lanes 6 - 9) and the gel slices Cerenkov counted (see 7.3.25). The counts per minute of input DNA in the H1-F14/14 mer complex were used as a measure of the H1-F14 binding activity in that fraction relative to the amount of protein in the fraction.

A



B

Fraction	Total Protein (mg)	Volume (ml)	Protein Concentration (mg/ml)	Total Activity (units)	Specific Activity (units/mg)	Fold Purification	Cumulative Yield (%)
9DCE	7,500	750	10	750,000	100	–	100
9DCE HS	960	120	8	655,000	700	7	87
9DCE HSS-300	30	60	0.5	370,000	12,300	18	49

studies appeared to be the same as H1-F14 detected with the 14 mer probe. These gel retardation experiments were repeated and the results confirmed (see 3.3.11 (iii)). The G/C, and CCAAT box competitors were able to compete out H1-F14 almost as efficiently as the 14 mer itself, suggesting that the H1-F14 binding activity was not H1 box specific but represented a non-specific DNA binding activity. Why the H1-F14 activity did not bind the Δ14 mer probe was not clear, although fortuitous sequence changes in the Δ14 mer may have simply reduced its capacity to interact with this type of DNA binding protein. The Δ14 mer probe does begin to compete out the H1-F14/14 mer complex at 10-fold higher concentrations than the 14 mer (see Figure 2.8B). In hindsight, the small size of the 14 mer may simply have not allowed H1-SF to bind, and therefore favoured a non-specific DNA binding protein interaction. Furthermore southwestern and gel filtration size estimations of H1-F14 at less than 40 kDa (see 3.3.11 (i) and (ii)) suggested, that even if the H1-F14 binding activity still represented a specific, but perhaps very low affinity, interaction with the H1 box, H1-F14 was not the 90 kDa H1-SF binding activity originally identified by Dalton (1987). As a result of these findings, the purification of H1-F14 was abandoned.

CHAPTER 4

DETECTION OF H1-F25 AND ISOLATION OF H1-F25 CANDIDATES: Z5 AND Z12

4.1 INTRODUCTION

With the abandonment of H1-F14 as an H1-SF candidate (see Chapter 3), a third set of oligonucleotide probes (25 mers) were designed to try to detect a genuine H1 box-specific binding activity. This chapter reports on the detection of an H1-SF candidate (H1-F25), with the 25 mer probes, and the isolation of potential H1-F25 clones by oligoscreening of a ten day chicken embryo cDNA λgt11 expression library.

4.2 STUDY OF H1-SF/H1 BOX INTERACTION WITH 25 MER AND Δ25 MER

4.2.1 Design of 25 mer and Δ25 mer

Two double stranded oligonucleotide probes were designed, both 25 mers, one containing an intact H1 box (25 mer) and the other a disrupted version of the H1 box (Δ25 mer) (Figure 4.1A). It had been noted that the 14 mer had been prone to breakdown whereas the 40 mer appeared reasonably stable. It was hoped that the intermediate size 25 mers would be stable and at the same time allow as few non-specific interactions to occur as possible. Furthermore, it was reasoned that sequences outside of the H1 box, absent from the 14 mer, may be important for H1-SF binding. Thus, the 25 mer would now provide enough flanking sequence for H1-SF binding. Both the 40 mer and 14 mer had single stranded overhangs engineered at their termini to aid in their concatamerisation. However, these extra sequences may have promoted non-specific interactions in both cases, at the expense of H1-SF binding. Thus, the 25 mers were made so that they were blunt ended and only contained the H1 box and immediate flanking H1 histone promoter sequence. Additionally, the Δ25 mer involved changes to six out of the possible ten bases of the AAGAAACACA sequence (H1 box) compared to the original four base changes in the Δ40 and Δ14mers. It was hoped that the above changes when incorporated into the 25 mers would help avoid the problems experienced with the 40 mer and 14 mer probes, and permit them to bind H1-SF.

Figure 4.1

A 25 mer and Δ25 mer gel retardation probes

The proximal promoter region of the chicken histone H1.01 gene is shown (see Figure 1.4 for further chicken histone H1.01 5' noncoding sequence). Conserved promoter elements are highlighted and numbering is with respect to the cap site. The sequences of both the 25 mer and Δ25 mer gel retardation probes are also shown. Base changes made to the Δ25 mer probe disrupting the H1 box are designated with asterisks.

B 25 mer vs Δ25 mer gel retardation

Labelled 25 mer or Δ25 mer probe (100 pg) was incubated with 9DCE nuclear extract (1 μl; 10 μg/μl) with or without heterologous competitor DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). Non-specific (NS) and candidate specific (S; H1-F25) complexes formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe, were detected by autoradiography and are indicated. Lanes 1 - 3 contain 25 mer probe and lanes 4 - 6 Δ25 mer probe. Lanes are as follows;

1; 25 mer probe alone
2; 9DCE nuclear extract
3; 9DCE nuclear extract plus 100-fold excess of poly (dI-dC)
4; Δ25 mer probe alone
5; 9DCE nuclear extract
6; 9DCE nuclear extract plus 100-fold excess of poly (dI-dC)

Note: an extra band appears in the 25 mer and Δ25 mer probe alone tracks (lanes 1 and 4). This band appears to be probe specific and not caused by a DNA-protein interaction but the exact origin of the band is unknown. Similar bands are seen in the other gel retardation figures in this chapter.

A

-220 -210 -200 -190 -180 -170 -160
GCTGGAAACTCTCCCGAACGCAAGTACCTGCTCTTCTCTTCCCTCACCGAGGAGACGGGGCGATTTGGTGGCAGAA

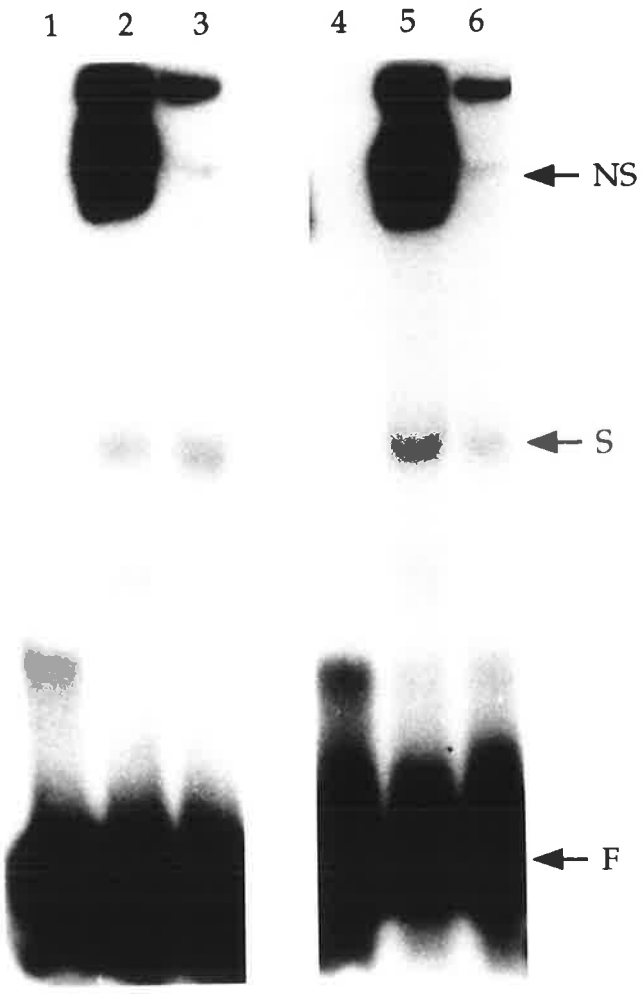
-150 -140 -130 -120 -110 -100 -90 -80
TTCCGAGGAAAATACACTTTTGTTAGTCCAAAGAAACACAAATCGAGCACACCGAAGGGCTCCCCGGCCGTGCAGC
H1 box

-70 -60 -50 -40 -30 -20 -10 +1
GGGGCGGGCTTAGCAACGCACCAATCACCGCGCGGCTCCTCTCTAAAAATACGAGCATCTGACCCGCGCCAGCCCA
G/C CCAAT TATA

H1 box
5'- GTCCAAAGAAACACAAATCGAGCAC - 3' 25 mer
3'- CAGGTTTCTTTGTGTTTAGCTCGTG - 5'

*** ***
5'- GTCCAAAGTCTCCTCAATCGAGCAC - 3' Δ25 mer
3'- CAGGTTTCAGAGGAGTTAGCTCGTG - 5'

B



4.2.2 Characterisation of H1-SF candidate: H1-SF25

(i) 25 mer vs Δ25 mer

The 25 mer and Δ25 mer probes were prepared as described in 7.3.12 (i). Labelled 25 mer or Δ25 mer probe was incubated with nine day chicken embryo (9DCE) nuclear extract, in standard gel retardation buffer, with, or without, unlabelled heterologous DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by non-denaturing polyacrylamide gel electrophoresis (7.3.12 (i)).

Several retarded DNA-protein complexes were detected in the gel retardation assay, resulting from the binding of proteins present in the nuclear extract to the labelled 25 mer probe (Figure 4.1B). The specificity of DNA-protein complexes formed were tested by incubation of labelled DNA with nuclear extract in the presence of unlabelled heterologous DNA. Addition of an excess of unlabelled heterologous DNA removed all bands, with the exception of one. However the same band was also seen with the Δ25 mer probe (Figure 4.1B). This result was very similar to that obtained with the 40 mer and Δ40 mer probes and their interaction with H1-F40 (2.2.5) which suggested, as this seemed to, that the binding activity identified (designated H1-F25) was not interacting specifically with the H1 box.

(ii) Heterologous and homologous competitors

However, as with H1-F40, there did seem to be a degree of sequence specificity, because the H1-F25 complex could not be competed out using an excess of heterologous DNA (poly(dI-dC) and also in the case of H1-F25: H5, G/C or CCAAT double standed oligonucleotides), but was competed out with equivalent amounts of homologous DNA (25 mer or Δ25 mer competitor) (Figure 4.2A).

Perhaps the H1 box is not the only sequence requirement for H1-SF binding. Other sequences immediately either side of the H1 box, which the 25 mer and Δ25 mer have in common, may also be critical. In this view, H1-F25 binding activity *in vitro* may still be specific but only slightly compromised by the mutation in the H1 box,

Figure 4.2

A Heterologous and homologous competitor analysis of H1-F25

Labelled 25 mer probe (100 pg) was incubated with 9DCE nuclear extract (1 µl; 10 µg/µl) with or without various competitor DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F25) complex formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 25 mer probe, while all lanes except lane 1 contain 9DCE nuclear extract. Lanes are as follows;

1; 25 mer probe alone

2; 100-fold excess of poly (dI-dC)

3; 20-fold excess of unlabelled 25 mer

4; 20-fold excess of unlabelled Δ25 mer

5; 20-fold excess of unlabelled H5 50 mer

6; 20-fold excess of unlabelled G/C box competitor

7; 20-fold excess of unlabelled CCAAT box competitor

B H1-F25 vs H1-F14

Gel retardation analysis of heparin-Sepharose chromatography fractions saved from the bulk purification of H1-F14 (3.3.3). Labelled 25 mer probe (100 pg) was incubated with 10 µl samples of each of the remaining pooled fractions from the heparin-Sepharose column purification (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F25) complex in addition to free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 25 mer probe. Lanes are as follows;

1; 9DCE nuclear extract control

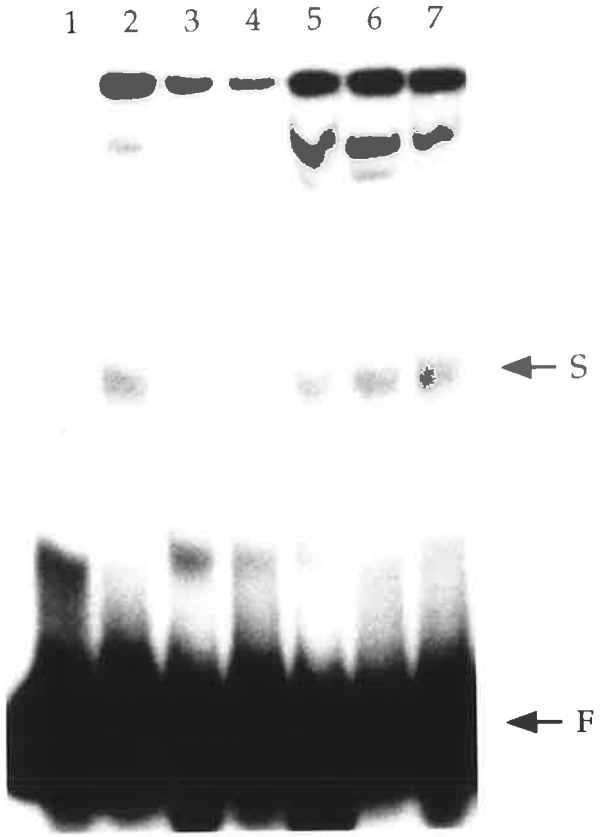
2; 0.3 M NaCl fraction originally identified as containing H1-F14 binding activity

3 and 4; 0.2 M NaCl fractions

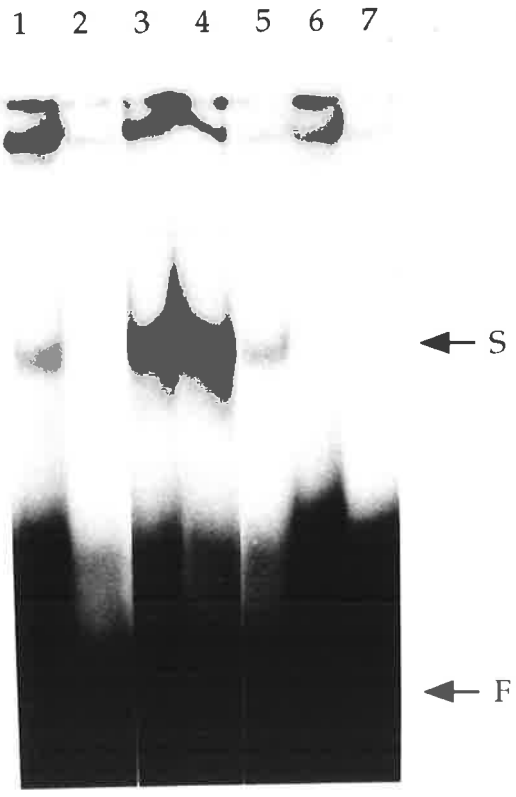
5; 0.3 M NaCl fraction discarded originally (i.e. no H1-F14 binding activity)

6 and 7; 0.4 M NaCl fractions

A



B



although biologically it is difficult to sustain this argument given the absolute conservation of the H1 box. However, sub-optimal (relaxed) binding conditions in the gel retardation incubation reaction, may prevent discrimination of a totally H1 box specific interaction. The same argument could be put to explain the similar DNA binding characteristics of H1-F40 (see 2.2.5).

The fact that the G/C or CCAAT competitors did not compete out the H1-F25 complex suggested that H1-F25 was not the same binding activity as H1-F14, which bound non-specifically to the 14 mer, G/C, and CCAAT probes (3.3.11 (iii)). H1-F40 binding activity also seems to be resistant to competition from the G/C and CCAAT box oligonucleotide competitors (K. Duncliffe; personal communication).

(iii) H1-F25 vs H1-F14

The difference between H1-F25 and H1-F14 (and similarity between H1-F25 and H1-F40) was demonstrated further by assaying the heparin-Sepharose chromatography fractions, saved from the bulk purification of H1-SF14 (3.3.3), for H1-F25 binding activity, by gel retardation analysis (Figure 4.2B). Minimal H1-F25 binding activity was found in the pooled 0.3 M NaCl heparin-Sepharose fraction containing H1-F14 activity. Most of the H1-F25 activity was found in the 0.2 M heparin-Sepharose fractions and the early 0.3 M NaCl fractions, all of which had been found not to contain H1-F14 activity (3.3.3). These same fractions had previously been found to contain H1-F40 binding activity (see 3.4). Although circumstantial, their overlapping elution profiles and similar DNA binding properties suggest that H1-F25 and H1-F40 represent the same DNA binding protein/activity.

(iv) Double stranded vs single stranded binding protein

To rule out the possibility that H1-F25 binding activity may have been due to a non-specific single stranded binding protein, two further sets of gel retardations were carried out. Firstly, labelled single stranded 25 mer probes (+ or - strand) were incubated with 9DCE nuclear extract and the reactions analysed by gel retardation

assay. Retarded complexes were formed with both single stranded probes (see Figure 4.3A), but these were clearly not H1-F25, and presumably involved non-specific single stranded binding protein(s). Secondly, labelled 25 mer (double stranded) probe was incubated with 9DCE nuclear extract and the specificity of the H1-F25 complex tested with the addition of unlabelled single stranded 25 mer (+ or - strand) competitor. Neither the + or - strand 25 mer competitors were able to compete out H1-F25 (Figure 4.3B). These two sets of results suggested that H1-F25 represented a genuine double stranded binding activity and was not the result of a single stranded binding protein.

From the results presented in 4.4.2 it appeared that H1-F25 possessd a degree of sequence specificity that should be investigated further.

4.2.3 Modification interference analysis of H1-F25 binding specificity

One major disadvantage of the gel retardation assay is that it does not provide detailed information on how a protein binds to the DNA. That is, the specific bases in a DNA sequence that the protein makes contact with and that are critical for binding. However there are a number of techniques that can provide this detailed information. These include DNA protection or footprinting assays and modification interference assays.

DNA protection assays (Galas and Schmitz, 1978) work on the basis that sequence-specific DNA-binding proteins will protect the nucleotides involved in their binding to DNA from digestion by endonucleases, or chemical modification and cleavage. Therefore, when protein extract is incubated with radioactively labelled DNA (containing the binding site for the protein), the DNA-protein complex subjected to partial endonuclease digestion or partial chemical modification and cleavage, and the products of the protection assay analysed on a denaturing polyacrylamide gel, the region involved with protein binding appears as a gap ('footprint') in the ladder of DNA fragments produced by the digestion or modification and cleavage. To be

Figure 4.3 Double stranded vs single stranded DNA binding protein

A Labelled single stranded 25 mer probe (+/coding strand or -/noncoding strand; 100 pg) was incubated with 9DCE nuclear extract (1 µl; 10 µg/µl) (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). Single stranded binding protein (SS) complexes in addition to free single strand (F(SS)) probe were detected by autoradiography and are indicated. For comparison, labelled 25 mer (double stranded) probe was incubated with 9DCE nuclear extract (1 µl; 10 µg/µl) and assayed as above (7.3.12 (i)). The candidate specific (S; H1-F25) complex and free (F) unbound probe are indicated. Lanes are as follows;

1 and 2; 25 mer probe alone, 25 mer probe and 9DCE nuclear extract

3 and 4; Single stranded 25 mer probe (+ strand) alone, single stranded 25 mer probe (+ strand) and 9DCE nuclear extract

5 and 6; Single stranded 25 mer probe (- strand) alone, single stranded 25 mer probe (- strand) and 9DCE nuclear extract

B Labelled 25 mer probe (100 pg) was incubated with 9DCE nuclear extract (1 µl; 10 µg/µl) with or without various competitor DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F25) complex formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 25 mer probe, while all lanes except lane 1 contain 9DCE nuclear extract. Lanes are as follows;

1; 25 mer probe alone

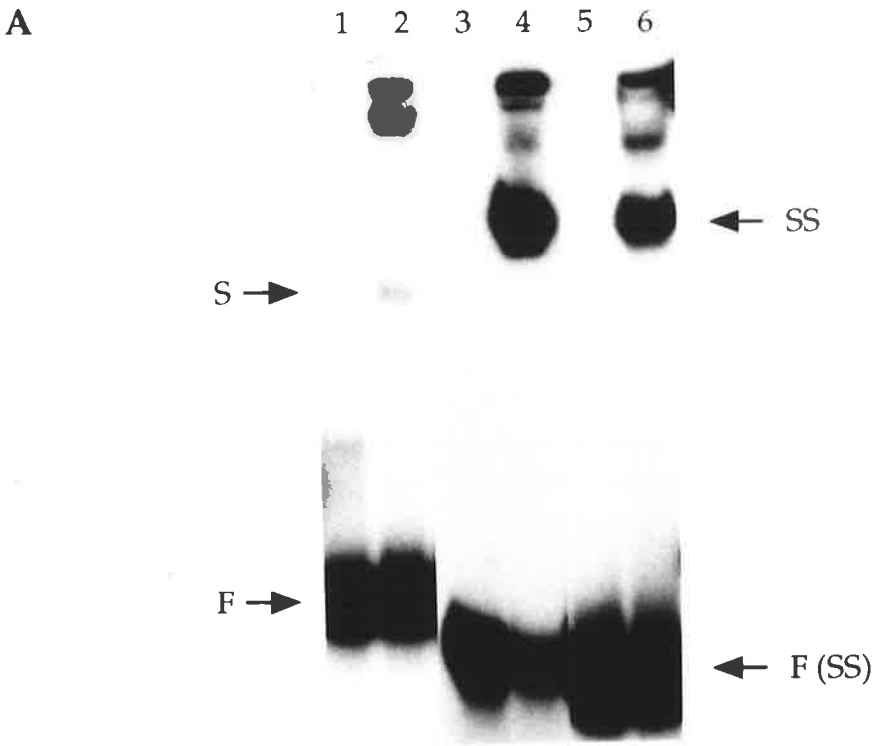
2; 100-fold excess of poly (dI-dC)

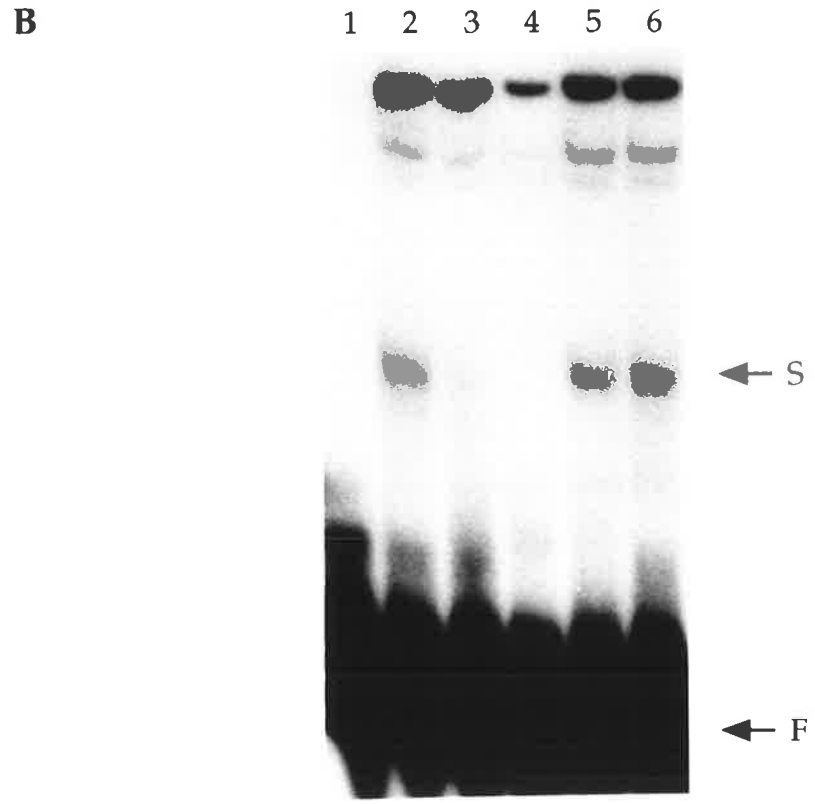
3; 20-fold excess of unlabelled 25 mer

4; 20-fold excess of unlabelled Δ25 mer

5; 20-fold excess of unlabelled single stranded 25 mer probe (+ strand)

6; 20-fold excess of unlabelled single stranded 25 mer probe (- strand)


A
1 2 3 4 5 6
SS
S
F
F (SS)


B
1 2 3 4 5 6
S
F

able to identify unambiguously the region of protection it is also necessary to include a control where the DNA is cleaved in the absence of protein. This is particularly important when DNase I (Galas and Schmitz, 1978) is used as the means of cleavage, because DNase I does not cleave completely at random and as a result produces some ambiguity when trying to determine the presence and nature of the protected region. DNase I, which cleaves single and double stranded DNA, is the cleavage agent most commonly used in protection analysis. Other cleavage methods include using methidiumpropyl-EDTA (Landolfi *et al.*, 1989) which cleaves randomly and dimethyl sulphate which has also been used to do *in vivo* protection analysis as dimethyl sulphate is able to freely and rapidly permeate cell membranes (Giniger *et al.*, 1985: Ephrussi *et al.*, 1985).

An alternative and relatively simple modification interference assay procedure was obtained from L. Coles (Institute of Medical and Veterinary Science, Adelaide; personal communication). This method was used instead of the DNA protection assay described above, to investigate H1-F25 binding to the H1 box and flanking regions.

Modification interference assays differ from protection assays as they involve partial chemical modification of the DNA prior to binding of the protein. Once the DNA has been modified it is then incubated with protein extract and the DNA-protein complexes separated from free DNA by gel retardation. Protein will only bind to DNA molecules which have not been chemically modified at the bases critical for protein binding. When the DNA from the DNA-protein complex is isolated, cleaved at the modified bases, and re-run on a denaturing polyacrylamide gel the DNA will show a region of protection from cleavage which represents where the protein is binding to the DNA. It is necessary to compare this cleavage pattern with modified and cleaved DNA that has not been selected for by its ability to bind the protein of interest, to precisely define the binding site. Reagents used for modification interference analysis include dimethyl sulphate (which methylates purines allowing

base cleavage of the DNA backbone at these positions with piperidine) and diethyl pyrocarbonate (which carboxyethylates purines; adenine modification is preferred over guanine, allowing preferential cleavage, again by piperidine, at these sites). The modification interference assay used in this study, involved using formic acid to modify purines by depurination and piperidine to then cleave the modified residues.

(i) Preparation of probe

The DNA fragment used as the probe for the modification interference assay was prepared as follows. Double stranded H1 box 25 mer was cloned into the *Sma*I site of pBluescript SK+ (7.3.3). Cloning of the 25 mer was confirmed by sequencing (7.3.21). The clone was linearised with *Hin*dIII and end-labelled with α-^{32}P-dATP & α-^{32}P-dCTP (7.3.12 (ii)). The clone was then cut with *Bam*HI releasing a 59 bp probe fragment, containing the H1 box 25 mer and flanking polylinker sequence, labelled at the 5' end of the coding strand. The 59 bp probe was gel purified (7.3.4 (ii)) and then partially depurinated using formic acid (7.3.12 (ii)). It was necessary to use a probe of at least this length, so that the protein interaction with the core 25 bp sequence could be determined accurately, even at the ends of the 25 mer sequence.

(ii) Modification interference assay

The Modified 59 bp probe was incubated with 9DCE nuclear extract, in standard gel retardation buffer, in the presence of unlabelled heterologous and homologous DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by non-denaturing polyacrylamide gel electrophoresis (7.3.12 (i)). A single retarded band was observed with the 59 bp probe (Figure 4.4A). This complex was shown to represent H1-F25 binding activity as adjudged by competitor studies (Figure 4.4A). Retarded DNA (H1-F25 binding activity) and also free (control) DNA were isolated from the gel. The isolated DNA samples were then cleaved at modified purine residues with piperidine and analysed on a 12% denaturing polyacrylamide gel (7.3.12 (ii)).

Figure 4.4 Modification interference analysis of H1-F25

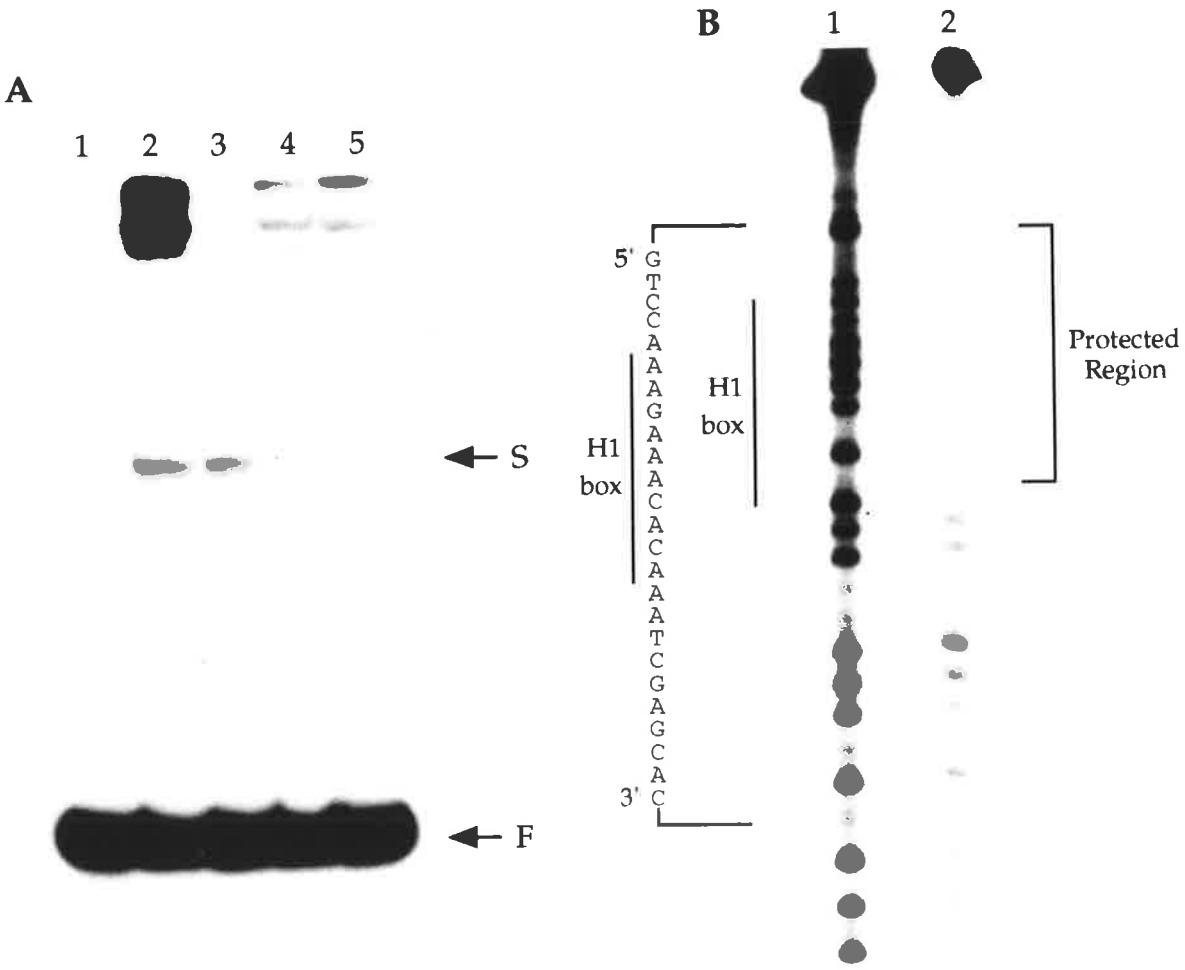
A Labelled 59 mer probe (100 pg), specially prepared (depurinated) for modification interference analysis (see 4.2.3 (i)), was incubated with 9DCE nuclear extract (1 μl; 10 μg/μl) with or without various competitor DNA (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 10% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; H1-F25) complex formed between nuclear extract components and labelled DNA, in addition to free (F) unbound probe were detected by autoradiography and are indicated. All lanes contain 59 mer probe, while all lanes except lane 1 contain 9DCE nuclear extract. Lanes are as follows;

1; 59 mer probe alone
2; 9DCE nuclear extract
3; 100-fold excess of poly (dI-dC)
4; 20-fold excess of unlabelled 25 mer
5; 20-fold excess of unlabelled Δ25 mer

B Modification interference analysis of H1-F25 binding specificity. Retarded H1-F25 complex (S) and free (control) DNA (F) were respectively isolated from lane 3 and lane 1 of gel retardation gel shown above. The DNA samples were then cleaved at modified purine residues with piperidine and analysed on a 12% denaturing polyacrylamide gel (7.3.12 (ii)). The cleavage ladders generated were detected by autoradiography. The protected region, H1 box and full sequence of the 25 mer oligonucleotide are indicated with respect to the ladders. Lanes are as follows;

1; Free (control) DNA
2; H1-F25 retarded complex DNA

Note the signal intensity between the free (control) DNA and that isolated from the retarded complex are not standardised as well as they could have been. However the protection pattern generated was completely reproducible; two further modification interference experiments identifying the same protected region.


A
1
2
3
4
5
S
F
B
1
2
5'
G
T
C
C
A
A
A
G
A
A
A
C
A
C
A
A
A
T
C
G
A
G
C
A
C
3'
H1
box
H1
box
Protected
Region

Comparing the DNA ladders obtained from the retarded complex and the free DNA, a region of protection was evident in the DNA ladder from the retarded complex (Figure 4.4B). This region covered the H1 box and imediate 5' flanking sequence. Thus it would appear that sequence outside the H1 box is also involved in H1-F25 binding. This is consistent with the observation that the Δ25 mer probe is able to bind H1-F25. The Δ25 mer probe contains the same 5' flanking region, and if as suggested (see 4.2.2 (ii)), H1-F25 binding activity *in vitro* is only slightly compromised by the mutation in the H1 box, this common flanking region may be sufficient under these conditions to maintain H1-F25 binding.

4.3 OLIGOSCREENING

A modified oligoscreening method developed by Kalionis and O'Farrell (B. Kalionis, Department of Biochemistry, University of Adelaide; personal communication) became available at the time H1-F25 was being characterised (note, method recently submitted to EMBO J., 1993). This modified method was similar to the oligoscreening protocol of Vinson *et al.* (1988), but included some important changes (7.3.15 (iii)), which had been shown to dramatically increase the chances of detecting positive clones (B. Kalionis; personal communication). These changes included plating phage from the library at a much lower density for the primary screening. This helped to ensure that production of fusion protein by phage was not inhibited by having plaques too close to each other. Eliminating antibiotic selection allowed larger plaques to form and produce more fusion protein compared to selection protocols. By incubating the plates with the second filter for 12 hours instead of 2 hours roughly equivalent signals from both the first and second lifts was achieved thereby increasing the effectiveness of the duplicate lifts. The use of a lower stringency washing step and the inclusion of Triton X-100 in the wash buffers to minimise background and non-specific interactions were also important in this protocol.

These changes were indeed important, and had proved successful where the Vinson *et al.* (1988) method had failed (B Kalionis; personal communication). In this light it seemed reasonable to use the method to screen a ten day chicken embryo cDNA λgt11expression library (Clontech library, catalogue number CL1001b; 1.06×10^6 independent clones) for H1-F25 candidate clones. An *engrailed* homeobox binding protein λgt11 clone (Desplan *et al.*, 1988) was obtained from B. Kalionis and used in this study as a positive control.

4.3.1 Probe preparation

Concatenated probes for both *engrailed* and H1-F25 oligoscreening were generated by kinasing the *engrailed* (see 7.2.8 for sequence of the oligonucleotide used to form the double stranded *engrailed* probe) and 25 mer oligonucleotides respectively with γ^{32}P-ATP, annealing, and ligating to form multimers of each of 5 - 10 copies. The procedure is described in more detail in 7.3.15 (i).

4.3.2 *Engrailed* control oligoscreening

The *engrailed* λgt11 clone was spiked into an aliquot of ten day chicken embryo cDNA λgt11 library, absorbed to *E.coli* strain Y1090, and plated onto a 150 mm L-agar plate so as to give approximately 50 *engrailed* plaques in a background of approximately 12,500 λgt11 library plaques per plate. After 4 hours of growth at 42°C, fusion proteins were induced by the addition of IPTG soaked filters, and incubated for an additional 6 hours (first protein lifts), or 12 hours (second protein lifts). Filters were air-dried at room temperature for 15 minutes before denaturation/renaturation processing. Filters were analysed with a concatenated *engrailed* probe. Duplicate *engrailed* positives were detected by autoradiography on the filters lifted from the *engrailed* spiked plate, suggesting that the oligoscreening technique was working satisfactorily (Figure 4.5).

Figure 4.5 ***Engrailed*** **control oligoscreening**

Shown are the results from the *engrailed* control oligoscreening. Oligoscreening was carried out according to the method described by Kalionis and O'Farrell (1993) (see 4.3.2 and 7.3.15 (iii)). Filters were probed with concatenated *engrailed* probe (see 4.3.1). Filters are as follows;

1 and 2; First and second duplicate filter lifts from a control plate containing approximately 12,500 ten day chicken embryo cDNA λgt11 library plaques spiked with approximately 50 engrailed λgt11 clone plaques.

1

2

4.3.3 H1-F25 oligoscreening

The ten day chicken embryo cDNA λgt11 library was chosen because, as described in Section 4.2.2, crude nuclear extract from nine day chicken embryos was shown to contain H1-F25 binding activity. Furthermore, to follow the sucessful method as closely as possible, oligoscreening for H1-F25 was carried out using the oligoscreening buffer recommended by Kalionis and O'Farrell (1993). This was done after it was demonstrated, by gel retardation analysis, that H1-F25 was able to efficiently bind the 25 mer in the oligoscreening buffer (data not shown).

H1-F25 oligoscreening was carried out as described for the *engrailed* control, except that the ten day chicken embryo cDNA λgt11 cDNA library was absorbed to *E.coli* strain Y1090, and plated onto 15 cm L-agar plates with approximately 12,500 plaques per plate and filters analysed with concatenated 25 mer probe. Additionally, in case H1-F25 binding activity was not recoverable after denturation/renaturation treatment, oligoscreening was performed omitting the denaturation/renaturation step. In this case the filters were blocked directly after they had been air dried (7.3.15 (iii)). Approximately 5×10^5 plaques, corresponding to roughly half of the ten day chicken embryo cDNA λgt11 library, were screened. However, no positive signals were detected with this 'native' oligoscreening. Therefore the filters were rescreened, but this time with the denaturation/renaturation processing step included by simply repeating the cycle of denaturation/renaturation and then following the normal protocol. A total of seven first round duplicate positives were identified with this repeat screening. Positive plaques were picked into 1ml PSB, the phage allowed to elute overnight at 4°C, and subjected to a second round of screening. Of the seven first round positives, only two were detected in the second round screening. These clones designated λ4.1 and λ8.1 were purified to homogeneity by a third round of screening. The λ4.1 and λ8.1 first round duplicate positives are shown in Figures 4.6A and 4.6B respectively.

Figure 4.6 H1-F25 oligoscreening first round duplicate positives
Shown are the first round duplicate positives, λ4.1 and λ8.1, identified as a result of oligoscreening a ten day chicken embryo cDNA λgt11 expression library with concatenated 25 mer probe. Oligoscreening was carried out according to the method described by Kalionis and O'Farrell (1993) (see 4.3.3 and 7.3.15 (iii)). Filters are as follows;

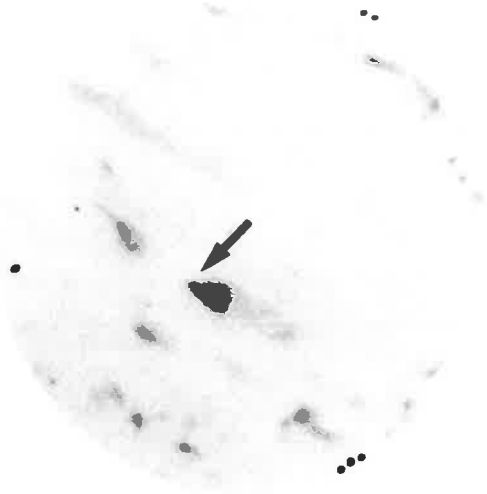
A 1 and 2; First and second duplicate filter lifts with the autoradiographic signal representing the λ4.1 plaque indicated on each filter by an arrow.

B 1 and 2; First and second duplicate filter lifts with the autoradiographic signal representing the λ8.1 plaque indicated on each filter by an arrow.

A

1

2



B

1

2

The DNA binding specificities of the purified λ4.1 and λ8.1 clones were tested by oligoscreening with the concatenated 25 mer probe and three additional probes: concatenated Δ25 mer and *engrailed* binding site probes and a single copy H5 probe (7.3.15 (iii)). Both the λ4.1 and λ8.1 clones hybridised with the 25 mer probe (Figures 4.7A and 4.7B). However the clones reacted differently with the Δ25 mer probe: λ4.1 binding the Δ25 mer with equal or slightly greater affinity than the 25 mer, and λ8.1 interacting only very weakly with the Δ25 mer probe. Neither clone hybridised with the H5 or engrailed negative control probes. These results suggest that the clones were specific DNA binding proteins and may indeed be related to the H1-F25 binding activity identified in 9DCE nuclear extract (4.2.2). The λ4.1 clone possessed similar binding characteristics to H1-F25, while the λ8.1 clone appeared to bind the 25 mer with an increased specificity not seen with H1-F25 in the gel retardation assay.

4.4 ANALYSIS OF λ4.1 AND λ8.1 CLONES

λ DNA was prepared from both λ4.1 and λ8.1 clones (7.3.17). Digestion with *Eco*RI revealed that the λ4.1 and λ8.1 inserts were approximately 2.1 kb and 1.4 kb in size repectively. To allow the orientation of the inserts to be determined, DNA from both clones was further digested with a number of restriction enzymes. A single *Sac*I site was identified, located approximately 1.4 kb from the 5' end of the λ4.1 clone, and two *Hin*dIII sites were identified, located approximately 150 bp and 1.0 kb from the 5' end of the λ8.1 clone.

