CONTROL MECHANISMS OF HIGHER EUKARYOTIC GENE TRANSCRIPTION - DIVERGENT HISTONE GENES

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

by

Richard Alan Sturm, B.Sc. (Hons.)

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

April, 1985.

CONTENTS

CONTROL MECHANISMS OF HIGHER EUKARYOTIC GENE TRANSCRIPTION

- DIVERGENT HISTONE GENES

This laboratory has isolated the entire gene complement of the histone multigene family from the chicken genome. The isolated chicken H2A and H2B protein encoding genes have been studied in detail in this thesis. Six divergent H2A/H2B gene pairs have been oriented and their intergene promoter regions sequenced. Comparative analysis has identified potential regulatory elements and a conserved spatial arrangement of the genes. Coding regions of the divergent genes are all approximately 350 bases apart. TATA boxes are separated by 180 bases that contain four CCAAT boxes and a 13 base motif that is also found near immunoglobulin gene promoter elements.

Functional testing of the intergene promoter region has been carried out by *in vitro* mutagenesis using pAT plasmid or M13 phage vectors and by gene transfer experiments in the nucleus of *Xenopus* oocytes. A 420 base-pair fragment containing the intergene region is sufficient to correctly promote and initiate the divergent transcripts. Mapping of the promoter region indicates that the two genes have overlapping promoter elements. The significance of this gene arrangement in the contol of histone protein expression is discussed.

STATEMENT

This thesis contains no material which 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 is made in the text. I consent to the thesis being made available for photocopying and for loan.

Richard Alan Sturm

ACKNOWLEDGEMENTS

I wish to acknowledge the assistance of and express my thanks to the following people:

Dr. Julian Wells, for his supervision throughout the course of this work.

Peter Wigley and Anna Koltunow, for their help, technical advice and many critical discussions.

Allan Robins, Dave Bird, Richard D'Andrea and Leeanne Coles and other members of the laboratory for their valuable practical advice during the initial stages of this work.

Professor W.H. Elliot, for permission to work in the Department of Biochemistry, University of Adelaide.

Allan Robins, for the use of his microcomputer in the preparation of this thesis, Deb Ehmann and Anne Chartier for their typing assistance and Jenny Cassady for the production of photographic diagrams.

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

CHAPTER 1: THESIS PERSPECTIVE - MOLECULAR LEVELS OF GENE CONTROL

1-1 Introduction

The theme of this thesis concerns the molecular mechanisms involved in cellular transcription of eukaryotic DNA. In particular, the control and co-ordination of a multigene family will be the focus of the experimental work presented in the main text of the thesis. Steps in the pathway from DNA to protein synthesis that may be involved in regulation in eukaryotes will be examined. An account of the histone gene systems of eukaryotic organisms and various techniques used to analyse DNA transcription will also be presented as a background to the experimental work of the chicken histone multigene family investigated here.

1-2 Levels in the Control of Gene Expression

The living biological cell is a marvellous machine (Alberts et al., 1983). Cells of higher metazoans are of many types with strikingly distinct morphology and function. Differential gene expression underlies many of the differences between cell types. Gene expression in turn is controlled at a number of levels. Cell types can develop differences by controlling:

(1) how or when a given gene is transcribed
(2) how the initial transcript is processed
(3) selecting which mRNAs in the cell nucleus are exported to the cytoplasm
(4) selecting which mRNAs in the cytoplasm are translated by ribosomes
(5) selectively stabilizing certain mRNA molecules in the cytoplasm
(6) once protein has been produced from the mRNA it can be post-translationally modified to impart different activities, properties or stabilities.

DNA transcription is especially subject to cellular regulation and a very important part of the control processes occurs at the first step, the synthesis of the primary RNA transcript (Derman et al., 1981). There are two general processes which need to be understood in the transcription of eukaryotic genes; firstly the nature of the DNA signals and molecular components required for the activation of any gene, and secondly how

transcription is regulated throughout development. That is, what determines whether a particular gene is transcribed at a specific time and in a specific cell.

1-3 Active Chromatin

The DNA of all eukaryotic organisms is closely associated with a wide variety of DNA-binding proteins of two general classes: the histones and the non-histone chromosomal proteins. The localised structure of the chromosome influences gene regulation since genes present in heterochromatin are not expressed (Weisbrod, 1982). On the other hand those genes which are being expressed in a tissue specific fashion exist in an active chromatin conformation which is different from that of the equivalent chromosome regions in tissues where the gene is inactive (Weisbrod, 1982). Active genes are readily accessible to reagents such as DNase I.

Irrespective of the characteristics of active genes described below, transcriptional activation requires an initial un-ravelling step of a domain of the chromosome containing the gene to be transcribed. The primary events involved in decondensation/condensation of chromatin domains are fundamental to the questions of gene control and differentiation but very little is known in this area. Thus, the secondary components directly involved in the act of transcription will be the major points of discussion.

1-3-1 The Nucleosome and Histone Proteins

The basic unit of chromatin is the nucleosome. This consists of 146 to 240 base pairs of DNA wound twice around a histone protein core. The core proteins are made up of four types H2A, H2B, H3 and H4 present in equal amounts in the form of an octamer. A fifth histone type, H1 and variants

such as $H1^0$ (Gjerset et al., 1982) and H5 (Aviles et al., 1978), is associated with the nucleosome as a linker between core particles. Major and minor subtypes of all the histone protein types except H4 have been found (Isenberg, 1979; Von Holt et al., 1979; Zweidler, 1980). The majority of these variants differ from the most common forms by only a few amino acids, however the variant residues could cause potentially important conformational differences between the subtypes.

Histone variants appearing in only one specialized tissue are found in sperm of several species (Zweidler, 1980). A histone which appears to be specific to erythroid cells of fish, amphibians, reptiles and birds is the H1 variant H5 (Aviles et al., 1978). $H1^0$, a mammalian histone closely related to H5 is found in a relatively wide range of terminally differentiated cells (Gjerset et al., 1982) and like H5 accumulates post-mitotically. Developmentally controlled variants also occur in a more general sense in some species. For example, in the sea urchin the forms of H2A, H2B and H1 which are expressed in the blastula are replaced by the late subtypes β, γ and ζ by the gastrula stage (Newrock et al., 1978).

The histones can exist in different forms by post-translational modifications. These modifications include acetylation, phosphorylation, ADP-ribosylation, methylation and ubiquitination (Isenberg, 1979). Histones in either their modified or unmodified forms have been associated with various states of activity of chromatin (Halleck and Gurley, 1982; Prentice et al., 1982). Phosphorylation of H2A for example has been found to be greater in the euchromatin fraction and a ubiquitinated form of the H2A protein has been shown to have a selective arrangement in the Drosophila genome (Levinger and Varshavsky, 1982). Whether these modifications are the cause or effect of changes in chromatin states is not known.

A histone variant which differs substantially from its major subtype is H2A.Z. This protein was initially isolated from mouse nucleosomes but has subsequently been found in all tissues and species studied (West and Bonner, 1980). Whether H2A.Z's sequence differences are related to its distribution in chromatin is yet to be determined. One histone variant which appears to be strongly associated with active chromatin is an H2A variant from Tetrahymena thermophila, known as hv-1. It is found exclusively in the transcriptionally active macronucleus and not at all in the transcriptionally inactive micronucleus (Allis et al., 1982).

As well as being linked with developmental stages, various histones are expressed differently during the cell cycle. Approximately 90% of histone synthesis is coupled to DNA synthesis but 10% proceeds throughout the cell cycle and is known as basal synthesis. This 10% includes the ongoing synthesis of the variants H2A.X, H2A.Z, H3.3 and H2B variants (Wu and Bonner, 1981).

The presence of these histone variants, their cross species conservation and developmental regulation suggests that they could well play a role in differentiation and selective gene expression.

A special type of nucleosomal subunit located on transcribed DNA regions with properties that facilitate transcription has been postulated (Prior et al., 1983) and termed the "lexosome". Applied to rDNA in particular, it is a speculative gene activation model involving a nucleosome - lexosome transition. The lexosome formation involves a symmetrical unfolding of the histone core and the binding of proteins in addition to the core histones. A loosening of DNA - histone interactions could lead to a transient displacement of histones during the passage of RNA polymerase molecules (see below).

1-3-2 Nucleosome Phasing and DNase Hypersensitive Sites

As mentioned, actively transcribed genes exhibit a structure different from that of non-transcribed regions when pancreatic DNase I is used to probe its accessibility. A sensitivity to DNase I of actively transcribed genes seems to be a general phenomenon (Weisbrod, 1982; Elgin, 1984) but only indicates a potential for gene transcription rather than transcription itself. The appearance of DNase I hypersensitive sites precedes the initiation of transcription and is apparent only in tissues that are committed to expressing a gene.

Studies of the mini-chromosome of SV40 and the chick β - globin gene indicate that hypersensitive sites correspond to regions of DNA that are possibly nucleosome free (Jakobovitis et al., 1980; Saragosti et al., 1980; McGhee et al., 1981). Data from rearranged SV40 genomes and from Drosophila mutants at the Sgs-4 locus indicate that sequences adjacent to genes can be critical for the formation of DNase I hypersensitive sites (Jongstra et al., 1984, Shermoen and Beckendorf, 1982). It is possible that nuclease sensitive sites are generated by the interaction of non-histone proteins with DNA by binding to the region itself. The Drosophila hsp70 gene once induced by heat shock binds a trans-acting factor (Wu, 1984a), however a DNase I sensitive chromatin structure is present at this 5' terminus of the coding sequence before activation (Wu, 1980).

The question of how nucleosomes are arranged on active and inactive DNA has three simplistic answers. They could be arranged randomly on the DNA, they may be located in unique positions that are identical in a homogeneous population of cells, or they may occupy a small number of distinct positions. The latter two of these arrangements are referred to as nucleosome "phasing". Phasing does exist for certain positions in the genome, however the extent of the nucleosome preference, its structural

reasons and its function is far from being understood (Zachau and Igo-Kemenes, 1981; Kornberg, 1981).

The functional significance of regular nucleosome arrangements on satellite DNA may be that, by modulating chromatin superstructure they affect chromosome recognition and pairing. The DNase I hypersensitive regions of protein coding genes, postulated to be nucleosome-free, may precisely position nucleosomes at their borders or influence the chromatin structure at the region of transcription and impose a short range order. However for genuine phasing of nucleosomes to exist, it would mean that the loss or addition of just a single base pair could have deleterious effects. A loosely preferred distribution of nucleosomes at specific sites would accommodate the genomic variations of individuals of a species that are found in nature (Jeffreys, 1979; Brown, 1982). At present research on nucleosome phasing is only at the level of nuclease digestion and blot hybridization.

1-3-3 Non-Histone Proteins Associated with Chromatin

The protein fraction most often associated with active nucleosomes is the high mobility group proteins (HMG). They are small highly conserved proteins and are present in 10% - 20% of nucleosomes. Two HMGs in particular have been associated with active chromatin, HMGs 14 and 17. These proteins can be eluted from chick erythrocyte chromatin with 0.35M NaCl, without any detectable change in the gross structure of individual nucleosomes. However with this depleted chromatin the globin gene is no longer hypersensitive to DNase I (Weisbrod and Weintraub, 1979). Reconstitution of the depleted chromatin with purified HMGs at a ratio of 1 mole HMG: 10 mole of nucleosomes successfully restores the DNase I sensitivity of the globin gene.

There is no tissue specificity associated with HMGs. HMGs 14 and 17 from brain nuclei are capable of restoring the hypersensitivity of the globin gene from salt depleted erythrocyte chromatin. The specificity of HMG binding resides in the recipient depleted chromatin; HMGs 14 and 17 from red blood cell chromatin fail to induce DNase I hypersensitivity to the globin gene of HMG - depleted chromatin from brain. The positioning of HMGs in active nucleosomes has been investigated and indicates a direct correspondence between chromosomal regions which are capable of interacting with HMGs and the regions which are highly sensitive to DNase I digestion. The general conclusion is that HMGs 14 and 17 occupy two binding sites at each end of the nucleosome core and interact with or cover the internucleosomal, histone - free DNA of the hypersensitive site (Weisbrod, 1982).

Much less is known about other non-histone proteins associated with active chromatin. Increasing attention is being paid to sequence specific DNA binding proteins. It has been suggested that specific regulatory proteins may bind to a gene promoter site independently of RNA polymerase. The polymerase could then bind directly to the promoter-recognition protein complex (Travers, 1983). General promoter recognizing proteins have indeed been purified and their role in transcription suggests that a pre-initiation complex is formed independently of the RNA polymerase (Davison et al., 1983; Gidoni et al., 1984).

The isolation and characterization of gene-specific regulatory proteins is just beginning. A 5S RNA gene transcription factor known as TFIIIA is required for the accurate initiation *in vitro* of Xenopus 5S genes (Korn, 1982; Brown, 1984). This factor interacts with residues 50-83 of the transcribed region and thus binds to an intragenic promoter. The large T - antigen of SV40 and polyoma acts as a positive regulator of virus

replication/late transcription and a negative regulator of its own early gene transcription by binding to three operator sites near its origin of replication (Schaffhausen, 1982). The E1A gene product of Adenovirus acts as a trans-acting transcriptional enhancer of the human β - globin gene transfected into Hela cells (Green, 1983b). Moreover a 36 base pair region of the β- globin promoter containing the TATA box is sufficient for such an interaction. Whether there is a direct contact between the E1A protein and the promoter has not been directly shown but it is a distinct possibility.

Positive control has also been shown by the sequence-specific binding of the glucocorticoid receptor complex to cloned fragments of murine mammary tumour virus DNA (Chandler et al., 1983). This binding mediates hormone responsiveness via the promoter. A partially purified fraction of chicken erythrocyte nuclei can confer hypersensitivity to the chick β - globin gene 5' flanking region in vitro, when allowed to form chromatin with purified histone proteins. One or more proteins may interact specifically with the sequence of the hypersensitive region (Emerson and Felsenfeld, 1984). Direct investigation of the binding of a heat shock activator protein to a specific DNA sequence, termed a heat-shock box, 5' to the Drosophila hsp 70 and hsp 82 genes has been done by mapping DNA exonuclease resistance sites in chromatin (Wu, 1984a and b).

1-3-4 Higher Order Structures

The nucleosome is only the first level of DNA packaging inside the cell. A hierarchy of superstructures compact the nucleosomal DNA chain culminating in the most condensed form of metaphase chromatin. A 100 Å fibre is generated by folding of the internucleosomal linker DNA; this fibre is then coiled into a 300 Å fibre probably by the crosslinking of

nucleosomes by H1 protein (Igo-Kemenes et al., 1982).

In both interphase and metaphase chromosomes the 300 Å chromatin fibre appears to be folded into loops or domains. Stretches of 35-85 kilobases of DNA are thought to be anchored to a supporting structure of the nucleus termed the matrix or scaffold. The distribution of DNA sequences and the location of proteins within chromatin domains are now being studied. The identification of a sequence-specific chromatin-loop organization for the tandemly repeated histone gene cluster of Drosophila has been reported (one loop for each repeated unit; Mirkovitch et al., 1984).

DNA endonucleases have been used to investigate higher order organization of active chromatin. Active genes have an overall DNase I sensitivity intermediate between that of the hypersensitive sites of active genes and a totally inactive gene (Weisbrod, 1982). Different genes are preferentially degraded in different cell types corresponding to the pattern of RNAs that the cells make. As stated above (1-3-2) DNase I sensitivity only indicates the potential for transcription of a gene and not transcription itself, therefore eukaryotic gene activation probably occurs in two stages. In the first stage a local region of chromatin, probably a domain, is modified to decondense it in preparation for transcription. Secondly, DNA transcription is activated in selected gene regions of decondensed loops by regulatory proteins binding to specific sites on the altered chromatin (1-3-3).

Domains or loops are essential elements of higher order chromatin structure and are related to units of transcription. Their structure, rearrangement, and time course of activation may control developmental regulation.

1-3-5 Role of Matrix

The nuclear matrix consists of a peripheral lamina, nuclear pore structures, a residual nucleolus, and an intranuclear matrix. It has an important role in the structural organization of DNA and is composed of three major polypeptides (Berezney, 1979). The DNA loops of interphase nuclei are anchored to the matrix possibly at specific sites along the chromosome. The preferential location of specific gene loci at specific sites has been examined in Drosophila polytene nuclei (Agard and Sedat, 1983). The topographical analysis reported was sufficient to determine an overall folding pattern, but could not align the genetic map with the topological arrangement.

The nuclear matrix is the site of DNA replication (Pardoll et al., 1980), possibly the site of nuclear RNA processing (Long et al., 1979) and has been shown to bind hormones in a specific manner. Isolated nuclear matrix has been found to preferentially contain actively expressed genes (Ciejek et al., 1983) indicating that the nuclear matrix is the site of nuclear transcription. The attachment of the hormone sensitive chicken ovalbumin gene is reversible when the hormone is withdrawn. Thus another level of control for gene expression may exist in the nucleus.

1-3-6 DNA Modification and Structure

In most higher organisms, DNA is modified by the conversion of a small percentage of cytosine residues to 5-methyl-cytosine. The major site of methylation is at CpG sequences and the methylation pattern is inherited automatically following DNA replication due to the specificity of eukaryotic methylases recognizing hemimethylated CpG dinucleotides. The control of gene activity by DNA methylation has been recognized as a logically attractive possibility. A region of DNA important for gene activation could be stably maintained through repeated cell divisions in

either a methylated or unmethylated form because of the inheritance of methylation patterns.

Methylation of DNA can supress transcription of some genes, however the activity of others are unaffected (Bird, 1984). In vitro site directed methylation of globin genes has shown that the methylation of specific CpG sequences between 760 base pairs upstream and 100 base pairs downstream of the 5' end of the human γ - globin gene can prevent transcription; methylation of CpG sequences elsewhere in the gene appear to be unimportant to transcription of the gene (Busslinger et al., 1983). In evaluating the effect of DNA methylation on gene expression it is significant that the levels of CpG methylation varies considerably between vertebrates and invertebrates. No 5-methyl-cytosine residues have been detected in Drosophila at all and no methylated gene as such has been detected in other invertebrates. It is therefore not clear at the moment what part DNA methylation plays in eukaryotic gene expression.

The structure and function of a DNA molecule can both be affected by the DNA sequence itself, torsional stress, base modification and the interaction of specific proteins with the DNA. Altered DNA conformations are likely to play an important part in DNA packaging into nucleosomes and in the sequence-specific recognition of DNA by proteins (Wu and Crothers, 1984). DNA is a dynamic molecule and alterations in its fine structure possibly play a role in controlling genes.

Conventional DNA is a right-handed double helix (the B-helix; B-DNA) but a range of right-handed structural helix families can exist (Cantor, 1981). One left-handed DNA structural variant has been described and is known as Z-DNA. Z-helix formation is favoured in alternating purine/pyrimidine stretches and is stabilized by the inclusion of the 5-methyl-cytosine base.

Supercoiled DNA is torsionally stressed and in response to this stress the molecule twists, bends or deforms to minimize its free energy. Changing a section of B-DNA to Z-DNA, right to left, untwists the supercoil, and thus supercoiling of DNA also favours Z-DNA formation. Another possible response of DNA to torsional stress is the formation of cruciforms, which requires the presence of inverted sequence repeats. This relieves some of the stress by denaturing several base pairs in the cruciform duplex junctions and loop outs at the end of the cruciform.

Direct evidence for Z-DNA formation in vivo comes from the binding of anti-Z antibodies to Drosophila polytene chromosomes. This binding has been shown to depend on the chromatin fixation techniques used and it is therefore not certain if Z-DNA does exist in vivo (Hill and Stollar, 1983). Many regions that have the potential to form Z-DNA are often located at important sites. Several eukaryotic enhancer elements contain potential Z-DNA regions (Nordheim and Rich, 1983).

Under relaxed conditions the structural conformation of normal right-handed DNA changes from place to place along the length of a DNA molecule depending on the base sequence. Specific sequence induced variation in the structure of Watson-Crick DNA may be of importance for proteins that bind to a consensus sequence. The recognition of the essential bases in the consensus may be influenced by the flanking sequences and influence protein binding affinity (Drew and Travers, 1984). Thus there may be a consensus of structure as well as sequence.

1-4 RNA Transcription and Transcriptional control

Transcriptional control, the rate of synthesis of nuclear RNA, is certainly of paramount importance in eukaryotic cells (Darnell, 1982). Differential stability of cytoplasmic mRNAs and RNA processing represent

important and as yet not fully investigated levels of control. The mechanical steps of mRNA formation are considerably more complex in eukaryotes than prokaryotes; the mRNA requires the addition of a 7-methyl-guanosine cap at the 5' end of primary transcripts, the addition of a 3' polyadenosine segment and the removal of introns/ligaton of exons (Nevins, 1983). All of these steps are likely to be subjected to regulation.

1-4-1 RNA Polymerases

The first level of transcriptional control in cells is in the choice of RNA polymerase. Three different RNA polymerases make RNA in eukaryotes. Polymerase II transcribes genes that encode proteins; polymerase I makes the large ribosomal RNAs 18S, 28S and 5.8S in a contiguous cistron; RNA polymerase III transcribes a variety of small RNAs including the tRNAs, 5S ribosomal RNA and some repetitive DNA sequences. The different polymerases can be identified by their sensitivity to the fungal toxin α - amanitin. RNA polymerase I is unaffected by α - amanitin, polymerase III is moderately sensitive and polymerase II is very sensitive. All three classes are multi-subunit structures of six to ten polypeptides and within each class heterogeneity of subunit composition has been observed. The three enzymes recognize different start and stop signals on the DNA.

1-4-2 Polymerase II Promoter Elements

A primary mRNA transcript is generated from a transcriptional unit of the genome containing appropriate signals for initiation and termination. A considerable body of information has been collected concerning the structure of eukaryotic promoters but as yet there is no data for a general definitive termination site for polymerase II transcription. The definition of gene promoters has relied on sequence comparisons to look for homologies, followed by surrogate genetics to probe the function of the sequence homologies as no one consensus promoter sequence, as such, exists

for eukaryotic genes. However a number of distinct, transcriptional control elements can be identified. This, in itself, may allow modulation of transcription of different genes by using different combinations of these elements. These elements are common to most genes, but are not all present in all cases; it appears that only a combination of them is required. Sequences essential for the transcription of polymerase II genes can be divided into those that act immediately 5' to the cap site of mRNA or those that can act in a cis fashion far upstream/downstream from the start of transcription. Sequences specific to individual genes or to a group of genes have been discovered and may be recognized in a tissue specific or inducible manner.

(i) Immediate 5' Elements

The region of initiation of an mRNA transcript usually consists of an A residue surrounded by pyrimidine residues (Breathnach and Chambon, 1981). Rather than having a single site of RNA initiation on the DNA sequence there may be heterogeneity in the start site of transcription (McKnight et al., 1981; Wigley et al., 1985; work presented in this thesis). It is assumed that the cap site of the mature mRNA is the transcription start site as well (Nevins, 1983).

An A-T rich region centered about 25-30 bases upstream from the mRNA start sites is found in most genes and is known as the Goldberg - Hogness or TATA box (Goldberg, 1979). It is identified by a consensus sequence TATA(A/T)A(A/T) (Corden et al., 1980) but the most important part of the box is an ATA stretch (Grosschedl et al., 1981). Virtually every gene analyzed to date, except the E2 and IVa2 genes of adenovirus and the late genes of SV40 and polyoma, have some form of the TATA sequence. The function of this sequence appears to be involved in the accurate

positioning of the start of transcription (Grosschedl and Birnstiel, 1980a).

A second region of homology, the CCAAT box, is positioned 70-90 bases upstream from the cap site of some genes. It has the canonical sequence GG PyCAATCT (Benoist et al., 1980). The generality of this box and the exact sequence (because of the heterogeneity of most CCAAT boxes identified in the literature) for its function are yet to be fully investigated. Work presented in this thesis shows that multiple CCAAT boxes are found in the promoter region of the divergent H2A/H2B histone gene pairs of the chicken.

It has been determined that this sequence is functionally important for the maximal expression of the rabbit β-globin gene (Dierks et al., 1983), but deletion of the CCAAT-like sequences of other genes has either had no effect (SV40; Benoist and Chambon, 1981) or led to a slight stimulation of transcription (H2A; Grosschedl and Birnstiel, 1980a). No functionally significant sequence with homology to the CCAAT box is found in the HSV TK gene (McKnight et al., 1981).

A third region, the -100 region, approximately 100 bases from the cap site of the rabbit β-globin gene has been identified as an upstream promoter element (Dierks et al., 1983). The important component of this region seems to be (T/A)CACCC which is present in almost all the β-like globin genes so far studied. Mapping of the promoter regions of HSV TK and the SV40 early genes suggest that they contain elements that are structurally and functionally related to the -100 region of the globin genes. The sequence CCGCCC is part of the recognition site of the second distal promoter of the TK gene (McKnight et al., 1984) and is located approximately 100 bases from the TK cap site. However the reverse of this sequence GGGCGG is found in the first distal promoter region of the TK gene and is present in multiple copies in the SV40 and a related monkey promoter

(Gidoni et al., 1984). These elements have variously been referred to as G-C rich elements (box). The SV40 early promoter contains six repeats of the CCGCCC box and the production of T antigen is proportional to the number of repeats present (Fromm and Berg, 1982; Everett et al., 1983).

Comparisons of gene sequences and the functional analysis of several gene promoters indicate that there is no direct role for DNA sequences more than 100-300 bases immediately upstream of mRNA cap sites. No evidence exists linking any of these above elements discussed to cell specific gene expression, they may only represent the binding sites for proteins that function in the mechanics of transcription. Upstream elements which have been implicated in tissue specific or inducible expression (see below) could function to modulate the efficiency of recognition or exposure of these immediate "core promoter" elements to the transcriptional apparatus.

(ii) Transcriptional Enhancers

Study of viral transcription units has identified a class of control elements, termed enhancers, that can activate transcription units independent of their orientation or precise position relative to a transcription start site (Khoury and Gruss, 1983). Enhancers are short stretches of DNA containing a core of about 8-11 bases, usually located 5' to promoters; they can act over a distance of up to several kilobases to increase transcription from promoter sequences. Viral enhancers operate in a variety of cell types, although not always with equal efficiency and can function not only when linked to their natural genes but also in association with heterologous genes. They are often present in short tandem repeats of 50 to 100 bases. Viruses with a single copy of the sequence are viable so the purpose of the duplicated element is unclear.

Several cellular genes have now been shown to possess their own

enhancer elements. Expression of cloned immunoglobulin genes introduced into myeloma cells has shown the presence of an enhancer upstream of the immunoglobulin heavy - chain switch region (Gillies et al., 1983; Banerji et al., 1983; Queen and Baltimore, 1983). During the generation of antibody diversity the immunoglobulin molecule is assembled by the translocation of one of several hundred variable - region (V) gene segments to the site of the constant - region (C) gene segment. Each V - region segment has its own promoter but is transcriptionally inert until linked to the constant region which has its low constitutive activity increased by the DNA rearrangement.

The sea urchin H2A gene modulator region located between -111 and -165 upstream of the initiation site can act in both orientations and has extended homology to several viral enhancers (Grosschedl et al., 1983). Upstream regulatory elements of the insulin and chymotrypsin genes may increase transcription in a cell type specific manner by a mechanism related to those used by the viral enhancer elements (Walker et al., 1983).

The human adult β-globin gene may contain a controlling element downstream of its transcription start site (Charnay et al., 1984; Wright et al., 1984). In experiments to define the DNA sequences responsible for the differential expression of human α- and β-globin genes a hybrid of the human α-globin upstream promoter region and the β-globin structural gene was produced. When introduced into a mouse erythroleukemia cell line the chimeric gene construct is inducible in the same manner as the complete β-globin gene upon terminal differentiation of the cell line. Similar results were obtained with fusions between the promoter of either the non-inducible human fetal globin gene or a mouse histocompatibility antigen gene and the human adult β-globin structural gene. It is suggested that the β-globin gene, which has a full complement of upstream promoter elements, also harbours an element within the structural gene that is crucial for

appropriate expression in erythroid cells.

How enhancers work is unknown but several possibilities have been suggested. They may act as chromatin organizers, bidirectional entry sites for RNA polymerase, sites of attachment to the nuclear matrix or regulators of local DNA supercoiling. The potential of enhancer regions of several DNA viruses and RNA retroviruses to form Z-DNA suggests that sequence-specific Z-DNA binding proteins may be involved in transcriptional enhancement (Nordheim and Rich, 1983).

(iii) Tissue-Specific and Inducible Gene Expression

The hypothesis that trans-activating molecules interact with gene specific cis recognition sequences of tissue specific or co-ordinately induced genes is the currently accepted model of gene activation. Thus a regulatory network of trans-activators could activate or repress genes in a cascade mechanism during the development and differentiation pathways of metazoans (Davidson et al., 1983). Regulatory signals of several higher eukaryotic genes have now been analyzed by in vitro mutagenesis and gene transfer. As discussed above eukaryotic promoters can extend over a large region and be composed of different functional elements. The enhancer elements are strong candidates for the cis acting sequences postulated for selective gene expression.

Sequence analysis has revealed no extensive homology between various heterologous enhancers. It is noted that the functional activities of different enhancers vary depending upon the host or the tissue into which they are introduced. This host/tissue dependence of enhancer activity has been examined in a competitive assay between several enhancer-containing molecules for cellular components that may interact with them (Scholer and Gruss, 1984). Indeed, a host cell preference of different enhancers was

reflected in these competitive experiments and an interaction between enhancer-containing DNA fragments and cellular components suggested.

Perhaps the best evidence for the gene activation model so far is the properties of the glucocorticoid responsive region in mouse mammary tumour virus (MMTV). Steroid hormones modulate the transcription from the MMTV long terminal repeats by an interaction between the steroid-receptor complex and a specific DNA sequence resembling an enhancer element (Parker, 1983). The glucocorticoid receptor has also been shown to interact with the human metallothionein-IIa gene and the region of interaction is homologous to the MMTV glucocorticoid responsive region (Karin et al., 1984). Therefore it has been shown that two structurally unrelated genes can be induced by an interaction with the same trans-acting molecule. Another example is the co-ordinate induction of the Drosophila heat-shock genes by the heat shock activator protein binding to the heat-shock box (Wu, 1984b).

Another facet of this proposed regulatory network is that one gene can be expressed from different promoters in different tissues if more than one cis acting element is present near the gene. The synthesis of different mRNAs from a single gene by the use of alternative promoters has been found in eukaryotes. The α-amylase gene of the mouse, the invertase gene of yeast, the Discoidin-1a gene of Dictyostelium and the Alcohol Dehydrogenase gene of Drosophila are examples of such transcriptional initiation complexity (Benyajati et al., 1983).

1-4-3 Transcription of Polymerase I Genes

RNA polymerase I acts exclusively on a single kind of transcription unit, the ribosomal RNA cistron. Comparing rRNA genes from yeast, Drosophila, Xenopus and mouse does not identify any common sequences that could potentially act as promoters (Sommerville, 1984). However homologies between more closely related species can be identified. From these short

range comparative data an important role for sequences located immediately around the initiation site is suggested.

By using rRNA genes with various sequence deletions in in vitro functional assays a wider region of DNA is implicated. The first few transcribed nucleotides and a stretch of T residues that may be analogous to the RNA polymerase II TATA box may act together in the initiation of transcription. It has been found that sequences that lie further upstream of the start site can enhance transcription of mouse rDNA. Xenopus rDNA contains duplicated promoters that extend several thousand base pairs upstream of the start site; the number of these repeated spacer elements is related to the production of rRNA in oocyte injection studies (Reeder et al., 1983). In vitro transcription assays of rDNA templates is dependent upon components supplied from cell extracts. Only if the rDNA and transcription factors are from the same species is correct initiation seen. Thus RNA polymerase I transcription factors and promoters appear to be species specific.

1-4-4 Transcription of Polymerase III Genes

In eukaryotic cells, RNA polymerase III is responsible for the transcription of 5S RNA, tRNA, some small cytoplasmic and nuclear RNAs and some virus associated RNAs. These genes are short compared with the average length of the genes transcribed by polymerase I and II and the region of DNA essential for promotion of transcription is always contained within the coding sequence (Korn, 1982; Lassar et al., 1983). The termination of transcription occurs at the first run of four or more T residues following initiation.

For the 5S RNA genes a 34 base pair segment has been mapped as the internal promoter and initiation of transcription occurs 50 bases upstream

of this internal control region. The promoter of tRNA genes is split into two essential regions of about 10 bases (Box A and B), separated by 30-40 bases, and these are highly conserved in all tRNA genes. Initiation of transcription occurs 11-18 bases upstream from box A. Other polymerase III transcribed genes appear to be very similar to tRNA genes containing two regions of homology to boxes A and B and also initiating a short distance 5' to the A box. The first 11 bases of the 5S RNA gene promoter have been shown to be structurally and functionally related to the A box component of the tRNA gene promoter (Ciliberto et al., 1983)

Multiple cellular factors are required for RNA polymerase III transcription. These form stable transcriptional complexes in association with the polymerase III genes and have been analyzed by chromatographic fractionation of crude cellular extracts. Two factors, IIIB and IIIC are necessary for transcription of the tRNA type genes where as 5S genes require these same factors plus a third gene-specific factor TFIIIA. TFIIIA also interacts with 5S RNA in oocytes apparently to stabilize the stored 5S RNA. This interaction has led to the suggestion that in the cell 5S gene transcription in the presence of limiting amounts of this factor may be subject to autoregulation. Pre-initiation complexes between purified genes and the fractionated factors persist even after many rounds of transcription or when challenged with a competing template. The establishment of stable complexes may represent a key step in the activation of these genes and provide a means for the maintenance and propagation of a specific set of activated polymerase III genes.

1-4-5 Processing and Stability of RNA

The events that convert a primary RNA transcript to a mature RNA molecule are collectively known as RNA processing. Changes in transcriptional initiation, termination and RNA splicing have been shown to

fluctuate in specific situations, in an apparently regulated fashion. Even after the formation of an mRNA is complete, changes in the stability of the mRNA can alter the effective expression of the gene. Thus the complex pathway of RNA biogenesis serves a regulatory role.

(i) Initiation and Termination

Almost immediately after initiation of an RNA chain an m^7Gp residue is added to the first base of the RNA. The central phosphate of the m^7GpppN cap is derived from the pppN that starts the RNA chain, which is evidence for the RNA transcript initiating at the cap site (Darnell, 1982). Precision of nascent RNA synthesis termination is not yet understood for polymerase II transcription. In every case where it has been examined, transcription does not terminate at the poly A addition site of the mRNA but at some distance downstream. Although termination can be mapped within a few hundred bases, a clearly defined site has yet to be shown.

(ii) 3' End Processing and Poly A Addition

The one clear example of a class of nonpolyadenylated mRNAs, the histone mRNAs, initially represented potential 3' end generation by termination (Birchmeier et al., 1982). However even in this instance it has been shown (using one of the chicken H2B genes described in this thesis), that the 3' terminus is generated by processing (Krieg and Melton, 1984). A sequence with dyad symmetry present at the 3' end of histone genes together with some immediate downstream elements serves as a recognition site for a small nuclear ribonucleoprotein (snRNP) to process at this point (Galli et al., 1983).

The Poly A containing mRNAs possess an AAUAAA or related sequence about 10-30 bases upstream from the poly A tail. The role of the AAUAAA hexanucleotide has been shown to be that of a recognition signal for the

processing of the primary transcript (Fitzgerald and Shenk, 1981). An endonucleolytic cleavage is thought to occur at the site of polyadenylation to generate a substrate for a poly A polymerase to polymerase A residues onto the 3' hydroxyl group (Montell et al., 1983). An intact AAUAAA sequence is required for efficient RNA cleavage, but not for polyadenylation. It appears that the AAUAAA sequence only forms part of the RNA cleavage signal as internal AAUAAA sequences are observed distant from 3' ends of mRNA. Other primary or secondary structures may be additional features of the RNA cleavage signal.

(iii) Splicing

The primary transcripts of many eukaryotic genes are interrupted by stretches of non-coding DNA called introns. These sequences are subsequently deleted by a cleavage-ligation process termed RNA splicing to generate the mature transcript of colinear exons. It appears there are three general classes of RNA splicing, one for tRNA, a second for mRNA and a third that includes rRNA and mitochondrial mRNA and rRNA (Cech, 1983). Nuclear mRNA splice sites are designated by short sequences at the intron-exon junctions, tRNA splicing is directly related to the structure of the exons and mitochondrial splicing is directed mainly by the structure of the intron. All three classes of splicing systems have different chemical mechanisms for the cleavage and ligation of the phosphodiester bonds.

In nuclear mRNA precursors, the splice sites are always marked by sequences resembling 5'(A/C)AG/GTAGT...intron...(T/C)nN(T/C)AG/G 3' (Mount, 1982). The first two 5' bases (GT) and the last two 3' bases (AG) of an intron (underlined) are strictly conserved. Mutations that disrupt one of these sequences interfere with splicing whereas mutations that make a nonjunction sequence a better match to the consensus can produce a new splice site. This indicates that the consensus sequence is the primary

determinant of the splice junctions. The small nuclear ribonucleoprotein U1 RNP has been shown to be involved in the selection of splice sites (Padgett et al., 1983). The introns are excised as lariat- like structures rather than as linear or circular RNA. The lariats are formed by attachment of the 5' end of the intron to a site close to the 3' end (Weissmann, 1984). The m^7Gp cap is required for efficient *in vitro* splicing and it is thought that recognition of the cap structure may be an important step in the formation of a specific ribonucleoprotein complex required for splicing (Konarska et al., 1984).

(iv) Differential Processing

There are several cases known in which various poly A addition sites or splice sites can be chosen on the same primary transcript. Such differential processing can depend on the physiological condition of the cell, on the cell type or the type of transcription unit. A number of virus transcription units, such as in adenovirus, SV40, polyoma and retroviral RNAs, contain multiple splicing arrangements of their RNA precursors. These viral genomes on integration into cellular DNA of a variety of cell types can still produce multiple mRNAs. Thus, it is clearly possible for DNA in the cell genome to encode primary transcripts that can be spliced in alternative ways. This suggests some sort of scanning model for the identification of splice sites of mRNA. If splice site selection were to be dissimilar in different tissues, different gene products could be obtained from the one gene. Tissue-specific RNA processing has been shown in the production of the calcitonin protein and a calcitonin gene related peptide from the unique calcitonin gene (Rosenfeld et al., 1983).

A differential choice of poly A addition sites has been found in the late adenovirus transcript and the immunoglobulin heavy chain

transcriptional unit. During late adenovirus infection a transcript encompassing five poly A sites occurs. The selection of a poly A site initially occurs before the polymerase has completed the transcription of distal sequences i.e. a poly A site has already been chosen and a poly A tail added before the downstream poly A sites have become available. As the infection proceeds, site selection does not follow a random pattern. Although one would expect a polar effect where the first site was selected most frequently, such is not the case. A relative efficiency of 1:2:3:2:2 progressing down the transcription unit is seen (Nevins and Wilson, 1981). These results indicate that poly A site utilization is controlled to alter gene expression.

During B-lymphocyte development and immunoglobulin synthesis, the heavy chain is first inserted as an integral protein in the plasma membrane. Later, a similar chain is found secreted as part of the immunoglobulin molecule. The membrane-bound form of the protein is generated by the transcription of all the coding sequences into nuclear RNA, which has a hydrophobic carboxyl protein terminus coded for at the end of this long transcript. Two poly A sites occur in this transcription unit, the long primary transcript covering both of them. Depending on the choice of poly A site, two heavy chain mRNAs with different 3' sequences (due to the presence of alternative exons) can be formed, one for the membrane protein and one for the secreted chain (Early et al., 1980).

1-5 Cytoplasmic Control of Gene Expression

1-5-1 Stability of mRNA

The final concentration of functional mRNA available for translation into protein depends not only on the rate of gene transcription, but also on the relative stability of the mRNA once in the cytoplasm. Immediate shutdown of expression of a given gene after cessation of transcription can

only be accomplished if the mRNA is also rapidly degraded.

