



THE ISOLATION AND CHARACTERIZATION
OF CHICKEN HISTONE GENES

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

by

RICHARD PAUL HARVEY, B.Sc.(Hons.)

Department of Biochemistry,
University of Adelaide,
South Australia.

June, 1982.

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in 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.

RICHARD PAUL HARVEY

ACKNOWLEDGEMENTS

I wish to thank:

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

Dr. J.R.E. Wells, for his excellent supervision.

Dave Bird, Sue Clark, Bob Crawford, Paul Krieg, Rob Richards and Allan Robins for creating a friendly and stimulating atmosphere in which to work, for critical discussions and technical advice.

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

SUMMARY

THE ISOLATION AND CHARACTERIZATION OF CHICKEN HISTONE GENES

1. The work in this thesis is directed towards an understanding of the structure and the control of eukaryote genes. In particular, it describes aspects of the chicken histone gene system.

The histone gene system although complex, offers several relevant pursuits in the study of gene expression. The genes are reiterated, and both protein and DNA sequences have been shown to be highly conserved. The different histone types are coordinately expressed. The existence of cell-cycle and developmentally regulated variant sub-types, argues for nucleosomal heterogeneity and a role for histones in differential chromatin structure and perhaps control of gene expression. Questions relating to co-ordinate gene expression, gene evolution and the influence of chromatin architecture on gene control may therefore be addressed through a detailed analysis of histone genes.

2. A cDNA probe was prepared from mRNA derived from 5-day chick embryos and assessed for its usefulness as a vertebrate histone gene probe. This probe was used to screen a chicken recombinant genomic DNA library. Positive plaques were screened with a "negative" probe containing globin, ribosomal and 4S RNA sequences to eliminate recombinants selected with known contaminants of the cDNA probe. After plaque purification, there were four possible histone gene containing clones.

3. One recombinant (λ 7.4; λ CH-01) was chosen for further analysis. This clone was shown to contain histone genes on the basis of three criteria: used as probe, λ CH-01 could detect histone genes in total sea urchin DNA; λ CH-01 cross-hybridized with a sea urchin histone gene recombinant, λ 55; shotgun DNA sequencing revealed histone H2A gene sequences within a coding domain of λ CH-01.

4. The organization of the genes within λ CH-01 was determined using homologous and cross-species gene-specific DNA probes. Genes detected with embryo cDNA which did not hybridize with the gene-specific probes available were identified by DNA sequencing. The overall arrangement of histone genes within λ CH-01 was distinctly disordered.

5. Another recombinant (λ 1-6; λ CH-02) was also shown to contain histone genes and the gene arrangement was determined. A disordered situation was also found for this clone.

6. Southern blotting analysis was performed to confirm the observations of "disordered arrangement" of chicken histone genes. No evidence for a conserved repeating cluster of histone genes was obtained from Southern blots of chicken DNA.

7. To examine the microstructure of chicken histone genes, two H2B genes were completely sequenced. The two genes coded for the same H2B protein sub-type yet were divergent

in nucleic acid sequences outside the protein coding portion. The implication to the evolution of histone genes is discussed.

A sequence of 9 bps was found in the 5'-region of both H2B genes which is conserved in all histone H2B genes that have been sequenced. This sequence is therefore implicated in the control of transcription of these H2B genes.

Other flanking sequences common to histone and other eukaryote genes were recognized, and their presence is discussed.

8. Based on the observations made on chicken histone genes, a proposal is put forward to explain the evolution of histone genes from a highly ordered to a considerably disordered state.

	<u>Page</u>
1.6 The Histone Gene System	34
1.6.1 The Histone Proteins	34
1.6.2 Histone Modifications	34
1.6.3 Primary structure subtypes	36
1.6.4 Histone mRNA	37
1.6.5 The Histone Genes	39
1.7 Aims of the Project	44

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials	46
2.2 Methods	47
2.2.1 Preparation of RNA	47
2.2.2 Restriction enzyme digestion and analysis of DNA	49
2.2.3 <i>In vitro</i> synthesis of labelled DNA	51
2.2.4 Subcloning of DNA fragments	53
2.2.5 Large-scale preparation of recombinant plasmid DNA	54
2.2.6 Isolation of clones from a recombinant gene library plating and screening (Benton and Davis, 1977)	57
2.2.7 Gilbert and Maxam DNA sequencing procedures	59

CHAPTER 3: APPROACHES TO THE ISOLATION OF CHICKEN HISTONE GENES - SELECTION OF RECOMBINANTS

3.1 Introduction	62
3.2 Results	65
3.2.1 Preparation of chick embryo RNA	65
3.2.2 Synthesis and use of 5 day embryo 7-11S cDNA	68
3.2.3 Primary selection of recombinants	70
3.2.4 Negative screening and plaque purification	72

CHAPTER 4: IDENTIFICATION OF A CHICKEN HISTONE GENE RECOMBINANT

4.1 Introduction	75
--------------------------	----

CHAPTER 8: THE MICROSTRUCTURE OF CHICKEN
HISTONE GENES

8.1	Introduction	105
8.2	Results	105
8.2.1	Sequencing procedures using the chemical degradation method	105
8.2.2	The H2B gene of pCH3.3E	111
8.2.3	The 5'-leader region of the pCH3.3E H2A gene	114
8.2.4	DNA sequencing procedures using the enzymatic chain termination method	115
8.2.5	The H2B gene of pCH4.6E	118
8.3	Discussion	121

CHAPTER 9: FINAL DISCUSSION

9.1	Introduction	131
9.2	Chicken histone genes	132
9.3	Histone variant genes	132
9.4	Chicken variant-histone genes	134
9.5	The evolution of "early" histone genes	136
9.5.1	Homogenization of reiterated genes	136
9.5.2	Histone gene number	138
9.5.3	Histone gene organization and evolution	140
	POSTSCRIPT	143
	BIBLIOGRAPHY	146

CHAPTER 1

INTRODUCTION



1.1 INTRODUCTION

The developing embryos of higher eukaryote organisms contain a myriad of cell types, each following a specific program of maturation. Each cell type finally reaches its fully-mature, specialized, "differentiated" state which has precise function in the organism as a whole. It is the nature of this process, whereby like cells in an embryo, diversify and multiply to become morphologically and functionally distinct cells in the adult, which remains one of the fundamental questions in modern biology. Experimental evidence suggests that "differentiation" occurs with precise programming of gene activation and inactivation. To understand exactly how selective gene expression is able to control the development of a single cell into a multi-cellular organism, we must understand how individual genes are structured and how such structure has bearings on control.

The work presented in this thesis focuses on this aspect of the process of differentiation. In particular, the structure and control of histone genes in the chicken. Although histone genes are not the specific end-product of any differentiated cell, their synthesis is regulated during the development of many cell types. Because histones have such a precise role in the structure of chromatin, selective regulation of histone variants may also influence the program of gene activity in those cells.

The experimental work in this thesis has a narrow perspective, however, this introduction considers the

process of differentiation in a wider context. It reviews evidence for the "variable gene activity" theory of differentiation and, at a more molecular level, current knowledge on the structure of eukaryote genes. Aspects of histones and histone genes are presented in detail, and finally the more specific aims of this project are discussed.

1.2 THE VARIABLE GENE ACTIVITY THEORY OF CELL DIFFERENTIATION

Two premises are required to establish the view that differentiation is a function of variable gene activity. The first is the now well accepted relationship between the nucleotide sequence of the DNA in the genome, and the amino acid sequence of cellular proteins. Because the functional and morphological characteristics of a cell are specified by its proteins, differentiation ultimately depends upon transcription of genomal information. Similarly, because DNA is the hereditary material, programs of differentiation are ultimately dictated by events in the nucleus.

The second premise is that every cell nucleus in an adult contains an equivalent complement of DNA. This distinguishes the notions of selective gene activity in differentiation, from possible mechanisms involving selective loss or amplification of genes.

Initial indications that differentiated cells still contained information which could specify the characteristics of other cell types was seen in the phenomenon of

"transdifferentiation". This is where, during normal development, cells change their specialized role and assume a new state of differentiation. For example, Stone (1950) showed that in the regenerating newt eye, neural retinal cells derive directly from pigment cells.

More elegant demonstrations of a constant genome come from nuclear injection studies. Gurdon (1962) reported that normal swimming tadpoles could be raised from enucleated *Xenopus* eggs which had been injected with a nucleus derived from a differentiated intestinal cell. This demonstrated that the information contained within a differentiated cell was sufficient to specify the development of the whole organism. More recently, nuclei derived from primary cultures of adult *Xenopus* kidney, lung, heart, testis and skin have been shown to give rise to normal tadpoles (Laskey and Gurdon, 1970). In general then, the differentiated state, despite its stability, is in fact, fully reversible, and information is not lost during development. Examples of sequence amplification (ribosomal genes in *Xenopus laevis*, Brown and Blackler, 1972), loss (enucleation in mammalian erythroid cells) or rearrangement (immunoglobulin genes, Hozumi and Tonegawa, 1976) are therefore exceptional cases.

The developmental process can be split into two convenient, but possibly arbitrary parts: determination and differentiation. Determination is the point at which a cell or group of cells becomes committed to a specific developmental pathway. Although not yet specialized,

the fate of their descendants is precisely defined. Differentiation is the process through which a determined cell changes into a specific cell type, in which specialized functions are expressed.

1.3 DETERMINATION

Determination has been defined by Hadorn (1965) as "a process which initiates a specific pathway of development by singling it out from among the various possibilities for which a cellular system is competent".

1.3.1 Accumulative Commitment

The complete organism develops from a single fertilized ovum. This zygote clearly has the potential to become committed to any developmental program. A particular program is initiated by the commitment of a cell to a broad path of development. During development however, cells become committed to more and more specific developmental sub-programs. In general, a normal cell always produces daughter cells that have a level of determination identical to or more specific than its own state.

A good example of an increasingly specific developmental pathway is the production of erythrocytes from their progenitor cells (review, Ingram, 1974). After several divisions of the zygote, one division gives rise to a precursor hematopoietic stem cell that is committed to produce the three major blood cell types; white blood

cells, platelet-producing cells, and red blood cells. A progeny cell of this stem cell subsequently divides to produce another stem cell committed specifically to differentiate into a mature erythrocyte. This occurs under the influence of the hormone erythropoietin.

The stage of embryonic development at which the reduction of a cell's developmental options is first observed, differs from species to species. Any one of the 16 cells of the hydromedusa blastula has the capacity to produce a complete adult organism (Wilson, 1925). No commitment to specific developmental pathways has occurred at this stage. After the next division however, none of the resultant 32 cells are capable of producing an adult.

In contrast the nematode zygote divides to produce two cells already limited in their developmental options; the anterior cell giving rise to primary ectoderm, while the posterior cell is the precursor of the germ-line tissue (Nigan *et al.*, 1960; Hirsh and Vanderslice, 1976).

Several stimuli which result in the determination of cells in the embryo have been recognized. These include cytoplasmic localization, intra-cellular interactions, extra-cellular influences, and specific transcription. It remains an axiom that whatever the mechanism of initiation of development, it is ultimately a result of actual gene activity in the embryo or the potential is generated by prior gene activity in the developing oocyte.

1.3.2 Cytoplasmic Localization

"Localization" is a widespread mechanism whereby signals for specific programs of development are sequestered in the egg cytoplasm. As the early embryo cells (blastomeres) divide up the egg cytoplasm, they appear to inherit "instructions" for various types of cell differentiation.

The early egg and embryo exhibit two poles distinguishable by their morphology and pigmentation. Early cleavages usually occur with a characteristic orientation to these poles, and the fate of the resulting blastomeres depends on the portion of the cytoplasm which they receive. For example the amphibian germ cell stem line originates early in cleavage from blastomeres forming at the vegetal pole of the egg. Smith (1966) showed that destruction of the vegetal pole cytoplasm by ultraviolet irradiation before migration of the cleavage nucleus into that region, always resulted in normal but sterile adults. Irradiation of the animal pole cytoplasm however produces no defects. Smith also showed that after vegetal pole irradiation, micro-injection of vegetal cytoplasm from unirradiated eggs was able to compensate for the destroyed cytoplasmic factors and produce embryos containing germ cells.

In *Chironomus*, division of the egg results in cells destined to become part of the posterior or anterior lineages. Yajima (1960) has produced symmetrical embryos by centrifugation of eggs. This is interpreted to be an

effect of equal diffusion of egg cytoplasm towards both poles of the egg after centrifugation has stopped, thus eliminating any effective sequestering of determinants.

Davidson *et al.*, (1965) and Britten and Davidson (1971) propose explanations of the localization phenomenon based on embryo gene activation. According to these interpretations, regulatory molecules or mRNAs synthesizing such molecules are topographically localized in the egg cytoplasm. In *Smittia* eggs, it is indeed likely that maternal RNA is the determining factor as introduction of ribonuclease into the cytoplasm induces symmetrical mutants (Kandler-Singer and Kalthoff, 1976). This view of localization implies that gene activity in the developing egg plays a latent role in the early development of the embryo.

1.3.3 Inter-cellular Interaction

The initial determination of a cell may depend upon its position in the embryo. Classical embryologists have traced cell lineages to recognise primordial stem cells and have established that these cells occupy specific locations in the embryo. This mechanism has been studied mostly in mammalian embryos where cytoplasmic localization does not seem to occur.

In preimplantation mammalian embryos only two forms of differentiation are evident. Cells of the early blastula may differentiate into either extraembryonic

trophoblast or inner-cell mass. Only the inner-cell mass goes on to form the actual embryo. Clearly no determination has occurred in the preimplantation precursors to these cells, as one group of cells or the other can be completely removed with^{-out} affecting the development of the embryo (Lin, 1969; Gardner, 1971). Ultimate differentiation into trophoblast as opposed to elements of the internal cell mass however, depends on whether blastomeres are on the inside or the outside of the embryo. Hillman *et al.*, (1972) has shown that blastomeres of the 4-cell embryo will always form trophoblast when placed on the outside of other embryos. Conversely when completely surrounded by embryo cells, they always form inner-cell mass. The complete engulfment of the cells destined to become inner cell mass is essential as partially exposed inner cells or inner cells exposed to a glass interphase will always form trophoblast tissue (Stern, 1973).

Elegant studies using allophenic mice have shown that several cells in the embryo can be determined independently and identically (Mintz, 1970). Allophenic mice are produced by combining two genetically different embryos at the eight cell stage before individual blastomeres have become determined. These embryos, upon reimplantation into a female, will give rise to chimeric mice made up of two genotypes. Phenotypic markers can be used to identify the donor origin of different adult tissues. Most tissues contain both geno-

types indicating that more than a single progenitor cell is determined independently to follow similar or identical developmental pathways. For example, allophenic mice from black and white donor mice show 34 alternating black and white stripes of coat colour. This suggests that 34 cells along the neural crest of the embryo are independently determined and maintain their topological relationship to each other during development.

In summary, the positioning of cells in the embryonic architecture influences the developmental program to which they become committed. The signal for determination may be directed simultaneously at a group of specifically positioned cells. This is evident at the very first determination event in mammalian embryos. In non-mammalian embryos, pattern formation and embryonic morphogenesis frequently occurs by means of interaction between cell lineages which were initially determined by means of cytoplasmic localization in eggs and early embryos.