4.4.1 Subcloning of λ4.1 and λ8.1 into pBSSK+

(i) Restriction endonuclease mapping

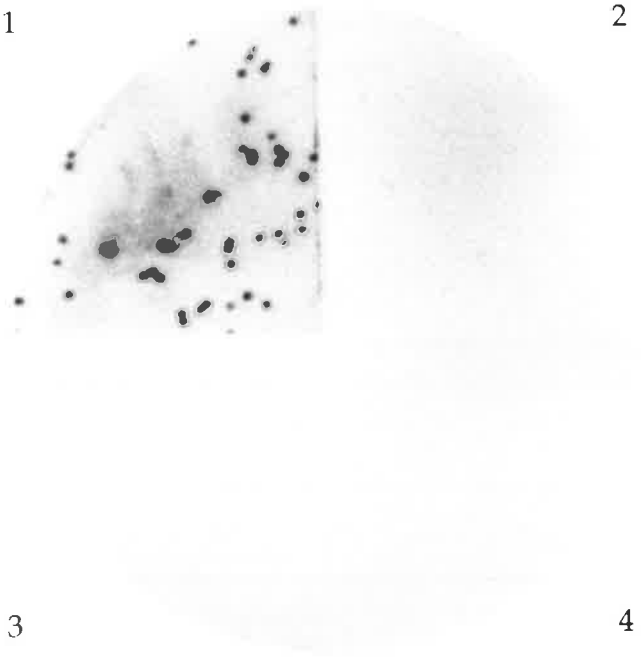
To make further manipulations of the λ4.1 and λ8.1 clones easier, DNA from each was digested with *Eco*RI, and ligated into *Eco*RI digested pBluescript SK+. The ligation reactions were transformed into *E.coli* strain DH5α, and colonies containing recombinants identified by colour selection (7.3.3). Plasmid DNA was prepared from recombinants as described in section 7.3.1 (ii). The DNA samples were digested

Figure 4.7 Oligoscreening analysis of λ4.1 and λ8.1 DNA binding specificities
The DNA binding specificities of the purified λ4.1 and λ8.1 plaques were tested by oligoscreening with the concatenated 25 mer probe and three additional probes; concatenated Δ25 mer and *engrailed* binding site probes and a single copy H5 probe (see 4.3.3). Oligoscreening was carried out according to the method described by Kalionis and O'Farrell (1993) (7.3.15 (iii)). Filters are as follows;

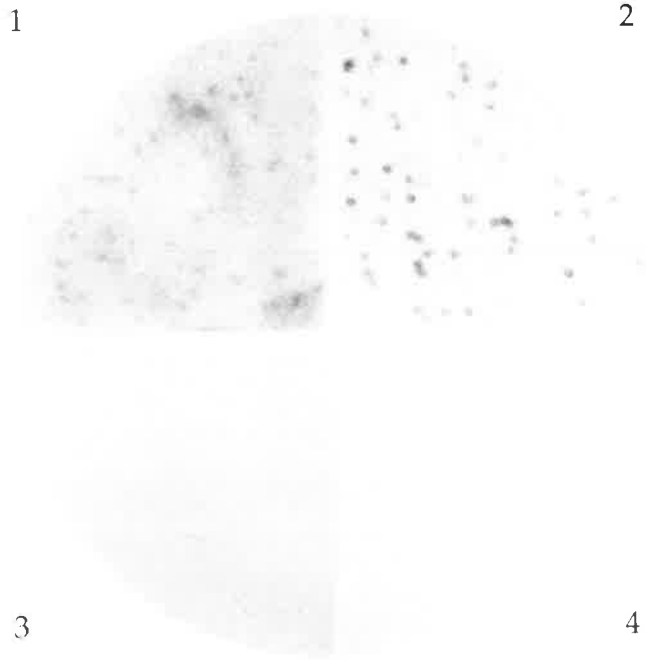
A λ8.1 filter; single protein lift from a plate containing approximately 200 λ8.1 plaques. The filter has been divided up into quarters and each quarter probed with a different probe.
1; 25 mer probe
2; Δ25 mer probe
3; *Engrailed* probe
4; H5 probe

B λ4.1 filter; single protein lift from a plate containing approximately 200 λ4.1 plaques. The filter has been divided up into quarters and each quarter probed with a different probe.
1; 25 mer probe
2; Δ25 mer probe
3; *Engrailed* probe
4; H5 probe

A



B



with the appropriate restriction enzymes (*Eco*RI to confirm the presence/size of the inserts, *Sal*I or *Hin*dIII to determine the orientation of the λ4.1 or λ8.1 inserts respectively) and analysed by electrophoresis on an agarose gel (7.3.2). Plasmids containing the λ4.1 or λ8.1 inserts in the two different orientations were identified and large scale plasmid preparations of each produced (7.3.1 (i)). The two orientations of the λ4.1 insert in pBluescript were called pBS4.11 and pBS4.12, while the two orientations of the λ8.1 insert in pBluescript were called pBS8.11 and pBS8.12.

A series of restriction enzyme digestions of the pBS4.11, pBS4.12, pBS8.11, and pBS8.12 constructs were carried out (7.3.2 (i)) and from analysis of these, restriction maps were generated for the λ4.1 and λ8.1 inserts (Figures 4.8A and 4.8B).

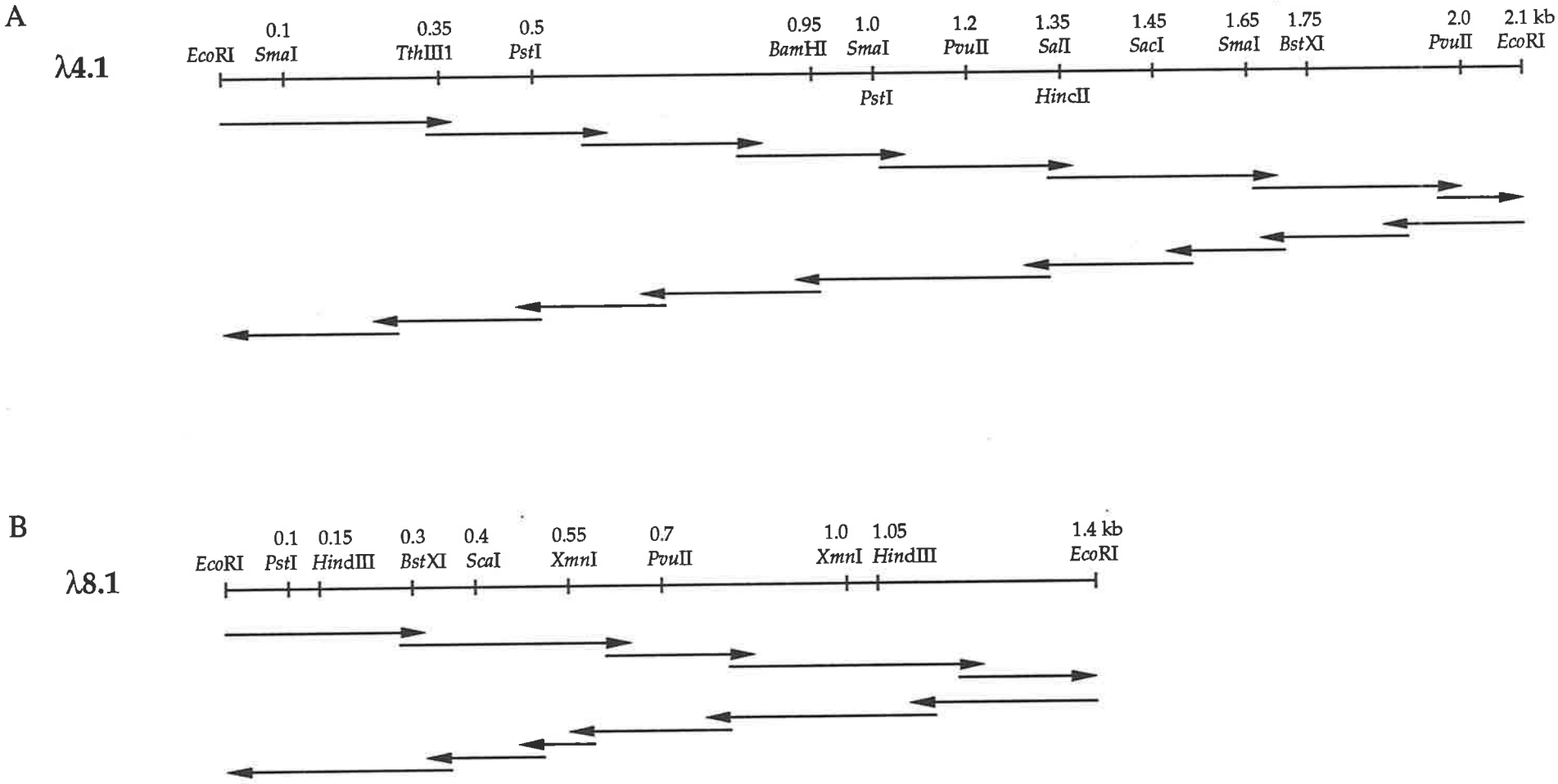
(ii) Dideoxy sequencing

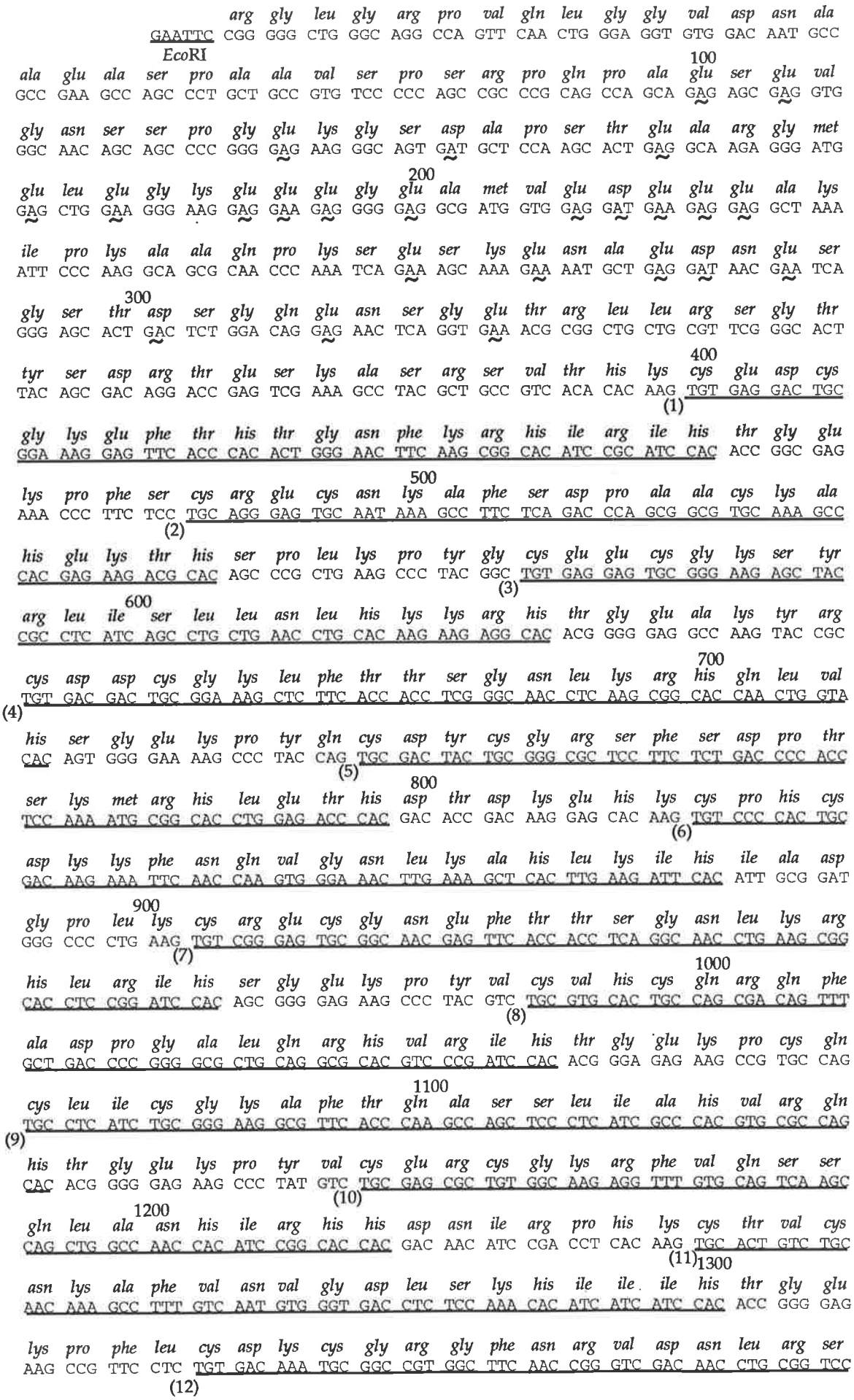
A Series of deletion subclones of pBS4.11, pBS4.12, pBS8.11, and pBS8.12 were made using the Promega Erase-a-base system (7.3.20). The pBS4.11 and pBS4.12 clones were digested with *Kpn*I and *Hin*dIII, while the pBS8.11 and pBS8.12 clones were digested with *Kpn*I and *Xho*I. These digestions linearised each of the clones in the polylinker and left the vector protected from deletion but the inserts open to exonuclease III action. A Series of exonuclease III deletion subclones were then generated according to the manafacturers instructions and transformed into *E.coli* strain DH5α. Sets of overlapping subclones for each of the four clones were size selected after electrophoresis on agarose gels. The four sets of deletion subclones were sequenced (Sanger *et al.*, 1977; 7.3.21). This enabled the entire sequence of both strands of the λ4.1 and λ8.1 inserts to be determined. The sequencing strategies are shown in (Figures 4.8A and 4.8B), and the sequences in (Figures 4.9 and 4.10). The reading frames determined from the original λgt11 clones, and the deduced amino acid sequences are also indicated.

Figure 4.8 Dideoxy sequencing of λ4.1 and λ8.1

A The *Eco*RI insert from λ4.1 was cloned in both orientations into *Eco*RI digested pBluescript SK+ to produce constructs designated pBS4.11 and pBS4.12 (see 4.4.1 (i)). A series of restriction enzyme digestions of the pBS4.11 and pBS4.12 constructs were carried out and from these a restriction map for the λ4.1 insert was generated. The positions of internal restriction sites are indicated, and numbered as measured from the 5' end of the insert. Deletion subclones from both pBS4.11 and pBS4.12 were created in both directions using the Erase-a-base system (see 4.4.1 (ii)). The entire cDNA clone was sequenced in both directions, as indicated by the arrows.

B The *Eco*RI insert from λ8.1 was cloned in both orientations into *Eco*RI digested pBluescript SK+ to produce constructs designated pBS8.11 and pBS8.12 (see 4.4.1 (i)). A series of restriction enzyme digestions of the pBS8.11 and pBS8.12 constructs were carried out and from these a restriction map for the λ8.1 insert was generated. The positions of internal restriction sites are indicated, and numbered as measured from the 5' end of the insert. Deletion subclones from both pBS8.11 and pBS8.12 were created in both directions using the Erase-a-base system (see 4.4.1 (ii)). The entire cDNA clone was sequenced in both directions, as indicated by the arrows.


A
λ4.1
0.1
EcoRI SmaI
0.35
TthIII1
0.5
PstI
0.95
BamHI
1.0
SmaI
PstI
1.2
PvuII
1.35
SalI
HincII
1.45
SacI
1.65
SmaI
1.75
BstXI
2.0
PvuII
2.1 kb
EcoRI
B
λ8.1
0.1
EcoRI PstI
0.15
HindIII
0.3
BstXI
0.4
ScaI
0.55
XmnI
0.7
PvuII
1.0
XmnI
1.05
HindIII
1.4 kb
EcoRI


arg gly leu gly arg pro val gln leu gly gly val asp asn ala
GAATTC CGG GGG CTG GGC AGG CCA GTT CAA CTG GGA GGT GTG GAC AAT GCC
EcoRI
ala glu ala ser pro ala ala val ser pro ser arg pro gln pro ala glu ser glu val
GCC GAA GCC AGC CCT GCT GCC GTG TCC CCC AGC CGC CCG CAG CCA GCA GAG AGC GAG GTG
100
gly asn ser ser pro gly glu lys gly ser asp ala pro ser thr glu ala arg gly met
GGC AAC AGC AGC CCC GGG GAG AAG GGC AGT GAT GCT CCA AGC ACT GAG GCA AGA GGG ATG
glu leu glu gly lys glu glu glu gly glu ala met val glu asp glu glu glu ala lys
GAG CTG GAA GGG AAG GAG GAA GAG GGG GAG GCG ATG GTG GAG GAT GAA GAG GAG GCT AAA
200
ile pro lys ala ala gln pro lys ser glu ser lys glu asn ala glu asp asn glu ser
ATT CCC AAG GCA GCG CAA CCC AAA TCA GAA AGC AAA GAA AAT GCT GAG GAT AAC GAA TCA
gly ser thr asp ser gly gln glu asn ser gly glu thr arg leu leu arg ser gly thr
GGG AGC ACT GAC TCT GGA CAG GAG AAC TCA GGT GAA ACG CGG CTG CTG CGT TCG GGC ACT
300
tyr ser asp arg thr glu ser lys ala ser arg ser val thr his lys cys glu asp cys
TAC AGC GAC AGG ACC GAG TCG AAA GCC TAC GCT GCC GTC ACA CAC AAG TGT GAG GAC TGC
400
(1)
gly lys glu phe thr his thr gly asn phe lys arg his ile arg ile his thr gly glu
GGA AAG GAG TTC ACC CAC ACT GGG AAC TTC AAG CGG CAC ATC CGC ATC CAC ACC GGC GAG
lys pro phe ser cys arg glu cys asn lys ala phe ser asp pro ala ala cys lys ala
AAA CCC TTC TCC TGC AGG GAG TGC AAT AAA GCC TTC TCA GAC CCA GCG GCG TGC AAA GCC
500
(2)
his glu lys thr his ser pro leu lys pro tyr gly cys glu glu cys gly lys ser tyr
CAC GAG AAG ACG CAC AGC CCG CTG AAG CCC TAC GGC TGT GAG GAG TGC GGG AAG AGC TAC
(3)
arg leu ile ser leu leu asn leu his lys lys arg his thr gly glu ala lys tyr arg
CGC CTC ATC AGC CTG CTG AAC CTG CAC AAG AAG AGG CAC ACG GGG GAG GCC AAG TAC CGC
600
cys asp asp cys gly lys leu phe thr thr ser gly asn leu lys arg his gln leu val
TGT GAC GAC TGC GGA AAG CTC TTC ACC ACC TCG GGC AAC CTC AAG CGG CAC CAA CTG GTA
700
(4)
his ser gly glu lys pro tyr gln cys asp tyr cys gly arg ser phe ser asp pro thr
CAC AGT GGG GAA AAG CCC TAC CAG TGC GAC TAC TGC GGG CGC TCC TTC TCT GAC CCC ACC
(5)
ser lys met arg his leu glu thr his asp thr asp lys glu his lys cys pro his cys
TCC AAA ATG CGG CAC CTG GAG ACC CAC GAC ACC GAC AAG GAG CAC AAG TGT CCC CAC TGC
800
(6)
asp lys lys phe asn gln val gly asn leu lys ala his leu lys ile his ile ala asp
GAC AAG AAA TTC AAC CAA GTG GGA AAC TTG AAA GCT CAC TTG AAG ATT CAC ATT GCG GAT
gly pro leu lys cys arg glu cys gly asn glu phe thr thr ser gly asn leu lys arg
GGG CCC CTG AAG TGT CGG GAG TGC GGC AAC GAG TTC ACC ACC TCA GGC AAC CTG AAG CGG
900
(7)
his leu arg ile his ser gly glu lys pro tyr val cys val his cys gln arg gln phe
CAC CTC CGG ATC CAC AGC GGG GAG AAG CCC TAC GTC TGC GTG CAC TGC CAG CGA CAG TTT
1000
(8)
ala asp pro gly ala leu gln arg his val arg ile his thr gly glu lys pro cys gln
GCT GAC CCC GGG GCG CTG CAG GCG CAC GTC CGG ATC CAC ACG GGA GAG AAG CCG TGC CAG
cys leu ile cys gly lys ala phe thr gln ala ser ser leu ile ala his val arg gln
TGC CTC ATC TGC GGG AAG GCG TTC ACC CAA GCC AGC TCC CTC ATC GCC CAC GTG CGC CAG
1100
(9)
his thr gly glu lys pro tyr val cys glu arg cys gly lys arg phe val gln ser ser
CAC ACG GGG GAG AAG CCC TAT GTC TGC GAG CGC TGT GGC AAG AGG TTT GTG CAG TCA AGC
(10)
gln leu ala asn his ile arg his his asp asn ile arg pro his lys cys thr val cys
CAG CTG GCC AAC CAC ATC CGG CAC CAC GAC AAC ATC CGA CCT CAC AAG TGC ACT GTC TGC
1200
(11)
asn lys ala phe val asn val gly asp leu ser lys his ile ile ile his thr gly glu
AAC AAA GCC TTT GTC AAT GTG GGT GAC CTC TCC AAA CAC ATC ATC ATC CAC ACC GGG GAG
1300
lys pro phe leu cys asp lys cys gly arg gly phe asn arg val asp asn leu arg ser
AAG CCG TTC CTC TGT GAC AAA TGC GGC CGT GGC TTC AAC CGG GTC GAC AAC CTG CGG TCC
(12)

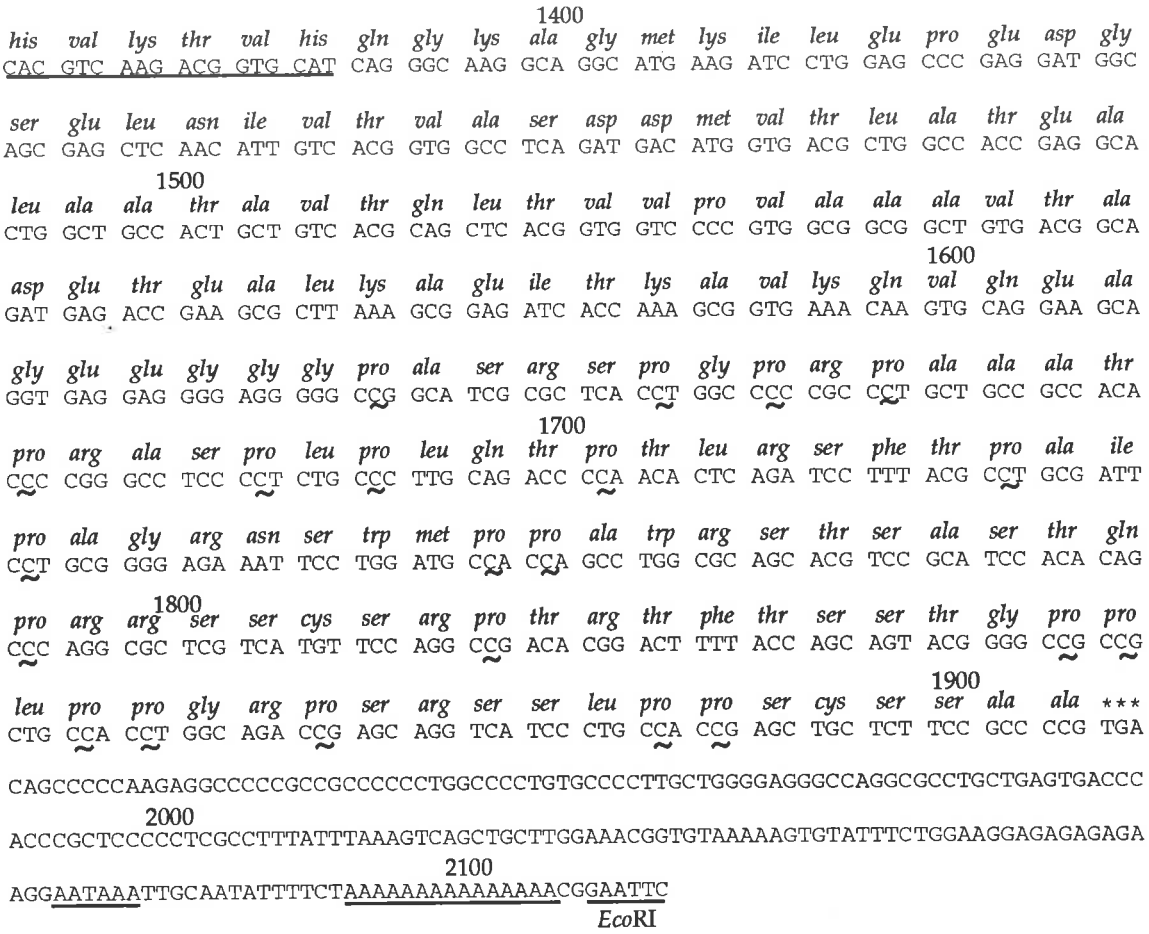


Figure 4.9 Nucleotide sequence of λ4.1

The complete nucleotide sequence of the insert from the λ4.1 clone is shown along with the deduced amino acid sequence. The clone codes for an open reading frame of 1,908 bp but appears not to contain the full 5' coding region. The poly(A) tract and polyadenylation signal (AAUAAA) sequences are underlined in the 3' noncoding region. Zinc fingers are underlined and numbered (1) - (12). Approximately 50 bp 5' of the zinc fingers, a cluster of acidic residues, constituting a possible acidic activation domain, are underlined with ~ symbols. While immediately upstream of the termination codon, a cluster of proline residues, constituting a possible proline-rich activation domain, are also underlined with ~ symbols.

Figure 4.10 Nucleotide sequence of λ8.1

The complete nucleotide sequence of the insert from the λ8.1 clone is shown along with the deduced amino acid sequence. The clone codes for an open reading frame of 1,299 bp but appears not to contain the full 5' coding region and 3' noncoding region. Zinc fingers are underlined and numbered (1) - (5). Approximately 300 bp 5' of the zinc fingers, a cluster of acidic residues, constituting a possible acidic activation domain, are underlined with ~ symbols.

```
                          arg lys thr val phe leu lys thr leu asn glu gln
                   GAATTC CGC AAA ACT GTG TTC TTG AAA ACA TTA AAT GAA CAA
                   EcoRI
                                                                       100
arg leu glu gly glu phe cys asp ile ala ile val val glu asp val lys phe arg ala
CGT TTG GAA GGA GAA TTT TGT GAC ATA GCT ATC GTG GTT GAA GAT GTT AAA TTC CGA GCA

his arg cys val leu ala ala cys ser thr tyr phe lys lys leu phe lys lys leu glu
CAT AGG TGT GTG CTT GCT GCC TGC AGT ACC TAC TTC AAA AAG CTT TTC AAA AAA CTA GAA
                                       200
val asp ser ser ser val ile glu ile asp phe leu arg ser asp ile phe glu glu val
GTC GAT AGT TCT TCA GTA ATA GAA ATA GAT TTT CTC CGT TCT GAT ATT TTT GAG GAA GTT

leu asn . tyr met tyr thr ala lys ile ser val lys lys glu asp val asn leu met met
CTT AAT TAC ATG TAC ACC GCA AAA ATT TCT GTT AAG AAG GAA GAT GTA AAT TTG ATG ATG
                   300
ser ser gly gln ile leu gly ile arg phe leu asp lys leu cys ser gln lys arg asp
TCT TCA GGC CAG ATT CTT GGT ATT CGT TTT CTG GAT AAA CTC TGC TCT CAA AAA CGT GAT
                                                                       400
val ser ser pro glu glu asn thr gln ser lys ser lys tyr cys leu lys ile asn arg
GTA TCT AGT CCT GAA GAG AAC ACA CAG TCC AAG AGC AAG TAC TGT CTG AAA ATA AAC CGT

pro ile gly glu pro asn asp thr gln asp asp glu val glu glu ile gly asp his asp
CCT ATT GGG GAA CCT AAC GAT ACC CAG GAT GAT GAG GTA GAA GAG ATT GGA GAT CAC GAT
                                       500
asp ser pro ser asp val thr val glu gly thr pro pro ser gln glu asp gly lys ser
GAC AGT CCA TCA GAT GTG ACA GTG GAA GGA ACT CCC CCC AGT CAG GAA GAT GGG AAA TCA

pro thr thr thr leu arg val gln glu ala ile leu lys glu leu gly ser glu glu val
CCA ACC ACT ACC CTG AGA GTT CAG GAG GCG ATT CTG AAG GAG CTG GGA AGC GAA GAG GTT
                   600
arg lys val asn cys tyr gly gln glu val glu ser met glu thr thr glu ser lys asp
CGA AAG GTA AAC TGC TAT GGC CAA GAA GTA GAG TCT ATG GAA ACA ACG GAA TCG AAA GAC
                                                                       700
leu gly ser gln thr pro gln ala leu thr phe asn asp gly ile ser glu val lys asp
TTA GGA TCT CAA ACA CCT CAG GCT TTG ACG TTT AAT GAT GGC ATA AGT GAA GTG AAA GAT

glu gln thr pro gly trp thr thr ala ala gly asp met lys phe glu tyr leu leu tyr
GAA CAG ACA CCA GGC TGG ACA ACA GCA GCT GGG GAT ATG AAG TTT GAA TAT TTG CTT TAT
                                       800
gly his arg glu his ile val cys gln ala cys gly lys thr phe ser asp glu ala arg
GGT CAT AGG GAA CAC ATT GTA TGT CAG GCT TGT GGT AAG ACC TTT TCT GAT GAA GCA CGA
                           (1)
leu arg lys his glu lys leu his thr ala asp arg pro phe val cys glu met cys thr
CTG AGA AAA CAT GAA AAG CTA CAC ACT GCT GAT AGA CCA TTT GTT TGT GAA ATG TGT ACA
                   900                                  (2)
lys gly phe thr thr gln ala his leu lys glu his leu lys ile his thr gly tyr lys
AAG GGC TTT ACC ACA CAA GCT CAT TTG AAA GAG CAT CTG AAA ATA CAC ACA GGT TAC AAG
                                                                       1000
pro tyr ser cys glu val cys gly lys ser phe ile arg ala pro asp leu lys lys his
CCT TAC AGT TGC GAG GTA TGT GGA AAG TCT TTT ATT CGT GCA CCG GAC TTA AAA AAG CAT
           (3)
glu arg val his ser asn glu arg pro phe ala cys his met cys asp lys ala phe lys
GAA AGA GTT CAC AGT AAT GAG AGG CCA TTT GCA TGC CAC ATG TGC GAT AAA GCT TTC AAG
                                           (4)  1100
his lys ser his leu lys asp his glu arg arg his arg gly glu lys pro phe val cys
CAC AAG TCC CAC CTC AAA GAC CAT GAA AGA AGA CAC CGG GGA GAG AAA CCT TTT GTC TGC
                                                                        (5)
gly ser cys thr lys ala phe ala lys ala ser asp leu lys arg his glu asn asn met
GGT TCC TGC ACT AAA GCA TTT GCT AAA GCA TCG GAT CTA AAA AGG CAT GAG AAC AAT ATG
                   1200
his ser glu arg lys gln val thr thr ala asn ser ile gln ser glu thr glu gln leu
CAC AGT GAA AGA AAA CAA GTT ACT ACA GCC AAT TCC ATC CAG AGT GAA ACA GAA CAG TTA
                                                                       1300
gln ala ala ala met ala leu glu ala glu gln gln leu glu thr ile ala cys ser ***
CAG GCA GCA GCT ATG GCC TTG GAA GCA GAG CAG CAA TTG GAA ACT ATA GCT TGT AGT TAA

AAAAAAAAAAAAAAAAAAGAAAAAAATTCAAAGGACAGAAATTTACAGTCTAATTTACAGTCTTTGGCCTATCAAAGA
                1400
ACATGCAATTCAGTGCTGAACAAGGTAACTTAGCATACCGGAATTC
                                        EcoRI
```

4.4.2 Sequence analysis of λ4.1

The sequence of λ4.1 shows an open reading frame of 1,908 bp that contains a 15 residue poly(A) tract and a polyadenylation signal (AAUAAA) in the 3' noncoding region. The lack of an ATG codon with a consensus Kozak ribosome binding site (Kozak, 1987) in the 5' region of the clone suggests that the clone does not contain the full 5' coding sequence. The nucleotide and deduced amino acid sequences for the clone were compared to the GENBANK nucleic acid database (June 1991), EMBL nucleic acid database (June 1991), GENPEPTIDE protein database (June 1991) and SWISSPROT protein database (June 1991 and most recently July 1992) using the FASTA, MAIL-BLAST and MAIL-DAP search programs (Devereux *et al.*, 1984). At the amino acid level one region of significant homology was identified which included a block of 12 Cys_2-His_2 type zinc fingers (see 1.6.1 (i) for general description of the zinc finger motif). However outside of this region no significant homologies were detected, indicating that this cDNA clone (designated Z12 for future reference) encoded a novel zinc finger DNA binding protein. Further inspection of the Z12 amino acid sequence resulted in the identification of a region outside, and immediately 5', of the zinc fingers which appeared to be unusually rich (~ 30%) in acidic amino acids. This region may represent an acidic activation domain similar to those found in the yeast activator proteins GCN4 and GAL4 and the HSV protein VP16 (see 1.6.2 (i)). Acidic activation domains often involve an amphipathic α-helix with all the negatively charged residues displayed along one surface of the helix. If the Z12 acidic rich domain does involve α-helical structure, then the acidic residues spaced approximately three to four residues apart in places in this domain would be consistent with such a periodic arrangement. An additional region immediately upstream of the termination codon in Z12 was also identifed that was rich (~ 25%) in proline residues. This region may represent a proline-rich activation domain similar to those found in the CTF/NF1, AP-2, Oct2 and SRF transcription factors (see 1.6.2 (ii)).

4.4.3 Sequence analysis of λ8.1

The sequence of λ8.1 shows an open reading frame of 1,299 bp that is incomplete at the 3' end. Poly(A) tract or polyadenylation signal sequences are not evident in the putative 3' noncoding region of the clone suggesting that the cloning procedure used during the cDNA library construction has truncated the 3' end of the mRNA. The lack of an ATG codon with a consensus Kozak ribosome binding site (Kozak, 1987) in the 5' region of the clone also suggests that this clone does not contain the full 5' coding sequence. The nucleotide, and deduced amino acid sequences for the clone were compared to the same databases as for Z12 (see above). At the amino acid level one region of significant homology was identified which included a block of 5 Cys_2-His_2 type zinc fingers. However outside of the zinc fingers no significant homologies were detected, indicating that this cDNA clone (designated Z5 for future reference) encoded a novel zinc finger DNA binding protein. As with Z12, further inspection of the Z5 amino acid sequence resulted in the identification of a region outside and 5' of the zinc fingers which also appeared to be unusually rich (~ 40%) in acidic amino acids and likewise may represent an acidic activation domain.

A more detailed discussion of the structural features of both the Z12 and Z5 clones follows in 4.5.

4.4.4 Expression of Z5 and Z12 proteins

Expression of the Z5 and Z12 proteins in sufficient quantities was desirable, firstly to be able to determine whether or not the binding properties of Z5 and Z12 toward single copy 25 mer probe, in the gel retardation assay, were similar to those for H1-F25 and secondly to be able to produce antibodies to the Z5 and Z12 proteins which would aid further biological characterisation of Z5 and Z12.

UNIVERSITY OF ADELAIDE LIBRARY

(i) Production of λ4.1 and λ8.1 lysogens

Since the λ4.1 and λ8.1 clones were λgt11 derivatives attempts were made to make bacteria lysogenic for the recombinant phage, and use IPTG to induce production of the β-galactosidase fusion proteins.

E.coli strain Y1089 was lysogenised with either λ4.1 or λ8.1 clones (7.3.22 (i)). The lysogenic strains were grown to OD_{600} 0.5, before induction of the β-galactosidase fusion protein by heat shock at 42°C for 20 minutes, followed by addition of IPTG to 10 mM, and incubation at 37°C for a further 1-2 hours. Cells were harvested by centrifugation, and lysed by heating to 100°C for 5 minutes in lysis buffer (7.3.22 (ii)). Total cell lysates were analysed on an SDS-polyacrylamide gel (7.3.13 (i)), proteins visualised by silver staining (7.3.13 (iii)), and compared with extracts prepared from uninduced λ4.1 and λ8.1 lysogens, and uninduced and induced wild type λgt11 lysogen. The levels of the fusion proteins produced were very low, with bands corresponding to the fusion proteins only just visible on an SDS-polyacrylamide gel after silver staining (data not shown). This may be due to the size of the β-galactosidase portion of the fusion protein (114 kDa). For this reason it was decided to use an alternative fusion protein expression system.

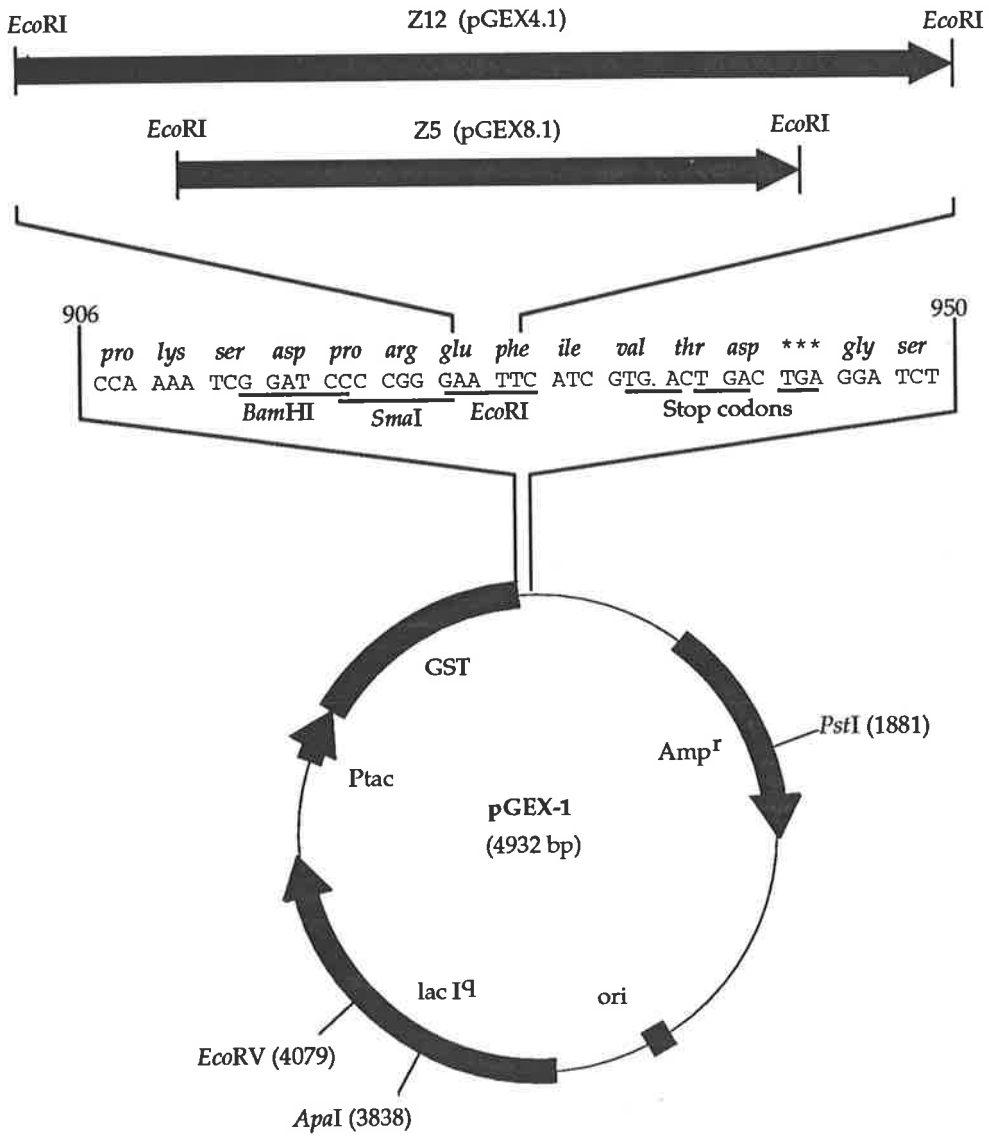
(ii) Subcloning of λ4.1 and λ8.1 into pGEX-1

The pGEX fusion protein expression system (Smith and Johnson, 1988) was chosen to express the Z5 and Z12 products. The pGEX-1 expression plasmid possesses an IPTG inducible *tac* promoter, adjacent to the coding sequence of Glutathionine S Transferase (GST) from *S. japonicum*, with a normal termination codon replaced by a polylinker followed by TGA termination codons in all three reading frames (Figure 4.11). Therefore, a protein coding region of interest can be cloned into the polylinker and fused in frame with the GST protein. Because GST is relatively small (26 kDa), most GST fusion proteins are also soluble and easily purified. The GST portion of the fusion protein enables the use of glutathionine affinity purification of the fusion protein if desired.

Figure 4.11 Z5 and Z12 pGEX-1 expression constructs

Shown is a diagramatic representation of the pGEX-1 expression vector (Smith and Johnson, 1988). The pGEX-1 plasmid contains an IPTG inducible *tac* promoter, followed by the coding region of the Glutathionine S Transferase (GST) gene from *S. japonicum*, in which the normal termination codon has been replaced by a polylinker (containing the unique sites for *Bam*HI, *Sma*I and *Eco*RI) followed by TGA termination codons in all three reading frames. Also present on the pGEX-1 plasmid is the β-lactamase coding gene (Amp^r), a bacterial replication origin (*ori*) and the $lacI^q$ gene coding for the *lac* repressor. The *lac* repressor binds the *tac* promoter repressing expression of the GST fusion protein, but upon addition of IPTG derepression occurs and GST fusion protein is expressed.

The *Eco*RI inserts from the λ4.1 (Z12) and λ8.1 (Z5) clones were cloned into *Eco*RI site of pGEX-1 in the orientation indicated. This ensured that the coding regions of the inserts were in the same reading frame as the GST gene.