Changes in the half-life of specific mRNAs in response to the presence of hormones have been documented. Estrogen selectively stabilizes Xenopus liver vitellogenin mRNA against cytoplasmic degradation. The half-life of vitellogenin mRNA is 3 weeks in the presence of estrogen but only 16 hours after estrogen is withdrawn (Brock and Shapiro, 1983). Upon exposure of breast tissue to prolactin, the stability of the casein mRNA increases 25 fold (Guyette et al., 1979). It has been proposed that there are cytoplasmic hormone binding proteins that act to stabilize particular mRNAs by a direct and reversible protein-nucleic acid interaction.

Histone gene expression is closely co-ordinated with cellular DNA replication. In addition to the temporal coupling of histone transcription, inhibition of DNA replication results in a rapid destabilization of histone mRNA (Nurse, 1983).

Several studies have suggested that the 3' poly A tail plays a determining role in mRNA stability (Nevins, 1983). However there is a body of evidence which implies that the poly A does not regulate mRNA stability in certain cases but that its primary function is in protein synthesis (Jacobson and Favreau, 1983; Palatnik et al., 1984). It has been postulated that the two phenomena are indirectly related in that mRNA stability is in part a consequence of translational efficiency. The m^7Gp cap structure has been shown to be critical in determining the nuclear and cytoplasmic stabilities of mRNA in the Xenopus oocyte (Green et al., 1983a). Thus blocking of 5' and 3' ends of an mRNA transcript may significantly reduce degradation by RNA exonucleases.

1-5-2 Translation

Translation of eukaryotic mRNAs typically occurs at the first AUG

triplet from the 5' end of the message. This observation led to the hypothesis, termed the "scanning" mechanism of initiation, whereby the 40S subunit of eukaryotic ribosomes binds to the 5' cap structure of a message and subsequently migrates along the RNA chain searching for and stopping at the first AUG codon. Subsequently a 60S subunit joins this stalled complex and peptide bond formation begins (Kozak, 1980). Although it seems clear that the position of an AUG triplet relative to the 5' end of the mRNA has a major role in identifying it as a functional initiator codon, initiation of translation at internal AUG codons is possible (Liu et al, 1984).

The distribution of the bases flanking functional initiator codons has been examined and the sequence CC(A/G)CC AUG G has been tentatively proposed as a consensus sequence for eukaryotic initiation sites. To assess if the translational machinery is sensitive to the sequences around AUG initiator codons, single nucleotide changes have been introduced near the translational start site of a cloned pre-proinsulin gene (Kozak, 1984). Sequence changes around the initiator codon drastically affect the yield of proinsulin protein, with the presence of an A residue 3 bases upstream from the AUG being required for the most efficient utilization of the initiator codon.

Differential translation of specific mRNA molecules present in the cell cytoplasm is recognized as a level of control of gene expression. A characteristic of growing oocytes of all animal species is the storage of maternal mRNAs which are destined to be translated in the early embryo. The mechanism which regulates the translation of this mRNA is unknown but oocyte-specific proteins have been implicated in the reversible repression of embryonic stored maternal mRNAs (Richter and Smith, 1984). Major shifts in translation of mRNA have also been analyzed in non-embryonic cells. During heat trauma of most eukaryotic cells there is a decrease in total

protein synthesis and enhanced production of certain 'heat-shock' proteins (hsp's). This increase in the level of hsp's has both transcriptional and translational components, but selective translation of specific mRNAs plays a major part in the heat-shock response (Bienz and Gurdon, 1982; Ballinger and Pardue, 1983).

1-5-3 Post-Translation

Although all eukaryotic mRNAs are monocistronic, in that only one species of polypeptide can be translated per messenger molecule, different individual functional proteins can be encoded by the one mRNA species. In this case the mRNA is translated into a single, large "multifunctional" protein. This polyprotein is then cleaved by specific proteases to yield distinct proteins. Several hormonal and neural peptides are initially synthesized as larger precursors, which are then converted to individual peptides by post-translational cleavage.

The precursor to adrenocorticotropic hormone (ACTH) is also the precursor to β-endorphin and melanocyte-stimulating hormones (MSHS) and is called pro-opiomelanocortin (POMC). The precursor to the enkephalins contains six copies of Met-enkephalin and one copy of Leu-enkephalin (Herbert and Uhler, 1982). The major domains of the POMC molecule are flanked each side by pairs of basic amino acid residues and it is thought that a trypsin-like activity separates the domains from one another. POMC is present in both the anterior and the intermediate lobes of the pituitary and in a number of sites in the brain. POMC is processed to different hormones in these tissues and the synthesis and secretion of ACTH/endorphin peptides is regulated differently in these sites. Processing studies have indicated that differential tissue expression of POMC products can be affected by different post-translational processing enzymes in each tissue.

The biological significance of multi-peptide precursors may be to effect the simultaneous production of multiple, active peptides that perform co-ordinate functions.

1-6 Histone Genes

The first eukaryotic polymerase II transcribed genes to be cloned and studied were the histone genes of sea urchins and Drosophila. Studies were facilitated by the high histone gene copy number in these species and the extraordinary DNA sequence conservation of histone subtypes between species which has allowed subsequent isolation and examination of the histone multigene families from a wide range of organisms. Consequently, histone genes are one of the best characterized eukaryotic gene systems so far researched. Recent advances in the knowledge of histone gene organizations and their expression has been the subject of several extensive reviews (Hentschel and Birnstiel, 1981; Maxson et al, 1983a and b; Stein et al., 1984; Old and Woodland, 1984). Thus only a cursory review of histone gene systems of a variety of species will be given and only those details that are directly relevent to this thesis will be presented.

1-6-1 Tandemly Repeated Systems

The major class of histone genes of invertebrates occur in loci containing tens to hundreds of nearly identical repeating units. Such tandemly repeated clusters are found in Drosophila and several species of sea urchin. The newt and possibly Xenopus may also contain similar repeats (see Maxson et al., 1983). Several common features can be seen in this type of organization, viz., the repeating units contain the genes for all five histones; these genes all encode the "early" histone proteins (embryonically expressed); repeated units of such clusters are extremely homologous within species, but there may be distinct major and minor repeats which differ considerably. However, some organizational features of

tandem repeats show significant variation, such as gene orders, polarity of transcription and the lengths of the repeats.

1-6-2 Disorganized Clusters

The histone genes of the higher eukaryotes (mammals and birds) and the lower eukaryotes (yeast and protista) are present in much lower copy numbers than those of the "tandemly repeating" organisms. There is no immediately apparent gene order or arrangement in these organisms, except that the histone genes are loosely clustered into groups of several genes separated by large spacer DNA regions. The most detailed picture of vertebrate histone gene organisation has come from recent studies in the chicken genome (D'Andrea et al., 1985; see chapter two). Genes for the "late histone" subtypes of the sea urchin have recently been isolated and shown to have a similar disordered structure. However, upon closer examination of the gene organization tables in Maxson et al. (1983a) trends do become apparent, in that there is a close association of H2A genes with H2B genes and to a lesser extent an association of H3 and H4 genes. Work presented in this thesis goes on to define the extent of the H2A/H2B association in the chicken genome.

1-6-3 Gene Sequences

Gross examination of the histone gene architecture from several genera has not suggested any regulatory mechanism controlling these genes. More detailed structure-function relationships have come from DNA sequence analyses.

(i) mRNA Untranslated Region and Cap Site

Characteristically histone leader regions are short and pyrimidine rich but there is little conservation of sequences in the leader regions of any of the histone mRNA subtypes in general. There are, albiet short,

conserved regions between the mRNA leaders of adjacent chicken histone H1 and H2B genes (Coles, 1985). The 5' termini of most histone mRNAs are located within a conserved PyCATTCPu sequence.

(ii) Coding Regions and Spacers

Initial attempts to identify conserved sequences in the cloned major histone repeat of P. miliaris (clone h22) revealed that the coding regions were G-C rich and intronless. The spacers were A-T rich, could not code for proteins and contained stretches of alternating copolymeric DNA. Simple alternating copolymers are also present in the spacers of Drosophila but in a somewhat shorter form. Histone genes from a wide variety of organisms studied subsequently have also been shown to lack introns. Exceptions to this trend have been found in two chicken genes, one an H3 subtype variant gene and the other an H2A protein variant gene called H2AF (A. Robins personal communication).

(iii) Promoter Elements

The sequences 5' to histone genes are thought to be more important in the control of histone gene transcription than any gross organization of this multigene family (this will be further discussed in the main text). As outlined in 1-4-2 there are several independent promoter signals common to most polymerase II transcribed genes and these are present in the histone genes, though not present in all genes of the family. There also exist several histone-specific/histone-subtype specific sequences.

The TATA element (with variations) is present in all histone genes examined. In the sea urchin, the sequence GATCC is present 11 bases upstream of the TATA sequence of most genes. Work in this laboratory has shown that this sequence is not found close to the TATA box of any chicken histone gene except H4. Moreover this element is present only in H4 genes of several other higher eukaryotes. It is concluded that, at least for

these cases, this element is in fact H4 gene-specific (Wang et al., 1985).

Other subtype specific motifs have been described for H1 and H2B genes of several species. The most significant H1 motif is an AAACACA, found approximately 100 bases upstream from the cap site (Coles and Wells, 1985). A 13 base H2B-specific sequence (Harvey et al., 1982) is located 6-30 base pairs upstream from the TATA box and will be a major point of discussion in chapter 6 of this thesis. The sea urchin H2A gene promoter region has been extensively mapped and deletion of far-upstream sequences can modulate transcription (1-4-2 (ii)). This area contains two well conserved regions (with homology to several viral enhancers) between sea urchin species which are also present in two chicken H2A genes (Wang et al., 1985), although these two elements have been fused into one element in the chicken.

(iv) Conserved 3' Elements

The 3' flanking regions of histone mRNAs are remarkable in that most contain a 23 base canonical

5' AACGGC(C/T)CTTTTCAG(G/A)GCCACCA 3'

sequence which contains a 6 base hyphenated inverted repeat; 3' termini of histone mRNAs localized to date fall within the ACCA region. This sequence is responsible for the endonucleolytic cleavage event that generates the histone 3' end (1-4-5 (ii)). Those histone genes that lack this processing site, such as in yeast, contain an AATAAA sequence and are correspondingly polyadenylated.

1-6-4 Regulation of Histone Expression

Histone proteins associate with DNA as a nucleosomal particle containing an invariant 1:1 molar ratio of the four core histones - H2A, H2B, H3 and H4 - and about half equimolar amounts of the H1 histone subtype. Histone gene families therefore are regulated by controls that

must ensure proper stoichiometric relationships of histone to histone and total histone to DNA. Clearly these controls must ensure that dividing cells synthesize sufficient histones to organize newly replicated DNA into functional chromatin. The demand for histones may be periodic and light or sustained and heavy depending on the doubling time of the cell. For example, cultured cells exhibit a relatively long cell cycle whereas cleaving embryos of some species divide extremely rapidly. Different strategies for histone regulation are obviously called for in these two extreme cases.

A considerable body of evidence exists to indicate that histone production and DNA synthesis are temporally co-ordinated during the cell cycle. For most studies of expression in the cell cycle cultured mammalian cells have been employed. The tightness of the coupling between DNA synthesis and histone expression has been found to depend upon the cell type. Regulation of replication-dependent histone synthesis operates at two levels, that of transcription of the genes and of destabilization of the histone transcripts at the end of S-phase. Knowledge of the mechanisms of regulation at both levels remains at the hypothetical stage but is currently an intensive area of research.

The coupling of histone and DNA synthesis of eukaryotes is not found throughout their life cycles. A well documented example of uncoupled histone synthesis occurs during embryogenesis. Upon fertilization, the eggs of many organisms undergo a period of rapid cell division in the course of which distinct cell lineages are established. Large quantities of histones are used in chromatin formation and the developing zygote can use pre-existing stores of histone protein and newly synthesised histone from mRNA accumulated during oogenesis. These stores can be supplemented by de novo transcription and translation during cleavage. The presence of many copies

of histone genes in several organisms may permit unusually rapid synthesis of large quantities of histone mRNA during cleavage compared to that seen during adult expression.

The unfertilized sea urchin egg contains a significant store of histone mRNA and protein. During the development of the sea urchin three major classes of histone proteins appear. A succession of switches occurs between these histone gene classes; from the first cell division cleavage stage (cs) variants are found in chromatin and their synthesis continues until the 8-cell stage whereupon the α variants are synthesized and used throughout early cleavage after which they are largely replaced by the γ class. The maternal stockpile of histone contains both α and cs variants. Histone genes coding for early embryonic protein forms (α) are the ones that are organized in tandem repeats. There is no evidence of any developmental classes of histone genes in Drosophila or Xenopus which have the tandemly repeated component of sea urchin organization, so major histone class switching may be an echinoderm speciality.

Thus histone gene expression is regulated temporally, qualitatively and quantitatively. Such regulation must also include mechanisms to ensure that the appropriate stoichiometry of all five histone subtypes is maintained in newly synthesized chromatin. One way of achieving this would be through the assembly of nucleosomes from appropriately sized histone protein pools. Alternatively, nucleosome histone stoichiometry may result from specific regulatory controls exerted during chromatin assembly. It has been observed that the sea urchin histone mRNAs are synthesized in amounts that reflect the molar ratio of histones in chromatin. Conversely gene dosage experiments with H2A/H2B genes in yeast apparently shows no deleterious effects when the cell only contains one of its normal two

copies of these genes. Synthesis of stoichiometric quantities of the various histone mRNAs does not alone explain how chromatin assembly is achieved during embryogenesis; such large maternal stores of histones occur that further mechanisms must exist to regulate their accumulation and use. How stoichiometry is achieved in these two completely different biological circumstances (i.e. the stoichiometry of the large embryonic store of histone produced when histone synthesis is uncoupled to DNA replication compared to that achieved when there is tight coupling of histone synthesis and DNA replication that is characteristic of non-embryonic cells) will be further investigated in the thesis discussion.

1-7 Functional Testing of Eukaryotic Genes

Gene structure and sequence analyses alone do not enable the mechanisms involved in the control of gene expression in higher organisms to be elucidated. For this to be achieved, gene function must be studied.

Classical genetic studies of higher organisms have been relatively restricted for several reasons viz., the mammalian genome is approximately 1000-fold more complex than that of bacteria; most animals and plants have long generation times so experiments take a long time to complete; they are usually diploid, which hinders the isolation of recessive mutations. To solve these problems a procedure of "reversed" genetics has been developed. In this approach, alterations are first introduced into specific regions of DNA believed to be involved in some aspect of gene regulation or expression. Secondly, the "phenotype" conferred by these sequence manipulations is tested by assaying for gene function. A primary requirement in this assay is that the unmutated gene is expressed in the system, in vivo or in vitro, in as physiological a manner as is possible.

Several experimental systems for the study of eukaryotic gene expression have been developed. Each has its own advantages or special uses

and all the systems can be grouped into three main categories. In vitro assays are obviously limited by their nature and to the specific purpose for which they were developed. Whole cell analysis can examine each level of a gene's control mechanism. The production of transgenic organisms can investigate the complete spectrum of a gene's physiological significance.

1-7-1 In Vitro Transcription

Transcriptional systems that operate in vitro are available for the anaysis of templates for all three types of RNA polymerase genes (Wickens and Laskey, 1981). The advantage of cell-free systems lies in the fractionation and identification of active components. When a soluble cell -free system performs a function efficiently it becomes possible to fractionate away that activity and to characterize the factors that restore it. When purified eukaryotic RNA polymerases are incubated with purified DNA templates, transcription begins at non-specific sites. When crude cellular extracts are added it is possible to identify sequences and fractions that direct accurate initiation.

1-7-2 Transfection of DNA into Cultured Cells

Genes can be introduced into eukaryotic cells by the use of various chemicals or electrical gradients which facilitate DNA uptake. DEAE-dextran, calcium phosphate precipitation, glycerol/PEG shock, protoplast or liposome fusion and electrointroduction have all been used to mediate the trasfection of DNA into tissue culture cells. The particular technique to be utilized depends on the cell type being used for analysis, the efficiency of uptake or survival rate required and the selection procedure being used. How the DNA actually enters the cell nucleus using these techniques is not known.

There are two main experimental approaches used with tissue culture

cells, either cellular "transformation" (stable inheritance of genes) can be bestowed upon the transfected cell, or transient expression of the introduced gene can be obtained.

If cellular "transformation" is aimed for, the efficiency of transformation and thus transfection is not necessarily important; a selectable marker is usually co-introduced; and a clonal or homogeneous population of cells can be obtained, but the time period required to do this can be long (weeks or months). Transient gene expression does not require the production of a permanent cell line and the time course of the experiment is limited only by the survival time of the cells or the DNA introduced into them (usually days). However a high transfection efficiency is normally required, and a sensitive assay (such as the production of a marker protein) is commonly used to analyze the expression from a gene promoter.

The fate of the DNA once it has entered the cell nucleus depends upon the form in which it was introduced. Circular or linear molecules are usually used for transformation experiments, and will be integrated into the cells' chromosomes in variable numbers and in a random manner unless the gene is specifically transfected on a vector that resists integration; that is one that naturally exists as an episomal DNA molecule. An example of this type of vector is Bovine Papilloma Virus (Lusky and Botchan, 1984). Retroviral vectors have been designed for transformation experiments because the copy number and fidelity of integration is controlled by the virus (Williams et al., 1984).

Transient assays usually do not allow time for the DNA to be integrated or if it does the integrated copies do not contribute significantly to the detected expression. It is a great advantage to the detection of expression if, once an introduced molecule has been delivered

to the cell nucleus, it can be replicated or propagated to provide a number of templates for transcription. A series of SV40-derived vectors have been developed for this purpose (Spandidos and Wilkie, 1984) making use of the SV40 origin of replication.

The applications of this technology make almost every cellular operation amenable to study. There is no better example of the power of this new genetics than the transfection of yeast cells. In this lower eukaryote it is possible to replace precisely specific genes that have been cloned, with *in vitro* mutagenized copies and even to introduce new minichromosomes that exist at single copy number and segregate properly on replication (Struhl, 1983).

1-7-3 Microinjection of DNA into Cells and the Construction of Transgenic Organisms

The only practical method for introducing genetic material into the nuclei of some somatic cells, oocytes and fertilized eggs of Xenopus and the mouse, and into Drosophila embryos is by physical means. Direct microinjection procedures have been developed for all these systems.

DNA can be introduced into single tissue-culture cells without the help of chemical treatment (Capecchi, 1980). A high proportion of stable transformants can be obtained using these methods, possibly because the DNA is directly applied to the nucleus whereas other techniques only allow initial DNA entry into the cell cytoplasm. The advantages of this route of entry are that the efficiency of transfection is so high (approaching unity) that a selection procedure need not be used, the fidelity of DNA integration is good and the copy number of insertion can be controlled by controlling the number of DNA molecules injected per cell.

The expression of cloned genes in Xenopus oocytes and eggs is a common and rapid method for analysis (Gurdon and Melton, 1981; Wickens and Laskey,

1981; Gurdon and Wickens, 1983; see methods section). Its advantages are that the injections are technically easy, nearly all the steps in gene expression take place in injected oocytes, and because the amphibian ovum is such an enormous cell very few cells have to be injected to obtain enough material for analysis. The vast molecular knowledge available on the development of the frog egg and the control of injected DNA during embryogenesis is one of its special advantages (Newport and Kirschner, 1982).

Microinjection of plasmid DNA into fertlized mouse ova has been shown to be a convenient and also rapid system for studying transcriptional regulation (Brinster et al., 1982). If the male pronucleus of a fertilized mouse egg is injected it is possible to get retention of the gene in every cell of the developing mouse (Palmiter et al., 1982; Lacey et al., 1983). Such transgenic animals often transmit the gene through the germ line to the next generation. This technique promises to be invaluable in the investigation of tissue specific gene expression.

Drosophila embryos can be made transgenic when injected with genes contained in P-element vectors. P-elements are transposons which are highly mobile in the Drosophila germ line when in non P-strain genetic backgrounds, and are responsible for the phenomenon of hybrid dysgenesis. The injected DNA specifically integrates into the germ line cells of embryos, thus the phenotype conveyed by the injected gene does not show up until the progeny of the injected fly is produced. Using this system three tissue specific genes have been shown to be controlled correctly in the Drosophila chromosomes (Flavell, 1983). The added attractions in Drosophila are the short generation time (about 2 weeks) and the availability of mutants.

NOTE TO READER:

The number and the names of histone gene containing recombinant clones discussed in the text can become overwhelming. Most of the work can be followed in a logical progression. Whenever a group of clones is discussed (such as the six divergent H2A/H2B gene clones) they are always presented in the same order, which is usually the order of their isolation. The position of a histone clone within the chicken histone gene loci can always be traced by reference to figures 2-1, 2-2 and 2-3.

CHAPTER 2: ORGANIZATION OF THE CHICKEN HISTONE MULTIGENE FAMILY - DIVERGENTLY ORIENTATED H2A/H2B GENES

2-1 Introduction

The entire complement of core and H1 histone genes have been isolated from the chicken genome (D'Andrea, 1985; Coles, 1985; D'Andrea et al., 1985). Fourty-two genes are located within four as yet unlinked loci. The total region of the chicken genome covered is 175 kilobases, encompassed in three genomal cosmid clones and ten λ clones (figures 1, 2, and 3). The number of each type of histone gene is approximately equal i.e. H1, 6; H2A, 10; H2B, 8; H3, 10; H4, 8.

The overall pattern of histone gene organisation from gross restriction mapping and southern blotting is one of disorder, with no long range repeat being evident. Dispersed clusters is the most apt description of the distribution of the genes. There are however preferred associations, H1 with H2A and H2B genes, and a coupling of H3 and H4 genes. The extent of the H2A/H2B associations will be examined in this chapter.

2-2 Orientation of H2A/H2B Genes

The starting points for Richard D'Andrea and Leeanne Coles chromosome walk through the chicken histone gene loci were λ CH-01 and λCH-02 (Harvey et al., 1981; figures 2-1 and 2-2). There is therefore a temporal relationship between their isolation of clones, and my orientation of the H2A/H2B gene pairs contained within them. The initial stimulus for looking specifically at the coupling of H2A/H2B genes was the orientation of such a pair in the λCH-01 clone and my subsequent realization that the promoter region of each gene was extremely close together. Moreover it was obvious that the two genes could share common promoter elements.

(1) λCH-01

A 3.3 kilobase-pair (kb) fragment generated by an EcoRI digestion of

Figure 2-1: Structure of recombinant clones surrounding λCH-01

The overall organization of the region surrounding λCH-01 is shown. The position and direction of histone genes are shown in the top of the diagram by the arrows and the clones derived from this region are indicated. Detailed restriction maps of subclones derived from the genomal clones are also shown with the precise location of genes illustrated (figure taken from D'Andrea et al., 1985).

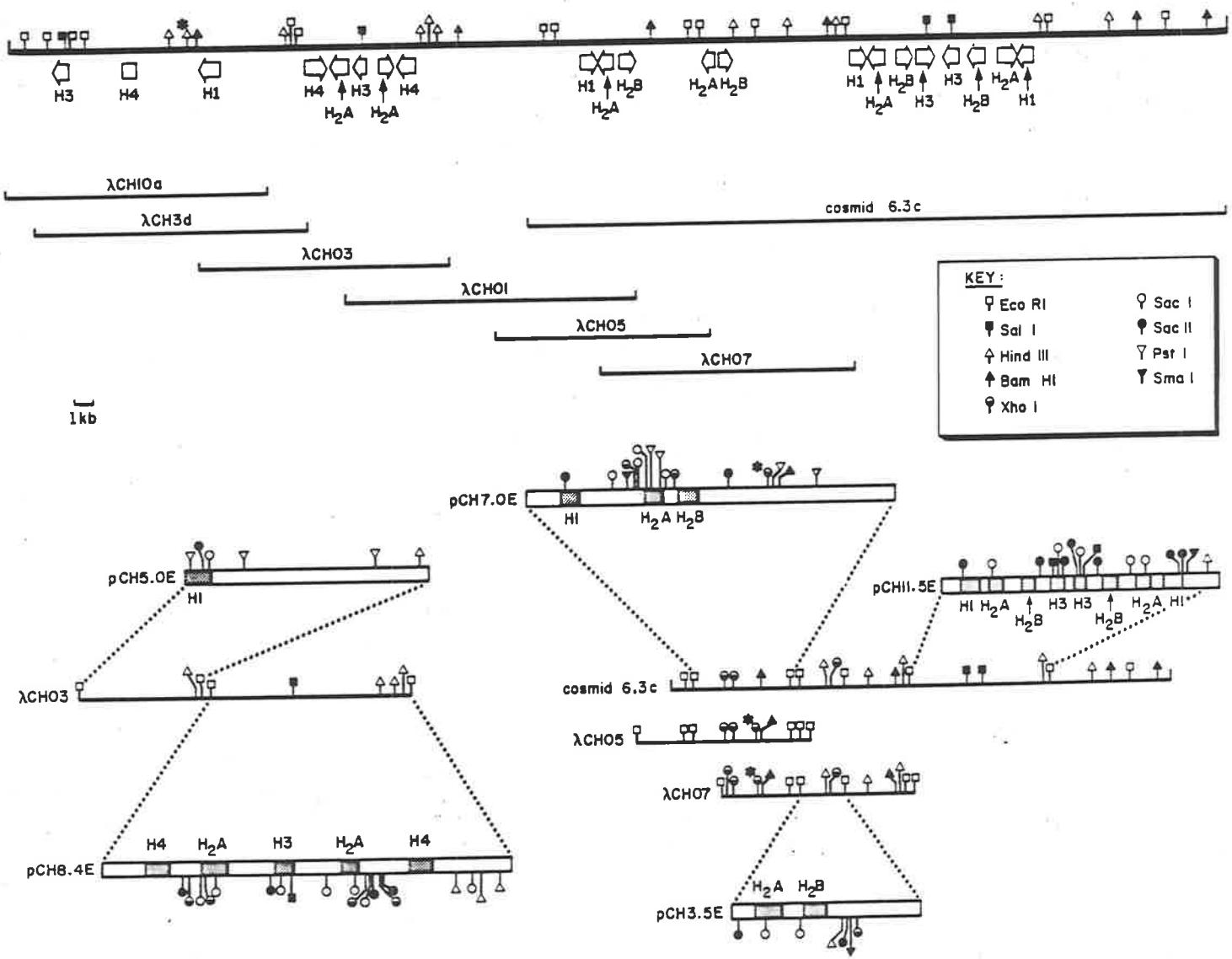

H3
H4
H1
H4
H3
H4
H2A
H2A
H1
H2B
H2A
H2A
H2B
H1
H2B
H3
H2A
H1
H2A
H3
H2B
H3
H2A
H1
λCH10a
λCH3d
λCH03
λCH01
λCH05
λCH07
cosmid 6.3c
1kb
KEY :
Eco RI
Sal I
Hind III
Bam HI
Xho I
Sac I
Sac II
Pst I
Sma I
pCH7.0E
HI
H2A H2B
pCH5.0E
HI
λCH03
pCH8.4E
H4
H2A
H3
H2A
H4
pCH11.5E
HI H2A
H2B
H3 H3
H2B
H2A HI
cosmid 6.3c
λCH05
λCH07
pCH3.5E
H2A H2B

Figure 2-2: Structure of recombinant clones surrounding λCH-02

The genes present in this region are shown above the overall restriction map. Subclones pCH22.0B and pCH2.7B were constructed from cosmid 6.1C and detailed restriction maps of these are shown indicating precise gene locations (figure taken from D'Andrea et al., 1985).

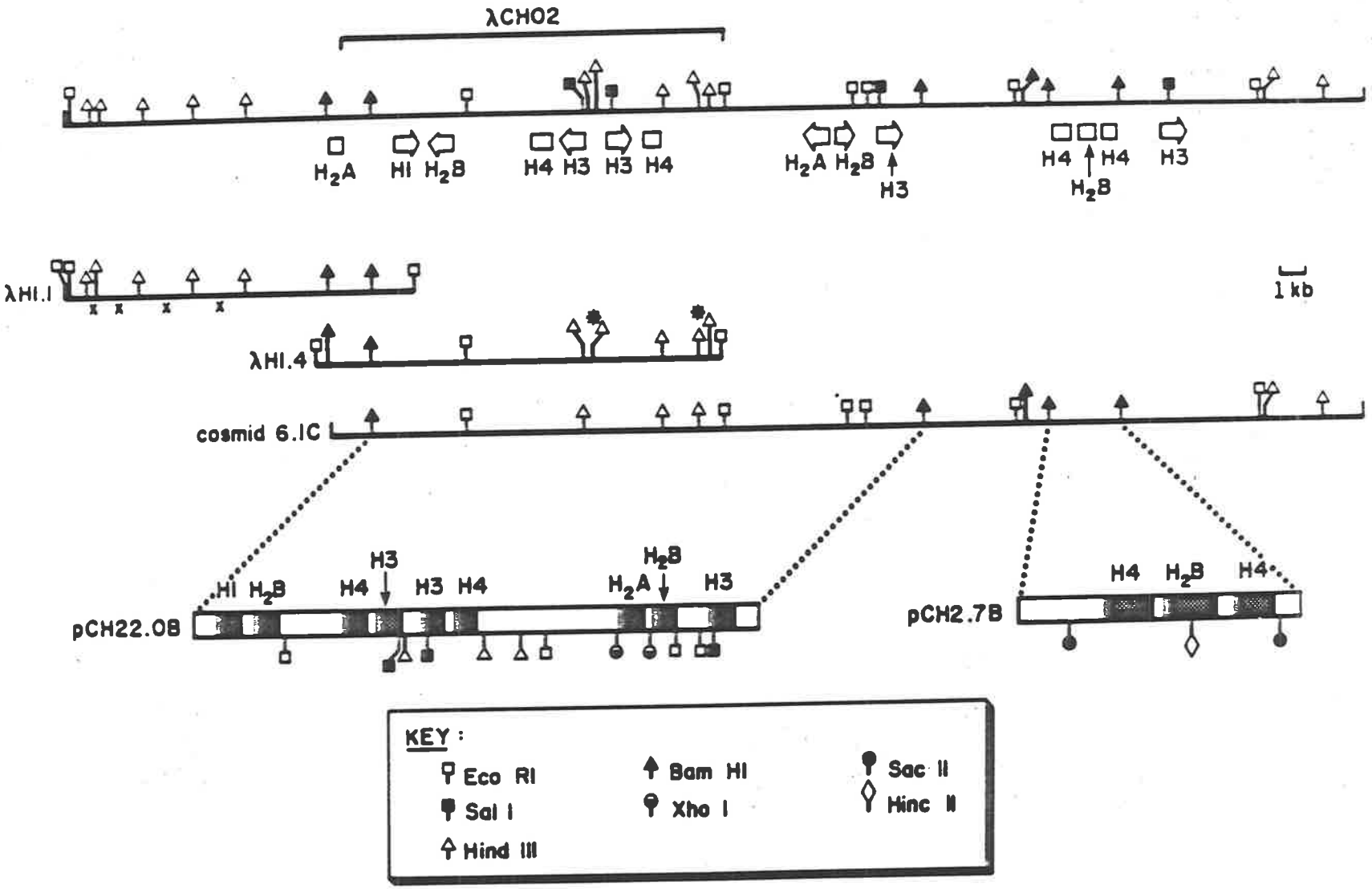

λCH02
H2A
H1 H2B
H4 H3 H3 H4
H2A H2B
H3
H4 H4
H2B
H3
λH1.1
1 kb
λH1.4
cosmid 6.1C
H1 H2B
H3
H4
H3 H4
H2B
H2A
H3
pCH22.0B
H4 H2B H4
pCH2.7B
KEY :
Eco RI
Sal I
Hind III
Bam HI
Xho I
Sac II
Hinc II

Figure 2-3: Structure of genomal clone λ CH1-10

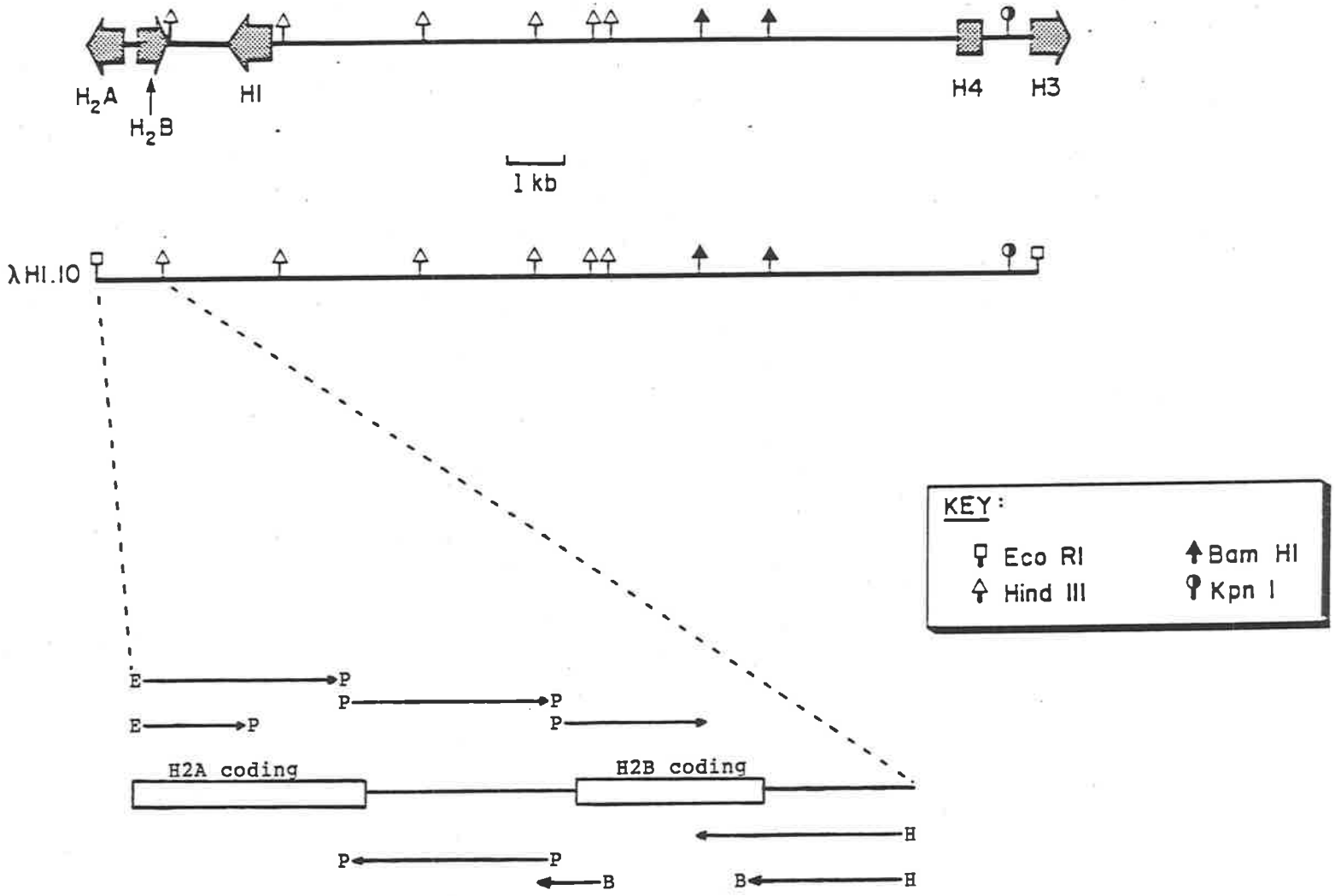
An overall restriction map of this histone gene containing region is presented. Terminal EcoRI cleavage sites are derived from synthetic DNA linkers used during the construction of the clone. The sequencing strategy of the H2A/H2B pair used to orientate the two genes is shown below the restriction map (data taken in part from D'Andrea et al., 1985). The total sequence of the EcoRI-HindIII fragment is presented in figure 3-6.

E = EcoRI

P = PstI

B = BstEII

H = HindIII


H2A
H2B
H1
H4
H3
1 kb
λH1.10
KEY:
Eco RI
Bam HI
Hind III
Kpn I
E
P
P
P
E
P
P
H2A coding
H2B coding
H
P
P
B
B
H

λCH-01 contains the entire H1, H2A and H2B coding regions. This 3.3 kb fragment has been inserted into the EcoRI site of pBR325 and is known as pCH3.3E (this fragment will later be described in a 7 kb genomal EcoRI fragment from λCH-05, known as p7AT). A restriction map of it is shown in figure 2-4. The complete sequence of the 3.3 kb insert has been determined initially using Maxam and Gilbert technology, then M13 dideoxy sequencing (D'Andrea et al., 1981; Harvey et al., 1982; Coles and Wells, 1985). The H2A gene is contained completely within a 706 b.p. XhoI fragment and the H2B gene within a 533 b.p. XhoI-EcoRI fragment.

At this stage the orientation of the XhoI fragment in pCH3.3E had not been determined, and in fact had been assumed to be oriented such that the H2A gene was transcribed in the same direction as the H1 and H2B genes. From the sequence analysis of the 3.3 kb insert, three BstEII restriction sites were predicted. All were in coding regions, one in the H2A and two in the H2B. The contiguous sequence of the H2B gene meant that there had to be a 300 b.p. BstEII coding fragment. Depending on the orientation of the XhoI fragment a 407 b.p. (co-directional transcription) or 702 b.p. (divergent transcription) BstEII band was predicted. Upon analysis of the DNA (figure 2-4) a band of approximately 700 b.p. was found.

The discovery of the 702 b.p. BstEII band meant that the H2A/H2B genes of λCH-01 were divergently transcribed, and sequencing from the SmaI site of pCH3.3E through the XhoI site 140 b.p. away (figure 2-4) confirmed this. Subsequently, sequencing of a DNA fragment containing the intergenic XhoI site confirmed the junction of the H2A and H2B genes (A. Robins, this laboratory).

(2) Cosmid 6.1C

Having the sequence of the λCH-01 H2A/H2B gene pair available did not

Figure 2-4: Organization of H2A/H2B genes contained within λ CH-01

A schematic representation of the genomal clone λ CH-01 and the plasmid subclone derived from it pCH3.3E, illustrates the position and orientation of the H1 gene, and the H2A/H2B divergent gene pair. The position of BstEII restriction sites used to orientate the genes of the divergent pair are shown along with the unique SmaI and the two XhoI restriction sites.

B = BstEII

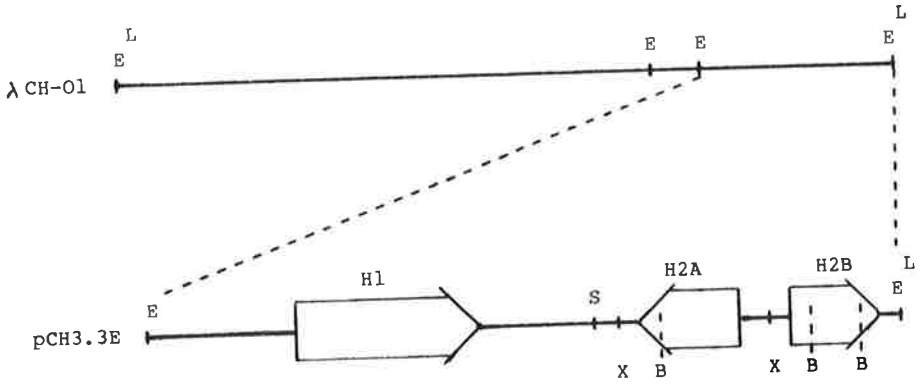
S = SmaI

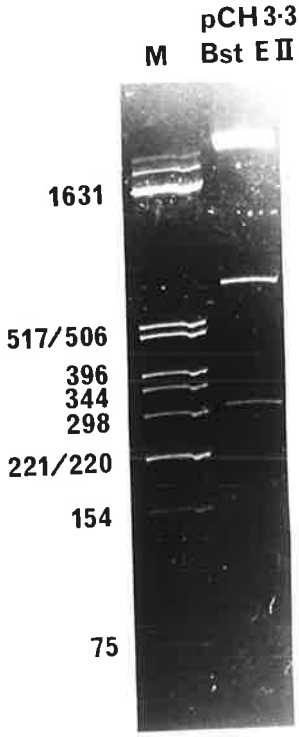
X = XhoI

E = EcoRI

E^{L} = EcoRI linker

The photograph of the polyacrylamide gel containing the BstEII digest of pCH3.3E and a pBR322 HinfI marker track is shown below the figure. The 700 base BstEII fragment indicates that the H2A and H2B genes must be directed away from each other. The 350 base BstEII band results from two internal H2B gene BstEII recognition sites.