1.3.4 Extracellular Interactions

Although ultimately the control of differentiation is nuclear, in multicellular animals nuclear differentiation is often seen to be regulated by extracellular signals. Control of the ionic environment for instance, may play a crucial role in the initiation of new programs of gene expression. Exposure of porcine lymphocytes to

the calcium ionophore A23187, causes commitment to lymphoblast transformation of more than 70% of cells in 24 hours. This can also be achieved with Concanavalin A. Prior to the commitment induced by ConA, an influx of calcium ions from the medium is observed. When calcium ions are chelated with EGTA, lymphocytes no longer undergo the transformation (Henske *et al.*, 1977). A similar influx of calcium ions precedes commitment in the Friend cell system (Levenson *et al.*, 1980).

Perhaps the most simplistic mechanism of initiation of differentiation to be studied is that of aggregation in the slime mold *Dictyostelium discoideum* (review, Robertson and Grutsch, 1981). This ameboid organism becomes multicellular for part of its life-cycle where the unicellular form aggregates and differentiates to form a fruiting body. Both the aggregation and the initiation of the differentiation program occur in response to cAMP. Cells respond chemotactically to short pulses of cAMP secretion at specific concentrations and migrate towards the aggregative centre. Cells also respond to cAMP with their own pulse of cAMP synthesis and secretion. The program of differentiation, triggered also by cAMP is independent of aggregation. Patterns of protein synthesis normally associated with aggregation occur in response to cAMP when aggregation is inhibited (Parish, 1979).

Although no morphogen has been identified in the embryos of multicellular organisms, provocative facts

suggest that the concentration dependence and dynamics of the slime mold cAMP signal, may represent a special case of a universal mechanism. For example in the chick embryo, contact formation and perhaps cell movement can be controlled by a propagated cAMP signal (Gingle, 1977).

Further understanding of the action of extracellular factors in determination may be precipitated by the study of tumour promoters. The most potent of these are the phorbol esters, which are not carcinogenic in themselves, but can induce tumour growth after a subthreshold dose of carcinogen. One action of phorbol esters is to delay differentiation and may do so by borrowing the receptor system for natural cellular signals (Weinstein and Wigler, 1977).

1.3.5 Nuclear Events

One strategy used to study the signal that initiates development is a clonal assay approach (review; Levenson and Housman, 1981). After exposure to a signal that induces differentiation, cells are cloned from the population in the absence of the stimulus. The colonies observed are only one of two basic types: all cells in a colony are undifferentiated, or all cells express all features of the terminal differentiated state. The implication of such results is that a single cellular event occurs when the stimulus is added and the cells become committed. The introduction of inhibitors of macromolecular synthesis

into the clonal assay has been used to determine the nature of the signal. In Concavelin A stimulated lymphocytes, by exposure of cells to drugs for various times and examination of the proportion of committed cells in the absence of drugs, a mRNA with a short half-life (3-4 hr.) has been implicated as the transformation signal (Milner, 1978).

The homoeotic mutants of *Drosophila*, in which there is an abnormal substitution of one adult body part for another, also suggest that the signal for developmental programs are single transcription event. Most homoeotic mutations are located in a small region of the third chromosome, and analysis has shown that single gene defects can switch one developmental program to another.

One particularly well characterized homoeotic gene system is the bithorax locus (review; Lewis, 1978) and its controlling gene *esc* (Struhl, 1981). Lewis has proposed a model for the action of the bithorax complex, where at least eight genes normally promote particular programs of thoracic and abdominal segment development. The ground state represents no gene expression, and corresponds to the meso-thoracic segment. In each of the successive abdominal segments one additional gene is expressed so that by the eight abdominal segment, all genes are expressed. The continued expression of sets of bithorax genes maintain the pathways of development associated with each body segment. Mutations in the bithorax genes produce phenotypes with body segments

in altered states of differentiation. For example if the first bithorax gene is inactive then all body segments resemble the mesothorax. Similarly, if the eighth gene is mutant, then the eighth abdominal segment resembles the seventh.

The *esc* gene, acting only in early development, has been observed to control the selective expression of sets of bithorax genes (and other regulatory genes) during thoracic and abdominal development. In general mutations in the *esc* gene allow expression of all of the bithorax genes. All segments therefore, resemble the eighth. In double *esc* and bithorax mutations, all segments resemble the eighth segment of the equivalent bithorax only mutation.

Three levels of developmental control have therefore been recognized in this system. The *esc* gene active in the early embryo (commitment), regulates the correct expression of the bithorax genes (development), which maintain each body segment to the correct program of gene expression (differentiation).

1.4 DIFFERENTIATION

The process of cell differentiation is, in general, achieved by variable gene expression. This section reviews the direct evidence for this assertion together with the exceptional cases of gene movement and examples of post transcriptional control.

1.4.1 Evidence for Transcriptional Control

Evidence has been presented above (section 1.2) suggesting that the DNA content of all of the somatic cells of an organism is identical and that there is little evidence that gene amplification, deletion or rearrangement plays a significant part in the process of cell differentiation. Observations from a number of other studies suggest that transcriptional control is the major mode of gene expression.

Hybridization experiments have shown that in general the RNAs of a differentiated cell type hybridize with less than 10% of single copy sequences in the genome. For example, mouse liver RNA hybridizes to 2-5% of mouse single copy DNA (Hahn and Laird, 1971; Brown and Church, 1972). Similar figures of 4% are obtained for rat ascites cells (Holmes and Bonner, 1974), and 0.6-0.9% for *Xenopus* oocytes (Davidson and Hough, 1971).

Hybridization experiments using RNA from different cell types suggest that qualitatively diverse sets of RNA molecules exist in different cell types of the same organism. For example, there is greater than 70% difference between non-repetitive transcripts of mouse liver, spleen and kidney (Grouse *et al.*, 1972). This indicates that overlapping but clearly distinct sets of RNA sequences are synthesized in various differentiated cell types. *In vitro* chromatin transcription experiments support these results (review, Bonner *et al.*, 1968). Only a small fraction of the DNA is transcribed by added RNA

polymerase. Template activity increases, however, as the chromatin is deproteinized.

Evidence of transcriptional control of gene expression can be obtained from the cytological examination of dipteran polytene chromosomes. The polytene puffs have been shown to be sites of intense RNA synthesis, and these features can be localized to specific chromosomal regions, varying according to the state of differentiation (Delling, 1964; McKenzie *et al.*, 1975). A good example is the Balbiani rings of *Chironomus* salivary gland chromosomes, (Daneholt and Hosick, 1974). The synthesis of specific RNA and protein products has been associated with specific puffs and three of these chromosomal puffs are responsible for about 80% of the RNA synthesis in the cell, while apparently representing only a few percent of total genomic DNA. Both the secretory peptides arising from the puff transcripts and the puffs themselves are specific to only certain salivary gland cells, indicating that the activity of a limited set of genes is responsible for the differentiated function of these cells.

A number of developmental systems have been studied where the rate of synthesis of a protein product may be followed from a low or undetectable level to a high level characteristic of the mature differentiated cell type. Measurements on haemoglobin, (Hunt, 1974) and ovalbumin, (Rhoads *et al.*, 1973) indicate that the increase in the concentration of the specific protein is wholly due to

an increase in the concentration of the respective mRNA. In the uninduced chicken oviduct and other cell types such as liver ovalbumin mRNA is present in extremely low concentrations (Harris *et al.*, 1975; Axel *et al.*, 1976). In the chick fibroblast, ovalbumin mRNA and globin mRNA are absent (Groudine and Weintraub, 1975). Thus, when these mRNAs are not in polysomes, transcription of the structural genes is repressed.

1.4.2 Gene Amplification

Experiments have been performed to measure the copy number of specific genes to see whether the regulation of their mRNA synthesis is achieved by gene amplification. In many systems, including globin (Harrison *et al.*, 1974) and ovalbumin (Sullivan *et al.*, 1973), it was found that no difference occurred in the number of genes present in cells in which the protein is normally synthesized, and other cell types normally not expressing the protein.

In a few systems, however, gene amplification has been shown to play the major role in regulation of the synthesis of RNA. The most well-known case is that of the ribosomal cistron amplification in amphibian oocytes (Gall, 1968) to augment ribosomal RNA synthesis during rapid development in oogenesis. While this mechanism is not unique to amphibia, and is observed in a number of different phyletic groups (Hourcade *et al.*, 1974), it is far from being a general biological mechanism for ribosomal RNA regulation. In the chicken,

for example, DNA extracted from a wide range of tissues always contain the same number of ribosomal cistrons, even though the rates of ribosomal RNA synthesis differ markedly between these tissues (Ritossa *et al.*, 1966).

In a number of other systems, when cells are placed under "abrupt selection" pressure they can rapidly synthesize detoxifying products by amplification of the appropriate genes. In the presence of methotrexate and N-phospho-nacetyl-L-aspartate, cells amplify the genes for dehydropholate reductase and transcarbamylase respectively (Alt *et al.*, 1978; Wahl *et al.*, 1979). Similarly in the presence of cadmium, Friend leukemia cells amplify the genes for metallothionein-I genes (Beach and Palmiter, 1981). Evidence suggests that this is a single copy gene in normal mouse cells. Under cadmium selection, however, there is a three-fold increase in the rate of synthesis of metallothionein-I mRNA, and this correlates well with the 3-5 fold amplification of the genes.

While mechanisms do exist for the amplification of genes to regulate transcription, the extreme selective pressure placed on each system suggests that these mechanisms are not generally used in the normal course of cell development.

1.4.3 Gene Rearrangement

Hozumi and Tonegawa, (1976), using restriction and hybridization analysis of mouse DNA, have demonstrated

that the DNA coding for the lambda light chain immunoglobulin protein is not in the same arrangement in cells synthesizing immunoglobulin as it is in germ-line DNA. These observations were verified by cloning and sequencing experiments (Bernard *et al.*, 1978), and have since been extended to the DNA coding for kappa light chain proteins (Max *et al.*, 1979; Sakano *et al.*, 1979).

The immunoglobulin light and heavy chain molecules consist of two regions, a constant (C) region and a variable (V) region, and in each case the V-region is involved in antigen recognition. Restriction enzyme analysis and cloning experiments showed that the DNA sequences coding for the V and C-regions of kappa and lambda chains were well separated in DNA from germ-line tissue, but closer together (1.2 Kb in lambda light chain genes), in DNA from tissue synthesizing antibodies. The final gene still includes intervening sequences within the coding region of the lambda and kappa chains, (Brack and Tonegawa, 1977; Max *et al.*, 1979, respectively), however, the V and C-regions are now apparently included in the same RNA transcript, and the intervening sequences are removed by RNA splicing (review, Abelson, 1979). Thus, DNA rearrangement during development appears to be necessary to form the functional gene in the differentiated anti-body producing cell. This may have evolved as a mechanism of generating at least part of the remarkable variability required by the immune system.

1.4.4 Movable Genetic Elements

Mobile genetic elements have been identified in maize (McClintock, 1956), yeast (Cameron *et al.*, 1979) and in *Drosophila* (Finnegan *et al.*, 1978). Examples of known Mendelian genes, and disperse repeated gene families have been observed, together with other elements of unknown nature. On the basis of genetic characterizations it can be concluded that in *Drosophila*, mutants are produced at specific loci at high frequency, as a result of transposition of genetic elements. Although it is not known whether this is a random or directional effect, there is evidence that it may occur under the influence of disperse repetitive gene sequences which are found to abound in the *Drosophila* genome. Among these are the "copia" sequences (Finnegan *et al.*, 1978). These genes code for an abundant mRNA present in a variety of cells, and studies on morphologically and cytologically distinct *Drosophila* isolates, indicate that they are capable of transposition from one chromosomal location to another.

Although it has been predicted that mobile sequences occur in all eukaryote genomes (McClintock, 1956), there is no evidence to suggest that they have an active role in control of differentiation or are representative of a general mechanism of control. In a number of gene systems including rabbit globin (Jeffreys and Flavell, 1977), chicken ovalbumin (Breathnach *et al.*, 1977) and chicken keratin genes (Saint, 1979), restriction enzyme

studies have shown the arrangement of DNA sequences coding for the protein to be identical in tissues actively synthesizing the protein, and in tissues in which the protein is not produced. Therefore, expression must involve the activation of pre-existing control regions rather than insertion or rearrangement of control regions adjacent to the genes.

1.4.5 Post-transcriptional Control

Post-transcriptional control of gene expression may occur at the level of mRNA maturation, recognition of mRNA by the translation machinery and post-translational modification (review, Revel and Groner, 1978). Untranslated mRNAs are found as ribonucleoproteins (mRNPs) in the nucleus and cytoplasm of many cells, indicating that a direct control is placed on mRNA translation as well as transcription. One obvious example of this is the stored maternal mRNA in sea-urchin eggs (Gross *et al.*, 1973). They contain a dominant inhibitor of aminoacyl tRNA binding (Gambino *et al.*, 1973). It is not until fertilization of the egg and subsequent development of the embryo, that the concentration of the inhibitor diminishes to allow translation of the stored mRNA (Hille, 1974).

The control of globin expression in reticulocytes is a system which exhibits both transcriptional and translational regulation during development. When nucleated, reticulocytes synthesize equal amounts of

α and β globin chains. Transcription is presumably coordinately regulated, however details of the mechanism are not known.

Translation of globin mRNA is under the control of a translation inhibitor, the activation of which is controlled by hemin (review, Revel and Groner, 1978). Without hemin the (met-tRNA_f.40S) ribosomal subunit is lost from the initiation complex. The inhibitor is a protein kinase and acts by phosphorylating the smallest subunit of the trimeric initiation factor eIF-2. The biological significance of this mode of control is not entirely clear, but it may represent a mechanism whereby temporary suppression of protein synthesis can be achieved under conditions of energy starvation. Lodish (1976) has postulated that under such conditions translation of mRNAs most important to a particular cell type would be more resistant to the inhibitory effects.

Thus, post-transcriptional gene regulation can be seen to be a mechanism of "fine-tuning" the control of gene expression on a transcriptional level.

1.5 LEVELS OF TRANSCRIPTIONAL CONTROL

The advent of recombinant DNA technology has brought about a new era in biochemical, medical and genetic research. The application of this technology to the question of eukaryote gene structure and function has resulted in rapid progress in understanding. The ability to produce probes representative of specific gene

sequences has enabled the examination of eukaryote genes and their control at the level of the DNA sequence, chromosomal organization and cellular location.

This section reviews the current knowledge on the structure of eukaryotic genes and the levels on which gene transcription is controlled.

1.5.1 The Britten and Davidson Model of Gene Control

The most comprehensive model which attempts to explain the control of eukaryote genes is that of Britten and Davidson (1969). It is basically an extension of the prokaryotic operon model of Jacob and Monod (1961), and is designed to account for two distinctive features of eukaryotes: the temporally controlled expression of different sets of genes during development, and the organization of repetitive and unique sequences in the genome.

The sequence of gene expression is postulated to result from a cascade of positive regulatory events at a transcriptional level. In such a sequence, each activated group of structural genes produces one or more regulatory proteins that activate subsequent groups of structural genes. For groups of structural genes, together with their regulator gene to be coordinately expressed, requires that all of these genes have a common controlling element adjacent to those genes. Every set of coordinately expressed genes would need to have a different controlling element. Expression of a gene on

more than one occasion would require more than one control element adjacent to it, each corresponding to the set of genes expressed at different stages.