EcoRI
Z12 (pGEX4.1)
EcoRI
EcoRI
Z5 (pGEX8.1)
EcoRI
906
950
pro lys ser asp pro arg glu phe ile val thr asp *** gly ser
CCA AAA TCG GAT CCC CGG GAA TTC ATC GTG ACT GAC TGA GGA TCT
BamHI
SmaI
EcoRI
Stop codons
GST
Ptac
pGEX-1
(4932 bp)
Amp^r
PstI (1881)
lac I^q
ori
EcoRV (4079)
ApaI (3838)

The pGEX-1 plasmid possesses an *Eco*RI site in it's polylinker in the same reading frame as the EcoRI site used to clone the λgt11 inserts (see Figure 4.11). Therefore the λ4.1, and λ8.1 clones were digested with *Eco*RI, and the inserts ligated into *Eco*RI digested pGEX-1. The ligation reactions were transformed into *E.coli* strain ED8799. Colonies were taken from the plate, and plasmid DNA prepared (7.3.1 (ii)). This DNA was digested (*Eco*RI to confirm the presence/sizes of the inserts and *Pst*I to determine the orientation of the inserts) and analysed by electrophoresis on an agarose gel (7.3.2). Plasmids containing either λ4.1 or λ8.1 inserts in the correct orientation were identified. These plasmids were designated pGEX4.1 and pGEX8.1 respectively.

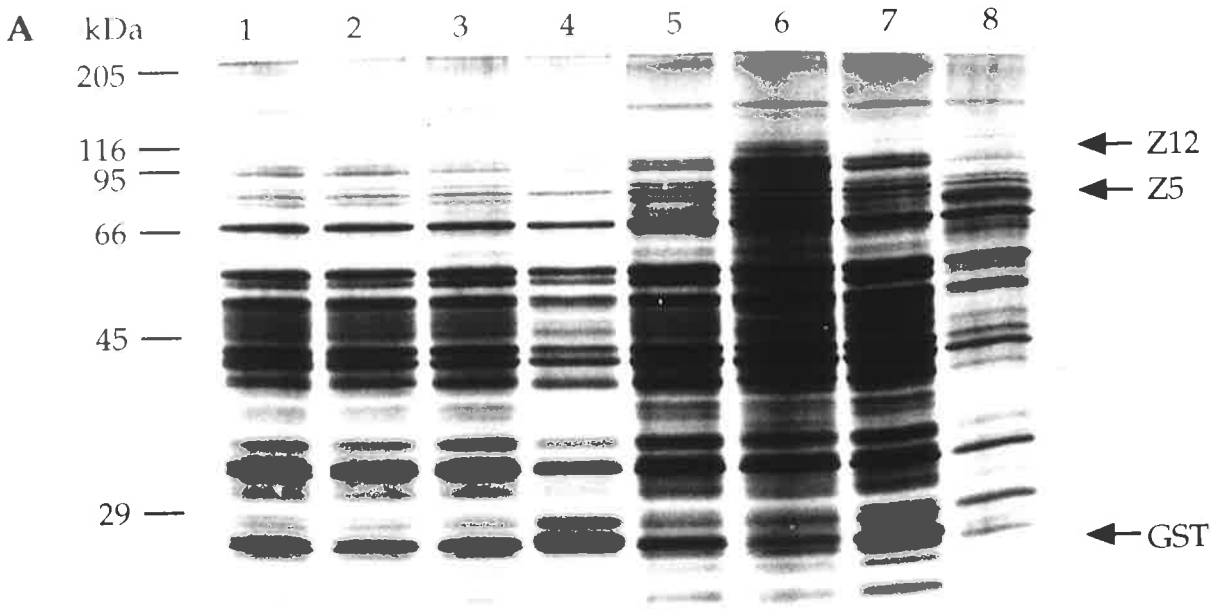
(iii) Production of fusion proteins

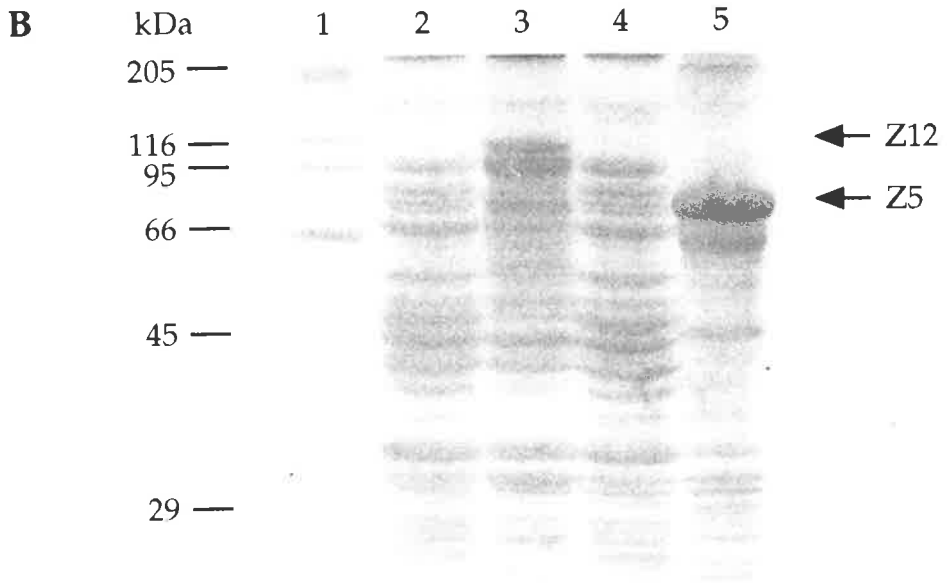
Induced and uninduced cultures of *E.coli* strain ED8799 transformed with the pGEX4.1 or pGEX8.1 constructs were prepared along with control cultures of *E.coli* strain ED8799 and *E.coli* strain ED8799 transformed with pGEX-1 (7.3.23 (i)). Cells were harvested by centrifugation, lysed by heating to 100°C for 5 minutes in lysis buffer, recentrifuged to pellet cell debris and supernatants recovered (7.3.23 (i)). These samples were analysed on an SDS-polyacrylamide gel (7.3.13 (i)), and protein visualised by silver staining (7.3.13 (iii)). Specific induced proteins were detected in the extracts from the pGEX4.1 and pGEX8.1 constructs (Figure 4.12A). A 105 kDa fusion protein was detected in pGEX4.1 extract giving an estimated size of 79 kDa for the Z12 protein (as compared to the deduced Z12 encoded protein size of 69,123 Da), and a 75 kDa fusion protein detected in pGEX8.1 extract giving an estimated size of 49 kDa for the Z5 protein (as compared to the deduced Z5 encoded protein size of 49,203 Da). The discrepancy between the estimated size of the Z12 encoded protein from the gel (79 kDa) and size from the deduced Z12 protein sequence (~ 69 kDa) was believed to be the result of the Z12 fusion product running slightly anomalously during SDS-polyacrylamide gel electrophoresis (i.e. some proteins do not always run 'true' to size because of their amino acid composition). However as will be explained in 5.4.3 a sequencing error was later found to be the source of this discrepancy.

Figure 4.12 Production of Z5 and Z12 GST fusion proteins

A Protein extracts were prepared from ED8799 cells alone and ED8799 cells containing the pGEX-1, pGEX4.1 or pGEX8.1 contructs (7.3.23 (i)). Samples of these extracts were subjected to SDS-polyacrylamide gel electrophoresis and protein visualised by silver staining (7.3.13 (i) and (iii)). The cultures had been uninduced or induced with 0.1 mM IPTG as indicated (7.3.23 (i)). Lanes are as follows;
1 and 2; Uninduced and induced ED8799 extract
3 and 4; Uninduced and induced pGEX-1 extract
5 and 6; Uninduced and induced pGEX4.1 extract
7 and 8; Uninduced and induced pGEX8.1 extract

B Z12 and Z5 fusion protein inclusion bodies were prepared from ED8799 cells containing the pGEX4.1 or pGEX8.1 constructs respectively (7.3.23 (ii)). The inclusion bodies were processed (washed, solubilized and refolded), samples subjected to SDS-polyacrylamide gel electrophoresis and protein visualised by coomassie staining (7.3.13 (i) and (iii)). The cultures had been uninduced or induced with 0.1 mM IPTG as indicated (7.3.23 (ii)). Lanes are as follows;
1; Protein molecular size markers (sizes of the markers are indicated)
2 and 3; Uninduced and induced Z12 fusion protein inclusion body extract
4 and 5; Uninduced and induced Z5 fusion protein inclusion body extract


A
kDa
1
2
3
4
5
6
7
8
205
116
95
66
45
29
Z12
Z5
GST


B
kDa
1
2
3
4
5
205
116
95
66
45
29
Z12
Z5

The levels of production of the Z12 and Z5 fusion proteins were still quite low compared to the pGEX-1 GST control (Figure 4.12A). The reason for the low levels of production of both fusion proteins was investigated. Microscopic examination of induced bacterial cells containing either pGEX4.1 or pGEX8.1 constructs (performed by C. Senn; Bresatec, Adelaide) revealed the presence of inclusion bodies in each case. This explained why comparatively low levels of both fusion proteins were detected in the extracts, since the bulk of the inclusion bodies would have been lost with the pelletted cell debris.

(iv) Preparation of inclusion bodies

Induced and uninduced cultures of *E.coli* strain ED8799 transformed with the pGEX4.1 or pGEX8.1 constructs were prepared (7.3.23 (ii)). Inclusion bodies were prepared, washed, solubilized, and refolded essentially as described in Sambrook *et al.* (1989) (7.3.23 (ii)). These samples were analysed on an SDS-polyacrylamide gel (7.3.13 (i)), and protein visualised by coomassie staining (7.3.13 (ii)). Specific induced proteins of 105 kDa and 75 kDa were detected in the extracts from the pGEX4.1 and pGEX8.1 constructs respectively (Figure 4.12B). Much larger amounts of the proteins were present as a result of isolating them by the inclusion body isolation method. However there was still a significant amount of contaminating protein co-purifying with both the Z5 and Z12 fusion protein inclusion bodies. This was undesirable but it was thought that preliminary characterisation of Z5 and Z12 DNA binding activities could still be undertaken using these extracts.

4.4.5 Characterisation of Z5 and Z12 binding activities

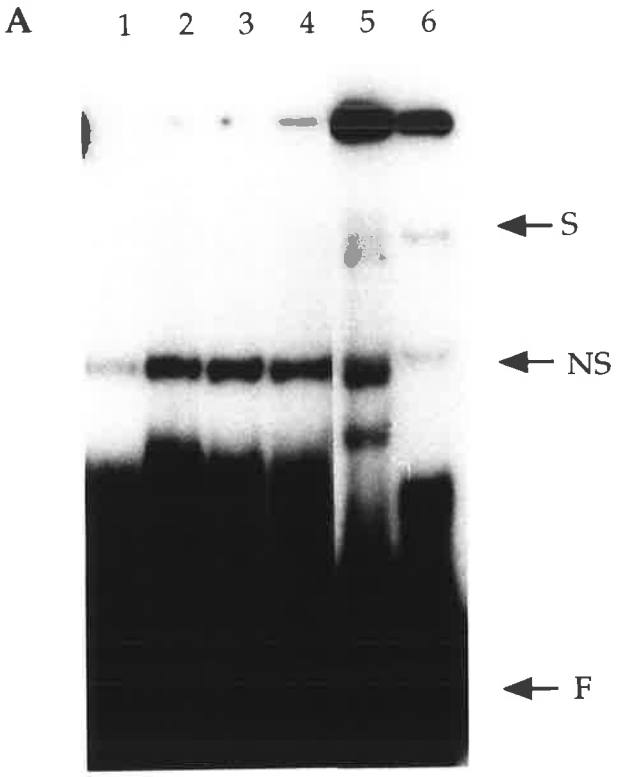
Preliminary characterisation of Z5 and Z12 DNA binding specificities was carried out by gel retardation analysis with the 25 mer probe. A number of control protein extracts were prepared and assayed in addition to the solubilized and refolded Z12 and Z5 fusion protein inclusion body extracts (prepared from the induced cultures of *E.coli* strain ED8799 transformed with the pGEX4.1 or pGEX8.1 constructs). The control extracts included extracts from the ED8799 strain alone (uninduced and

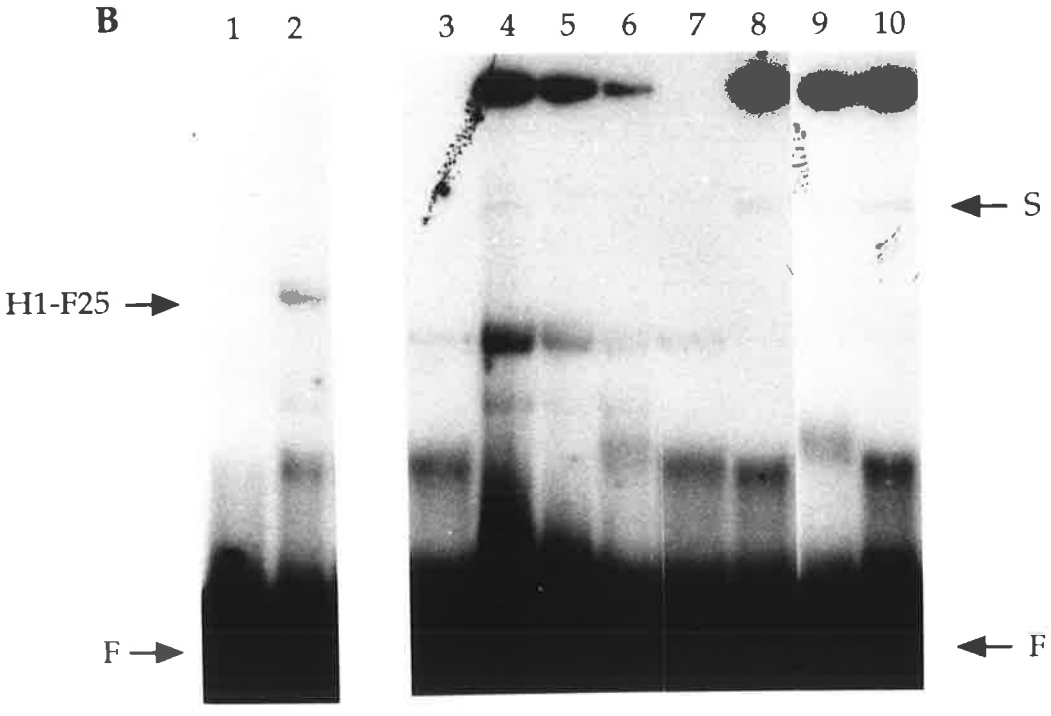
induced) and ED8799 transformed with the parental pGEX-1 vector (uninduced and induced). Labelled 25 mer probe was incubated with each of the extracts, in the oligoscreening buffer used originally to isolate the Z5 and Z12 clones (7.3.15 (iii)). DNA-protein complexes were separated from free DNA by non-denaturing polyacrylamide gel electrophoresis (7.3.12 (i)). Several different retarded DNA-protein complexes were detected in this initial gel retardation assay (Figure 4.13A). A single band was detected, that was common to all the extracts, which presumably represented a nonspecific interaction involving an ED8799 host strain protein. However, in the reactions containing induced Z5 or Z12 fusion protein extract, additional bands were observed. There was one unique band in the case of Z5 and at least four different bands in the case of Z12. These extra complexes were not present in uninduced Z5 and Z12 fusion protein extract reactions (Figure 4.13B). The specificities of these retarded complexes were next tested by gel retardation assay by incubation of 25 mer probe with Z5 or Z12 fusion protein extract in the presence of unlabelled heterologous or homologous DNA (Figure 4.13B). In the case of Z5, addition of an excess of unlabelled heterologous DNA (poly(dI-dC)) removed all bands, with the exception of the one unique Z5/25 mer complex mentioned above. Double stranded 25 mer competitor competed out all bands including the Z5/25 mer complex, while Δ25 mer competitor did not compete out the Z5/25 mer complex. This is consistent with the earlier oligoscreening results (4.3.3) where the λ8.1 (Z5) clone bound the 25 mer strongly but interacted very weakly with the Δ25 mer. In the case of the Z12 fusion protein extract reactions, it appears that the 25 mer and Δ25 mer competitors have both competed out the candidate Z12/25 mer complexes. This again is consistent with the earlier oligoscreening results (4.3.3), where the λ4.1 (Z12) clone interacted with both the 25 mer and the Δ25 mer almost equally.

Figure 4.13 Gel retardation analysis of Z5 and Z12 DNA binding specificities

A Labelled 25 mer probe (100 pg) was incubated with solubilized and refolded Z12 and Z5 fusion protein inclusion body extracts (1 µl; 10 µg/µl) and also control extracts from ED8799 cells alone (uninduced and induced) and ED8799 cells transformed with the parental pGEX-1 vector (uninduced and induced), in Kalionis and O'Farrell oligoscreening buffer (7.3.15 (iii)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). Non-specific (NS) and candidate specific (S; Z5/25 mer and Z12/25 mer) complexes formed between extract components and labelled DNA, in addition to free (F) unbound probe, were detected by autoradiography and are indicated. All lanes contain 25 mer probe. Lanes are as follows;
1 and 2; Uninduced and induced ED8799 extract
3 and 4; Uninduced and induced pGEX-1 extract
5; Induced Z12 fusion protein inclusion body extract
6; Induced Z5 fusion protein inclusion body extract

B Labelled 25 mer probe (100 pg) was incubated with solubilized and refolded Z12 and Z5 fusion protein inclusion body extracts (uninduced and induced), in Kalionis and O'Farrell oligoscreening buffer (7.3.15 (iii)), with heterologous or homologous DNA competitor (7.3.12 (i)). DNA-protein complexes were separated from free DNA by electrophoresis on a 12% non-denaturing polyacrylamide gel (7.3.12 (i)). The candidate specific (S; Z5/25 mer and Z12/25 mer) complexes and free (F) unbound probe were detected by autoradiography and are indicated. For comparison, labelled 25 mer probe (100 pg) was incubated with 9DCE nuclear extract (1 µl; 10 µg/µl) and assayed as above. The candidate specific (S; H1-F25) complex and free (F) unbound probe are indicated. All lanes contain 25 mer probe. Lanes are as follows;
1; 25 mer probe alone
2; 9DCE nuclear extract plus 100-fold excess poly(dI-dC)
3; Uninduced Z12 fusion protein inclusion body extract
4-6; Induced Z12 fusion protein inclusion body extract plus 100-fold excess poly(dI-dC), 20-fold excess 25 mer and 20-fold excess Δ25 mer competitors respectively
7; Uninduced Z5 fusion protein inclusion body extract
8-10; Induced Z5 fusion protein inclusion body extract plus 100-fold excess poly(dI-dC), 20-fold excess 25 mer and 20-fold excess Δ25 mer competitors respectively


A
1 2 3 4 5 6
S
NS
F


B
1 2 3 4 5 6 7 8 9 10
H1-F25
S
F
F

"These results are preliminary only. Further studies would require more precise standardisation of DNA from 'bound' and 'free' tracks (Fig. 4.4) and a longer probe to enable proper definition of the end-point of protein protection."

4.5 DISCUSSION

This chapter describes further efforts to identify, isolate and characterise a genuine H1 box-specific binding activity. Two 25 mer oligonucleotide probes were specially designed to maximize the possibility of detecting such an interaction. However, a totally H1 box-specific binding activity was not detected using these probes. Instead, a DNA binding activity, H1-F25, was detected which interacted with both the 25 mer (intact H1 box) and Δ25 mer (disrupted H1 box) probes. H1-F25 behaved like the previously identified H1-F40 binding activity (see Chapter 2), and for all intents and purposes the two complexes appear to represent the same DNA binding protein activity. However, further characterisation of H1-F25 binding specificity by modification interference analysis, demonstrated that H1-F25 actually bound both the H1 box and the immediate 5' flanking sequence (see 4.2.3 (ii)). Therefore H1-F25 binding activity may still represent a specific interaction with the H1 box if, as discussed in 4.2.2 (ii) and 4.2.3 (ii), H1-F25 binding activity, *in vitro*, is not completely compromised by the mutation in the H1 box and the interaction with the 5' flanking sequence is sufficient under the 'relaxed' *in vitro* binding conditions to maintain H1-F25/Δ25 mer binding.

Interestingly Dalton and Wells (1988*a*), using DNase I protection analysis, demonstrated protection of at least a 15 bp region over the H1 box including sequences both immediately 3' and 5' to H1 box. Additional 5' flanking sequence may also be involved in binding, however the published data does not cover this extended region. Therefore the protection analysis shown may suggest that the H1 box is not the only sequence requirement for H1-SF binding. With this in mind, and taking into account an as yet unexplained possibility that the binding conditions used throughout the course of this work may not have ever approximated the conditions that allowed the original (Dalton and Wells 1988*a*) discrimination of H1-SF, it was considered that H1-F25 binding activity could still represent H1-SF.

As a result, the isolation of H1-F25 was pursued by oligoscreening a ten day chicken embryo cDNA λgt11 expression library with concatenated 25 mer probe (see 4.3). Two potential H1-F25 clones, λ4.1 and λ8.1, were isolated that possessed similar DNA binding characteristics to H1-F25 as judged by both rescreening (oligoscreening) the purified clones with various oligonucleotide probes (4.3.3) and also expressing the clones as fusion products and subjecting the protein extracts to gel retardation analysis with the single copy 25 mer probe (see 4.4.4 and 4.4.5). Interestingly, the clones possessed slightly different binding characteristics. The λ4.1 clone appeared to bind the Δ25 mer probe as well or slightly better than the 25 mer while λ8.1 interacted only very weakly with the Δ25 mer probe. The gel retardation analysis will be discussed below.

The oligoscreening protocol used to isolate these clones involved a number of modifications compared to the method used in the original attempts to isolate H1-F14 (see Chapter 3). However, it is difficult to judge whether these changes were critical to the isolation of the H1-F25 clones or whether or not they would have allowed the detection of H1-F14, as a direct comparison between the two methods was not done. All of the changes involved with the Kalionis and O'Farrell (1993) method would potentially increase the sensitivity of the oligoscreening technique. Although in the case of the λ4.1 and λ8.1 clones the denaturation/renaturation process (common to both oligoscreening methods) appears to have been the most important step for their detection. This was concluded since no duplicate positive signals were detected when the filters were screened without this step. Vinson *et al.* (1988) discuss the merits of the denaturation/renaturation cycle and suggest that the process among other things, may 'melt' any insoluble precipitated form of fusion protein expressed from the the λgt11 clones. This may have been important for both the λ4.1 and λ8.1 clones as fusion proteins generated from these clones, albeit GST fusion products (see 4.4.4) were later shown to form inclusion bodies.

The λ4.1 and λ8.1 clones were found to encode novel zinc finger-containing proteins containing 12 zinc fingers and 5 zinc fingers respectively. The best matches found with the λ4.1 (Z12) and λ8.1 (Z5) clones, in the protein databases searched (see 4.4.2 and 4.4.3) were other zinc finger-containing proteins, and in these cases the observed amino acid sequence homology (at best 45%) was found only within the zinc finger regions of these proteins. The Z12 and Z5 clones appear to be incomplete. Both lack an ATG codon with a consensus Kozak ribosome binding site (Kozak, 1987) in their 5' region, suggesting that the clones do not contain full 5' coding sequences. No other readily identifiable DNA binding motifs appear to be encoded by the Z5 and Z12 clones. Therefore it would appear that the zinc fingers are soley responsible for the DNA binding activities of the Z5 and Z12 clones.

The zinc fingers in both Z12 and Z5 are of the Cys_2-His_2 finger type that match the consensus sequence motif (Xaa_5-Tyr/Phe-Xaa-Cys-$Xaa_{2,4}$-Cys-Xaa_3-Phe-Xaa_5-Leu-Xaa_2-His-$Xaa_{3\text{-}5}$-His). The amino acid sequences of the zinc fingers from the Z12 and Z5 clones along with the consensus sequence have been aligned for comparison in Figure 4.14. Extended homology exists between the blocks of fingers in each clone with several additional residues highly conserved. However these highly conserved residues are apparently not involved directly in determining DNA binding specificity for the fingers. Studies of Cys_2-His_2 zinc fingers have revealed that each finger domain interacts with 3 adjacent base pairs of DNA through three key residues (Xaa^{17}, Xaa^{20}, Xaa^{23}; see Figure 4.14) located in the α-helix of the finger domain (Berg, 1988; Nardelli *et al.*, 1992; Pavletich and Pabo, 1991). Interaction of these residues with the DNA involves the protein sequence (fingers) in the NH_2 to COOH orientation running antiparallel to the contacted bases in the 5' to 3' direction (i.e. Xaa^{23} recognises the first base pair, Xaa^{20} the central base pair, and Xaa^{17} the third base pair in the triplet). The relationship between the different amino acid residues at these positions and the DNA triplet sequences they recognise has been the subject of several recent investigations (Nardelli *et al.*, 1992; Thukral *et al.*, 1992; Desjarlais and Berg, 1992). Some general associations between particular amino

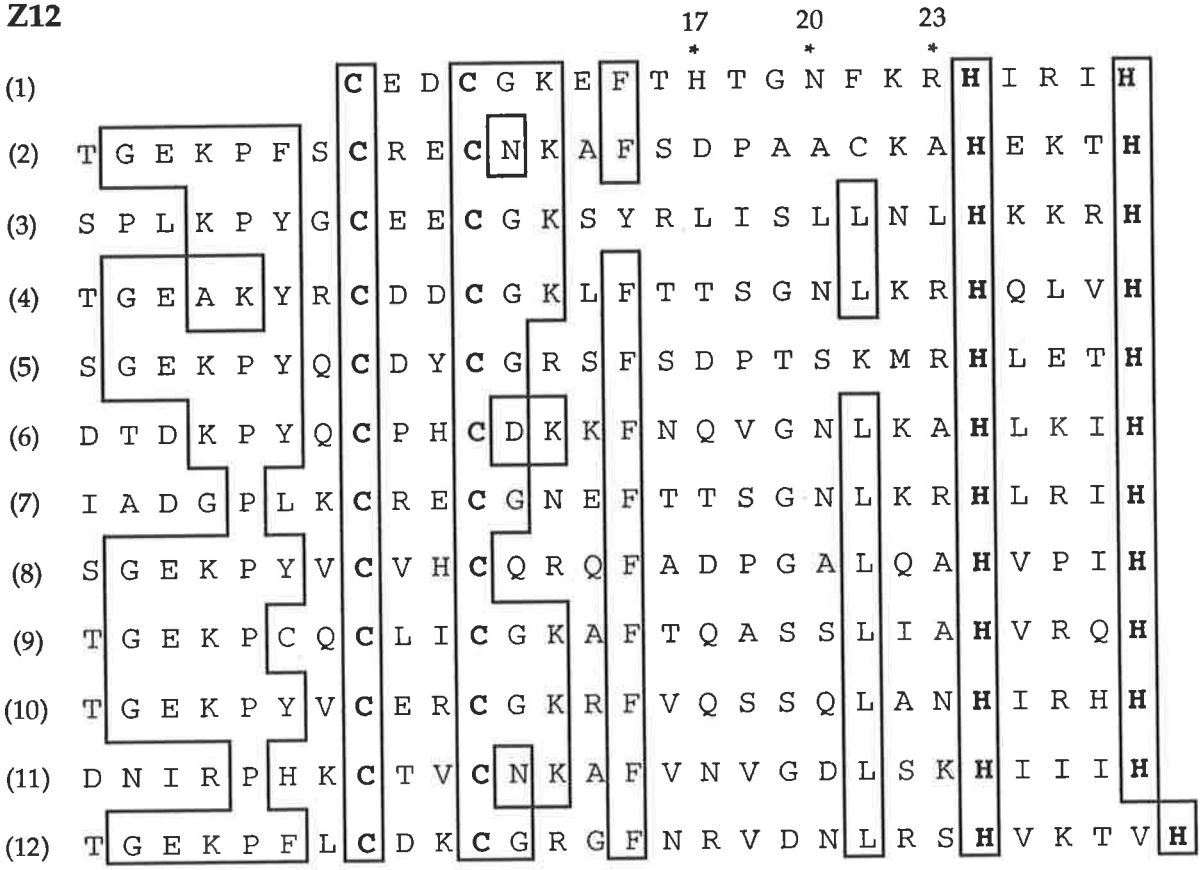
Figure 4.14 Amino acid sequence comparison of the Z12 and Z5 zinc fingers
Displayed is the amino acid sequence comparison of the Z12 and Z5 zinc fingers. The numbers in the brackets indicate the number of the zinc finger relative to its position in the primary sequence (see Figures 4.9 and 4.10). A consensus sequence for the Cys_2-His_2 type finger domain is also shown, with highly conserved residues noted and dashes marking positions which are more variable. Conserved cysteines and histidines are in bold and boxed while other conserved residues are boxed. Amino acid residues implicated in DNA recognition are indicated with asterisks with numbers indicating position in finger repeat unit.

The designations for amino acid residues are in the standard single-letter code; A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

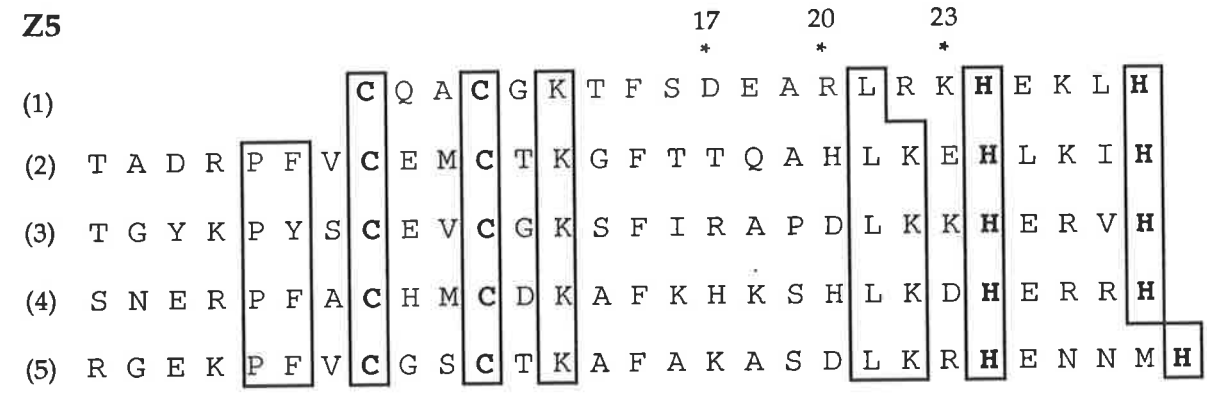
C_2-H_2 Finger consensus

- - - - - $^{Y}/_{F}$- **C** - - **C** - - - F - - - - - L - - **H** - - - **H**

Z12



Z5



acids and bases have been identified. However in many cases no simple relationship appears to exist, with more complex interactions between neighbouring residues probably involved in determining the exact DNA binding specificity. Additionally most of the work that has been done in this area has involved characterisation of zinc finger proteins with G-rich binding sites. Therefore it is difficult to comment on whether or not the amino acid residues in these key DNA recognition positions in the Z5 and Z12 zinc fingers are consistent with a specific interaction with the A-rich H1 box. However, interestingly in both the Z5 and Z12 fingers there is under representation of arginine or histidine residues in positions 17, 20 and 23 which seem to be the predominant residues in these positions in zinc fingers that recognise G-rich binding sites (Thukral *et al.*, 1992). Also the number of zinc fingers in the clones may suggest the size of the binding site recognised by the proteins. With 5 zinc fingers, Z5 would be expected to specifically recognise 15 bp of DNA. This is consistent with the size of the H1 box and immediate flanking sequence. The 12 zinc fingers of Z12 suggests an interaction with a larger binding site of 36 bp (assuming that every finger interacts with the DNA at one time). A further observation is that the last finger in each block of fingers in the Z12 and Z5 clones contains an extra amino acid before the last histidine residue. The evolution and the significance of this structure to the common binding specificities of these two clones cannot be resolved at this time.

It was important to elucidate whether or not the Z12 and Z5 clones actually represent proteins that specifically interact with the H1 box. Work involving the expression of the Z12 and Z5 proteins as fusion products and analysis of these products by gel retardation assay attempted to resolve this question. Although only preliminary gel retardation data was obtained (see 4.4.5), when considered along with the oligoscreening results it appeared that there was some sequence specificity displayed towards the H1 box. Indeed in the case of Z5, a marked increase in binding specificity to the H1 box, not previously seen with H1-F25, was observed. However, as these studies were all done using Z12 and Z5 fusion protein products, the binding

specificities of the clones may not truly represent the binding specificities of the native proteins. Further studies should include processing Z12 and Z5 products to remove the fusion portion. It was observed that the Z12 and Z5 products behaved similarly whether they were associated with β-gal (oligoscreening) or GST (pGEX-1 expression products) fusions suggesting that Z12 and Z5 DNA binding properties were not modified or compromised in the fusion proteins. Other zinc finger proteins have been shown to require zinc for coordinating the finger structures and DNA binding (Miller *et al.*, 1985; Kadonaga *et al.*, 1987; Eisen *et al.*, 1988; Lee *et al.*, 1991). The fact that Z12 and Z5 proteins were able to be detected by oligoscreening and also able to bind the 25 mer probe in the gel retardation assay using buffer without added Zn^{2+} requires explaination. Trace amounts of zinc (or perhaps some other divalent ion able to substitute for zinc) may be present in the buffer (and/or extract in the case of the gel retardation analysis) at a concentration required to permit the finger proteins to bind DNA. Interestingly Dailey *et al.* (1987) have observed that the H4TF1 and H4TF2 histone H4 specific transcription factors (see 1.11.1 (ii)) require zinc for DNA binding suggesting that they may represent zinc finger containing proteins. Obviously further studies need to be done to evaluate the affect of zinc on Z12 and Z5 DNA binding. Finally, the DNA binding specificities displayed by the incomplete Z12 and Z5 clones may not be a true representation of the DNA binding preferences of the complete Z12 and Z5 proteins. Therefore isolation of the complete Z12 and Z5 clones may be required before further characterisation of their DNA binding specifities can be achieved.

CHAPTER 5

ISOLATION AND CHARACTERISATION OF FURTHER Z5 AND Z12 cDNA AND GENOMIC CLONES

5.1 INTRODUCTION

In the previous chapter the isolation and preliminary characterisation of two novel zinc finger protein clones, Z12 and Z5, was described. However as both clones were incomplete, it was decided to obtain full length cDNA clones to enable the full characterisation of Z12 and Z5 proteins. This chapter describes work concerned firstly with the isolation of full length Z5 and Z12 cDNA clones from a ten day chicken embryo cDNA λgt11 library, and secondly the screening of a chicken genomic λ Charon 4a library for Z5 and Z12 genomic clones. The reason for trying to isolate the genes for Z5 and Z12 was that their characterisation may help determine their identity and function. In particular, the availability of the full length genomic sequence of Z5 and Z12 and identification of their promoters could lead to studies concerned with the mechanism of their transcriptional control and provide a basis for further investigation of the biology of the proteins. Screening for full length Z5 and Z12 cDNA clones and the Z5 and Z12 genes were carried out in parallel.

5.2 NORTHERN ANALYSIS

Sequence analysis of the Z5 and Z12 clones suggested that both clones were incomplete (4.4.2 and 4.4.3). It was hoped that Northern analysis of RNA from nine day chicken embryos would indicate the size of the full length mRNAs for both clones. RNA from nine day chicken embryos was used since this was the source material used to construct the cDNA λgt11 expression library from which the Z5 and Z12 clones were isolated. Both total RNA and poly(A)$^+$ RNA were prepared as described in 7.3.18 (i) and (ii). Approximately 8 mg total RNA was generated from 6 g of embryos, and 5 mg of this processed to yield approximately 75 μg of poly(A)$^+$ RNA. Total RNA (10 μg) and poly(A)$^+$ RNA (1 μg) were run on a 1.2% agarose/formaldehyde gel along with 5 μg of RNA markers (7.3.18 (iii)). The RNA was vacuum blotted onto a Genescreen membrane, and UV-crosslinked (7.3.18 (iv)). A 380 bp *Eco*RI/*Sca*I restriction fragment from the Z5 cDNA clone and a 367 bp *Eco*RI/*Tth*III1 restriction fragment from the Z12 cDNA clone (see Figures 4.8A and 4.8B), each representing the 5' ends of their respective cDNAs, were oligo-labelled

with α-^{32}P-dATP (7.3.4 (iii)) and used to probe Z5 and Z12 mRNAs. The filter was first hybridised with the Z5 probe and processed as described in section (7.3.18 (iv)). After autoradiography (7.3.24) the filter was stripped, rehybridised with the Z12 probe, and processed as before. However no specific signals representing Z5 and Z12 mRNA species were seen even after 1-2 weeks autoradiographic exposure.

A control chicken β-Actin probe was used to reprobe the filter, to check the integrity of the RNA and that the Northern transfer technique was working. The chicken β-Actin probe, a 220 bp fragment, was prepared by *Kpn*I and *Pvu*II digestion of a construct containing the chicken β-Actin cDNA (pPAc; 7.2.6 (i)), and oligo-labelling the fragment with α-^{32}P-dATP (7.3.4 (iii)). Northern analysis was carried out as above. A specific signal was detected, representing β-Actin mRNA (1.8 kb), after only 1 hour autoradiographic exposure (Figure 5.1A). Therefore it appeared that RNA was still intact on the filter and the Northern transfer technique was working, but indicated that both the Z5 and Z12 mRNA species were at a very low abundance and not able to be detected (at least with the amount of RNA used for this Northern analysis).

The Northern analysis was repeated using increased amounts of both total (25 μg) and poly(A)$^+$ (10 μg) RNA. The filter was first hybridised with the Z5 probe and processed as described in Section 7.3.18 (iv). After autoradiography (7.3.24) the filter was stripped, rehybridised with the Z12 probe, and processed as before. The results are shown in Figure 5.1B. A faint band corresponding to a mRNA of approximately 3.0 kb in size was detected with the Z12 probe, but only in the poly(A)$^+$ RNA track and only after 1 week autoradiographic exposure. Using similar conditions it was not possible to detect a specific signal representing the Z5 mRNA species. These results suggested that both Z5 and Z12 RNA species are of very low abundance, and, given the size of the mRNA (3.0 kb) detected by the Z12 probe, that the 2.1 kb Z12 cDNA clone was incomplete.

Figure 5.1

A Chicken β-Actin control Northern analysis

Total RNA (10 μg) and poly (A)+RNA (1 μg) isolated from nine day chicken embryos were run on a denaturing 1.2% agarose gel along with 5 μg of RNA markers (7.2.7 (iii)), transferred to Genescreen membrane and probed with a ^{32}P-labelled 220 bp *Kpn*I/*Pvu*II chicken β-Actin cDNA fragment (7.3.18). A specific signal representing β-Actin mRNA, detected after 1 hour autoradiographic exposure, is indicated. Sizes of the RNA markers are shown. Lanes are as follows;

1; Total RNA

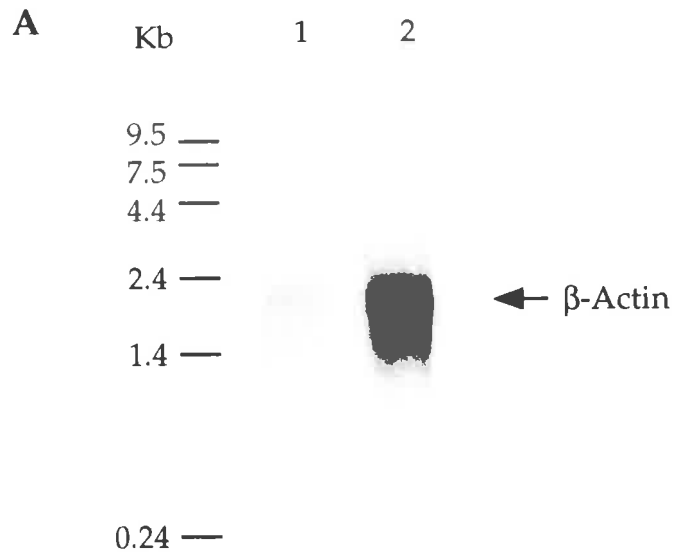
2; Poly(A)+ RNA

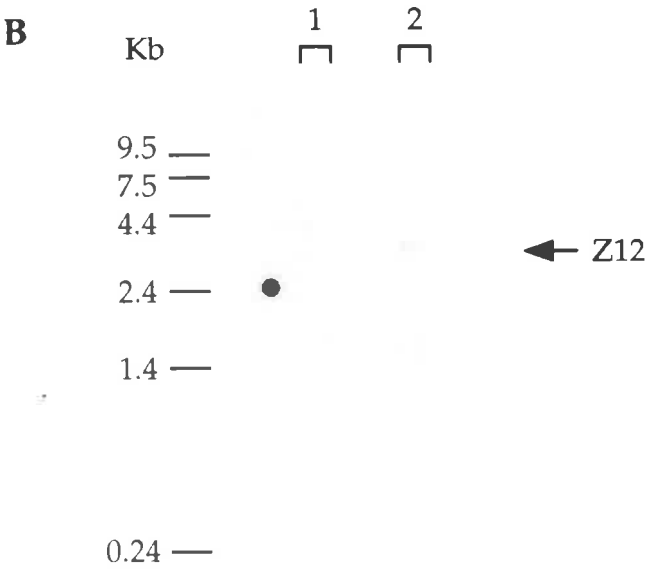
B Z12 Northern analysis

Total RNA (25 μg) and poly (A)+RNA (10 μg) isolated from nine day chicken embryos were run on a denaturing 1.2% agarose gel along with 5 μg of RNA markers (7.2.7 (iii)), transferred to Genescreen membrane and probed with a ^{32}P-labelled 5' *Eco*RI/*Tth*III1 restriction fragment from the Z12 cDNA clone (7.3.18). A signal representing Z12 mRNA was detected after 1 week autoradiographic exposure is indicated. Sizes of the RNA markers are shown. Lanes are as follows;

1; Total RNA

2; Poly(A)+ RNA


A
Kb
1
2
9.5
7.5
4.4
2.4
1.4
0.24
β-Actin


B
Kb
1
2
9.5
7.5
4.4
2.4
1.4
0.24
Z12

5.3 cDNA LIBRARY SCREENING

In an attempt to isolate the full length Z5 and Z12 cDNA clones it was decided to rescreen the ten day chicken embryo cDNA λgt11 library (Clontech library, catalogue number CL1001b; 1.06 x 10^6 independent clones) by conventional means.