λ CH-01
L
E
E
E
E
L
pCH3.3E
E
H1
S
H2A
H2B
L
E
X
B
X
B
B


pCH3·3E
M
Bst EII
1631
517/506
396
344
298
221/220
154
75

conclusively identify the promoter elements contained within the intergene region although likely TATA and CCAAT box elements (1-4-2(i)) were suggested. By comparison to a single independent chicken H2B gene contained within λCH-02 (figure 2-2), and several H2B genes of other species a so called "ubiquitous H2B-gene specific 5' element" was identified in front of the H2B TATA box (Harvey et al., 1982).

Given this lack of readily identifiable motifs that could potentially be shared by the two genes it was decided to examine other chicken H2A/H2B gene pairs in order to determine their respective orientations and, if appropriate, to do comparative sequence analysis. Examination of the restriction map of cosmid 6.1C (Lesnikowski, 1983; D'Andrea et al., 1985; figure 2-2) identified a H2A/H2B gene system containing a XhoI site which separated the two coding regions when southern blot analysis was done. Because a H2A/H2B intergenic XhoI site was found in λCH-01 (figure 2-4) it seemed reasonable to investigate, by sequence analysis, the nature of the cosmid 6.1C H2A/H2B intergenic region.

A subclone from cosmid 6.1C, named pCH22.0B (figure 2-2) was used as the source of DNA for this H2A/H2B pair. A 5.0 kb EcoRI fragment (figure 2-5) was isolated preparatively from this clone and redigested with XhoI. A 1.0 kb XhoI-EcoRI H2B containing fragment and a 2.3 kb partially digested XhoI-EcoRI fragment were subcloned into an M13mp8 SalI-EcoRI directional sequencing vector. Upon analysis of the 1.0 kb fragment, reading from the XhoI site, the start of the H2B coding region was readily identified (figure 3-3) thus orientating the gene toward the EcoRI site. The presence and position of the XhoI recognition sequence was found to be co-incidental, in that it was not in the same position as the λCH-01 intergene XhoI site (also I beleive it to be of no consequence when comparative sequence analysis is done; see chapter 3).

Figure 2-5: Analysis of the H2A/H2B gene pair of cosmid 6.1C

The diagram shows a gross restriction map of the 22.0 kb BamHI subclone from cosmid 6.1C and the 5 kb EcoRI restriction fragment used in the analysis of the H2A/H2B gene pair. The strategy used in obtaining the partial sequence of the two genes (figure 3-3) is also outlined.

B = BamHI

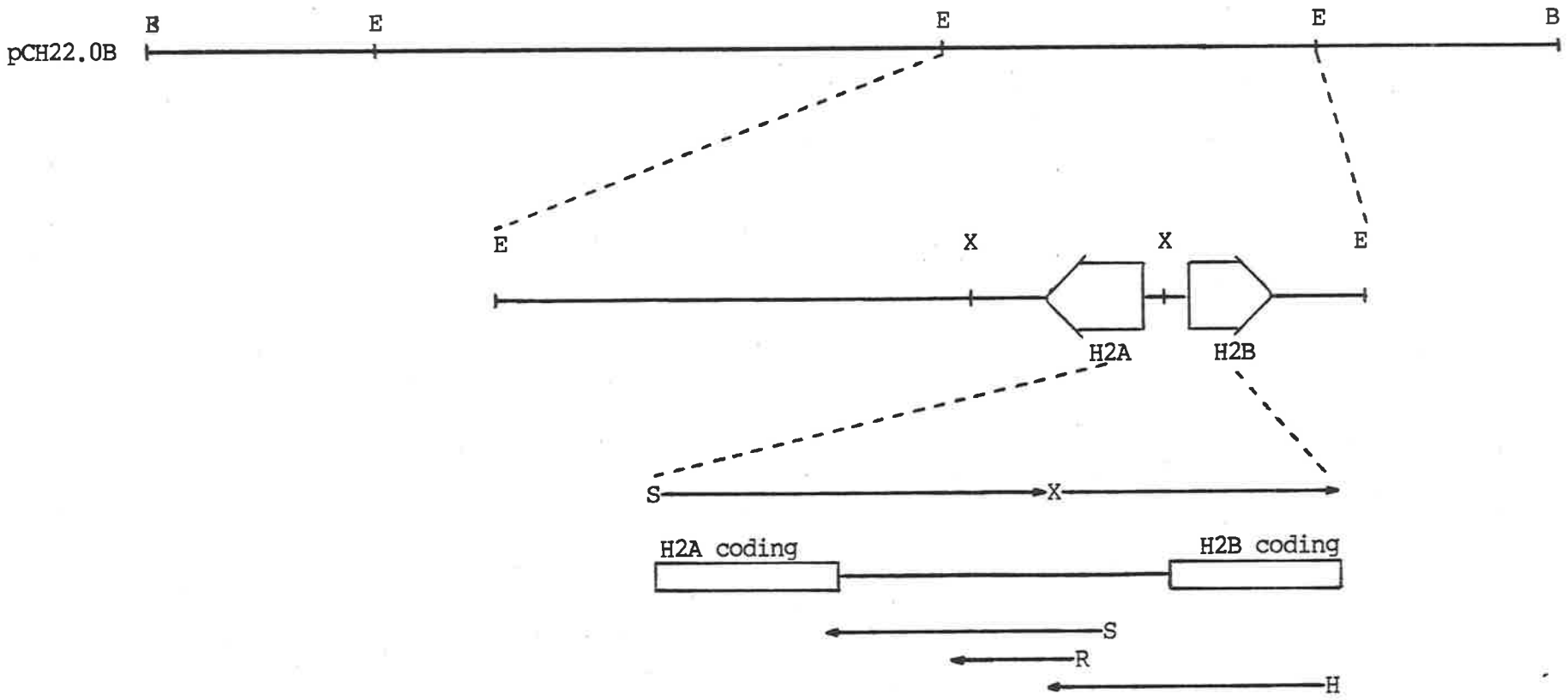
E = EcoRI

X = XhoI

S = SauIIIA

R = RsaI

H = HaeIII


pCH22.0B
E
E
E
E
B
E
X
X
E
H2A
H2B
S
X
H2A coding
H2B coding
S
R
H

The 2.3 kb XhoI-EcoRI clone containing both the H2A and H2B coding regions was restricted with Sau3A as a recognition site for this enzyme was discovered close to the intergenic XhoI site. It was then run on an acrylamide gel next to a track of M13mp8 Sau3A digested DNA in an attempt to identify bands that were unique to the 2.3 kb partial XhoI-EcoRI insert. A band of approximately 500 b.p. was seen to be unique to the 2.3 kb XhoI-EcoRI clone, it was therefore excised and cloned into an M13mp8 BamHI vector. Sequence analysis of one direction of this Sau3A band was found to go through the intergenic XhoI site and into the start of the H2A coding region. A second divergently orientated chicken H2A/H2B gene pair was thus characterized as shown in figure 2-5.

(3) λCH-07

A third H2A/H2B gene pair was mapped on a clone overlapping λCH-01, termed λCH-07 (figure 2-1). A 3.5 kb EcoRI fragment containing this pair had been subcloned into pBR325 by Richard D'Andrea to facilitate mapping of the chicken histone gene loci. Fine restriction mapping of this pCH3.5E clone however was not intended to orientate the gene pair.

From mapping data a SacI restriction site was seen to separate the two coding regions (figure 2-6). In addition another SacI site was found in the H2A coding region. This unique SacI band of approximately 550 b.p. was purified through an acrylamide gel and cloned by blunt ending the fragment with DNA polymerase I Klenow fragment and ligating into the SmaI site of the M13mp9 sequencing vector. The complete sequence of this fragment was then obtained. The H2A coding region was indeed found but in itself this fragment did not orientate the H2A gene.

Utilizing the sequence generated from the cloned SacI band and with the liklihood that a divergent pair was indeed being studied a PstI-BstEII

Figure 2-6: Analysis of the H2A/H2B gene pair of λ CH-07

The position of the 3.5 kb EcoRI restriction fragment of λ CH-07 containing the H2A/H2B gene pair along, with the position of the SacI restriction sites used in the initial sequencing/orientation of the gene pair are shown. The strategy used to obtain the sequence data presented in figure 3-4 is outlined below the restriction maps.

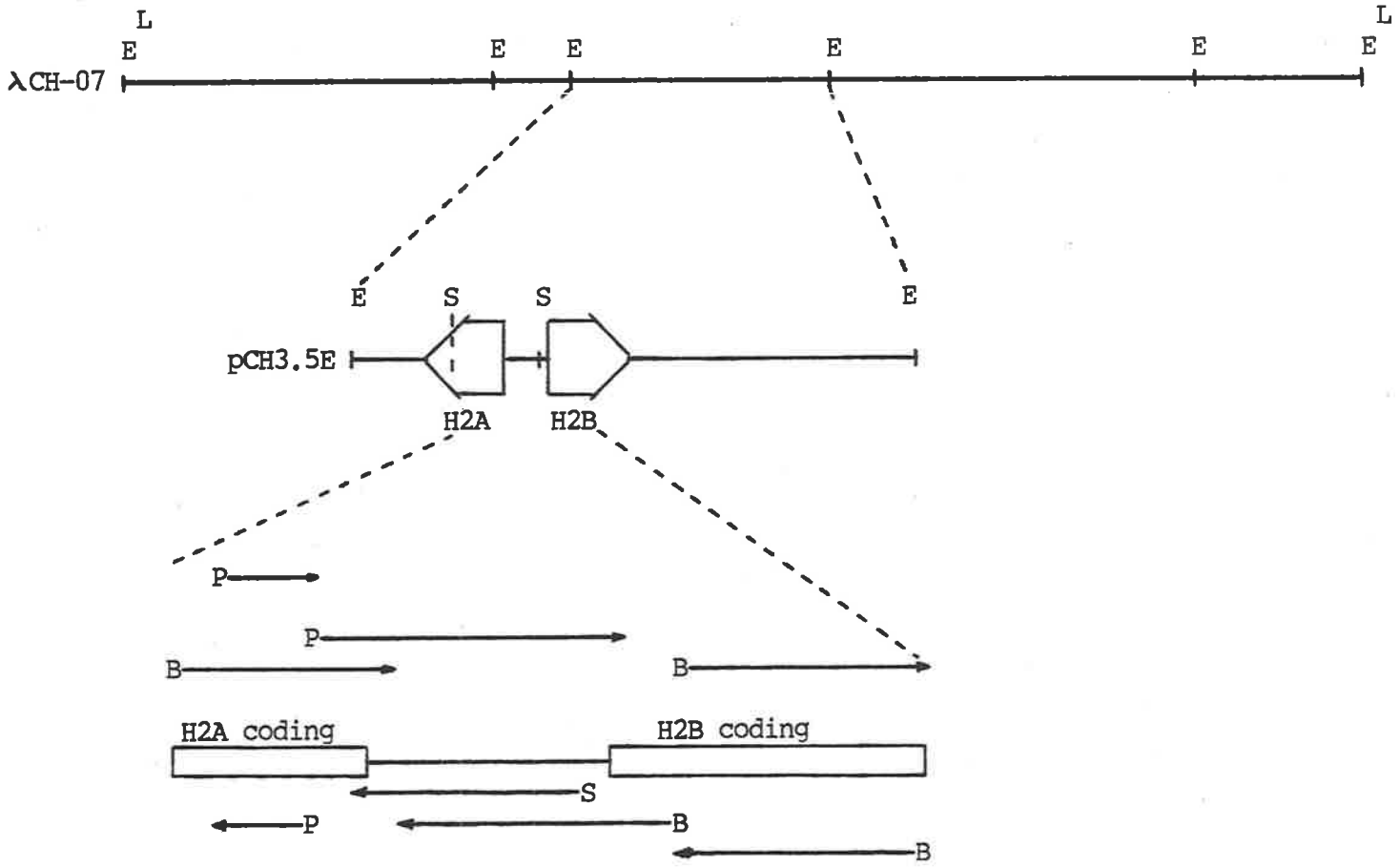
E = EcoRI

E^{L} = EcoRI linker

S = SacI

B = BstEII

P = PstI


L
E
λCH-07
E
E
E
E
L
E
E
S
S
E
pCH3.5E
H2A
H2B
P
P
B
B
H2A coding
H2B coding
S
P
B
B

fragment of approximately 470 b.p. generated upon digestion of pCH3.5E was purified, blunt-ended and cloned. The coding regions of every iso-coding chicken histone gene, excepting H1, is very highly conserved. From other data it seemed likely that there was a BstEII site near the beginning of the H2B coding region, and it was known (from the SacI sequence) that a PstI site was present near the start of the H2A coding region. A band of about 470 b.p. was consistent with the belief that a divergent pair was being examined. The sequencing data provided by this insert spanned the intergenic SacI site and entered the start of the H2B coding region. The hypothesis of divergent gene orientation was thus confirmed (figure 2-6).

(4) Cosmid 6.3C

Two additional H2A/H2B gene pairs were discovered in the λCH-01 overlapping cosmid 6.3C (figure 2-1). It was decided to restrict pCH11.0E, a subclone from cosmid 6.3C (figure 2-1), with EcoRI and SalI to separate these two pairs from each other and so not confuse restriction fragments generated from them. To completely remove any doubt of identification, the left-hand 4.5 kb EcoRI-SalI fragment of pCH11.0E was cloned into the SalI-EcoRI sites of M13mp8. The right handed SalI-EcoRI fragment was obtained in sufficient quantity and purity so as not to necessitate cloning.

The left-handed and right-handed SalI-EcoRI isolates of pCH11.0E were digested with PstI-BstEII. Each generated a 470 b.p. fragment on an acrylamide gel. As discussed above (and will become apparent from the data discussed in chapter 3) the size of these PstI-BstEII bands is consistent with divergent H2A/H2B genes and as such these size bands were good candidates for sequencing. These bands were isolated separately, treated with Klenow enzyme and ligated into the SmaI site of M13mp9. Sequencing of each fragment identified both pairs as divergent H2A/H2B genes as shown in figure 2-7.

Figure 2-7: Analysis of the H2A/H2B gene pairs of cosmid 6.3C

An 11.0 EcoRI subclone from cosmid 6.3C (figure 2-1) was used as the source of DNA for the two pairs of divergent H2A/H2B genes of this cosmid clone. The restriction map of the left (4.5 kb) and right (5.5 kb) EcoRI-SalI fragments of pCH11.0E and the position and orientation of the gene pairs in them is presented. The restriction sites used for sequencing of the two intergene regions (figure 3-5) are shown.

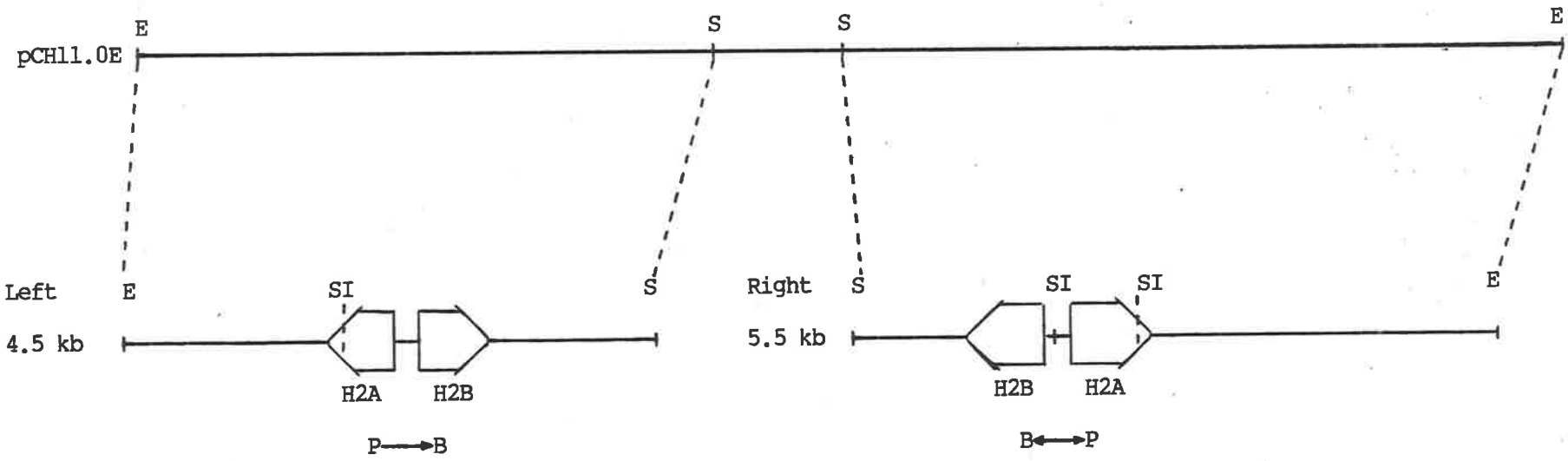
E = EcoRI

S = SalI

SI = SacI

P = PstI

B = BstEII


pCH11.0E
E
S
S
E
Left
4.5 kb
E
SI
S
H2A
H2B
P⟶B
Right
5.5 kb
S
SI
SI
E
H2B
H2A
B⟷P

(5) λ CH1-10

The last H2A/H2B gene pair isolated is contained within the cloneλCH1-10 (figure 2-3). The genes in this clone form a separate cluster, with no apparent linkage to the λCH-01 or λCH-02 loci. An EcoRI-HindIII fragment from this clone was inserted into the EcoRI-HindIII sites of M13mp8 and M13mp9 by Leeanne Coles. Sequencing of these two clones readily identified the H2A coding region reading from the EcoRI linker site and the H2B coding region reading from the HindIII site. A divergent gene orientation was thus determined as shown in figure 2-3.

2-3 Discussion

The histone genes present in the chicken genome can be divided, by close association, into 11 clusters (figures 2-1, 2-2 and 2-3). None of these arbitrary regions contain an entire complement of core and H1 genes and there is no evidence for a repeating unit between the clusters. Within 5 of the clusters, however, there are closely associated and divergently transcribed H2A/H2B gene pairs. This leaves four of the H2A and two of the H2B gene complement (of 10 H2A and 8 H2B) as separate entities.

One of these independent gene clusters without a divergent H2A/H2B pair retains the H1, H2A , H2B gene association (figure 2-2). The remaining H2B gene is associated with an H3 gene and two H4 genes in an apparent inverted duplication (figure 2-1; Wang et al., 1985). The last H2A gene is contained within a fourth histone gene loci represented by a single cosmid clone (not presented here) in association with a H3 gene.

A form of a repeating unit has thus been discovered within the apparently randomly organized chicken histone multigene family. It is interesting to consider the gene associations of other non-repeating histone gene systems. Upon examination of the gene organization tables in

Maxson et al. (1983a) it is clear that a divergent organisation of H2A/H2B pairs is present in a number of species such as Yeast, Sea Urchin (late transcripts), Xenopus and Human. This arrangement is also present in the repeating systems of Drosophila and Newt. This association of H2A/H2B genes through evolution suggests a functional significance of this arrangement.

CHAPTER 3: COMPARATIVE SEQUENCE ANALYSIS OF CHICKEN H2A/H2B GENE PAIRS

3-1 Introduction

Concurrently with the orientation of the six divergent H2A/H2B pairs the sequence of the intergene regions could be compared. The sequence comparison for all six pairs was initially intended to analyze the promoter regions of these genes, but in addition several of the protein coding regions were extensively sequenced during orientation, and therefore were also available for comparison. The complete sequence of pCH3.3E (figure 2-4) had been obtained previously, but with the isolation of overlapping clones to λCH-01 it became possible to sequence the 3' end of the H2B gene, which was not contained within pCH3.3E, to complete this gene in its entirety.

3-2 Sequence of six chicken H2A/H2B gene pairs

(1) pCH3.3E/p7AT

The construction of the chicken genomal λ library (screened by Richard D'Andrea and Leeanne Coles) involved a partial HaeIII/AluI restriction digestion of chicken DNA, ligation of a 12 base EcoRI linker onto the blunt termini generated by these enzymes and cloning into the EcoRI site of the charon 4A vector. pCH3.3E has an EcoRI linker just after the end of the H2B protein-coding region. With the isolation of λCH-05 and the discovery that it overlapped the right hand end of λCH-01 it became possible to extend the sequence of pCH3.3E and so find the 3' histone terminator of this H2B gene.

A 7 kb genomal EcoRI fragment generated from λCH-05 contained the entire 3.3 kb EcoRI fragment of λCH-01 and pCH3.3E (figure 3-1). This was subcloned into the EcoRI site of pBR325 by Richard D'Andrea for mapping purposes (figure 2-1) and into the EcoRI site of pAT153 by myself to facilitate the isolation of the H2B 3' end (figure 3-1) and the gene expression work to be discussed in chapters 4 and 5. From mapping data a

Figure 3-1: Isolation of the 3' end of the pCH3.3E/p7AT H2B gene

The overlapping chicken histone genomal clones λ CH-01 and λ CH-07 and a gross restriction map of p7AT are presented. The BstEII-SacII restriction fragment used to generate the sequence of the H2B 3' end is shown.

L
E = EcoRI linker

E = EcoRI

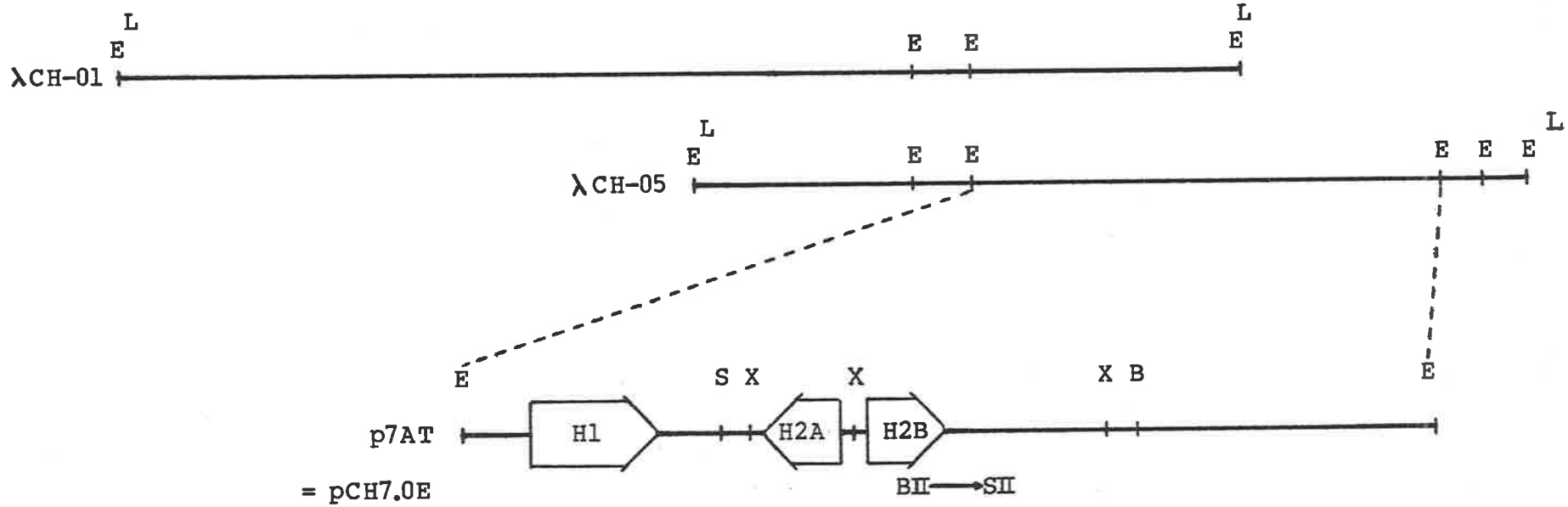
S = SmaI

X = XhoI

B = BamHI

BII = BstEII

SII = SacII


L
E
λCH-01
E E
L
E
L
E
λCH-05
E E
E E E
L
E
S X
X
X B
E
p7AT
H1
H2A
H2B
= pCH7.0E
BII→SII

SacII site was found 3' to the H2B gene. Sequence of the H2B gene predicted a BstEII restriction site very close to the right-hand EcoRI linker of λ CH-01. A BstEII-SacII digestion of p7AT allowed isolation of the required fragment of about 700 b.p. which was purified through an acrylamide gel. The termini were blunt ended with Klenow enzyme and the fragment ligated into the HincII site of M13mp7. Sequencing in from the BstEII site generated the 3' end of the H2B gene discussed by Harvey et al. (1982).

The sequence of p7AT from the unique SmaI site to the end of the H2B gene is presented in figure 3-2. This includes the complete sequence of both the H2A and H2B mRNA transcription units.

(2) pCH22.0B

The sequencing strategy for the H2A/H2B pair of cosmid 6.1C is shown in figure 2-5. 751 b.p. of continuous DNA sequence is presented in figure 3-3 which spans the intergenic region of the genes and includes the first 63/72 N-terminal amino acids of H2A/H2B respectively.

(3) pCH3.5E

The sequencing strategy for the H2A/H2B pair contained in pCH3.5E is shown in figure 2-6. A continuous sequence of 1017 b.p. is presented in figure 3-4. The first 102/120 amino acids of H2A/H2B are contained within this sequence.

(4) pCH11.0E

The sequencing strategy for the left and right divergent pairs of pCH11.0E is shown in figure 2-7. A non-contiguous 488 b.p. for the left H2A/H2B and a complete 463 b.p. for the right H2A/H2B is shown in figure 3-5. In each pair 25/20 N-terminal amino acids of H2A/H2B are presented.

(5) λCH1-10

The complete sequence of the EcoRI linker-HindIII H2A/H2B containing

Figure 3-2: Sequence of the pCH3.3E/p7AT H2A/H2B gene pair

The complete sequence of the H2A and H2B transcription units from p7AT is shown (data taken from Harvey et al., 1982). Coding regions have been translated and the intergene region is numbered on both strands starting from each ATG initiation codon. The restriction enzyme recognition sites for SmaI, XhoI, BstEII and the two HpaII sites that encompass the intergene region are indicated, as are the dyad symmetry elements responsible for mRNA 3' end formation. The elements identified in Table 3-1 are highlighted in colour (see legend to Table 3-1). The beginning and end of each mRNA is marked by vertical arrows.

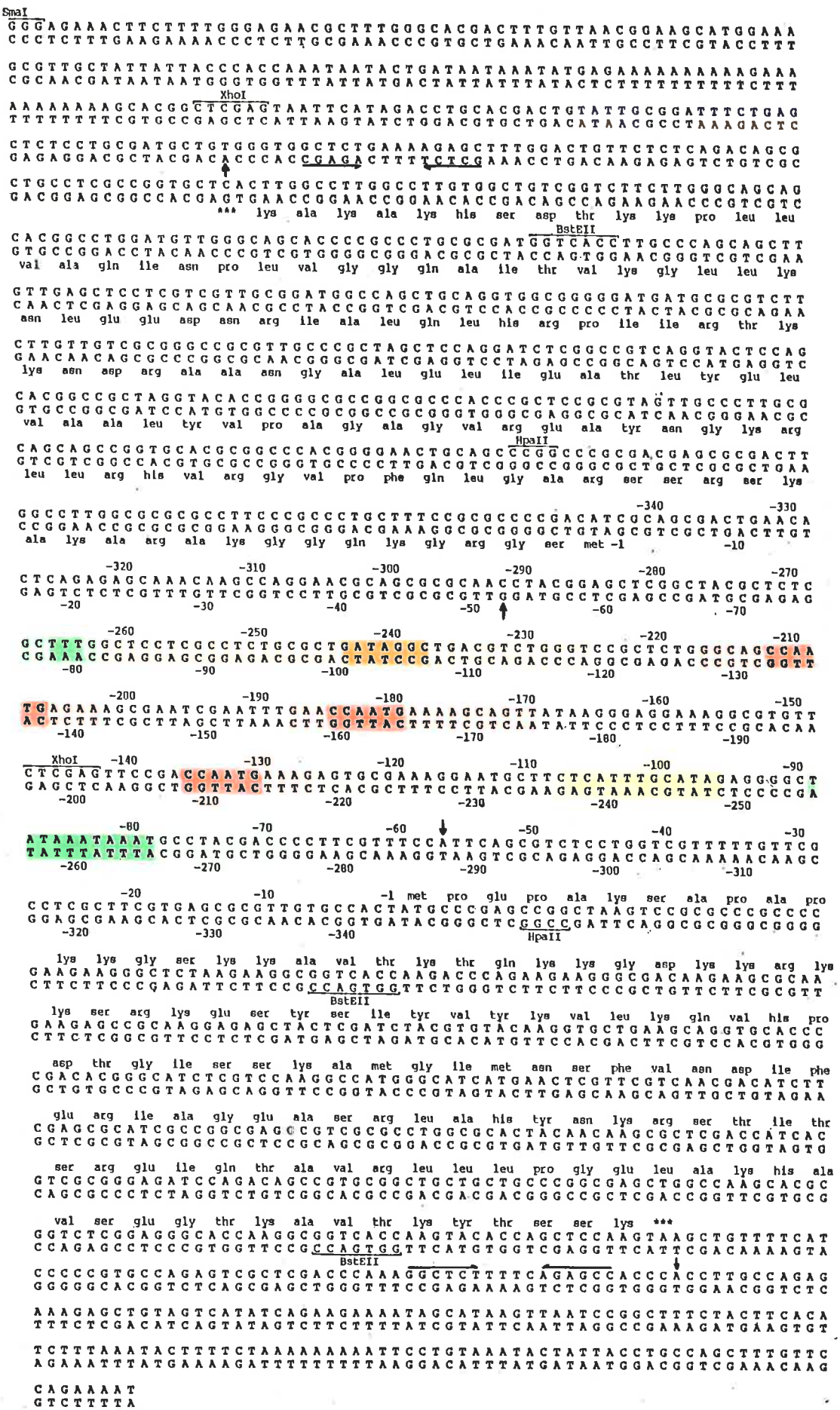

SmaI
XhoI
*** lys ala lys ala lys his ser asp thr lys lys pro leu leu
BstEII
val ala gln ile asn pro leu val gly gly gln ala ile thr val lys gly leu leu lys
asn leu glu glu asp asn arg ile ala leu gln leu his arg pro ile ile arg thr lys
lys asn asp arg ala ala asn gly ala leu glu leu ile glu ala thr leu tyr glu leu
val ala ala leu tyr val pro ala gly ala gly val arg glu ala tyr asn gly lys arg
HpaII
leu leu arg his val arg gly val pro phe gln leu gly ala arg ser ser arg ser lys
ala lys ala arg ala lys gly gly gln lys gly arg gly ser met -1 -10
-340 -330 -320 -310 -300 -290 -280 -270
-20 -30 -40 -50 -60 -70
-260 -250 -240 -230 -220 -210
-80 -90 -100 -110 -120 -130
-200 -190 -180 -170 -160 -150
-140 -150 -160 -170 -180 -190
XhoI -140 -130 -120 -110 -100 -90
-200 -210 -220 -230 -240 -250
-80 -70 -60 -50 -40 -30
-260 -270 -280 -290 -300 -310
-20 -10 -1 met pro glu pro ala lys ser ala pro ala pro
-320 -330 -340 HpaII
lys lys gly ser lys lys ala val thr lys thr gln lys lys gly asp lys lys arg lys
BstEII
lys ser arg lys glu ser tyr ser ile tyr val tyr lys val leu lys gln val his pro
asp thr gly ile ser ser lys ala met gly ile met asn ser phe val asn asp ile phe
glu arg ile ala gly glu ala ser arg leu ala his tyr asn lys arg ser thr ile thr
ser arg glu ile gln thr ala val arg leu leu leu pro gly glu leu ala lys his ala
val ser glu gly thr lys ala val thr lys tyr thr ser ser lys ***
BstEII

Figure 3-3: Sequence of the pCH22.0B H2A/H2B gene pair

The partial sequence of the H2A/H2B transcription units of pCH22.0B is shown. The protein coding regions are indicated above or below the DNA sequence. The SauIIIA, RsaI, XhoI and HaeIII restriction sites used in the analysis of this gene pair (figure 2-5) are outlined in the figure. The two amino acid differences of the H2B protein encoded by this clone at positions 31 and 33 are boxed. The intergene elements identified in Table 3-1 are highlighted in colour (see legend to Table 3-1).

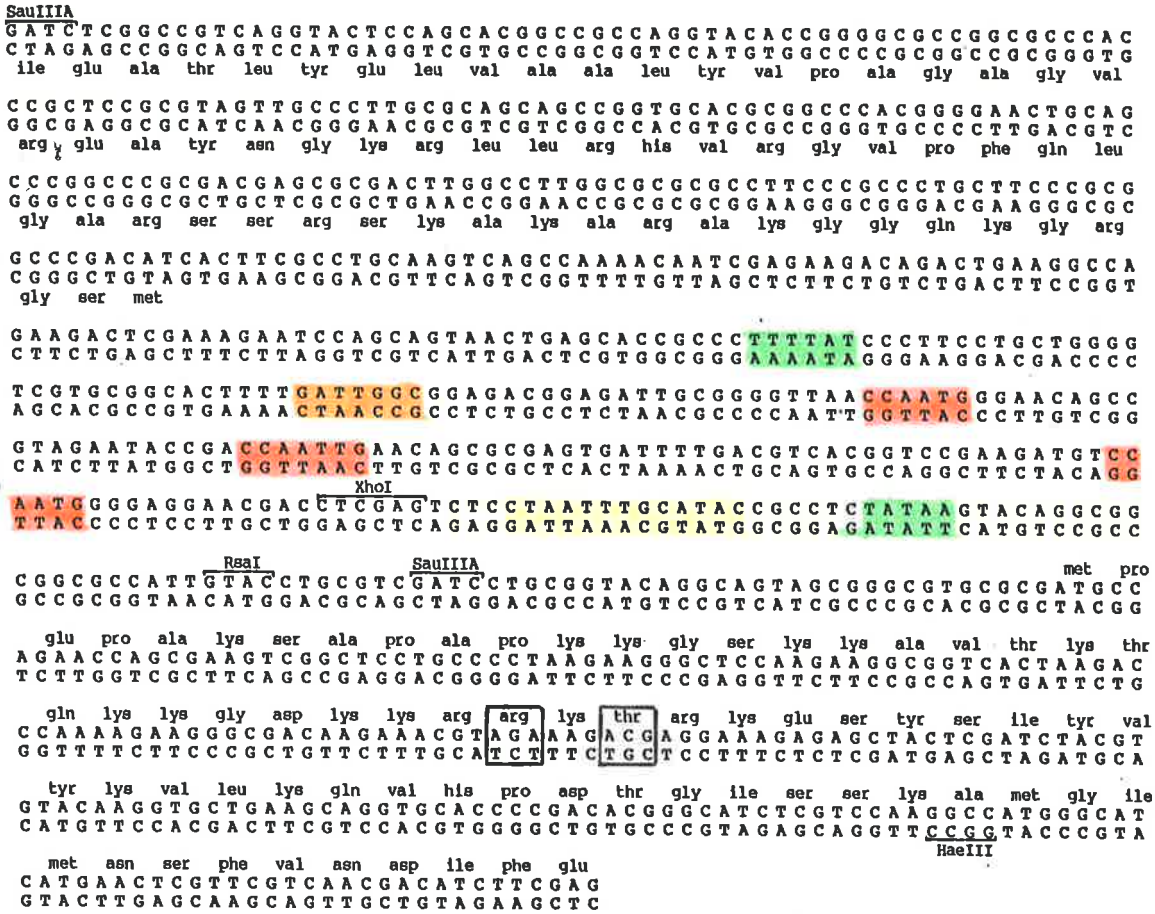

SauIIIA
GATCTCGGCCGTCAGGTACTCCAGCACGGCCGCCAGGTACACCGGGGCGCCGGCGCCCAC
CTAGAGCCGGCAGTCCATGAGGTCGTGCCGGCGGTCCATGTGGCCCCGCGGCCGCGGGTG
ile glu ala thr leu tyr glu leu val ala ala leu tyr val pro ala gly ala gly val
CCGCTCCGCGTAGTTGCCCTTGCGCAGCAGCCGGTGCACGCGGCCCACGGGGAACTGCAG
GGCGAGGCGCATCAACGGGAACGCGTCGTCGGCCACGTGCGCCGGGTGCCCCTTGACGTC
arg glu ala tyr asn gly lys arg leu leu arg his val arg gly val pro phe gln leu
CCCGGCCCGCGACGAGCGCGACTTGGCCTTGGCGCGCGCCTTCCCGCCCTGCTTCCCGCG
GGGCCGGGCGCTGCTCGCGCTGAACCGGAACCGCGCGCGGAAGGGCGGGACGAAGGGCGC
gly ala arg ser ser arg ser lys ala lys ala arg ala lys gly gly gln lys gly arg
GCCCGACATCACTTCGCCTGCAAGTCAGCCAAAACAATCGAGAAGACAGACTGAAGGCCA
CGGGCTGTAGTGAAGCGGACGTTCAGTCGGTTTTGTTAGCTCTTCTGTCTGACTTCCGGT
gly ser met
GAAGACTCGAAAGAATCCAGCAGTAACTGAGCACCGCCCTTTTATCCCTTCCTGCTGGGG
CTTCTGAGCTTTCTTAGGTCGTCATTGACTCGTGGCGGGAAAATAGGGAAGGACGACCCC
TCGTGCGGCACTTTTGATTGGCGGAGACGGAGATTGCGGGGTTAACCAATGGGAACAGCC
AGCACGCCGTGAAAACTAACCGCCTCTGCCTCTAACGCCCCAATTGGTTACCCTTGTCGG
GTAGAATACCGACCAATTGAACAGCGCGAGTGATTTTGACGTCACGGTCCGAAGATGTCC
CATCTTATGGCTGGTTAACTTGTCGCGCTCACTAAAACTGCAGTGCCAGGCTTCTACAGG
XhoI
AATGGGGAGGAACGACCTCGAGTCTCCTAATTTGCATACCGCCTCTATAAGTACAGGCGG
TTACCCCTCCTTGCTGGAGCTCAGAGGATTAAACGTATGGCGGAGATATTCATGTCCGCC
RsaI
SauIIIA
met pro
CGGCGCCATTGTACCTGCGTCGATCCTGCGGTACAGGCAGTAGCGGGCGTGCGCGATGCC
GCCGCGGTAACATGGACGCAGCTAGGACGCCATGTCCGTCATCGCCCGCACGCGCTACGG
glu pro ala lys ser ala pro ala pro lys lys gly ser lys lys ala val thr lys thr
AGAACCAGCGAAGTCGGCTCCTGCCCCTAAGAAGGGCTCCAAGAAGGCGGTCACTAAGAC
TCTTGGTCGCTTCAGCCGAGGACGGGGATTCTTCCCGAGGTTCTTCCGCCAGTGATTCTG
gln lys lys gly asp lys lys arg arg lys thr arg lys glu ser tyr ser ile tyr val
CCAAAAGAAGGGCGACAAGAAACGTAGAAAGACGAGGAAAGAGAGCTACTCGATCTACGT
GGTTTTCTTCCCGCTGTTCTTTGCATCTTTCTGCTCCTTTCTCTCGATGAGCTAGATGCA
tyr lys val leu lys gln val his pro asp thr gly ile ser ser lys ala met gly ile
GTACAAGGTGCTGAAGCAGGTGCACCCCGACACGGGCATCTCGTCCAAGGCCATGGGCAT
CATGTTCCACGACTTCGTCCACGTGGGGCTGTGCCCGTAGAGCAGGTTCCGGTACCCGTA
HaeIII
met asn ser phe val asn asp ile phe glu
CATGAACTCGTTCGTCAACGACATCTTCGAG
GTACTTGAGCAAGCAGTTGCTGTAGAAGCTC

Figure 3-4: Sequence of the pCH3.5E H2A/H2B gene pair

The partially determined amino acid sequences and nucleotide sequence of the intergene region of the pCH3.5E H2A/H2B gene pair are presented. BstEII, PstI, and SacI restriction enzyme recognition sites used in the generation of M13 clones for sequencing are outlined (figure 2-6). The intergene elements identified in Table 3-1 are highlighted in colour (see legend to Table 3-1).