Since these controlling elements will be present next to more than one gene, these sequences will appear as moderately repeated sequences in the genome. While there is no direct evidence for the model, there is considerable data showing that short, moderately repeated sequences are interspersed between larger unique sequences in the eukaryote genome (Davidson *et al.*, 1973; Schmid and Deininger, 1975), as proposed by the model. Extensive DNA sequencing data on groups of coordinately expressed genes should provide information with which the model may be further evaluated (see below).

1.5.2 Intervening Sequences

The interruption of genes with DNA sequences (introns) which do not code for part of the protein product was first detected in the 28S ribosomal RNA genes of *Drosophila* (Glover and Hogness, 1977; White and Hogness, 1977). It is now obvious that the majority of eukaryotic genes are arranged in this way, with introns splitting each gene into separate exons. Introns are mostly localized in the region coding for the amino acids however they have also been found in the 5'-untranslated regions of insulin (Lomedico *et al.*, 1979), and ovalbumin (Breathnach *et al.*, 1978) genes. There is no apparent rule regarding the number of introns and exons.

Some genes contain only a single intron (e.g. yeast actin gene; Galwitz and Sures, 1980), while others are highly complex. The chick α -2 collagen gene, for instance, contains 50 introns (Ohkubo *et al.*, 1980; Wozney *et al.*, 1980). In the mouse dihydrofolate reductase gene (Nunberg *et al.*, 1980), introns account for 95% of the DNA sequence.

Split genes are transcribed into colinear precursors from which the intron transcripts are removed by a cellular splicing mechanism, the details of which are unknown.

The function and origin of introns has been the subject of much speculation (review, Breathnach and Chambon, 1981). Popular interpretations suggest that RNA splicing has evolved to facilitate the duplication of genes by unequal crossing-over events, and the evolution of new genes by fusion of ancestral genes coding for specific protein domains.

Whatever the evolutionary significance of RNA splicing, in a few cases alternative modes of splicing have evolved to specify different transcripts. In the mouse, liver and salivary gland α -amylase are identical proteins but are translated from tissue-specific mRNAs. These mRNAs are transcribed from a single α -amylase gene, the tissue specificity being generated by alternative splicing events in the 5'-nontranslated region (Hagenbuchle *et al.*, 1981).

Readthrough of the termination signal of the gene

specifying the secretory form of IgM, and an alternative splicing program, generate the mRNA for the membrane bound form of IgM (Rogers *et al.*, 1980; Early *et al.*, 1980).

It is possible from these examples, that two sets of signals exist for the splicing machinery. If this were so then presumably regulation of each transcript may be achieved by limiting that aspect of the splicing machinery that recognises one set of signals.

A further level of control complexity is seen in the cytochrome b genes of yeast mitochondria (review; Borst and Grivell, 1981). In laboratory yeast strains there are two cytochrome b genes, one "long" and one "short" gene. The "long" gene contains the two introns present in the "short" gene, but also an additional three introns. In at least one of these additional introns there is an open reading frame which codes for a maturase function which is catalytic in promoting splicing of that particular intron. Processing of the primary transcript involves splicing of the first intron to produce an intermediate which then acts as an mRNA to produce a chimeric protein containing amino acids encoded in the first two exons and amino acids encoded in the open reading frame of the second intron. This protein has the maturase function, and controls its own production by catalysing splicing of the second intron.

Whether such mechanisms exist for nuclear genes is unclear, however it exemplifies control at the splicing level which is independent of the general availability of the splicing machinery.

1.5.3 Pseudogenes

Globin loci often contain sequences complementary to globin probes that cannot, however, encode functional globin polypeptides because the translational reading frames have mutated. Changes at splice junctions may also occur. An example is a goat β -globin pseudogene (Cleary *et al.*, 1980). It has a frameshift mutation leading to a premature termination signal; the histidines linked to the heme binding site have been substituted; the termination codon has changed; one splice junction has changed from GT to GC and the Goldberg-Hogness box (see below) has mutated.

Pseudogenes appear not to give rise to stable transcripts, although one interferon cDNA clone has been isolated which displays a frameshift mutation (Goeddel *et al.*, 1981).

The widespread occurrence of pseudogenes in the globin gene systems indicates that they may have a role in control of globin gene expression. Although no evidence exists for this claim, it is interesting to note that in most globin systems a pseudogene separates those genes expressed in different developmental periods (Cleary *et al.*, 1981).

1.5.4 RNA Polymerases

Three types of RNA polymerases, designated I, II and III, have been found in eukaryote cells. RNA polymerase I transcribes genes for ribosomal cistrons and

RNA polymerase II, transcribes most other gene sequences. RNA polymerase III catalyses the synthesis of tRNA and 5S ribosomal RNA.

The levels of polymerases I and III seem to vary in parallel according to the physiological state of the cell. This suggests that tRNA and ribosomal gene activity may be regulated by the availability of their respective RNA polymerase.

1.5.5 Hormone Action

A great deal of evidence has shown that steroid hormones are capable of specifically altering the expression of certain genes. A system which corresponds well to the Britten and Davidson (1969) idea of gene control is the ecdysone induction of pupal development in insects. Hormones (effectors) activating hormone receptors (sensor protein), become regulatory elements which activate structural genes and some regulatory proteins (activation proteins) for subsequent programs of expression.

In *Drosophila* (Grossbach, 1973), there are only ten puffs visible in the salivary gland chromosomes during the larval stage. When ecdysone is added these puffs begin to regress while a number of new puffs suddenly appear. During the transition from the larval to the pupal stage about 120 new puffs eventually appear, each puff arising at a specific time after ecdysone addition, and lasting for a specific period. The first puffs appear

to be directly induced by the hormone as their appearance is not affected by inhibitors of protein synthesis, whereas the appearance of late puffs is blocked.

The control of synthesis of egg proteins by oestrogen or progesterone in the chick oviduct, has been investigated at the molecular level (O'Malley *et al.*, 1976). Upon entering the cell, steroid hormones are initially bound to specific cytoplasmic receptor proteins. This complex moves into the nucleus where it binds to sites on the target cell chromatin.

This is followed by activation of specific genes resulting in new mRNA species. The progesterone receptor of chicken oviduct cells has been purified and its properties extensively investigated (Vedeckis *et al.*, 1978). The receptor has been shown to be a protein-dimer, one subunit locating the hormone-receptor complex to specific regions of chromatin, while the other subunit alters the structure or conformation of the chromatin so that initiation of new RNA synthesis can occur (O'Malley *et al.*, 1978).

The fact that the local structure of chromatin must be modified to allow new transcription implies that template conformation may play a significant role in the control of transcription. Genes that are to be expressed in the same developmental period are often arranged together in pairs. For example, this is the case for the δ/β and G_{γ}/A_{γ} human globin pairs, α_1/α_2 chicken globin pairs (Breathnach and Chambon, 1981),

and chorion gene A/B pairs (Jones and Kafatos, 1980). Embryonic sea urchin histone gene variants are also clustered (Busslinger *et al.*, 1981). This suggests that specific events may occur such that areas of chromatin containing sets of developmentally-related genes, are only opened up for transcription during particular developmental periods.

1.5.6 Phasing of Nucleosomes

Using techniques of nuclease digestion of chromatin and DNA blotting, genes coding for proteins, satellite DNAs and ribosomal and tRNA genes have been found to have a specific spacial relationship with nucleosome particles (review: Zachau and Igo-Kemenes, 1981). Particularly notable examples are that of the chicken embryonic β -globin gene of red blood cells (Weintraub *et al.*, 1981).

Nucleosomes were found to be in preferential locations on and around the β -globin gene when the gene was in the non-expressed state in red blood cells. In brain cells, however, nucleosomal distribution was random.

In the *Drosophila* histone gene clusters, nucleosomes were precisely positioned on the non-transcribed spacer. The 5'-ends of all five histone genes were in an exposed configuration.

Thus, nucleosome phasing may be related to the inter-conversion of active and inactive states of a gene.

1.5.7 DNase Sensitivity of Active Genes

Digestion of chromatin with nucleases suggests that transcriptionally active regions are organized into nucleosome like structures but that these have a different conformation from bulk chromatin (Review; Lilley and Pardon, 1979). Regions of chromatin which code for the products of a particular cell type are selectively sensitive to DNaseI or DNaseII. The degree of sensitivity is the same whether the gene is transcribed frequently or occasionally, suggesting that this state reflects "competency for transcription".

The active conformation appears to correspond to elevated levels of two non-histone proteins HMG14 and HMG17. In the absence of HMG proteins, active genes are no longer selectively sensitive to DNase. Weisbrod *et al.*, (1980) have demonstrated that the two HMG proteins can restore the DNaseI sensitivity of chicken globin genes when reconstituted with salt washed chromatin or nucleosomal cores.

1.5.8 DNA-Methylation

Riggs (1975) and Holliday and Pugh (1975) postulated that changes in DNA methylation may provide a means of controlling gene expression. Recent evidence indicates an inverse correlation between the extent of DNA methylation and expression of both viral (Desrosiers *et al.*, 1979) and cellular (Mandel and Chambon, 1979) genes.

Agents which inhibit methylation have been used to

study this effect. The most specific of these agents is 5-azacytidine. Exposure of cell lines to low levels of 5-azacytidine has resulted in the re-activation of previously silent and methylated endogenous retrovirus genes (Groudine *et al.*, 1981) and metallothionein-1 genes (Compete and Palmiter, 1981). The acquisition of DNaseI hypersensitive sites in retrovirus genes in response to 5-azacytidine, suggests that undermethylation of a given gene precedes a conformational change in the chromatin of that gene which permits transcription to occur (Groudine *et al.*, 1981).

1.5.9 Eukaryotic Promoters: PolymeraseII Genes

By analogy with prokaryote genes, transcription of eukaryotic genes should be initiated at a well-defined promoter region, located somewhere before the start of the mRNA sequence. The 5'-regions of a number of eukaryote genes have been sequenced in the hope of finding a common sequence which might act as a eukaryote promoter.

For polymeraseII genes, at least two elements may be required for specific initiation of transcription at the mRNA start sight: the "selector" sequence and the "modulator" sequence (review: Breathnach and Chambon, 1981).

With few exceptions, polymeraseII genes contain an A/T rich homology region, known as the "TATA box", centered about 25-30 bps upstreat from the mRNA start.

This sequence shows similarities to the prokaryote 'Pribnow box' promoter sequence, however, *in vitro* and *in vivo* transcription studies have indicated that the "TATA box" has more of a selector function than a promoter function. Manipulation or deletion of this region generally results in heterogeneous 5' mRNA termini and/or less efficient transcription initiation.

Transcription studies have also recognized other regions of sequence upstream from the "TATA box" which can drastically modulate the rate of transcription initiation. These regions often overlap evolutionarily conserved regions specific to gene types. They have been called "modulator" sequences. Manipulation in surrogate genetics systems can result in an increase or decrease in initiation efficiency.

Although sequence comparisons and transcription studies have recognized candidates for the eukaryote promoter, little is known of their actual interaction with RNA polymerase or other regulatory molecules.

1.5.10 The Extent of Promoter Sequences

Although sequences immediately preceding genes have a major role in promotion of their expression, it is largely unknown how far the regulatory sequences extend. Transformation of mammalian cultured cell with cloned genes has indicated that sequences relevant to hormonal regulation (Kurts, 1981) and heat shock induction (Corces *et al.*, 1981) are located in the immediate vicinity of

the gene. However, an analysis of deletions associated with β -thalassemias in humans has revealed regions that control the developmental expression of HbF ($\alpha_2\gamma_2$) are located at least 15 Kb downstream from the γ genes in the vicinity of the δ and β genes (Maniatis *et al.*, 1980). It is not known whether these mutations are within genes which code for regulatory proteins.

1.5.11 Eukaryote Promoters - PolymeraseIII Genes

In contrast to polymeraseII genes, polymeraseI and III genes do not have the homologous sequences found upstream from mRNA initiation points. At least one aspect of the polymeraseIII promoter has, however, been recognized as intragenic (Bogenhagen *et al.*, 1980). Transcription of *Xenopus* 5S RNA genes by purified RNA polymeraseIII, is facilitated by the addition of soluble factors, so it is likely that the intragenic sequence represents a "modulator" region. RNA polymeraseIII may employ various binding factors specific for different control sequences within tRNA genes and 5S RNA genes. Such a mechanism would permit the differential control of tRNA and 5S RNA synthesis that occurs under different physiological conditions (Brown and Littna, 1966).

1.6 THE HISTONE GENE SYSTEM

1.6.1 The Histone Proteins

Histones are a set of five small basic proteins. They are the fundamental protein component of chromatin and are associated with the DNA of all eukaryotes. The fifth histone, H1, is bound to the DNA between successive nucleosome core particles, and plays a role in the higher order structure of chromatin. It is present at about half the molar concentration of the other histones (Kornberg, 1974; Noll and Kornberg, 1977).

The core histones have been very conserved through evolution. This is especially true of H3 and H4. The H4 sequence of cows and pea differ in only 2 of 102 amino acid residues. Partial sequencing of the H4 protein of *Tetrahymena* (Glover and Gorovsky, 1978) has, however, shown substantial divergence from the calf sequence. Nevertheless, compared to either globular proteins or other histones, H4 still remains a highly conserved molecule.

The other histones exhibit different structural domains which in some cases show independent evolution. The H2B protein for instance, is an evolutionary hybrid. Approximately the first third of the molecule is variable; the rest is highly conserved. These distinctions are presumably related to the importance of particular regions of the histone molecules to their respective interactions with other histones, DNA and non-histone proteins.

1.6.2 Histone Modifications

Histones become modified post-synthetically in a

number of ways. These include acetylation, phosphorylation, methylation, poly(ADP) ribosylation and the formation of protein A24. In most cases only poor correlations exist which suggest a function for these modifications. For example, acetylation has been associated with transcriptionally active regions of the genome (Vidali *et al.*, 1978), and phosphorylation with chromatin condensation before mitosis (Bradbury *et al.*, 1974).

Of particular current interest is the formation of protein A24. This is a modification of H2A, whereby ubiquitin, a highly conserved protein found in animal cells, bacteria, yeast and higher plants (Goldstein *et al.*, 1975), becomes linked to H2A proteins via an isopeptide bond (Goldknopf and Busch, 1977). All H2A variants so far characterized have ubiquitinated counterparts (Wu and Bonner, 1981).

Recently, Levinger and Varshavsky (1982) have found that active genes in *Drosophila* cells are greatly enriched in ubiquitinated nucleosomes. In addition, Kleinschmidt and Martinson (1981) have observed that uH2A can be reconstituted with nucleosomes which have identical nuclease digestion properties to native chromatin. The ubiquitinated nucleosomes do, however, have the capacity to bind HMG proteins 14 and 17. These proteins are responsible for maintaining the DNaseI sensitive configuration of active chicken globin genes (Weisbrod *et al.*, 1980; section 1.5.7). The incorporation of uH2A into the nucleosomes of active genes, may therefore define certain

structural changes in chromatin, which are necessary for gene transcription.

1.6.3 Primary Structure Subtypes

Although histones are extremely conserved through evolution, structural variants occur in all species that have been characterized. The variants may be changed in only one or a few amino acid positions, or may be massively substituted. Often, the expression of these subtypes is regulated through development or the cell cycle.