5.3.1 Probe preparation

A 380 bp *Eco*RI/*Sca*I restriction fragment and 292 bp *Hin*dIII/*Eco*RI restriction fragment from the Z5 cDNA clone (see Figure 4.8B), representing the 5' and 3' ends of the clone, were oligo-labelled with α-^{32}P-dATP (7.3.4 (iii)) and used as a probes to screen for full length Z5 cDNA. Both 5' and 3' Z5 probes were used because it appeared that the Z5 clone was incomplete at both 5' and 3' ends. A 367 bp *Eco*RI/*Tth*III1 restriction fragment from the Z12 cDNA clone (see Figure 4.8A), representing the 5' end of the clone, was oligo-labelled with α-^{32}P-dATP (7.3.4 (iii)), and used as a probe to screen for full length Z12 cDNA.

5.3.2 Z5 and Z12 cDNA library screening

Approximately 2 x 10^6 plaques were screened as described in Section 7.3.16. Briefly, the ten day chicken embryo cDNA λgt11 cDNA library was absorbed to *E.coli* strain Y1090, and plated onto 15 cm L-agar + 50 μg/ml ampicillin plates with approximately 50,000 plaques per plate. The plates were incubated at 37°C for approximately 6 hours and then chilled at 4°C before the phage were transferred in duplicate onto Plaquescreen filters. The phage were lysed by a 2 minute treatment in an autoclave set at 105°C and the DNA baked onto the filters for 2 hours at 80°C. Prehybridization, hybridization and washing conditions were as described in 7.3.16. The filters were first probed with the Z5 probes, then after being stripped (7.3.16), probed with the Z12 probe. A total of three Z5 and eight Z12 first round duplicate positives were identified. A representative Z12 first round duplicate positive (λ16.1) is shown in Figure 5.2. Positive plaques were picked into 1ml PSB and the phage allowed to elute overnight at 4°C.

Figure 5.2 Z12 cDNA library screening first round duplicate positive
Shown is an example of a Z12 first round duplicate positive, λ16.1, identified as a result of screening a ten day chicken embryo cDNA λgt11 library with the ^{32}P-labelled 5' *Eco*RI/*Tth*III1 restriction fragment, from the Z12 cDNA clone (7.3.16). Filters are as follows;

1 and 2; First and second duplicate filter lifts with the autoradiographic signal representing the λ16.1 plaque indicated on each filter by an arrow.

1

2

5.3.3 PCR analysis of Z5 and Z12 cDNA clones

Preliminary characterisation of the Z5 and Z12 cDNA clones that were isolated above, was carried out by PCR. This was done using two general primers that flanked the *Eco*RI site of λgt11, in combination with three clone specific primers that hybridised near the 5' and 3' ends of the Z5 cDNA, and near the 5' end of the Z12 cDNA respectively. A diagramatic representation of the primers, the postions they prime from, and the sizes of the PCR products they should produce is shown in Figure 5.3A (see Section 7.2.8 for sequences of the PCR primers). The short PCR products indicated in Figure 5.3B were diagnostic for the incomplete Z5 and Z12 cDNA clones. It was hoped that full length Z5 and Z12 candidate clones could be quickly identified by the longer PCR products expected to be generated from these clones.

(i) Z5 cDNA clones

Each of the Z5 cDNA clones were analysed by PCR (7.3.19) with both the 5' and 3' Z5 specific primers (λ8.1 PCR 1 and λ8.1 PCR 2) in combination with each of the general λgt11 primers (λgt11 PCR 1 or λgt11 PCR 2). All three clones produced identical size PCR products (both 3' and 5' products) to those generated from the original Z5 clone isolated by oligoscreening (data not shown). This suggested that these clones were identical to the original Z5 clone. These clones were not investigated further.

(ii) Z12 cDNA clones

Each of the Z12 cDNA clones were analysed by PCR (7.3.19) with the 5' Z12 specific primer (λ4.1 PCR 1) in combination with each of the general λgt11 primers (λgt11 PCR 1 or λgt11 PCR 2). Five of the clones produced identical size 5' PCR products to that generated from the original Z12 clone, which suggested that these clones were identical to the original Z12 clone. A further two clones failed to produce PCR products. This suggested that these clones were either, smaller than the original Z12 clone and lacked the 5' sequence needed to hybridise with the 5' Z12 specific primer, or were perhaps not related to Z12 at all. These seven clones were not investigated

Figure 5.3 PCR analysis of candidate full length Z5 and Z12 cDNA clones

A Shown is a diagramatic representation of the λ4.1 (Z12) and λ8.1 (Z5) clones and the PCR primers (represented by horizontal arrows) designed to detect full length Z5 and Z12 clones. Two λgt11 specific primers (λgt11 PCR1 and λgt11 PCR2) were used in combination with three clone specific primers (λ8.1 PCR1, λ8.1 PCR2 and λ4.1 PCR1) to PCR 'screen' candidate Z5 and Z12 clones isolated from a nine day chicken embryo cDNA λgt11 library. The sizes of the PCR products generated from the original Z5 and Z12 clones, using the different combinations of primers, are indicated.

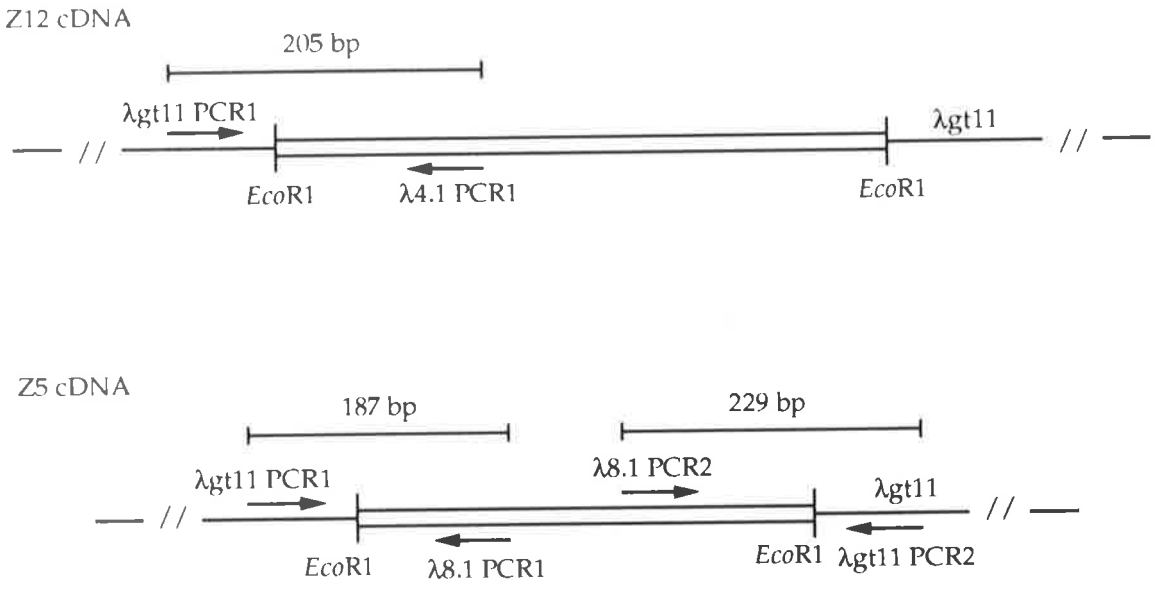
B Control PCR reactions were carried out with the λ4.1 (Z12) and λ8.1 (Z5) clone templates (10 ng) and the different combinations of primers described above (7.3 19). Samples (5 μl) from each of the completed reactions along with 1 μg of *Hpa*II digested pUC19 DNA markers were run on a 4% agarose gel and stained with ethidium bromide. Lanes are as follows;

1; *Hpa*II digested pUC19 DNA markers
2; λ4.1 (Z12) template plus λgt11 PCR1/λ4.1 PCR1 primers
3; λ4.1 (Z12) template plus λgt11 PCR2/λ4.1 PCR1 primers
4; λ8.1 (Z5) template plus λgt11 PCR1/λ8.1 PCR1 primers
5; λ8.1 (Z5) template plus λgt11 PCR2/λ8.1 PCR1 primers
6; λ8.1 (Z5) template plus λgt11 PCR1/λ8.1 PCR2 primers
7; λ8.1 (Z5) template plus λgt11 PCR2/λ8.1 PCR2 primers

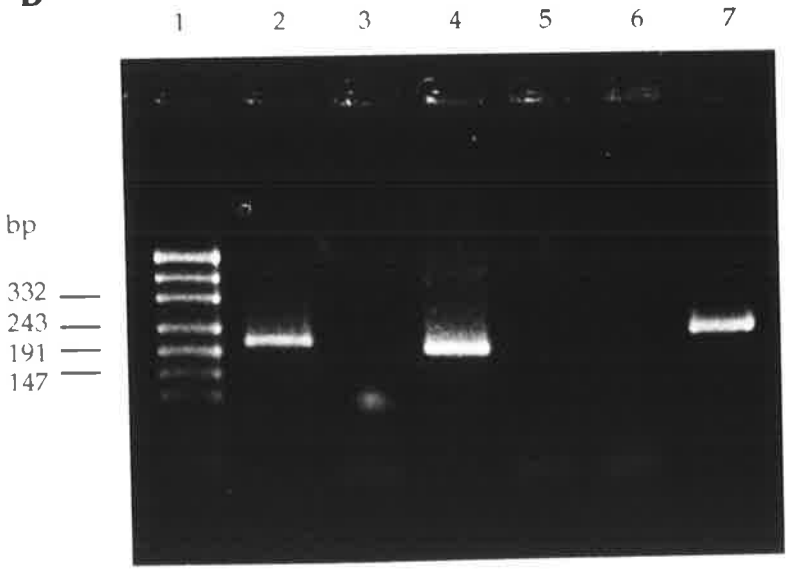
C PCR reactions were carried out with the candidate Z12 clone, λ16.1 template (10 ng) and the different combinations of primers described above (7.3 19). Samples (5 μl) from each of the completed reactions along with 1 μg of *Hpa*II digested pUC19 DNA markers were run on a 4% agarose gel and stained with ethidium bromide. Lanes are as follows;

1; *Hpa*II digested pUC19 DNA markers
2; λ16.1 template plus λgt11 PCR1/λ4.1 PCR1 primers (see Figure 5.6B for more accurate sizing of λ16.1 PCR product).
3; λ16.1 template plus λgt11 PCR2/λ4.1 PCR1 primers

A


Z12 cDNA
205 bp
λgt11 PCR1
λgt11
EcoR1
λ4.1 PCR1
EcoR1
Z5 cDNA
187 bp
229 bp
λgt11 PCR1
λ8.1 PCR2
λgt11
EcoR1
λ8.1 PCR1
EcoR1
λgt11 PCR2

B


1 2 3 4 5 6 7
bp
332
243
191
147

C


1 2 3

further. However, a 5' PCR product approximately 1.1 kb in size was generated from the final clone, λ16.1 (Figure 5.3C). This suggested that λ16.1 contained approximately 900 bp extra 5' sequence compared to the Z12 clone. This was consistent with the λ16.1 insert containing the rest of the Z12 5' sequence taking into account that the Z12 mRNA was shown to be approximately 3.0 kb in size by Northern analysis (see 5.2).

5.4 ANALYSIS OF λ16.1 CLONE

λ DNA was prepared from the λ16.1 clone (7.3.17). However digestion with *Eco*RI generated two fragments of approximately 2.1 kb and 0.6 kb in size. This suggested that there was either an internal *Eco*RI site in the full length Z12 cDNA or perhaps the two fragments represent individual unrelated clones that were ligated together when the library was constructed.

5.4.1 Subcloning of λ16.1 into pBSSK+

To make further manipulation of the λ16.1 clone easier, DNA from the clone was digested with *Eco*RI, and the two fragments ligated separately into *Eco*RI digested pBluescript SK+. The ligation reactions were transformed into *E.coli* strain DH5α, and colonies containing recombinants identified by colour selection (7.3.3). Plasmid DNA was prepared from recombinants as described in Section 7.3.1 (ii). The DNA samples were digested with the appropriate restriction enzymes (*Eco*RI to confirm the presence and size of the inserts, and *Sal*I to determine the orientation of the λ16.1 2.1 kb insert) and analysed by electrophoresis on an agarose gel (7.3.2). Plasmids containing the 2.1 kb insert in the two different orientations (designated pBS16.11 and pBS16.12) or the 0.6 kb insert (designated pBS16.13) were identified and large scale plasmid preparations of each produced (7.3.1 (i)).

(i) Restriction endonuclease mapping

A series of restriction enzyme digestions of the 2.1 kb and 0.6 kb inserts from λ16.1 were carried out (7.3.2 (i)) and from analysis of these, restriction maps were

generated for the inserts (Figures 5.4A and 5.4B). The λ16.1 2.1 kb insert possessed a very similar restriction pattern to the original Z12 clone but appeared to possess an extra 250 - 300 bp of 5' sequence (and lack approximately 150 - 200 bp of 3' sequence) compared to the original λ4.1 Z12 clone.

(ii) Dideoxy sequencing

A series of deletion subclones of pBS16.11 and pBS16.12 were made using the Promega Erase-a-base system (7.3.20). The pBS16.11 and pBS16.12 clones were digested with *Kpn*I and *Hin*dIII. These digestions linearised the clones in the polylinker and left the vector protected from deletion but the inserts open to exonuclease III action. Exonuclease III deletion subclones were then generated according to the manafacturers instructions and transformed into *E.coli* strain DH5α. Sets of overlapping subclones were size selected after electrophoresis on agarose gels. The two sets of deletion subclones were sequenced (Sanger *et al.*, 1977; 7.3.21). This enabled the entire sequence of both strands of the λ16.1 2.1 kb insert to be determined. The sequencing strategy is shown in Figure 5.4A, and the sequence in Figure 5.5. The reading frame determined from the original λgt11 clone, and the deduced amino acid sequence are indicated.

The 0.6 kb insert was sequenced (Sanger *et al.*, 1977; 7.3.21) from both ends (approximately 400 bp of sequence in each direction generating 200 bp of overlapping sequence). These overlapping sequences were combined to generate the complete sequence of the 0.6 kb insert (data not shown).

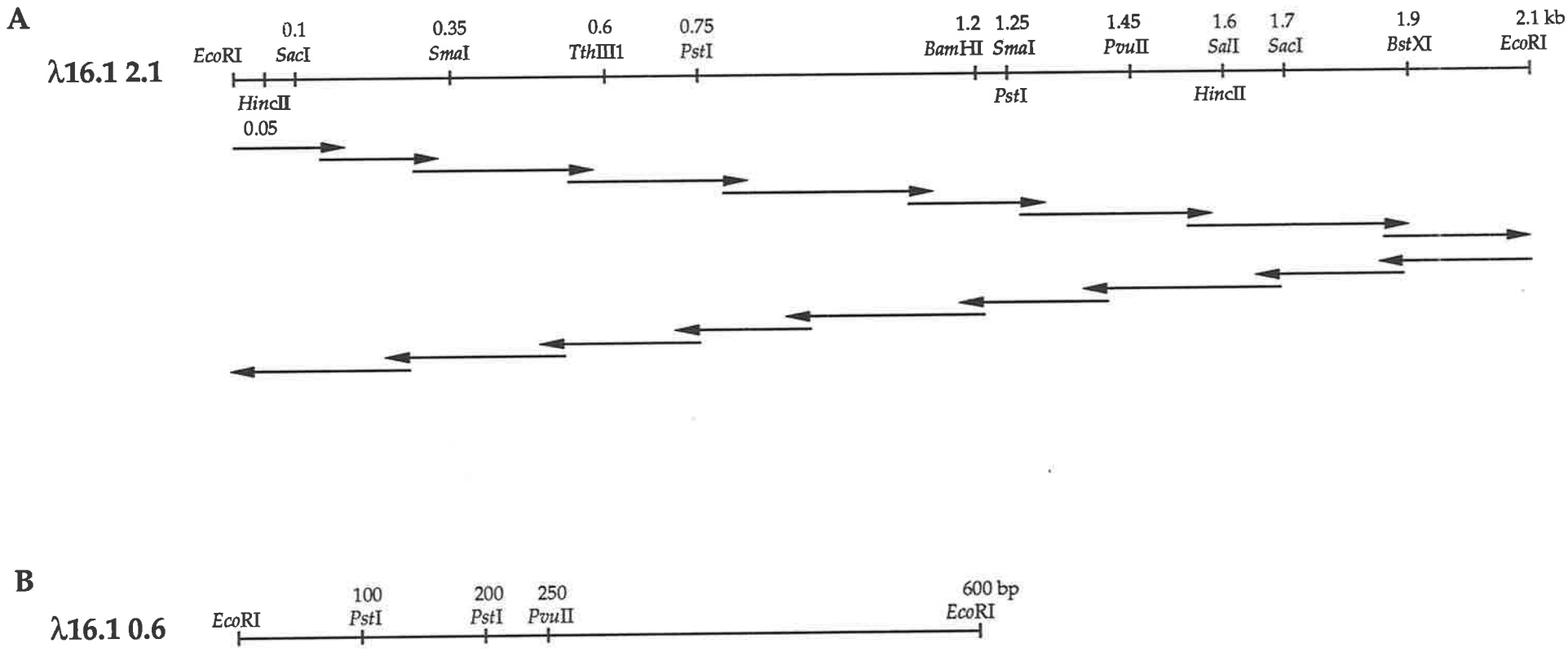
(iii) PCR aided orientation of λ16.1 2.1 kb and 0.6 kb inserts

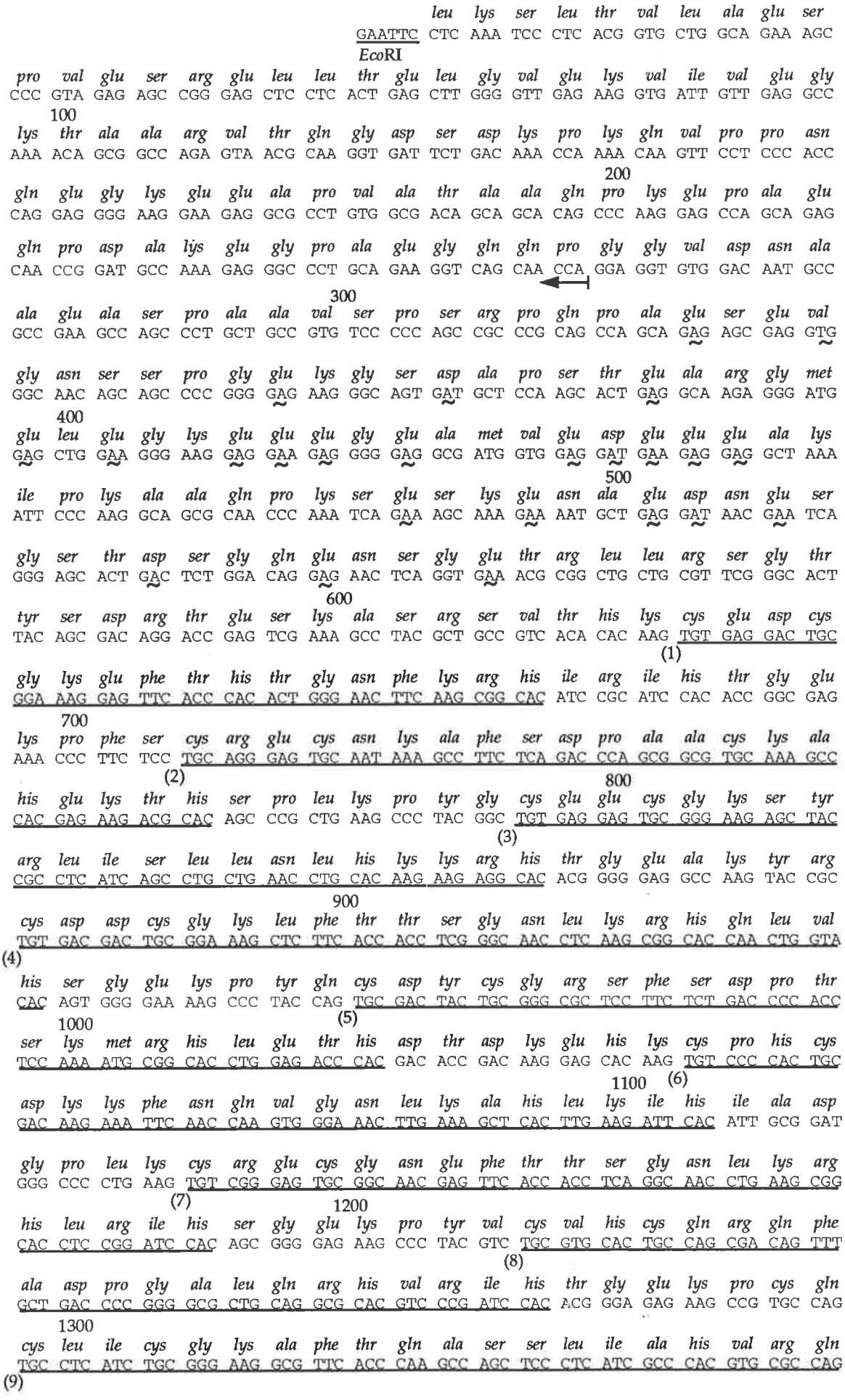
A primer (λ16.1 PCR 1) was designed that hybridised approximately 200 bp from the end of the 0.6 kb insert. The λ16.1 PCR 1 primer was used in combination with either the λgt11 PCR 1 primer or the 5' Z12 specific primer (λ4.1 PCR 1) to orientate the 0.6 kb and 2.1 kb inserts. A diagramatic representation of the primers, the positions they prime from, and the sizes of the PCR products expected, depending on the

Figure 5.4 Dideoxy sequencing of λ16.1 2.1 kb and 0.6 kb inserts

A The 2.1 kb *Eco*RI insert from λ16.1 was cloned in both orientations into *Eco*RI digested pBluescript SK+ to produce constructs designated pBS16.11 and pBS16.12 (see 5.4.1). A series of restriction enzyme digestions of the pBS16.11 and pBS16.12 constructs were carried out and from these a restriction map for the λ16.1 2.1 kb insert was generated. The positions of internal restriction sites are indicated, and numbered as measured from the 5' end of the insert. Deletion subclones from both pBS16.11 and pBS16.12 were created in both directions using the Erase-a-base system (see 5.4.1 (ii)). The entire cDNA clone was sequenced in both directions, as indicated by the arrows.

B The 0.6 kb *Eco*RI insert from λ16.1 was cloned into *Eco*RI digested pBluescript SK+ to produce a construct designated pBS16.13 (see 5.4.1). A series of restriction enzyme digestions of the pBS16.13 construct was carried out and from these a restriction map for the λ16.1 0.6 kb insert was generated. The positions of internal restriction sites are indicated, and numbered as measured from the 5' end of the insert. The 0.6 kb insert was sequenced from both ends with approximately 400 bp of sequence obtained in each direction. These overlapping sequences were combined to generate the complete sequence of the 0.6 kb insert.


A
λ16.1 2.1
EcoRI
0.1
SacI
HincII
0.05
0.35
SmaI
0.6
TthIII1
0.75
PstI
1.2
BamHI
1.25
SmaI
PstI
1.45
PvuII
1.6
SalI
HincII
1.7
SacI
1.9
BstXI
2.1 kb
EcoRI
B
λ16.1 0.6
EcoRI
100
PstI
200
PstI
250
PvuII
600 bp
EcoRI


leu lys ser leu thr val leu ala glu ser
GAATTC CTC AAA TCC CTC ACG GTG CTG GCA GAA AGC
EcoRI
pro val glu ser arg glu leu leu thr glu leu gly val glu lys val ile val glu gly
CCC GTA GAG AGC CGG GAG CTC CTC ACT GAG CTT GGG GTT GAG AAG GTG ATT GTT GAG GCC
100
lys thr ala ala arg val thr gln gly asp ser asp lys pro lys gln val pro pro asn
AAA ACA GCG GCC AGA GTA ACG CAA GGT GAT TCT GAC AAA CCA AAA CAA GTT CCT CCC ACC
200
gln glu gly lys glu glu ala pro val ala thr ala ala gln pro lys glu pro ala glu
CAG GAG GGG AAG GAA GAG GCG CCT GTG GCG ACA GCA GCA CAG CCC AAG GAG CCA GCA GAG
gln pro asp ala lys glu gly pro ala glu gly gln gln pro gly gly val asp asn ala
CAA CCG GAT GCC AAA GAG GGC CCT GCA GAA GGT CAG CAA CCA GGA GGT GTG GAC AAT GCC
300
ala glu ala ser pro ala ala val ser pro ser arg pro gln pro ala glu ser glu val
GCC GAA GCC AGC CCT GCT GCC GTG TCC CCC AGC CGC CCG CAG CCA GCA GAG AGC GAG GTG
gly asn ser ser pro gly glu lys gly ser asp ala pro ser thr glu ala arg gly met
GGC AAC AGC AGC CCC GGG GAG AAG GGC AGT GAT GCT CCA AGC ACT GAG GCA AGA GGG ATG
400
glu leu glu gly lys glu glu glu gly glu ala met val glu asp glu glu glu ala lys
GAG CTG GAA GGG AAG GAG GAA GAG GGG GAG GCG ATG GTG GAG GAT GAA GAG GAG GCT AAA
500
ile pro lys ala ala gln pro lys ser glu ser lys glu asn ala glu asp asn glu ser
ATT CCC AAG GCA GCG CAA CCC AAA TCA GAA AGC AAA GAA AAT GCT GAG GAT AAC GAA TCA
gly ser thr asp ser gly gln glu asn ser gly glu thr arg leu leu arg ser gly thr
GGG AGC ACT GAC TCT GGA CAG GAG AAC TCA GGT GAA ACG CGG CTG CTG CGT TCG GGC ACT
600
tyr ser asp arg thr glu ser lys ala ser arg ser val thr his lys cys glu asp cys
TAC AGC GAC AGG ACC GAG TCG AAA GCC TAC GCT GCC GTC ACA CAC AAG TGT GAG GAC TGC
(1)
gly lys glu phe thr his thr gly asn phe lys arg his ile arg ile his thr gly glu
GGA AAG GAG TTC ACC CAC ACT GGG AAC TTC AAG CGG CAC ATC CGC ATC CAC ACC GGC GAG
700
lys pro phe ser cys arg glu cys asn lys ala phe ser asp pro ala ala cys lys ala
AAA CCC TTC TCC TGC AGG GAG TGC AAT AAA GCC TTC TCA GAC CCA GCG GCG TGC AAA GCC
(2)
800
his glu lys thr his ser pro leu lys pro tyr gly cys glu glu cys gly lys ser tyr
CAC GAG AAG ACG CAC AGC CCG CTG AAG CCC TAC GGC TGT GAG GAG TGC GGG AAG AGC TAC
(3)
arg leu ile ser leu leu asn leu his lys lys arg his thr gly glu ala lys tyr arg
CGC CTC ATC AGC CTG CTG AAC CTG CAC AAG AAG AGG CAC ACG GGG GAG GCC AAG TAC CGC
900
cys asp asp cys gly lys leu phe thr thr ser gly asn leu lys arg his gln leu val
TGT GAC GAC TGC GGA AAG CTC TTC ACC ACC TCG GGC AAC CTC AAG CGG CAC CAA CTG GTA
(4)
his ser gly glu lys pro tyr gln cys asp tyr cys gly arg ser phe ser asp pro thr
CAC AGT GGG GAA AAG CCC TAC CAG TGC GAC TAC TGC GGG CGC TCC TTC TCT GAC CCC ACC
1000
(5)
ser lys met arg his leu glu thr his asp thr asp lys glu his lys cys pro his cys
TCC AAA ATG CGG CAC CTG GAG ACC CAC GAC ACC GAC AAG GAG CAC AAG TGT CCC CAC TGC
1100
(6)
asp lys lys phe asn gln val gly asn leu lys ala his leu lys ile his ile ala asp
GAC AAG AAA TTC AAC CAA GTG GGA AAC TTG AAA GCT CAC TTG AAG ATT CAC ATT GCG GAT
gly pro leu lys cys arg glu cys gly asn glu phe thr thr ser gly asn leu lys arg
GGG CCC CTG AAG TGT CGG GAG TGC GGC AAC GAG TTC ACC ACC TCA GGC AAC CTG AAG CGG
(7)
1200
his leu arg ile his ser gly glu lys pro tyr val cys val his cys gln arg gln phe
CAC CTC CGG ATC CAC AGC GGG GAG AAG CCC TAC GTC TGC GTG CAC TGC CAG CGA CAG TTT
(8)
ala asp pro gly ala leu gln arg his val arg ile his thr gly glu lys pro cys gln
GCT GAC CCC GGG GCG CTG CAG GCG CAC GTC CGG ATC CAC ACG GGA GAG AAG CCG TGC CAG
1300
cys leu ile cys gly lys ala phe thr gln ala ser ser leu ile ala his val arg gln
TGC CTC ATC TGC GGG AAG GCG TTC ACC CAA GCC AGC TCC CTC ATC GCC CAC GTG CGC CAG
(9)

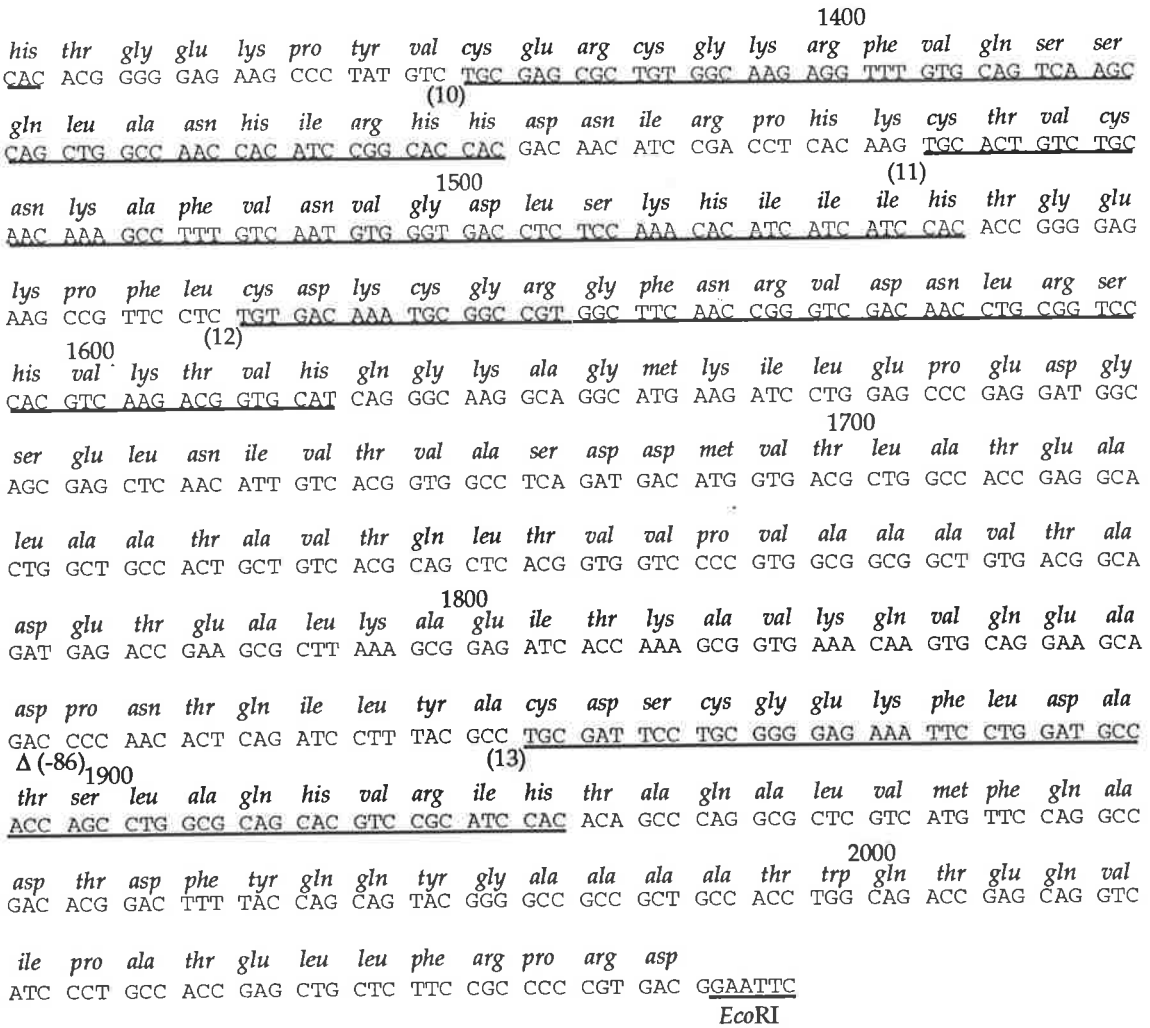


Figure 5.5 Nucleotide sequence of λ16.1 2.1

The complete nucleotide sequence of the 2.1 kb insert from the λ16.1 clone is shown along with the deduced amino acid sequence. The clone codes for an open reading frame of 2,058 bp, appears not to contain the full 5' coding region and does not contain the complete 3' coding and noncoding region. The 12 zinc fingers initially identified in the Z12 clone (λ4.1) are underlined and numbered (1) - (12) along with an additional zinc finger (13). Immediately 5' of the zinc fingers, a cluster of acidic residues (also originally identified in the Z12 clone), constituting a possible acidic activation domain, are underlined with ~ symbols. The point at which the λ16.1 2.1 5' sequence diverges from the Z12 5' sequence is indicated by a horizontal arrow. The position of a 86 bp insert found in Z12 (λ4.1) which is not in λ16.1 2.1 is indicated by (Δ (-86)).

orientation of the 0.6 kb and 2.1 kb inserts, is shown in Figure 5.6A (see Section 7.2.8 for sequences of the PCR primers). The λ16.1 clone was characterised by PCR (7.3.19) with the combinations of the primers mentioned above. A product approximately 450 bp in size was generated with the λ16.1 PCR 1 and the λgt11 PCR 1 primers (Figure 5.6B). This result suggested that the two inserts were orientated as represented in the upper alternative in Figure 5.6A.

5.4.2 Sequence analysis of λ16.1

The 0.6 kb insert from the λ16.1 clone (λ16.1 0.6) was found to contain numerous stop codons in all three reading frames. It also contained a stretch of sixteen A residues (at it's 3' end in relation to the λ16.1 2.1 kb insert) which therefore provided it with a means of being independently copied at the time of making the library. As a result of these findings and another experiment described in Section 5.7, it would seem highly unlikely that the 0.6 kb fragment represents a genuine 5' extension of the λ16.1 2.1 kb insert.

The 2.1 kb insert from the λ16.1 clone (λ16.1 2.1) codes for an open reading frame of 2,058 bp. There are a number of differences between λ16.1 2.1 and Z12. Apart from a difference in the 5' ends of the two clones (discussed later in this section), λ16.1 2.1 lacks, near its 3' end, an 86 bp region of sequence that is present in Z12. This sequence difference introduces a 13th zinc finger in λ16.1 2.1 immediately 3' of the where the 86 bp sequence occurs in Z12. The extra 86 bp sequence in Z12 puts this zinc finger out of frame in the Z12 sequence. The 13th zinc finger is separated by approximately 250 bp from the other contiguous 12 zinc fingers in λ16.1 2.1 (see Figure 5.5).

Additionally, λ16.1 2.1 lacks the 200 bp of sequence at the 3' end of Z12, and does not include a termination codon. The 3' end of λ16.1 2.1 finishes at a point just before where the termination codon was believed to exsist in Z12. However, on further examination of the Z12 sequence a sequencing error was found to have been made in

Figure 5.6 PCR aided orientation of the λ16.1 2.1 kb and 0.6 kb inserts

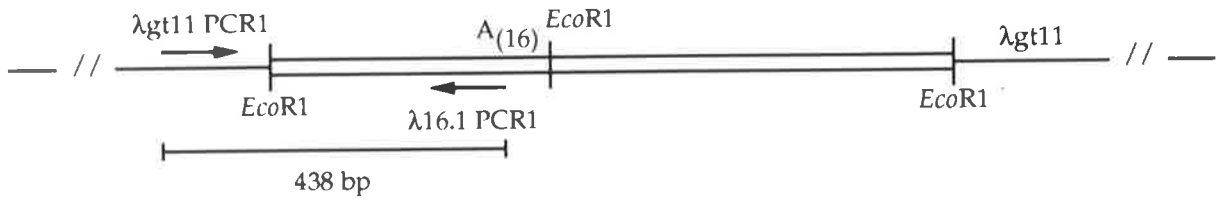
A Shown are diagramatic representations of the λ16.1 clone. The 2.1 kb and 0.6 kb *Eco*RI fragments, that constitute the full λ16.1 insert, are not drawn to scale. The 0.6 kb fragment is identified by $A_{(16)}$ representing a poly(A) sequence located at the end of the 0.6 kb fragment. The two alternative orientations of the 2.1 kb and 0.6 kb fragments, along with the combinations of PCR primers (represented by horizontal arrows; λgt11 PCR1/λ16.1 PCR1 and λ4.1 PCR1/λ16.1 PCR1) used to determine the orientation of the 2.1 kb and 0.6 kb fragments, are shown. The sizes of the PCR products expected to be generated from these two orientations are indicated.

B PCR reactions were carried out with the λ16.1 clone template (10 ng) and the different combinations of primers described in A above (7.3 19). Samples (5 μl) from each of the completed reactions along with 1 μg each of *Hpa*II digested pUC19 DNA markers and *Eco*RI digested SPP1 DNA markers were run on a 4% agarose gel and stained with ethidium bromide. Lanes are as follows;

1; *Hpa*II digested pUC19 DNA markers
2; λ16.1 template plus λgt11 PCR1/λ4.1 PCR1 primers
3; λ16.1 template plus λgt11 PCR1/λ16.1 PCR1 primers
4; λ16.1 template plus λ4.1 PCR1/λ16.1 PCR1 primers
5; *Eco*RI digested SPP1 DNA markers

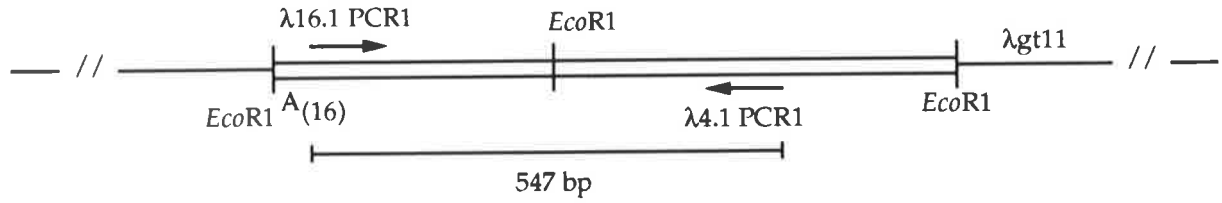
A

λ16.1 clone

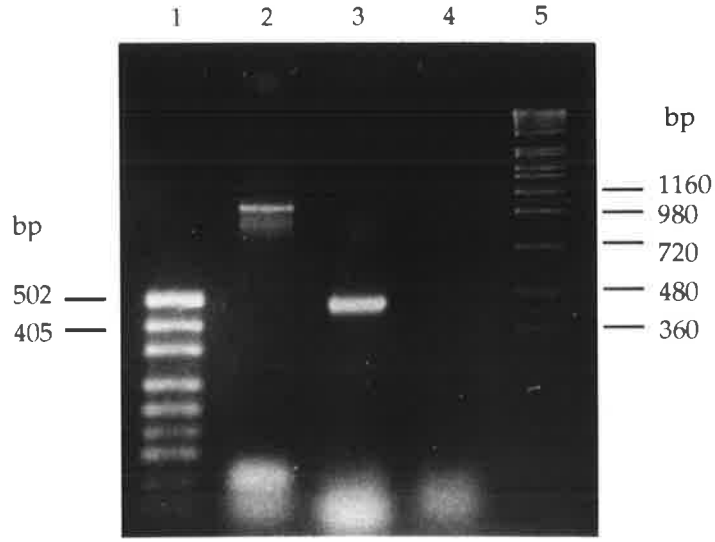


or

λ16.1 clone



B



the original λ4.1 Z12 sequence. A single base had been omitted (4 C's instead of 5 C's) just before the putative stop codon in Z12 (see Figure 4.9). On inclusion, this extra residue puts the originally designated (but incorrect) Z12 termination codon out of frame and the sequence now runs right to the end of the clone without interruption from any further termination codons (Figure 5.7A). Thus, Z12 codes for an open reading frame of 2,116 bp which can be translated to generate a potential protein product of 77,115 Da. This predicted size was now consistent with the size of the Z12 pGEX4.1 expression product, which was found to be approximately 105 kDa (i.e. 26 kDa GST + approximately 79 kDa Z12 product; see 4.4.4 (iii)).

The extra 86 bp sequence present in Z12 but absent in λ16.1 2.1 appears to represent an intron that has failed to be correctly spliced out of Z12 and has been included in the Z12 cDNA when the library was constructed. It is also possible that the 86 bp sequence difference actually represents an alternative exon splicing between Z12 and λ16.1 2.1. However a number of lines of evidence would seem to argue against this possibility and support the intron proposal. Firstly, both donor and acceptor splice site sequences have been identified coincident with the 5' and 3' junctions of the 86 bp sequence in Z12 (see Figure 5.7B). Secondly, preliminary sequence obtained from a mouse Z13 genomic clone recently isolated using the chicken Z12 cDNA as a probe (T. Schulz, this laboratory; personal communication) has shown a high degree of conservation with chicken Z12 and λ16.1 2.1. However where sequence divergence does occur, these regions appear to correspond to introns. Consistent with this, the mouse genomic clone sequence diverges from the Z12 sequence throughout the 86 bp sequence and appropriate donor and acceptor splice site sequences are present in the mouse genomic clone at the 5' and 3' junctions of the 86 bp sequence (see Figure 5.7B). The evidence presented above suggests that Z12 represents an incompletely spliced cDNA, while λ16.1 2.1 is actually the correctly processed cDNA (see 5.7 for further discussion).