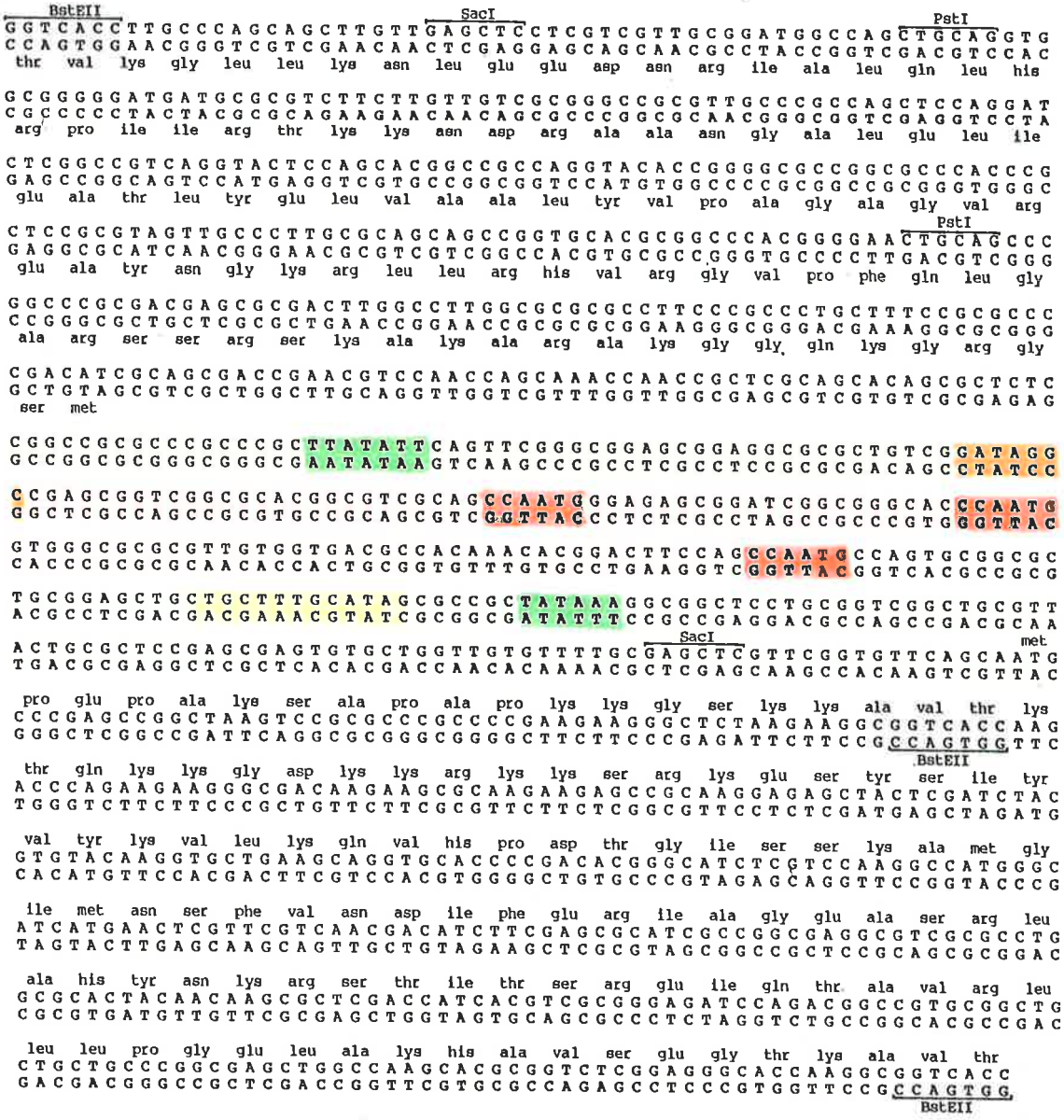

BstEII
SacI
PstI
PstI
SacI
met
BstEII
BstEII

Figure 3-5: Sequence of the pCH11.0E H2A/H2B gene pairs

The nucleotide and amino acid sequences of (a) the left H2A/H2B gene pair and (b) the right H2A/H2B gene pair of pCH11.0E. The sequence of the left gene pair was obtained by several sequencing runs of a PstI-BstEII restriction fagment clone orientated in only one direction. The dashes in the sequence represent undetermined bases. Both orientations of a Pst-BstEII intergenic fragment of the right hand gene pair were used in the determination of this sequence. The intergene sequences identified in Table 3-1 are highlighted in colour (see legend to Table 3-1).

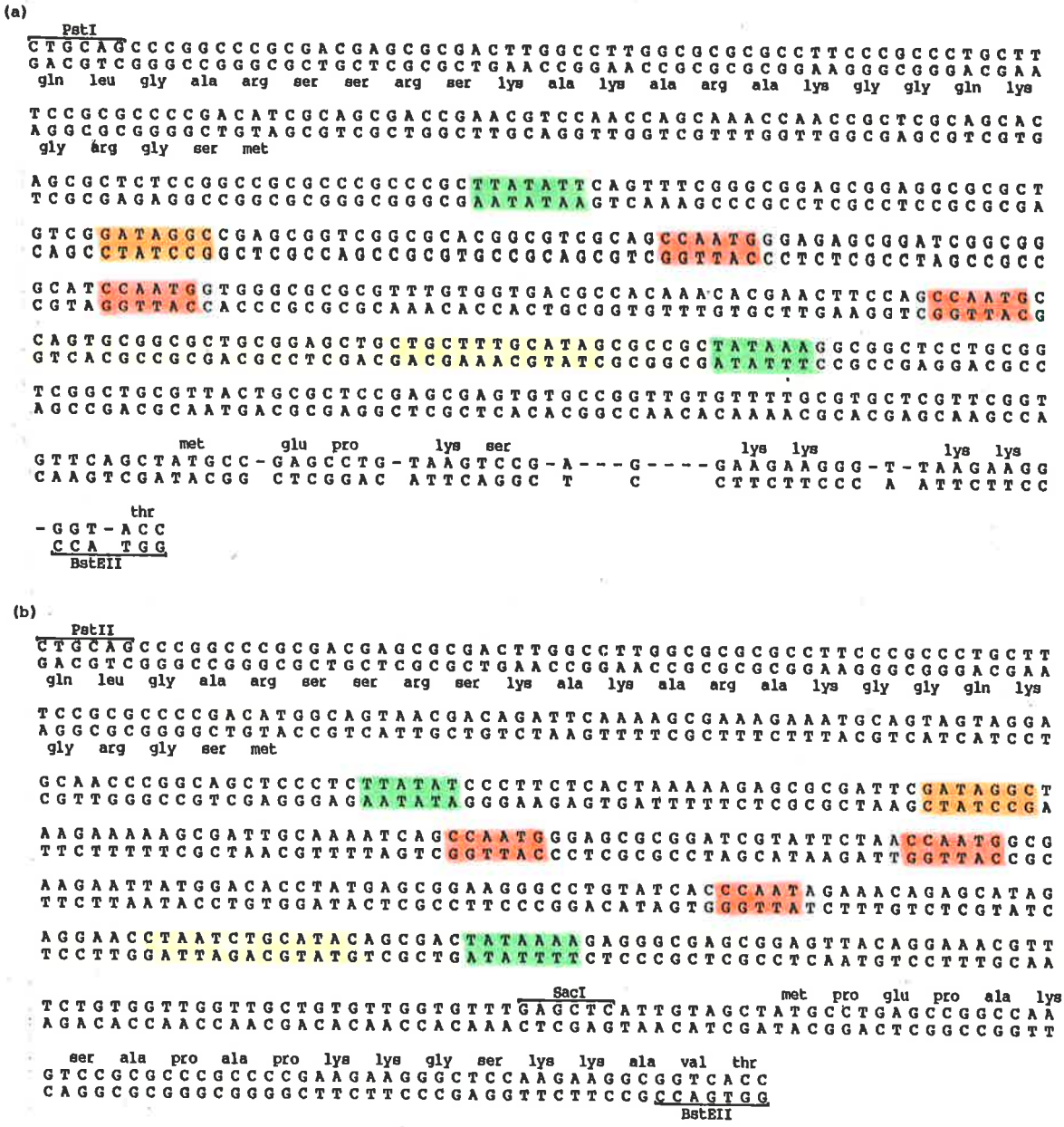
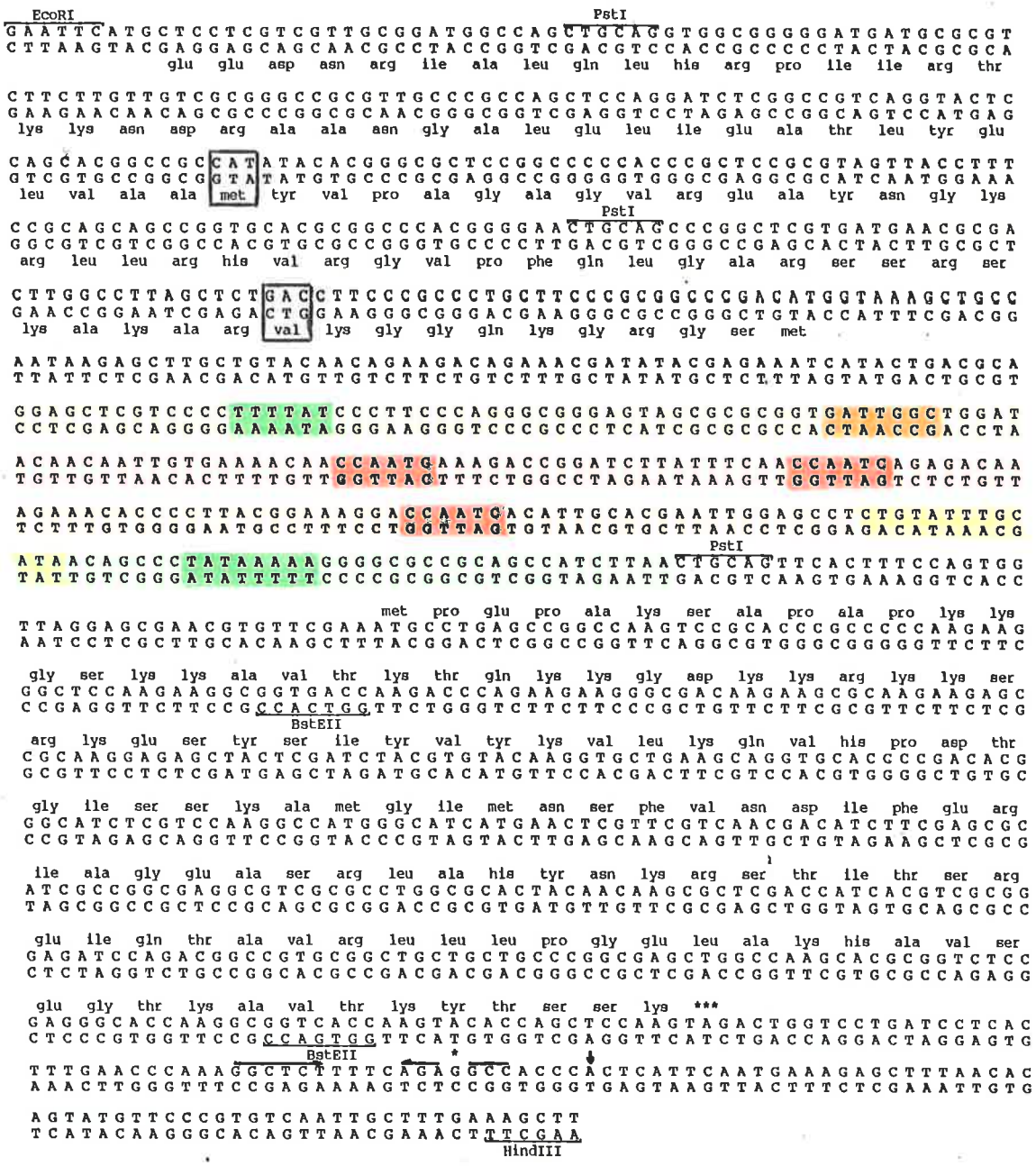

(a)
PstI
CTGCAGCCCGGCCCGCGACGAGCGCGACTTGGCCTTGGCGCGCGCCTTCCCGCCCTGCTT
GACGTCGGGCCGGGCGCTGCTCGCGCTGAACCGGAACCGCGCGCGGAAGGGCGGGACGAA
gln leu gly ala arg ser ser arg ser lys ala lys ala arg ala lys gly gly gln lys
TCCGCGCCCCCGACATCGCAGCGACCGAACGTCCAACCAGCAAACCAACCGCTCGCAGCAC
AGGCGCGGGGGCTGTAGCGTCGCTGGCTTGCAGGTTGGTCGTTTGGTTGGCGAGCGTCGTG
gly arg gly ser met
AGCGCTCTCCGGCCGCGCCCCGCCCCGCTTATATTCAGTTTCGGGCGGAGCGGAGGCGCGCT
TCGCGAGAGGCCGGCGCGGGCGGGCGAATATAAGTCAAAGCCCGCCTCGCCTCCGCGCGA
GTCGGATAGGCCGAGCGGTCGGCGCACGGCGTCGCAGCCAATGGGAGAGCGGATCGGCGG
CAGCCTATCCGGCTCGCCAGCCGCGTGCCGCAGCGTCGGTTACCCTCTCGCCTAGCCGCC
GCATCCAATGGTGGGCGCGCGTTTGTGGTGACGCCACAAACACGAACTTCCAGCCAATGC
CGTAGGTTACCACCCGCGCGCAAACACCACTGCGGTGTTTGTGCTTGAAGGTCGGTTACG
CAGTGCGGCGCTGCGGAGCTGCTGCTTTGCATAGCGCCGCTATAAAGGCGGCTCCTGCGG
GTCACGCCGCGACGCCTCGACGACGAAACGTATCGCGGCGATATTTCCGCCGAGGACGCC
TCGGCTGCGTTACTGCGCTCCGAGCGAGTGTGCCGGTTGTGTTTTGCGTGCTCGTTCGGT
AGCCGACGCAATGACGCGAGGCTCGCTCACACGGCCAACACAAAACGCACGAGCAAGCCA
met glu pro lys ser lys lys lys lys
GTTCAGCTATGCC-GAGCCTG-TAAGTCCG-A---G----GAAGAAGGG-T-TAAGAAGG
CAAGTCGATACGG CTCGGAC ATTCAGGC T C CTTCTTCCC A ATTCTTCC
thr
-GGT-ACC
CCA TGG
BstEII
(b)
PstII
CTGCAGCCCGGCCCGCGACGAGCGCGACTTGGCCTTGGCGCGCGCCTTCCCGCCCTGCTT
GACGTCGGGCCGGGCGCTGCTCGCGCTGAACCGGAACCGCGCGCGGAAGGGCGGGACGAA
gln leu gly ala arg ser ser arg ser lys ala lys ala arg ala lys gly gly gln lys
TCCGCGCCCCGACATGGCAGTAACGACAGATTCAAAAGCGAAAGAAATGCAGTAGTAGGA
AGGCGCGGGGCTGTACCGTCATTGCTGTCTAAGTTTTCGCTTTCTTTACGTCATCATCCT
gly arg gly ser met
GCAACCCGGCAGCTCCCTCTTATATCCCTTCTCACTAAAAAGAGCGCGATTCGATAGGCT
CGTTGGGCCGTCGAGGGAGAATATAGGGAAGAGTGATTTTTCTCGCGCTAAGCTATCCGA
AAGAAAAAGCGATTGCAAAATCAGCCAATGGGAGCGCGGATCGTATTCTAACCAATGGCG
TTCTTTTTCGCTAACGTTTTAGTCGGTTACCCTCGCGCCTAGCATAAGATTGGTTACCGC
AAGAATTATGGACACCTATGAGCGGAAGGGCCTGTATCACCCAATAGAAACAGAGCATAG
TTCTTAATACCTGTGGATACTCGCCTTCCCGGACATAGTGGGTTATCTTTGTCTCGTATC
AGGAACCTAATCTGCATACAGCGACTATAAAAGAGGGCGAGCGGAGTTACAGGAAACGTT
TCCTTGGATTAGACGTATGTCGCTGATATTTTCTCCCGCTCGCCTCAATGTCCTTTGCAA
SacI
met pro glu pro ala lys
TCTGTGGTTGGTTGCTGTGTTGGTGTTTGAGCTCATTGTAGCTATGCCTGAGCCGGCCAA
AGACACCAACCAACGACACAACCACAAACTCGAGTAACATCGATACGGACTCGGCCGGTT
ser ala pro ala pro lys lys gly ser lys lys ala val thr
GTCCGCGCCCGCCCCGAAGAAGGGCTCCAAGAAGGCGGTCACC
CAGGCGCGGGCGGGGCTTCTTCCCGAGGTTCTTCCGCCAGTGG
BstEII

Figure 3-6: Sequence of the λ CH1-10 H2A/H2B gene pair

The partial sequence of the H2A gene and the complete sequence of the H2B gene transcription unit and the intergene region of λ H1-10 are shown. The two amino acid changes in the H2A coding region at positions 11 and 52 are boxed. The position of the EcoRI, PstI, BstEII and HindIII restriction sites used to generate M13 clones for sequencing are outlined (figure 2-3). The 3' dyad symmetry element of the H2B gene showing the one base loopout (base shown by asterisk) is indicated by the arrows. The intergene sequences identified in Table 3-1 are highlighted in colour (see legend to Table 3-1).


EcoRI
PstI
glu glu asp asn arg ile ala leu gln leu his arg pro ile ile arg thr
lys lys asn asp arg ala ala asn gly ala leu glu leu ile glu ala thr leu tyr glu
leu val ala ala met tyr val pro ala gly ala gly val arg glu ala tyr asn gly lys
PstI
arg leu leu arg his val arg gly val pro phe gln leu gly ala arg ser ser arg ser
lys ala lys ala arg val lys gly gly gln lys gly arg gly ser met
PstI
met pro glu pro ala lys ser ala pro ala pro lys lys
gly ser lys lys ala val thr lys thr gln lys lys gly asp lys lys arg lys lys ser
BstEII
arg lys glu ser tyr ser ile tyr val tyr lys val leu lys gln val his pro asp thr
gly ile ser ser lys ala met gly ile met asn ser phe val asn asp ile phe glu arg
ile ala gly glu ala ser arg leu ala his tyr asn lys arg ser thr ile thr ser arg
glu ile gln thr ala val arg leu leu leu pro gly glu leu ala lys his ala val ser
glu gly thr lys ala val thr lys tyr thr ser ser lys ***
BstEII
HindIII

fragment of λCH1-10 has been obtained (strategy presented in figure 2-3). The entire H2B transcription unit and the first 93 amino acids of the H2A gene is shown in the 1113 b.p. of figure 3-6.

3-3 Results

The first H2A/H2B divergent gene pair isolated and examined was that from λCH-01. Because of this it was used as the "standard" divergent pair to which all the others were compared in computer analysis and my own analysis by eye. The material discussed below is summarized in Table 3-1.

(1) Protein Coding Regions

One of the striking aspects of the "core" chicken histone genes is the extraordinary conservation of DNA sequence in coding domains. Third base changes are usually rare. When the coding regions of the divergent pairs are each compared to the complete coding regions of pCH3.3E/p7AT, several third base and very few first and second base changes are seen. Most of these base changes result in identical amino acid sequences

The proteins encoded by the H2B genes of pCH3.3E/p7AT, pCH3.5E and λCH1-10 correspond to the H2B.1 variant described by Urban et al. (1979). Not enough DNA sequence was obtained to classify the pCH11.0E H2B encoded proteins. Cosmid 6.1C H2B gene has two amino acid changes (viz. lys to arg at position 31, and ser to thr at position 33) as shown in the amino acid sequence of figure 3-3 (boxed residues). These amino acid changes are conservative and are probably of no functional significance given that the N-terminus of the histone H2B protein is the most variable part of the protein when comparisons between species are made (Isenberg, 1979). In addition N-terminal residues 14 to 31 of the yeast H2B protein can be deleted without apparent adverse effect on the function of the protein (Wallis et al., 1983). These amino acid changes do not correspond with the H2B.2 subtype variant sequence (Urban et al., 1979) and moreover an H2B

Table 3-1: Comparative sequence analysis of chicken H2A/H2B genes

The six divergent chicken H2A/H2B intergene regions have been compared in an attempt to identify common regulatory sequences. A consensus of conserved regions has been compiled in the Table with distances between them indicated. Two independent H2A genes and one separate H2B gene are also presented in the comparison. The conserved regions identified are highlighted in colour (see figures 3-2 to 3-6), i.e. TATA boxes in green; CAT1, 2, 3 boxes in red; CAT4 in orange; H2B box in yellow). The distance between divergent pair ATG initiation codons, the protein subtype encoded and the number of amino acids determined from the DNA sequence for each gene pair are listed in the Table.

Table 3-1 Comparative sequence analysis of chicken histone H2A/H2B genes

Gene system	Bases between ATG	Subtype Variant	Number of Amino Acids	H2A ATG	TATA	CAT4	CAT3	CAT2	CAT1	H2B BOX	TATA	H2B ATG	Number of Amino Acids	Subtype Variant
pCH3.3E/p7AT	343	H2A.1	129 complete	CAT GTA	...77... TTT AAA	...19... GATAGGC CTATCCG	...25... CCAATG GGTTAC	...21... CCAATG GGTTAC	...43... CCAATG GGTTAC	...22... CTCATTTGCATAG GAGTAAACGTATC	...6... TATAAATAAAT ATATTTATTTA	...78... ATG TAC	126 complete	H2B.1
pCH22.0B	346	H2A.1	63		...90... TTTTAT AAAATA	...30... GATTGGC CTAACCG	...23... CCAATG GGTTAC	...21... CCAATTG GGTTATC	...39... CCAATG GGTTAC	...22... CTAATTTGCATAC GATTAAACGTATG	...6... TATAA ATATT	...65...	72	not yet described
pCH3.5E	351	H2A.1	102		...71... TTATATT AATATAA	...30... GATAGGC CTATCCG	...26... CCAATG GGTTAC	...21... CCAATG GGTTAC	...42... CCAATG GGTTAC	...22... CTGCTTTGCATAG GACGAAACGTATC	...6... TATAAA ATATTT	...82...	120	H2B.1
pCH11.0E L	353	H2A.1	25		...71... TTATATT AATATAA	...31... GATAGGC CTATCCG	...26... CCAATG GGTTAC	...21... CCAATG GGTTAC	...43... CCAATG GGTTAC	...22... CTGCTTTGCATAG GACGAAACGTATC	...6... TATAAA ATATTT	...82...	20	insufficient sequence
pCH11.0E R	328	H2A.1	25		...64... TTATAT AATATA	...27... GATAGGC CTATCCG	...25... CCAATG GGTTAC	...21... CCAATG GGTTAC	...43... CCAAT GGTTA	...21... CTAATCTGCATAC GATTAGACGTATG	...6... TATAAAA ATATTTT	...71...	20	insufficient sequence
λCH1.10	333	H2A.2	93		...85... TTTTAT AAAATA	...29... GATTGGC CTAACCG	...24... CCAATG GGTTAC	...21... CCAATC GGTTAG	...31... CCAATC GGTTAG	...21... CTGTATTTGCATA GACATAAACGTAT	...6... TATAAAAA ATATTTTT	...63...	126 complete	H2B.1
CONSENSUS	342				...76...	...28... GATAGGC CTATCCG	...25... CCAATG GGTTAC	...21... CCAATG GGTTAC	...40... CCAATG GGTTAC	...22... CTGATTTGCATAC GACTAAACGTATG C G	...6...	...74...		
λCH.02		no H2B gene			no sequence data available ... 4...		CCAGT GGTCA	...23... CCAATT GGTTAA	...42... CCAATC GGTTAG	...21... CTCGATTTGCATAC GAGCTAAACGTATG	...6... [redacted]	...59..	126 complete	H2B.1
pCH8.4E L		H2A.1	129 complete		...83... TTTATAT AAATATA	...30... GATAGGC CTATCCG	...27... CCAATG GGTTAC	...21... CCAATA GGTTAT	...43... GCAATG CGTTAC	...22... no homology				no H2B gene
pCH8.4E R		H2A.1	129 complete		...83... TTTATAT AAATATA	...30... GATAGGC CTATCCG	...27... CCAATG GGTTAC	...21... CCAATA GGTTAT	...43... GCAATG CGTTAC	...22... no homology				no H2B gene

protein containing these amino acid changes has not yet been reported in the literature.

Both of the H2A subtype variants reported in the literature can be identified from the sequence of the divergent pair H2A genes. pCH3.3E/p7AT, pCH22.0B, pCH3.5E and as far as the DNA sequence of pCH11.0E left and right H2A genes extend, all correspond to the H2A.1 subtype variant (Urban et al., 1979). The H2A protein of λCH1-10 has two amino acid changes (viz. ala to val at position 11, and leu to met at position 52) as shown in figure 3-6. These are protein changes in the H2A.2 variant identified by Urban et al., (1979).

(2) Intergene Region

Each of these independent chicken H2A/H2B gene pairs is extremely tightly linked. The coding regions are on average, 342 b.p. apart. Table 3-1 shows a comparative analysis of the six intergene regions to illustrate the position and spacing of the conserved elements identified. The organization of TATA boxes and hence predicted mRNA leader sequence lengths are similar for all the genes; an average of 180 b.p. separate the H2A and H2B TATA boxes.

The most significant homologies identified between the gene pair intergene regions are four CCAAT boxes (1-4-2(ii)). These are all regularly spaced with three of them having the consensus "CCAATG" (CAT 1, CAT 2 and CAT 3 of Table 3-1) positioned at approximately 40, 90 and 110 b.p. upstream of and directed towards the H2B TATA boxes. Each juxtaposed H2A gene has a "GCCTATC" sequence (CAT 4 of Table 3-1) orientated towards it and about 135 b.p. from the H2B TATA boxes. This fourth CCAAT box is an average of 28 b.p. upstream from the H2A TATA boxes.

The position of the pCH3.3E/p7AT H2A TATA box designated in Table 3-1

differs from that assigned by D'Andrea et al. (1981). The only A rich region to be found at a position consistent with the other divergent pairs is a run of three A residues. This is an extremely poor match to the consensus TATA box (1-4-2(i)) whereas the region identified by D'Andrea is an extremely good match. However the mRNA cap site of this H2A gene has been determined (chapter 4) and is consistent with my assignment of the TATA box. In addition, no true TATA box consensus for either the H2A or H2B gene can be derived from the comparative sequence analysis. It is possible that the full sequence of a TATA region is not of over-riding importance, but an ATA string can usually be identified in most TATA boxes.

Two of the intergene regions have almost exactly the same nucleotide sequence. pCH3.5E H2A/H2B and the left H2A/H2B pair of pCH11.0E have only seven b.p. differences between them. The sequence of the pCH3.3E H2A mRNA leader sequence is almost the same as that of the predicted H2A mRNA leader sequences of the pCH3.5E/pCH11.0E L H2A genes. All three of these pairs are in gene clusters that are adjacent on the chicken chromosome (figure 2-1). This may be significant in that gene conversion or gene duplication processes may have been involved in generating these homologies.

A fifth region of homology with almost exact conservation for 13 b.p., present 6 b.p. upstream of H2B TATA boxes, has been described previously (Harvey et al., 1982) and is known as the H2B box (Table 3-1). This will be discussed in length in chapter 6.

The sequence of two separate chicken H2A genes (pCH8.4E L and R, figure 2-1) and one separate chicken H2B gene (λCH-02, figure 2-2) is available for comparison (Table 3-1). These genes contain the same CCAAT elements as those identified in the divergent H2A/H2B pairs. They are less well conserved in sequence but the spacing and orientation of CCAAT boxes are still highly conserved. The independent H2A genes of pCH8.4E contain

almost the same sequence up to about 250 b.p. 5' to their ATG initiation codons and it is thought that this is the result on a duplication event (Wang et al., 1985). Their sequences diverge almost immediately after the most distal CCAAT box (CAT 1) and no evidence of a H2B box can be found. This suggests that the H2B box is truely H2B-specific.

The single H2B gene of λCH-02 contains all the elements of the divergent pairs up to the end point of the available sequence data (Table 3-1; Grandy et al., 1982). The three CCAAT boxes orientated toward the H2B are present but insufficient sequence data is available to determine whether the H2A CCAAT box (CAT 4) is present. This leaves open the question as to whether CAT 4, oriented toward the H2A gene, is gene specific or not.

(3) 3' Gene Elements

Of the divergent pairs discussed here only the 3' end of the H2B genes of pCH3.3E/p7AT and λCH1-10 and the 3' end of the H2A gene of pCH3.3E have been determined. The total sequence of many chicken histone genes within this cluster has however been obtained (Wang et al., 1985; Coles and Wells, 1985; L. Coles and J. Powell personal communication). All of these genes have been shown to contain the dyad symmetry required for the processing of histone mRNAs (1-6-3(iv)). An additional chicken specific 3 base extension to this 3' element has been noted previously (Harvey et al., 1982).

The 3' dyad symmetry of the λCH1-10 H2B gene is unusual in that the base pairing is not complete (figure 3-6). How this loop out in the stem would affect the processing of the 3' termini of this gene is not known, but the correct base pairing of the stem has been shown to be essential for the formation of a sea urchin H2A gene 3' end in the Xenopus oocyte (Birchmeier et al., 1983).

The H2B gene described in pCH3.3E/p7AT was given to Krieg and Melton

(1984) to conduct experiments to determine whether histone 3' termini are generated by RNA polymerase termination or RNA processing at the 3' hyphenated symmetry element. A XhoI fragment of p7AT was cloned into an Sp6 in vitro RNA synthesis vector and extended RNA transcripts of the H2B gene were obtained. Upon injection of this RNA into the nucleus of Xenopus oocytes the correct 3' termini was generated, indicating 3' processing was involved in the formation of histone 3' termini.

3-4 Discussion

The linkage of H2A/H2B genes in the chicken has been shown by sequence comparison to have a very strict association with respect to the distance between protein coding regions. Four CCAAT-like sequences have been identified in the intergene regions that also have a strict spatial association. This sequence has been shown to be important for eukaryotic gene expression (1-4-2(i)), although its exact role is not known for certain due to conflicting experimental results. So far only one CCAAT box has usually been emphasised in promoter regions of eukaryotic genes positioned 40-60 b.p. from the TATA box (Breathnach and Chambon, 1981). The nature of this element (or possibly a group of individual elements) termed the CCAAT box must be rethought given the number, orientation and specific sequences of them found in the chicken H2A/H2B genes. It may be that the CCAAT box can function in both orientations much like the eukaryotic enhancer elements (Khoury and Gruss, 1983) and the hexanucleotide CCGCCC element of the H.S.V. Thymidine kinase gene (McKnight et al., 1984).

Divergent, closely linked H2A/H2B genes are relatively common in histone gene families. Examination of the few that have been sequenced; Yeast, 600 b.p. intergene region (Wallis et al., 1980; Choe et al., 1982); Drosophila, 200 b.p. intergene region (Goldberg, 1979); Xenopus, 300-400 b.p. intergene region (Moorman et al.,1982) and a human H2A/H2B pair (Zhong

et al., 1983) has failed to identify the same promoter element arrangement. Such large evolutionary comparisons with yeast and Drosophila may be limited in their usefulness in trying to identify controlling elements other than TATA boxes. Insufficient sequencing data was obtained for the Xenopus divergent pair to make the comparison but it is potentially comparable. The identification of the human divergent pair referred to seems premature since no H2A coding region was identified.

Because the promoter regions of these divergent pairs are so close it is tempting to make an analogy with prokaryotic systems. Some prokaryotic genes have been found to be closly associated, allowing for co-ordinate regulation by common regulatory molecules. Two divergent gene pairs in E. coli have been shown to be regulated by proteins binding to sites between the genes (Brandsma et al., 1983). Gene promoter and operator interactions commonly occur in prokaryotes. The λ bacteriophage leftward and rightward transcripts being controlled by three binding sites for the CI repressor is such an example (Hendrix et al., 1983).

The control of the yeast mating type locus involves divergent transcription from a central promoter region (Siliciano and Tatchell, 1984). Moreover, deletions in the intergenic region (263 b.p. separate the ATG start codons) have mapped a single promoter region essential for transcription of each of the two genes of the locus. There is therefore a precedent in a eukaryotic organism for the sharing of a common regulatory region(s).

The possibility of regulatory elements being present in the coding regions of either gene of the chicken H2A/H2B pairs can be argued given the close and consistent distance between the coding regions of the two genes. This comparative sequence analysis presented here does not allow a

determination of the functional role of the conserved intergene elements or the coding regions. For this a high fidelity transcriptional assay has to be used in combinaton with *in vitro* mutagenesis of the divergent pair intergene/promoter region.

CHAPTER 4: ANALYSIS OF H2A/H2B TRANSCRIPTS IN 5 DAY CHICK EMBRYO RNA AND EXPRESSION OF AN ISOLATED H2A/H2B GENE PAIR IN THE XENOPUS OOCYTE

4-1 Introduction

Because of the unusually close apposition of the divergently transcribed H2A/H2B gene pairs, and with the aim of determining the nature of this linkage in the control of the expression of these genes, I initiated a study of transcription from one of the cloned gene pairs. Various systems are available for the analysis of gene expression (1-7), but a prerequisite of using any surrogate genetic system is that introduced genes must be expressed with fidelity. This seems to be true for genes microinjected into Xenopus oocyte nuclei (Gurdon and Wickens, 1983) and in particular the work done by Birnstiel's group on expression of sea urchin histone genes in Xenopus oocytes (Hentschel and Birnstiel, 1981) suggested that it was a generally useful system to study transcriptional control of chicken H2A/H2B divergent gene pairs.

Histone genes in the chicken exist in clusters in relatively low copy number in the genome. Their transcription is usually coupled to S-phase in dividing cells (J. Coleman, personal communication). When cloned chicken histone genes are microinjected into frog oocytes, the genes are present in very high copy number and as circular plasmid nucleoprotein rather than low copy number chromosomal structures. In addition in the developing oocyte expression of histone mRNA and protein is uncoupled from cell division , as this specialised cell builds a maternal store of histone to be used during embryonic development. For these reasons some caution must be exercised in extrapolating conclusions from heterologous systems.

In conjunction with the functional testing of a H2A/H2B pair, the endogenous chicken transcripts from these genes (in 5 day embryo extracts) were also examined to determine the fidelity and the relative level of each

transcript compared to that in microinjected oocytes. The gene pair chosen for study was the one that was first isolated and orientated, and for which the complete gene sequences were available. This gene system was the one from λCH-01. As discussed in chapter 3, the H2B 3' termini is missing from the 3.3 kb fragment generated by EcoRI digestion of λCH-01. A 7 kb genomal EcoRI fragment overlapping (and containing in full) the 3.3 kb fragment was obtained from λCH-05 (figure 2-1 and 3-1). This was subcloned into the EcoRI site of pAT153 for propagation and ease of isolation of the 7 kb insert. This clone was named p7AT and in addition to containing the complete divergent pair, an H1 gene is also contained within the insert.

4-2 Mapping of 5' and 3' termini of the H1, H2A and H2B genes of p7AT in 5 day chick embryo RNA

To examine whether histone gene transcripts obtained in the oocyte from injected templates (see below) were the same as is seen in vivo S1 nuclease protection experiments and primer extension assays for the H2A and H2B gene transcipts of p7AT present in 5 day chick embryo RNA were carried out. The DNA restriction fragments used for S1 analysis and the position of synthetic primers used in primer extension analysis are shown in figure 4-1. The H1 gene of p7AT was analyzed in collaboration with Leeanne Coles; the 5' and 3' mapping of this gene is presented by her in her Ph.D. thesis. It will suffice here to produce only the 5' termini mapping data with the H1 synthetic primer (figure 4-2).

Iso-coding chicken histone genes are highly homologous in the coding region but often diverge immediately in flanking sequences (compare the leader sequences of the divergent pairs presented in chapter 2; L. Coles and J. Powell, personal communication), it was therefore expected that two kinds of S1 resistant DNA products would be observed for each type of gene

Figure 4-1: DNA restriction fragments used for S1 analysis and extension of synthetic primers for the H2A/H2B genes of p7AT

The products of the H2A and H2B 26 base synthetic DNA primers after hybridisation to mRNA and extension with reverse transcriptase are indicated above the gene representations (see materials and methods for synthetic primer sequences). Below the gene figures are the positions and lengths of restriction fragments used in the S1 assays and the protected fragments after hybrisation of them to RNA and digestion with S1 nuclease. The restriction fragments used were as follows:

3' H2A = 205 bp XhoI-BstEII

5' H2A = 420 bp HpaII-HpaII

3' H2B = 650 bp BstEII-SacI

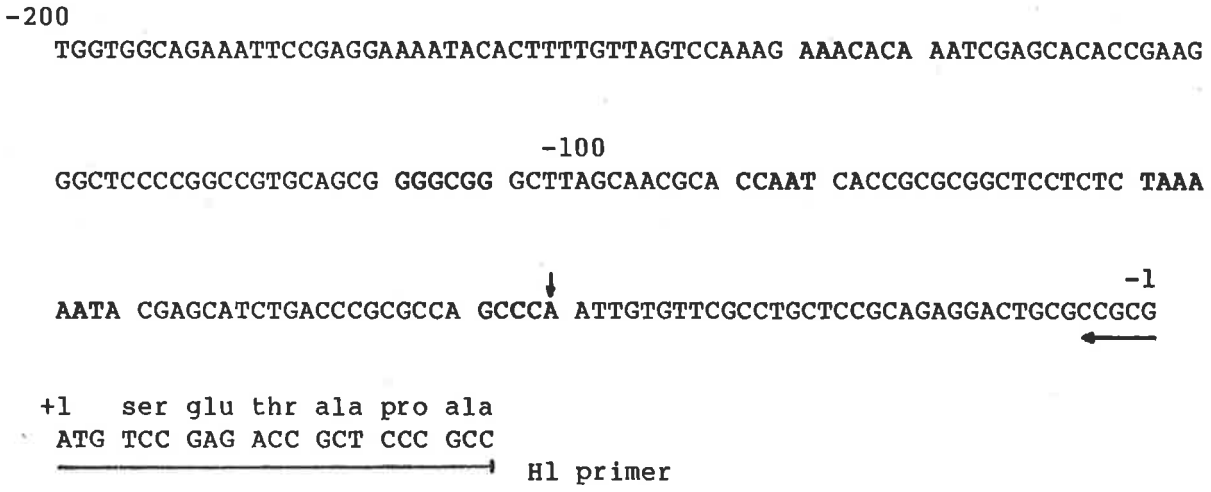
5' H2B = 202 bp XhoI-BstEII

PRIMER EXTENSION ANALYSIS

S1 NUCLEASE ANALYSIS

Figure 4-2: Sequence of the 5' region of the p7AT H1 gene

The 5' nucleotide sequence of the p7AT H1 gene sense strand is presented starting at -200 from the ATG initiation codon (data taken from Coles and Wells, 1985). Bases in bold type are H1 gene conserved elements thought to be involved in the promotion of the H1 transcript which begins at -37 in the sequence (shown by the vertical arrow). The position of binding of the H1 gene specific 26 base primer is indicated below the gene.


-200
TGGTGGCAGAAATTCCGAGGAAAATACACTTTTGTTAGTCCAAAG AAACACA AATCGAGCACACCGAAG
-100
GGCTCCCCGGCCGTGCAGCG GGGCGG GCTTAGCAACGCA CCAAT CACCGCGCGGCTCCTCTC TAAA
-1
AATA CGAGCATCTGACCCGCGCCA GCCCA ATTGTGTTCGCCTGCTCCGCAGAGGACTGCGCCGCG
+1 ser glu thr ala pro ala
ATG TCC GAG ACC GCT CCC GCC
Hl primer

transcript. One of these corresponds to the homologous hybrid, the other(s) to hybrids which are only homologous in the coding regions.

In addition, this heterogeneity in 5' mRNA leader sequences makes it possible to synthesise gene-specific primers. Empirically it was found that 26 base primers in the 5' untranslated region or spanning the 5' non-coding and the start of the coding region gave unambiguous extension products. That is, they generally permitted detection of specific transcripts from genes microinjected into Xenopus oocytes, and in total chick embryo RNA could be used to define transcripts from precise genes. Well after the oocyte transcription studies were initiated, sequencing of other H2A/H2B gene pairs revealed that the leader of the H2A gene in p7AT is homologous enough to the H2A leaders of pCH3.5E and pCH11.0E L (figures 3-4 and 3-5(a)) to allow cross-hybridization of any synthetic primer that could be designed. This was not realized until the sequence of these H2A genes became available, and after the synthesis of the p7AT H2A synthetic 26 base primer. Consequently the p7AT H2A primer also detected transcripts from other H2A genes in RNA from chick embryos.