Sea urchins contain specific classes of variants which correspond to different developmental periods (Brandt *et al.*, 1979; Brandt and Von Holt, 1978) or tissues (Wouters-Tyrou *et al.*, 1978). A distinct shift in histone gene activity is seen to accompany hatching in the sea urchin *Lytechinus pictus* (Grunstein, 1978).

Recently, Wu and Bonner (1981) have characterized the variants associated with a Chinese hamster ovary cell line and found specific variant subtypes which are synthesized throughout the cell cycle (basal level synthesis) as distinct from the predominant histone subtypes which are only synthesized during S-phase.

There is no real indication of the function of these variants. Some variants exhibiting only few substitutions may be inconsequential to function, having arisen through gene mutation and become fixed in the gene pool. In support of this, Rykowski *et al.*, (1981) have mutated each of the two H2B variant genes in yeast and concluded

that neither is essential to function.

Extreme variation between the subtypes within a species, and stage specific switching, would however argue for the existence of heterogeneous nucleosome particles, and a role for these in development and control of gene expression. Nevertheless, some evolutionary data would question this as a general assumption. The sea urchin *P. miliaris* has a maxi-family of histone genes (repeated several hundred fold in the genome) of which clone h22 is representative (Schaffner *et al.*, 1978). Clone h19 (Busslinger *et al.*, 1980), however, is representative of a minor class of developmentally regulated gene variants. From DNA sequencing studies it was found that h19 was very closely related to the maxi-family of repeats in another species of sea urchin *S. purpuratus*, and that h19 was more related to the *S. purpuratus* maxi-family than it was to clone h22 (the *P. miliaris* maxi-family). This indicates that at the time of divergence of *S. purpuratus* and *P. miliaris*, both h22 and h19 classes existed and that *P. miliaris* adopted the h22 class as its maxi-family, with the h19 class as a minor, developmentally regulated group, whereas *S. purpuratus* adopted the h19 class as its maxi-family. This tends to question the relevance of switching in development of the sea urchin.

1.6.4 Histone mRNA

Histone protein synthesis is closely coupled to DNA replication in the somatic cells of animals (Robbins and

Borun, 1967), yeast (Moll and Wintersberger, 1976) and protozoans (Prescott, 1966), with the exception of basal level synthesis (Wu and Bonner, 1981) and other specific examples of uncoupled synthesis (Zlatanowa and Swetly, 1978; Groppi and Coffino, 1980; Marian and Wintersberger, 1980; Sanders, 1981). Histone mRNA is rapidly lost from the cytoplasm and polyribosomes when DNA replication is stopped (Robbins and Borun, 1967; Gallwitz and Mueller, 1969).

Mature histone mRNAs are about 9S in size and in general are not polyadenylated (Adesnick and Darnell, 1972). Since the poly-A-tract has no function in protein synthesis (Williamson *et al.*, 1974; Bard *et al.*, 1974), it has been proposed that it acts as a nuclease inhibitor (Levy *et al.*, 1975), and could therefore effect mRNA stability. The demonstration that other naturally occurring mRNAs that lack poly-A-tails are also rapidly processed (Milcarek *et al.*, 1974; Nemer *et al.*, 1974), supports this view. This could explain the rapid turnover of cytoplasmic histone mRNAs at the end of S-phase (Gallwitz, 1975), which is not observed for other mRNAs. Evidence has however accumulated that at least a subfraction of histone mRNAs may be polyadenylated (Borun *et al.*, 1977; Ruderman and Pardue, 1978), and it has been postulated that these differences may be related to stage specific switching of histone mRNA sequences.

There is some evidence that histone mRNA may be transcribed into a high molecular weight precursor form in sea urchin gastrulae (Kunkel *et al.*, 1978), but not

in cleavage stage embryos (Childs *et al.*, 1979). Evidence for a high molecular weight precursor of histone mRNA has also been found in HeLa cells (Melli *et al.*, 1977; Hackett *et al.*, 1978), but this has not been substantiated by more recent studies (Rickles *et al.*, 1982).

Although it is generally accepted that histone mRNA expression is regulated at the level of transcription, there are examples of post-transcriptional control. The stored maternal histone mRNA in *Xenopus* oocytes (Adamson and Woodland, 1974) would be a specific example. Perhaps a more general situation is seen in yeast strains that have a duplicated H2A/H2B pair (wild type yeast only contain 2 copies of H2A and H2B). These strains show double the transcription rate of genes, however steady-state mRNA levels are unaffected (Osley and Hereford, 1981).

1.6.5 The Histone Genes

(1) Sea Urchin

The histone genes of sea urchin are perhaps the best characterized of all eukaryotic structural genes (review; Kedes, 1979; Hentschel and Birnstiel, 1981), with almost complete DNA sequences available for two species, *Strongylocentrotus purpuratus* (Sures *et al.*, 1978) and *Psammechinus miliaris* (Schaffner *et al.*, 1978).

All five genes are present in the same order (H1, H2A, H2B, H3, H4) and are transcribed off the same DNA strand. Although the length of the repeat varies

between the various sea urchin species (6.6Kb for *P. miliaris*; 6.5 Kb for *S. purpuratus*), the five genes are proportionately arranged within AT-rich spacer DNA (Portmann *et al.*, 1976).

Direct DNA sequencing of cloned repeats (Schaffner *et al.*, 1978; Sureset *et al.*, 1978) has indicated that the coding sequences are colinear with the amino acid sequences, and so no intervening sequences are present within these histone genes. DNA sequencing of histone genes from other species (Hentschel and Birnsteil, 1981) and S1-mapping data (Hentschel *et al.*, 1980) suggest that the lack of intervening sequences is a general feature of histone genes. Recently however, Engel *et al.*, (1982) have isolated a chicken H3 gene containing two intervening sequences. This indicates that there may be an as yet unexplored class of histone genes which have a structure more typical of other eukaryote genes.

Several examples of gene heterogeneity have been seen in the sea urchin histone gene family. *Lytechinus pictus* contains two classes of repeats which are not intermingled on the DNA (Cohn and Kedes, 1979a, 1979b). Both classes code for similar, if not identical, early embryonic mRNAs, but have totally divergent spacer regions. *Echinus esculentus* also contains two classes of repeat (6 and 7 Kb respectively) but these have been determined to be allelic in nature and differing only in a 1 Kb insertion near the H4 gene.

Isolated histone genes, called "orphons", have been detected in *L. pictus* (Childs *et al.*, 1981). Analysis of one H3 "orphon" revealed that it contained

an H3 gene together with some spacer DNA which was present outside its "parental" repeat and imbedded in moderately repeated DNA. Further analysis has shown that "orphan" histone genes or pseudogenes are present 5-20 times in sea urchin genomes in a highly idiosyncratic fashion. One H2B orphan (pseudogene) contained a transposon-like intron (Hentschel and Birnstiel, 1981).

A minor class of histone variant genes has been isolated from *P. miliaris* (Busslinger *et al.*, 1980). These genes retain the general format typical of the major classes of sea urchin histone genes, but encode variants which correspond to the sea urchin embryonic period.

(ii) *Drosophila melanogaster* histone genes

The histone genes of *D. melanogaster* are the only ones, other than the sea urchin genes, that have been extensively characterized. They are present at about 100 copies per haploid genome. A number of clones of *D. melanogaster* DNA containing histone genes have been isolated (Lifton *et al.*, 1977). These were selected by their ability to hybridize to sea urchin histone mRNA.

Two major types of repeating unit were found. In both cases, all five histone coding sequences were contained within a repeat and, like in the sea urchin, genes were separated by AT-rich spacer DNA. The two types of repeat are 4.8 and 5 Kb long and differ only by the presence, in the longer repeat, of an insertion into the spacer DNA.

Analysis of the position and polarity of the genes has shown that, contrary to the sea urchin situation, the direction of transcription is not the same for all histone mRNA species, and that the mRNA is transcribed from both strands of the DNA. The arrangement of the genes is

5' \rightarrow \leftarrow \rightarrow \leftarrow \rightarrow 3'

 H1 H2B H2A H4 H3

This arrangement makes polycistronic transcription unlikely, as multiple initiation points are required to produce mRNAs from the correct strand.

Like the sea urchin gene system, microheterogeneity is observed superimposed on the major repeat pattern. Several infrequent repeats have been isolated, but so far none have shown to code for structural variants (Goldberg, 1979). *Drosophila* histone gene "orphons" have also been observed (Childs *et al.*, 1981).

(iii) Other invertebrate species

The size of the histone gene repeat has been determined for a number of marine invertebrates (Freigan *et al.*, 1976) including horseshoe crab (4.1 Kb), clam (4.5 Kb), oyster (6.3 Kb) and worm (5.2 Kb). Like in the sea urchins, all five genes are present in the repeating unit.

(iv) Vertebrates

Xenopus: Hybridization studies and analysis of cloned DNA indicate that some *Xenopus* histone genes exist in conserved tandem repeats, but that other genes are arranged in a highly idiosyncratic fashion. Dongen *et al.*, (1981) have isolated a clone containing a gene cluster with a different gene order to that of the major repeat. Zernik *et al.*, (1980) have found a clone with the same gene order as the major repeat but which shows restriction sites heterogeneity between repeats. In

addition, another genomic clone shows a cluster of genes with a different gene order and which contains a variant H1 gene. This suggests that the notion of a conserved, tandem, repeat as in sea urchins and *Drosophila*, does not apply for the *Xenopus* histone genes.

Newt: Histone genes from the newt *N. viridescens* have been isolated and characterized (Stephenson *et al.*, 1981). The genes in this species are reiterated 600-800 fold in a genome 15X larger than that of *Xenopus*. The gene structure is completely different to the *Xenopus* genes. The gene order is H1-H3-H2B-H2A-H4 and all genes are transcribed from one strand except the H2B gene. Although these clusters show homogeneity in restriction sites, they are not tandemly repeated, but rather embedded in stretches of up to 50 Kb of a repeating (225 bps) satellite DNA. Transcription loops of *N. viridescens* lampbrush chromosomes contain these satellite sequences produced by termination readthrough (Diaz *et al.*, 1981).

Chicken: In addition to the work described in this thesis, Engel and Dodgson (1981) have isolated and partially characterized chicken histone genomic clones. They also conclude that chicken histone genes are clustered but that no repeating unit exists.

Mouse: Two clones have been isolated from a mouse genomic library that show different gene compositions (Sittman *et al.*, 1981). One clone contains the genes H3-H4-H1, and the other, H3-(H2B,H4)-H3-H2A, indicating

some randomization of gene order. Another fragment of mouse DNA has been found to have only an H4 gene (Seiler, 1981).

Human: Clones containing H4 genes only have also been isolated from a human genomic library (Heintz *et al.*, 1981). Seven independent clones have been isolated by Sierra *et al.*, (1982a) containing three different gene arrangements, indicating again, that human histone genes are clustered but not tandemly repeated. Such an arrangement is also found by Heintz *et al.*, (1981) on 5 independent histone gene clones.

Other features of the human system have been revealed from characterization of the human clones. Sierra *et al.*, (1982a) have found that one of the histone gene containing clones also encodes an RNA of approximately 330 nucleotides, which is only synthesized in G1 phase of the HeLa cell cycle. Sierra *et al.*, (1982b) have also found that some human histone genes are interspersed with multiple copies of the Alu DNA family (Tashima *et al.*, 1981).

In general, the rigidly conserved arrays of histone genes observed in the lower eukaryotes are not maintained in vertebrate species. Gene order appears to have been randomized to the extent that histone genes are interspersed with both middle repetitive DNA and other non-histone genes.

1.7 Aims of the Project

At the time that the work in this thesis was initiated,

only the structure of the sea urchin and *Drosophila* histone genes were known. No vertebrate histone genes had been isolated.

Previous work in this laboratory had focused on the tissue specific histone H5 and its mRNA. H5 is a histone specific to nucleated erythroid cells. Attempts were being made to enrich for chicken H5 mRNA so that the genes could be isolated and studied. In order to understand the specific control of H5 expression, it seemed necessary to be able to compare its gene sequences and structure with that of the normal complement of core histone genes. This thesis describes the isolation and partial characterization of chicken histone genes. Subsequent to this work the H5 genes have been isolated and characterized in this laboratory.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

Agarose: for analytical purposes and for transfer of DNA from agarose gels to nitrocellulose, agarose was purchased from Sigma Chemical Co., Saint Louis Missouri, U.S.A.

Low Melting Temperature Agarose: for preparative agarose gel electrophoresis, was purchased from Bethesda Research Laboratories Inc. (BRL), Galthersburg Maryland, U.S.A.

Chemicals: All chemicals were of analytical reagent grade or of highest obtainable purity.

Column Chromatography Materials: Sephadex G-50 was purchased from Pharmacia Fine Chemicals, Uppsala Sweden.

Nitrocellulose Filter Paper: for most purposes was purchased from Sartorius GmbH, Göttingen, Germany. For transfer of small DNA fragments (less than 1 Kb) from agarose gels, grade BA-85 nitrocellulose was purchased from Schleicher and Schuell Inc., Keene New Hampshire, U.S.A.

Enzymes: The enzymes used in the course of this work were obtained from the sources listed below.

Bacterial alkaline phosphatase: Worthington Biochemical Corp., Freehold New Jersey, U.S.A.

E. coli Deoxyribonuclease I: Sigma Chemical Co.

E. coli DNA-polymerase I: Boehringer Mannheim GmbH, Mannheim Germany.

Exonuclease III: supplied by P. Molloy.

Polynucleotide Kinase: Boehringer, Mannheim.

Proteinase K: E. Merck, Darmstadt Germany.

RNA-dependent DNA-polymerase (reverse transcriptase):
gift from J.W. Beard and the N.I.H. Cancer Program.

Ribonuclease A: Sigma Chemical Co.

T4-DNA ligase: Boehringer, Mannheim.

Restriction Enzymes: EcoRI was a gift from P. Krieg and S. Clark. All other restriction enzymes were purchased from New England Biolabs Inc., Beverly Maryland, U.S.A., or Boehringer, Mannheim.

Biological Reagents:

RNA: chicken scale ribosomal and 4S RNAs were supplied by S. Wilton. The 10S globin mRNA was supplied by P. Krieg.

Sea urchin histone gene recombinant, λ 55: was supplied by K. Gross.

Chicken recombinant library: was constructed by J. Dodgson, D. Engel and R. Axel and was supplied by J. Dodgson.

2.2 METHODS

2.2.1 Preparation of RNA

All procedures involving RNA were carried out at 0°C using sterile solutions and glassware.

Method 1: Preparation of 7-11S RNA from chick embryos -
Seeberg et al., (1977)

Thirty 5 day-old chick embryos were snap-frozen in

liquid nitrogen, then homogenized in 7M guanidinium-HCl, 20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1% (w/v) Sarkosyl in a Dounce homogeniser in a final volume of 30 mls. Total RNA from this homogenate was recovered as material centrifuged through 5.7 M CsCl. The clear RNA pellets were resuspended in 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 5% Sarkosyl and 5% phenol, then made 0.1 M in NaCl and extracted with an equal volume of phenol:chloroform (1:1 v/v). RNA from the aqueous phase was collected after ethanol precipitation, resuspended in 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.5% SDS, heated at 65°C for 5 minutes, chilled and centrifuged on 10-40% sucrose gradients in 100 mM NaCl, 10 mM Tris-HCl pH 7.4, 1 mM EDTA, (210,000g for 16 hours). The 7-11S RNA was collected and then refractionated on a second 10-40% sucrose gradient.