Figure 5.7

A Corrected Z12 nucleotide sequence

The corrected Z12 nucleotide sequence has been abbreviated to include only the most 5' region and all of the 3' region immediately downstream of the block of 12 zinc fingers, with the bulk of the sequence ommitted represented by (— //—). The additional base added to correct the Z12 sequence is indicated by a # symbol. This extra residue puts the original stop codon in Z12 (situated 2 bases downstream; see Figure 4.9) out of frame with the result that the corrected Z12 sequence now codes for an open reading frame of 2,108 bp.

The point at which the Z12 5' sequence diverges from the λ16.1 2.1 5' sequence is indicated by a horizontal arrow, while the 86 bp region present in Z12 but absent from λ16.1 2.1 is underlined. Approximately 250 bp 3' of the zinc fingers (in this diagram 250 bp 3' of (— //—)), a large cluster of proline residues, constituting a possible proline-rich activation domain, are underlined with ~ symbols.

B Comparison of Z12 and λ16.1 2.1 splice sites surrounding an 86 bp 3' region

The sequences from the mouse Z13 genomic clone, Z12 cDNA clone and λ16.1 2.1 cDNA clone surrounding the putative splice sites of an 86 bp region identified in the Z12 cDNA clone are displayed. The consensus sequence is also displayed for both donor and acceptor splice sites (Shapiro and Senapathy, 1987; Penotti, 1991).

A

```
                                                arg  gly  leu  gly  arg  pro  val  gln  leu  gly  gly  val  asp
                                  GAATTC CGG GGG CTG GGC AGG CCA GTT CAA CTG GGA GGT GTG GAC
                                  EcoRI                                                  100
asn  ala  ala  glu  ala  ser  pro  ala  ala  val  ser  pro  ser  arg  pro  gln  pro  ala  glu  ser
AAT GCC GCC GAA GCC AGC CCT GCT GCC GTG TCC CCC AGC CGC CCG CAG CCA GCA GAG AGC

                                                1400
glu  val  ——— // ———  gln  gly  lys  ala  gly  met  lys  ile  leu  glu  pro  glu  asp  gly
GAG GTG               CAG GGC AAG GCA GGC ATG AAG ATC CTG GAG CCC GAG GAT GGC

ser  glu  leu  asn  ile  val  thr  val  ala  ser  asp  asp  met  val  thr  leu  ala  thr  glu  ala
AGC GAG CTC AAC ATT GTC ACG GTG GCC TCA GAT GAC ATG GTG ACG CTG GCC ACC GAG GCA

          1500
leu  ala  ala  thr  ala  val  thr  gln  leu  thr  val  val  pro  val  ala  ala  ala  val  thr  ala
CTG GCT GCC ACT GCT GTC ACG CAG CTC ACG GTG GTC CCC GTG GCG GCG GCT GTG ACG GCA

                                                                                1600
asp  glu  thr  glu  ala  leu  lys  ala  glu  ile  thr  lys  ala  val  lys  gln  val  gln  glu  ala
GAT GAG ACC GAA GCG CTT AAA GCG GAG ATC ACC AAA GCG GTG AAA CAA GTG CAG GAA GCA

gly  glu  glu  gly  gly  gly  pro  ala  ser  arg  ser  pro  gly  pro  arg  pro  ala  ala  ala  thr
GGT GAG GAG GGG AGG GGG CCG GCA TCG CGC TCA CCT GGC CCC CGC CCT GCT GCC GCC ACA

                                                1700
pro  arg  ala  ser  pro  leu  pro  leu  gln  thr  pro  thr  leu  arg  ser  phe  thr  pro  ala  ile
CCC CGG GCC TCC CCT CTG CCC TTG CAG ACC CCA ACA CTC AGA TCC TTT ACG CCT GCG ATT

pro  ala  gly  arg  asn  ser  trp  met  pro  pro  ala  trp  arg  ser  thr  ser  ala  ser  thr  gln
CCT GCG GGG AGA AAT TCC TGG ATG CCA CCA GCC TGG CGC AGC ACG TCC GCA TCC ACA CAG

          1800
pro  arg  arg  ser  ser  cys  ser  arg  pro  thr  arg  thr  phe  thr  ser  ser  thr  gly  pro  pro
CCC AGG CGC TCG TCA TGT TCC AGG CCG ACA CGG ACT TTT ACC AGC AGT ACG GGG CCG CCG

                                                                      1900      #
leu  pro  pro  gly  arg  pro  ser  arg  ser  ser  leu  pro  pro  ser  cys  ser  ser  ala  pro  val
CTG CCA CCT GGC AGA CCG AGC AGG TCA TCC CTG CCA CCG AGC TGC TCT TCC GCC CCC GTG

thr  ala  pro  lys  arg  pro  pro  pro  pro  pro  trp  pro  leu  cys  pro  leu  leu  gly  arg  ala
ACA GCC CCC AAG AGG CCC CCG CCG CCC CCC TGG CCC CTG TGC CCC TTG CTG GGG AGG GCC

                                                2000
arg  arg  leu  leu  ser  asp  pro  pro  ala  pro  pro  arg  leu  tyr  leu  lys  ser  ala  ala  trp
AGG CGC CTG CTG AGT GAC CCA CCC GCT CCC CCT CGC CTT TAT TTA AAG TCA GCT GCT TGG

lys  arg  cys  lys  lys  cys  ile  ser  gly  arg  arg  glu  arg  arg  asn  lys  leu  gln  tyr  phe
AAA CGG TGT AAA AAG TGT ATT TCT GGA AGG AGA GAG AGA AGG AAT AAA TTG CAA TAT TTT

          2100
leu  lys  lys  lys  lys  asn
CTA AAA AAA AAA AAA AAC GGAATTC
                       EcoRI
```

B

	Donor (5' boundary)	Acceptor (3' boundary)
Consensus	$^{C}_{A}$AG | gt$^{a}_{g}$agt	$(^{c}_{t})_{n}$ n $^{c}_{t}$ag | G
Mouse Z13 genomic clone	AAG | gtgaga	acag | ACC
Z12 cDNA	CAG | GTGAGG	GCAG | ACC
λ16.1 2.1 cDNA	CAG | ACC	

The Z12 and λ16.1 2.1 sequences were also found to diverge at the 5' end (see Figures 5.7A and 5.5 for comparison). The different 5' regions coded for by the two clones may, in this case be caused by an alternative splicing event, rather than representing the failure of another intron to be removed from the Z12 cDNA sequence . Further investigation is required to determine which of these two possibilities is correct (no genomic sequence for this region was available at this time). However, the Z12 sequence immediately 5' to the putative splice site does not match the 3' acceptor splice site consensus which suggests that the Z12 5' region does actually represent coding sequence and not an intron (see Figure 5.7B).

Shown in Figure 5.8A is a composite sequence (designated Z13 for future reference) combining the sequence of the λ16.1 2.1 clone and the extra 3' sequence from the corrected Z12 clone. The Z12 and λ16.1 2.1 sequence were assembled in such a contiguous manner as the mouse Z13 genomic sequence revealed that no further introns were present in this 3' region (i.e. 3' of the 13^{th} zinc finger until the stop codon). A diagramatic representation of the respective protein structures of the Z12 and Z13 clones is shown in Figure 5.8B. The deduced amino acid sequences for the Z13 and corrected Z12 clones were compared to the GENPEPTIDE protein database (July 1992) and SWISSPROT protein database (July 1992) using the FASTA, MAIL-BLAST and MAIL-DAP search programs (Devereux *et al.*, 1984). However no other regions of significant amino acid homology were identified apart from the block of 12 zinc fingers originally identified with the Z12 search.

5.5 GENOMIC LIBRARY SCREENING

A λ Charon 4a chicken genomic library (5×10^5 independent clones, 10 - 20 kb in size; Dodgson *et al.*, 1979) was screened to try to isolate the Z5 and Z12 genes (7.3.16). The number of independent clones required in the library to have at least a 99% probability of containing any sequence present in the genome (chicken genome approximately 2×10^9 bp; review, Old and Woodland, 1984) is approximately 5×10^5 (Clarke and Carbon, 1976).

Figure 5.8

A Z13 nucleotide sequence

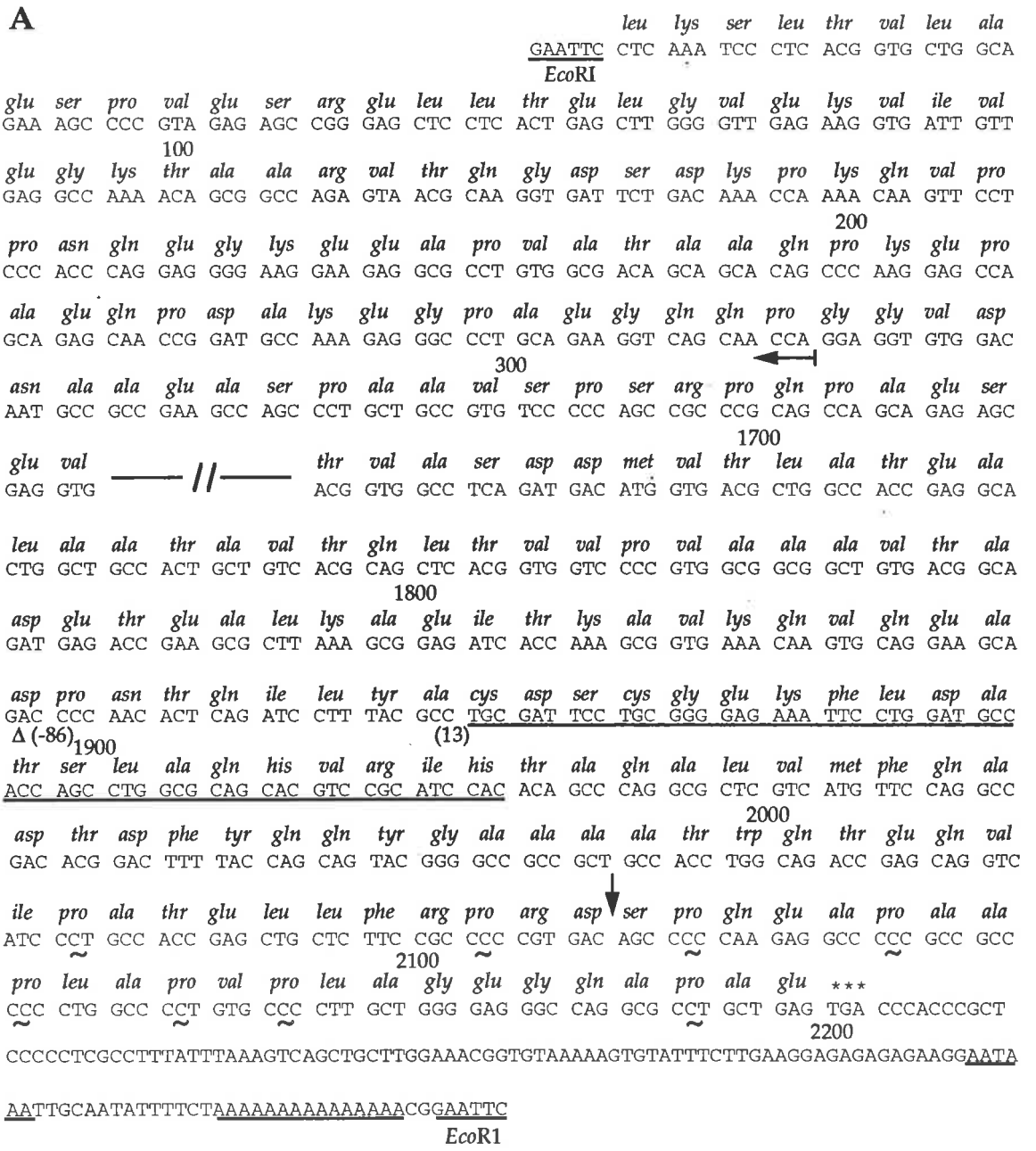
Shown is a composite sequence, designated Z13, combining the sequence of the λ16.1 2.1 clone and the extra 3' sequence information from the corrected Z12 clone is shown. The Z13 nucleotide sequence has been abbreviated to include only the 5' most 350 bp and all of the 3' region from a point approximately 50 bp downstream of the block of 12 zinc fingers, with the bulk of the sequence ommitted represented is by (— //—). The point at which the λ16.1 2.1 and corrected Z12 sequence have been combined is indicated by a vertical arrow.

A region, located immediately 5' of the stop codon, although considerably smaller in size and containing many fewer proline residues (underlined with ~ symbols), compared to the proline-rich region identified in the corrected Z12 sequence (see Figure 5.7A), may still represent a possible proline-rich activation domain.

The point at which the λ16.1 2.1 5' sequence diverges from the Z12 5' sequence is indicated by a horizontal arrow. Additionally where the extra 86 bp region present in Z12 is positioned relative to the λ16.1 2.1 sequence is indicated by (Δ (-86)). The 13th zinc finger encoded by the λ16.1 2.1 clone is underlined and numbered (13), while the poly(A) tract and polyadenylation signal (AAUAAA) sequences are underlined in the 3' noncoding region.

B Diagramatic representation of Z12 and Z13 protein structures

Zinc finger modules are represented as solid boxes and non-finger coding region represented by open boxes. Where the amino acid sequences of Z12 and Z13 diverge at their NH_2 termini are indicated by vertical arrows. The proposed 86 bp intronic sequence in Z12 is represented by a hatched box. This sequence actually forms part of the coding region of the COOH terminal non-finger coding portion of the Z12 protein. The position of this 86 bp sequence with respect to the Z13 coding region is indicated by (Δ (-86 bp). The amino acid sequences of Z12 and Z13 diverge from the point of the 86 bp insertion in Z12.


A
leu lys ser leu thr val leu ala
GAATTC CTC AAA TCC CTC ACG GTG CTG GCA
EcoRI
glu ser pro val glu ser arg glu leu leu thr glu leu gly val glu lys val ile val
GAA AGC CCC GTA GAG AGC CGG GAG CTC CTC ACT GAG CTT GGG GTT GAG AAG GTG ATT GTT
100
glu gly lys thr ala ala arg val thr gln gly asp ser asp lys pro lys gln val pro
GAG GCC AAA ACA GCG GCC AGA GTA ACG CAA GGT GAT TCT GAC AAA CCA AAA CAA GTT CCT
200
pro asn gln glu gly lys glu glu ala pro val ala thr ala ala gln pro lys glu pro
CCC ACC CAG GAG GGG AAG GAA GAG GCG CCT GTG GCG ACA GCA GCA CAG CCC AAG GAG CCA
ala glu gln pro asp ala lys glu gly pro ala glu gly gln gln pro gly gly val asp
GCA GAG CAA CCG GAT GCC AAA GAG GGC CCT GCA GAA GGT CAG CAA CCA GGA GGT GTG GAC
300
asn ala ala glu ala ser pro ala ala val ser pro ser arg pro gln pro ala glu ser
AAT GCC GCC GAA GCC AGC CCT GCT GCC GTG TCC CCC AGC CGC CCG CAG CCA GCA GAG AGC
1700
glu val
GAG GTG
thr val ala ser asp asp met val thr leu ala thr glu ala
ACG GTG GCC TCA GAT GAC ATG GTG ACG CTG GCC ACC GAG GCA
leu ala ala thr ala val thr gln leu thr val val pro val ala ala ala val thr ala
CTG GCT GCC ACT GCT GTC ACG CAG CTC ACG GTG GTC CCC GTG GCG GCG GCT GTG ACG GCA
1800
asp glu thr glu ala leu lys ala glu ile thr lys ala val lys gln val gln glu ala
GAT GAG ACC GAA GCG CTT AAA GCG GAG ATC ACC AAA GCG GTG AAA CAA GTG CAG GAA GCA
asp pro asn thr gln ile leu tyr ala cys asp ser cys gly glu lys phe leu asp ala
GAC CCC AAC ACT CAG ATC CTT TAC GCC TGC GAT TCC TGC GGG GAG AAA TTC CTG GAT GCC
Δ (-86)
1900
(13)
thr ser leu ala gln his val arg ile his thr ala gln ala leu val met phe gln ala
ACC AGC CTG GCG CAG CAC GTC CGC ATC CAC ACA GCC CAG GCG CTC GTC ATG TTC CAG GCC
2000
asp thr asp phe tyr gln gln tyr gly ala ala ala ala thr trp gln thr glu gln val
GAC ACG GAC TTT TAC CAG CAG TAC GGG GCC GCC GCT GCC ACC TGG CAG ACC GAG CAG GTC
ile pro ala thr glu leu leu phe arg pro arg asp ser pro gln glu ala pro ala ala
ATC CCT GCC ACC GAG CTG CTC TTC CGC CCC CGT GAC AGC CCC CAA GAG GCC CCC GCC GCC
2100
pro leu ala pro val pro leu ala gly glu gly gln ala pro ala glu ***
CCC CTG GCC CCT GTG CCC CTT GCT GGG GAG GGC CAG GCG CCT GCT GAG TGA CCCACCCGCT
2200
CCCCCTCGCCTTTATTTAAAGTCAGCTGCTTGGAAACGGTGTAAAAAGTGTATTTCTTGAAGGAGAGAGAGAAGGAATA
AATTGCAATATTTTCTAAAAAAAAAAAAAAACGGAATTC
EcoR1
B
Z12
NH2
COOH
Z13
NH2
COOH
-86 bp

5.5.1 Probe preparation

The 1.4 kb Z5 cDNA and 2.1 kb Z12 cDNA fragments, after being oligo-labelled with α-^{32}P-dATP (7.3.4 (iii)), were used as probes to screen for the Z5 and Z12 genes respectively.

5.5.2 Z5 and Z12 genomic library screening

The genomic library was absorbed to *E.coli* strain LE392, and plated onto twenty 15 cm L-agar + 50 μg/ml ampicillin plates with approximately 50,000 plaques per plate. The λ phage were transferred in duplicate onto Plaquescreen filters and processed as described in Section 7.3.16. The filters were first probed with the Z5 probe, then stripped and probed with the Z12 probe. A total of seven Z5 and four Z12 first round duplicate positives were identified. Representative Z5 and Z12 first round duplicate positives (λG2.1 and λG7.1) are shown in Figures 5.9A and 5.9B respectively. Positive plaques were picked into 1ml PSB, the phage allowed to elute overnight at 4°C, and subjected to a further two rounds of screening to purify the clones to homogeneity.

5.6 ANALYSIS OF λG2.1 AND λG7.1 CLONES

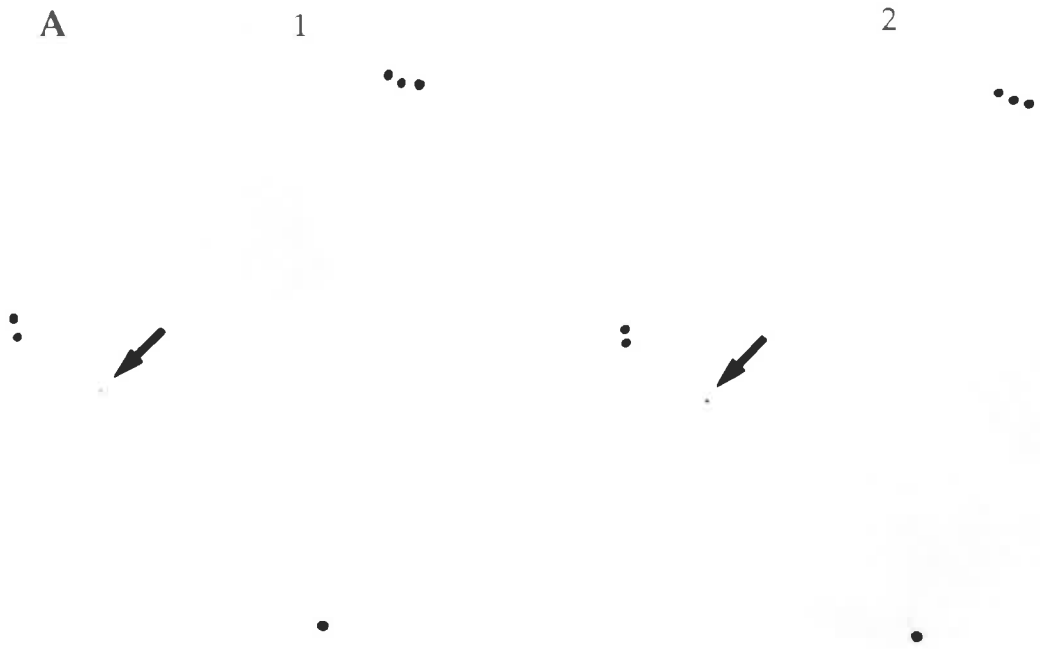
λ DNA was prepared from the seven Z5 clones (λG2.1, λG3.1, λG10.1, λG11.1, λG14.1, λG18.1, and λG20.1) and four Z12 clones (λG6.1, λG7.1, λG16.1, and λG16.2) (7.3.17). Digestion with *Eco*RI revealed that there were actually only five different Z5 genomic clones (λG2.1, λG3.1, λG10.1 were identical) and one distinct Z12 genomic clone (λG6.1, λG7.1, λG16.1, and λG16.2 were identical). Southern analysis of the *Eco*RI digested Z5 and Z12 genomic clones was carried out with the 1.4 kb Z5 and 2.2 kb Z12 cDNA probes respectively (7.3.5) (Figures 5.10A and 5.10B).

A 5.2 kb *Eco*RI fragment from the λG2.1 Z5 clone was found to hybridise with the Z5 cDNA probe. In addition as this fragment was shown to hybridise with both 5' (380 bp *Eco*RI/*Sca*I restriction fragment) and 3' (292 bp *Hin*dIII/*Eco*RI restriction fragment; see Figure 4.8B) Z5 cDNA probes, it may contain the entire Z5 gene.

Figure 5.9 Z5 and Z12 genomic library screening first round duplicate positives
Shown are examples of Z5 and Z12 genomic first round duplicate positives, λG2.1 and λG7.1, identified as a result of screening a λCharon 4a chicken genomic library with ^{32}P-labelled 1.4 kb Z5 and 2.1 kb Z12 cDNA fragments (7.3.16). Filters are as follows;

A 1 and 2; First and second duplicate filter lifts with the autoradiographic signal representing the λG2.1 plaque indicated on each filter by an arrow.

B 1 and 2; First and second duplicate filter lifts with the autoradiographic signal representing the λG7.1 plaque indicated on each filter by an arrow.


A
1
2

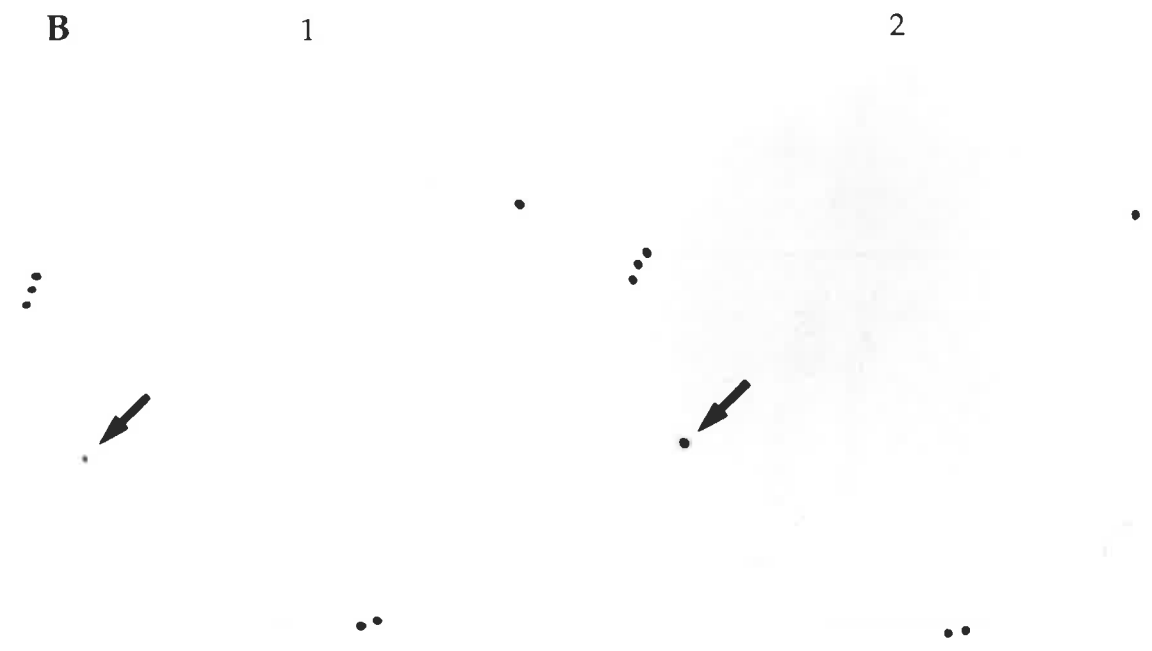

B
1
2

Figure 5.10 Southern analysis of Z5 and Z12 genomic clones

A λ DNA (~ 2 μg) from each of the λG7.1, λG2.1, λG10.1, λG11.1, λG14.1 and λG20.1 clones were digested with *Eco*RI and along with DNA molecular size markers separated by gel electrophoresis on a 1% agarose gel. DNA was visualised by staining with ethidium bromide. Lanes are as follows;

1; *Hin*dIII digested λ DNA markers

2; λG7.1

3; *Eco*RI digested SPP1 phage markers

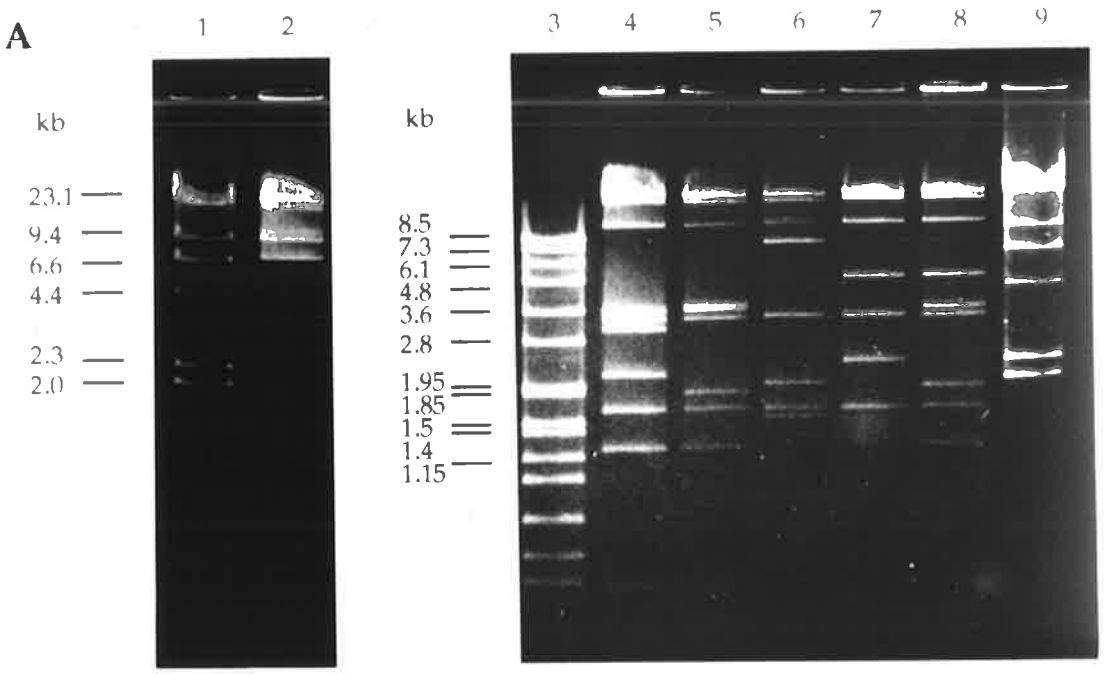
4 - 8; λG2.1, λG10.1, λG11.1, λG14.1 and λG20.1

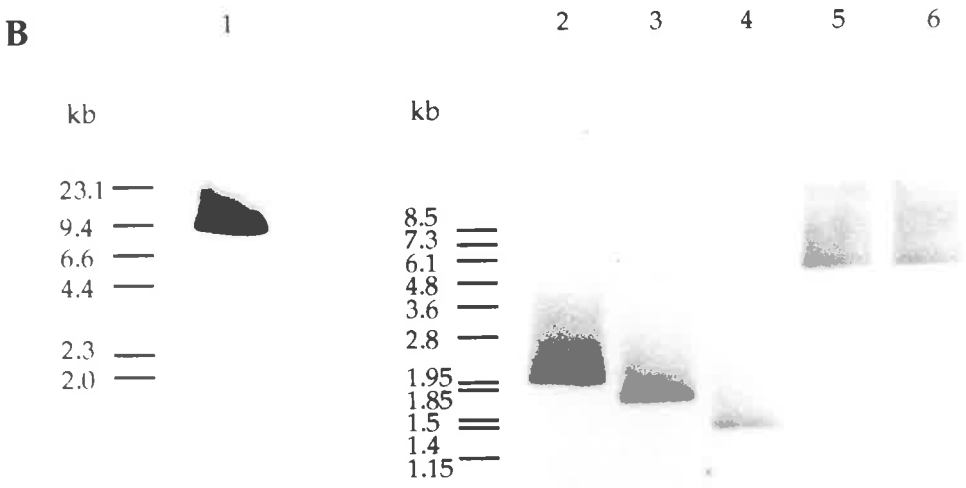
9; *Hin*dIII digested λ DNA markers

B DNA was transferred from the above gel to a Genescreen membrane (7.3.5), probed with either the ^{32}P-labelled 2.1 kb Z12 cDNA fragment (λG7.1) or the ^{32}P-labelled 1.4 kb Z5 cDNA fragment(λG2.1, λG10.1, λG11.1, λG14.1 and λG20.1) and hybridizing bands detected by autoradiography. Sizes of the markers are shown. Lanes are as follows;

1; λG7.1

2 - 6; λG2.1, λG10.1, λG11.1, λG14.1 and λG20.1


A
1
2
3
4
5
6
7
8
9
kb
23.1
9.4
6.6
4.4
2.3
2.0
kb
8.5
7.3
6.1
4.8
3.6
2.8
1.95
1.85
1.5
1.4
1.15


B
1
2
3
4
5
6
kb
23.1
9.4
6.6
4.4
2.3
2.0
kb
8.5
7.3
6.1
4.8
3.6
2.8
1.95
1.85
1.5
1.4
1.15

Whereas, the other four Z5 genomic clones (λG10.1, λG11.1, λG14.1, and λG20.1) did not hybridize with the 5' Z5 cDNA probe suggesting that none possessed the 5' end of the Z5 gene. No further work was done with these clones.

A 9.0 kb *Eco*RI fragment from the λG7.1 Z12 clone was found to hybridise with the Z12 cDNA probe. However, this fragment only hybridised with a 5' (367 bp *Eco*RI/*Tth*III1 restriction fragment) Z12 cDNA probe and did not hybridise with a 3' (350 bp *Bst*XI/*Eco*RI restriction fragment; see Figure 4.8A) Z12 cDNA probe. This suggested that the λG7.1 Z12 clone did not possess the 3' end of the Z12 gene.

5.6.1 Subcloning of λG2.1 and λG7.1 into pBSSK+

To make further manipulations of the λG2.1 and λG7.1 clones easier, DNA from each was digested with *Eco*RI, separated by agarose gel electrophoresis and the 5.2 kb and 9.0 kb fragments isolated from the gel (7.3.3 (i)). The purified 5.2 kb and 9.0 kb fragments were then ligated into *Eco*RI digested pBluescript SK+. The ligation reactions were transformed into *E.coli* strain DH5α, and colonies containing recombinants identified by colour selection (7.3.3). Plasmid DNA was prepared from recombinants as described in section (7.3.1 (ii)). The DNA samples were digested with *Eco*RI to confirm the presence and sizes of the inserts, and analysed by agarose gel electrophoresis (7.3.2). Plasmids containing the λG2.1 5.2 kb or λG7.1 9.0 kb inserts (designated pBSG2.1 and pBSG7.1 respectively) were identified and large scale plasmid preparations of each produced (7.3.1 (i)).

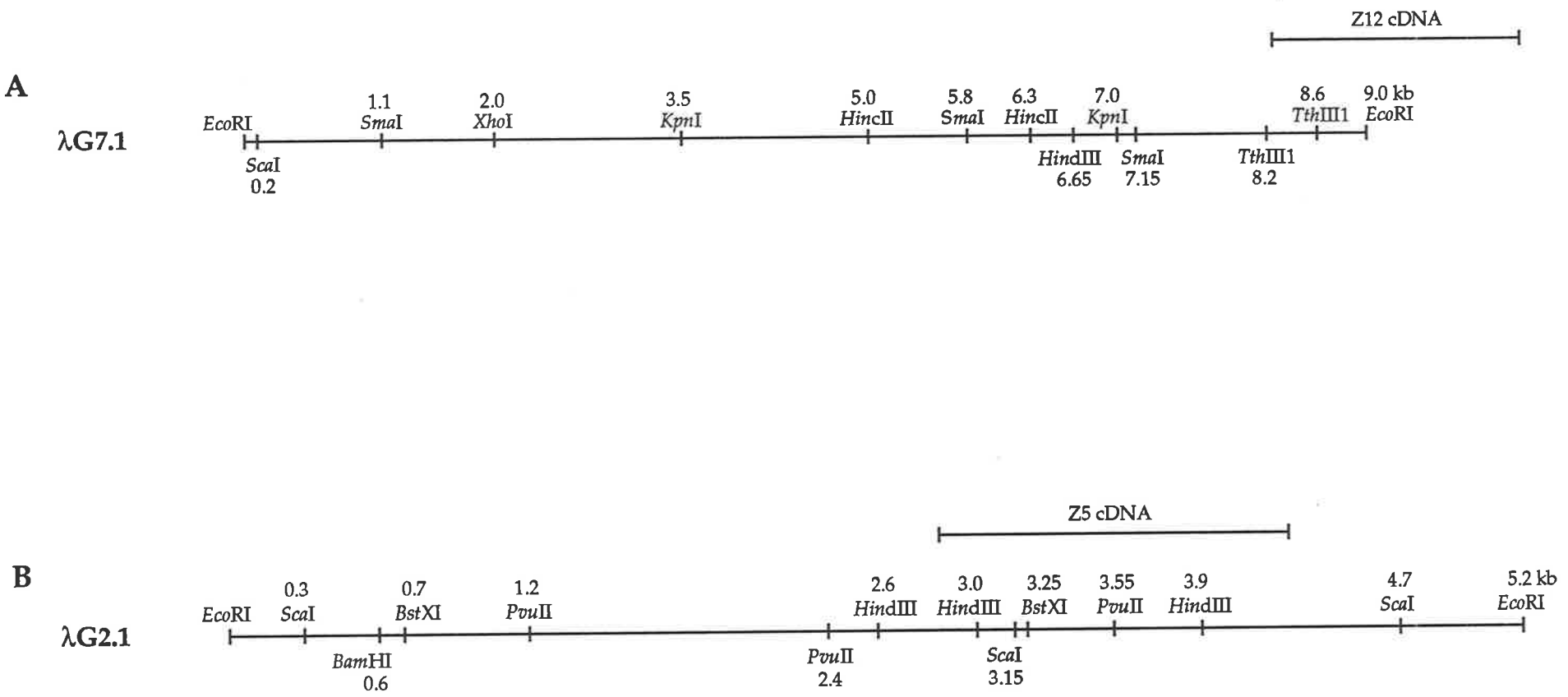
(i) Restriction endonuclease mapping

The two pBluescript clones, pBSG2.1 and pBSG7.1, were digested with various combinations of restriction enzymes. From these digestions and subsequent Southern analysis with both 5' and 3' Z5 and Z12 cDNA probes, restriction maps were constructed for the λG2.1 5.2 kb and λG7.1 9.0 kb fragments (Figures 5.11A and 5.11B).

Figure 5.11 Restriction maps of λG7.1 and λG2.1 genomic clones

A The 9.0 kb *Eco*RI insert from λG7.1 was cloned in into *Eco*RI digested pBluescript SK+ to produce a construct designated pBSG7.1 (see 5.6.1). A series of restriction enzyme digestions of the pBSG7.1 construct and subsequent Southern analysis with both 5' and 3' Z12 cDNA probes enabled a restriction map for the λG7.1 9.0 kb insert to be generated. The positions of internal restriction sites are indicated, and numbered as measured from the 5' end of the insert. The relative position of the Z12 cDNA sequence to the genomic fragment is also indicated.

B The 5.2 kb *Eco*RI insert from λG2.1 was cloned in into *Eco*RI digested pBluescript SK+ to produce a construct designated pBSG2.1 (see 5.6.1). A series of restriction enzyme digestions of the pBSG2.1 construct and subsequent Southern analysis with both 5' and 3' Z5 cDNA probes enabled a restriction map for the λG2.1 5.2 kb insert to be generated. The positions of internal restriction sites are indicated, and numbered as measured from the 5' end of the fragment. The relative position of the Z5 cDNA sequence to the genomic fragment is also indicated.


A
λG7.1
Z12 cDNA
EcoRI
1.1
SmaI
2.0
XhoI
3.5
KpnI
5.0
HincII
5.8
SmaI
6.3
HincII
7.0
KpnI
8.6
TthIII1
9.0 kb
EcoRI
ScaI
0.2
HindIII
6.65
SmaI
7.15
TthIII1
8.2
B
λG2.1
Z5 cDNA
EcoRI
0.3
ScaI
0.7
BstXI
1.2
PvuII
2.6
HindIII
3.0
HindIII
3.25
BstXI
3.55
PvuII
3.9
HindIII
4.7
ScaI
5.2 kb
EcoRI
BamHI
0.6
PvuII
2.4
ScaI
3.15

A set of subclones of pBSG2.1 were constructed to make the analysis of pBSG2.1 more managable. The pBSG2.1 clone was digested with *Hin*dIII and *Eco*RI generating a set of 0.4, 0.9, 1.3 and 2.6 kb fragments, which were separated by agarose gel electrophoresis and isolated from the gel (7.3.3 (i)). The purified 0.4, 0.9 and 1.3 kb DNA fragments were ligated into *Hin*dIII digested pBluescript SK+ and the 2.6 kb fragment ligated into *Hin*dIII/*Eco*RI digested pBluescript SK+. Ligation reactions were transformed into *E.coli* strain DH5α and colonies containing recombinants identified by colour selection (7.3.3). Plasmid DNA was prepared from recombinants as described in section 7.3.1 (ii). The DNA samples were digested with *Hin*dIII (0.4, 0.9 and 1.3 kb) and *Hin*dIII/*Eco*RI (2.6 kb), to confirm the presence and sizes of the inserts, and analysed by agarose gel electrophoresis (7.3.2).

(ii) Preliminary dideoxy sequencing of pBSG2.1 subclones and pBSG7.1

The 5' and 3' ends of pBSG2.1, the four pBSG2.1 subclones and pBSG7.1 were sequenced (Sanger *et al.*, 1977; 7.3.21). Approximately 250 - 300 bp of DNA sequence, in each case, was obtained (data not shown). This enabled confirmation that the λG2.1 and λG7.1 clones were indeed genomic clones for Z5 and Z12 repectively, and permitted the accurate alignment of the Z5 and Z12 cDNA sequences with the pBSG2.1 and pBSG7.1 sequences (see Figures 5.11A and 5.11B).

5.7 DISCUSSION

This chapter describes the efforts to isolate full length cDNA clones for both Z5 and Z12 and also genomic clones containing the Z5 and Z12 genes. As discussed in 4.4.2 and 4.4.3 both the Z5 and Z12 clones appeared to be incomplete at the 5' end. Consistent with this, Northern analysis (see 5.2) revealed that the Z12 mRNA was approximately 3.0 kb in length. The Z5 mRNA was unable to be detected by Northern analysis. Therefore both the Z5 mRNA and Z12 mRNA represent very low abundance messages. This is consistent with their possible roles as transcription factors, since these factors are usually present in vanishingly small quantities, of the order of 0.001% of the total cellular protein (Kadonaga and Tijan, 1986).

A ten day chicken embryo cDNA λgt11 library was screened, with Z5 and Z12 probes resulting in three Z5 and eight Z12 duplicate positives being isolated (see 5.3.2). Only one of the candidate Z12 clones (λ16.1), was found to contain additional 5' sequence (see below for discussion). All three Z5 clones appreared to be identical to the original Z5 clone isolated by oligoscreening, while five of the eight Z12 clones seemed to be identical to the original Z12. Since the cDNA library was obtained commercially, it is likely that the library has been amplified. This may explain why multiple copies of the same clone are present in the library. However the numbers of Z5 and Z12 clones isolated are what would be expected for low abundance mRNAs, and in fact may be be totally representative of the relative abundance of the mRNAs coding for the proteins (e.g. eight Z12 clones isolated from approximately 2×10^6 clones corresponds to a relative abundance of approximately 1:250,000). Further, mRNA secondary structure, peculiar to either of the Z5 or Z12 clones, may have blocked complete first strand cDNA synthesis at particular positions in the sequences which could explain the multiple copies of identical incomplete clones. Screening a randomly primed or 'stretch' cDNA library may be useful in overcoming such a problem. Alternatively, a PCR based approach to isolate the remaining 5' sequences of the Z5 and Z12 clones may be more successful than library screening considering the low abundance of the mRNAs. PCR amplification from a cDNA library as described by Friedman *et al.* (1988) or anchored PCR as described by Frohman *et al.* (1988) from first strand cDNA generated from chicken embryo RNA could be attempted.