(i) 3' Termini

The results of S1 analysis to determine 3' termini of H2A and H2B transcripts in chick embryo RNA are shown in figure 4-3. As discussed above, more than one set of protected fragments is to be expected from a heterologous mixture of RNAs (multiple, closely migrating bands within each region are often encounted in S1 analysis and are not the point of heterogeneity just referred to). For H2A (lane 2), the 148 b.p. protected fragment maps the 3' terminus of the transcript derived from the gene equivalent to the H2A in p7AT. Other fragments include the undigested probe (205 b.p.), a 109 b.p. protected fragment from a closely related but non-homologous H2A gene 3' end and a 90 b.p. coding-region protected fragment.

Figure 4-3: S1 mapping and primer extension analysis of p7AT gene transcripts in 5 day chick embryo RNA

5' and 3' S1 analysis of the endogenous p7AT H2A/H2B gene pair transcripts in size fractionated chick embryo RNA is shown in tracks 1-4:

1 = H2B 3'

2 = H2A 3'

3 = H2B 5'

4 = H2A 5'

M = pBR322 HpaII markers

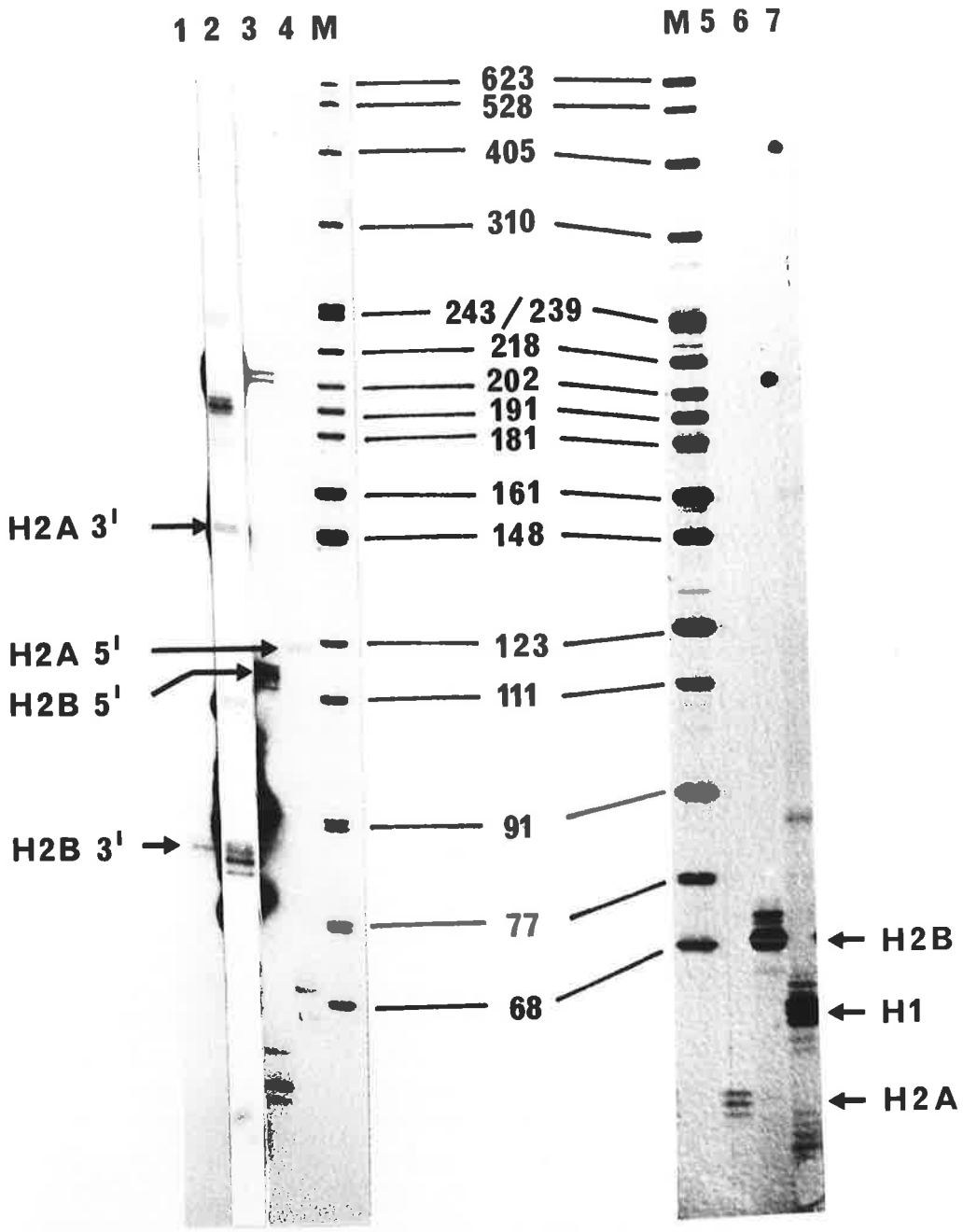
Primer extension on size fractionated chick embryo RNA using the 26 base synthetic DNA primers designed for the genes of p7AT is shown in tracks 5-7:

M = pBR322 HpaII markers

5 = H2A primer

6 = H2B primer

7 = H1 primer


1 2 3 4 M
M 5 6 7
623
528
405
310
243 / 239
218
202
191
181
161
148
123
111
91
77
68
H2A 3'
H2A 5'
H2B 5'
H2B 3'
H2B
H1
H2A

For H2B (lane 1), only one major band of 87 b.p. is seen and this maps the 3' terminus at the expected position. The predicted coding region fragment is not seen for H2B, but the hybrid would only be 26 bases long and therefore unstable in the S1 digestion conditions used (Grosschedl and Birnstiel, 1980b). Because the 3' dyad symmetry element is involved in the generation of correct 3' termini, the mapping of p7AT H2A and H2B 3' termini to the ACCCA adjacent to the stem of the hairpin loop (see figure 3-2) is expected and confirmed.

(ii) 5' Termini

The H2B-specific primer shows a major extension product of 68 bases, consistent with a cap site at -57 in figure 3-2 (figure 4-3, lane 6). This is 28 b.p. downstream from the H2B TATA box and therefore maps in the expected position. Less abundant bands seen in this analysis may be due to heterogeneous start sites in vivo or pause sites in reverse transcription in vitro. S1 analysis (figure 4-3, lane 3) also shows some heterogeneity, but protection of fragments of approximately 115 b.p. confirm the designation of a major cap site at -57. The smaller bands (58 b.p.) are coding-region protected fragments.

The results for 5' mapping of H2A genes is also shown in figure 4-3. Both S1 mapping and primer-extension analysis clearly locate the cap site to -52 in figure 3-2. That is, the 420 b.p. HpaII fragment (figure 4-1) is protected by embryonic RNA to yield an S1 product of 118 b.p. and the H2A specific primer yields an extension product of 49 bases (figure 4-3, lanes 4 and 5 respectively). The 69 b.p. S1 protected fragment in lane 4 detects the H2A coding region. These independent assays therefore place a cap site for H2A at -52. As discussed above, the primer extension assay here probably detects the H2A genes from pCH3.5E and pCH11.0E L in addition to

that contained within p7AT. However the S1 experiment is expected to be completely specific for the H2A gene of p7AT.

An extension assay using the H1-specific primer detects a product of 58 bases (figure 4-3, lane 7). This is consistent with the H1 cap site being at -37 in figure 4-2.

4-3 Characterisation of histone gene trascripts from p7AT microinjected into Xenopus oocyte nuclei

Initially, vector-free supercoiled 7 kb EcoRI insert of p7AT was injected into the Xenopus oocyte as described for optimal expression of the sea urchin h22 repeat genes (Kressman et al., 1977; Probst et al., 1979). However, side-by-side experiments in which vector-containing and vector-free DNA preparations were microinjected showed there was little difference in expression of the chicken histone genes as assayed by S1 nuclease analysis or primer extension (see below). The recombinant plasmid, containing pAT153 sequences, was therefore used routinely.

The restriction fragments outline in figure 4-1 and the synthetic primers were used to analyse transcripts from each histone gene after injection of p7AT into Xenopus oocyte nuclei. The major 3' termini of H2B and H2A transcripts and the 5' termini of H1 from the injected genes mapped to the same position as found for transcripts in chick embryo RNA (figure 4-4, lanes 2,3 and 6).

The major H2A 5' S1-protected fragment as well as the major H2A primer extension product always mapped at the site seen for chick embryo H2A 5' termini, however, other distinct minor bands are also seen. A doublet at 215 and 220 b.p. for S1 analysis (figure 4-4, lane 1) and primer extension to 135 and 140 bases (figure 4-4, lane 4) maps additional H2A cap sites to A residues four b.p. apart, at -148 and -153 in figure 3-2. These sites are 21 and 26 b.p. respectively downstream from a TATA sequence at -173.

Figure 4-4: S1 mapping and primer extension analysis of p7AT gene transcripts from micro-injected Xenopus oocytes

5' and 3' S1 analysis of the H2A/H2B genes of p7AT expressed in Xenopus oocytes is shown in tracks 1-3:

1 = H2A 5'

2 = H2B 3'

3 = H2A 3'

M = pBR322 HpaII markers

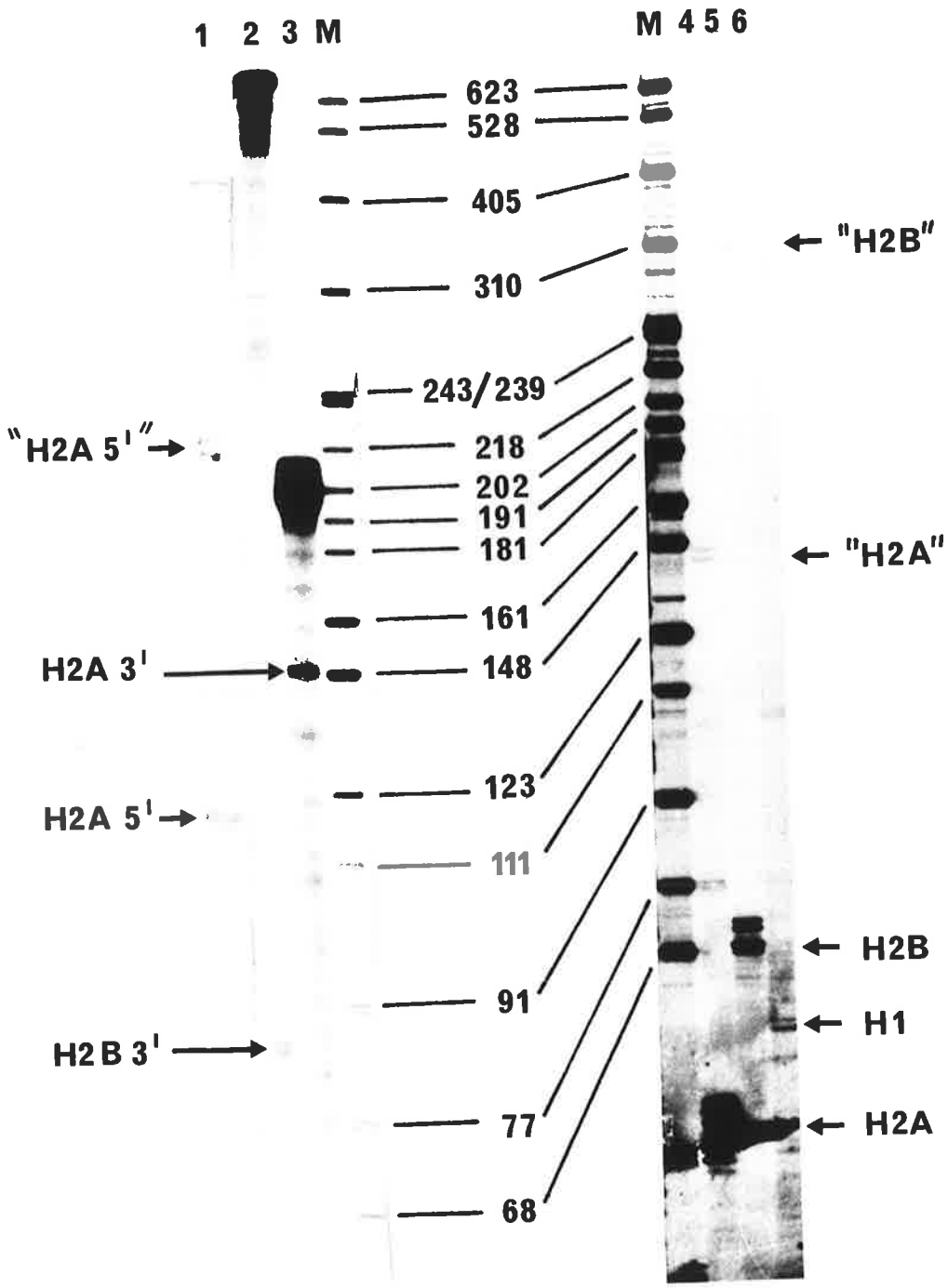
Primer extension analysis on RNA equivalent to that of one oocyte injected with p7AT using the 26 base synthetic DNA primers is shown in tracks 4-6:

M = pBR322 HpaII markers

4 = H2A primer

5 = H2B primer

6 = H1 primer


1 2 3 M
M 4 5 6
623
528
405
310
243/239
218
202
191
181
161
148
123
111
91
77
68
"H2A 5'"
H2A 3'
H2A 5'
H2B 3'
"H2B"
"H2A"
H2B
H1
H2A

An additional minor extension product to that seen in 5 day chick embryo RNA is also seen with the H2B gene-specific primer (figure 4-4, lane 5). A band of approximately 305 bases is produced in addition to the major extension product of 68 bases. No potential TATA element can be found upstream of this spurious H2B cap site.

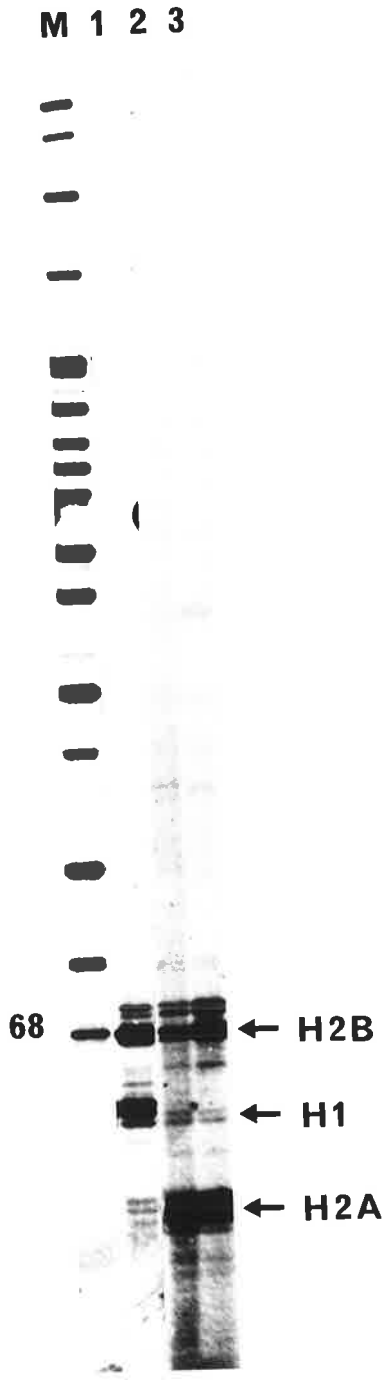
4-4 Quantitation of p7AT transcripts in chick embryo RNA and from injected Xenopus oocytes

The three gene specific synthetic primers can be used to assay the expression of all three genes in the same primer extension reaction due to their (purposely designed) different extension lengths (H2A: 49 bases; H1: 58 bases ; H2B: 68 bases). When the primers are used in excess the assay becomes quantitative (McKnight et al., 1981) and the level of each genes extension product can be directly related to one another. The data in figure 4-5 shows directly comparable (i.e. the same specific activity of individual gene-specific primers used in each extension) quantitative extension results on p7AT gene transcripts from chick embryo RNA (lane 1), RNA from oocytes injected with p7AT (lane 2) and RNA from oocytes injected with vector-free closed circular 7 kb insert from p7AT (lane 3).

The relative abundance of each message differs between chick embryo and p7AT oocyte transcripts. In the chicken, the H1 and H2B transcripts are about equally abundant and both are present at a higher level than the H2A transcript (H1 = H2B > H2A). When p7AT is microinjected into the oocyte the various levels of the three genes is quite different. The H2A mRNA is now the most abundant followed by the H2B transcript and the H1 gene is at a much lower detectable level (H2A > H2B >> H1). This pattern of expression is identical in oocyte RNA from 7 kb EcoRI closed circles, however the level of H2B extension product is slightly increased.

Figure 4-5: Quantitation of p7AT H1, H2A and H2B gene transcripts in chick embryo RNA and in micro-injected Xenopus oocytes

Quantitative primer extension analysis is shown using the three 26 base p7AT gene specific primers on size fractionated 5 day chick embryo RNA (track 1) and on RNA equivalent to that from one oocyte injected with p7AT (track 2) or 7 kb EcoRI circularised insert of p7AT (track 3). A pBR322 HpaII marker lane is shown to indicate the different extension lengths of the three primers; H2A extends to a major band at 49 bases, H1 to 58 bases and H2B to 68 bases.


M 1 2 3
68
H2B
H1
H2A

4-5 Discussion

Despite the production of additional minor mRNA transcripts from p7AT DNA templates injected into the Xenopus oocyte, all of the histone genes of p7AT are transcibed efficiently and with fidelity. Differences in the levels of each transcript compared with that in the chick, however, can be discerned. The H1 gene of p7AT is not expressed in the oocyte at anywhere near the relative level that it is in chicken embryonic RNA (approximately 10 fold relative reduction). This is consistent with the finding that the sea urchin H1 gene is not expressed well in Xenopus oocytes (Hentschel et al., 1980). Perhaps this poor level of expression of exogenous H1 genes is not surprising considering the Xenopus H1 protein (and therfore probably the mRNA) itself is made at a low rate during oogenesis and remains relatively low even into the cleavage stage of embryogenesis (Flynn and Woodland, 1980).

The difference in the relative level of H2A/H2B transcripts in the chick embryo (approximately 1 : 10) and the oocyte (approximately 3 : 1) RNA analysed is surprising and made even more so when it is realised that transcripts from three H2A genes are actually being analysed in the chick RNA. Notwithstanding the change in relative levels of the chicken histone transcripts of p7AT in the Xenopus cell the fidelity of each transcript is good. The use of minor aberrant cap sites constitutes only between 1 to 5 % of detectable transcripts (i.e. compared to the correctly initiated transcripts). These spurious initiation sites may be due to the unusually large number of gene copies injected into the oocyte ($\simeq 10^9$ gene copies), allowing some of the plasmid molecules to be assembled into a wrong chromatin formation and thus direct the initiation of these artifactual transcripts.

The accurate and efficient initiation of transcripts from the p7AT

H2A/H2B divergent histone gene pair in the Xenopus oocyte shows the potential of this system to map the promoter elements contained within the intergene region. The expression of the H1 gene should be unaffected by any *in vitro* mutation introduced into the divergent gene region and should act as an internal "expression level indicator" of any modified p7AT templates i.e. it can be used to directly compare the levels of H2A/H2B transcripts from modified p7AT templates (see next chapter). Manipulation of the p7AT plasmid and expression of such templates in the oocyte system is the next step in determining the role of the chicken divergent H2A/H2B gene structure in the control of expression of these genes.

CHAPTER 5: MAPPING OF A CHICKEN H2A/H2B DIVERGENT PAIR PROMOTER REGION

5-1 Introduction

Mapping DNA sequences responsible for the efficient and accurate initiation of gene transcription can be done in several ways (1-7). All involve selective deletion/insertion of sequences or single/multiple base transitions followed by functional testing of the mutants in suitable transcription assay systems. As shown in chapter 4 the Xenopus oocyte is a suitable functional testing system for the chicken H2A/H2B gene pair contained within p7AT. In deciding what mutational approach (deletion/insertion/base mutation) should be taken in dissecting the promoter of the divergent pair, three aspects were considered,

(1) can the H2A/H2B genes be separated from each other and still be expressed ?

(2) what is the minimum amount of sequence required for the expression of both genes at the optimum level ?

(3) from comparative sequence analysis (chapter 3) it can be predicted that any of the four CCAAT elements of the intergene region could be promoter elements. Are these shared by the two genes ?

In addition to these considerations it was important to determine if there is co-ordinate regulation of the two genes at the level of transcription. The most precise mutational analysis of the intergene region would be one in which single base changes were introduced into the conserved regions identified by the comparative sequence analysis, such as the CCAAT boxes. Another all encompassing approach would be to linker scan the intergene region (McKnight and Kingsbury, 1982). The insertion of a linker through every part of the intergene region would be unnecessary because of the identification of the conserved sequence elements. A gross approach, and one that is commonly used, would be to construct progressive

deletions from both ends of the H2A/H2B intergene region.

For reasons of simplicity, a combination of gross deletional mutagenesis and directional mutagenesis was used in the analysis of the H2A/H2B promoter region.

5-2 Expression of separated H2A/H2B genes

The H2A and H2B genes of p7AT are each contained entirely on separate XhoI restriction fragments (figure 3-1). Moreover the intergene region of the two genes is split by a XhoI recognition sequence (figure 3-2) and the H1 gene of p7AT is unaffected by XhoI digestion. By independent removal of each of the XhoI fragments of p7AT it is possible to examine the expression of each gene in the absence of the other.

p7AT DNA was partially digested with XhoI restriction enzyme. This DNA was then passed through a 1% low melting point agarose gel to purify the 8.9 kb (p7AT minus the H2B gene) and 10.0 kb (p7AT minus the H2A gene) partially digested fragments away from the 1.8 kb H2B XhoI band and 0.7 kb H2A XhoI band. The resolution of high molecular weight DNA is poor in this type of agarose gel system, so no attempt was made to get the single gene deletion bands away from linear p7AT (10.7 kb) or the fully XhoI digested band (8.2 kb). This purified DNA was then ligated to reform circular plasmid molecules and transformed into bacterial cells for propagation. It was expected that p7AT parental, p7AT-H2A (minus the H2A gene), p7AT-H2B (minus the H2B gene) and p7AT-(H2A, H2B) molecules would be obtained. Upon miniscreening of antibiotic resistant bacterial colonies obtained after transformation all of these types of molecules were indeed found (figure 5-1). The two single gene deletion constructs leave one CCAAT box present in p7AT-H2A (containing H1 and H2B genes), and three in p7AT-H2B (containing H1 and H2A genes).

Figure 5-1: Miniscreening of potential p7AT partial XhoI fragment deletion plasmids

Plasmid DNA extracted from bacteria transformed with a partial XhoI digestion of p7AT were digested with XhoI and electrophoresed on a horizontal 1% agarose gel. Four types of plasmid molecules were expected to result from the partial digestion and transformation of p7AT:

(1) no deletion = p7AT

(2) deletion of the 0.7 kb XhoI fragment = p7AT-H2A

(3) deletion of the 1.85 kb XhoI fragment = p7AT-H2B

(4) deletion of both XhoI fragments = p7AT-(H2A, H2B)

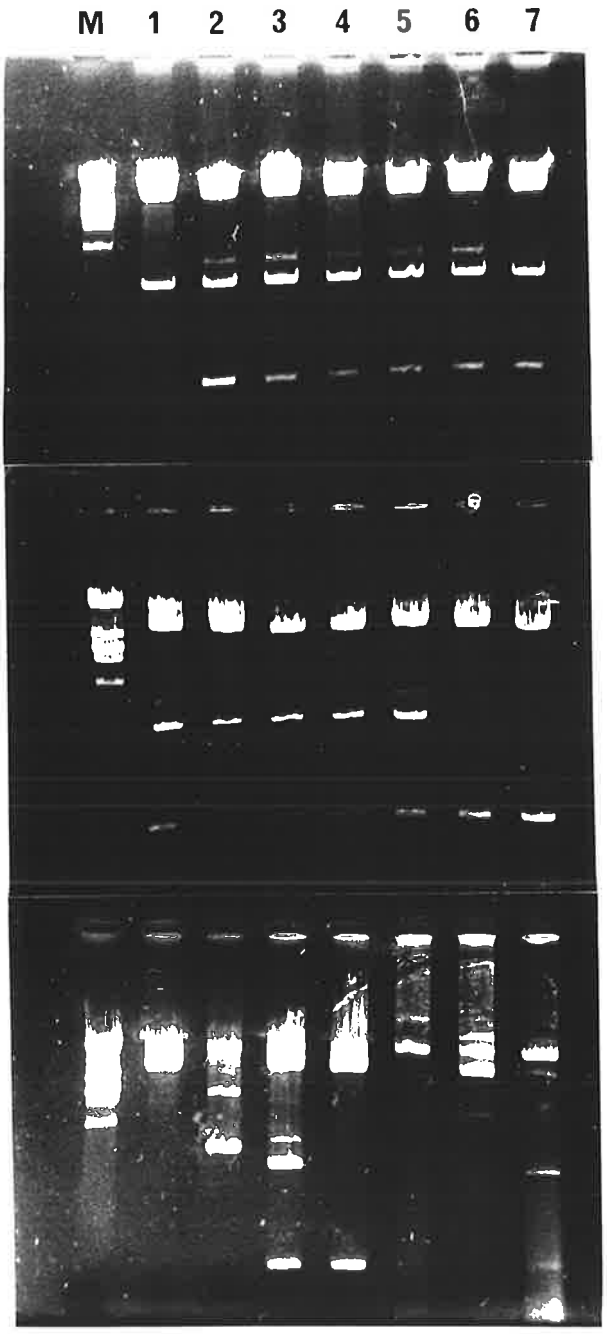
The photograph of the miniscreen gels show examples of all these types of molecules along side lambda EcoRI marker tracks:

p7AT = upper gel tracks 2 to 7
middle gel tracks 1, 3, 4 and 5
lower gel tracks 3 and 7

p7AT-H2A = upper gel track 1
middle gel track 2
lower gel track 2

p7AT-H2B = middle gel tracks 6 and 7
lower gel track 4

p7AT-(H2A, H2B) = lower gel tracks 1 and 5


M 1 2 3 4 5 6 7

Quantitative primer extension analysis on RNA extracted from Xenopus oocytes injected with either p7AT, p7AT-H2A or p7AT-H2B is shown in figure 5-2. The expression of the H1 gene in these constructs is not affected by the gross deletions and acts as the internal control in the expression experiments. The data in figure 5-2 shows that the level of the H1 primer extension product is approximately the same for all constructs indicating a consistent microinjection technique. Thus a direct comparison between H2A and H2B primer extension bands can be made on this gel. The H2A and H2B genes are expressed when separated at the intergenic XhoI site, but are trascribed at much lower levels as independent entities compared with their expression in a divergent pair configuration.

The H2A gene decreases its expression by approximately three fold whereas the H2B gene decreases its transcription rate by about 10-fold when the genes are isolated from one another. This may reflect the number of CCAAT boxes present in each of the separated promoter regions; three for H2A and one for H2B. This experiment demonstrates that in this transcription system, the two genes have overlapping promoter elements.

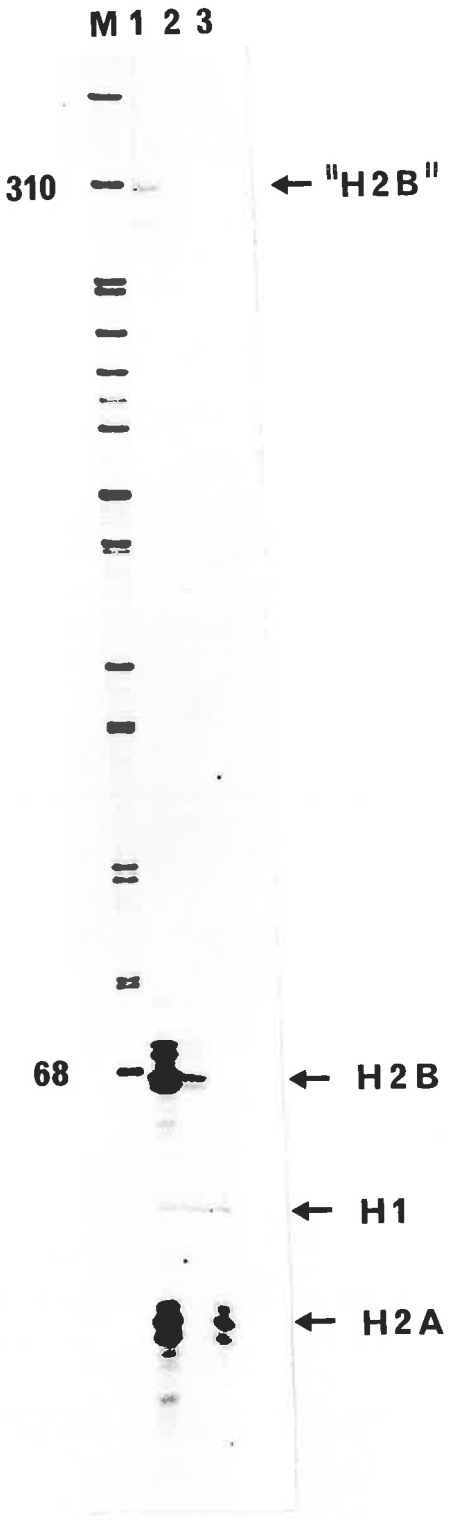
5-3 Expression from a H2A/H2B intergenic restriction fragment

In an attempt to delineate the minimum sequence requirement for promotion of the divergent transcripts, and to exclude the possibility of the protein coding regions (which are also well conserved, chapter 3) effecting transcription, I decided to take an intergenic restriction fragment from p7AT and clone it into the polylinker cloning sites of an M13 vector. The ability of the intergene region, plus a small amount of coding region, to direct transcription could then be tested in the Xenopus oocyte.

A practical requirement of such a restriction fragment is that it must contain the priming sites complementary to the 26 base gene specific primers for H2A and H2B to allow transcriptional analysis to be performed.

Figure 5-2: Quantitative primer extension analysis of separated H2A/H2B gene constructs of p7AT injected into Xenopus oocytes

The photograph shows primer extension on oocyte RNA from p7AT (track 1), p7AT-H2A (track 2) and p7AT-H2B (track 3) injected oocytes using all three of the histone gene specific primers. The identity of each group of primer extension products is indicated on the figure. A pBR322 HpaII marker track is included next to the primer extension tracks.


M 1 2 3
310
"H2B"
68
H2B
H1
H2A

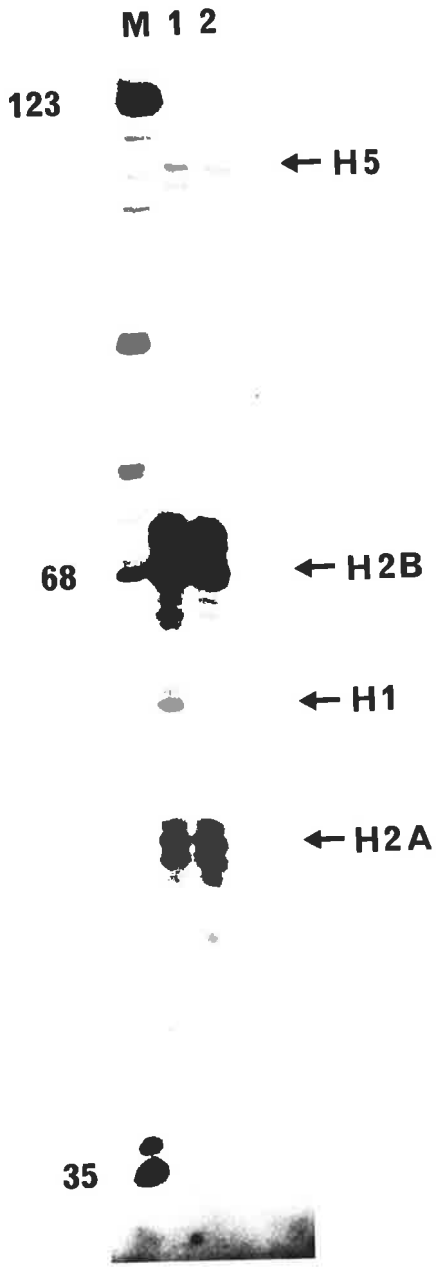
The isolation of the intergene fragment in an M13 vector removes most of the coding regions and also the normal 3' end of each transcript. It was already known that the 3' nontranslated region of histone genes is required for 3' end formation (1-6-3(iv)) but it was not known what effect the removal of the 3' conserved symmetry element would have on the stability, and therefore the level of histone transcripts in the ooctye. Because the intergene region was to be introduced into a polylinker cloning site it was possible to introduce other fragments on either side of it easily. If necessary the histone 3' dyad symmetry could be introduced either side of the M13-intergene region clone. However, this was found to be unnecessary. As judged by primer extension assays, histone transcripts without their 3' end are found at levels comparable to transcripts of the entire genes and it is likely that the truncated RNA is stable in the oocyte (see below; figure 5-3).

The restriction fragment spanning the p7AT H2A/H2B intergene region chosen for analysis was a 420 b.p. HpaII fragment (figure 3-2). This was cloned into the AccI site of M13mp9 and sequenced to confirm the insert and determine the orientation. This clone contains 68 b.p. of the H2A coding region and 13 b.p. of the H2B coding region and the HpaII restriction site in the H2B gene is immediately adjacent to the H2B synthetic primer site.

When this intergene constuct is injected into the Xenopus oocyte and compared by quantitative primer extension to transcripts from p7AT injected oocytes the H1 gene cannot be used as the internal expression control. To circumvent this problem a control gene was coinjected in equal amounts with each of the above constructs and the level of this genes expression used as an internal control. The expression of the H5 gene has been examined by primer extension in the Xenopus oocyte (Wigley et al., 1985) and gives an

Figure 5-3: Quantitative primer extension analysis of p7AT/pH5 and M13 HpaII intergene clone/pH5 constructs injected into Xenopus oocytes

RNA extracted fom oocytes that had been co-injected with either p7AT/pH5 (track 1) or the 420 base HpaII H2A/H2B intergene region M13 clone/pH5 (track 2) DNA templates were analysed by primer extension in the presence of excess H1, H2A, H2B and H5 synthetic primers. The primer extension products of each 26-mer is indicated on the figure and a pBR322 HpaII marker track is included for size determination.


M 1 2
123
H5
68
H2B
H1
H2A
35

extension of 111 bases using a gene specific 26 base primer. This H5 gene was therefore used as the internal expression level indicator.

Primer extension on RNA extracted fom oocytes injected with p7AT/pH5 and M13 HpaII intergene clone/pH5 is shown in figure 5-3. The extensions are almost quantitatively identical for the two injection series. Obviously the H1 gene is not detected in the M13 HpaII intergene clone injected oocytes. The 305 base spurious H2B primer extension transcript (discussed in section 4-3, see figure 4-4) is also absent when the intergene fragment is expressed in isolation; somehow the aberrant promoter must be rendered inactive in this fragment. This result indicates that the majority of the H2A/H2B coding regions are not required for the promotion of transcription of the H2A/H2B mRNA's in the oocyte.

5-4 Effect of the removal of an intergenic CCAAT element on expression of the H2A/H2B genes

The demonstration that the intergene region is the promoter of the divergent H2A/H2B genes, and that separation of the genes in the intergene region effects both transcripts, does not determine whether the two genes share any of the conserved elements of the intergene region (chapter 3). It was decided to initiate a site-directed mutagenesis study of the H2A/H2B intergene region to see if the genes not only overlap their promoter elements but actually share them. The CAT1 box (Table 3-1) was chosen as an element to be removed from the intergene region as it was already known that the gross XhoI deletion construct p7AT-H2B, which lacked this CCAAT element, had a reduced H2A gene transcription rate. It seemed likely that this mutation would cause a detectable phenotypic change in H2A transcription when removed. It was also possible that the H2B gene transcription efficiency would be affected if there was a common usage of this promoter element.

Deletion of short segments of DNA is facilitated by the use of single stranded M13 bacteriophage vectors (Adelman et al., 1983; Chan and Smith, 1984). The procedure involves hybridisation of a short oligodeoxyribonucleotide to sequences that flank either side of the specific sequence to be deleted. The "deletion primer" is incorporated into heteroduplex DNA by Klenow DNA polymerase synthesis on a suitable M13 single stranded template. The heteroduplex DNA, when transformed into bacterial cells, gives rise to mutant recombinant phage by utilizing the repair mechanisms of the bacterial cell. Mutant bacteriophage are identified by hybridization of the deletion primer ^{32}P-oligodeoxyribonucleotide to 'phage DNA bound to nitrocellulose filters. Appropriate cycles of washing at successively higher temperatures and autoradiography, clearly identifies mutant phage which contain inserts colinear with the deletion primer sequence.

To commence the site-directed deletion of the CAT 1 box sequence a 5 kb BamHI fragment of p7AT was excised and cloned into the BamHI site of the M13mp8 phage vector (see restriction map in figure 3-1). This fragment includes the 375 b.p. BamHI-EcoRI fragment of the pAT153 vector sequence and still contains the complete transcription units of the H1, H2A and H2B genes. Both orientations of the 5 kb insert were obtained in the M13mp8 vector but the orientation containing the coding strand of H2B and H1 genes was chosen for mutagenesis. This clone, named M135kb.B-parental was chosen for mutagenesis as this allowed the H2B and H1 (for H1 mutagenesis work not described in this thesis) 26 base synthetic primers (figure 4-1 and 4-2) to be used as sequencing primers to verify potential deletion mutants.

A 22 base synthetic DNA deletion primer homologous for 11 bases either side of the CAT1 sequence "CCAATG" was used for the deletion mutagenesis. The sequence of this primer and the seqence of the expected CAT1 deletion

mutant M135kb.BΔCAT1 is shown in figure 5-4. The sequence of the M135kb.B-parental genome is compared to that of M135kb.BΔCAT1 using the H2B synthetic 26 base primer as a sequencing primer in figure 5-5. The single CAT1 deletion identified is not a clean removal of the CCAATG sequence, but includes an additional base substitution as shown in the actual deletion sequence of figure 5-4. It is not fully understood how this extra mutation has occured but a similar base mutation has been seen in other site-directed mutagenesis work undertaken on the H1 gene of M135kb.B-parental (B. Younghusband, personal communication). This additional base change was thought to be incidental to the purpose of the experiment and so M135kb.B ΔCAT1 was used for subsequent expression work.

Functional testing of M135kb.B-parental and M135kb.BΔCAT1 phage R.F. DNA in the Xenopus oocyte is shown in figure 5-6. Unexpectedly it appears as if the CAT1 deletion has made essentially no impact on the expression of either the H2A or H2B genes (compare parental and ΔCAT1 in figure 5-6; note that the level of H1 control transcripts is virtually identical in these tracks). The minor H2A extension products from the pseudo TATA box at -173 (figure 3-2) and giving rise to the 135-140 base primer extension products are increased when CAT1 is deleted. The major H2A extension transcript(s) however remain unaffected. The H2B aberrant primer extension transcript of 305 bases (4-3) is reduced to 299 bases in the M135kb.BΔCAT1 injections as this spurious transcript encompasses the CAT1 element, and is therefore reduced by 6 bases when CAT1 is deleted.

This absence of a phenotypic effect of the CAT1 box deletion is unexpected and is made even more so given the extaordinary sequence conservation of the element in all six of the H2A/H2B divergent pairs (Table 3-1). One possible answer to the anomaly between the expression of

Figure 5-4: Deletion of the CAT1 box of the H2A/H2B intergene region

The sequence of the 22-base synthetic oligonucleotide designed to delete the CAT1 box of the H2A/H2B intergene region is shown against the corresponding sequence of the single stranded M135kb.B-parental DNA used in the mutagenesis procedure. The sequence is numbered according to that presented in figure 3-2 for the intergene region. The primer is homologous for 11 bases either side of the sequence to be deleted. The sequence expected to be obtained after the deletion procedure and that actually obtained (figure 5-5) is also presented. Note the additional one base transition between the parental sequence and the deletion sequence (shown with asterisk).