Method 2: Preparation of 7-11S RNA from chick embryos - Chirgwin *et al.*, (1979), (J. Brooker, personal communication).

Thirty 5 day-old embryos were homogenized in a Dounce homogenizer in 7 mls of 6 M guanidinium HCl, 0.2 M Na-acetate pH 5.2 and 1 mM 2-mercapto-ethanol. Acetic acid (1.6 mls) was added to the homogenate, then nucleic acid was precipitated by adding 1/10th volume of 2 M K-acetate pH 5 and an equal volume of ethanol. The precipitate was centrifuged at 10,000g for 15 minutes to recover the insoluble material. The pellet was suspended in half the original homogenate volume of 6 M guanidinium-HCl,

0.2 M Na-acetate pH 5.2, 10 mM EDTA and again precipitated by addition of an equal volume of ethanol, (-20°C 1 hr). Nucleic acid was recovered, resuspended and precipitated again using the same procedure. After recovery the pellet was dissolved in 10 mls of 25 mM EDTA and extracted with an equal volume of phenol:chloroform (1:1 v/v). Two volumes of 4.5 M K-acetate pH 6 were added to the aqueous phase, the DNA was left to precipitate overnight, then recovered by centrifugation, (16,000g, 40 minutes). The pellet was dissolved in water and precipitated by adjusting the solution to 0.2 M K-acetate pH 5, and adding two volumes of ethanol (-80°C 1 hr). After recovery the pellet was dissolved in 3 mls of 10 mM Tris-HCl pH 7.4, 1 mM EDTA, 0.1% (w/v) SDS and fractionated on four 10-40% sucrose gradients in 100 mM NaCl, 10 mM Tris-HCl pH 7.4, 1 mM EDTA (210,000g, 16 hours, 4°C). The 7-11S fraction was refractionated on two 10-40% sucrose gradients.

2.2.2 Restriction Enzyme Digestion and Analysis of DNA

Restriction digestion conditions

Restriction endonuclease digestion of DNA with either commercial or non-commercial preparations of enzyme were carried out using the conditions recommended for each enzyme by New England Biolabs Inc. All reactions were stopped by the addition of EDTA to a final concentration of 25 mM. Protein was removed by extraction with phenol:

chloroform (1:1 v/v) and DNA recovered by ethanol precipitation of the aqueous phase.

Agarose gel electrophoresis

Electrophoresis of DNA for analytical or preparative purposes, or for transfer to nitrocellulose was carried out on 14cm x 14cm x 0.3cm slab gels containing agarose (0.7 - 1.5%) dissolved in 40 mM Tris-acetate pH 8.2, 20 mM Na-acetate, 1 mM EDTA. Electrophoresis was carried out at 60 mA for about 3 hours. DNA was visualized by staining with 0.02% ethidium bromide solution for 15 minutes and examination under UV light.

Polyacrylamide gel electrophoresis

Electrophoresis of DNA species less than about 1 Kb in length was carried out on vertical 14cm x 14cm x 0.5 mm gels containing 6% acrylamide polymerized in 90 mM Tris-borate pH 8.3, 2.5 mM EDTA. Electrophoresis was performed at 150V for 2 hours. DNA was visualized by ethidium bromide staining.

Transfer of DNA to nitrocellulose and hybridization with a labelled probe

Restricted DNA fractionated on agarose slab gels was transferred to nitrocellulose filter paper using the method of Southern (1975), as modified by Wahl *et al.*, (1979). The times used for partial hydrolysis of DNA in the gel with 0.25 N HCl were halved for 0.3cm thick agarose gels.

The prehybridization, hybridization and washing

conditions for nitrocellulose filters were exactly as described by Wahl *et al.*, (1979). The washed, dried nitrocellulose filter was placed in contact with X-ray film and exposed at -80°C in the presence of an intensifying screen.

Low melting temperature agarose gel electrophoresis

DNA fragments to be isolated from agarose gels were electrophoresed on vertical or horizontal 0.3mm thick gels cast from low melting temperature agarose dissolved in 40 mM Tris-acetate pH 8.2, 20 mM Na-acetate, 1 mM EDTA. Electrophoresis was carried out at a maximum of 7.5V/cm. DNA was detected by ethidium bromide staining and bands to be isolated, excised from the gel with a blade. The agarose slice was melted at 65°C in an Eppendorf tube containing 200 μl of 200 mM NaCl, 10 mM Tris-HCl pH 7.4, 1 mM EDTA. An equal volume of phenol, saturated with the same buffer was added, the tube removed from the heat and vortexed hard. Phases were separated by centrifugation for 5 minutes in an Eppendorf centrifuge. The phenol extraction was repeated on the aqueous phase twice at room temperature. Residual phenol was removed by extraction with ether. DNA was precipitated twice with ethanol.

2.2.3 *In vitro* Synthesis of Labelled DNA

Random-primed reverse transcription

Priming of RNA lacking a 3' poly-A tract was achieved by the random hybridization of oligo-nucleotides of

salmon-sperm DNA, prepared as described by Taylor *et al.*, (1976). Extension of the random primers using reverse transcriptase (20 units) was carried out in a 20 μ l reaction mix containing 1 mM each of dATP, dGTP and dTTP, about 0.1 mM $\alpha^{32}\text{P}$ -dCTP (400 Ci/mmol), 50 mM Tris-HCl pH 8.3, 10 mM MgCl_2 , 10 mM 2-mercaptoethanol and 2 mg/ml oligo-nucleotides. The solution was incubated at 37°C for 90 minutes, then after phenol extraction, made 0.1 N NaOH and incubated further to digest the RNA template. The aqueous phase was loaded onto a 0.4cm x 4cm Sephadex G-50 column, and eluted with 10 mM Tris-HCl pH 7.4, 1 mM EDTA. The excluded volume fractions were pooled and added directly to hybridization mixes.

Nick-translation of double-stranded DNA

Labelling double-stranded DNA using *E. coli* DNA polymerase I was carried out essentially as described by Maniatis *et al.*, (1975). The 50 μ l incubation mix contained 50 mM Tris-HCl pH 7.8, 5 mM MgCl_2 , 10 mM 2-mercaptoethanol, 50 μ g/ml bovine serum albumin, 5 μ M each of $\alpha^{32}\text{P}$ -dCTP and ^{32}P dGTP (400 Ci/mmol) and unlabelled dATP and dTTP. The DNA was nicked by the addition of 20 pg of *E. coli* DNaseI and the reaction started by the addition of 2 units of *E. coli* DNA polymerase I. Incubation was for 90 minutes at 14°C and the reaction mix was phenol extracted and passaged through a Sephadex G-50 column to remove unincorporated nucleotides.

2.2.4 Subcloning of DNA Fragments

The ligation of DNA fragments into plasmid vectors was found to be size selective depending on the concentration of the DNA species. For this reason fragments to be subcloned were isolated from sucrose gradients or agarose gels before ligation.

Ligation

The insertion of specific fragments into plasmid vectors could be optimized by adjusting the concentration of both insert and vector DNA according to the mathematical treatment of Dugaiczky *et al.*, (1975). Ligation conditions were as specified by Boehringer, Mannheim.

Transformation of bacteria with plasmid DNA

E. coli strain ED8654: $r_k^- m_k^+$ *supE supF trpR* was grown overnight at 37°C in L-broth and then diluted 1/50 into fresh L-broth and grown to an A_{600} of 0.6 - 0.8. The cells were chilled on ice for 30 minutes, pelleted by centrifugation and resuspended in $\frac{1}{2}$ volume of ice-cold 0.1 M $MgCl_2$. The cells were pelleted immediately and resuspended in 1/20 of the original volume of ice-cold 0.1 M $CaCl_2$. The suspension was kept on ice for at least 1 hour. 0.2 ml of these "competent" cells were added to 0.1 ml of the plasmid DNA in 0.1 M Tris-HCl pH 7.4, and stood on ice for 30 minutes, with occasional stirring. The cells were heat-shocked at 42°C for 2 minutes, kept on ice for 30 minutes and then warmed slowly to room temperature. 0.5 mls of L-broth was

added to the transformed cells and incubated at 37°C for 30 minutes. Transformed cells were mixed with 3 mls of 0.7% L-agar and plated on 1.5% L-agar plates containing 15 µg/ml tetracycline. These were incubated overnight at 37°C.

Plasmid preparation - miniscreen

Isolation of plasmid DNA from small cultures was carried out as follows. 1.5 ml cultures of each recombinant were grown overnight in L-broth containing 15 µg/ml tetracycline (or appropriate antibiotic). Cells were pelleted by centrifugation for 30 seconds in an Eppendorf centrifuge, washed in 10 mM Tris-HCl pH 9, 1 mM EDTA, pelleted again and resuspended in 150 µl of 15% sucrose, 50 mM Tris-HCl pH 9 and 50 mM EDTA. 50 µl of freshly prepared lysozyme solution (4 mg/ml) was added and the solution incubated at room temperatures for 15 minutes and at 0°C for 30 minutes. 200 µl of ice-cold water was added and the solution heated to 72°C for 15 minutes. After centrifugation at 30,000g for 20 minutes the supernatant was removed from the white gelatinous pellet and ethanol precipitated. Plasmid DNA was resuspended in water and an aliquot removed for restriction endonuclease digestion in a 200 µl digestion mix. DNA was extracted with phenol/chloroform and ethanol precipitated before electrophoresis on agarose slab gels.

2.2.5 Large-scale Preparation of Recombinant Plasmid DNA

500 ml cultures of recombinant cells were grown in

L-broth to an A_{600} of 1.0 and then chloramphenicol was added to a final concentration of 150 $\mu\text{g/ml}$. The cells were incubated overnight at 37°C to allow amplification of plasmid DNA (Clewell, 1972). The cells were pelleted by centrifugation (1000g, 5 minutes), and resuspended in 10 mls of 15% sucrose, 50 mM Tris-HCl pH 9, 50 mM EDTA. 2 mls of 6 mg/ml lysozyme solution was added and the cells were incubated at room temperature for 15 minutes and at 0°C for 30 minutes. 10 mls of boiling-hot 50 mM Tris-HCl pH 9, 50 mM EDTA was added with vigorous mixing and the solution was heated at 72°C for 15 minutes. The lysate was centrifuged at 30,000g for 30 minutes and the supernatant was removed and treated with DNAase-free RNAase A (20 $\mu\text{g/ml}$), for 30 minutes at 37°C , then Proteinase K (50 $\mu\text{g/ml}$), for 30 minutes at 37°C . The solution was extracted with an equal volume of phenol:chloroform (1:1, v/v) and the aqueous phase ethanol precipitated. Plasmid DNA was resuspended in 16 mls of water and 16 gms of CsCl and 4 mls of 10 mg/ml ethidium bromide were added and dissolved. This solution was centrifuged at 210,000g for 40 hours at 15°C . Plasmid bands were viewed under UV illumination and the lower supercoiled band withdrawn with a syringe. The solution was extracted five times with isopropanol which had been saturated with 9.5 gm/ml CsCl, then dialysed extensively against 10 mM Tris-HCl pH 7.4, 1 mM EDTA. DNA was stored as an ethanol precipitate.

L-broth to an A_{600} of 1.0 and then chloramphenicol was added to a final concentration of 150 $\mu\text{g/ml}$. The cells were incubated overnight at 37°C to allow amplification of plasmid DNA (Clewell, 1972). The cells were pelleted by centrifugation (1000g, 5 minutes), and resuspended in 10 mls of 15% sucrose, 50 mM Tris-HCl pH 9, 50 mM EDTA. 2 mls of 6 mg/ml lysozyme solution was added and the cells were incubated at room temperature for 15 minutes and at 0°C for 30 minutes. 10 mls of boiling-hot 50 mM Tris-HCl pH 9, 50 mM EDTA was added with vigorous mixing and the solution was heated at 72°C for 15 minutes. The lysate was centrifuged at 30,000g for 30 minutes and the supernatant was removed and treated with DNAase-free RNAase A (20 $\mu\text{g/ml}$), for 30 minutes at 37°C , then Proteinase K (50 $\mu\text{g/ml}$), for 30 minutes at 37°C . The solution was extracted with an equal volume of phenol:chloroform (1:1, v/v) and the aqueous phase ethanol precipitated. Plasmid DNA was resuspended in 16 mls of water and 16 gms of CsCl and 4 mls of 10 mg/ml ethidium bromide were added and dissolved. This solution was centrifuged at 210,000g for 40 hours at 15°C . Plasmid bands were viewed under UV illumination and the lower supercoiled band withdrawn with a syringe. The solution was extracted five times with isopropanol which had been saturated with 0.5 gm/ml CsCl, then dialyzed extensively against 10 mM Tris-HCl pH 7.4, 1 mM EDTA. DNA was stored as an ethanol precipitate.

2.2.6 Isolation of Clones from a Recombinant Gene Library Plating and Screening (Benton and Davis, 1977)

0.25 mls of a suspension containing 1.6×10^4 phage in 10 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, were added to 0.5 mls of a mid-log phase culture of LE392 (*E. coli*, $rk^- mk^+ supE supF trpR^-$) in L-broth. This mixture was incubated at 37°C for 10 minutes. 8 mls of 0.7% L-agar at 45°C were added, and the mixture poured onto 13cm agar plates containing 10% tryptone, 5% yeast extract, 5% NaCl, 2% glucose, 10 mM Tris-HCl pH 7.5, and 1 mM $MgCl_2$. Plates were incubated at 37°C overnight, then dried without lids for 30 minutes at 37°C, and placed at 4°C to harden the agar. Two 12cm nitrocellulose discs were layered in succession onto each plate until wet. Orientation marks were made with ink and a needle. The filters were carefully peeled from the plates and dipped into 0.5 M NaOH, 0.75 M NaCl solution for 30 seconds, then into 0.5 M Tris-HCl pH 7, 1.5 M NaCl solution for 30 seconds. Filters were blotted dry, stacked between layers of paper and baked at 80°C for 2 hours *in vacuo*. In a culture plate, filters were then immersed in a solution of 50 X Denhardt's solution (Denhardt, 1966), 50 X SSC (1 X SSC is 0.15 M NaCl, 0.015 M Na-citrate pH 7.0), 0.5 M EDTA, 10% SDS, and incubated at 65°C for 4 hours. Subsequent pretreatment and hybridization of filters with ^{32}P -labelled probes was carried out as specified by Wahl *et al.*, (1979). Filters were washed with 2 X SSC, 10 X Denhardt's solution and 0.1% SDS,

briefly at room temperature, then several times at 65°C for 5 hours. Autoradiography was carried out for 3 days.

Growth of phage - liquid culture

Positive plaques were picked with a sterile pasteur pipette and resuspended in 1 ml of 10 mM Tris pH 7.4, 10 mM MgCl₂ and diluted appropriately to titre the stock. Plate stock of phage were generated by plating enough phage so that total lysis of the bacterial lawn occurred. The plates were then flooded with 10 mls of 0.1 M NaCl, 10 mM Tris-HCl pH 7.4, 10 mM MgCl₂ (PSB). The layer of top agar was scraped off into a centrifuge tube together with the PSB and agitated vigorously. Agar was pelleted by centrifugation at 10,000g for 5 minutes, and the clarified phage stock sterilized by the addition of a drop of chloroform.