The λ16.1 clone, isolated from the ten day chicken cDNA λgt11 library with the 5' Z12 probe, was found to contain 0.6 kb and 2.1 kb *Eco*RI fragments. The 2.1 kb fragment contained sequence overlapping the original Z12 clone, whereas, the 0.6 kb *Eco*RI fragment appeared to represent an unrelated insert that presumably was ligated with the 2.1 kb insert when the inserts were packaged during library construction. This explaination that the 0.6 kb fragment was unlikely to be legitimately associated with the λ16.1 2.1 kb is further supported by the observation

that the 0.6 kb fragment did not hybridise with the 9.0 kb fragment from the Z12 genomic clone (λG7.1) shown to contain the 5' portion of the Z12 gene (data not shown).

Although the λ16.1 2.1 kb insert (λ16.1 2.1) was a similar size to the Z12 clone and displayed considerable sequence colinearity with the Z12 clone, two significant sequence differences between λ16.1 2.1 and Z12 were identified (see 5.4.2). The first is a 86 bp insert in Z12 that is absent from λ16.1 2.1 and causes a frame shift to generate an extra zinc finger in λ16.1 2.1. Second, additional sequence divergence between Z12 and λ16.1 2.1 was observed at their 5' ends and this will be discussed further below.

As discussed in 5.4.2, the 86 bp sequence may represent intronic sequence that has failed to be removed from the Z12 cDNA clone, rather than representing an alternatively spliced region of coding sequence. That is, the 2.1 kb insert in the λ16.1, clone was likely to be the correct cDNA sequence at this position. However it is possible that the inclusion of intronic sequence in Z12 a genuine product of the Z12 gene.

Alternative splicing generating proteins with slightly different binding specificities, has recently been observered in the regulation of other zinc finger proteins. Examples involving elimination of entire zinc fingers (Zarkower and Hodgkin, 1992) and also insertion of additional sequence between zinc fingers (Bickmore *et al.*, 1992) have been described. However, in both of these cases no shift of reading frame was involved. Although not involving an intronic sequence, a similar strategy to that possibly occurring with Z12 and λ16.1 2.1 clones to generate transcript diversity by shifting zinc finger(s) out of frame, has been described by Hsu *et al.* (1992) with the *Drosophila* CF2 transcription factor. In this example, splicing omits an exon that puts the coding sequence downstream of this splice site out of frame such that two zinc fingers are lost and an increased open reading frame, beyond the termination codons

of smaller alternative CF2 transcripts, is formed. The Z12 clone (see Figure 5.7A) with the 86 bp insert, codes for an open reading frame that runs through the termination codon for λ16.1 2.1. However as the open reading frame also includes a correctly positioned polyadenylation signal sequence (relative to the termination codon in the λ16.1 2.1 sequence) and also a poly (A)+ tract, it would seem unlikely that the Z12 transcript represents a correctly processed alternative transcript. The question of whether or not the 86 bp sequence does form part of a correctly processed Z12 mRNA could be resolved by carrying out Northern analysis using the 86 bp sequence to probe chicken embryo mRNA, or using specifically designed primers to PCR amplify across the 86 bp region, to determine if it is present, in chicken embryo mRNA.

A further observation is that the 86 bp insert in Z12 produces a large proline-rich region associated with and immediately downstream of the 86 bp insertion (see Figure 5.7A). This region may represent a proline-rich activation domain (see 1.6.2 (ii)). Whereas, in Z13 (see Figure 5.8A) a smaller proline-rich region is formed in the 3' region of the sequence, immediately upstream of the termination codon. The relative functional significance of this smaller proline-rich region is unknown.

The different 5' regions coded for by the two clones (see Figures 5.5 and 5.7A for comparison) may actually be caused by an alternative splicing event that involves coding region in Z12, rather than the non-removal of intronic sequence from the Z12 cDNA. A further possibility is that the two clones may be encoded by different genes. Full genomic sequence from the chicken Z12 genomic clone (λG7.1) covering this region should help to clarify this matter. Southern analysis of *Eco*RI digested chicken genomic DNA with the Z12 cDNA probe detected a single band (~ 11.0 kb) suggesting a single gene (data not shown). However, Southern analysis with probes for the different 5' regions of the Z12 and λ16.1 2.1 cDNA clones, and also a Z12 genomic clone containing flanking noncoding sequence, will need to be carried out to confirm whether or not Z12 and λ16.1 2.1 are encoded by a single gene.

Arguing against the existence of alternatively spliced transcripts was the finding that only one Z12 transcript was detected by Northern analysis. However the alternate transcripts may have similar size, or perhaps because of their extremely low abundance, additional transcripts may not have been detected.

The 13th zinc finger in λ16.1 2.1 is separated by approximately 250 bp from the other contiguous 12 zinc fingers (see Figure 5.5). How this extra isolated finger may influence DNA binding specificity is of some interest. Expression of λ16.1 2.1 as a pGEX-1 expression product may allow the binding specificity of the Z13 clone to be examined.

Screening of a chicken genomic library for Z5 and Z12 was also carried out. Genomic clones containing Z5 and Z12 gene sequence were isolated and preliminary characteristion of these clones commenced. A 5.2 kb fragment from the Z5 genomic clone, λG2.1, would appear to contain the entire Z5 gene considering that the Z5 cDNA sequence has been mapped within the fragment with approximately 2.5 kb 5' and 1.0 kb 3' flanking sequence either side. However, as the Z5 cDNA is not complete, and the size of the full length Z5 cDNA is unknown, it is possible that this fragment may not contain the entire gene. If this is the case, it would be hoped that other *Eco*RI fragments from the λG2.1 clone would contain the necessary sequence to be able to fully characterise the Z5 gene.

A *Eco*RI 9.0 kb fragment from the Z12 genomic clone (λG7.1) was found to contain 5' cDNA sequence. However, it did not contain Z12 cDNA 3' downstream sequence from the fifth zinc finger in the Z12 cDNA. This was apparent after approximately 300 bp of sequence was determined from the 3' end of the 9.0 kb λG7.1 fragment. When compared to the Z12 cDNA clone, the immediate sequence in from the 3' end of the 9.0 kb λG7.1 fragment was found to code for an intron that eventually (~ 75 bp upstream) ran into coding sequence for the fifth zinc finger.

The *Eco*RI retriction site at the 3' end of the 9.0 kb λG7.1 fragment cannot represent a genuine *Eco*RI site in the intron but rather, to be consistent with the preliminary Southern analysis (see above), represent an *Eco*RI site introduced at the end of the fragment during construction of the library (see Dodgson *et al.*, 1979 and reference therein; Maniatis *et al.*, 1978). Otherwise, at least two bands would have been detected in the Southern analysis of *Eco*RI digested chicken genomic DNA with the Z12 cDNA probe.

The quality of the λ Charon 4a chicken genomic library used in this work is questionable having been amplified a number of times, but may still contain a complete Z12 genomic clone. However, the best approach may be to screen a more recently constructed chicken genomic library.

CHAPTER 6

FINAL DISCUSSION

6.1 DISCUSSION

The primary aim of work presented in this thesis was to isolate a chicken *trans*-acting factor (H1-SF) that has been reported to bind a histone H1 specific promoter element called the H1 box. Previous work had indicated that the interaction between H1-SF and the H1 box modulated cell cycle control of histone H1 gene transcription (Dalton and Wells, 1988*a*). Isolation and characterisation of H1-SF would contribute to a better understanding of the molecular mechanisms which control transcription of histone H1 genes during the cell cycle. Furthermore, elucidation of how H1-SF is regulated may help move one step closer to an understanding of how coordinate expression of the different classes of histone genes is achieved, and perhaps how histone gene transcription is linked to the process of DNA replication.

The H1 box is not the only regulatory element involved in cell cycle control of histone H1 gene transcription. Another element, the H1-CCAAT box, which binds the factor, H1TF2, has also been found to be involved in this process (Gallinari *et al.*, 1989; La Bella *et al.*, 1989). Furthermore, the conserved spatial arrangement of the chicken histone H1 proximal promoter elements, the H1 box, H1-CCAAT box and also a G/C box (potential Sp1 binding site) with respect to the TATA box (see Figure 1.3) suggests that direct interactions between the bound complexes at these sites are critical for appropriate transcriptional regulation of histone H1 genes. Although the proximal promoter elements appear to be the major components in the regulation of the histone H1 genes, regulatory regions further upstream of these proximal promoter elements have recently been observed to also influence histone H1 transcription (Kremer and Kistler, 1992). These regions have not been as well characterised but have been reported to confer some degree of cell cycle control on H1 transcription. Therefore, although the H1 box appears to be a critical element to cell cycle control of H1 transcription it should be recognised that it functions within an complex regulatory framework.

Although the H1 box is not found in the promoters of any of the other classes of histone genes (the other classes of histone genes possess different elements that regulate their cell cycle expression; see 1.11.1 (ii)), the H1 box is not totally histone H1 gene specific. For instance the AAACACA sequence (core sequence of the H1 box) has been identified in the promoter of the cell cycle controlled dihydrofolate reductase (DHFR) gene (Farnham and Schimke, 1986) where it may play a similar regulatory role in controlling transcription of this gene as it does for the histone H1 gene. DHFR is a key enzyme in nucleotide metabolism and is thus involved in DNA replication. Additionally, the H1 box is strikingly similar to the yeast autonomously replicating sequences (ARS) which represent putative origins of DNA replication (Montiel *et al.*, 1984; Younghusband *et al.*, 1986). Several examples of ARS have been found associated with yeast histone genes and there is evidence that the ARS may be required for the periodic expression (Osley and Hereford, 1982), or influence absolute transcript levels (Osley *et al.*, 1986) of the nearby histone genes. However, the significance of putative origins of replication being closely associated with histone and other cell cycle regulated genes (Miller and Nasmyth, 1984; Brand *et al.*, 1987) and whether the similarity of the H1 box to the ARS is biologically significant, remains unresolved.

Another interesting correlation between histone gene transcription and DNA replication involves the histone H2B gene and its cell cycle regulatory promoter element, the H2B box. The H2B box (octamer sequence) which binds the Oct-1 transcription factor is also found to function as a DNA replication element in the adenovirus origin of replication. Oct-1 (also designated NF-III representing the binding activity originally identified binding the adenovirus octamer sequence) binds both octamer elements and possesses both transcription and replication activities respectively in appropriate *in vitro* transcription and DNA replication assays (Fletcher *et al.*, 1987; O'Neill and Kelly, 1988; O'Neill *et al.*, 1988). This precedent suggests that H1-SF, which binds the H1 box, may also have multiple

regulatory roles, where apart from playing a key role in S-phase transcription of the H1 genes it may also be important for DNA replication.

H1-SF binding activity as originally identified by Dalton (1987) proved elusive, with initial attempts to identify and confirm H1-SF binding activity, using the gel retardation assay unsuccessful. The double stranded oligonucleotide probes used contained the same H1 box (40 mer) and disrupted version of the H1 box (Δ40 mer) used to identify H1-SF. For reasons still unclear, an H1 box specific binding activity (H1-SF) was not detected. However, a binding activity (H1-F40) was identified and shown to interact with both the 40 mer and Δ40 mer probes. H1-F40 binding activity did appear to show some sequence specificity even though it bound both 40 mer and Δ40 mer. However S. Dalton (personal communication) reported to have never observed a complex with the DNA binding characteristics described for H1-F40. Furthermore, in agreement with Dalton and Wells (1988*a*), Gallinari *et al.* (1989) using similarly designed probes, but corresponding to the human histone Hh9 H1 gene, also detected a completely H1 box specific binding activity (H1TF1) in HeLa cells. Therefore, at this point in time, H1-F40 was dismissed as an H1-SF candidate.

Subsequently, another set of double stranded oligonucleotide probes containing an intact H1 box (14 mer) and disrupted H1 box (Δ14 mer) were designed in an attempt to detect a genuine H1-SF candidate. The rationale for these changes was to prevent as many non-specific interactions from occurring as possible by making the oligonucleotide probes as small as possible while still including the relevant DNA recognition sequence. An H1-SF candidate (H1-F14), which bound the 14 mer but not the Δ14 mer was identified by gel retardation assay. By these binding criteria and other competitor studies H1-F14 was believed to be H1 box specific and represent H1-SF (see 2.3.3). Furthermore H1-F14 seemed to represent a different binding activity compared to H1-F40 (see 2.4 and 4.2.2 (iii)). With the belief that H1-F14 represented an H1 box specific binding activity (and possibly represented the

H1-SF activity detected by Dalton (1987)), purification of H1-F14 was commenced. Two different approaches were taken.

The first approach involved oligoscreening of a ten day chicken embryo λgt11 cDNA expression library for H1-F14 with a concatenated 14 mer probe. The ten day chicken embryo cDNA λgt11 library was chosen because nuclear extract from nine day chicken embryos was shown to contain H1-F14 binding activity. It was envisaged that the oligoscreening approach would avoid the more difficult and labour-intensive approach of protein purification. However no H1-F14 clones were isolated after extensive oligoscreening of the library. An appropriate control clone was successfully detected with the oligoscreening method. Therefore the failure to detect potential H1-F14 clones was likely to be due to biological characteristic(s) peculiar to H1-F14 that prevented it from being identified with this technique (other limitations of the oligoscreening technique with respect to the detection of DNA binding proteins were discussed in 3.2) rather than a general operational problem with the technique.

Therefore a planned second approach was pursued, in which, a large scale protein column chromatography purification for H1-F14 was used to isolate enough pure protein to allow determination of amino acid sequence. This would enable specific oligonucleotide probes to be designed and used to screen recombinant DNA libraries for the H1-F14 gene. Crude nuclear extract was prepared from 600 dozen nine day old chicken embryos, and H1-F14 partially purified by heparin-Sepharose column chromatography. Extract was then concentrated by membrane filtration concentration and then subjected to Sephacryl S-300 gel filtration column chromatography (see Figure 3.10 for relative purification of H1-F14). Affinity chromatography purification of H1-F14 was to be performed next to purify H1-F14 to homogenity. However, before affinity chromatography was carried out, experiments were completed that demonstrated that H1-F14 represented a non-specific binding protein (K. Duncliffe, this laboratory; personal communication).

K. Duncliffe was able to show that histone H1 specific G/C and CCAAT box probes each form a retarded complex that, by competitior studies, appeared to be the same as H1-F14 detected with the 14 mer probe. These gel retardation experiments were repeated and the results confirmed (see 3.3.11 (iii)). The G/C, and CCAAT box competitors were able to compete out H1-F14 almost as efficiently as the 14 mer itself, suggesting that the H1-F14 binding activity was not H1 box specific but represented a non-specific DNA binding activity. Why the H1-F14 activity did not bind the Δ14 mer probe was not clear, although fortuitous sequence changes in the Δ14 mer may have simply reduced its capacity to interact with this type of DNA binding protein. In hindsight, the small size of the 14 mer may simply have not allowed H1-SF to bind, and therefore favoured a non-specific DNA binding protein interaction. Furthermore southwestern and gel filtration size estimations of H1-F14 at less than 40 kDa (see 3.3.11 (i) and (ii)) suggested, that H1-F14 was not the 90 kDa H1-SF binding activity originally identified by Dalton (1987). As a result of these findings, the purification of H1-F14 was abandoned. Although there are precendents for DNA binding proteins binding specifically to more than one unrelated site (Pfeifer *et al.*, 1987; Johnson *et al.*, 1987), it is difficult to imagine H1-F14 representing such a protein, particulary considering the differences in the three oligonuclotide sequences that H1-F14 recognised.

It was then reasoned that sequences outside of the H1 box, and therefore absent from the 14 mer, may be important for H1-SF binding. A further set of double stranded oligonucleotide probes containing an intact H1 box (25 mer) and a disrupted H1 box (Δ25 mer) were designed using a rational described in 4.2.1. A H1-SF candidate (H1-F25) was identified by gel retardation assay using these probes (see 4.2.2) and like H1-F40 which bound the 40 mer and Δ40 mer probes, also bound both the 25 mer and Δ25 mer probes. More rigorous testing of H1-F25 binding specificity, by modification interference analysis, demonstrated that sequences outside of the H1 box along with those contained within it, were involved in binding (see 4.2.3). As a

result of these findings H1-F25 was not dismissed as a H1-SF candidate and its further characterisation was pursued.

The demonstration that additional flanking sequence to the H1 box was involved in H1-F25 binding was consistent with the earlier suggestion that H1-SF would perhaps be unable to bind the 14 mer because the 14 mer lacked the required flanking regions. But conversly, why was H1-F14 binding activity not detected with the 25 mer probe (or 40 mer probe)? One possible explaination is that H1-F25 (H1-F40) binds the H1 box with a degree of affinity consistent with a specific interaction (at least at a higher affinity than the non-speific interaction prescribed to H1-F14), such that H1-F25 (H1-F40) binding would compete and exclude the relatively weaker binding H1-F14 from interacting with the 25 mer and 40 mer probes.

The probes utilised in this study were based on the H1.01 gene. However, this is not the major H1 subtype in dividing cells (Shannon and Wells, 1987). In fact, H1.01 represents a partially replication-dependent subtype (see 1.10.3) which becomes the predominantly expressed histone H1 subtype in differentiated tissues. Therefore in hindsight, H1.01 may not have been the best choice of histone H1 promoter sequence to isolate H1-SF, particularly when it is considered that the H1 box may influence the relative expression of the different H1 subtypes. It has been suggested that although the H1 box is involved in cell cycle regulation of the histone H1 genes it may also be involved in determining the relative abundance of each subtype in different cell types (Heintz, 1991). Taking this into account, the H1.01 gene may possess a sub-optimal binding H1 box (and immediate flanking region) that binds H1-SF with less avidity than for example the H1 box in the promoter of the H1.02 gene, which appears to be the major replication dependent subtype in dividing cells. Other differences between the H1.01 and H1.02 genes may also be involved. For instance, a sub-optimal TATA box, with a base substitution in the third position in the TATA box (Coles *et al.*, 1987) is also found in the H1.01 gene and may explain the significantly reduced levels of H1.01 transcripts compared to H1.02 transcripts

(Corden *et al.*, 1980). Furthermore differences in the levels of expression between the two subtypes may also be significantly controlled at the level of RNA processing. The H1.01 gene contains a less than optimal 3' processing element (Coles *et al.*, 1987).

While the H1.01 and H1.02 H1 motifs both contain the sequence AAGAAACACA, the immediate flanking sequences are divergent. Experiments by others in this laboratory are in progress to determine whether the H1 box from the H1.02 gene binds a more convincing H1-SF candidate or H1-F25 with greater affinity/sepecificty than the H1.01 H1 box (K. Duncliffe, personal communication). Taking into account the results obtained with the modification interference analysis demonstrating that flanking sequence is involved in H1-F25 binding, this investigation would seem to be particularly relevant.

However the observation remains that the original H1-SF binding activity was identified using H1.01 H1 box sequence. Therefore the H1-F25 binding activity characterised using the H1.01 H1 box was still of interest as it represented the best H1-SF candidate identified since the original description of H1-SF by Dalton and Wells (1988*a*).

A modified oligoscreening method (Kalionis and O'Farrell, 1993), different in a number of important aspects to the method used in the original attempts to isolate H1-F14, was used to screen a ten day chicken embryo λgt11 cDNA expression library for H1-F25. Two candidate H1-F25 clones were isolated (λ4.1 and λ8.1). These clones were subsequently found, by DNA sequence analysis, to code for novel zinc finger DNA binding proteins with 12 and 5 zinc fingers respectively (nucleic acid and protein databases searched; July, 1992) (see 4.5 for further discussion of Z5 and Z12 structures).

The Z12 and Z5 clones were subcloned into the pGEX-1 expression vector and expressed in *E. coli*. Extracts were prepared and Z12 and Z5 fusion proteins shown

to bind 25 mer probe in the gel retardation assay, in a manner analogous to H1-F25. The Z5 clone only bound the Δ25 mer probe very weakly in the oligoscreening assay (see 4.3.3). Consistent with this result the Z5/25 mer complex was not competed out by Δ25 mer competitor in the gel retardation assay (see 4.4.5). Z12 behaved more like H1-F25 by interacting with both 25 mer and Δ25 mer but seemed to bind the Δ25 mer more strongly. Neither Z5 or Z12 bound the non-specific control probes in the oligoscreening assay. In particular, the behaviour of Z5 in particular was very encouraging in terms of representing a possible H1-SF candidate.

As both λ4.1 (Z12) and λ8.1 (Z5) were incomplete cDNA clones for Z5 and Z12 respectively (see 4.4.2, 4.4.3 and 5.2) further screening to isolate the full length cDNA clones forthese genes was undertaken. Three further Z5 clones were isolated but all of these were found to be the same as the original Z5 clone. Another Z12 cDNA clone λ16.1 (Z13) was isolated and found to contain different 5' coding sequence in addition to an extra zinc finger located near the 3' end of the clone (see 5.4.2).

The original Z12 clone would appear to be an incorrectly spliced cDNA copy. A 86 bp intronic sequence appears to not have been removed from the Z12 sequence creating a frameshift in the Z12 sequence downstream of this point. The λ16.1 (Z13) clone lacks the 86 bp sequence (presumably correctly spliced) and as a result codes for a 13th zinc finger approximately 30 bp downstream of the proposed splice site. The 5' ends of Z12 and λ16.1 (Z13) also diverge. However in this instance the differences could be the result of alternative splicing event. See Figures 4.9, 5.5, 5.7A, 5.8A and 5.8B for sequence comparisons and diagramatic representation of Z12 vs λ16.1 (Z13).

Further binding studies with Z13 should be done to examine the influence that the extra finger has on binding specificity. To make sure that the complete DNA binding domains and thus DNA binding specificities of the clones are characterised, these

studies and a more complete characterisation of Z5 binding specificity should perhaps wait until the full length Z12/Z13 and Z5 clones are isolated .

Screening of a λCharon 4a chicken genomic library for genomic clones containing the Z12 and Z5 genes was carried out at the same time as the cDNA library screening. Isolation and characterisation of the genes for Z5 and Z12 should help to determine their identity and function. In particular, the availability of the full length genomic sequence of Z5 and Z12 and identification of their promoters could lead to studies concerned with the mechanism of their transcriptional control and provide a basis for further investigation of the biology of the proteins. Genomic clones were isolated for both Z12 (λG7.1) and Z5 (λG2.1). The λG2.1 clone contains the entire Z5 cDNA sequence and it is quite possible it contains the entire Z5 gene. The λG7.1 clone was found to contain 5' cDNA sequence but it did not contain Z12 cDNA 3' downstream sequence from the fifth finger in the Z12 cDNA. The Z12 genomic clone may prove useful in the immediate future to resolving the origins of the 5' sequence differences between the Z12 and λ16.1 (Z13) cDNA clones.

A mouse genomic clone for Z13 has also been recently isolated in this laboratory (T. Shulz, personal communication) which shows considerable sequence homology to the chicken Z12 and λ16.1 (Z13) cDNA clones. This suggests that the Z12/Z13 protein may carry out an important function in the cell considering the high degree of conservation between the mouse and chicken sequences. Mouse Z5 genomic clones have also been isolated (T. Shulz, personal communication) however no extensive sequence information is available to be able to compare the mouse and chicken Z5 sequences at this time.

6.2 FUTURE WORK

The uncertainty surrounding the identities of the Z12 and Z5 clones and their relationship to H1-SF, as originally identified by Dalton (1987), could be addressed with the following experiments.

Considering that the isolation of full length Z5 and Z12/Z13 cDNA clones is important to be able to fully characterise the proteins and that the isolation of full length clones for both Z5 and Z12 has proven difficult, other approaches should be considered. Given the low abundance of the mRNAs (see 5.7) a PCR based method to isolating the remaining 5' sequences for Z5 and Z12 may be successful (Friedman *et al.*, 1988; Frohman *et al.*, 1988). Complete Z5 and Z12 clones may be obtained by screening different libraries. Randomly primed or 'stretch' cDNA libraries may be useful in overcoming the problems experienced with the oligo d(T) primed library used in this study.

Investigation of the exact nature of the DNA sequences recognised by Z5 and Z12 would also help to further identify and characterise the proteins. A recently developed method that enables the DNA binding specificities of proteins to be determined could be used (Pollock and Treisman, 1990; Payre and Vincent, 1991; Prendergast and Ziff, 1992). In this method protein is incubated with a pool of random oligonucleotides (or restriction enzyme digested genomic DNA), complexes purified by immunoprecipitation (or preparative gel retardation) and the bound DNA amplified by PCR. The DNA is then used in further rounds of binding, immunoprecipitation (gel retardation) and PCR amplification until specific binding is detectable and a consensus nucleotide sequence for binding with the protein can be generated.

Following the isolation of compete Z5 and Z12/Z13 clones the influence of Z5 and Z12/Z13 proteins on H1 gene transcription in an *in vitro* transcription assay, (perhaps similar to that described by Gallinari *et al.* (1989)) could be investigated. Additionally, Dalton and Wells (1988*a*) were able to demonstrate a marked reduction in histone H1 gene transcription when multiple copies of the 40 mer H1 box oligonucleotide were introduced into Hela cells expressing an histone H1 gene construct. While introduction of the Δ40 mer containing the mutated H1 box had no detectable effect. These results suggested that a titratable factor (H1-SF) was

involved in histone H1 gene transcription that specifically bound the H1 box. A similar experiment that may prove informative to whether Z5 or Z12 represents H1-SF could be done by introducing anti-sense Z5 and Z12 RNA into cells transfected with a similar H1 histone gene construct and following the affects on H1 gene transcription.

Generation of antibodies to the Z5 and Z12 proteins may also prove useful for further characterisation of the proteins by Western and *in situ* analysis. These studies could provide evidence to whether or not Z5 or Z12 protein expression patterns are consistent with either being H1-SF.

Finally, if Z5 or Z12 are shown not to be H1-SF, they are novel proteins in their own right with highly conserved sequences between chicken and mouse and therefore likely to have important functions and as such warrant further investigation.

CHAPTER 7

MATERIALS AND METHODS

7.1 ABBREVIATIONS

Abbreviations are as described in 'Instructions to authors' (1991) Biochem. J. **273**, 1-19. In addition:

A_{xxx}:	absorbance at xxx nm
APS:	ammonium persulphate
BCIG:	5-bromo-4-chloro-3-indolyl-β-D-galactoside
bisacrylamide:	N, N'-methylene-bisacrylamide
bp:	base pair
BSA:	bovine serum albumin
CAPS:	3-(cyclohexylamino)-1-propanesulphonic acid
Da:	dalton
DEAE:	diethylaminoethyl
DEPC:	diethylpyrocarbonate
DMEM:	Dulbecco's Modified Eagle's Medium
DMF:	dimethylformamide
DTT:	dithiothreitol
EDTA:	ethylenediaminetetra-acetate
EGTA:	ethyleneglycol bis(amino-ethyl ether)tetra-acetate
FCS:	foetal calf serum
HEPES:	4-(2-hydroxyethyl)-1-piperazine-ethanesulphonic acid
HPLC:	High performance liquid chromatography
IPTG:	isopropyl-β-D-thio-galactopyranoside
kb:	kilobase pair
kDa:	kilodalton
MOPS:	4-morpholinepropanesulphonic acid
PMSF:	phenylmethylsulphonyl fluoride
poly(dI-dC):	polydeoxyinosinic-deoxycytidylic acid
PVDF:	polyvinylene difluoride
RSP:	reverse sequencing primer
SDS:	sodium dodecyl sulphate
TEMED:	N, N, N', N'-tetramethylethylenediamine
USP:	universal sequencing primer
UV:	ultraviolet

7.2 MATERIALS

7.2.1 Chemicals, reagents and kits

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

Acrylamide, agarose, ampicillin, ATP: Sigma
Bisacrylamide: Bio-Rad
BCIG, bromophenol blue, BSA: Sigma
Chicken serum: Flow Laboratories
CsCl: Boehringer Mannheim
DEAE-Sepharose, Dextran sulphate: Pharmacia
DMEM: Gibco
dNTPs, DTT: Sigma
Erase-a-base kit: Promega
Ethidium bromide: Sigma
Foetal Calf Serum: Commonwealth Serum Laboratories
Genescreen: New England Nuclear
Gentamycin: Schering Corporation
Heparin-Sepharose CL-6B: Pharmacia
Immobilon-P: Millipore
IPTG: Sigma
Mixed bed resin AG 501-X8 (D): Bio-rad
mRNA purification kit: Pharmacia
Nitrocellulose: Schleicher and Schuell
Oligo-labelling kit: Bresatec
Plaquescreen: New England Nuclear
PMSF, poly(dI-dC): Sigma
Polaroid film (type 667): Polaroid
Salmom sperm DNA, SDS: Sigma
Sephacryl S-300, Sepharose CL-6B: Pharmacia
Sequencing kit: Bresatec
TEMED: Bio-Rad
Ultrafiltration membranes (YM10): Amicon
Urea (ultra-pure): Merck
X-ray film: Fuji, Kodak
Xylene cyanol: Sigma

7.2.2 Enzymes

Enzymes were obtained from the following sources:

Calf intestinal phosphatase: Sigma
Deoxyribonuclease I: Sigma
E.coli DNA polymerase I, Klenow fragment: Bresatec
Lysozyme: Sigma
Proteinase K: Boehringer Mannheim
Restriction endonucleases: New England Biolabs, Pharmacia
Ribonuclease A: Sigma
T4 DNA ligase: Bresatec
T7 DNA polymerase: Bresatec
Taq polymerase: Bresatec
T4 polynucleotide kinase: Bresatec

7.2.3 Isotypically labelled compounds

α-^{32}P-dNTP's (3000Ci/mmole): Bresatec
α-^{35}S-dATP (1500 Ci/mmole): Bresatec
γ-^{32}P-ATP (4000 Ci/mmole): Bresatec

7.2.4 Bacterial strains, phage and cell lines

(i) *E.coli* strains

DH5α: *supE44, Δlac U169 (f80 lacZΔM15), hsdR17, recA1, endA1, gyrA96 thi-1, relA1* (Sambrook *et al.*, 1989). Gift from H. Richardson (Department of Biochemistry, University of Adelaide).

ED8799: *hsdS, metB7, supE, (glnV)44, supF, (tyrT)58, Δ(lacZ)M15*, r_k^-, m_k^- Gift from S. Clarke (Biotechnology Australia).

LE392: F^- *supE44, supF58, hsdR514* (r_k^-, m_k^-), *lacY1* or *Δ(lacIZY)6, galK2, galT22, metB1, trpR55*, λ^- (Sambrook *et al.*, 1989). Gift from V. Thonglairoam (Department of Biochemistry, University of Adelaide).

Y1089: *araΔ139, ΔlacU169, proA*$^+$, *Δlon, rpsL, hflA150*[*chr::Tn10(tet*^{r}*)*], pMC9 (Young and Davis, 1983*a*). Gift from B. Kalionis (Department of Biochemistry, University of Adelaide).

Y1090: *supF, hsdR, araΔ139, Δlon, ΔlacU169, rpsL, proA*$^+$, *strA*, [*trpC22::Tn10*], pMC9 (Young and Davis, 1983*a*). Gift from B. Kalionis.

(ii) Phage

Ten day chicken embryo cDNA λgt11 library CL1001b: Clontech

CCAAT enhancer binding protein (C/EBP) λgt11 clone (Landschultz *et al.*, 1988): Gift from S.McKnight (Howard Hughes Research laboratories, Carnegie Institution of Washington).

Engrailed homeobox binding protein λgt11 clone (Desplan *et al.*, 1988): Gift from B. Kalionis

Chicken genomic λ Charon 4a library (Dodgson *et al.*, 1979): Gift from A. Robins (Department of Biochemistry, University of Adelaide).

Wild type λgt11 lysogen (Y1089): Gift from B. Kalionis

(iii) Tissue culture cell lines

LSCC HD2 (Beug *et al.*, 1982): Temperature sensitive avian erythroblastosis virus (ts34-AEV) transformed chicken erythroid cell line. Gift from T. Graf (Institute of Virology, German Cancer Research Centre, Heidelberg).

7.2.5 Media and buffers

(i) Media

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

L-broth:	1% (w/v) amine A, 0.5% (w/v) yeast extract and 1% (w/v) NaCl, pH 7.0
LMM-broth:	L-broth, 0.2% (w/v) maltose and 10 mM $MgSO_4$
L-agar plates:	L-broth and 1.5% (w/v) bacto-agar
LMM-agarose:	LMM-broth and 0.7% (w/v) agarose

All bacteria were grown in L-broth, LMM broth or on L-agar plates. Ampicillin (50 μg/ml) was added where required for selection.

Tissue culture cells were maintained at 37°C in an atmosphere of 5% CO_2. LSCC HD2 cells were grown in DMEM medium supplemented with 20 mM HEPES, 5 x

10^{-5} M β-mercaptoethanol, sodium bicarbonate (2.4 g/Litre), gentamycin (50 mg/Litre), 2% chicken serum and 10% FCS.

(ii) Buffers

Commonly used buffers were:

Denhardts:	0.02% (w/v) polyvinyl pyrolidone, 0.02% (w/v) BSA and 0.02% (w/v) Ficoll
PBS:	7.5 mM Na_2HPO_4, 2.5 mM $NaH_2PO_4.2H_2O$ and 145 mM NaCl
PSB:	10 mM Tris-HCl pH 7.4, 10 mM NaCl and 10 mM $MgCl_2$
TBE:	50 mM Tris-borate pH 8.3 and 1 mM EDTA
TE:	10 mM Tris-HCl pH 7.4 and 1 mM EDTA
TAE:	40 mM Tris-acetate pH 8.2 and 1 mM EDTA
Agarose load buffer:	50% (v/v) glycerol, 50 mM Tris-HCl pH 7.4, 5 mM EDTA and 0.1% (w/v) Bromophenol blue
Acrylamide load buffer:	50% (v/v) glycerol, 50 mM Tris-HCl pH 7.4, 5 mM EDTA, 0.1% (w/v) Bromophenol blue and 0.1% (w/v) Xylene cyanol
Gel retardation load buffer:	50% (v/v) glycerol, 5 mM DTT and 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 and 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
RNA load buffer:	50% (v/v) glycerol, 1 mM EDTA and 0.1% (w/v) Bromophenol blue
TM Buffer:	50 mM Tris-HCl pH 7.9, 12.5 mM $MgCl_2$, 1 mM EDTA, 20% (v/v) glycerol, 1 mM DTT and 0.5 mM PMSF
SSC:	150 mM NaCl and 15 mM sodium citrate
SSPE:	180 mM NaCl, 10 mM sodium phosphate pH 7.4 and 1 mM EDTA

7.2.6 Recombinant DNA clones and cloning vectors

(i) Recombinant DNA clones

pPAc (Kost *et al.*, 1983): Gift from K. Duncliffe (Department of Biochemistry, University of Adelaide).

(ii) Cloning vectors

pBluescript SK+ (Stratagene): Gift from A. Robins

pGEX-1 (Smith and Johnson, 1988): Gift from R. B. Saint (Department of Biochemistry, University of Adelaide).

7.2.7 Molecular size markers

(i) DNA markers

DMW-S1: *Eco*RI digested SPP1 phage DNA (0.38 - 7.84 kb): Bresatec

DMW-P1: *Hpa*II digested pUC19 DNA (26 - 502 bp): Bresatec

DMW-L1: *Hin*dIII digested λ DNA (0.56 - 23 kb): Bresatec

(ii) Protein markers

MW-SDS-200: high molecular mass (29 - 205 kDa): Sigma

(iii) RNA markers

5620SA: RNA standards (0.24 - 9.49 kb): B.R.L.

7.2.8 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:

H5 +:	5' dTCGAGAGCCAGGAGGAGGAGAGGGGACTCCTCCTTGTCCATAGGA
H5 + cont.	GTGAGGCACAGCCCG 3'
H5 -:	5' dGATCCGGGCTGTGCCTCACTCCTATGGACAAGGAGGAGTCCCCTC
H5 - cont.	TCCTCCTCCTGGCTC 3'
H1 Box 40 +:	5' dTATTCTTTTGTTAGTCCAAAGAAACACAAATCGAGCACAG 3'
H1 Box 40 -:	5' dGATCCTGTGCTCGATTTGTGTTTCTTTGGACTAACAAAAG 3'
H1 Box Δ40 +:	5' dTATTCTTTTGTTAGTCCAAAGGACTACGAATCGAGCACAG 3'
H1 Box Δ40 -:	5' dGATCCTGTGCTCGATTCGTAGTCCTTTGGACTAACAAAAG 3'
H1 Box 14 +:	5' dAAGAAACACAAGCT 3'

H1 Box 14 -:	5' dTGTGTTTCTTAGCT 3'
H1 Box Δ14 +:	5' dAAGGACTACGAGCT 3'
H1 Box Δ14 -:	5' dCGTAGTCCTTAGCT 3'
H1 G/C +:	5' dTGCAGCGGGGCGGGCTTAGC 3'
H1 G/C -:	5' dGCTAAGCCCGCCCCGCTGCA 3'
H1 CCAAT +:	5' dCAACGCACCAATCACCGCG 3'
H1 CCAAT -:	5' dCGCGGTGATTGGTGCGTTG 3'
H1 Box 25 +:	5' dGTCCAAAGAAACACAAATCGAGCAC 3'
H1 Box 25 -:	5' dGTGCTCGATTTGTGTTTCTTTGGCA 3'
H1 Box Δ25 +:	5' dGTCCAAAGTCTCCTCAATCGAGCAC 3'
H1 Box Δ25 -:	5' dGTGCTCGATTGAGGAGACTTTGGAC 3'
RSP:	5' dAACAGCTATGACCATG 3'
USP:	5' dGTAAAACGACGGCCAGT 3'
λgt11 PCR 1:	5' dATCCAAGCTTCGGTGGCGACGACTCCTGG 3'
λgt11 PCR 2:	5' dCGGGATCCTATTGACACCAGACCAACTGGTA 3'
λ8.1 PCR 1:	5' dGCGAATTCGAAGTAGGTACTGCAGGCAG 3'
λ8.1 PCR 2:	5' dGCGAATTCTGGAAGCAGAGCAGCAATT 3'
λ4.1 PCR 1:	5' dGCGAATTCAGTGCTTGGAGCATCACT 3'
λ16.1 PCR 1:	5' dCTTATGTGCAGTACAGG 3'
Eng.:	5' dTCAATTAATTGA 3'
C/EBP +:	5' dAATTCAATTGGGCAATCAGG 3'
C/EBP -:	5' dAATTCCTGATTGCCCAATTG 3'

7.3 METHODS

7.3.1 Isolation of plasmid DNA

(i) Large scale method

A single bacterial colony was picked into 100 ml of L-broth plus the appropriate antibiotic and grown overnight at 37°C with vigorous shaking. The cells were harvested by centrifugation at 5,000 rpm (HB-4 rotor) for 10 minutes and resuspended in 2 ml 15% (w/v) sucrose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA and lysozyme (final concentration 2 mg/ml). The suspension was kept on ice while 4 ml 0.2 M NaOH/0.1% (w/v) SDS was added, followed by 2.5 ml 3 M sodium acetate pH 4.6. After mixing gently, the chromosomal DNA and cellular debris were removed by centrifugation at 15,000 rpm (HB-4 rotor) for 15 minutes. The supernatant was collected, and the DNA precipitated by adding 5 ml ice cold isopropanol, and spinning at 15,000 rpm (HB-4 rotor) for 10 minutes. The DNA pellet was resuspended in 1.38 ml TE buffer, and added to 1.5 g CsCl in a 10 ml tube. After the

CsCl was dissolved, 120 µl of 10 mg/ml ethidium bromide was added to give a total volume of 2.2 ml which was placed in a small Beckman TL-100 heat sealable tube. The sample was then centrifuged at 80,000 rpm for
16 hours in a Beckman TL-100 centrifuge. The lower band in the CsCl gradient containing supercoiled plasmid DNA was removed from the tube using a needle and syringe. To remove ethidium bromide, 1 volume of water-saturated butanol was added, mixed, and the top layer discarded. This step was repeated three times. One volume of water, and three volumes of ethanol were then added, and the DNA precipitated. After centrifugation to pellet the DNA, the DNA was resuspended in 0.2 ml of TE buffer, and stored at -20°C. Approximately 2 µg of plasmid was recovered for each ml of overnight culture used.

(ii) Miniscreen method

Colonies were grown overnight in 1.5 ml of L-broth plus the appropriate antibiotic at 37°C with continual shaking. The cells were then pelleted by centrifugation for one minute in a microfuge and the cell pellet resuspended in 200 µl of STET buffer (8% sucrose, 5% Triton X-100, 50 mM Tris-HCl pH 8.0, 50 mM EDTA). Lysozyme was added (0.5 mg/ml final concentration), the sample mixed by gentle inversion, boiled for 30 seconds, and then spun in a microfuge for 15 minutes at 4°C. The pellet was removed with a sterile toothpick and 200 µl of cold isopropanol added to the remaining supernatant. The sample was put on ice for 10 minutes and the DNA precipitated by spinning in a microfuge for 15 minutes at 4°. The DNA pellet was washed with 70% (v/v) nuclease-free ethanol, and dried *in vacuo* before being resuspended in 20 µl TE buffer. DNA obtained was adequate for restriction enzyme digestion and gel electrophoresis.

7.3.2 Restriction enzyme digestion and analysis of DNA

(i) Restriction enzyme digestion

Restriction endonuclease digestions were performed using the conditions for each enzyme as recommended by the supplier. Analytical digests were performed in 10 µl reactions containing 200 - 500 ng of DNA, and a two to five-fold excess of enzyme for 1 hour. Preparative digests were performed in 100 µl reactions containing 2 - 10 µg DNA. Reactions were stopped by adding a quarter volume of agarose loading buffer or acrylamide loading buffer and loading onto appropriate gel. Alternatively, a phenol/chloroform extraction was performed by adding an equal volume of phenol/chloroform (1:1) and vortexing, followed by centrifugation for three minutes in a microfuge, and recovery of the aqueous phase. DNA was

ethanol precipitated from the aqueous supernatant by adjusting the reaction mix to 0.3 M sodium acetate pH 5.5, and the addition of 2.5 volumes of cold, nuclease-free ethanol. The DNA was pelleted by centrifugation for 10 minutes in a microfuge. The DNA pellet was washed with 70% (v/v) nuclease-free ethanol, and dried *in vacuo* before being resuspended in an appropriate volume of water.