Sequence surrounding the CAT1 box:

```
      XhoI
-150 ____-140        -130      -120      -110
   TTCTCGAGTTCCGA CCAATG AAAGAGTGCGAAAGGAATGCT

     AGCTCAAGGCT        TTTCTCACGCT                    CAT1 deletion
   3'                              5'                    primer
```

Expected deletion sequence:

```
   TTCTCGAGTTCCGA DELETED AAAGAGTGCGAAAGGAATGCT

   AAGAGCTCAAGGCT          TTTCTCACGCTTTCCTTACGA       strand read on
3'                                            5'      sequencing gel
```

Actual deletion sequence:

```
   TTCTCGAGTTCCGA DELETED AAAAAGTGCGAAAGGAATGCT

   AAGAGCTCAAGGCT          TTTTTCACGCTTTCCTTACGA       strand read on
3'                            *               5'      sequencing gel
```

Figure 5-5: Sequence analysis of the M135kb.BΔCAT1 box deletion

The H2B gene specific 26 base primer was used to sequence across the H2A/H2B intergene region of the M135kb.B-parental genome and the CAT1 box deleted genome. The sequences are listed beside the gel to indicate the deletion and the additional one base transition of M135kb.BΔCAT1 box (shown by the asterisk).

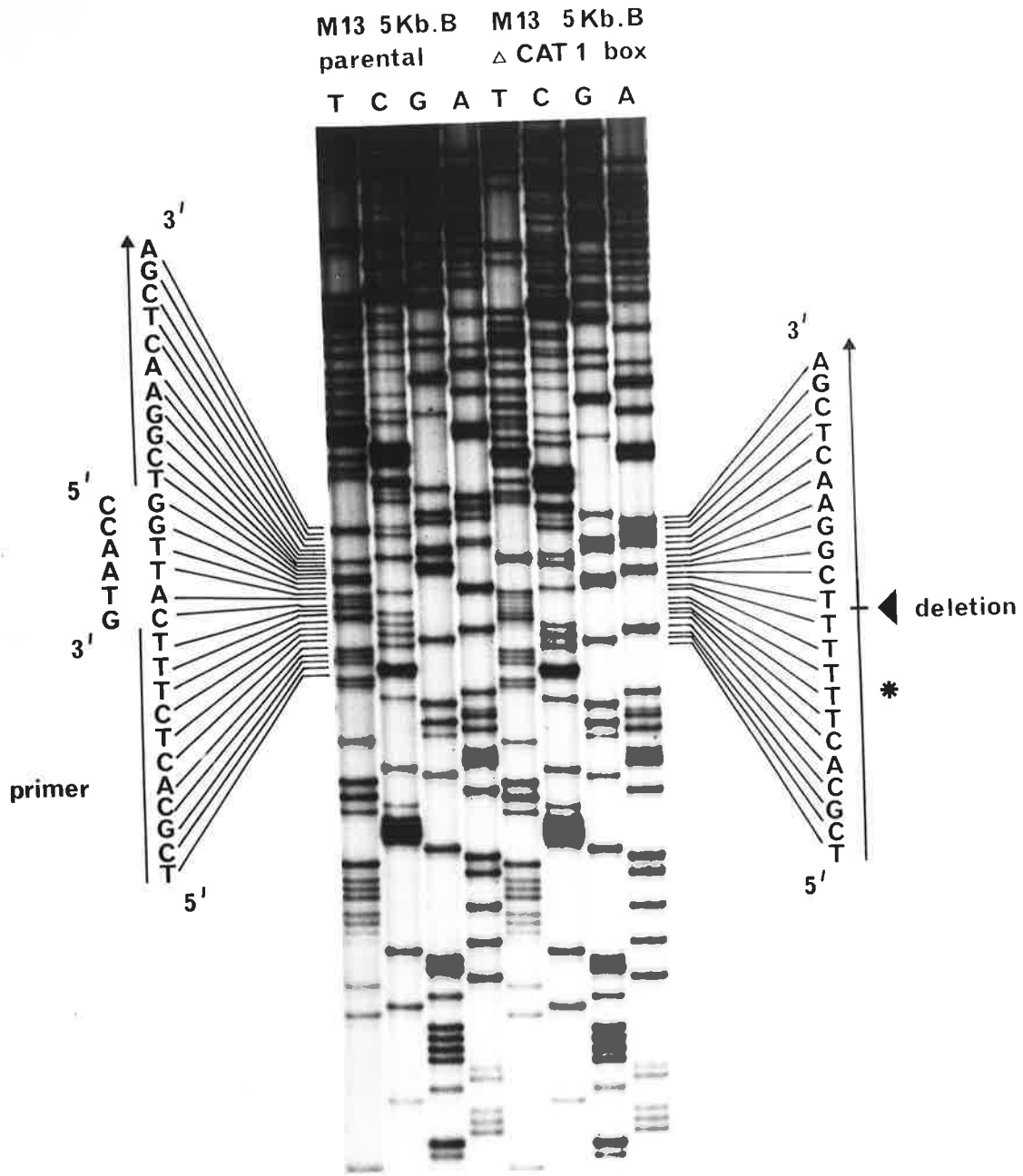
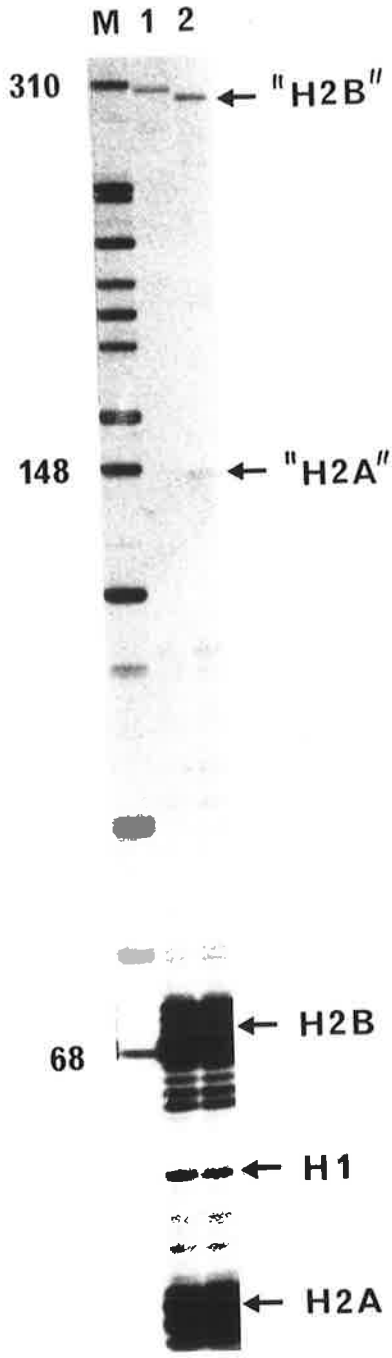

M13 5Kb.B
parental
M13 5Kb.B
△ CAT 1 box
T C G A T C G A
3'
AGCTCAAGGCTGGTTACTTTTCTCACGCT
5'
5' CCAATG
3'
primer
3'
AGCTCAAGGCTTTTTTTCACGCT
5'
deletion
*

Figure 5-6: Quantitative primer extension analysis of RNA from M135kb.B-parental and M135kb.BΔCAT1 injected Xenopus oocytes

The figure shows primer extension on M135kb.B-parental (track 1) and M135kb.BΔCAT1 (track 2) injected oocyte RNA and a pBR322 HpaII marker lane. The bands corresponding to each gene of the injected templates are labelled.


M 1 2
310
"H2B"
148
"H2A"
68
H2B
H1
H2A

p7AT-H2B (three fold decrease in H2A expression) and M135kb.BΔCAT1 (no effect on H2A or H2B expression) which both lack the CAT1 box is that p7AT-H2B also lacks the conserved 13 b.p. H2B box (Table 3-1). This element and its possible involvement in the lowering of H2A gene expression will be discussed in chapter 6.

Other feasible but unlikely explanations for the lack of effect of the CAT1 box mutation include the possibility that the resultant mutant sequence may be able to substitute for a CCAAT element (new sequence CCGAAA) and thus represent a null mutation. Also, there may be synergistic effects of the remaining CCAAT elements that compensate for the removal of the CAT1 box. A simplistic explanation for the normal expression of the M135kb.BΔCAT1 recombinant is that it could be an artifact of the oocyte system.

5-5 Discussion

Mapping the DNA sequence of the H2A/H2B genes of p7AT responsible for the correct initiation of mRNA transcripts from these two genes in the Xenopus oocyte transcription system, has shown that a fragment of 420 b.p. is sufficient for promotion of transcripts from both genes. Moreover splitting the two genes within this 420 b.p. region reduces expression of each transcript demonstrating that the divergent genes overlap promoter elements. This promoter arrangement has an obvious potential to co-ordinate expression of the two genes at the transcriptional level leading to balanced production of each protein within the cell.

Despite the failure to demonstrate a transcriptional effect from deletion of the CAT1 box, it is still probable that this element has a significant biological role *in vivo* given the conserved nature of this sequence in six divergent gene pairs and two isolated H2A and one single H2B gene. Sequence comparison data makes it most likely that at least some

of the three remaining CCAAT elements not investigated in this study could be used to promote both of the divergent transcription units.

CHAPTER 6: ANALYSIS OF THE 13 BASE H2B GENE-SPECIFIC ELEMENT

6-1 Introduction

Sequence comparison of two chicken H2B genes by Harvey et al. (1982) showed a highly conserved 13 b.p. sequence, termed the H2B-box, located in the 5' "promoter" region of two chicken H2B genes. It was also found to be present in H2B genes from other organisms. The sequence of seven chicken H2B 5' regions is summarized in Table 3-1 and a consensus 13 b.p. H2B-box sequence is presented. This element is extaordinarily well conserved in sequence and position in the chicken H2B genes, always being 6 b.p. from the H2B TATA boxes. A summary of potentially analogous sequences to the chicken H2B-box at similar positions in H2B genes of several different organisms are presented in Table 6-1, and are compared to the chicken H2B-box consensus sequence.

Ubiquitous 5' histone gene-specific sequences have also been observed for the H1 gene (Coles and Wells, 1985) and the H4 gene (Clerc et al., 1983). In the latter case the sequence was shown to be a functional component for the efficient transcription of the H4 gene in the Xenopus oocyte transcription system. In the case of the H1 gene, the specific element is about 70 b.p. from the genes TATA box (-149 to -155 of figure 4-2). The H4 gene-specific element is approximately 20 b.p. upstream from the TATA box, and is therefore more like the H2B-box in having a close association with the TATA box.

The presence of the 13 b.p. H2B element in the divergent gene pairs gives it the potential to act on the H2A genes that accompany the H2B genes. The 5' sequence of the isolated H2A genes shown in Table 3-1 lack this element indicating that it is probably specific to the H2B genes of the H2A/H2B pairs. The close association of the H2B-box with the TATA box suggests that it may be an element involved in the binding of a trans-

Table 6-1: H2B gene specific sequence element across species and comparison to the Drosophila heat-shock box consensus

Sequences potentially equivalent to the chicken H2B-box in a variety of species are tabulated and compared to the Drosophila heat-shock box consensus. Sequences have been aligned around the central triple T bases in an attempt to highlight maximal homology. The chicken consensus was taken from Table 3-1, the HSB consensus from Pelham (1982) and Davidson et al. (1983), the Xenopus sequence from Moorman et al. (1982), the mouse sequence from Sittman et al. (1983), the human sequence from Zhong et al. (1983), the Sea Urchin sequences from Sures et al. (1978) and Busslinger et al. (1980), the yeast sequence from Wallis et al. (1982) and the Drosophila sequence from Goldberg (1979).

Species	H2B box sequence		
Xenopus laevis	CTTATTTGCATGG	... 6 bp ...	TATA box
Mouse	CTTGACGTTTGCAGA	... 24 bp ...	
Human	CTTATTTGCATAAG	... 6 bp ...	
Sea Urchin Sp/2	CTCATTTGCATAC	... 28 bp ...	
H19	CTCATTTGCATAC	... 31 bp ...	
H22	CTCAACATTTGCATAC	... 25 bp ...	
Yeast H2B 1	CTCTTTCCGCAT	... ? bp ...	
H2B 2	CTAATTTTGTTTAT	... ? bp ...	
Drosophila	CTGAAT-TGACTG	... 12 bp ...	
Chicken consensus	CTGATTTGCATAG/C	... 6 bp ...	
Drosophila HSB consensus	CTNGAATNTTCTAGA	... 10 bp ...	

acting factor capable of modulating H2B transcription perhaps independently of the H2A gene and indeed all other histone genes. The report of H2B as a heat-shock protein in Drosophila (Sanders, 1981) prompted a close look at the possible relationship between the H2B box and the heat-shock box as discussed below.

The heat-shock response in eukaryotes is characterised by the rapid synthesis of a small number of specific proteins (heat-shock proteins: hsp's) and a repression of the synthesis of most other proteins (Schlesinger et al., 1982). The mechanism by which heat-shock genes are activated has been shown to have both transcriptional and translational components (Ashburner and Bonner, 1979; Bienz and Gurdon, 1982; Di Domenico et al., 1982; Ballinger and Pardue, 1983). Heat induction of hsp genes at the transcriptional level has been best characterised using the Drosophila hsp 70 and hsp 82 genes (Pelham, 1982; Wu, 1984a and b). Pelham identified an upstream element, termed a heat-shock box (HSB), present 5' to hsp genes that has been shown by Wu to bind a heat-shock activator protein (HAP) which acts as a positive regulator of the genes.

In separate experiments, the H2B protein was identified as an hsp in Drosophila (Sanders, 1981). Further analysis suggested that there is a rapid transcriptional response mediating the increased synthesis of H2B during heat-shock (Tanguay et al., 1983) but post-translational control has also been shown to be important (Farrell-Towt and Sanders, 1984).

Table 6-1 compares the H2B gene-specific sequence of diverse species and the chicken consensus H2B-box to a consensus HSB from Drosophila (Pelham, 1982; Davidson et al., 1983). There is an homology between the two consensus sequences and also the same approximate position relative to the TATA box. The H2B-specific box is well conserved between species with the

exception of yeast and Drosophila H2B genes. The consensus HSB is derived from 8 Drosophila HSB sequences which in themselves show considerable variation. The H2B-box sequence is certainly more highly conserved than the HSB sequences listed by Pelham but this is not surprising since different genes with different heat-shock responses were considered. Looking across species it is notable that the HSB consensus sequence from Drosophila can be aligned directly with a HSB in the slime mold (Zuker et al., 1983) indicating that the Drosophila HSB consensus is functionally significant.

The combination of three independent pieces of information, namely that histone H2B is a heat-shock protein in Drosophila, that a consensus sequence for the heat-shock promoter can be defined and that H2B genes have a highly conserved element with homology to the heat-shock promoter sequence suggested that the chicken 13 b.p. H2B-box may be functionally equivalent to the HSB. Although it is not known whether the H2B protein in the chicken or any other species besides Drosophila is a heat-shock protein it is most likely to be given the conserved nature of the response to thermal stress (Schlesinger et al., 1982).

The reason that the Drosophila H2B-box shown in Table 6-1 is not homologous to the consensus Drosophila HSB could have two possible explanations. First, the heat-shock genes used to compile the Drosophila HSB consensus are those that give rise to the major hsp's of the cell, whereas the stimulation of H2B production by heat-shock is only about two fold (Camato et al., 1982). Different transcriptional stimulation of hsp genes could be mediated by different binding affinities of the respective HSB for the trans-acting HAP i.e. the weaker the homology to the consensus the lower the binding efficiency and the less tanscriptional stimulation. Alternatively it has been suggested that a different H2B gene from the one present in the tandem histone repeat that has been sequenced could be

transcribed during heat shock (Farrell-Towt and Sanders, 1984).

Experiments to test the ability of the H2B-box to function as a heat-shock box were undertaken utilizing the Xenopus oocyte transcription system as it had been shown to be a suitable system to analyse the heat-shock phenomenon i.e. the Drosophila hsp70 gene is only transcribed during heat-shock treatment (Voellmy and Rungger, 1982).

6-2 Effect of heat-shock on transcription of a chicken H2B gene in the Xenopus oocyte

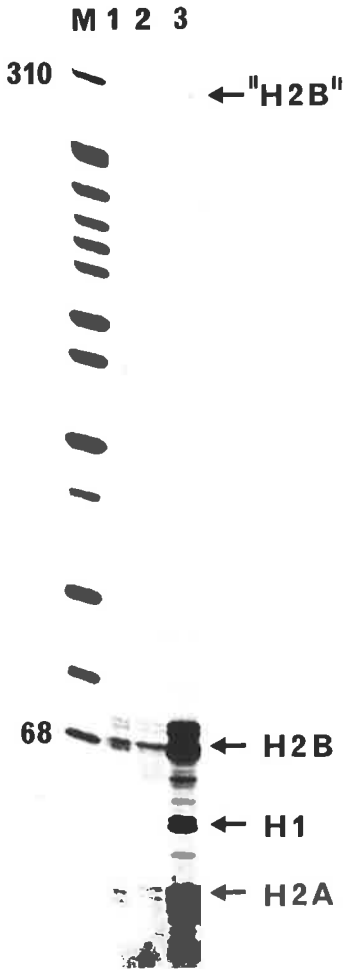
The H2B gene utilized in this study was the one whose expression had already been established in the Xenopus oocyte i.e. the H2B gene of the p7AT divergent H2A/H2B gene pair. With the H2B-box sequence from this gene alone it is possible to align 12 out of 14 bases with the Drosophila HSB allowing two deletions and two insertions:-

Drosophila HSB	CTNGAATNTTCTAGA
p7AT H2B-box	CTC-A-TTT C TAGA G A

p7AT DNA was injected into thirty Xenopus oocyte nuclei and incubated at 20 C for 6 hours to allow nuclear protein assembly on the injected DNA. Ten such oocytes were incubated for another 2 hours at 37 C (heat-shock), another ten continued incubation at 20 C for 2 hours and the third set of ten allowed to incubate for a total of 24 hours at 20 C as in the normal injection procedure. RNA was extracted from the first two pools of oocytes at 8 hours and the final pool at 24 hours. The level of transcription initiation for H1, H2A and H2B genes was estimated by quantitative primer extension analysis using the 26-mer synthetic primers. Two exposures of the data are presented in figure 6-1. The shorter exposure shows the major cap sites of each gene after the usual 24 hour incubation. In the longer exposure, a comparison of the two 8 hour incubated oocyte tracks shows that

Figure 6-1: Quantitative primer extension on control and heat shocked oocytes injected with p7AT

Two exposure times are shown of the one primer extension gel used to analyse the RNA from oocytes used in the heat-shock experiment. The exposure on the left shows the normal level of the histone gene primer extension products after 24 hours at normal temperature (track 3) next to pBR322 HpaII markers, the bands corresponding to each gene of p7AT are indicated on the figure. A longer exposure of this primer extension analysis is shown on the right panel to allow the 8 hour injected oocyte RNA samples to become visible. Track 1 shows the effect of heat-shock on the expression of the H2B gene compared to that in track 2 which contains RNA from control oocytes kept at normal temperature.


M 1 2 3
310
"H2B"
68
H2B
H1
H2A


M 1 2 3
310
"H2B"
"H2A"
68
H2B
H1
H2A

H1 and H2A transcripts appear unaltered by the heat shock treatment. On the other hand, there is a clear alteration in cap site usage for H2B transcripts such that the lower band of the H2B primer extension doublet is markedly increased (this effect is not due to the primer extension protocol).

There is about a five-fold increase in the use of the cap site giving the lower band of the H2B primer extension doublet and although this is small, it is consistent with the small (two-fold increase) in H2B expression during heat-shock treatment of Drosophila Kc cells (Camato et al., 1982). It appears that a heat-shock response for H2B genes injected into Xenopus oocytes may be brought about by a change in cap site usage. Whatever the mechanism, it was thought likely that it was mediated via the H2B-box given its similarity to the Drosophila HSB. Alternatively the data of figure 6-1 could have been a relatively non-specific effect of the two hour heat treatment on Xenopus oocyte functions.

6-3 Analysis of a H2B-box deletion mutant

In conjunction with the deletion of the CAT1 box of the p7AT H2A/H2B pair (5-4) the H2B box was also deleted. The reasons for this were two fold; to see if the heat shock effect could be abolished with the removal of the H2B-box and thus establish that this sequence was responsible for the phenomenon, and also to see if this element caused the decrease in H2A gene expression in the p7AT-H2B construct (see section 5-2, figure 5-2). The procedure for deleting the 13 b.p. H2B gene-specific sequence was identical to that already outlined in 5-4. The 22 base deletion primer used to hybridize to the M135kb.B-parental single stranded phage DNA is outlined in figure 6-2. The resultant sequence after the removal of the H2B-box is also shown in figure 6-2, the deletion being named M135kb.BΔH2B-box. Figure

Figure 6-2: Deletion of the H2B-box contained within the H2A/H2B intergene region

The sequence of the 26-base synthetic oligonucleotide designed to delete the H2B-box of the H2A/H2B intergene region is shown against the corresponding sequence of the single stranded M135kb.B-parental DNA used in the mutagenesis procedure. The sequence is numbered according to that presented in figure 3-2 for the intergene region. The primer is homologous for 13 bases either side of the sequence to be deleted. The sequence expected to be obtained after the deletion procedure is also presented.

Sequence surrounding the H2B-box:

```
-130      -120      -110       -100       -90       -80
   AAAGAGTGCGAAAGGAATGCTT CTCATTTGCATAGA GGGGCTATAAATAAA

            CTTTCCTTACGAA                CCCCGATATTTAT      H2B-box
          3'                                          5' deletion primer
```

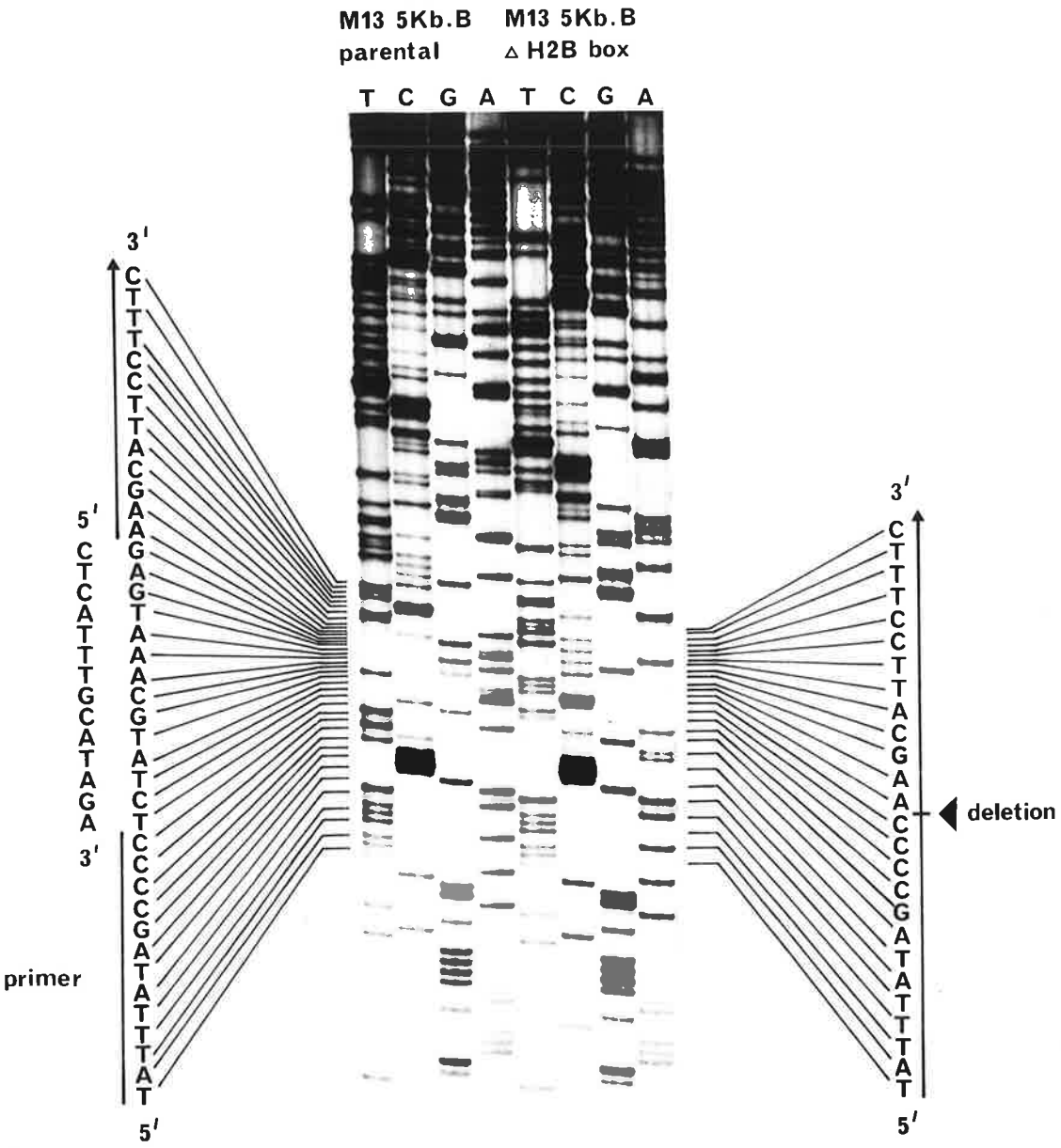
Expected deletion sequence:

```
   AAAGAGTGCGAAAGGAATGCTT      DELETED     GGGGCTATAAATAAA

   TTTCTCACGCTTTCCTTACGAA                  CCCCGATATTTATTT    strand read on
 3'                                                      5'  sequencing gel
```

Figure 6-3: Sequence analysis of the M135kb.BΔH2B-box deletion

The H2B gene specific 26 base primer was used to sequence across the H2A/H2B intergene region of the M135kb.B-parental genome and the H2B-box deleted genome. The sequences are listed beside the gel to indicate the deletion.


M13 5Kb.B
parental
M13 5Kb.B
Δ H2B box
T C G A T C G A
3'
CTTTCCTTACGAAGAGTAAACGTATCTCCCCGATATTTAT
5'
5' CTCATTTGCATAGA 3'
primer
3'
CTTTCCTTACGAACCCCCGATATTTAT
5'
deletion

6-3 compares the sequence of the M135kb.B-parental genome to that of M135kb.BΔH2B-box using the H2B 26-mer as a sequencing primer.

The first experiment using the H2B-box deletion construct was to compare the expression of the divergent H2A/H2B genes to the that of the parental genome under normal conditions. As can be seen on the primer extension assay presented in figure 6-4 the 305 base H2B aberrant transcript, spanning the divergent intergene region, is reduced by 14 bases due to the H2B-box deletion being part of this transcript. Besides this obvious change there is no immediately apparent effect on the level of either the H2A or H2B transcript produced by the removal of the H2B-box. There may be a slight relative reduction of the H2A gene transcript but nowhere near the extent seen in the p7AT-H2B construct (5-4). It is possible that the removal of both the CAT1 and H2B-box regions together in the p7AT-H2B construct contribute synergistically to the poor expression of the H2A gene. A point not previously considered is that the H2B TATA box, though not conserved in sequence to the extent that the CAT1 and H2B-box of the intergene region are, may affect the usage of the H2A transcription unit by concentrating polymerase molecules in the intergene region or even allowing the entry of polymerase on to the DNA in the opposite direction to the H2B transcription unit.

The heat-shock protocol on injected oocytes discussed above was repeated on the M135kb.B-parental and M135kb.BΔ H2B-box constucts. The results of this experiment are shown in figure 6-5. The primer extension analysis of the H1 and H2A transcripts are virtually identical in both injection series irrespective of the heat treatment. The H2B primer extension product indicates a change in cap site usage induced by the temperature shift of the oocytes regardless of the presence or absence of the 13 b.p. H2B gene specific sequence. The H2B gene resonse to thermal

Figure 6-4: Quantitative primer extension analysis of RNA from M135kb.B-parental and M135kb.BΔH2B-box injected Xenopus oocytes

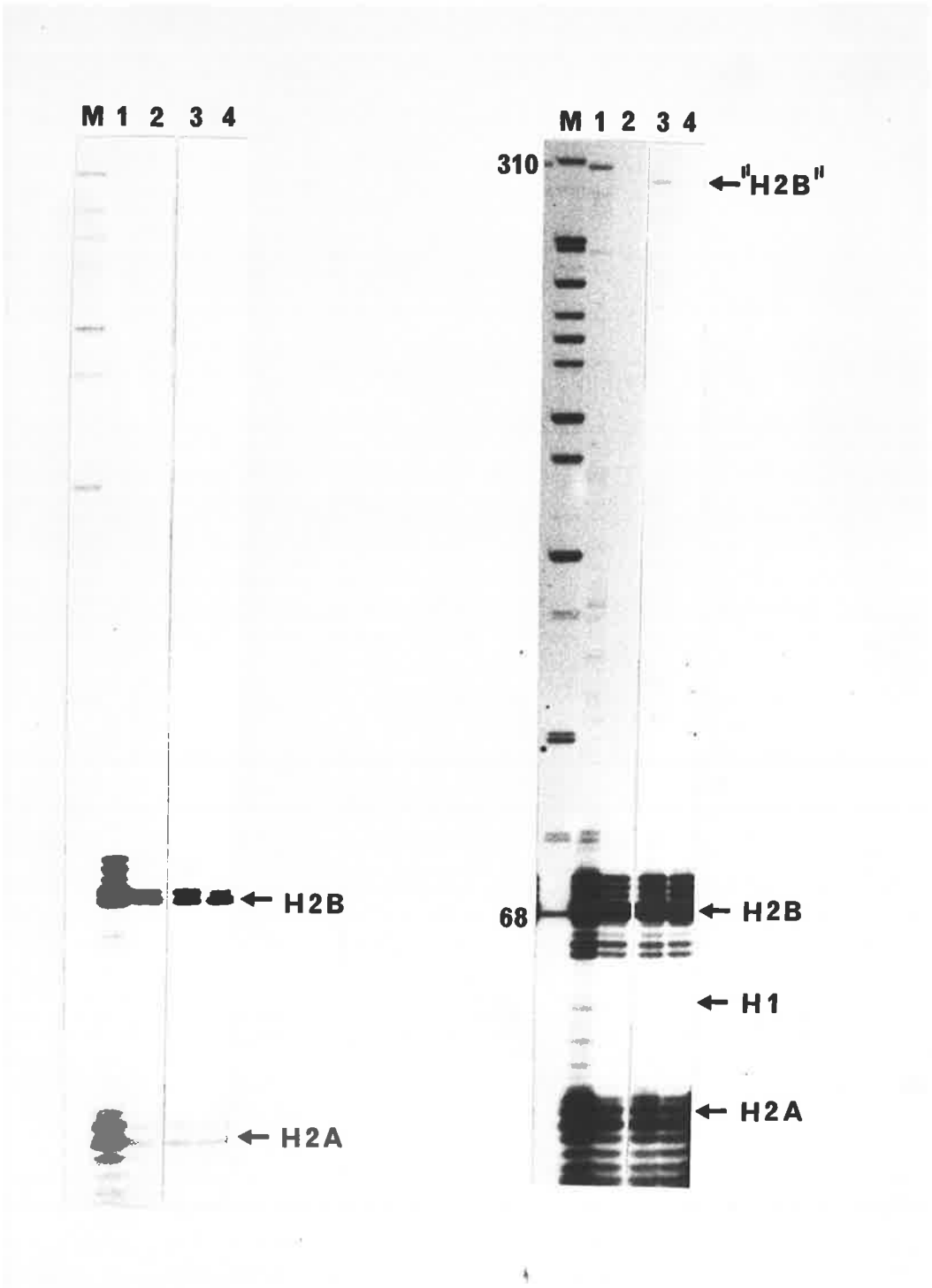
The figure shows primer extension on M135kb.B-parental (track 2) and M135kb.B Δ H2B-box (track 1) injected oocyte RNA and a pBR322 HpaII marker lane. The bands corresponding to each gene of the injected templates are labelled.


1 2 M
310
"H2B"
H2B
68
H1
H2A

Figure 6-5: Quantitative primer extension on control and heat shocked oocytes injected with M135kb.B-parental and M135kb.BΔH2B-box

Two exposures are shown of the primer extension gel used to analyse RNA from oocytes injected with the parental histone construct or the H2B-box deletion mutant. The exposure on the left shows a short exposure time of histone gene primer extension products after 8 hours incubation at normal temperature (tracks 1 and 3) or 37 ^{0}C (tracks 2 and 4) of M135kb.B-parental and M135kb.BΔH2B-box injected oocytes respectively, next to pBR322 HpaII markers. Bands corresponding to each gene are indicated on the figure. This exposure allows the change in H2B cap site usage to be seen most clearly.

A longer exposure of this primer extension analysis is shown on the right panel to allow the other gene primer extension products to become visible.


M 1 2 3 4
H2B
H2A
M 1 2 3 4
310
"H2B"
68
H2B
H1
H2A

stress may therefore be an artifactual result due to the thermal lability of some factor(s) involved in the transcription of the gene.

Given the homology of the H2B-box to the HSB consensus of Drosophila but the negative result upon experimentation, shows that care must be taken in extrapolating the function of DNA sequences at low homologies. This will be discussed in detail below.

6-4 The H2B-box has homology to a conserved 5' immunoglobulin gene sequence

The tightly conserved 13 b.p. sequence, present in the H2B genes of all species, initially had not been found in the 5' regions of other histone genes, nor any other genes transcribed by polymerase II. It was therefore thought to be an element specific to H2B genes. Later work has shown it to be present in a number of genes.

Falkner and Zachau (1984) in the course of their work on immunoglobulin genes defined a decanucleotide (dc) sequence, TNATTTGCAT, found upstream of all human and mouse immunoglobulin Kappa and lambda chain variable region genes, and within the mouse heavy-chain enhancer. An inverted form of the dc element, ATGCAAATNA, was found to occur upstream of all immunoglobulin heavy-chain variable region genes. Their computer search of the EMBL nucleic acid sequence library for related sequences at a 89% homology level found the dc element in the 72-b.p. repeat of SV40, the 5' flanking region of the chicken ovalbumin gene, in the sea urchin histone gene cluster, the dispersed repetitive R sequences, and also upstream of the genes for mouse and human class II histocompatability antigens.

Parslow et al. (1984) independently identified the same sequence elements when studying a mouse lymphocytoid pre-B cell line 70Z/3 containing a rearranged immunoglobulin kappa light chain gene. The octanucleotide sequence, ATTTGCAT, was found to lie approximately 70 b.p.

upstream from the cap site in every light chain gene examined; in the heavy genes at the corresponding location the precise inverse of the above sequence, ATGCAAAT, was discovered.

It was suggested by both groups of workers that the conserved immunoglobulin sequence block may serve as a recognition locus for factors regulating immunoglobulin gene expression in a tissue specific fashion. These equivalent 8-mer and 10-mer sequences are contained within the 13 base H2B-box (Table 6-2). At the time of Falkner's search of the EMBL data base the sequence of the two chicken H2B genes reported by Harvey et al. (1982) to identify the 13 b.p. H2B-box were not included, thus explaining their failure to identify it. Subsequent scientific correspondence purports to having identified the dc element in the homoetic gene ftz of Drosophila (see below; Tsonis, 1984).

Table 6-2 is a compilation of sequences cited by either Falkner or Parslow and by a computer search (initiated here but not comprehensive) of gene regions with homology to the dc consensus sequence. The Table lists the sequence with the identifiable homology and the relative position of the sequence to the recognised TATA box or one with suitable homology if this has not been determined. In the case of the immunoglobulin heavy chain genes where the element is inverted with respect to the TATA box, the inverse sequence is presented to facilitate comparisons.

The first points of note from this sequence compilation is that for most of the regions identified (except for those listed at the bottom of the Table) the homology extends further than the 10 base dc consensus sequence and there is a tight coupling to a TATA box of the same order of distance to that of the H2B-box and its TATA box. Indeed the homology of each box extends to something like the inter-species homology between the histone H2B genes themselves shown in Table 6-1. However, the homology is

Table 6-2: Gene systems containing the U-box sequence

Upstream sequences of a number of gene systems containing the U-box sequence, TTTGCA, are compiled and their position relative to a TATA box indicated. The sequences are aligned around the central U-box homology to show the almost invariant nature of this element.

In the case of the heavy chain immunoglobulin genes and one of the Drosophila ftz sequences, the reverse orientation of the element is presented so the sequence tabulation is consistent. The distance from a TATA box divergent from the real TATA box is also presented for the heavy chain immunoglobulin genes.

References for the immunoglobulin data are in Falkner and Zachau (1984) or Parslow et al. (1984). The U-box containing sequences in the chicken feather keratin gene A, chick H5 and H1 genes, the Herpes Simplex Virus Thymidine Kinase gene and the human B-globin gene were found by computer analysis. The Drosophila ftz gene U-box sequences were from Tsonis (1984), the human metallothionein IIa gene from Karin and Richards (1982), the consensus metallothionein heavy metal binding sequences were from Stuart et al. (1984) and Carter et al. (1984).

Gene system	distance from TATA box	Sequence	distance from TATA box
Immunoglobulin light chain			
mouse Kappa			
70Z/3		CTTGCGATTTGCATAT	17 bp
K2		CTGATTTGCATG	10 bp ?
M173B		CTTAATGATTTGCATAT	28 bp
M41		CTTAATAATTTGCATAC	26 bp
M167		CAATGGATTTGCATAA	> 10 bp ?
T1		CAATGATTTGCATGC	28 bp
T2		GTATGATTTGCATAA	34 bp
K21C		CTGATTTGCATAT	10 bp
M11		CTCCTTTGCATAG	> 20 bp
human Kappa			
HK101		CTCATTTGCATGT	29 bp
Vd		CTGCCGATTTGCTT	? bp
h101		CTCATTTGCATG	31 bp
h122		CTGATTTGCCTG	31 bp
h100		AATTTTCATT	? bp
murine Lambda			
I and II		GTAATTTGCATT	28 bp
Immunoglobulin heavy chain			
M141	14 bp	CTGATTTGCATTG	18 bp
101	14 bp	CTGCTTTGCATTG	17 bp
108A	15 bp	GTAATTTGCATAT	21 bp
HG3	17 bp	GTTATTTGCATAT	11 bp
167	14 bp	CTAATTTGCATGT	8 bp
105, 111, 104	15 bp	GTGATTTGCATAT	? bp
HLA-DR	23 bp	TTGATTTGCATTT	14 bp
Immunoglobulin heavy chain enhancer		GTAATTTGCAT	
Immunoglobulin dc element		TNATTTGCAT	
SV40 enhancer		CATGCTTTGCATAC	
Chicken H2B-box consensus		CTGATTTGCATAG/C	
Human MT IIa gene		CTTTTGCACTC	18 bp
MT consensus			
Stuart et al., 1984		CYTTTGCRYYCG	15-17 bp
Carter et al., 1984		TGCGCCCGC(C/T)C	14-17 bp
U-box consensus		TTTGCA	10-30 bp
Chick FK gene A		ATGATTTGCATG	80 bp
HSV TK gene		ATATTTGCATG	102 bp
p7AT H1 gene		TTAATTTGCAATG	284 bp
Human B-globin		TATTTGCATA	390 bp
Chick H5 gene		CTTTTTGCAGAA	600 bp
Drosophila ftz	794 bp (to ATG)	GTATTTGCCT ATGTTTGCAT	854 bp (to ATG)

not absolute and there is always a "wobble" in homology around an absolutely conserved 6 base sequence, TTTGCA, that I have termed the U-box (ubiquitous-box). It looks as if there is a family of elements (much like the family of heat-shock boxes identified in each Drosophila hsp gene) that bind a common trans-acting factor with different efficiencies depending on the exact sequence of the box. It is known that each Drosophila HSB has a different binding efficiency to the HSAP identified by Wu (1984b), presumably depending on the precise nucleotide sequence.