For large scale phage preparations, approximately 4×10^{10} cells of *E. coli* were infected with phage at a multiplicity of infection of approximately 0.4, then added to 1 litre of NZCYM medium (10% NZ amine, 0.5% NaCl, 0.5% yeast extract, 0.1% casamino acids, 10% MgSO₄). This culture was agitated overnight at 37°C. A few drops of chloroform were added and the culture agitated for a further 15 minutes. 1 µg/ml RNase A and 0.25 µg/ml DNaseI were added and the cultures placed on ice for 30 minutes. NaCl was added to a final concentration of 1 M and the culture clarified by centrifugation at 10,000g for 10 minutes. Phage particles were

precipitated by the addition of 100 gms of polyethylene glycol 6000 and chilling to 4°C for 1 hour. The flocculated phage were pelleted (10,000g, 10 minutes) and then resuspended in 16 mls of PSB. This suspension was layered onto discontinuous CsCl gradients containing 1.9 ml portions of CsCl solutions in PSB with densities of 1.7, 1.5 and 1.45 gms/cc. Gradients were centrifuged at 210,000g for 1.5 hours at 4°C. Phage particles banding between the 1.5 and 1.45 gms/cc interface were collected, dialysed against PSB, and stored at 4°C. DNA was isolated from phage stocks by phenol/chloroform extraction in the presence of 0.1% SDS, and concentrated by ethanol precipitation.

Preparation of phage DNA from plate stocks

Phage stocks were prepared from many plates as described above. Phage were concentrated by polyethylene glycol precipitation and purified by banding on CsCl step gradients. DNA was isolated by SDS/phenol/chloroform extraction.

Biohazard Containment

All manipulations involving recombinant DNA were carried out in accordance with the guidelines of the Biohazard Safety Committee (ASCOR, Australia).

2.2.7 Gilbert and Maxam DNA Sequencing Procedures

Most procedures were as described by Maxam and Gilbert (1980), however some modifications used are described.

Phosphatase treatment of DNA

DNA fragments were dephosphorylated directly in a restriction endonuclease reaction mix by the addition of 1/10 volume of 1 M Tris-HCl pH 8, 1/15 volume of 10% SDS and 2-10 units of bacterial alkaline phosphatase or calf intestinal phosphatase. Incubation was at 37°C for 2 hours. Phosphatases were inactivated by three phenol/chloroform extractions.

Elution of DNA from polyacrylamide gels

Excised DNA bands were placed in an Eppendorf tube containing 0.5 mls of 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA and 0.1% SDS. The mixture was incubated at 37°C overnight. Buffer was separated from the gel slice and the DNA precipitated from solution by the addition of two volumes of ethanol. DNA was then pelleted by centrifugation in an Eppendorf centrifuge, resuspended in 0.2 M NaCl, 10 mM Tris-HCl pH 7.4, 1 mM EDTA, extracted with phenol/chloroform and re-precipitated with two volumes of ethanol.

The forward polynucleotide-kinase reaction

Eluted, dephosphorylated DNA was end-labelled in a reaction volume of 10 µl containing 50 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol, 0.1 mM spermidine, 0.1 mM EDTA, approximately 50 pmoles of α(³²P)-ATP (4000 Ci/mmol) and 2 units of T4 polynucleotide kinase. Incubation was for 30 minutes at 37°C. The reaction was stopped by the addition of EDTA to 25 millimolar and

phenol/chloroform extraction. DNA was ethanol precipitated twice before secondary restriction endonuclease cleavage.

The exonuclease III reaction

DNA fragments were digested with exonuclease III in a reaction volume of 50 μ l containing 66 mM Tris-HCl, pH 8.0, 0.66 mM MgCl₂, 1 mM 2-mercaptoethanol and 10 units of exonuclease III. Incubation was for 2-5 minutes at 4°C. After incubation, the fragments were dephosphorylated by the addition of Tris-HCl pH 8, SDS and phosphatase as described above.

Strand-separation of DNA fragments

Dried, end-labelled DNA to be strand-separated was dissolved in 30 μ l of a denaturation buffer containing 30% dimethyl sulphoxide, 1 mM EDTA and electrophoresis dyes. An aliquot of 1.5 μ l was removed to use as an undenatured control. The remaining sample was heated at 90°C for 2 minutes, then chilled on ice. Both samples were loaded immediately on a 5% polyacrylamide gel polymerized from a 50:1 acrylamide:BIS stock. Electrophoresis was at 100V for three hours. Labelled fragments were detected by autoradiography and strand-separated fragments excised and eluted as described above.

CHAPTER 3

APPROACHES TO THE ISOLATION OF CHICKEN HISTONE GENES - SELECTION OF RECOMBINANTS

(N.B. Some of the data presented in
this chapter has been published:
Crawford, R.J., Krieg, P., Harvey, R.P.,
Hewish, D.A. and Wells, J.R.E. (1979)
Nature 279, 132-136.
Harvey, R.P. and Wells, J.R.E. Nucleic
Acids Res. (1979) 7, 1787-1798)

3.1 INTRODUCTION

At the onset of the research described in this thesis, no vertebrate histone genes had been described. Considerable detail had been accumulated on the structure and organization of the "early" class of histone genes in sea urchins and *Drosophila* (see Chapter 1).

Sea urchin histone DNA could be purified away from total genomic DNA before the advent of recombinant DNA technology. Because of its overall A/T rich composition and high reiteration frequency, it would migrate as a satellite species on CsCl density equilibrium gradients. Analysis of sea urchin histone genes was possible using material prepared in this way (Schaffner *et al.*, 1976; Gross *et al.*, 1976). When the appropriate cloning technology became available, the purified native histone genes were committed to recombinant form (Clarkson *et al.*, 1976; Kedes *et al.*, 1975) and further characterized.

The histone genes of *Drosophila* were isolated from a recombinant library by using a cross-species hybridization probe (Lifton *et al.*, 1977). This approach avoided the development of homologous probes for the selection of the genes.

In considering approaches to the isolation of higher eukaryote histone genes there were several options. The immediate problems focused on the selection, not the construction of recombinants. Although *Drosophila* histone genes were successfully selected using cross-species (sea urchin) mRNA probes, other attempts to

isolate vertebrate histone genes using cloned sea urchin histone DNA probes had been unsuccessful, (K. Gross, K. Murray, personal communications). Cross-hybridization of simple repeating sequences within the inter-gene regions of the sea urchin gene repeat, with genomic satellite DNA, resulted in selection of a predominance of non-histone clones. This could perhaps have been avoided by the use of coding-specific probes, however, in relation to this work, the restriction endonuclease data necessary to construct such probes had not at that time been compiled for the cloned histone gene repeat available to us (*Echinus esculentus*). An additional conceptual problem in this approach, was the use of probes across large species barriers. Accumulated nucleotide substitutions within the genes may prohibit effective cross-hybridization, especially to genes which encode variant protein sub-types. This may have been reflected in an inability in this laboratory to detect chicken histone genes on Southern blots using sea urchin probes (R.J. Crawford, J.R.E. Wells, personal communications). It seemed desirable therefore, to pursue a source of homologous probe for the selection of chicken histone recombinants, and to use cloned sea urchin histone DNA for secondary identification.

The difficulties in isolation of highly enriched histone mRNAs have been considerable. Because histones are "cell cycle" proteins, they are never the final "luxury product" of any differentiated tissue. In addition,

their synthesis only during S-phase of the cell cycle (Prescott, 1966), and the instability of their messengers outside S-phase (Perry and Kelley, 1973), prevent accumulation of histone mRNAs in most cell types. Nevertheless, because of the association of histone mRNA synthesis with DNA synthesis, rapidly dividing, or synchronous populations of cells have provided useful sources of histone mRNAs. Pulse-labelling of rapidly-cleaving sea urchin embryos produces histone mRNA of high radiochemical purity (Kedes and Gross, 1969). Histone mRNAs have also been enriched in a number of other systems, however attaining high levels of sequence or radiochemical purity has been tedious or difficult (Burckhardt and Birnstiel, 1978; Wilson and Melli, 1977).

By analogy with rapidly dividing sea urchin embryos, Crawford *et al.*, (1979) in this laboratory, have investigated rapidly developing 5-day old chick embryos as a source of vertebrate histone mRNAs. Enrichment of total embryo RNA for histone sequences was carried out on sucrose velocity gradients and by polyU-Sepharose chromatography. The resultant 7-11S, polyA⁻ fraction programmed the synthesis of only histones in the wheat germ cell free translation system, indicating that the chick embryo was indeed an extremely good source of histone mRNA (Crawford *et al.*, 1979).

This chapter describes techniques for the preparation of total embryo RNA, a partial assessment of the usefulness of cDNA made to embryo mRNA as a "positive" histone gene probe, and the use of cDNA in the selection of chicken recombinants.

3.2 RESULTS

3.2.1 Preparation of Chick Embryo RNA

Techniques for the isolation of total cellular RNA have been progressively refined to accommodate isolation of RNAs from tissues high in ribonuclease. Although ribonucleases are effectively sequestered from RNA in intact cells, disruption of the cells inevitably and rapidly mixes ribonucleases and RNA. Most competitive inhibitors of ribonuclease have proven inadequate for arresting the subsequent RNA degradation in tissue with high levels of ribonuclease (Chirgwin *et al.*, 1979). An active site inhibitor, diethyl pyrocarbonate (Ehrenberg *et al.*, 1976) which is effective in inhibition of RNA degradation (Harding *et al.*, 1977), also modifies single stranded nucleic acids (Ehrenberg *et al.*, 1976) and decreases the translational activity of mRNA.

Methods employing protein denaturants such as detergents (Schmidt, 1957) or phenol (Kirby, 1956), are only useful in tissues with low ribonuclease levels. However, an extremely potent denaturant, guanidinium thiocyanate has been used by Chirgwin *et al.*, (1979) to isolate undegraded cellular RNA from rat pancreas, a tissue extremely high in ribonuclease. Both anionic and cationic components of this reagent are strong chaotropes (Jencks, 1969) and can be solubilized to high concentration. A similar reagent, guanidinium hydrochloride (Cox, 1968), where only the guanidinium cation is chaotropic, is not

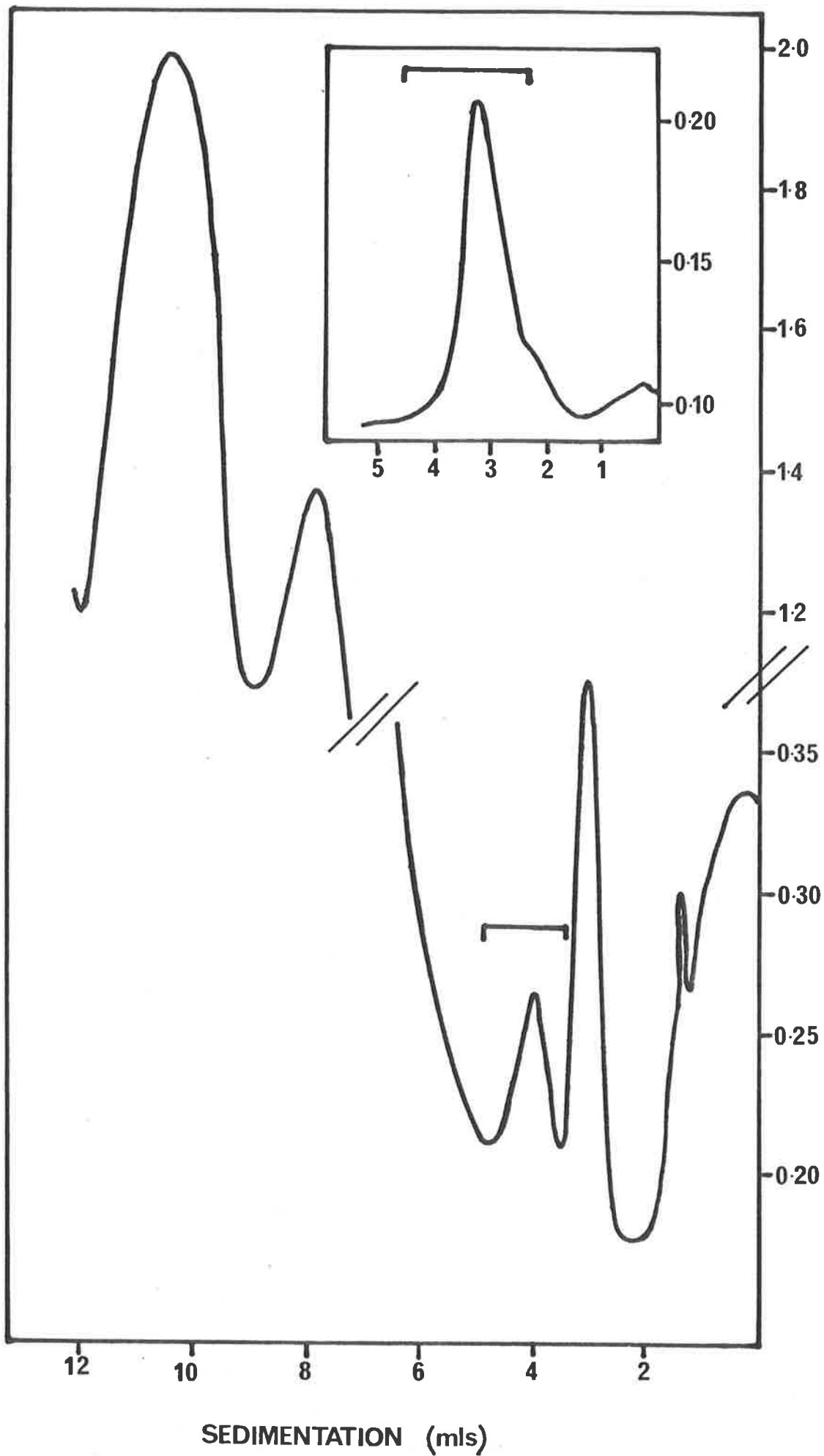
effective for RNA isolation from the rat pancreas (Chirgwin *et al.*, 1979).

RNA from 5-day chick embryos was first prepared (R. Crawford, D. Hewish, this laboratory) by the method of Seeberg *et al.*, (1977), using guanidinium hydrochloride as protein denaturant. The guanidinium homogenate was centrifuged through a CsCl pad to separate RNA from DNA, protein and guanidinium (section 2.2.1). After sucrose gradient centrifugation, a 7-11S fraction was selected and re-run on another gradient (Fig. 3.1). Although this method had yielded excellent RNA preparations (Crawford *et al.*, 1979), variability was seen. This was judged by low yields of 28S relative to 18S RNA, and accumulation of some degradation products in the slower migrating regions of sucrose gradient profiles. In an attempt to improve the reliability of this technique, several parameters were tested. Chirgwin *et al.*, (1979) stressed the importance of the efficiency and rapidity of the initial homogenization procedures. Mechanical (Sorval Omnimix) and manual (Dounce homogenizer) homogenization were tested in both guanidinium-hydrochloride and guanidinium-thiocyanate buffers, using both fresh embryos and embryos frozen in liquid nitrogen and pulverized before use. None of the more stringent applications improved the reliability of the preparations. This implied that the chick embryos were relatively low in ribonuclease and that the source of the variability might lie in some other aspect of the technique.