(ii) Agarose gel electrophoresis

Agarose was dissolved in TAE to the final concentration required (0.8 - 3.0%), and poured onto 7.5 cm x 5 cm microscope slides, to be used as horizontal gels. The gels were submerged in 400 ml of the appropriate buffer, and a current of 80 - 100 V was applied. DNA samples were dissolved in water, 3 µl of agarose loading buffer was added, and samples of 5 - 10 µl loaded into the gel slots. The DNA was visualised after running, with 0.02% Ethidium bromide, and exposure to short wave UV light (254 nm). Note, if the DNA sample was to be used for cloning or further experimental manipulation the DNA was visualised under long wave UV light (365 nm).

(iii) Polyacrylamide gel electrophoresis

Electrophoresis of DNA species of less than about 1 kb in length was carried out on vertical 14 cm x 14 cm x 0.5 mm slab gels containing 5 - 20% (w/v) acrylamide/bisacrylamide (30:1), which had been deionised with mixed bed resin. The acrylamide was polymerised in 1 x TBE buffer by the addition of 0.1% (w/v) APS, and 0.1% (v/v) TEMED. The gel was pre-electrophoresed for 30 minutes at 400 V prior to loading. A quarter volume of acrylamide load buffer was added to the samples which were directly loaded into gel slots approximately 1 cm wide. All acrylamide gels were electrophoresed at 400V until the bromophenol blue had migrated the desired distance. DNA was visualised under UV light (254 nm) after ethidium bromide staining, or by autoradiography (7.3.24) if the DNA was radioactively labelled.

(iv) Denaturing polyacrylamide gel electrophoresis

Where analysis of single-stranded DNA was required, 6% (w/v) polyacrylamide /8 M Urea gels containing TBE buffer were used. The gels used were 40 cm x 40 cm x 0.35 mm, and were pre-electrophoresed for 30 minutes before use. The gels were run in TBE buffer at 30 mA constant current. Gels were fixed with 200 ml of 10% (v/v) acetic acid, and washed with 4 litres of 20% (v/v) aqueous ethanol. After drying, gels were autoradiographed (7.3.24) overnight at room temperature. For

detection of low levels of radioactivity, a tungsten intensifying screen was used, and the gel exposed at -80°C .

7.3.3 Subcloning of DNA fragments into plasmid vectors

(i) Isolation of DNA from agarose gels

DNA was isolated from agarose gels using a spin column procedure (Heery *et al.*, 1990) as described below. A 0.5 ml Eppendorf tube was pierced through the bottom with a 25 gauge syringe needle and plugged with siliconized sterile glass wool. The agarose gel slice containing the DNA fragment was transferred to this tube which was then placed inside a 1.5 ml Eppendorf tube and the device spun at 6,000 rpm in an microfuge for 10 minutes. The eluate, containing the DNA fragment, was recovered in the 1.5 ml Eppendorf tube. The eluate was then extracted with phenol/chloroform extracted and DNA precipitated with ethanol (7.3.2 (i)).

(ii) Preparation of plasmid vectors

Vector DNA was linearised with a suitable restriction enzyme, then dephosphorylated with calf intestinal phosphatase in a 100 µl reaction mixture containing 100 mM Tris-HCl pH 8.0, 0.5% (w/v) SDS and 1.0 unit of enzyme. After a one hour incubation at 37°C, the linearised, dephosphorylated vector was purified from uncut vector by running the DNA on an agarose gel and isolating the appropriate band (7.3.2 (ii)).

(iii) Ligation into plasmid vectors

Restriction fragments to be subcloned were preparatively isolated from agarose gels (7.3.3 (i)). Ligation of insert into the vector was performed in a 20 µl volume containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1mM DTT, 0.5 mM ATP and 1 unit of T4 DNA ligase, at 4°C for 16 hours, or at room temperature for 4 hours. Sufficient insert to give a three fold molar excess over 20 ng of vector was normally used in a ligation mixture. When the DNA fragment had unsuitable protruding 5' termini, it was treated with DNA polymerase I, Klenow fragment, to end-fill the single-stranded regions to blunt-ends. This was done prior to purification of the fragment by agarose gel electrophoresis in a 20 µl reaction mixture containing 10 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 100 mM of each dNTP and 1 unit of *E.coli* DNA-polymerase I Klenow fragment.

(iv) Transformation of recombinants into bacteria

An overnight culture (1 ml) of *E.coli* ED8799, or *E.coli* DH5α, was diluted 50 fold in 50 ml of L-broth, and grown with shaking to an A_{600} of 0.6. The cells were pelleted by gentle centrifugation (HB-4 rotor; 5,000 rpm for 5 minutes), washed in a half volume of ice-cold 0.1 M $MgCl_2$, then resuspended in 2.5 ml of ice-cold 0.1 M $CaCl_2$, and left on ice for at least 1 hour. Usually one quarter of the ligation reaction was mixed with 200 µl of the competent cells, and left on ice for 30 minutes. The transformation mix was heat shocked at 42°C for 2 minutes, then 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. After this time, the cells were gently pelleted, resuspended in 100 µl of L-broth, and spread directly onto an L-agar plate (with the appropriate antibiotic). When colour selection was required with *E.coli* ED8799, or *E.coli* DH5α cells, L-agar plates (with the appropriate antibiotic) were used with 60 µl of BCIG (20 mg/ml in DMF) and 60 µl of IPTG (10 mg/ml) spread directly onto the surface of each plate.

7.3.4 Labelling of DNA fragments

(i) Kinasing of synthetic oligonucleotides

Oligonucleotides were 5' end-labelled with T4 polynucleotide kinase and γ-^{32}P-ATP. Normally 50 - 100 ng of oligonucleotide was kinased in a 10 µl reaction mixture containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 50 µCi of γ-^{32}P-ATP and 1 unit of enzyme. The reaction mixture was incubated at 37°C for 30 minutes, then 5 µl of acrylamide loading buffer was added and the mixture loaded onto a 20% (w/v) polyacrylamide gel for purification. The band corresponding to the labelled oligonucleotide was located by autoradiography (7.3.24), excised using a scalpel blade, and eluted by incubation at 37°C overnight into 400 µl of TE buffer.

(ii) Endfilling of DNA fragments

DNA fragments (500 ng) with 5' overhangs were incubated in a 10 µl reaction mixture containing 10 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 30 mM NaCl, 100 µM of each dNTP (excluding the appropriate labelled dNTP(s)) and 50 µCi of the appropriate α-^{32}P-dNTP (dATP and/or dCTP). One unit of *E.coli* DNA polymerase I, Klenow fragment was added to the reaction, and the mixture incubated at 37°C for 30 minutes. The reaction was terminated by phenol/chloroform extraction, and the labelled DNA ethanol precipitated. The DNA was resuspended in 10 µl of water, and 5 µl of acrylamide load buffer (7.2.5 (ii)) added. The sample was then loaded on a 10% acrylamide gel and purified as described above (7.3.4 (i)).

(iii) Oligo-labelling of DNA fragments

Oligo-labelling of DNA fragments was performed using the reagents and a protocol provided by Bresatec, based on a method described by Feinberg and Vogelstein (1983). Approximately 50 - 100 ng of a DNA restriction fragment was taken up in 6 µl of water, and denatured by heating at 100°C for 5 minutes and snap chilling on ice. The DNA was then mixed with 17 µl of a solution containing 25 µCi α-^{32}P-dATP, 40 mM dCTP, dGTP, dTTP, 100 mM Tris-HCl pH 7.6, 100 mM NaCl, 20 mM $MgCl_2$, 200 µg/ml BSA and 7.2 µg chemically synthesised hexamer primer. DNA Polymerase I, Klenow fragment (5 units) was added and the mixture incubated at 37°C for 20 minutes. The reaction mixture volume was then increased to 100 µl, and the reaction terminated by phenol/chloroform extraction. Carrier tRNA was added to 1 µg/ml, and 10 µl 7 M ammonium acetate and 250 µl ethanol added to precipitate the labelled DNA fragments. The probe was resuspended in water, and stored at -20°C until required. Before use, the probe was denatured by boiling for 5 - 10 minutes.

7.3.5 Transfer and hybridisation of DNA on membranes

DNA fragments fractionated by agarose gel electrophoresis were transferred to Genescreen membranes by a modified method of Southern (1975), using an LKB 2016 VacuGene vacuum blotting apparatus. Gels were placed on the apparatus, covered with 0.25 M HCl, and transfer carried out for 10 minutes under 50 cm/H_2O pressure. The acid solution was removed, and the gel washed with distilled water. The gel was then covered with 0.4 M NaOH solution, and the transfer continued for 60 minutes at 50 cm/H_2O pressure. Filters were washed twice in 2 x SSC for 5 minutes, before the DNA was UV-crosslinked to the membrane using a Stratagene UV Stratalinker 1800 set on 'auto cross link'. Filters were sealed in plastic bags, and prehybridised at 42°C for at least 4 hours, in a solution containing 50% (v/v) deionised formamide, 5 x Denhardts, 5 x SSPE, 0.5 % (w/v) SDS, and 100 µg/ml denatured sonicated salmon sperm DNA. Heat denatured probes were added, and allowed to hybridise for 12 - 16 hours at 42°C. Filters were washed at a stringency appropriate for the probe in use. Usually filters were washed in 2 x SSC/0.1% (w/v) SDS at room temperature for 30 minutes and then in 0.2 x SSC/0.1% (w/v) SDS at 65°C for an additional 30 - 60 minutes depending on the signal. The filters were then autoradiographed at -80°C with a tungsten intensifying screen. If filters were to be reprobed, they were first stripped by boiling in 0.5% (w/v) SDS for 10 minutes, and then prehybridised as above, before being probed again.

7.3.6 Manipulation and harvesting of LSCC HD2 cells

Approximately 16 hours prior to experimental manipulation, LSCC HD2 tissue culture cells were seeded in fresh media at approximately 1 x 10^6 cells per ml, in order to ensure that they were in an exponential phase of growth. The cells were harvested by sequential washing steps involving centrifugation (1,500 rpm, 5 minutes), decanting of the supernatant, and resuspension in 200 volumes of ice-cold PBS for a total of three times before being finally pelleted.

7.3.7 Nuclear extract preparation

(i) Small scale nuclear extract preparation from LSCC HD2 cells

Crude nuclear extract was prepared from LSCC HD2 tissue culture cells by the method of Strauss and Varshavsky (1984). Cells were harvested (7.3.6), washed in PBS, and pelleted by centrifugation (JA-10 rotor; 3,000 rpm for five minutes). All subsequent procedures were done at 4°C to minimise proteolytic degradation. The pellet was resuspended in 0.25 M sucrose plus 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, 1 mM DTT and 0.5 mM PMSF (Buffer A), and homogenised with fifteen strokes in a Dounce homogeniser. The cell lysate was diluted with three volumes of 2.0 M sucrose in Buffer A, and layered over a pad of 1.7 M sucrose in Buffer A. The nuclei were pelleted by centrifugation (HB-4 rotor; 12,000 rpm for 45 minutes) through the sucrose pad. Nuclear extract was made by resuspending the nuclear pellet in 0.4 M NaCl, 5 mM EDTA, 10 mM Tris-HCl pH 7.5, 1 mM DTT and 0.5 mM PMSF, and incubating at 4°C for 60 minutes, with occasional gentle mixing. The solution was transfered 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), and the extract was snap frozen in liquid nitrogen, and stored at -80°C.

(ii) Small scale nuclear extract preparation from chicken embryos

Nuclear extract was prepared using a combination of methods from Panyim *et al.* (1971) and Lee *et al.* (1988). Embryos (1 dozen) were removed from the egg, and placed in 30 ml of ice-cold 0.25 M sucrose, plus 0.01 M Tris-HCl pH 8.0, 0.01 M $MgCl_2$, 0.05 M $NaHCO_3$, 1 mM DTT and 0.5 mM PMSF (Grinding buffer). All subsequent procedures were done at 4°C to minimise proteolytic degradation of the extract. Embryos were homogenised for 1 minute in a Sorvall Omni-Mixer, then the homogenate filtered through four layers of cheesecloth, and a single layer of Miracloth to remove large particulate matter. Cells were pelleted by centrifugation (HB-4 rotor; 3,000 rpm for 10 minutes), washed once by resuspension in 30 ml of

grinding buffer plus 0.2% (v/v) Triton X-100 (washing buffer), and pelleted again by centrifugation (HB-4 rotor; 3,000 rpm for 10 minutes). The washed pellet was resuspended in an equal volume of
10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT and 0.5 mM PMSF (swelling buffer), and incubated at 4°C for 30 minutes. Cells were lysed by 5 passes through a 23 gauge needle and the cell lysate centrifuged in a microfuge for 1 minute. Nuclear extract was made by resuspending the nuclear pellet in an equal volume of 20 mM HEPES pH 7.9, 25% (v/v) glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1 mM DTT and 0.5 mM PMSF (extraction buffer), and incubating at 4°C for 60 minutes, with occasional gentle mixing. The solution was centrifuged (microfuge, 5 minutes), the supernatant collected, snap frozen in liquid nitrogen, and stored at -80°C.

(iii) Large scale nuclear extract preparation from chicken embryos

Crude nuclear extract was prepared from large numbers of chicken embryos by a scaled up and slightly modified version of the procedure described above (7.3.7(ii)). Approximately 100 dozen chicken embryos were removed from the eggs, and placed in 3 Litres of ice-cold 0.25 M sucrose, plus 0.01 M Tris-HCl pH 8.0, 0.01 M $MgCl_2$, 0.05 M $NaHCO_3$, 1 mM DTT and 0.5 mM PMSF (grinding buffer). All subsequent procedures were done at 4°C to minimise proteolytic degradation of the extract. Embryos were homogenised for 1 minute in a Sorvall Omni-Mixer, then the homogenate filtered through four layers of cheesecloth, and a single layer of Miracloth to remove large particulate matter. Cell were pelleted by centrifugation (JA-10 rotor; 5,000 rpm for 10 minutes), and washed twice by resuspension in 3 Litres of grinding buffer plus 0.2% (v/v) Triton X-100 (washing buffer), further homogenisation for 30 seconds, and pelleting by centrifugation (JA-10 rotor; 5,000 rpm for 10 minutes). The pellet was resuspended in an equal volume of 10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT and 0.5 mM PMSF (swelling buffer), and incubated at 4°C for 1 hour. Cells were lysed by 5 passes through a 23 gauge needle and the cell lysate centrifuged (SS-34 rotor; 8,000 rpm for 10 minutes). Nuclear extract was made by resuspending the nuclear pellet in an equal volume of 20 mM HEPES pH 7.9, 25% (v/v) glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1 mM DTT and 0.5 mM PMSF (extraction buffer), and incubating at 4°C for 60 minutes, with occasional gentle mixing. The solution was centrifuged (SS-34 rotor; 8,000 rpm for 10 minutes), and the supernatant snap frozen in liquid nitrogen, and stored at -80°C.

7.3.8 Protein concentration determination

Protein concentration of nuclear extracts was routinely determined by the method of Bradford (1976), using BSA as a standard.

7.3.9 Ammonium sulphate precipitation

Protein was precipitated from solution by addition of a saturated solution of ammonium sulphate to varying final concentrations of 0 - 80%. The precipitate was pelleted by centrifugation (microfuge, for 20 minutes), and resuspended in 0.5 ml TM buffer (7.2.5 (ii)) plus 0.1 M NaCl. The solution was dialysed extensively against the same buffer at 4°C, to remove ammonium sulphate.

7.3.10 Concentration of protein extract

(i) Amicon ultrafiltration stirred cell concentration

Protein was concentrated with the use of an Amicon ultrafiltration stirred cell (model 8200; 200 ml capacity) and ultrafiltration membrane (YM10; 10 kDa molecular mass cut off rating). Concentration was carried out according to the manufacturers instructions.

(ii) Amicon centriprep spin colum concentration

Small scale concentration of protein extract was carried out using an Amicon centriprep spin column concentrator (15 ml capacity) which incorporated a filtration membrane with a 10 kDa molecular mass cut off rating. Concentration was carried out according to the manafacturers instructions.

7.3.11 Chromatographic enrichment of nuclear extract components

(i) Heparin-Sepharose chromatography

Heparin-Sepharose CL-6B beads were rehydrated according to the manufacturers protocol, and packed into either a 10 ml column (for pilot scale chromatography), or into a 200 ml column 200 mm x 35 mm (for preparative scale chromatography), and allowed to settle under gravity. The column was then equilibrated with TM buffer plus 0.2 M NaCl. Crude nuclear extract was diluted to a final NaCl concentration of 0.2 M with TM buffer, and loaded onto the column. The column was washed with 2 - 3 column volumes of TM buffer plus 0.2 M NaCl, at a flow rate of 0.5 ml/minute. The washing step was repeated three times, with increasing concentrations of NaCl: 0.3, 0.4 and 1.0 M. In all cases, fractions of the washes were collected, snap frozen in liquid nitrogen, and stored at -80°C.

(ii) DEAE-Sepharose chromatography

DEAE-Sepharose beads were rehydrated according to the manufacturers protocol, and packed into a 10 ml column (for pilot scale chromatography) and allowed to settle under gravity. The column was then equilibrated with TM buffer plus 0.1 M NaCl. Heparin-Sepharose purified extract was diluted to a final NaCl concentration of 0.1 M with TM buffer, and loaded onto the column. The column was washed with 2 - 3 column volumes of TM buffer plus 0.1 M NaCl, at a flow rate of 0.5 ml/minute. The washing step was repeated four times, with increasing concentrations of NaCl: 0.2, 0.3, 0.4 and 1.0 M. In all cases, fractions of the washes were collected, snap frozen in liquid nitrogen, and stored at -80°C.

(iii) Sephacryl S-300 gel filtration chromatography

Sephacryl S-300 beads were rehydrated according to the manufacturers 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.

7.3.12 Detection and analysis of DNA binding proteins

(i) Gel retardation assay

For gel retardations involving oligonucleotide probes, one of the two complementary oligonucleotides containing the binding site of interest was kinased with γ^{32}P-ATP (7.3.4 (i)). The two strands were then annealed to form a double-stranded molecule by addition of 2 - 5 fold excess of cold complementary oligonucleotide to 100 ng of labelled oligonucleotide. Annealing was accomplished by heating the oligonucleotide mixture to 100°C for 3 minutes in a heating block, turning the heating block off, and allowing the probe to cool slowly to room temperature. For gel retardations involving restriction fragments, the fragments were labelled by endfilling with α^{32}P-dNTPs (7.3.4 (ii)). Cold double stranded competitor oligonucleotides, used in gel retardations, were made by annealing equal amounts of the two complementary strands as above. The concentrations of the oligonucleotides, used to prepare the probes and competitor DNAs, were determined by measuring their UV absorbance at a wavelength of 260 nm (7.3.26).

Aliquots of the oligonucleotides and cold competitor DNAs were also checked by visualising the DNA samples under UV light (254 nm) after running on polyacrylamide gels (7.3.2 (iii)) against markers of known concentration and staining with ethidium bromide. The specific activities of the ^{32}P-labelled oligonucleotide probes were determined by Cerenkov counting (7.3.25). DNA binding activity was assayed essentially as described (Schneider *et al.*, 1986). Protein (usually 5 - 10 µg) was added to a 20 µl mixture containing 100 pg of ^{32}P labelled probe, 25 mM HEPES pH 7.5, 1 mM EDTA, 150 mM NaCl, 5 mM DTT, 0.5 mM PMSF and 10% (v/v) glycerol. If cold competitor DNA was used in an assay, it was added to the mixture before the protein. After incubation at room temperature for 30 minutes, 5 µl of gel retardation load buffer was added to the sample, and the sample immediately loaded onto a polyacrylamide gel. The products of a DNA binding reaction were resolved on a 10 - 12% (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 pH 8.5 at 250 - 400 V at 4°C. After the dye had run the required distance, the gel was dried using a Hoefer slab gel drier, and autoradiographed (7.3.24) at -80°C with a tungsten intensifying screen, and Kodak X-OMAT AR-50 film.

(ii) Modification interference assay

A restriction fragment was used as the probe for the modification interference assay. It was labelled by endfilling with α^{32}P-dNTPs (7.3.4 (ii)) and gel purified (7.3.2 (iii)). The probe was then partially depurinated by incubation for 10 minutes at room temperature with 2.5 volumes of formic acid. The reaction was stopped by adjusting the reaction mix to 0.3 M sodium acetate pH 7.0 and 0.1 mM EDTA and the DNA ethanol precipitated (7.3.2 (i)). The modified probe was then used for gel retardation experiments with crude nuclear extract as described above (7.3.12 (i)). The position of bound probe in the gel retardation was identified by autoradiography (7.3.24) and purified by excising the appropriate band from the gel and eluting in 400 µl TE buffer overnight at 37°C. The isolated DNA, along with control probe DNA, were cleaved at the modified purine residues by treating with 10% (v/v) piperidine for 30 minutes at 90°C. Samples were then dried *in vacuo* and resuspended in 50 µl of water. Samples were once more dried *in vacuo* and resuspended in 50 µl of water and dried *in vacuo* again. The dried pellets were resuspended in an amount of loading buffer such that each sample contained a comparable amount of radioactivity. Samples were analysed on a 12% denaturing polyacrylamide gel after boiling for 3 minutes at 100°C (7.3.2 (iv)).

7.3.13 SDS-polyacrylamide gel electrophoresis of proteins

(i) Gel electrophoresis

Protein was electrophoresed on vertical 14 x 14 x 0.05 cm SDS-polyacrylamide gels (Laemmli, 1970). The gel mix contained 10% (w/v) acrylamide (38:1 acrylamide: bisacrylamide), 2.5% (w/v) SDS, 375 mM Tris-HCl pH 8.8, 0.2% (w/v) APS and 0.1% (v/v) TEMED. The gels also comprised a 2 cm 4% (w/v) acrylamide stacking gel containing 125 mM Tris-HCl pH 6.8, 2.5% (w/v) SDS, 0.2% (w/v) APS and 0.1% (v/v) 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 and 0.1% (w/v) SDS), and gels were run at 20 mA until the dye reached the bottom of the gel.

(ii) Coomassie staining

For large amounts of protein, the protein in the gel was visualized by staining with 0.1% (w/v) Coomassie brilliant blue in 50% (v/v) methanol, 10% (v/v) 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 washed with two changes of distilled water, for 20 minutes, before the gel was dried onto a sheet of 3 MM Whatman paper, on a Hoefer slab gel drier.

(iii) Silver staining

When low levels of protein (< 200 ng) were present in the polyacrylamide gels, the protein was stained with silver by a modification of the simplified method described by Heueshoven and Dernick (1985). Gels were fixed overnight in 30% (v/v) ethanol, 10% (v/v) acetic acid. The gel was washed thoroughly with distilled water to attain a near neutral pH, then sensitized prior to staining in Farmer's Reducer (30% (w/v) potassium ferricyanide, 60% (w/v) sodium thiosulphate, 10% (w/v) sodium carbonate), 1 g in 200 ml water, for 30 seconds. The Farmer's Reducer was removed by washing with two changes of distilled water, 10 minutes for 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, containing 0.5 ml per litre formaldehyde. The developer was changed twice, when it became yellow-brown (after 10 - 60 seconds). Protein bands appeared after the last change of developer. The reaction was stopped with 1% (v/v) acetic acid, and the gel washed with two changes of distilled water, 10 minutes for each change, before the gel was dried between two sheets of cellophane paper on a Hoefer slab gel drier.

7.3.14 Southwestern analysis

Protein was transferred to PVDF membrane (Immobilon-P) by the method of Matsudaira (1988), using a Hoefer transfer electrophoresis unit. Briefly, protein to be transferred was fractionated on a 10% SDS-polyacrylamide gel (7.3.13 (i)). The gel was transferred to an Immobilon-P membrane in 10 mM CAPS and 10% (v/v) HPLC-grade methanol at 400 mA for 30 minutes. Extra tracks containing MW-SDS-200 markers (7.2.7 (ii)) and duplicates of the protein samples to be transferred were separated from the bulk of the gel before transfer and silver stained (7.3.13 (iii)). The membrane was processed essentially as described by Miskimins *et al.* (1985). The membrane was rinsed with binding buffer (25 mM HEPES pH 7.5, 1 mM EDTA, 150 mM NaCl, 5 mM DTT, 0.5 mM PMSF and 10% (v/v) glycerol) and allowed to air dry overnight at 4°C. The membrane was then blocked with binding buffer supplemented with 5% (w/v) Carnation nonfat dry milk for 3 hours at 4°C. This was followed by a rinse in binding buffer plus 0.25% (w/v) Carnation nonfat dry milk. The filter was probed in binding buffer plus 0.25% (w/v) Carnation nonfat dry milk, 100 µg/ml sonicated, denatured salmon sperm DNA and 10^6 cpm/ml probe, for 18 hours. After hybridisation, the membrane was washed twice for 15 minutes each time at 4°C in binding buffer plus 0.25% (w/v) Carnation nonfat dry milk and 0.1% (w/v) Triton X-100, then air dried, and autoradiographed (7.3.24). The size of the hybridizing band was estimated by comparison with markers from the silver stained half of the gel.

7.3.15 Oligoscreening of cDNA λgt11 expression libraries

(i) Probe preparation

Ladder probes for this type of screen were concatenated by ligation to 5 - 1 0 copies of the double stranded binding site probes (H1 box 14 + and -, C/EBP + and -, H1 Box 25 + and -, and *eng*; 7.2.8). Probes were prepared by kinasing 100 - 500 ng of each strand with γ-^{32}P-ATP (7.3.4 (i)), and gel purifying the labelled DNA away from the unincorporated label. The complementary strands were annealed by heating the oligonucleotide mixture to 100°C for 3 minutes in a heating block, turning the heating block off, and allowing the probe to cool slowly to room temperature. Double stranded oligonucleotides were ligated in a 20 µl volume containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 0.5 mM ATP and 1 unit of T4 DNA ligase at room temperature for 4 hours. The reaction was terminated by phenol/chloroform extraction, and the probe ethanol precipitated in the presence of ammonium acetate to remove any remaining unincorporated label. The probe was resuspended in 200 µl of TE buffer. A small aliquot of ligated probe was run on a

denaturing acrylamide gel (7.3.2 (iv)) to ensure ligation products were of the required size.

(ii) Vinson *et al.* (1988) method

λ Libraries were plated onto 150 mm L-agar plates at a density of approximately 50,000 plaques/plate. A culture of *E.coli* strain Y1090 was grown overnight in LMM-broth plus 50 µg/ml ampicillin. The phage were absorbed to 0.5 ml of the overnight culture for 15 minutes at room temperature, before 10 ml of LMM-agarose was added, and the sample plated onto L-agar plates supplemented with 50 µg/ml ampicillin. Plates were incubated at 42°C for four hours, overlayed with IPTG-impregnated nitrocellulose filters, and incubated for an additional 6 hours at 37°C. Duplicate filters, were prepared by overlaying a second IPTG-impregnated filter onto the plate after removal of the first filter. The second filter was left in place for an addition 2 hours at 37°C. After removal from the plates, filters were air dried for 15 minutes at room temperature. All additional processing steps were carried out at 4°C. Filters were placed in 200 ml of 25 mM HEPES pH 7.9, 5 mM $MgCl_2$, 40 mM NaCl, 0.5 mM DTT (binding buffer), plus 6 M guanidine-HCl, and shaken gently for 5 minutes. Half of the solution was decanted, and diluted by the addition of an equal volume of binding buffer. The filters were transfered to the diluted solution (3 M guanidine-HCl), and shaken gently for a further 5 minutes. This dilution procedure was repeated 4 times, followed by two washes in binding buffer. The filters were blocked by incubation for 30 minutes in binding buffer supplemented with 5% (w/v) Carnation nonfat dry milk, followed by a 5 minute wash in binding buffer plus 0.25% (w/v) Carnation nonfat dry milk. Filters were then probed in binding buffer plus 0.25% (w/v) Carnation nonfat dry milk and 100 µg/ml sonicated, denatured salmon sperm DNA with approximately 10^6 cpm/ml probe for 12 hours. Filters were washed three times, 5 minutes each time, in binding buffer plus 0.25% (w/v) Carnation nonfat dry milk. The filters were then allowed to air dry and were autoradiographed (7.3.24) at -80°C with a tungsten intensifying screen.

(iii) Kalionis and O'Farrell (1993) method

Screening was performed essentially as described in Sambrook *et al.* (1989), with modifications from Kalionis and O'Farrell (B. Kalionis, personal communication; note, method recently submitted for publication to EMBO J., 1993). Briefly, λ libraries were plated onto 150 mm L-agar plates at a density of not more than 12,500 plaques/plate. A culture of *E.coli* strain Y1090 was grown overnight in LMM-broth plus 50 µg/ml ampicillin. The phage were absorbed to 0.5 ml of the overnight

culture for 15 minutes at room temperature, before 10 ml of LMM-agarose was added, and the sample plated onto L-agar plates without ampicillin. Plates were incubated at 42°C for 4 hours, overlayed with IPTG-impregnated Optibond nylon reinforced nitrocellulose filters (Schleicher and Schuell), and incubated for an additional 6 hours at 37°C. Duplicate filters were prepared by overlaying a second IPTG-impregnated filter onto the plate after removal of the first filter. The second filter was left in place for a further 12 hours at 37°C. After removal from the plates, filters were air dried for 15 minutes at room temperature. All additional processing steps were carried out at 4°C. Filters were placed in 25 ml per filter of 25 mM HEPES pH 7.9, 3 mM $MgCl_2$, 40 mM KCl, 1 mM DTT (Binding buffer), plus 6 M guanidine-HCl, and shaken gently for 5 minutes. Half of the solution was decanted, and diluted by the addition of an equal volume of binding buffer. The filters were transfered to the diluted solution (3 M guanidine-HCl), and shaken gently for a further 5 minutes. This dilution procedure was repeated 4 times, followed by two washes in binding buffer. The filters were blocked by incubation for at least 60 minutes in binding buffer supplemented with 5% (w/v) Carnation nonfat dry milk, and 100 µg/ml sonicated, denatured salmon sperm DNA, followed by a 5 minute wash in binding buffer plus 0.25% (w/v) Carnation nonfat dry milk. Filters were probed in 2.5 ml per filter of binding buffer plus 0.25% (w/v) Carnation nonfat dry milk and 100 µg/ml sonicated, denatured salmon sperm DNA with 25 ng/ml (~ 10^6 cpm/ml) probe for 12 hours. Filters were washed twice, 10 minutes each time, in binding buffer plus 0.25% (w/v) Carnation nonfat dry milk, and 0.1% (w/v) Triton X-100. The filters then allowed to air dry and were autoradiographed (7.3.24) at -80°C with a tungsten intensifying screen.

7.3.16 Screening of cDNA and genomic libraries

λ Libraries were plated onto 150mm L-agar plates to give an optimum number of plaques for screening purposes. For both the λgt11 cDNA and λ Charon 4a genomic libraries approximately 5 x 10^4 plaques/plate were screened. A culture of the appropriate strain of *E. coli* (7.2.4 (i)) was grown overnight in LMM-broth. The phage were absorbed to 0.5 ml of the overnight culture for 15 minutes at room temperature 10 ml of LMM-agarose was added and the sample plated onto L-agar plates. Growth was continued throughout the day until almost a confluent plate of lysis was obtained (usualy 6 hours for cDNA libraries and 10 hours for genomic libraries). Plates were chilled at 4°C and the phage lifted in duplicate onto plaquescreen filters. The phage were lysed by a 2 minute treatment in an autoclave set at 105°C and the DNA baked onto the filters for 2 hours at 80°C. Filters were

prehybridised at 42°C for at least 4 hours, in a solution containing 50% (v/v) deionised formamide, 5 x Denhardts, 5 x SSPE, 0.1 % (w/v) SDS, and 100 µg/ml denatured sonicated salmon sperm DNA. Heat denatured probes were added, and allowed to hybridise for 12 - 16 hours at 42°C. Filters were washed at a stringency appropriate for the probe in use. Usually filters were washed in 2 x SSC/0.1% (w/v) SDS at room temperature for 30 minutes and then in 0.2 x SSC/0.1% (w/v) SDS at 65°C for an additional 30 - 60 miutes depending on the signal. The filters were then autoradiographed at -80°C with a tungsten intensifying screen. If filters were to be reprobed, they were first stripped by boiling in 0.5% (w/v) SDS for 10 minutes, and then prehybridised as above, before being probed again.

7.3.17 Purification of λ DNA

Approximately 10^5 λ phage were absorbed to 0.5 ml of an overnight culture of *E.coli* strain Y1090 at room temperature for 15 minutes. This was added to 50 ml LMM-broth + 50 µg/ml ampicillin, and incubated overnight at 37°C with shaking. Cells were removed by centrifugation (SS-34 rotor; 5,000 rpm for 5 minutes), and the supernatant treated with 50 µg crude Deoxyribonuclease I, and 100 µg Ribonuclease A at room temperature for 30 minutes. The suspension was then spun (SS-34 rotor; 18,000 rpm for 15 minutes), and the supernatant collected. The phage were pelleted by centrifugation (SS-34 rotor; 18,000 rpm for 3 hours), and resuspended in 0.1 M Tris-HCl pH 8.0, 0.3 M NaCl and 0.2% (w/v) SDS. Proteinase K (200 µg) was added, and the sample incubated at 37°C for 60 minutes. The sample was then extracted with phenol/chloroform three times before the DNA was precipitated with ethanol and resuspended in 100 µl of TE. The yield of DNA from this method was approximately 100 µg, and the DNA was sufficiently clean for restriction enzyme digestion.

7.3.18 Isolation and analysis of RNA

(i) Isolation of RNA from chicken embryos

RNA was extracted from nine day chicken embryo tissue following the method of Chomczynski and Sacchi (1987). Tissue was placed in an ice-cold solution of 4 M guanidinium isothiocyanate, 25 mM sodium citrate pH 7.0, 0.5% (w/v) Sarkosyl, and 0.1 M β-mercaptoethanol (1 g of tissue/10 ml solution). Tissue was immediately homogenised for 1 minute in a Sorvall Omni-Mixer, and 1 ml of 2 M sodium acetate pH 4.0, 2 ml chloroform/isoamyl alcohol (49:1), and 10 ml phenol added per gram of tissue. The solution was mixed, then placed on ice for 15 minutes, followed by centrifugation (HB-4 rotor; 8,000 rpm for 20 minutes at 4°C). The supernatant was

collected, and RNA precipitated by mixing in one volume of isopropanol and standing at -20°C for 15 minutes. After centrifugation (HB-4 rotor; 8,000 rpm for 20 minutes at 4°C), the RNA pellet was dissolved in 4M guanidinium isothiocyanate, 25 mM sodium citrate pH 7.0, 0.5% (w/v) Sarkosyl and 0.1 M β-mercaptoethanol (1 ml/g of tissue), and reprecipitated with one volume of isopropanol. The RNA was stored at -20°C as a precipitate.

(ii) Isolation of poly(A)$^+$ RNA

Polyadenylated RNA (poly(A)$^+$) was isolated from total RNA using a Pharmacia mRNA purification kit. Approximately 5 mg of total RNA was processed (4 x 1.25 mg aliquots). Each 1.25 mg aliquot was dissolved in 1 ml TE buffer, and denatured by heating to 65°C for 5 minutes. A volume of 200 µl of 10 mM Tris-HCl pH 7.4, 1 mM EDTA and 3.0 M NaCl was added to the RNA sample, which was then loaded onto an oligo(dT)-cellulose spun column, and allowed to soak in under gravity. The spun column was then washed with 2 x 250 µl 10 mM Tris-HCl pH 7.4, 1 mM EDTA and 0.5 M NaCl, followed by 3 x 250 µl 10 mM Tris-HCl pH 7.4, 1 mM EDTA and 0.1 M NaCl. Finally, the poly(A)$^+$ RNA bound to the column was eluted by 4 x 250 µl of TE buffer, prewarmed to 65°C. Glycogen (20 µg) was added to the eluate, and the RNA precipitated with ethanol. The Poly(A)$^+$ RNA precipitate was stored at -20°C.

(iii) Agarose gel electrophoresis of RNA

Electrophoresis of RNA was performed on 1.2% (w/v) agarose gels, containing 2.2 M formaldehyde and 1 x formaldehyde gel-running buffer (20 mM MOPS pH 7.0, 8 mM sodium acetate and 1 mM EDTA). RNA samples were ethanol precipitated, and resuspended in 4.5 µl DEPC-treated water, 2 µl 5 x formaldehyde gel-running buffer, 3.5 µl formaldehyde, and 10 µl formamide. Samples were heated to 65°C for 15 minutes prior to loading. The gel was submerged in 1 x formaldehyde gel-running buffer and run at 4 volts/cm.

(iv) Northern analysis of RNA

RNA fractionated on agarose gels (7.3.18 (iii)) was transferred onto Genescreen membranes using an LKB 2016 VacuGene vacuum blotting apparatus, with 10 x SSC as the transfer buffer. Transfer was continued for 60 minutes at 50 cm/H_2O. Filters were washed twice in 2 x SSC for five minutes, before the RNA was UV-crosslinked to the membrane using a Stratagene UV Stratalinker 1800 set on 'auto cross link'. Filters were sealed in plastic bags, and prehybridised at 42°C for at least 4 hours, in a

solution containing 50% (v/v) deionised formamide, 5 x Denhardts, 5 x SSPE, 0.5 % (w/v) SDS, 10% (w/v) Dextran sulphate, and 100 µg/ml denatured sonicated salmon sperm DNA. Heat denatured probes were added, and allowed to hybridise for 12 - 16 hours at 42°C. Filters were washed in 2 x SSC/0.1% (w/v) SDS at room temperature for 30 minutes and then in 0.2 x SSC/0.1% (w/v) SDS at 65°C for an additional 30 - 60 miutes depending on the signal. The filters were then autoradiographed at -80°C with a tungsten intensifying screen. If filters were to be reprobed, they were first stripped by boiling in 0.5% (w/v) SDS for 10 minutes, and then prehybridised as above, before being probed again.

7.3.19 Polymerase chain reaction (PCR) analysis

PCR was carried out using a Perkin-Elmer/Cetus DNA thermal cycler (model 0993-8412). PCR reactions were performed in a 50 µl volume containing 10 mM Tris-HCl pH 8.3, 50 mM KCl, 0.001% (w/v) gelatin, 2 mM $MgCl_2$, 200 µM dNTP's, 1 µg each primer and 1 unit *Taq* polymerase. The reaction protocol involved 30 cycles, each made up of 1 minute 95°C (denaturation), 1 minute 56°C (annealing), and 2 minutes 72°C (extension).

7.3.20 Preparation of exonuclease III deletion subclones

Series of exonuclease III deletion subclones were generated using a Promega Erase-a-base kit according to the manufacturers instructions.

7.3.21 Dideoxy sequencing

(i) Preparation of template

Approximately 2 µg of pBluescript clone DNA, prepared using the miniscreen procedure (2.3.1 (ii)), diluted in 18 µl TE buffer, was treated with 2 µl of Ribonuclease A (10 mg/ml) for 15 minutes at 37°C. The sample was incubated for an additional 15 minutes after 5 µl of 1 M NaOH/1 mM EDTA had been added. Finally the sample was dialysed by spinning through a Sepharose CL-6B spin column. Sepharose CL-6B spin columns were prepared as follows. A 0.5 ml Eppendorf tube was pierced with a 21 gauge syringe needle, so that only only two-thirds of the needle bevel emerged through the bottom. Approximately 20 µl of sterile acid washed glass beads (200 µm diameter) were added to the tube, followed by 500 µl of Sepharose slurry. The tube was then placed inside a 2 ml Eppendorf tube and the column spun at 3,000 rpm for 5 minutes in an microfuge. The 2 ml tube was replaced with a 1.5 ml Eppendorf tube, the sample added to the top of the packed Sepharose, and the column spun at

3,000 rpm for 5 minutes. The eluate recovered in the 1.5 ml tube was used immediately for annealing or was stored at -20°C.

(ii) Template/primer annealing

Template (7 µl) as prepared above (7.3.21 (i)) was incubated with 5 ng of USP or RSP (7.2.8), in a 10 µl reaction mixture containing 40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$ and 50 mM NaCl for 30 minutes at 37°C and then 30 minutes at room temperature. Primed tempates were used immediately for sequencing or stored at -20°C.

(iii) Polymerisation

Primed template DNA (10 µl) was extended at room temperature for 3 - 5 minutes in a reaction containing 2 µl (1.5 µM dGTP, 1.5 µM dCTP and 1.5 µM dTTP), 10 µCi α-^{35}S-dATP, 1 µl 100 mM DTT and 2 units T7 DNA polymerase. Aliquots (3.5 µl) of the extension mixture were then added to 2.5 µl of each termination mix, A (80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddGTP and 50 mM NaCl), C (80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddGTP and 50 mM NaCl), G (80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddGTP and 50 mM NaCl), and T (80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddGTP and 50 mM NaCl) and incubated at 37°C for 3 - 5 minutes. Reactions were terminated by adding 4 µl of formamide loading buffer, and heating to 100 °C for 3 minutes. A volume of 1 µl was loaded onto a denaturing polyacrylamide gel (7.3.2 (iv)).

7.3.22 λgt11 Fusion protein production

(i) Generation of λgt11 lysogens

E.coli strain Y1089 was grown to an OD_{600} of 0.8 in LMM-broth supplemented with 50 µg/ml ampicillin. The Y1089 cells were infected with the λgt11 recombinant phage at a multiplicity of infection of approximately 5 for 20 minutes at 30°C. Cells were diluted, and plated on an L-agar plate at a estimated density of 200 cells per plate. Plates were incubated overnight at 30°C. Single colonies were picked, and tested for temperature sensitivity at 42°C. Colonies that showed retarded growth at 42°C compared to 30°C were assumed to be lysogens.