The finding of an inverse U-box in front of all heavy chain immunoglobulin gene regions suggests that the sequence maybe able to act irrespective of orientation when linked to a TATA box. I have indicated the distance between the inverse U-box and a TATA element just downstream of it with respect to the normal orientation presented in Table 6-2. Thus, the U-box may not act irrespective of orientation but possibly works on a transcription unit divergent from but very closely associated with the heavy-chain gene locus. This is pure speculation and no data for a divergent transcription unit in the immunoglobulin heavy-chain locus has been reported.

The expected occurance of any 6 base sequence in a single strand of DNA can be calculated statistically to be once every 4096 bases (4^6). Therefore the identification of the U-box in an average length gene should be relatively common, however the strict positioning of the 6 base U-box relative to other known promoter elements futher refines the definition of a U-box. The U-box must be very tightly associated with the TATA element of a gene to be called a U-box (in my definition) as the following examples show.

Two genes with good U-boxes with respect to sequence and position in

relation to a TATA box is a HLA-DR gene and a human metallothionein gene (Table 6-2). To illustrate the need for sequence homology and position associations in identifying a U-box, sequences with homology to the U-box in the chick feather keratin gene A, HSV-TK gene, the p7AT H1 gene, human β-globin gene, the chick H5 gene and the Drosophila ftz gene are presented in Table 6-2. The homology of these elements to the dc consensus is slightly less than any of the other genes presented in Table 6-2 but their distances from known TATA boxes bears no resemblence to their function as a U-box, in the sense of the other genes whose U-box is always tightly associated with a TATA box. The same may be true for the U-boxes identified in the mouse immunoglobulin enhancer and the SV40 enhancer identified by Falkner and Zachau (1984).

The work of Parslow et al. (1984) showed that the 70Z/3 rearranged immunoglobulin kappa light chain gene exhibits all the known properties of a functionally competent transcription unit. However mRNA transcripts from this gene are only detected after exposure (i.e. induction) of the cells containing the gene to bacterial lipopolysaccharide. The induction of this gene and also the presence of a potential U-box in the consensus DNA sequence thought to be responsible for heavy metal induction of the metallothionein genes (Stuart et al., 1984; Carter et al., 1984; Table 6-2), has lead me to the following speculative hypothesis to the function of the U-box. I propose that the U-box may be the recognition sequence for a trans-acting factor induced by general stress of a eukaryotic cell be it heat-shock, heavy metals or a challenge with bacterial lipopolysaccharide. This is despite my failure to demonstrate the induction of the p7AT H2B gene by thermal stress in the oocyte being caused by the H2B/U-box (6-3).

6-5 Discussion

The search for a biological function of the chicken H2B-box has not

been successful. It is difficult to imagine that such an element, conserved in sequence and in its position relative to the TATA box, does not have a role in the expression of H2B genes. Perhaps the non-dividing oocyte is an inappropriate host cell to reveal the function of this motif. The presence of the H2B/U-box in several other gene systems leads to the idea of a general stress response mechanism whereby induction of genes containing the U-box next to their TATA boxes occurs. It may also be that the role of such a motif will be more mundane and may be seen in a multitude of diverse genes as a general promoter element, much like the TATA box. I doubt this will be the case.

Other roles of the U-box sequence besides one involved directly in transcriptional control include a possible function as a specific origin of DNA replication, which could subsequently activate genes containing it. The synthesis of histone transcripts during the replication of DNA (1-6-4) must be considered as evidence for this hypothesis. One final option to the function of the U-box is that it may be a nuclear matrix binding sequence, and be involved in higher order active chromatin functions (1-3-5). It has been recently determined that there are specific DNA sequences that bind to the nuclear matrix (Mirkovitch et al., 1984).

CHAPTER 7: FINAL DISCUSSION

7-1 Histone gene architecture

One feature of a designed system, in this case the control of gene expression of two genes of a multigene family, is that it is likely to be easier to control the system if the component parts are together.

The apparently random organisation of histone genes in the chicken genome (figures 2-1, 2-2 and 2-3) belies the extraordinarily similar architecture of two of the genes of the multigene family. The coupling of H2A and H2B genes cannot be mistaken for a remnant of evolution of the gene family in the chicken. The DNA sequence of five of the six pairs shows that some 5' regions have been evolving independently, but forces have been operating to keep the genes not only together but with a consistent spacing between each other. Nor can the divergent H2A/H2B gene architecture be viewed as a peculiarity of the chicken histone gene system. As discussed in 2-3 this association, though not spacing or sequence homology, can be seen in a number of diverse species.

The selection pressure keeping these two genes together is probably the strict stoichiometric requirement of the proteins coded for by each gene. The equimolar appearance of H2A and H2B in the nucleosome can be achieved by assembly from the appropriately sized pools or by some regulatory control operating during nucleosome assembly. Given the divergent gene architecture of H2A and H2B in the chicken its seems more than likely it is a coupling of mRNA production resulting in the accumulation of each protein at an equal or an appropriate level so that the nucleosome is assembled from the correct stoichiometric pool of each protein. The expression work undertaken in this thesis supports the idea that the genes are coupled for the purpose of coupling the production of each protein. When the genes are separated the production of each

transcript would be affected. This does not preclude the idea of a fine tuning of nucleosome assembly by mechanisms that operate during nucleosome formation. Indeed it is possible that the levels of the other histones relative to H2A and H2B are controlled in just such a manner.

Why only two of the chicken histone genes have such a regular gene arrangment and the others not is puzzling. So far research on H3 and H4 genes have shown no rigid association of these two, however there is a loose coupling of these two genes in several species (see gene organisational tables in Maxson et al., 1983a). A more comprehensive explanation for the coupling of only the H2A and H2B histone genes may be gained with futher investigation into the biology of the histone gene system of the chicken.

There are H2A and H2B genes that are not in a divergent arrangement in the chicken. There are three individual H2A genes and a single H2B gene in addition to the six divergent H2A/H2B gene pairs. The effect of these additional non proportional gene copies would be expected to increase the level of H2A protein relative to that produced by the H2B genes due to the higher gene copy number. However relative rates of mRNA transcription and translation must also be taken into account before the eeffect of gene copy number on protein production can be considered. In addition, some of these genes may be non-functional although sequence analysis of 27 of the 42 core plus H1 chicken histone genes has not yet revealed a pseudogene.

One of the best characterised histone gene systems, besides that of the chicken, contains the embryonically-expressed sea urchin histone genes. It is noteworthy here that the H2A and H2B genes are not linked but are separated by an H3 gene in a tandem gene repeat of the 5 histone gene types (Hentschel and Birnstiel, 1981). A possible explanation for the uncoupling

of the H2A and H2B genes in this system is that there may be only loose transcriptional control exercised in this gene system. These highly reiterated genes of the sea urchin are used to produce the maternal histone mRNA store during oogenesis and the newly synthesised histone gene transcripts produced during the early stages of cleavage of the zygote. It may be better when dealing with a maternal store of histone mRNA and protein to control the translation and/or the nucleosomal assembly of the histone proteins to achieve the stoichiemetric requirement of the nucleosome. In contrast to these embryonic type histone genes, the lower copy number late histone subtype genes in sea urchin are more like the chicken histone gene system, with a random gene architecture and contain at least one divergent H2A/H2B gene pair.

7-2 Divergent gene promoter sequences: multiple repeating sequence elements and the role of the CCAAT box

The chicken H2A/H2B pairs contain four CCAAT sequences in their intergene regions. The presence of multiple CCAAT boxes may not be uncommon in eukaryotic genes although to date it has only been customary to identify one CCAAT element in a genes promoter sequence. Indeed there is a duplication of the CCAAT element in the human $^{A}\gamma$-globin gene. Moreover a single G to A substitution in the more distal of the two CCAAT elements can be correlated to the hereditary persistence of fetal haemoglobin in some β-thalassaemics (Gelinas et al., 1985; Collins et al., 1985). It has been speculated that this CCAAT box region of the γ-globin gene is the binding site of a repressor that turns off the gene in the normal human adult. The G to A substitution would then inhibit the DNA-protein interaction and allow the inappropriate expression of the fetal gene in the adult. Alternatively the γ-gene may depend on binding of a positive trans-activating factor normally present in fetal cells, the single base change

would then not affect this binding but may change the specificity of the CCAAT element so that another trans-activator can act upon the sequence in the adult. Whatever the mechanism, it has been proposed that the CCAAT box is one of the elements which contribute to the developmental control of γ-globin expression.

This suggested function for the CCAAT element must be considered to be its role in only one specific case; the element is by no means specific to the globin genes and therefore must act differently in other genes. The CCAAT element is in all probability a family of protein/factor binding sequences, its role changing with slight changes in the CCAAT sequence itself and sequences surrounding it. This example of a semi-conserved sequence block identified in front of diverse genes is similar to the identification of the U-box in front of the several different types of genes discussed in chapter 6. Again a definitive functional role for an apparently consistent sequence block is elusive.

The appearance of multiple repeating sequence blocks, like the CCAAT boxes of the H2A/H2B pair, has also been seen in the intergene region of the SV40 virus divergent gene promoter region. However in this case it is the G-C box element (1-4-2(i)) that is repeated. A transcription factor termed Sp1, one of the few that have so far been examined, has been shown to interact with this G-C region (Gidoni et al., 1984). It has been found that the factor binds only to the DNA strand that contains the six GGCGGG repeated sequences. There is evidence that the Sp1 dependent transcription of the SV40 genome results in the initiation of divergent transcripts on opposite DNA strands at locations approximately equidistant from the binding region.

Gidoni et al. have proposed two models to explain how the binding of a factor to an asymmetric sequence results in symmetrical activation of

transcription of the SV40 genome. Activation could preceed via protein-protein contact. This would require the Sp1 protein to have recognition sites at both ends of the asymmetric molecule which is unlikely. Alternatively the activation of the SV40 divergent transcripts by Sp1 has been suggested to proceed by a perturbation of the DNA helix that propagates in both directions. I favour this second hypothesis and imagine the same thing could happen in the H2A/H2B intergene region with the multiple repeating CCAAT elements. In this case the activation must be more complex in that a single CCAAT box is orientated toward one gene and three toward the other.

No three dimensional conformation for any eukaryotic gene promoter has been described in any detail, therefore what can be said about the structure of the promoter region of genes must be taken as speculation. The effect of the binding of a regulatory protein to a gene promoter sequence may be mediated through direct protein-protein interactions or through an "action at a distance" whereby the structure of the DNA helix at some disance from the actual protein binding site undergoes a conformational change. It is possible that both mechanisms could be used at the same time by different DNA binding proteins that bind to the same promoter region. The presence of multiple asymmetrical sequence units in both the divergent H2A/H2B and SV40 transcription units suggests that the hypothesis of a signal being propagated through the DNA backbone is more likely to be correct.

7-3 Functional analysis of a H2A/H2B divergent gene promoter region

The failure to find effects of the CAT1 and H2B-box deletions on the expression of either the H2A or the H2B gene (5-4; 6-3) in the oocyte transcription system is paradoxical given the marked effect of separation

of the two genes (5-2), the finding that the intergene region with only small amounts of the gene coding region is sufficient for promotion of both transcripts (5-3), and the conserved sequence and spatial arrangement of these two elements in the intergene region (Table 3-1). Perhaps in the separation of the genes at the intergenic XhoI site there is a loss of the postulated "perturbation signal" transmitted through the helix (7-2) which is not seen in the small site directed deletions. This loss of control on gene separation could mean that the remaining CCAAT boxes would not be able to act synergistically to promote transcription.

The remarkable spatial relationship of the conserved intergenic elements (Table 3-1) suggests that spacing could be important in the interaction of the CCAAT boxes proposed to be involved in the promotion of transcription. Future work on the analysis of the divergent gene promoter could concentrate on the significance of the spacing of these elements. Experiments would involve the introduction of exogenous DNA of varying lengths between the conserved elements, followed by functional testing. This would be most easily done by utilising the unique XhoI site between the CAT1 and CAT2 boxes.

CHAPTER 8: MATERIALS AND METHODS

8-1 Abbreviations

Abbreviations were as described in "Instructions to authors" (1978). In addition:

BCIG : 5-bromo-4-chloro-3-indolyl-β-D-galactoside

bisacrylamide : N,N'-methylene-bisacrylamide

ddNTP : dideoxynucleoside triphosphate

DTT : dithiothreitol

IPTG : isopropyl-β-D-thio-galactopyrosine

PIPES : piperazine-N,N'-bis(2-ethnane-sulfonic acid)

PEG : polyethylene glycol

SDS : sodium dodecyl sulphate

TEMED : N,N,N',N'-tetramethylethylenediamine

8-2 Materials

8-2-1 Chemicals and Reagents

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

Acrylamide, agarose, ATP, ddNTPs, dNTPs, DTT and bisacrylamide - Sigma

Low melting point agarose - B.R.L.

Mixed bed resin AG 501-X8 (D) - Bio-rad

TEMED and xylene cyanol - Tokyo Kasei

Urea (ultra pure) - Merck

Chloramphenicol - gift from Parke-Davis

Bromo cresol purple, formamide and PEG 6000 - BDH

M13 universal primer, γ-^{32}P-ATP, α-^{32}P-dCTP and α-^{32}P-dATP

- Biotechnology Research Enterprises of S.A. (BRESA)

8-2-2 Enzymes

Enzymes were obtained from the following sources:

AMV reverse transcriptase - Molecular Genetic Resources

Calf intestinal phosphatase and Ribonuclease A - Sigma

E. coli DNA-polymerase I, Klenow fragment - Boehringer Mannheim

BRESA

Proteinase K - Boehringer Mannheim

Restriction endonucleases - Boehringer Mannheim

New England Biolabs

S1 nuclease - Boehringer Mannheim

T4 DNA ligase - New England Biolabs

BRESA

T4 polynucleotide kinase - Boehringer Mannheim

8-2-3 Buffers and Media

All bacteria, except JM101, were grown in L-broth or on L-agar plates. JM101 was grown in minimal medium, 2 x YT broth and on minimal plus glucose plates.

L-broth : 1% (w/v) amine A, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, pH 7.0

L-agar plates contained L-broth with 1.5% (w/v) bacto-agar

Minimal medium : 2.1% (w/v) K_2HPO_4, 0.9% (w/v) KH_2PO_4, 0.2% (w/v) $(NH_4)_2SO_4$, 0.1% (w/v) tri-sodium citrate

Minimal plus glucose plates contained minimal medium, 0.4% (w/v) glucose, 0.0001% (w/v) thiamine and 1.5% (w/v) bacto-agar

2 x YT broth : 1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl, pH 7.0.

All buffers and media were prepared with distilled and deionised water

and sterilised by autoclaving, except heat labile reagents, which were filter sterilised.

8-2-4 Bacterial strains, clones and cloning vehicles

Bacterial stocks

E392 - gift from Dr. J.B. Egan

MC1061 - gift from Dr. R.P. Harvey

JM101 - gift from Dr. A. Robins

Cloned DNA Sequences

pCH3.3E - gift from Dr. R.P. Harvey

pCH22.0B - gift from C. Lesnikowski

λCH-05, pCH3.5E and pCH11.0E - gift from R. D'Andrea

λCH1-10 EcoRI-HindIII H2A/H2B fragment - gift from L.S. Coles

Cloning Vectors

pBR322 - gift from R. D'Andrea

pAT153 - gift from Dr. J.B. Egan

M13mp7, M13mp8, M13mp9, M13mp18, M13mp19 - gift from Dr. A. Robins

8-3 Methods

8-3-1 Isolation of plasmid DNA

(i) Growth and amplification of plasmid DNA

A loopful of a glycerol stock of E. coli, containing the plasmid to be grown, was streaked on an L-agar plate (usually supplemented with an appropriate antibiotic to maintain selective pressure for the retention of the plasmid; 30 ug/ml) and incubated overnight at 37 C. A single colony was used to infect 5 ml of L-broth (supplemented with antibiotic), which was then grown overnight at 37 C, with vigorous shaking if amplification of the plasmid was intended. If amplification was not required a 100 ml solution of L-broth was infected with a single colony and the broth incubated overnight with shaking.

To amplify plasmid in the 5 ml overnight culture it was diluted 100 fold into 500 ml of fresh broth (without antibiotic) and grown with aeration at 37 C to an A_{600} of 1.0, at which time chloramphenicol was added to a final concentration of 150 ug/ml, and incubation continued overnight (Clewell, 1972).

(ii) Large scale isolation

The cells from the 100 ml overnight or the 500 ml amplified culture were harvested by centrifugation (JA-10 rotor, 6000 rpm for 10 minutes). The plasmid DNA was isolated by a modified procedure of the alkaline extraction procedure of Birnboim and Doly (1979).

Cell pellets were resuspended in 4 ml of 15% (w/v) sucrose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA, 2 mg/ml lysozyme and incubated on ice for 20 minutes. 8 ml of freshly made 0.2 M NaOH, 1% (w/v) SDS was added and gently mixed until the suspension became almost clear and slightly viscous. The solution was then left on ice for a further 10 minutes. 5 ml of 3 M sodium acetate, pH 4.6, was added and gently mixed by inversion for a few seconds. The tube was maintained on ice for 30 minutes to allow protein, high molecular weight RNA and chromosomal DNA to precipitate, then centrifuged for 25 minutes at 15,000 rpm (JA-20 rotor) to remove the clot. The supernatant containing supercoiled plasmid DNA was carefully aspirated, avoiding lumps of the precipitate, treated with Ribonuclease A (heat-treated at 80 C for 20 minutes to inactivate deoxyribonucleases; 50 ul of a 10mg/ml stock) for 20 minutes at 37 C, phenol-chloroform extracted, then ethanol precipitated (8-3-3).

After centrifugation, the plasmid DNA pellet was redisolved in 1.6 ml of water and 0.4 ml of 4 M NaCl, 2 ml of 13% PEG 6000 added and the solution was then placed on ice for 1 hour (half of these volumes were used

when the nucleic acid from a 100 ml bacterial culture was to be PEG precipitated). The plasmid DNA was collected as a pellet after a 10 minute centrifugation in an Eppendorf microfuge. The PEG supernatant was removed, the pellet washed in 70% ethanol and then dried in vacuo. The pellet was then resuspended, ethanol precipitated, washed, dried and resuspended in an appropriate volume of water.

The yield of DNA, prepared by this method, was approximately 300 to 600 ug per 500 ml amplified culture and 100 ug per 100 ml culture, as assayed by electrophoresis (8-3-4).

(iii) Miniscreen procedure

Colonies were grown overnight in 2 ml of L-broth plus an appropriate antibiotic at 37 C with continual shaking. The cells were then pelleted by centrifugation for 5 minutes in an Eppendorf microfuge. Plasmid DNA was extracted by the same, but scaled down, procedure as described above except that the PEG precipitation procedure was omitted. The DNA prepared in this way was pure enough to be cut with restriction enzymes.

8-3-2 Isolation of M13 replicative form

A single plaque was toothpicked into 100 ml of 2 x YT broth containing 10 ml of a fresh JM101 overnight culture, grown from a single colony picked from a minimal plus glucose plate. This culture was incubated at 37 C with vigorous aeration for 6 hours. The method of isolation of M13 replicative form from this culture was the same as that for the isolation of plasmid DNA (8-3-1). The yield of DNA prepare in this way was approximatly 200 ug.

8-3-3 Restriction endonuclease digestions

All restriction endonuclease digestions were performed using the conditions for each enzyme described by Davis et al. (1980). ATP (50 uM) was also included if the restricted DNA was to be ligated. A two-fold excess of enzyme generally was used and the reactions were run for an hour,

although this time was increased for preparative digestions.

Reactions were stopped by the addition of EDTA, pH 7.4 to 5 mM, and protein removed by phenol/chloroform extraction or by the addition of a quarter volume of urea load buffer (4 M urea, 50% (w/v) sucrose, 50 mM EDTA, pH 7.4, 0.1% (w/v) bromo cresol purple). In phenol/chloroform extractions one half volume of phenol saturated with 100 mM Tris-HCl, pH 9.0, 5mM EDTA, 50 mM 2-mercaptoethanol, was added, mixed and a half volume of chloroform added. After vortex mixing and centifugation (Eppendorf microfuge, 1 minute; JA-20 rotor 5 minutes at 7,000 rpm) the upper aqueous phase was recovered.

DNA was ethanol precipitated from the aqueous supernatant by adjusting the reaction mix to 0.2 M NaCl or 0.3 M sodium acetate, pH 5.5 and addition of 2.5 volumes of nuclease-free ethanol, followed by freezing in an ethanol/dry ice bath. After thawing, the DNA was pelleted by centrifugation for 10 minutes in an Eppendorf microfuge or a JA-20 rotor (10,000 rpm). The DNA pellet was washed with 70% nuclease-free ethanol and dried *in vacuo* before being redissolved in an appropriate volume of water.

8-3-4 Routine gel electrophoresis

(i) 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% acrylamide/bisacrylamide (30:1), which had been deionized with mixed bed resin.

A 15 ml mixture of acrylamide and bisacrylamide in TBE buffer (100 mM Tris-borate, 2.5 mM EDTA) was prepared, 250 ul of 10% (w/v) ammonium persulphate and 12.5 ul of TEMED added, and the solution poured into a gel mould and allowed to polymerise. Gel reservoir tanks contained

approximately 1 litre of TBE buffer and gels were pre-electrophoresed at 50 mA for 20 minutes before loading. Samples were dissolved in 10 ul of water and a quarter volume of loading buffer (50% (w/v) sucrose, 50 mM Tris-HCl, pH 7.4, 5 mM EDTA, 0.1% (w/v) bromo cresol purple and 0.1% (w/v) xylene cyanol) and layered directly into gel slots approximately 1 cm wide. When DNA samples of greater than 10 ug were electrophoresed, the DNA samples were dissolved in a larger volume of water and loading buffer and layered into an appropriately sized gel slot.

All acrylamide gels were electrophoresed at 30 mA until the dyes had moved the desired distance. DNA was visualised under UV light after ethidium bromide staining or by autoradiography (8-3-5) if the DNA contained ^{32}P.

(ii) Agarose gel electrophoresis

Agarose was dissolved in TEA (40mM Tris-acetate, 20 mM sodium-acetate, 1 mM EDTA, pH 8.2) to 0.7-2% (w/v) and cast either in 14 cm x 14 cm x 0.3 cm vertical slab gel templates or on to 7.5 cm x 5 cm microscope slides, for horizontal gels. Vertical gels were electrophoresed between reservoirs each containing 500 ml of TEA at 65 mA, usually for 3 hours. Horizontal gels were submerged in 400 ml of TEA buffer and a current of 125 mA was applied for approximately 20 minutes.

DNA samples were dissolved in 10 ul of water and 2.5 ul of loading buffer (50% (v/v) glycerol, 50 mM Tris-HCl, pH 7.4, 5 mM EDTA, 0.1% (w/v) bromo cresol purple) and loaded directly into gel slots for vertical gels. Samples of 5-8 ul were loaded into the horizontal gel slots.

The DNA was visualised with ethidium bromide as described above (8-3-4(i)).

(iii) Low melting point agarose gel electrophoresis

When a DNA sample was to be recovered from an agarose gel, low melting

point agarose was used in place of normal agarose in the gel systems described above. To prevent cracking due to shrinkage, the gel was poured after leaving agarose and the gel mould to equilibriate at 37 C. The gel was then allowed to set at 4 C and electrophoresis was also carried out at 4 C.

Prior to loading the gel was pre-electrophoresed for 10 minutes. Electrophoresis was at similar voltages and for similar times to that for normal agarose gel electrophoresis. The bands were visualised and then cut from the gel for elution (8-3-7).

8-3-5 Autoradiography

^{32}P-labelled DNA, which had been electrophoresed on polyacrylamide gels was visualised by autoradiography. Gels were covered with a thin sheet of plastic-wrap and a sheet of Fuji X-ray 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.

When DNA bands were to be eluted from a gel, one of the glass plates of the gel mould was removed and plastic wrap placed over the gel. Two strips of tape were adhered to the plastic wrap each side of the tracks concerned and spotted with radioactive ink (to act as markers). A sheet of X-ray film was placed on the gel and autoradiographed for the required amount of time. After developing and fixing the film, the radioactive ink spots were lined up with the X-ray film and bands cut from the gel for elution (8-3-6).

8-3-6 Elution of DNA from polyacrylamide gels

The gel slice containing the DNA to be eluted was placed in an Eppendorf tube and 400 ul of TE buffer (10 mM Tris-HCl, pH 7.4, 1 mM EDTA) was added and incubated overnight at 37 ^{0}C with shaking. The buffer was aspirated from the gel slice and the DNA precipitated with nuclease-free ethanol.

8-3-7 Isolation of DNA from low melting point agarose

The smallest slice of agarose possible, containing the DNA fragment, was placed in an Eppendorf tube. The agarose was melted at 65 ^{0}C, then 200 ul of 0.2 NET buffer (200 mM NaCl, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA) added. This mixture was phenol extracted twice without the addition of chloroform, then once with phenol/chloroform. The aqueous phase was then ethanol precipitated. Approximately 50-70% of the fragment loaded onto the gel was recovered in this way.

8-3-8 Subcloning of DNA fragments into plasmid and M13 vectors

(i) End-filling of DNA fragments and blunt-end ligations

Vector DNA was linearised with a suitable restriction enzyme then dephosphorylated with calf-intestinal phosphatase in a 50 ul reaction mix containing 50 mM Tris-HCl, pH 9.0, 1 mM $MgCl_2$, 0.1 mM $ZnSO_4$ and 0.18 units of enzyme. After a 1 hour incubation at 37 ^{0}C the enzyme was heat inactivated at 68 ^{0}C for 15 minutes in the presence of 0.5% SDS and the protein removed by phenol/chloroform extraction. The linearised dephosphorylated vector was purified from uncut vector by passaging the DNA through a low melting point agarose gel.

Restriction fragments to be subcloned were preparatively isolated from either agarose or polyacrylamide gels. When the DNA fragment had protruding 5' or 3' termini, it was treated with DNA polymerase I Klenow fragment to end-fill or digest back the single-strand regions to blunt-ends. This was done, before purification of the fragment by gel electrophoresis, in a 20

ul reaction mix containing 10 mM Tris-HCl, pH 7.4, 50 mM NaCl, 10 mM MgCl, 1 mM DTT, 30 uM of each dNTP and 1 unit of Klenow fragment.

Ligation of insert into vector was done in a 10 ul volume containing 50 mM Tris-HCl, pH 7.4, 10 mM MgCl, 1 mM DTT, 0.5 mM dATP and 0.5 units of T4 DNA ligase at 4 ^{0}C for 16 hours. Sufficient insert to give a three fold molar excess over vector was generally used, 50 ng of plasmid or 10 ng of M13 vector was normally contained in a ligation mix. Recombinant molecules were transformed into bacteria as described below (8-3-9, 8-3-10).

(ii) Sticky-end ligations

These were performed as described above except 0.1 units of ligase was used in the reaction and incubation carried out at 14 ^{0}C for 4-16 hours.

8-3-9 Transformation procedure for plasmid recombinants

A single colony of E. coli strain E392 or MC1061 was used to infect 5 ml of L-broth which was then grown overnight at 37 ^{0}C with aeration. One ml of the overnight culture was diluted 50 fold in 50 ml of L-broth and grown with shaking to an A_{600} of 0.8. The cells were then rapidly chilled on ice for 30 minutes. The cells were pelleted by gentle centrifugation (HB-4 rotor, 5000 rpm for 2 minutes), washed in a half volume of ice cold 0.1 M $MgCl_2$, then resuspended in one twentieth volume of ice cold 0.1 M $CaCl_2$ and left on ice for at least 1 hour. 200 ul of this cell suspension was mixed with 100 ul of ligation mix diluted in 100 mM Tris-HCl, pH 7.4 and left on ice for 30 minutes with occasional mixing. The transformation mix was heated to 42 ^{0}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 ^{0}C for 30 minutes. After this time 3 ml of L-broth containing 0.7% agar was added to the transformation mix and poured onto an L-agar plate containing an appropriate antibiotic, depending on the

resistance carried by the plasmid. The plate was incubated overnight at 37 C.

8-3-10 Transformation procedure for M13 recombinants

A loopful of E. coli strain JM101, from a minimal plus glucose plate, was used to infect 5 ml of minimal medium which was then grown overnight at 37 C with aeration. The overnight culture was diluted 50 fold into 50 ml of 2 x YT and grown with shaking to an A_{600} of 0.4. The cells were pelleted by gentle centrifugation (HB-4 rotor, 5000 rpm for 2 minutes), then resuspended in one-twentieth volume of ice cold 50 mM $CaCl_2$ and left on ice for at least 1 hour. 200 ul of this cell suspension was mixed with a sample of the ligation mix and left on ice for 40 minutes. The transformation mix was heated to 42 C for 2 minutes. 3 ml of L-broth containing 0.7% agar, 20 ul of BCIG, 20 mg/ml in dimethylformamide, 20 ul of IPTG, 24 mg/ml in water, and 0.2 ml of a JM101 overnight culture diluted 1:5 in 2 x YT broth was then added and the mixture plated directly onto a minimal plus glucose plate. Incubation was overnight at 37 C.

8-3-11 M13 chain terminator sequencing of DNA (EMBO sequencing notes, 1980)

(i) Preparation of template

M13 phage plaques were toothpicked into 1 ml of a 1:40 dilution of a fresh JM101 overnight culture (grown in minimal medium) in 2 x YT broth. After incubation at 37 C with vigorous shaking for 5 hours, the culture was centrifuged for 5 minutes in an Eppendorf microfuge. The supernatant was poured into an Eppendorf tube containing 200 ul of 2.5 M NaCl, 20% PEG 6000 and left at room temperature for 15 minutes. The single-stranded M13 phage particles were collected as a pellet after centrifugation for 5 minutes. The supernatant was aspirated and the pellet resuspended in 100 ul of 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, phenol/chloroform extracted and the DNA

recovered by the addition of 10 ul of 3 M sodium acetate, pH 5.5 and 250 ul of nuclease-free ethanol (as for an ethanol precipitation 8-3-3). The DNA pellet was resuspended in 25 ul of 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA and stored frozen at -20 ^{0}C.

(ii) Sequencing reactions

Four separate reactions, each specific for one of the bases in DNA, were used in the sequence analysis of the insert of the M13 single-stranded template. In each of the sequencing reactions, the M13 universal primer or one of the histone gene specific primers (8-3-12; chapter 4) was extended in the presence of a different ddNTP such that there was a partial incorporation of each, which resulted in termination of synthesis.

The method described below is for the sequencing of one M13 clone but is readily expanded to allow the concurrent sequencing of 8-16 clones.

(a) Hybridization

2.5 ng of universal primer (17-mer) or 5 ng of one of the histone 26-mers was annealed to 5 ul of M13 single-stranded template (prepared as above) in a 10 ul volume containing 10 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$ by heating the solution to 70 ^{0}C for 3 minutes and then hybridizing at room temperature for 45 minutes.

(b) Polymerisation

1 ul of α-^{32}P-dATP (approximately 4 uCi) was lyophilized, the hybridization mix added, vortexed to dissolve the labelled dATP and then 1 ul of 10 mM DTT added. 1.5 ul of each of the appropriate zero mixes (T^{0} for ddTTP: 10 uM dTTP, 200 uM dCTP, 200 uM dGTP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA; C^{0} for ddCTP: 200 uM dTTP, 10 uM dCTP, 200uM dGTP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA; G^{0} for ddGTP: 200 uM dTTP, 200 uM dCTP, 10 uM dGTP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA; A^{0} for ddATP: 200 um of dTTP, dCTP and dGTP,

5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) and ddNTP solutions (0.1 mM for ddCTP and ddATP, 0.5 mM for ddTTP and ddGTP, each in water) were added together. 2 ul of the zero - ddNTP mixtures were added separately to four Eppendorf "reaction tubes".

0.5 ul of DNA polymerase I, Klenow fragment (1 unit/ul) was added to the hybridization mixture - label - DTT solution. 2 ul of this was then added to each of the four reaction tubes and the solutions were mixed by centrifugation for 1 minute. After 10 minutes incubation at 37 ^{0}C, 1 ul of dATP chase (500 uM dATP in 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) was added to each of the four tubes, mixed by a 1 minute centrifugation and incubated for a further 15 minutes at 37 ^{0}C.

3 ul of formamide loading buffer (100% formamide, 0.1% (w/v) bromo cresol purple, 0.1% (w/v) xylene cyanol and EDTA to 20 mM) was added to stop the reactions and mixed by a short centrifugation. Samples were boiled for 3 minutes and then loaded onto a sequencing gel.

(iii) Sequencing gels

Products of the dideoxy-chain terminator sequencing reactions were separated by electrophoresis on polyacrylamide gels which included 7 M urea as a denaturant. The gels used were 40 cm x 40 cm x 0.35 mm. A 6% gel, which was normally run, was made in the following way. A 85 ml mixture of acrylamide monomer (20:1, acrylamide to bisacrylamide) in TBE buffer containing 7 M urea, was prepared, 800 ul of 10% (w/v) ammonium persulfate and 65 ul of TEMED added, the mixture poured into a gel mould and allowed to polymerise.

Reservoir tanks contained about 4 litres of TBE buffer and these gels were pre-electrophoresed for 45 minutes at 30 mA. Debris and urea were removed from sample wells prior to loading by flushing with buffer from a syringe. All gels were run at 25-40 mA and kept at high temperature during

electrophoresis to facilitate DNA denaturation.

Gels were fixed with 200 ml of 10% (v/v) acetic acid and washed with 2 litres of 20% (v/v) aqueous ethanol. After drying, the gels were autoradiographed overnight at room temperature.

8-3-12 Preparation and kinasing of synthetic oligonucleotides

Synthetic DNA primers were generously provided by Dr. D. Skingle and S. Rogers. Oligodeoxyribonucleotides were synthesised on a silica gel support using deoxynucleoside morpholinophosphoramidite monomers (McBride and Caruthers, 1983), and protecting groups were removed as described by Metterici and Caruthers (1981). The tritylated DNA was purified by reverse-phase HPLC and then de-tritylated (Metterici and Caruthers, 1981). The products were provided as triethylammonium DNA salts in water and were both single-stranded and devoid of 5' phosphate groups making them ideal substrates for kinasing and sequencing by chemical degradation methods of Maxam and Gilbert (8-3-13).

The histone specific 26 base primers prepared were as follows:

H2A primer: (residues -340 to -315, figure 3-2, upper strand)

5' dAGCGACTGAACACTCAGAGAGCAAAC 3'

H2B primer: (residues -354 to -329, figure 3-2, lower strand)

5' dGGCTCGGGCATAGTGGCACAACGCGC 3'

H1 primer: (residues complementary to 21 to -5, figure 4-2)

5' dGGCGGGAGCGGTCTCGGACATCGCGG 3'

Primers were 5' end-labelled with T4 polynucleotide kinase and γ-^{32}P-ATP. Normally 50-100 ng of synthetic primer was kinased in a 10 ul reaction containing 50 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 5 mM DTT, 7 ul of lyophilized γ-^{32}P-ATP (approximately 35 uCi) and 1 unit of enzyme. The reaction was incubated for 45 minutes at 37^{0}C were upon 5 ul of formamide

loading buffer was added (8-3-11(ii)b) and the mixture loaded onto a 20% polyacrylamide gel for purification (8-3-4(i), 8-3-6).

8-3-13 Sequencing of synthetic oligonucleotides

Initially the sequences of some of the synthetic oligonucleotides were verified by a modified Maxam-Gilbert procedure (Banaszuk et al., 1983). 100 ng (in a volume of 50 ul) of radioactively labelled primer prepared as described above was subjected to the following reactions.

Guanine cleavage

5 ul of ^{32}P-primer containing 4 ug of carrier tRNA and 200 ul of 50 mM sodium cacodylate, pH 8.0, 10 mM $MgCl_2$ and 0.1 mM EDTA were combined and 1 ul of Dimethyl sulfate added and the mixture incubated at 37 ^{0}C for 15 minutes. The mixture was chilled on ice, and then 50 ul of "stop" solution containing 3 M sodium acetate, pH 5.5, 2.5 M 2-mercaptoethanol, 1 mM EDTA and 0.1 ug/ul tRNA added. The DNA was precipitated by the addition of 750 ul of ethanol and chilling at -80 ^{0}C. The dried pellet was re-ethanol precipitated then dissolved in 1 M piperidine and incubated at 90 ^{0}C for 30 minutes. After cooling the mixture was dried in vacuo then resuspended in 100 ul of water and transferred to a new Eppendorf tube, adjusted to 200 mM NaCl and the sample precipitated with ethanol. The residue was dissolved in 5 ul of formamide loading buffer.

Adenine and guanine cleavage

10 ul of ^{32}P-primer containing 4 ug of carrier DNA and 10 ul of 3% diphenylamine, 1.5 mM EDTA in 98% formic acid were combined and incubated at 37 ^{0}C for 12 minutes. The reaction was chilled on ice and 250 ul of 0.3 M sodium acetate, pH 5.5, 0.1 mM EDTA and 0.025 ug/ul tRNA then added. The mixture was then treated as described for the guanine cleavage above.

Cytosine and thymine cleavage

10 ul of ^{32}P-primer containing 4 ug of carrier tRNA was combined with

10 ul of water and 30 ul of of hydrazine and incubated at 45 C for 18 minutes. The reaction was stopped by the addition of 200 ul of 0.3 M sodium acetate, pH 5.5, 0.1 mM EDTA and 0.025 ug/ul tRNA. The DNA was precipitated with ethanol and then treated with piperidine as described for the guanine cleavage.

Cytosine cleavage

5 ul of ^{32}P-primer containing 4 ug of carrier tRNA was combined with 15 ul of 5 M NaCl and 30 ul of hydrazine and the reaction incubated at 45 C for 18 minutes. The reaction was stopped by the addition of 200 ul of 0.1 mM EDTA, 0.025 ug/ul tRNA. The DNA was precipitated with ethanol and treated with piperidine as described for the guanine cleavage.

Thymine cleavage

5 ul of ^{32}P-primer containing 4 ug of carrier tRNA was mixed with 20 ul of freshly prepared $KMnO_4$ (20 ug/ul) and incubated at 37 C for 1 hour. The reaction was stopped by the addition of 10 ul of allyl alcohol then dried *in vacuo* and treated with piperidine as described for the guanine cleavage.

The cleavage samples were run on 20% sequencing gels (8-3-11(iii)) which were frozen during autoradiography.

8-3-14 Isolation of chick embryo RNA

5-day chick embryo RNA was prepared by a guanidine-HCl method (Brooker et al., 1980). Twenty 5-day chicken eggs were cracked open and the embryos lifted out and dropped into 25 ml of 6 M guanidine-HCl, 0.2 M sodium acetate, pH 5.2, 1 mM 2-mercaptoethanol at 4 C and homogenised. The homogenate was passed through a French pressure cell at 843.7 kg/cm^2 and 4 C, and RNA was precipitated with 0.5 volume of ethanol (HB-4 rotor, 10,000 rpm for 10 minutes). The RNA pellet was resuspended in 12.5 ml of 6

M guanidine-HCl, 0.2 M sodium acetate, pH 5.2, 10 mM EDTA and reprecipitated with 7 ml of ethanol. After centrifugation the pellet was resuspended in 10 ml of 7 M urea, 0.1 M Tris-HCl, pH 8.0, 0.1mM EDTA, 0.1% SDS and phenol/chloroform extracted twice followed by precipitation with 2.5 volumes of ethanol. Total RNA isolated by this procedure was washed in 5 ml of 2 M LiCl, dried *in vacuo* and resuspended in 2 ml of sterile water and store at -80°C. Approximately 10 mg of RNA was yielded by this procedure.

1 ml of the RNA was made to 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, heated to 65°C for 5 minutes, snap chilled on ice and centrifuged (Beckman SW-41 rotor at 37,000 rpm for 16 hours at 4°C) on four 10-40% (w/v) linear sucrose gradients in 10 mM Tris-HCl, pH7.4, 1 mM EDTA. Gradients were fractionated by upward displacement with a 50% (w/v) sucrose solution using an ISCO Density Gradient Fractionator. The RNA fractions containing the histone mRNA (about 9 S) were collected, adjusted to 200 mM NaCl and ethanol precipitated.

8-3-15 *Injection of Xenopus oocytes*

(i) *Animals*

Xenopus laevis females were obtained from Dr. Ray Harris (South Australian Institute of Technology, Pharmacology Department) or Dr. Keith Dixon (Flinders University, Department of Biological Sciences). A breeding stock is maintained at Flinders University whereas the Institute colony of 50 females was brought from Ring Shipping Pty. Ltd., Cape Town, South Africa. For best results, wild-type females were maintained under laboratory conditions for at least six months prior to harvesting of oocytes.