FIGURE 3.1

Preparation of 7-11S RNA from 5-day old chick embryos (section 2.2.1). 10-40% sucrose gradient fractionation profile (A_{254}) of total cellular RNA.

INSET: Rerun of 7-11S RNA fraction from the first gradient.



On average 30 μg of 7-11S RNA could be recovered from 30 embryos, however, variability in yield also occurred. This could partly be attributed to the stage of harvest of the embryos. A rapid burst in growth of whole embryos occurred at approximately 5 days, but with some variability due to season and prior cold storage. Embryos harvested before the burst yielded low amounts of total RNA.

A. Robins and R. Richards (this laboratory) have also observed low recovery of RNA from homogenates of anaemic chicken blood using this technique. This was attributed to difficulties in centrifugation of RNA through homogenates which had been made viscous with deproteinated DNA. Harding *et al.*, (1978) have used the technique for isolation of only small amounts (100 μg or less) of RNA. It therefore seemed that scaling up the tissue volume would lead to serious losses using this method.

In an attempt to overcome variability in degradation and yield, the method of J. Brooker (personal communication; see section 2.2.1 for details) was subsequently adopted using either guanadinium-hydrochloride or guanidinium-thiocyanate as denaturant. This method, based on that of Chirgwin *et al.*, (1979), involves homogenization in a guanidinium salt, ethanol precipitation of nucleic acid to remove protein, and potassium acetate precipitation to remove DNA. Greater yields (80 μg per 30 embryos) and RNA integrity have been experienced routinely using this method.

3.2.2 Synthesis and Use of 5 day Embryo 7-11S cDNA

Although the 7-11S mRNA fraction from 5 day chick embryos could be shown to translate into predominantly histone protein products as analysed on polyacrylamide tube gels (Crawford *et al.*, (1979)), this is a poor indication of the relative abundance of histone sequences in this fraction. The most likely contaminant, limited ribosomal RNA breakdown (Taylor, 1979), would not programme the synthesis of products in cell-free translation systems. Thus the synthesis of radio-labelled cDNA to this mRNA fraction may produce a histone probe of low radiochemical purity.

To assess directly the use of chick embryo 7-11S cDNA as a histone probe, two hybridization experiments were performed. Highly labelled cDNA was synthesized with AMV reverse transcriptase using the random primer method of Taylor *et al.*, (1976). This method was employed because histone mRNAs generally lack a 3' poly-A track. Much of the ribosomal sequences could be removed from this cDNA by an RNA excess hybridization to 18S and 28S ribosomal RNA, then fractionation of unhybridized material on sucrose gradients (Crawford *et al.*, 1979). Approximately 30-50% of radioactivity could be removed in the form of rDNA/RNA hybrids with this method.

Using this enriched cDNA as hybridization probe, R. Crawford (this laboratory) was able to detect bands in Southern blots of restriction endonuclease digested sea urchin DNA corresponding in size to bands detected

when cloned sea urchin histone genes were used as probe. These bands were distinct from bands detected with ribosomal cDNA probes.

In an extension of this experiment, radiolabelled cDNA was hybridized to Southern blots of restriction endonuclease digests of the sea urchin histone gene clone, λ 55 (Fig. 3.2). λ 55 contains a 7 Kb fragment of histone DNA from the sea urchin *Echinus esculentus* equivalent to one of the tandemly repeated histone gene clusters seen in the sea urchin genome (Kedes, 1979). The fragment contains one each of all five sea urchin histone genes (K. Gross, personal communication). Figure 3.2 shows the hybridization signals detected when chick embryo cDNA was used to challenge digests of λ 55 DNA.

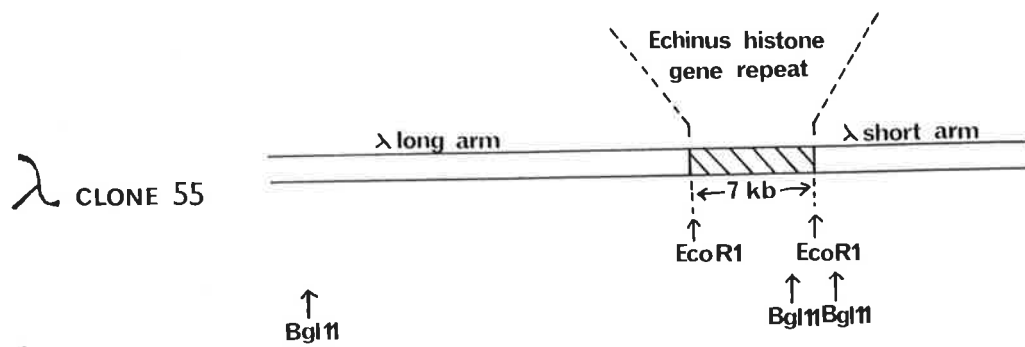
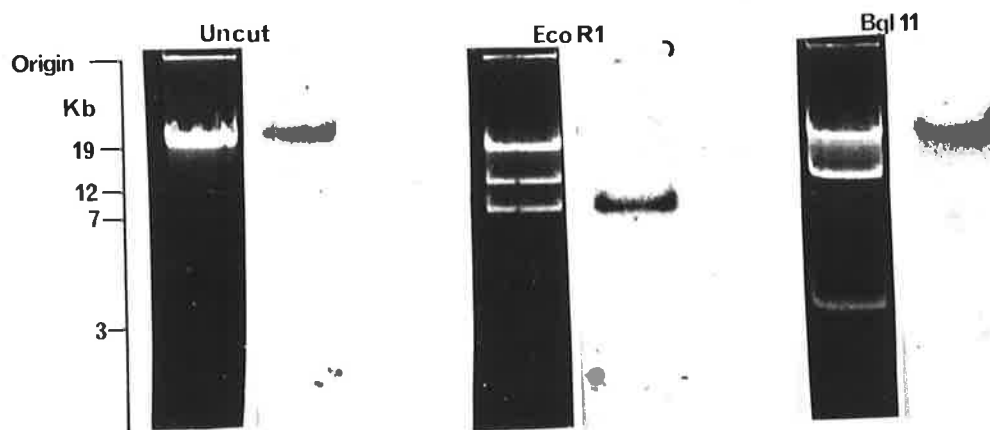
Although these experiments again, do not indicate the absolute abundance of histone sequences in the mRNA, they do demonstrate the ease of use of chick embryo cDNA as a direct histone gene probe without prior rigorous purification of the mRNA. The cDNA could readily detect histone genes in both cloned sea urchin DNA and total genomic sea urchin DNA, and was therefore considered to be of sufficient radiochemical purity for use in the selection of histone recombinants from a chicken gene library. These experiments also demonstrate the cross-reactivity of sea urchin and chicken histone DNAs. Presumably the ease at which histone sequences are detected in sea urchin genomic DNA with chicken cDNA is

FIGURE 3.2

Hybridization of 5-day embryo 7-11S cDNA to the sea urchin histone gene clone λ 55. λ 55 DNA was digested with EcoRI and BglII and, together with undigested DNA, electrophoresed on a 1% agarose gel. Bands were detected with EtBr staining. DNA was transferred to nitrocellulose (section 2.2.2) and hybridized to cDNA made from 5-day embryo 7-11S RNA. Filters were washed in 2XSSC and autoradiographed.

A map showing the relevant restriction sites in λ 55 DNA is included.

CHICK EMBRYO cDNA and S.U. HISTONE GENE X-REACTION



to some extent related to the large copy number of histone genes in sea urchins (300-1000X). The inability to detect histone sequences in chicken DNA using sea urchin probes (section 3.1) may therefore partly be a function of the reduced histone gene copy number in the vertebrate genomes (Kedes, 1979).

3.2.3 Primary Selection of Recombinants

When using chick embryo cDNA to directly screen a recombinant library, positive signals resulting from non-histone sequences in the probe would be expected. As the major contaminant of the cDNA is ribosomal RNA (Crawford *et al.*, 1979), many of these would be ribosomal recombinants. In fact, because of the high reiteration frequency of ribosomal sequences in the chick genome (200x; McClements and Skalka, 1977), these should outnumber the histone recombinants.

The embryo vascular system was also harvested during RNA preparation from embryos. The embryo mRNA would therefore also be expected to contain some significant level of 9S globin mRNA. However, these sequences would only select a relatively low number of globin recombinants because there are only few globin genes in the chick genome (Crawford *et al.*, 1977). It was not considered an advantage to attempt to remove these known contaminants by the RNA excess hybridization of Crawford *et al.*, (1979) for ribosomal sequences, or by hybrid-affinity columns for globin mRNAs, because low levels of these contaminants

may persist and still contribute to selection of recombinants containing non-histone genes. In any case, recombinants selected with known contaminants of the probe could be identified in subsequent screenings with the appropriate probes. Of greater significance are those recombinants which are selected with unknown non-histone constituents of the embryo cDNA. However, these were expected to be few, based on the lack of predominant non-histone translation products when 5 day embryo mRNA is used to programme the wheat germ translation system. It was anticipated that these recombinants could be distinguished from true histone recombinants by cross-species hybridization probes.

Crude cDNA made to the 7-11S fraction of 5-day embryo RNA was therefore used to score a chicken genomic library. This library was made available by J. Dodgeson and was constructed by the methods of Maniatis *et al.*, (1978) from partial AluI and HaeIII digests of chicken genomic DNA. Fragments of chick DNA with an average size of 17 Kb were inserted into the lambda charon vector 4a (Blattner *et al.*, 1977) by addition of synthetic EcoRI linkers. At this time the library had been used to examine chicken globin (Dodgson *et al.*, 1979) and ovalbumin (Gannon *et al.*, 1979) genes, and subsequently has been used in the analysis of a variety of other gene systems. Major rearrangement of DNA that can be attributed to the cloning procedures has not been found in the gene systems that have been analysed.

The chick embryo cDNA had previously been used to probe blots of restriction digests of chick DNA, and banding patterns were detected which suggested that chicken histone genes were clustered into a 15 Kb repeat (Crawford *et al.*, 1979 and section 4.3). Estimates of histone gene number also using this cDNA indicated a reiteration frequency of approximately 10 copies per haploid genome. Based on this data it was calculated that one recombinant in approximately 13,000 would contain histone sequences. The recombinant DNA library was plated at high density and screened in duplicate, with radiolabelled embryo cDNA, by the methods of Benton and Davis (1977). Hybridization conditions were modified from those of D. Kemp (personal communication and section 2.2.6) to include 10% dextran sulphate to accelerate the rate of hybridization (Wahl *et al.*, 1979). Library screenings without dextran sulphate resulted in few signals of weak intensity.

A total of 77 positive signals were observed in duplicate at the first round screening. Figure 3.3 shows positive signals produced from plates containing about 16,000 plaques. The area under the positive signals was picked from the culture plates and resuspended in storage buffer (section 2.2.6). Only plaques whose signals appeared in duplication were selected.

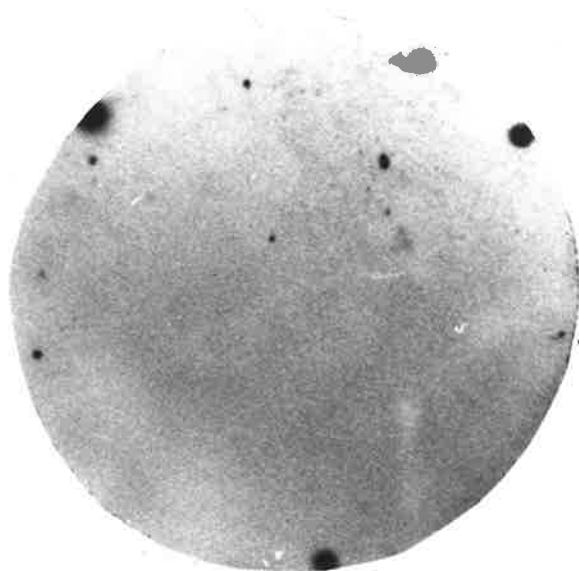
3.2.4 Negative Screening and Plaque Purification

In the primary screening, because plaques were in

FIGURE 3.3

Screening of a chicken recombinant library with 5-day embryo 7-11S cDNA. The chicken library was plated at a density of 16,000 plaques per plate, and from these Benton and Davis filters were prepared (section 2.2.6). Filters were probed with cDNA (10^6 cpm per filter) made from 5-day chick embryo 7-11S RNA, washed in 0.5XSSC and autoradiographed.

Many positive plaques were detected on each plate. The arrows show those positives corresponding to λ CH-01 and λ CH-02 (chapters 4 and 6).



$\lambda 1.6$
(ACH-02)



$\lambda 7.4$
(ACH-01)

contact (confluent lysis of the bacterial lawn), isolates would contain many recombinants in addition to the one producing the positive signal to embryo cDNA. It was therefore necessary to repeat screenings at lower plaque density to a point where a single plaque could be shown to give a positive response and harvested to the exclusion of any contaminating plaques. This procedure usually required an extra two rounds of screening.

It was also necessary to identify and discard those recombinants that were selected with known non-histone contaminants of the cDNA (section 3.2.3). These contaminants were ribosomal, globin and 4S sequences. It seemed advantageous to incorporate this "negative" selection into the second round "positive" screening.

Primary isolates were titred and replated to give distinguishable single plaques. Filters were prepared to 31 of the original isolates and screened with a mixed negative probe. This probe was radiolabelled cDNA (random primed) to chick reticulocyte 9S globin mRNA, and chick scale 18S, 28S, 5S and 4S RNAs. RNAs were selected from these tissues because histone mRNA levels would be expected to be minimal. Of those screened 15 were positive with the "negative" probe and therefore discarded, and 16 were negative. Those 16 filters that showed no response to the "negative" probe were then re-screened with the chick embryo cDNA so that the appropriate single plaques could be picked. Of these, 4 isolates responded to the embryo cDNA. Those that did not respond

could have been either false positives on the primary screen, or true positives that were missed when plaques were picked from the plates. Figure 3.4 shows the signals from second round screening of primary isolates with the negative cDNA probe and embryo cDNA.