(ii) Induction of fusion protein

LMM-Broth (2 ml) plus 50 µg/ml ampicillin was innoculated with a single colony of the Y1089 recombinant lysogen, and grown to OD_{600} 0.5. Fusion protein was induced by heat shock at 42°C for 20 minutes, followed by the addition of IPTG to

10 mM, and incubation at 37°C for a further 1 - 2 hours. Cells from 2 ml of culture were harvested by centrifugation (microfuge, 5 minutes), and lysed by heating to 100°C for 5 minutes in 100 µl 2% (w/v) SDS and 10% (v/v) β-mercaptoethanol. Total cell lysate (10 ml) was analysed on a SDS-polyacrylamide gel (7.3.13 (i)), and proteins visualised by silver staining (7.3.13 (iii)).

7.3.23 pGEX-1 Fusion protein production

(i) Induction of fusion protein

L-Broth (2 ml) plus 50 µg/ml ampicillin was innoculated with a single colony of *E.coli* strain ED8799 transformed with a recombinant pGEX-1 construct, and grown at 37°C overnight. The overnight culture was diluted 1 in 10 into fresh L-broth plus 50 µg/ml ampicillin, and allowed to grow at 37°C for 1 hour. Fusion protein was induced by the addition of IPTG to 0.1 mM, and further incubation at 37°C for 6 - 7 hours. For analysis of fusion protein production, cells from 1 ml of culture were harvested by centrifugation (microfuge, 5 minutes), and lysed by heating to 100°C for 5 minutes in 100 µl 2% (w/v) SDS and 10% (v/v) β-mercaptoethanol. The cell lysate was then spun (microfuge, 10 minutes), the supernatant recovered, snap frozen in liquid nitrogen, and stored at -80°C. Supernatant samples (10 µl) were analysed on SDS-polyacrylamide gels (7.3.13 (i)), and protein visualised by silver staining (7.3.13 (iii)).

(ii) Inclusion body preparation of fusion protein

L-Broth (100 ml) plus 50 µg/ml ampicillin was innoculated with a single colony of *E.coli* strain ED8799 transformed with a recombinant pGEX-1 construct, and grown at 37°C overnight. The overnight culture was diluted 1 in 10 into fresh L-broth plus 50 µg/ml ampicillin, and allowed to grow at 37°C for 1 hour. Fusion protein was induced by the addition of IPTG to 0.1 mM, and further incubation at 37°C for 6 - 7 hours. Inclusion bodies were prepared, washed, solubilized, and refolded essentially as described in Sambrook *et. al.* (1989) (adapted from Marston (1987)). The inclusion bodies after being solubilized (~ 10 ml) were dialysed extensively overnight against 0.1 M NaCl TM buffer (3 changes of 1 Litre each) at 4°C. The extract was snap frozen in liquid nitrogen, and stored at -80°C. Extract (10 µl) was analysed on SDS-polyacrylamide gels (7.3.13 (i)), and protein visualised by coomassie staining (7.3.13 (ii)).

7.3.24 Autoradiography

Radioactively 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 3 MM Whatman 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 amount 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 using an Agfa Curix 60 model X-ray film developer.

7.3.25 Cerenkov counting

The specific activities of ^{32}P-labelled DNA probes and amounts of ^{32}P-labelled DNA-protein complexes excised from gel retardation gels were determined by Cerenkov counting (Ross and Rasmussen, 1974) using a LKB 1214 Rackbeta model liquid scintillation counter according to the manufacturers instructions.

7.3.26 UV spectrophotometry

The concentrations of nucleic acid samples were determined by measuring their UV absorbance at a wavelength of 260 nm using a Shimadzu UV-160A recording spectrophotometer.

7.3.27 Containment facilities

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.

BIBLIOGRAPHY

BIBLIOGRAPHY

Allis, C.D., Glover, C.V.C., Bowen, J.K. and Gorovsky, M.A. (1980) Cell **20**, 609-617

Allis, C.D., Richman, R., Gorovsky, M.A., Ziegler, Y.S., Touchstone, B., Bradley, W.A. and Cook, R.G. (1986) J. Biol. Chem. **261**, 1941-1948

Alterman, R.M., Ganguly, S., Schulze, D.H., Marzluff, W.F., Schildkraut, C.L. and Skoultchi, R.I. (1984) Miol. Cell. Biol. **4**, 123-132

Artishevsky, A., Wooden, S., Sharma, A., Resendez, E. and Lee, A.S. (1987) Nature **328**, 823-827

Auwerx, J. and Sassone-Corsi, P. (1991) Cell **64**, 983-993

Aziz, N. and Munro, H.N. (1987) Proc. Natl. Acad. Sci. USA. **84**, 8478-8482

Baker, B.S. (1989) Nature **340**, 521-524

Baniahmad, A., Steiner, C., Köhne, A.C. and Renkawitz, R. (1990) Cell **61**, 505-514

Bannon, G.A. and Gorovsky, M.A. (1984) in 'Histone Genes: Structure, Organisation, and Regulation' (Stein, G.S., Stein, J.L. and Marzluff, W.F. eds.), J. Wiley and Sons, New York, 163-179

Bannon, G.A., Bowen, J.K., Yao, M. and Gorovsky, M.A. (1984) Nucl. Acids Res. **4**, 1961-1975

Beato, M. (1989) Cell **56**, 335-344

Bell, S.P., Learned, M., Jantzen, H.M. and Tjian, R. (1988) Science **241**, 1192-1197

Bengal, E., Flores, O., Krauskopf, A., Reinberg, D. and Aloni, Y. (1991) Mol. Cell. Biol. **11**, 1195-1206

Benne, R. (1990) Trends Genet. **6**, 177-181

Benoist, C. and Chambon, P. (1981) Nature **290**, 304-310

Berg, J.M. (1988) Proc. Natl. Acad. Sci. USA. **85**, 99-102

Beug, H., Doederlein, G., Freudenstein, C. and Graf, T. (1982) J. Cell. Physiol. Suppl. **1**, 195-207

Bickmore, W.A., Oghene, K., Little, M.H., Seawright, A., van Heyningen, V. and Hastie, N.D. (1992) Science **257**, 235-237

Biedenkapp, H., Borgmeyer, U., Sipple, A.E. and Klempnauer, K.H. (1988) Nature **335**, 835-837

Bienz, M. and Pelham, H.R.B. (1986) Cell **45**, 753-760

Biggin, M.D. and Tijan, R. (1988) Cell **53**, 699-711

Biggin, M.D. and Tijan, R. (1989) Trends Genet. **5**, 377-383

Bird, A.P. (1986) Nature **321**, 209-213

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. and Noll, M. (1986) Cell **47**, 1033-1040

Boyle, W.J., Smeal, T., Defize, L.H., Angel, P., Woodgett, J.R., Karin, M. and Hunter, T. (1991) Cell **64**, 573-584

Bradford, M.M. (1976) Anal. Biochem. **72**, 248-254

Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R. and Nasmyth, K. (1985) Cell **41**, 41-48

Brand, A.H., Micklem, G. and Nasmyth, K. (1987) Cell **51**, 709-719

Brawerman, G. (1987) Cell **48**, 5-6

Brent, R. and Ptashne, M. (1985) Cell **43**, 729-736

Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tijan, R. (1986) Science **234**, 47-52

Brown, D.T., Wellam, S.E. and Sittman, D.B. (1985) Mol. Cell. Biol. **5**, 2879-2886

Brush, D., Dodgson, J.B., Choi, O., Stevens, P.W. and Engel, J.D. (1985) Mol. Cell. Biol. **5**, 1307-1317

Buratowski, S., Hahn, S., Guarente, L. and Sharp, P.A. (1989) Cell **56**, 549-561

Busch, S.J. and Sassone-Corsi, P. (1990) Trends Genet. **6**, 36-40

Butler, W.B. and Mueller, G.C. (1973) Biochim. Biophys. Acta **294**, 481-496

Carbon, P., Murgo, S., Ebel, J., Krol, A., Tebb, G. and Mattaj, I.W. (1987) Cell **51**, 71-79

Carey, M. (1991) Curr. Opin. Cell Biol. **3**, 452-460

Chang, D.D. and Sharp, P.A. (1989) Cell **59**, 789-795

Chen, M., Shimada, T., Moulton, A.D., Cline, A., Humphries, R.K., Maizel, J. and Nienhuis, A.W. (1984) J. Biol. Chem. **259**, 3933-3943

Cheng, G., Nandi, A., Clerk, S. and Skoultchi, A.I. (1989) Proc. Natl. Acad. Sci. USA. **86**, 7002-7006

Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) Cell **53**, 11-24

Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156-159

Churchill, M.E.A. and Travers, A.A. (1991) Trends Biochem. Sci. **16**, 92-97

Clarke, L. and Carbon, J. (1976) Cell **9**, 91-99

Clerc, R.G., Corcoran, L.M., Le Bowitz, J.H., Baltimore, D. and Sharp, P.A. (1988) Genes Dev. **2**, 1570-1581

Cole, K.D., Kandala, J.C. and Kistler, W.S. (1986) J. Biol. Chem. **261**, 7178-7183

Coles, L.S. and Wells, J.R.E. (1985) Nucl. Acids Res. **13**, 585-594

Coles, L.S., Robins, A.J., Madley, L.K. and Wells, J.R.E. (1987) J. Biol. Chem. **262**, 9656-9663

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science **209**, 1406-1414

Cormack, B.P. and Struhl, K. (1992) Cell **69**, 685-696

Courey, A.J. and Tjian, R. (1988) Cell **55**, 887-898

Dailey, L., Hanly, S.M., Roeder, R.G. and Heintz, N. (1986) Proc. Natl. Acad. Sci. USA. **83**, 7241-7245

Dailey, L., Roberts, S.B. and Heintz, N. (1987) Mol. Cell. Biol. **7**, 4582-4584

Dailey, L., Roberts, S.B. and Heintz, N. (1988) Genes Dev. **2**, 1700-1712

Dalton, S., Coleman, J.R. and Wells, J.R.E. (1986) Mol. Cell. Biol. **6**, 601-606

Dalton, S. (1987) Ph.D. Thesis, University of Adelaide, South Australia

Dalton, S. and Wells, J.R.E. (1988*a*) EMBO J. **7**, 49-56

Dalton, S. and Wells, J.R.E. (1988*b*) Mol. Cell. Biol. **8**, 4576-4578

Dalton, S., Robins, A.J., Harvey, R.P. and Wells, J.R.E. (1989) Nucl. Acids Res. **15**, 1745-1756

D'Andrea, R.J., Coles, L.S., Lesnikowski, C., Tabe, L. and Wells, J.R.E. (1985) Mol. Cell. Biol. **5**, 3108-3115

Darnell, J.E. (1982) Nature **297**, 365-371

DeLisle, A.J., Graves, R.A., Marzluff, W.F. and Johnson, L.F. (1983) Mol. Cell. Biol. **3**, 1920-1929

Desjarlais, J.R. and Berg, J.M. (1992) Proc. Natl. Acad. Sci. USA. **89**, 7345-7349

Desplan, C., Theis, J. and O'Farrell, P.H. (1988) Cell **54**, 1081-1090

Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucl. Acids Res. 12, 387-395

Dodgson, J.B., Strommer, J. and Engle, J.D. (1979) Cell **17**, 879-887

Dunaway, M. and Dröge, P. (1989) Nature **341**, 657-659

Dynan, W.S. (1989) Cell **58**, 1-4

Dynan, W.S. and Tjian, R. (1983) Cell **35**, 79-87

Eisen, A., Taylor, W.E., Blumberg, H. and Young, E.T. (1988) Mol. Cell. Biol. **8**, 4552-4556

Engel, J.D. and Dodgson, J.B. (1981) Proc. Natl. Acad. Sci. USA. **78**, 2856-2860

Ephrussi, A., Church, G.M., Tonegawa, S. and Gilbert, W. (1985) Science **227**, 134-140

Evans, R.M. and Hollenberg, S.M. (1988) Cell **52**, 1-3

Fairall, L., Rhodes, D. and Klug, A. (1986) J. Mol. Biol. **192**, 577-591

Faisst, S. and Meyer, S. (1992) Nucl. Acids Res. **20**, 3-26

Farnham, P.J. and Schimke, R.T. (1986) Mol. Cell. Biol. **6**, 2392-2401

Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. **132**, 6-13

Felsenfeld, G. and McGhee, J.D. (1986) Cell **44**, 375-377

Finney, M., Ruvkun, G. and Horvitz, H.R. (1988) Cell **55**, 757-769

Fletcher, C., Heintz, N. and Roeder, R.G. (1987) Cell **51**, 773-781

Franza, B.R., Josephs, S.F., Gilman, M.Z., Ryan, W. and Clarkson, B. (1987) Nature **330**, 391-395

Fried, M. and Crothers, D.M. (1981) Nucl. Acids Res. **7**, 6505-6526

Friedman, K.D., Rosen, N.L., Newman, P.J. and Montgomery, R.R. (1988) Nucl. Acids Res. **16**, 8718

Friedman, A.D. and McKnight, S.L. (1990) Genes Dev. **4**, 1416-1426

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. and Noll, M. (1986) Cell **47**, 735-746

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Proc. Natl. Acad. Sci. USA. **85**, 8998-9002

Galas, D.J. and Schmitz, A. (1978) Nucl. Acids Res. **5**, 3157-3170

Gallinari, P., La Bella, F. and Heintz, N. (1989) Mol. Cell. Biol. **9**, 1566-1575

Garner, M.M. and Revzin, A. (1981) Nucl. Acids Res. **9**, 3047-3060

Gehring, W.J., Müller, M., Affolter, M., Percival-Smith, A., Billeter, M., Qian, Y.Q., Otting, G. and Wüthrich, K. (1990) Trends Genet. **6**, 323-329

Georgiev, O. and Birnstiel, M.L. (1985) EMBO J. **4**, 481-489

Geyer, P.K., Green, M.M. and Corces, V.G. (1990) EMBO J. **9**, 2247-2256

Ghosh, S. and Baltimore, D. (1990) Nature **344**, 678-682

Gill, G. and Ptashne, M. (1987) Cell **51**, 121-126

Gillies, S.D., Morrison, S.L., Oi, V.T. and Tonegawa, S. (1983) Cell **33**, 717-728

Giniger, E., Varnum, S.M. and Ptashne, M. (1985) Cell **40**, 767-774

Giniger, E. and Ptashne, M. (1987) Nature **330**, 670-672

Gjerset, C., Gorka, C., Hasthorpe, S., Lawrence, J.J. and Eisen, H. (1982) Proc. Natl. Acad. Sci. USA. **79**, 2333-2337

Gonzalez, G.A. and Montminy, M.R. (1989) Cell **59**, 675-680

Goodbourn, S., Burstein, H. and Maniatis, T. (1986) Cell **45**, 601-610

Gralla, J.D. (1989) Cell **57**, 193-195

Gralla, J.D. (1991) Cell **66**, 415-418

Graves, R.A. and Marzluff, W.F. (1984) Mol. Cell. Biol. **4**, 351-357

Greaves, D.R., Wilson, F.D., Lang, G. and Kioussis, D. (1989) Cell **56**, 979-986

Grosveld, G.C., Rosenthal, A. and Flavell, R.A. (1982) Nuc. Acids Res. **10**, 4951-4971

Grosveld, F., Blom van Assendelft, G., Greaves, D.R. and Kollias, G. (1987) Cell **51**, 975-985

Grunstein, M. (1990) Ann. Rev. Cell. Biol. **6**, 643-678

Hanly, S.M., Bleecker, G.C. and Heintz, N. (1985) Mol. Cell. Biol. **5**, 380-389

Harrison, S.C. and Aggarwal, A.K. (1990) Ann. Rev. Biochem. **59**, 933-969

Hartshorne, T.A., Blumberg, H. and Young, E.T. (1986) Nature **320**, 283-287

Harvey, R.P., Krieg, P.A., Robins, A.J., Coles, L.S. and Wells, J.R.E. (1981) Nature **294**, 49-53

Harvey, R.P., Robins, A.J. and Wells, J.R.E. (1982) Nucl. Acids Res. **10**, 7851-7863

Harvey, R.P., Whiting, J.A., Coles, L.S., Krieg, P.A. and Wells, J.R.E. (1983) Proc. Natl. Acad. Sci. USA. **80**, 2819-2823

Heery, D.M., Gannon, F. and Powell, R. (1990) Trends Genet. **6**, 173

Heintz, N., Zernik, M. and Roeder, R.G. (1981) Cell **24**, 661-668

Heintz, N., Sive, H.L. and Roeder, R.G. (1983) Mol. Cell. Biol. **3**, 539-550

Heintz, N. (1991) Biochim. Biophys. Acta **1088**, 327-339

Hentschel, C.C. and Birnstiel, M.L. (1981) Cell **25**, 301-313

Hentzen, P.C., Rho, J.H. and Bekhor, I. (1984) Proc. Natl. Acad. Sci. USA. **81**, 304-307

Heueshoven, J. and Dernick, R. (1985) Electrophoresis **6**, 103-112

Hoffmann, I. and Birnstiel, M.L. (1990) Nature **346**, 665-668

Hoopes, B.C. and McClure, W. (1987) in '*Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology.' Volume 2 (Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E. eds.), American Society for Microbiology, Washington, D.C., 1231-1240

Hope, K.A. and Struhl, K. (1986) Cell **46**, 885-894

Hsu, T., Gogos, J.A., Kirsh, S.A. and Kafatos, F.T. (1992) Science **257**, 1946-1950

Huang, H. and Cole, R.D. (1984) J. Biol. Chem. **259**, 14237-14242

Hunter, T. and Karin, M. (1992) Cell **70**, 375-387

Ingraham, H.A., Chen, R., Mangalam, H.A., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G. (1988) Cell **55**, 519-529

Jackson, D.A. and Cook, P.R. (1985) EMBO J. **4**, 919-925

Jackson, S.P. and Tijan, R. (1988) Cell **55**, 125-133

Johnson, P.F., Landshulz, W.H., Graves, B.J. and McKnight, S.L. (1987) Genes Dev. **1**, 133-146

Johnston, S.A., Salmeron, J.M. and Dincher, S.S. (1987) Cell **50**, 143-146

Jones, K.A., Yamamoto, K.R. and Tijan, R (1985) Cell **42**, 559-572

Jones, N.C., Rigby, P.W.J. and Ziff, E.B. (1988) Genes Dev. **2**, 267-281

Kadonaga, J.T. and Tjian, R. (1986) Proc. Natl. Acad. Sci. USA. **83**, 5889-5893

Kadonaga, J.T., Jones, K.A. and Tijan, R. (1986) Trends Biochem. Sci. **11**, 20-23

Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tijan, R. (1987) Cell **51**, 1079-1090

Kaftos, F.T., Orr, W. and Delidakis, C. (1985) Trends Genet. **1**, 301-305

Kakkis, E., Riggs, K.J., Gillespie, W. and Calame, K. (1989) Nature **339**, 718-721

Kalionis, B. and O'Farrell, P.H. (1993) submitted for publication to EMBO J.

Kamakaka, R.T. and Thomas, J.O. (1990) EMBO J. **9**, 3997-4006

Karim, F.D., Urness, L.D., Thummel, C.S., Klemsz, M.J., McKercher, S.R., Celada, A., Van Beveren, C., Maki, R.A., Gunther, C.V., Nye, J.A. and Graves, B.J. (1990) Genes Dev. **4**, 1451-1453

Kato, G.J. and Dang, C.V. (1992) FASEB J. **6**, 3065-3072

Kedes, L.H. (1979) Ann. Rev. Biochem. **48**, 837-70

Kirsh, A.L., Groudine, M. and Challoner, P.B. (1989) Genes Dev. **3**, 2172-2179

Klug, A. and Rhodes, D. (1987) Trends Biochem. Sci. **12**, 464-469

Kost, T.A., Theodorakis, N. and Hughes, S.H. (1983) Nucl. Acids Res. **11**, 8287-8301

Kouzarides, T. and Ziff, E.B. (1989) Nature **340**, 568-571

Kozak, M. (1987) Nucl. Acids Res. **15**, 8125-8148

Krasnow, M.A., Saffman, E.E., Kornfeld, K. and Hogness, D. (1989) Cell **57**, 1031-1043

Kraulis, P.J., Raine, A.R.C., Gadhavi, P.L. and Laue, E.D. (1992) Nature **356**, 448-450

Kremer, E.J. and Kistler, W.S. (1992) Gene **110**, 167-173

Krieg, P.A., Robins, A.J., D'Andrea, R. and Wells, J.R.E. (1983) Nucl. Acids Res. **11**, 619-627

La Bella, F., Sive, H.L., Roeder, R.G. and Heintz, N. (1988) Genes Dev. **2**, 32-39

La Bella, F., Gallinari, P., McKinney, J. and Heintz, N. (1989) Genes Dev. **3**, 1982-1990

Laemmli, U.K. (1970) Nature **227**, 680-685

Laird, P.W. (1989) Trends Genet. **5**, 204-208

Lamb, P. and McKnight, S.L. (1991) Trends Biochem. Sci. **16**, 417-422

Landau, N.R., St. John, T.P., Weissman, I.L., Wolf, S.C., Silverstone, A.E. and Baltimore, D. (1984) Proc. Natl. Acad. Sci. USA. **81**, 5836-5840

Landolfi, N.F., Yin, X., Capra, J.D. and Tucker, P.W. (1989) Biotechniques **7**, 500-504

Landschultz, W.H., Johnson, P.F., Adashi, E.Y., Graves, B.J. and McKnight, S.L. (1988) Genes Dev. **2**, 786-800

Lassar, A.B., Buskin, J.N., Lockshon, D., Davis, R.L., Apone, S., Hauschka, S.D. and Weintraub, H. (1989) Cell **58**, 823-831

Learned, R.M., Cordes, S. and Tijan, R. (1985) Mol. Cell. Biol. **5**, 1358-1369

Lee, K.A.W., Bindereif, A. and Green, M.R. (1988) Gene Anal. Techn. **5**, 22-31

Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. and Wright, P.E. (1989) Science **245**, 635-637

Lee, M.S., Gottesfeld, J.M. and Wright, P.E. (1991) FEBS Lett. **279**, 289-294

Leff, S.E., Rosenfeld, M.G. and Evans, R.M. (1986) Ann. Rev. Biochem. **55**, 1091-1117

Leff, S.E., Evans, R.M. and Rosenfeld, M.G. (1987) Cell **48**, 517-524

Lennox, R.W. and Cohen, L.H. (1984) in 'Histone Genes: Structure, Organisation, and Regulation' (Stein, G.S , Stein, J.L. and Marzluff, W.F. eds.), J. Wiley and Sons, New York, 373-395

Levine, B.J., Chodchoy, N., Marzluff, W.F. and Skoultchi, A.I. (1987) Proc. Natl. Acad. Sci. USA. **84**, 6189-6193

Levine, M., Rubin, G.M. and Tijan, R. (1984) Cell **38**, 667-673

Levine, M. and Hoey, T. (1988) Cell **55**, 537-540

Levine, M. and Manley, J.L. (1989) Cell **59**, 405-408

Lewin, B. (1990) Cell **61**, 1161-1164

Lichtsteiner, S., Wuarin, J. and U. Schibler, U. (1987) Cell **51**, 963-973

Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977) Cold Spring Harbour Symp. Quant. Biol. **42**, 1047-1051

Lillie, J.W., Green, M. and Green, M.R. (1986) Cell **46**, 1043-1051

Lin, Y. and Green, M.R. (1991) Cell **64**, 971-981

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P. and Yamamoto, K.R. (1991) Nature **352**, 497-505

Ma, J. and Ptashne, M. (1987*a*) Cell **48**, 847-853

Ma, J. and Ptashne, M. (1987*b*) Cell **51**, 113-119

Ma, J. and Ptashne, M. (1988) Cell **55**, 443-446

Mahajan, P.B. and Thompson, E.A. (1990) J. Biol. Chem. **265**, 16225-16233

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell **15**, 687-701

Mannironi, C., Bonner, W.M. and Hatch, C.L. (1989) Nucl. Acids Res. **22**, 9113-9126

Margottin, F., Dujardin, G., Gérard, M., Egly, J., Huet, J. and Sentenac, A. (1991) Science **251**, 424-426

Marmorstein, R., Carey, M., Ptashne, M. and Harrison, S.C. (1992) Nature **356**, 408-414

Marston, F.A. (1987) in 'DNA Cloning: A Practical Approach' volume 3 (Glover, D.M. ed.), IRL Press, Oxford, 59-88

Marzluff, W.F. and Graves, R.A. (1984) in 'Histone Genes: Structure, Organisation, and Regulation' (Stein, G.S., Stein, J.L. and Marzluff, W.F. eds.), J. Wiley and Sons, New York, 281-315

Marzluff, W.F. and Pandey, N.B. (1988) Trends Biochem. Sci. **13**, 49-52

Matsudaira, P. (1988) J. Biol. Chem. **262**, 10035-10038

Maxson, R., Cohn, R., Kedes, L. and Mohun, T. (1983) Ann. Rev. Genet. **17**, 239-277

McKnight, S.L. (1982) Cell **31**, 355-365

Meeks-Wagner, D. and Hartwell, L.H. (1986) Cell **44**, 43-52

Melton, D.W., McEwan, C., McKie, A.B. and Reid, A.M. (1986) Cell **44**, 319-328

Mermod, N., O'Neill, E.A., Kelly, T.J. and Tjian, R. (1989) Cell **58**, 741-753

Mezquita, J., Conner, W., Winkfein, R.J. and Dixon, G.H. (1985) J. Mol. Evol. **21**, 209-219

Miller, A.M. and Nasmyth, K.A. (1984) Nature **312**, 247-251

Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. **4**, 1609-1614

Miskimins, W.K., Roberts, M.P., McClelland, A. and Ruddle, F.H. (1985) Proc. Natl. Acad. Sci. USA. **82**, 6741-6744

Mitchell, P.J. and Tijan, R. (1989) Science **245**, 371-378

Mitchelson, K. and Moss, T. (1987) Nucl. Acids Res. **15**, 9577-9596

Montiel, J.F., Norbury, C.J., Tuite, M.F., Dobson, M.J., Mills, J.S., Kingsman, A.J. and Kingsman, S.M. (1984) Nucl. Acids Res. **12**, 1049-1067

Morris, T., Marashi, F., Weber, L., Hickey, E., Greenspan, D., Bonner, J., Stein, J. and Stein, G. (1986) Proc. Natl. Acad. Sci. USA. **83**, 981-985

Morse, R.H. and Simpson, R.T (1988) Cell **54**, 285-287

Müller, M.M., Ruppert, S., Schaffner, W. and Matthias, P. (1988) Nature **336**, 544-551

Müller, H., Sogo, J.M. and Schaffner, W. (1989) Cell **58**, 767-777

Müller, H. and Schaffner, W. (1990) Trends Genet. **6**, 300-304

Müllner, E.W., Neupert, B. and Kühn, L.E. (1989) Cell **58**, 373-382

Müller-Immerglück, M.M., Schaffner, W. and Matthias, P. (1990) EMBO J. **9**, 1625-1634

Murphy, S., Moorefield, B. and Pieler, T. (1989) Trends Genet. **5**, 122-126

Murre, C., Schonleber, P., McCaw, P.S. and Baltimore, D. (1989) Cell **56**, 777-783

Nacheva, G.A., Guschin, D.Y., Preobrazhenskaya, O.V., Karpov, V.L., Ebralidse, K.K. and Mirzabekov, A.D. (1989) Cell **58**, 27-36

Nardelli, J., Gibson, T. and Charnay, P. (1992) Nucl. Acids Res. **20**, 4137-4144

Nasmyth, K. (1982) Ann. Rev. Genet. **16**, 439-500

Nasmyth, K., Seddon, A. and Ammerer, G. (1987) Cell **49**, 549-558

Neelin, J.M., Callahan, P.X., Lamb, D.C. and Murray, K. (1964) Can. J. Biochem. **42**, 1743-1752

Norman, C., Runswick, M., Pollock, R. and Treisman, R. (1988) Cell **55**, 989-1003

Nunberg, J.H., Kaufman, R.J., Chang, A.C.Y., Cohen, S.N. and Schimke, R.T. (1980) Cell **19**, 355-364

O'Hare, P. and Goding, C.R. (1988) Cell **52**, 435-445

Old, R.W. and Woodland, H.R. (1984) Cell **38**, 624-626

O'Neill, E.A. and Kelly, T.J. (1988) J. Biol. Chem. **263**, 931-937

O'Neill, E.A., Fletcher, C., Burrow, C.R., Heintz, N., Roeder, R.G. and Kelly, T.J. (1988) Science **241**, 1210-1213

Orkin, S.H. (1990) Cell **63**, 665-672

Osley, M.A. and Hereford, L. (1982) Proc. Natl. Acad. Sci. USA. **79**, 7689-7693

Osley, M.A., Gould, J., Kim, S., Kane, M. and Hereford, L. (1986) Cell **45**, 537-544

Osley, M.A. (1991) Ann. Rev. Biochem. **60**, 827-861

Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affloter, M., Gehring, W.J. and Wüthrich, K. (1990) EMBO J. **9**, 3085-3092

Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M.C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. and Brown, L.G. (1987) Cell **51**, 1091-1104

Palmer, J.M. and Folk, W.R. (1990) Trends Biochem. Sci. **15**, 300-304

Pandey, N.B. and Marzluff, W.F. (1987) Mol. Cell. Biol. **7**, 4557-4559

Panyim, S. and Chalkley, R. (1969) Biochem. Biophys. Res. Commun. **37**, 1042-1049

Panyim, S., Bilek, D. and Chalkley, R. (1971) J. Biol. Chem. **246**, 4206-4215

Passmore, S., Maine, G.T., Elble, R., Christ, C. and Tye, B. (1988) J. Mol. Biol. **204**, 593-606

Paulson, J.C. (1989) Trends Biochem. Sci. **14**, 272-294

Pavletich, N.P. and Pabo, C.O. (1991) Science **252**, 809-817

Payre, F. and Vincent, A. (1991) EMBO J. **10**, 2533-2541

Peltz, S.W. and Ross, J. (1987) Mol. Cell. Biol. **7**, 4345-4356

Penotti, F.E. (1991) J. Theor. Biol. **150**, 385-420

Peterson, M.L. and Perry, R.P. (1986) Proc. Natl. Acad. Sci. USA. **83**, 8883-8887

Pfeifer, K., Prezant, T. and Guarente, L. (1987) Cell **49**, 19-27

Pollock, R. and Treisman, R. (1990) Nucl. Acids Res. **18**, 6197-6204

Powell, L.M., Wallis, S.C., Pease, R.J., Edwards, Y.H., Knott, T.J. and Scott, J. (1987) Cell **50**, 831-840

Prendergast, G.C. and Ziff, E.B. (1992) Trends Genet. **8**, 91-96

Ptashne, M. and Gann, A.A.F. (1990) Nature **346**, 329-331

Pugh, B.F. and Tijan, R. (1990) Cell **61**, 1187-1197

Reeder, R.H. (1990) Trends Genet. **6**, 390-395

Renkawitz, R. (1990) Trends Genet. **6**, 192-197

Reuter, G., Giarre, M., Farah, J., Gausz, J., Spierer, A. and Spierer, P. (1990) Nature 344, 219-223

Roeder, R.G. (1991) Trends Biochem. Sci. **16**, 402-407

Rosenberg, U.B., Schröder, C., Preiss, A., Kienlin, A., Côté, S., Riede, I. and Jäckle, H. (1986) Nature **319**, 336-339

Ross, H.H. and Rasmussen, G.T. (1974) in 'Liquid scintillation counting; recent developments' (Stanley, P.E. and Scoggins, B.A. eds.), Acedemic Press, New York, 363-382

Ruiz i Altaba, A., Perry-O'Keefe, H. and Melton, D.A. (1987) EMBO J. **6**, 3065-3070

Ryoji, M. and Worcel, A. (1985) Cell **40**, 923-932

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) 'Molecular Cloning: A Laboratory Manual' Second edition. Cold Spring Harbor Laboratory Press, New York

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA. **74**, 5463-5467

Sariban, E., Wu, R.S., Erickson, L.C. and Bonner, W.M. (1985) Mol. Cell. Biol. **5**, 1279-1286

Sassone-Corsi, P., Sisson, J. and Verma, I.M. (1988) Nature **334**, 314-319

Sawadogo, M. and Sentenac, A. (1990) Ann. Rev. Biochem. **59**, 711-754

Schleif, R. (1988) Science **241**, 1182-1187

Schneider, R., Gander, I., Müller, V., Mertz, R. and Winnacker, E.L.(1986) Nucl. Acids Res. **14**, 1303-1317

Schultz, M.C., Reeder, R.H. and Hahn, S. (1992) Cell **69**, 697-702

Schümperli, D. (1988) Trends Genet. **4**, 187-191

Schwabe, J.W.R. and Rhodes, D. (1991) Trends Biochem. Sci. **16**, 291-296

Shannon, M.F. and Wells, J.R.E. (1987) J. Biol. Chem. **262**, 9664-9668

Shapiro, M.B. and Senapathy, P. (1987) Nucl. Acids Res. **17**, 7155-7175

Sharma, A., Bos, T.J., Pekkala-Flagan, A., Vogt, P.K. and Lee, A.S. (1989) Proc. Natl. Acad. Sci. USA. **86**, 491-495

Sharp, P.A. (1987) Science **235**, 766-771

Shatkin, A.J. (1976) Cell **9**, 645-653

Shaw, G. and Kamen, R. (1986) Cell **46**, 659-667

Shepherd, J.C.W., McGinnis, W., Carrasco, A.E., De Robertis, E.M. and Gehring, W.J. (1984) Nature **310**, 70-71

Singh, H., LeBowitz, J.H., Baldwin, A.S. and Sharp, P.A. (1988) Cell **52**, 415-423

Singh, H., Clerc, R.G. and LeBowitz, J.H. (1989) Biotechniques **7**, 252-261

Sittman, D.B., Chiu, I., Pan, C., Cohn, R.H., Kedes, L.H. and Marzluff, W.F. (1981) Proc. Natl. Acad. Sci. USA. **78**, 4078-4082

Sittman, D.B., Graves, R.A. and Marzluff, W.F. (1983) Proc. Natl. Acad. Sci. USA. **80**, 1849-1853

Sive, H.L. and Roeder, R.G. (1986) Proc. Natl. Acad. Sci. USA. **83**, 6382-6386

Sive, H.L., Heintz, N. and Roeder, R.G. (1984) Mol. Cell. Biol. **4**, 2723-2734

Sive, H.L., Heintz, N. and Roeder, R.G. (1986) Mol. Cell. Biol. **6**, 3329-3340

Smale, S.T. and Baltimore, D. (1989) Cell **57**, 103-113

Smith, B.J., Walker, J.M. and Johns, E.W. (1980) FEBS Lett. **112**, 42-44

Smith, D.B. and Johnson, K.S. (1988) Gene **67**, 31-40

Sommer, H., Beltrán, J., Huijser, P., Pape, H., Lönnig, W., Saedler, H. and Schwarz-Sommer, Z. (1990) EMBO J. **9**, 605-613

Sorger, P.K. and Nelson, H.C.M. (1989) Cell **59**, 807-813

Sorger, P.K., Ammerer, G. and Shore, D. (1989) in 'Protein function: A practical approach' (Creighton, T.E. ed.), IRL press, Oxford, 199-223

Sorger, P.K. (1991) Cell **65**, 363-366

Southern, E.M. (1975) J. Mol. Biol. **98**, 503-517

Sparrow, D.B. (1991) Ph.D. Thesis, University of Adelaide, South Australia

Stein, G.S. and Stein, J.L. (1984) Mol. Cell. Biochem. **64**, 105-110

Stein, G.S., Stein, J.L. and Marzluff, W.F. (1984) 'Histone Genes: Structure, Organisation, and Regulation', J. Wiley and Sons, New York

Stern, S., Tanaka, M. and Herr, W. (1989) Nature **341**, 624-630

Stimac, E., Groppi, V.E. and Coffino, P. (1984) Mol. Cell. Biol. **4**, 2082-2090

Strauss, F. and Varshavsky, A. (1984) Cell **37**, 889-901

Stringer, K.F., Ingles, C.J. and Greenblatt, J. (1990) Nature **345**, 783-786

Struhl, K. (1989) Ann. Rev. Biochem. **58**, 1051-1077

Sturm, R.A. and Herr, W. (1988) Nature **336**, 601-604

Sturm, R.A., Das, G. and Herr, W. (1988*a*) Genes Dev. **2**, 1582-1599

Sturm, R.A., Dalton, S. and Wells, J.R.E. (1988*b*) Nucl. Acids Res. **16**, 8571-8586

Svaren, J. and Chalkley, R. (1990) Trends Genet. **6**, 52-56

Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, M. and Chambon, P. (1986) Nature **319**, 121-126

Tautz, D., Lehmann, R., Schnürch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jäckle, H. (1987) Nature **327**, 383-389

Theill, L.E., Castrillo, J., Wu, D. and Karin, M. (1989) Nature **342**, 945-948

Theodorakis, N.G. and Morimoto, R.I. (1987) Mol. Cell. Biol. **7**, 4357-4368

Thukral, S.K., Morrison, M.L. and Young, E.T. (1992) Mol. Cell. Biol. **12**, 2784-2792

Treisman, R. (1985) Cell **42**, 889-902

Triezenberg, S.J., Kingsbury, R.C. and McKnight, S.L. (1988) Genes Dev. **2**, 718-729

Turner, P.C., Aldridge, T.C., Woodland, H.R. and Old, R.W. (1983) Nucl. Acids Res. **11**, 4093-4107

Ullu, E. and Weiner, A.M. (1985) Nature **318**, 371-374

van Daal, A., White, E.M., Gorovsky, M.A. and Elgin, S.C.R. (1988) Nucl. Acids Res. **15**, 7487-7497

Van der Ploeg, L.H.T. (1987) Cell **51**, 159-161

van Holde, K.E., Lohr, D.E. and Robert, C. (1992) J. Biol. Chem. **267**, 2837-2840

van't Veer, L.J., Lutz, P.M., Isselbacher, K.J. and Bernards, R. (1992) Proc. Natl. Acad. Sci. USA. **89**, 8971-8975

van Wijnen, A.J., Massung, R.F., Stein, J.L. and Stein, G.S. (1988) Biochemistry **27**, 6534-6541

van Wijnen, A.J., Wright, K.L., Lian, J.B., Stein, J.L. and Stein, G.S. (1989) J. Biol. Chem. **264**, 15034-15042

van Wijnen, A.J., Owen, T.A., Holthuis, J., Lian, J.B., Stein, J.L. and Stein, G.S. (1991) J. Cell. Physiol. **148**, 174-189

Vinson, C.R., Sigler, P.B. and McKnight, S.L. (1989) Science **246**, 911-916

Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschulz, W.H. and McKnight, S.L. (1988) Genes Dev. **2**, 801-806

Weintraub, H. (1978) Nucl. Acids Res. **5**, 1179-1188

Weiss, M.A., Mason, K.A., Dahl, C.E. and Keutmann, H.T. (1990) Biochemistry **29**, 5660-5664

Wells, D. and McBride, C. (1989) Nucl. Acids Res. **17**, r311-r346

West, M.H.P. and Bonner, W.M. (1980) Biochemistry **19**, 3238-3245

White, R.J. and Jackson, S.P. (1992) Trends Genet. **8**, 284-288

Wickens, M. (1990) Trends Biochem. Sci. **15**, 277-281

Williams, T., Admon, A., Lüscher, B. and Tjian, R. (1988) Genes Dev. **2**, 1557-1569

Williams, T. and Tijan, R. (1991) Genes Dev. **5**, 670-682

Winslow, G.M., Hayashi, S., Krasnow, M., Hogness, D.S. and Scott, M.P. (1989) Cell **57**, 1017-1030

Winter, E., Levy, D. and Gordon, J.S. (1985) J. Cell Biol. **101**, 167-174

Wolffe, A.P. (1989) EMBO J. **8**, 527-537

Workman, J.L. and Roeder, R.G. (1987) Cell **51**, 613-622

Wu, R.S., Panusz, H.T., Hatch, C.L. and Bonner, W.M. (1986) CRC Crit. Rev. Biochem. **20**, 201-263

Xanthopoulos, K.G., Mirkovitch, J., Decker, T., Kuo, C.F. and Darnell, J.E. (1989) Proc. Natl. Acad. Sci. USA. **86**, 4117-4121

Yamamoto, K.R. (1985) Ann. Rev. Genet. **19**, 209-252

Yang, Y., Brown, D.T., Wellman, S.E. and Sittman, D.B. (1987) J. Biol. Chem. **262**, 17118-17125

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldman, K.A. and Meyerowitz, E.M. (1990) Nature **346**, 35-39

Yen, T.J., Machlin P.S. and Cleveland D.W. (1988) Nature **334**, 580-585

Young, R.A., Hagenbüchle, O. and Schibler, U. (1981) Cell **23**, 451-458

Young, R.A. and Davis, R.W. (1983*a*) Science **222**, 778-782

Young, R.A. and Davis, R.W. (1983*b*) Proc. Natl. Acad. Sci. USA. **80**, 1194-1198

Younghusband, H.B., Sturm, R. and Wells, J.R.E. (1986) Nucl. Acids Res. **14**, 635-644

Zarkower, D. and Hodgkin, J. (1992) Cell **70**, 237-249

Zinn, K. and Maniatis, T. (1986) Cell **45**, 611-618

Zlatanova, J. (1990) Trends Biochem. Sci. **15**, 273-276

Zweidler, A. (1984) in 'Histone Genes: Structure, Organisation, and Regulation' (Stein, G.S., Stein, J.L. and Marzluff, W.F. eds.), J. Wiley and Sons, New York 339-371