(ii) *Frog dissection*

Frogs were anaesthetised and the ovary, or part thereof, surgically

removed. Alternatively the animal was sacrificed (by pithing and decapitation) and the whole ovary obtained. A frog can be recycled up to four times if only a small portion of the ovary is removed during surgery.

In preparation for dissection the frog was placed in a plastic box containing 0.1% Ethyl-m-aminobenzoate (an anaesthetic) for approximately 20 minutes, removed using plastic gloves, washed with tap water and placed on its back. Alternatively, the frog was equally well immobilized by immersion in an ice water slurry for 20 minutes, placed on its back on an ice tray and covered in ice.

When no reflexes were evident the frog was dissected. The skin was swabbed in 0.5% Hibitane/70% ethanol to remove slime. Since the ovary of a mature female is the largest organ in the animal, the ovarian lobes are easily removed with forceps through a small incision either side of the ventral mid-line. The required number of lobes were cut off with scissors and the excised lobes placed immediately in Barths' saline. Once the incision had been sutured post-operative care entailed placing the animal in an angled dish, its nose just out of water, until it revived.

(iii) Ovary dissection and evaluation of oocytes

The excised ovary was rinsed in fresh Barths' saline then teased apart with (grade 5) watchmaker forceps. For long term storage the ovary must be divided into small clumps (less than 50 oocytes) and ideally into separate oocytes, which in any case are used for nuclear microinjection.

A good ovary contains over 30,000 large oocytes. They are at all stages of development. The ideal material is stage 5 and 6 oocytes (Dumont, 1972) in which the animal (brown) and vegetal (yellow) poles are separated by a white band and each cell is about 1 mm in diameter. The separated oocytes were kept in groups of 100 in 5 cm petri-dishes in a temperature-

controlled incubator at 17-21 °C; the medium was changed every two days.

Oocytes were kept submerged at all times. Healthy oocytes are evenly pigmented, not blotchy. Dimpled or deformed oocytes were removed. In order to segregate out oocytes damaged during the isolation procedure, oocytes were commonly injected the day after they had been separated.

(iv) Construction, calibration and handling of micro-pipettes and components of the injection system

Micro-injection needles were made from 100 ul micro-capillaries (BLAUBRAND, intraMARK). Figure 8-1 outlines the procedure of construction and gauging.

An Agla screw-controlled syringe (Wellcome Australia Ltd.) was connected to the micro-pipette by plastic tubing of 1 mm internal diameter. The needle was held and manoeuvered by means of a micromanipulator which gives a 4-5 fold reduction in the movement of the hand in all directions (Micro Techniques (Oxford) Ltd.). The whole system is shown in figure 8-2. The tubing and the needle was filled with medicinal paraffin coloured with Fast Red dye enabling discrimination of the paraffin/aqueous interface.

Injections were carried out under a dissecting stereozoom microscope at a magnification of about 15x. A cold light source for the microscope was essential as oocytes are prone to dehydration under a direct light source. A glass fibre optics system was suitable.

(v) Micro-injection technique

(a) Filling of micro-pipette

Usually 1 ul samples of injection solution were brought to the microscope stage on a piece of Parafilm. The tip of the paraffin filled needle was introduced below the surface of the droplet and the sample drawn into the pipette by screwing out the syringe. The only constraint on the amount of sample used to fill the pipette is that both the sample meniscus

Figure 8-1: Micro-injection pipettes

(a) A glass micro-capillary is drawn out by hand to an external diameter of approximately 300 um.

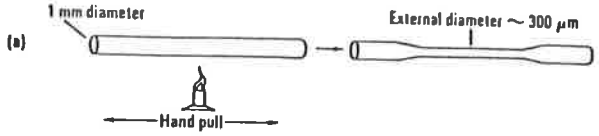
(b) The second pull is done using a heated coil, the needle being drawn out by attachment to a weighted pully. A 20 um diameter optimizes easy delivery of fluid and minimizes the damage to oocytes. Only one needle is obtained from each micro-capillary.

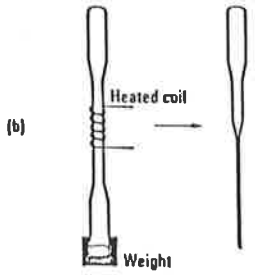
(c) The tip of the needle is broken off approximately 0.5-0.75 cm away from the first pull at an acute angle with watchmaker forceps.

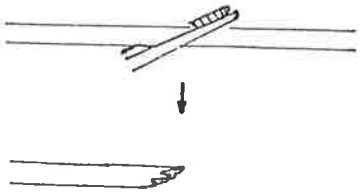
(d) Lengths and diameters of the fine and intermediate sections of a complete pipette.

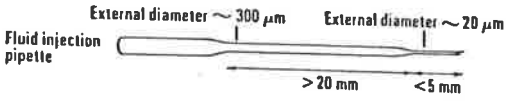
(e) The diameter of the expelled drop is measured by an eye piece graticule. The needle is then calibrated at 50 nl volumes on the pipette shaft with a fine (0.2 mm) texta drawing pen.

(figure adapted from Gurdon, 1974)


(a)
1 mm diameter
External diameter ~ 300 μm
Hand pull


(b)
Heated coil
Weight


(c)


External diameter ~ 300 μm
External diameter ~ 20 μm
Fluid injection pipette
> 20 mm
< 5 mm
(d)

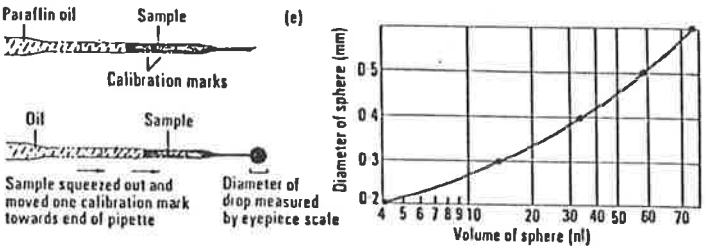
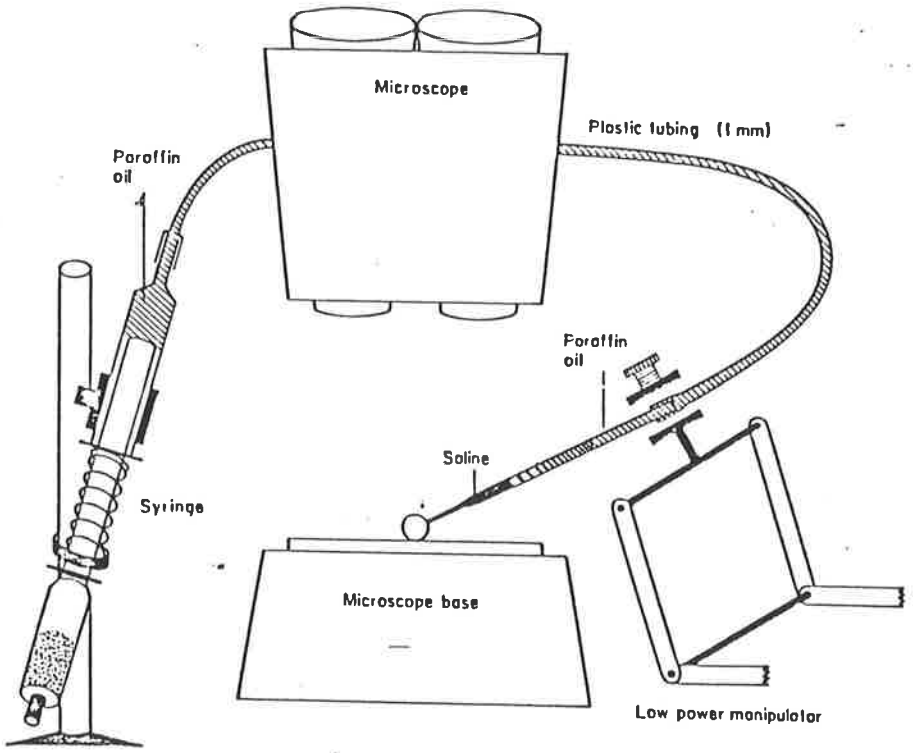

Paraffin oil
Sample
Calibration marks
Oil
Sample
Sample squeezed out and moved one calibration mark towards end of pipette
Diameter of drop measured by eyepiece scale
(e)
Diameter of sphere (mm)
0 5
0 4
0 3
0 2
4 5 6 7 8 9 10 20 30 40 50 60 70
Volume of sphere (nl)

Figure 8-2: Micro-injection equipment

A schematic representation of the micro-injection equipment is shown. (figure adapted from Gurdon, 1977)


Microscope
Plastic tubing (1 mm)
Paraffin oil
Paraffin oil
Saline
Syringe
Microscope base
Low power manipulator

and the needle tip have to remain within the same field of view.

During experiments the needle occasionally blocked and it was important not to put excessive positive or negative pressure on the sample as this would result in a spifflicating disaster. If repeated attempts to clear the needle failed a little of the tip of the needle was broken off with fine forceps; often the micro-pipette could be re-used. Of course, excessive trimming of a needle necessitated replacement.

(b) Preparation of DNA and oocytes for injection

As stated previously oocytes were usually used 24 hours after collection, DNA for injection was prepared by a clear lysate method (8-3-1), ethanol precipitated several times (all phenol must be removed), and resuspended in a buffer consisting of 88 mM NaCl, 10 mM Tris-HCl, pH 7.4. DNA was injected into the nucleus at a concentration of 200-400 ng/ul in a volume of 25-50 nl.

Initially, circularized, vector free insert was used for injection but later whole plasmid DNA was found to give identical results and was therefore routinely used. Circularization of insert DNA after purification through a low melting point agarose gel was done in a volume of 1 ml containing 50 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 1 mM DTT with the DNA at a concentration of 2 ug/ml and 2 units of T4 DNA ligase. This mixture was incubated overnight at 14 ^{0}C, ethanol precipitated, phenol extracted and re-ethanol precipitated before being resuspended in injection buffer.

A batch of 30 oocytes was usually used per DNA template injected. After injection the oocytes were generally incubated for 24 hours in Barths' saline at 20 ^{0}C.

(c) Nuclear injection

Because the oocyte is not transparent the nucleus cannot be seen. This

is only a minor problem as the nucleus occupies a consistent position in the oocyte and is readily injected with the following method.

Each oocyte was transferred from the petri-dish to a dry microscope slide (3-6 per slide) with a wide-mouthed pasteur pipette. Excess fluid was removed with the pasteur pipette as excessively wet oocytes were difficult to grasp with forceps. The slide was then transferred to the microscope stage.

Looking through the microscope the oocytes were individually manoeuvred with forceps and the end of the needle so the animal hemisphere was orientated towards the needle. The pipette was inserted at right angles to the oocyte equator exactly over the apex of the animal pole. The pipette was inserted about a quarter the depth of the oocyte. At the same time the oocyte was steadied with forceps. Once the needle had been positioned the syringe screw was turned to deliver the required volume by focussing on the calibrated shaft and watching the paraffin/aqueous meniscus. The oocyte was again grasped during needle withdrawal. The group of injected oocytes was then washed off the slide into another petri dish filled with fresh Barths' buffer.

(vi) Barths' saline

Solution A:	NaCl	206 gm
	KCl	3.0 gm
	$NaHCO_3$	8.1 gm
	Trizma base	73 gm

Made to one litre and pH to 7.6 with concentrated HCl.

Solution B:	$MgSO_4.7H_2O$	10.1 gm/500 ml
Soluition C:	$Ca(NO_3)_2.4H_2O$	3.9 gm
	$CaCl_2.H_2O$	3 gm

made to 500 ml.

Solution D: Penicillin 10 mg/ml

Streptomycin 10 mg/ml

Solutions A,B and C are stored at 4 ^{0}C. Solution D is stored frozen at -20 ^{0}C. To make Barths' saline, add 25 ml of A to 954 ml of water. Then add 10 ml of B and C and finally 1 ml of D.

8-3-16 Isolation of Xenopus oocyte RNA (Probst et al., 1979)

The oocytes were squashed in a loose-fitting glass homogenizer in a solution containing 10 mM Tris-HCl, pH 7.4, 1.5 mM $MgCl_2$, 10 mM NaCl, 1% SDS and 0.5 mg/ml proteinase K (0.5 ml of this solution was used per batch of oocytes). After incubation at room temperature for 25 minutes, EDTA was added to 10 mM and the mixture extracted three times with phenol/chloroform. The aqueous phase was made to 0.2 M NaCl and the oocyte RNA precipitated by the addition of 2.5 volumes of ethanol. Oocytes usually yielded 4-5 ug of total RNA each. They were stored dry at -80 ^{0} if the RNA extraction was not carried out immediately after incubation.

8-3-17 S1 nuclease analysis of RNA

For 5' analysis, DNA restriction fragments were treated with calf-intestinal phosphatase (8-3-8(i)) then labelled with γ-^{32}P-ATP using T4 polynucleotide kinase (8-3-12). For 3' end analysis, cohesive 3' termini of the appropriate restriction fragments were end-filled using DNA-polymerase I Klenow fragment in the presence of α-^{32}P-dNTPs (8-3-8(i)).

For S1 analysis both double stranded and single standed probes were used, though the single stranded fragments proved to be better reagents. Double stranded restriction enzyme fragments (figure 4-1) were heat denatured and annealed with an excess of M13 single stranded DNA (M135kb.B-parental genome and its reverse orientation; see 5-4) in a buffer containing 100 mM NaCl, 20 mM Tris-HCl, pH 7.4, 0.1% SDS at 60 ^{0}C to remove

the unwanted hybridizing strand. The unhybridized reagent strand was then purified by electrophoresis on a 6% non-denaturing polyacrylamide gel (8-3-4(i)).

Hybridisation reactions were carried out in a volume of 30 ul. The reaction mix contained 80% formamide, 40 mM PIPES, pH 6.4, 1 mM EDTA, 400 mM NaCl and incubations were at 50 ^{0}C for double standed probes and 30 ^{0}C for single stranded probes for 12-16 hours. Annealing reactions were diluted with 300 ul of ice-cold S1 buffer (200 mM NaCl, 2 mM $ZnSO_4$, 50 mM sodium acetate, pH 4.6) and 1,000-2,500 units of S1 nuclease added. Digestions were at 37 ^{0}C for 1 hour; the mixture was then phenol/chloroform extracted, nucleic acids precipitated with ethanol, and dissolved in formamide loading buffer and electrophoresed on 6% sequencing gels (8-3-11(iii)) prior to exposure.

8-3-18 Primer extension analysis of RNA (McKnight et al., 1981)

1 ng of each primer, 5' end-labelled (8-3-12), was separately or together added to the RNA equivalent to that of one injected oocyte (4-5 ug of total RNA) and ethanol precipitaed. The pellet was resuspended in 10 ul of 200 mM NaCl, 10 mM Tris-HCl, pH 8.3, heated to 70 ^{0} for 3 minutes then allowed to anneal at 42 ^{0}C for 1-3 hours. Following hybridization, the samples were incubated with reverse transcriptase (1 hour, 42 ^{0}C) in a volume of 34 ul containing 60 mM NaCl, 10 mM Tris-HCl, pH 8.3, 10 mM DTT, 500 uM dNTPs, 10 mM $MgCl_2$ and 8 units of enzyme.

Extension products were ethanol precipitated, washed in 70% aqueous ethanol and dried in vacuo. After resuspension in 5 ul of formamide loading buffer the extended primers were electrophoresed on 6% sequencing gels (8-3-11(iii)) and detected by exposure to X-ray film.

8-3-19 M13 site directed mutagenesis (Adelman et al., 1983; Chan and Smith, 1984)

(i) Formation, stabilization and enrichment of heteroduplex DNAs

For use as primers in the mutagenesis procedure, 10 pmoles of the CAT1 and H2B-box deleting synthetic oligonucleotides (figures 5-4 and 6-2 respectively) were each phosphorylated in the presence of 0.1 mM nonradioactive-ATP (8-3-12). For heteroduplex formation, 300 ng single-stranded M135kb.B-parental DNA (5-4) was heated to 70^{0} for 3 minutes and placed at room temperature for 10 minutes in a 20 ul volume containing 10 mM Tris-HCl, pH 7.4, 0.1 mM EDTA, 50 mM NaCl and 1 pmole (1 ul) of each phosphorylated deletion primer.

Primer extension was started by the addition of 30 ul of 50 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, 12 mM $MgCl_2$, 10 mM DTT, 0.7 mM ATP, 0.07 mM dATP, 0.2 mM each of dTTP, dCTP, dGTP and 2 units of DNA polymerase I Klenow fragment. After 30 minutes at room temperature, reaction mixtures were incubated for 4 hours at 37^{0}C, 2 units of T4 DNA ligase then added and incubated overnight at 4^{0}C. Heteroduplexes were recovered by phenol/chloroform extraction and ethanol precipitation then resuspended in 20 ul of water. 10 ul aliquots of the heteroduplexes were treated with 1000 units of S1 nuclease (8-3-17) for 5 minutes to digest single stranded parental M13 DNA, carrier tRNA (4 ug) was added and nucleic acid recovered after phenol/chloroform treatment by ethanol precipitation.

A 1:200 dilution of the S1 treated and non-S1 treated heteroduplexes were transformed into E. coli strain JM101 (8-3-10).

(ii) Phage screening (Benton and Davis, 1977)

Screening for mutant phage was done by in situ plaque hybridization using 100 ng of each of the ^{32}P-labelled deletion oligonucleotides (8-3-12). 100 of the S1 resistant recombinant M13 phage plaques obtained after bacterial transformation were picked in duplicate onto freshly plated JM101

on minimal plus glucose plates and grown overnight at 37 C. An unwashed, 5 cm nitrocellulose disc was layed onto one of the duplicate plates, orientation marks made with a needle, and when uniformly wet (1 minute later), peeled off and placed on to filter paper saturated with 0.5 M NaOH, 1.5 M NaCl for 2 minutes. The filter was then sequentially transfered to two filter papers soaked in 1.5 M NaCl, 0.5 M Tris, pH 8.0 for 4 minutes each.

The filter was air dried, baked at 80 C in vacuo for 1.5 hours then prehybridised in 4 ml of 0.9 x NET, 0.5% NP40 and 0.05% Blotto (0.9 x NET = 0.9 M NaCl, 0.09 M Tris-HCl, pH 8.0, 0.006 M EDTA; 1 x Blotto = 5% (w/v) non fat dry skim milk, 0.01% Antifoam A (Johnson et al., 1984)) in a sealed plastic bag at 65 C for 3 hours. Hybridization of the radioactively labelled primers to identify mutant phage was done in 4 ml of fresh 0.9 x NET, 0.5% NP40 and 0.05% Blotto at 42 C for 16 hours. Washing of the filter was done in 6 x SSC (1 x SSC = 0.15 M NaCl, 0.015 M tri-sodium citrate) twice for 5 minutes at room temperature, then at successively higher temperatures (45-65 C) each for 5 minutes following exposure to X-ray film.

Plaques that still hybridized strongly, over negative M135kb.B-parental controls included on the filter, at the high temperature washes were picked from the duplicate master plate and sequenced (8-3-11) with the H2B 26-mer to confirm the mutant phenotype. A mutation frquency of approximately 1% was observed using both the CAT1 and H2B-box deletion primers.

8-3-20 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

Adelman, J.P., Hayflick, J.S., Vasser, M. and Seeburg, P.H., DNA, 2, 183-193 (1983).

Agard, D.A. and Sedat, J.W., Nature, 302, 676-681 (1983).

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D., Molecular Biology of the Cell, Garland Publishing, Inc. New York and London (1983).

Allis, C.D., Ziegler, Y.S., Gorovsky, M.A. and Olmsted, J.B., Cell, 31, 131-136 (1982).

Ashburner, M. and Bonner, J.J., Cell, 17, 241-254 (1979).

Aviles, F.J., Chapman, G.E., Kneale, G.G., Crane-Robinson, C. and Bradbury, E.M., Eur. J. Biochem., 88, 363-371 (1978).

Ballinger, D.G. and Pardue, M.L., Cell, 33, 103-114 (1983).

Banaszuk, A.M., Deugau, K.V., Sherwood, J., Michalak, M. and Glick, B.R., Analytical Biochemistry, 128, 281-286 (1983).

Banerji, J., Olson, L., Schaffner, W., Cell, 33, 729-740 (1983).

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P., Nucl. Acids Res., 8, 127-142 (1980).

Benoist, C. and Chambon, P., Nature, 290, 304-310 (1981).

Benton, W.D. and Davis, R.W., Science, 196, 180-182 (1977).

Benyajati, C., Spoerel, N., Haymerle, H. and Ashburner, M., Cell, 33, 125-133 (1983).

Berezney, R., Cell Nucleus, 7, 413-456 (1979).

Bienz, M. and Gurdon, J.B., Cell, 29, 811-819 (1982).

Birchmeier, C., Grosschedl, R., Birnstiel, M.L., Cell, 28, 739-745 (1982).

Birchmeier, C., Folk, W. and Birnstiel, M.L., Cell, 35, 433-440 (1983).

Bird, A.P., Nature, 307, 503-504 (1984).

Birnboim, H.C. and Doly, J., Nucl. Acids Res., 7, 1513-1523 (1979).

Brandsma, J.A., Bosch, D., Backendorf, C. and van de Putte, P., Nature, 305, 243-245 (1983).

Breathnach, R. and Chambon, P., Ann. Rev. Biochem., 50, 349-383 (1981).

Brinster, R.L., Chen, H.Y., Warren, R., Sarthy, A. and Palmiter, R.D., Nature, 296, 39-42 (1982).

Brock, M.L. and Shapiro, D.J., Cell, 34, 207-214 (1983).

Brooker, J.D., May, B.K. and Elliott, W.H., Eur. J. Biochem., 106, 17-24 (1980).

Brown, A.L., Nature, 298, 793-794 (1982).

Brown, D.D., Cell, 37, 359-365 (1984).

Busslinger, M., Hurst, J. and Flavell, R.A., Cell, 34, 197-206 (1983).

Busslinger, M., Portmann, R., Irminger, J.C. and Birnstiel, M.L., Nucl. Acids Res., 8, 957-977 (1980).

Camato, R., Nicole, L. and Tanguay, R.M., Cold Spring Harbor Lab. of Quantitative Biology, Abstract of Papers Presented, 59 (1982).

Cantor, C.R., Cell, 25, 293-295 (1981).

Capecchi, M.R., Cell, 22, 479-488 (1980).

Carter, A.D., Felber, B.K., Walling, M., Jubier, M-F., Schmid, C.J. and Hamer, D.H., Proc. Natl. Acad. Sci. USA, 81, 7392-7396 (1984).

Cech, T.R., Cell, 34, 713-716 (1983).

Chan, V-L and Smith, M., Nucl. Acids Res., 12, 2407-2419 (1984).

Chandler, V.L., Maler, B.A. and Yamamoto, K.R., Cell, 33, 489-499 (1983).

Charnay, P., Treisman, R., Mellon, P., Chao, M., Axel, R. and Maniatis, T., Cell, 38, 251-263 (1984).

Choe, J., Kolodrubetz, D. and Grunstein, M., Proc. Natl. Acad. Sci. USA, 79, 1484-1487 (1982).

Ciejek, E.M., Tsai, M-J. and O'Malley, B.W., Nature, 306, 607-609 (1983).

Ciliberto, G., Raugei, G., Costanzo, F., Dente, L., and Cortese, R., Cell, 32, 725-733 (1983).

Clerc, R.G., Bucher, P., Strub, K. and Birnstiel, M.L., Nuc. Acids Res., 11, 8641-8657 (1983).

Clewell, D.B., J. Bacteriology, 110, 667-676 (1972).

Coles, L.S. and Wells, J.R.E., Nucl. Acids Res., 13, 585-594 (1985).

Coles, L.S., Ph.D. thesis, University of Adelaide, Adelaide, South Australia (1985).

Collins, F.S., Metherall, J.E., Yamakawa, M., Pan, J., Weissman, S.M. and Forget, B.G., Nature, 313, 325-326 (1985).

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P., Science, 209, 1406-1414 (1980).

D'Andrea, R., Harvey, R. and Wells, J.R.E., Nucl. Acids Res., 9, 3119-3128 (1981).

D'Andrea, R.J., Ph.D. thesis, University of Adelaide, Adelaide, South Australia (1985).

D'Andrea, R.J., Coles, L.S., Lesnikowski, C., Tabe, L. and Wells, J.R.E., Molec. and Cell. Biol., submitted (1985).

Darnell, J.E. Jr., Nature, 297, 365-371 (1982).

Davidson, E.H., Jacobs, H.T., Britten, R.J., Nature, 301, 468-470 (1983).

Davis, R.W., Botstein, D. and Roth, J.R., "Advanced Bacterial Genetics", Cold Spring Harbor Press, New York (1980).

Davison, B.L., Egly, J-M, Mulvihill, E.R. and Chambon, P., Nature, 301, 680-686 (1983).

Derman E., Krauter, K., Walling, L., Weinberger, C., Ray, M. and Darnell, J.E., Cell, 23, 731-739 (1981).

Di Domenico, B.J., Bugaisky, G.E. and Lindquist, S., Cell, 31, 593-603 (1982).

Dierks, P., Van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissmann, C., Cell, 32, 695-706 (1983).

Drew, H.R. and Travers, A.A. Cell, 37, 491-502 (1984).

Dumont, J.N., J. Morphol., 136, 153-180 (1972).

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L., Cell, 20, 313-319 (1980).

Elgin, S.C.R.,Nature, 309, 213-214 (1984).

EMBO dideoxy sequencing notes, In "Advanced Molecular Genetics " (1980).

Emerson, B.M. and Felsenfeld G., Proc. Natl. Acad. Sci. USA, 81, 95-99 (1984).

Everett, R.D., Baty, D. and Chambon, P., Nucl. Acids Res., 11, 2447-2464 (1983).

Falkner, F.G. and Zachau, H.G., Nature, 310, 71-74 (1984).

Farrell-Towt, J. and Sanders, M.M., Molec. and Cell. Biol., 4, 2676-2685 (1984).

Fitzgerald, M and Shenk, T., Cell, 24, 251-260 (1981).

Flavell, A., Nature, 305, 96-97 (1983).

Flynn, J.M. and Woodland, H.R., Developmental Biology, 75, 222-230 (1980).

Fromm, M. and Berg, P., J. Mol. Appl. Genet., 1, 457-481 (1982).

Galli, G., Hofstetter, H., Stunnenberg, H.G. and Birnstiel, M.L., Cell, 34, 823-828 (1983).

Gelinas, R., Endlich, B., Pfeiffer, C., Yagi, M. and Stamatoyannopoulos, G., Nature, 313, 323-325 (1985).

Gidoni, D., Dynan, W.S. and Tjian, R., Nature, 312, 409-413 (1984).

Gillies, S.D., Morrison, S.L., Oi, V.T., Tonegawa, S., Cell, 33, 717-728 (1983).

Gjerset, R., Gorka, C., Hasthorpe, S., Lawrence, J.J. and Eisen, H., Proc. Natl. Acad. Sci. USA, 79, 2333-2337 (1982).

Goldberg, M.L., Ph.D. thesis, Stanford University, Palo Alto, California (1979).

Grandy, D.K., Engel, J.D. and Dodgson, J.B., J. Biol. Chem. 257, 8577-8580 (1982).

Green, M.R., Maniatis, T and Melton, D.A., Cell, 32, 681-694 (1983a).

Green, M.R., Treisman, R. and Maniatis, T., Cell, 35, 137-148 (1983b).

Grosschedl, R. and Birnstiel, M.L., Proc. Natl. Acad. Sci. USA, 77, 1432-1436 (1980a).

Grosschedl, R. and Birnstiel M.L., Proc. Natl. Acad. Sci. USA, 77, 7102-7106 (1980b).

Grosschedl, R., Wasylyk, B., Chambon, P. and Birnstiel, M.L., Nature, 294, 178-180 (1981).

Grosschedl, R., Machler, M., Rohrer, U. and Birnstiel, M.L., Nucl. Acids Res., 11, 8123-8136 (1983).

Gurdon, J.B., In "Control of gene expression in animal development", Harvard University Press (1974).

Gurdon, J.B., Methods in Cell Biology, 16, 125-139 (1977).

Gurdon, J.B. and Melton, D.A., Ann. Rev. Genet., 15, 189-218 (1981).

Gurdon, J.B. and Wickens, M.P., Methods in Enzymology, 101, 370-386 (1983).

Guyette, W.A., Matuski, R.J. and Rosen, J.M., Cell, 17, 1013-1023 (1979).

Halleck, M.S. and Gurley, L.R., Experimental Cell R., 138, 271-285 (1982).

Harvey, R.P., Krieg, P.A., Robins, A.J., Coles, L.S. and Wells, J.R.E., Nature, 294, 49-53 (1981).

Harvey, R.P., Robins, A.J. and Wells, J.R.E., Nucl. Acids Res., 10, 7851-7863 (1982).

Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A., Lambda II, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory (1983).

Hentschel, C., Probst, E. and Birnstiel M.L., Nature, 288, 100-102 (1980).

Hentschel, C.C. and Birnstiel, M.L., Cell, 25, 301-313 (1981).

Herbert, E. and Uhler, M., Cell, 30, 1-2 (1982).

Hill, R.J. and Stollar, B.D., Nature, 305, 338-340 (1983).

Igo-Kemenes, T., Horz, W. and Zachau, H.G., Ann. Rev. Biochem., 51, 89-121 (1982).

Instructions to authors, Biochem. J., 169, 1-27, (1978).

Isenberg, I., Ann. Rev. Biochem., 48, 159-191 (1979).

Jacobson, A. and Favreau, M., Nucl. Acids Res., 11, 6353-6368 (1983).

Jakobovitis, E.B., Bratosin, S. and Aloni, Y., Nature, 285, 263-265 (1980).

Jeffreys, A.J., Cell, 18, 1-10 (1979).

Johnson, D.A., Gautsch, J.W., Sportsman, J.R. and Elder J.H., Gene Anal. Techn., 1, 3-8 (1984).

Jongstra, J., Reudelhuber, T.L., Oudet, P., Benoist, C., Chae, C-B, Jeltsch, J-M, Mathis, D.J. and Chambon, P., Nature, 307, 708-714 (1984).

Karin, M. and Richards, R.I., Nature, 299, 797-802 (1982).

Karin, M., Haslinger, A., Holtgreve, H, Richards, R.J., Krauter, P., Westphal, H.M., Beato, M., Nature, 308, 513-519 (1984).

Khoury, G. and Gruss, P., Cell, 33, 313-314 (1983).

Konarska, M.M., Padgett, R.A. and Sharp, P.A., Cell, 38, 731-736 (1984).

Korn, L.J., Nature, 295, 101-105 (1982).

Kornberg, R., Nature, 292, 579-580 (1981).

Kozak, M., Cell, 22, 7-8 (1980).

Kozak, M., Nature, 308, 241-246 (1984).

Kressman, A., Clarkson, S.G., Telford, J.L. and Birnstiel, M.L., Cold Spring Harbor Symp. Quant. Biol., 42, 1077-1082 (1977).

Krieg, P.A. and Melton, D.A., Nature, 308, 203-206 (1984).

Lacy, E., Roberts, S., Evans, E.P., Burtenshaw, M.D. and Costantini F.D., Cell, 34, 343-358 (1983).

Lassar, A.B., Martin, P.L. and Roeder, R.G., Science, 222, 740-748 (1983).

Lesnikowski, C., Honours thesis, University of Adelaide, Adelaide, South Australia (1983).

Levinger, L. and Varshavsky, A., Cell, 28, 375-385 (1982).

Liu, C-C., Simonsen, C.C. and Levinson, A.D., Nature, 309, 82-85 (1984).

Long, B.H., Huang, C-Y and Pogo, A.O., Cell, 18, 1079-1090 (1979).

Lusky, M. and Botchan, M.R., Cell, 36, 391-401 (1984).

Maxson, R., Cohn, R., Kedes, L. and Mohun, T., Ann. Rev. Genet., 17, 239-277 (1983a).

Maxson, R., Mohun, T. and Kedes, L., In "Eukaryotic Genes, Their Structure, Activity and Regulation" Editors, Maclean, N., Gregory, S.P. and Flavell, R.A., Butterworths, Chapter 16, 277-298 (1983b).

McBride, L.J. and Caruthers, M.H., Tetrahedron Letts., 24, 245 (1983).

McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D. and Felsenfeld, G., Cell, 27, 45-55 (1981).

McKnight, S.L., Gavis, E.R., Kingsbury, R., and Axel, R., Cell, 25, 385-398 (1981).

McKnight, S.L. and Kingsbury, R., Science, 217, 316-324 (1982).

McKnight, S.L., Kingsbury, R.C. Spence, A. and Smith, M., Cell, 37, 253-262 (1984).

Metterici, M.D. and Caruthers, M.H., J. Amer. Chem. Soc., 103, 3185 (1981).

Mirkovitch, J., Mirault, M-E. and Laemmli, U.K., Cell, 39, 223-232 (1984).

Montell, C., Fisher, E.F., Caruthers, M.H and Berk, A.J., Nature, 305, 600-605 (1983).

Moorman, A.F.M., de Boer, P.A.J., de Laaf, R.T.M. and Destree, O.H.J., FEBS Letters, 144, 235-241 (1982).

Mount, S.M., Nucl. Acids Res., 10, 459-472 (1982).

Nevins, J.R. and Wilson, M.C., Nature, 290, 113-118 (1981).

Nevins, J.R., Ann. Rev. Biochem., 52, 441-466 (1983).

Newport, J.W. and Kirschner, M.W., Cell, 30, 687-696 (1982).

Newrock, K.M., Cohen, L.H., Hendricks, M.B., Donnelly, R.J. and Weinberg, E.S., Cell, 14, 327-336 (1978).

Nordheim, A. and Rich, A., Nature, 303, 674-679 (1983).

Nurse, P., Nature, 302, 378 (1983).

Old, R.W. and Woodland, H.R., Cell, 38, 624-626 (1984).

Padgett, R.A., Mount, S.M., Steitz, J.A. and Sharp, P.A., Cell, 35, 101-107 (1983).

Palatnik, C.M., Wilkins, C. and Jacobson, A., Cell, 36, 1017-1025 (1984).

Palmiter, R.D., Brinster, R.L., Hammer, R.E., Trumbauer, M.E., Rosenfeld, M.G., Birnberg, N.C. and Evans, R.M., Nature, 300, 611-615 (1982).

Pardoll, D.M., Vogelstein, B. and Coffey, D.S., Cell, 19, 527-536 (1980).

Parker, M., Nature, 304, 687-688 (1983).

Parslow, T.G., Blair, D.L., Murphy, W.J. and Granner, D.K., Proc. Natl. Acad. Sci. USA, 81, 2650-2654 (1984).

Pelham, H.R.B., Cell, 30, 517-528 (1982).

Prentice, D.A., Loechel, S.C. and Kitos, P.A., Biochemistry, 21, 2412-2420 (1982).

Prior, C.P., Cantor, C.R., Johnson, E.M., Littau, V.C. and Allfrey, V.G., Cell, 34, 1033-1042 (1983).

Probst, E., Kressman, A. and Birnstiel, M.L., J. Mol. Biol., 135, 709-732 (1979).

Queen, C. and Baltimore, D., Cell, 33, 741-748 (1983).

Reeder, R.H., Roan, J.G. and Dunaway, M., Cell, 35, 449-456 (1983).

Richter, J.D. and Smith, L.D., Nature, 309, 378-380 (1984).

Rosenfeld, M.G., Mermod, J-J., Amara, S.G., Swanson, L.W., Sawchenko, P.E., Rivier, J., Wylie, W.V. and Evans, R.M., Nature, 304, 129-135 (1983).

Sanders, M.M., The Journal of Cell Biology, 91, 579-583 (1981).

Saragosti, S., Moyne, G. and Yaniv, M., Cell, 20, 65-73 (1980).

Schaffhausen, B., Critical Reviews in Biochemistry, 13, 215-265 (1982).

Schlesinger, M.J., Ashburner, M. and Tissieres, A., Heat Shock: From Bacteria to Man. Cold Spring Harbor, New York: Cold Spring Harbor

Laboratory (1982).

Scholer, H.R. and Gruss, P. Cell, 36, 403-411 (1984).

Shermoen, A.W. and Beckendorf, S.K., Cell, 29, 601-607 (1982).

Siliciano, P.G. and Tatchell, K., Cell, 37, 969-978 (1984).

Sittman, D.B., Graves, R.A. and Marzluff, W.F., Nucl. Acids Res., 11, 6679-6697 (1983).

Sommerville, J., Nature, 310, 189-190 (1984).

Spandidos, D.A. and Wilkie, N.M., Chapter 1, 1-48, In "Transcription and translation - a practical approach", Edited by Hames, B.D. and Higgins, S.J., IRL press (1984).

Stein, G.S., Stein, J.L. and Marzluff, W.F., Histone Genes: Structure, Organization, and Regulation, Wiley - Interscience, 1984.

Struhl, K., Nature, 305, 391-397 (1983).

Stuart, G.W., Searle, P.F., Chen, H.Y., Brinster, R.L. and Palmiter, R.D., Proc. Natl. Acad. Sci. USA, 81, 7318-7322 (1984).

Sures, I., Lowry, J. and Kedes, L.H., Cell, 15, 1033-1044 (1978).

Tanguay, R.M., Camato, R., Lettre, F. and Vincent, M., Can. J. Biochem. Cell Biol., 61, 414-420 (1983).

Travers, A., Nature, 303, 755 (1983).

Tsonis, P.A., Nature, 312, 314 (1984).

Urban, M.K., Franklin, S.G. and Zweidler, A., Biochemistry, 18, 3952-3960 (1979).

Voellmy, R. and Rungger, D., Proc. Natl. Acad. Sci. USA, 79, 1776-1780 (1982).

Von Holt, C., Strickland, W.N., Brandt, W.F., Strickland, M.S., FEBS Letters, 100, 201-218 (1979).

Walker, M.D., Edlund, T., Boulet, A.M., Rutter, W.J., Nature, 306, 557-561 (1983).

Wallis, J.W., Hereford, L. and Grunstein, M., Cell, 22, 799-805 (1980).

Wallis, J.W., Rykowski, M. and Grunstein, M., Cell, 35, 711-719 (1983).

Wang, S-W., Robins, A.J., D'Andrea, R. and Wells, J.R.E., Nucl. Acids Res., 13, number 4 (1985)

Weisbrod, S. and Weintraub, H., Proc. Natl. Acad. Sci. USA, 76, 630-634 (1979).

Weisbrod, S., Nature, 297, 289-295 (1982).

Weissmann, C., Nature, 311, 103-104 (1984).

West, M.W.P. and Bonner, W.M., Biochemistry, 19, 3238-3245 (1980).

Wickens, M.P. and Laskey, R.A., In "Genetic Engineering I" Edited by R. Williams, Academic Press London, 103-167 (1981).

Wigley, P.L., Sturm, R.A. and Wells, J.R.E., J. Mol. Biol., 181, 449-452 (1985).

Williams, D.A., Lemischka, J.R., Nathan, D.G. and Mulligan, R.C., Nature, 310, 476-480 (1984)

Wright, S., Rosenthal, A., Flavell, R. and Grosveld, F., Cell, 38, 265-273 (1984).

Wu, C., Nature, 286, 854-860 (1980).

Wu, C., Nature, 309, 229-234 (1984a).

Wu, C., Nature, 311, 81-84 (1984b).

Wu, H-M. and Crothers, D.M., Nature, 308, 509-513 (1984).

Wu, R.S. and Bonner, W.M., Cell, 27, 321-330 (1981).

Zachau, H.G. and Igo-Kemenes, T., Cell, 24, 597-598 (1981).

Zhong, R., Roeder, R.G. and Heintz, N., Nucl. Acids Res., 11, 7409-7425 (1983).

Zuker, C., Capello, J., Chisholm, R.L. and Lodish, H.F., Cell, 34, 997-1005 (1983).

Zweidler, A., In "Gene Families of Collagen and Other Structural Proteins". Eds. Prockop, D.J. and Champe, P.C., Amsterdam: Elsevier, 47-56 (1980).