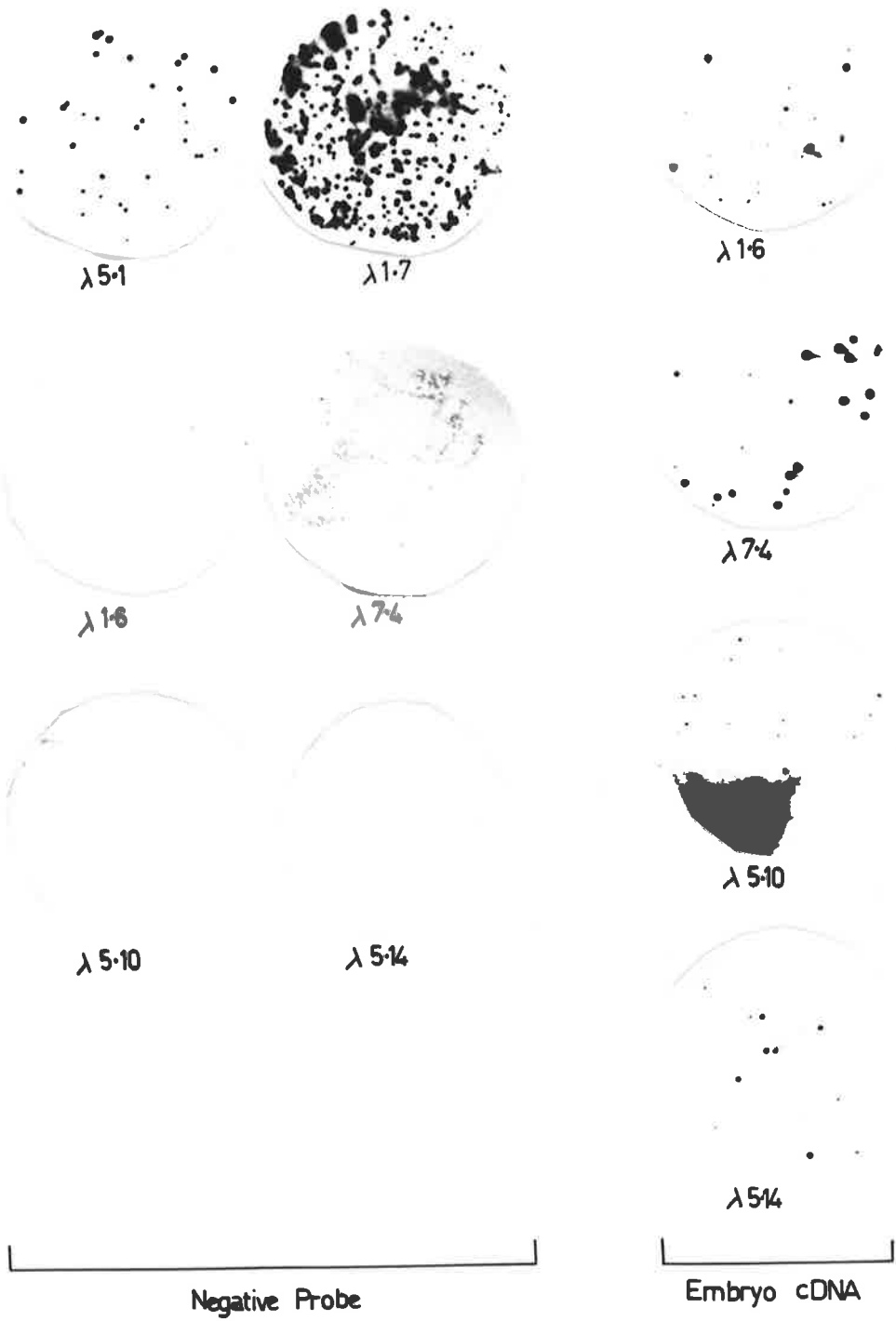
The 4 isolates that had responded in the second round screening to embryo cDNA but not to the negative probe became good candidates for containing chicken histone genes. These were subjected to a further round of screening with embryo cDNA to ensure purity of the plaque isolates.

FIGURE 3.4

Second round, and negative screening of primary isolates from a chicken library.

Primary isolates from screening the chicken genomic library were picked into storage buffer, titred and replated at low density (100-500 plaques per plate). Benton and Davis filters were prepared (section 2.2.6), and hybridized (section 2.2.6) with a "negative" probe consisting of globin, ribosomal and 4S cDNA (10^5 cpm per filter). Filters were washed in 0.5XSSC and autoradiographed. Those filters showing positive plaques were discarded as they were from non-histone recombinants. Those not responding to the "negative" probe were further hybridized to embryo cDNA.

Isolates which responded to embryo cDNA but not to the negative probe (λ 1.6, λ 7.4, λ 5.10 and λ 5.14, shown) became possible chicken histone clones.



CHAPTER 4

IDENTIFICATION OF A CHICKEN

HISTONE-GENE RECOMBINANT

(N.B. Some of the work described
in this chapter has been published;
Harvey, R.P. and Wells, J.R.E. (1979)
N.A.R. 7, 1787-1798)

4.1 INTRODUCTION

Although the most definitive identification of histone recombinants is provided by the DNA sequence of the genes themselves, hybridization studies using cloned histone DNA from other species allows an initial identification. The chick embryo "histone" cDNA had been shown to cross-hybridize with specific fragments in total sea urchin DNA and also with cloned sea urchin DNA (3.2.2). The use of sea urchin histone DNA to probe possible chicken histone recombinants, might therefore identify clones on which to focus a DNA sequencing analysis.

This chapter describes the identification of histone genes in one recombinant, λ 7-4, both by hybridization studies, and by DNA sequencing.

4.2 RESULTS

4.2.1 Subcloning λ 55 insert into pBR322

The sea urchin histone gene recombinant, λ 55, was constructed by K. Murray and E. Southern (personal communication) by insertion of sea urchin histone DNA into an EcoRI lambda cloning vector. In order to screen chicken recombinants using this sea urchin histone probe, it was necessary to subclone the λ 55 insert into a vector which would not cross hybridize with the lambda system. The plasmid vector pBR325 (Bolivar, 1978) was chosen because it allowed antibiotic selection of insertions into the EcoRI site.

λ 55 insert was released with EcoRI and isolated on a sucrose gradient. The insert was then ligated into EcoRI linearized and dephosphorylated pBR325, and a portion was transformed into competent *E. coli* (section 2.2.4). Seven colonies showing chloramphenicol sensitivity were grown in 1 ml cultures and the plasmid DNA was extracted (section 2.2.4), digested with EcoRI, and displayed on an agarose gel (Fig. 4.1(a)). Four of these transformants contained both insert and vector DNA. All inserts were 7 Kb, the size expected for the sea urchin histone fragment. To further check that these inserts contained histone DNA, the DNA in the gel was blotted onto nitrocellulose and probed with embryo cDNA (Fig. 4.1(b)). The inserts of all four putative sub-clones hybridized to the cDNA. One of these was chosen and plasmid DNA was isolated in bulk to use in the cross species hybridizations described in this chapter. The plasmid containing the sea urchin insert was called pSU55-E.

4.2.2 Identification of coding regions in λ 7.4

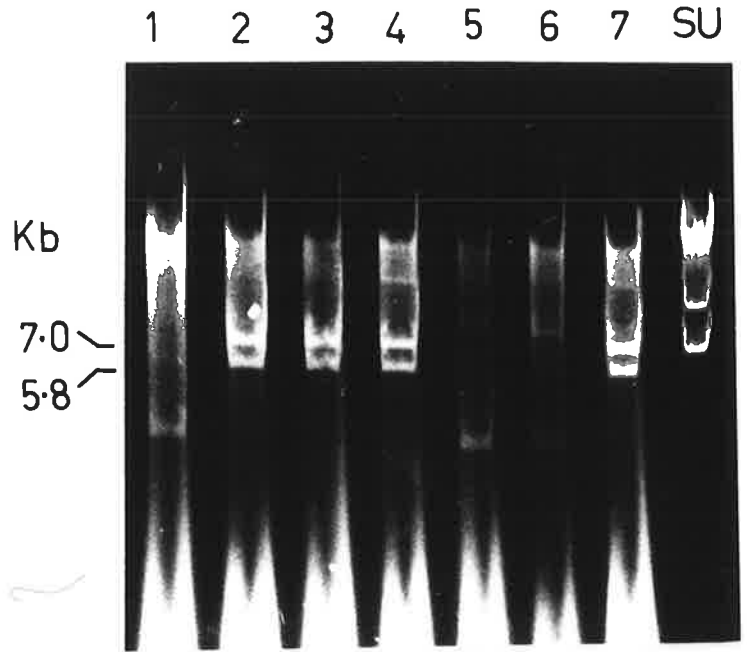
The recombinant giving the strongest signal with ^{32}P chick embryo cDNA on Benton and Davis filters, λ 7.4 (Fig. 3.4), was chosen for further analysis. Phage particles were amplified in liquid culture, concentrated by polyethylene glycol precipitation and purified on CsCl step gradients (section 2.2.6). DNA was isolated by phenol/chloroform extraction of concentrated phage stocks.

FIGURE 4.1

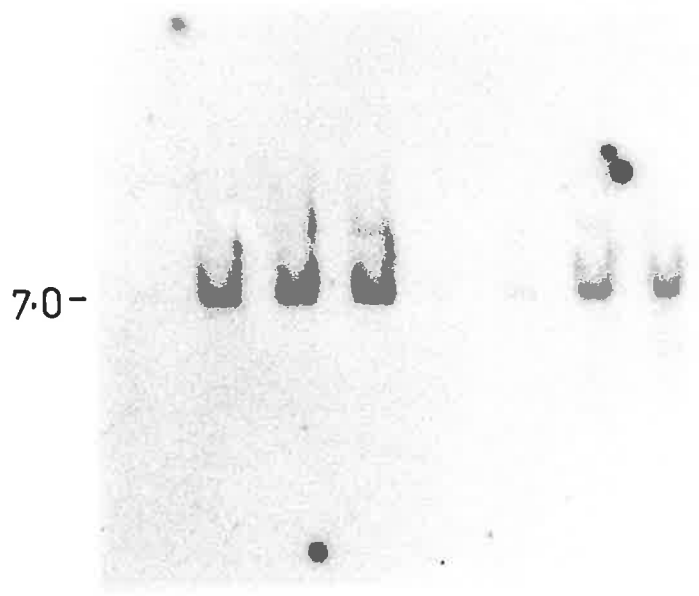
Subcloning of the λ 55 insert into pBR325
(section 4.2.1).

- (a) Mini-plasmid preparations of transformants electrophoresed on a 0.8% agarose gel and detected by EtBr staining. SU is EcoRI cut λ 55 DNA to serve as marker.

- (b) The gel shown in (a) was transferred to nitrocellulose, hybridized with 5-day embryo cDNA and autoradiographed.



(a)



(b)

Preliminary restriction endonuclease analysis was performed on the DNA to determine a suitable digestion pattern in which insert DNA migrated away from charon 4a vector DNA on agarose gels. Figure 4.2(a) shows EcoRI and HindIII digestions of λ 7.4 DNA. The EcoRI/HindIII double digest was suited to hybridization analysis as all insert DNA and vector DNA bands were well resolved. EcoRI/HindIII digests were then electrophoresed on an agarose gel and transferred to nitrocellulose (section 2.2.2). One filter was hybridized with chick embryo cDNA, and another with nick translated (section 2.2.3) pSU55-E DNA (Fig. 4.2(b)). Both probes detected fragments of 3.75 Kb and 3.3 Kb. Construction of an EcoRI and HindIII restriction map of λ 7.4 revealed that these two fragments were separated by 6 Kb of DNA that did not hybridize to either probe (Fig. 4.2(b)). The hybridization of the two fragments to embryo cDNA defined these regions as the general coding domains within λ 7.4. Although it was possible that non-histone coding regions of pSU55-E were participating in its hybridization to λ 7.4, this seemed unlikely as the sea urchin probe only detected the coding domains. These hybridizations strongly suggested that λ 7.4 contained histone genes.

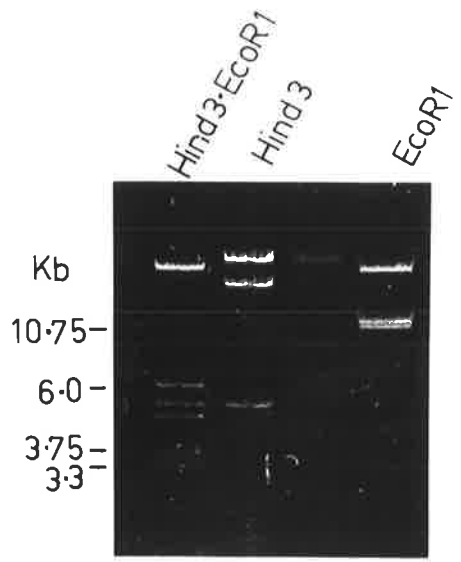
4.2.3 Hybridization of λ 7.4 probes to genomic DNA

The "Southern" hybridization patterns expected from probing histone DNA to total sea urchin DNA and to total

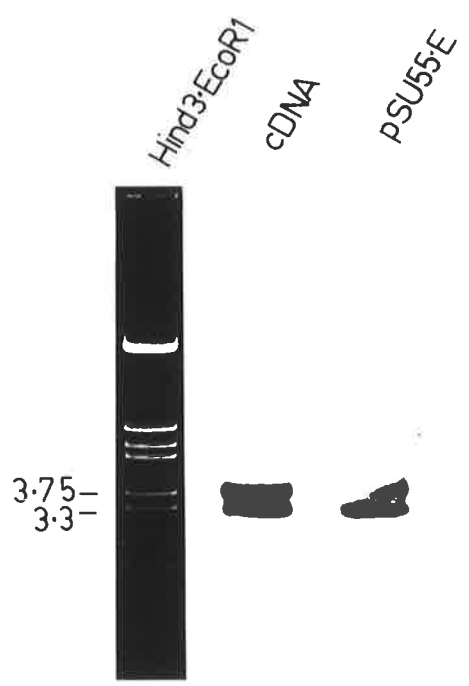
FIGURE 4.2

The coding domains of $\lambda 7.4$.

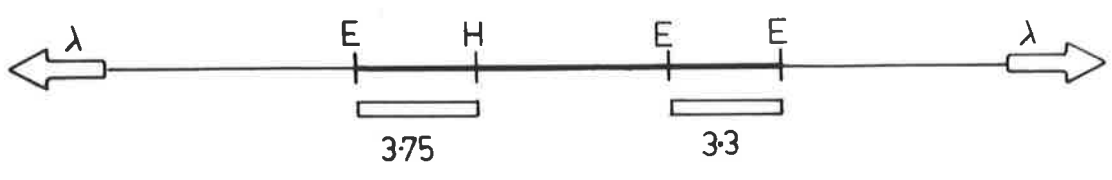
- (a) $\lambda 7.4$ DNA was restricted with EcoRI and HindIII, electrophoresed on a 1% agarose gel and detected by EtBr staining.
- (b) Duplicate HindIII/EcoRI digests of $\lambda 7.4$ DNA were electrophoresed on a 1% agarose gel and transferred to nitrocellulose (section 2.2.2). One filter was hybridized to embryo cDNA and the other to nick-translated pSU55-E (section 4.2.1). After washing in 2XSSC, filters were autoradiographed.
- (c) EcoRI (E) and HindIII (H) restriction map of $\lambda 7.4$ DNA determined from the digestions in (a). The blocked areas show the regions which hybridized to embryo cDNA and pSU55-E. Sizes are in Kilobase-pairs.



(a)



(b)



(c)

chicken DNA were known. In the sea urchin *E. esculentus*, histone DNA exists as tandemly repeated clusters with allelic sizes of 6.3 Kb and 7 Kb (K. Gross, personal communication). EcoRI cuts both repeat classes once. Therefore, in an EcoRI digest of *E. esculentus* DNA isolated from many individuals, histone DNA occurs predominantly in fragments of 6.3 Kb and 7 Kb.

Analysis of total chicken DNA with chick-embryo "histone" cDNA indicated that the histone genes in chickens were clustered into a 15 Kb tandem repeat (Crawford *et al.*, 1979). Southern blots showed a single 15 Kb fragment in EcoRI digests, and an extremely large fragment in BglII digests. This suggested that histone genes were arranged in 15 Kb repeats, all of which were cut once by EcoRI, and most of which were uncut by BglII.

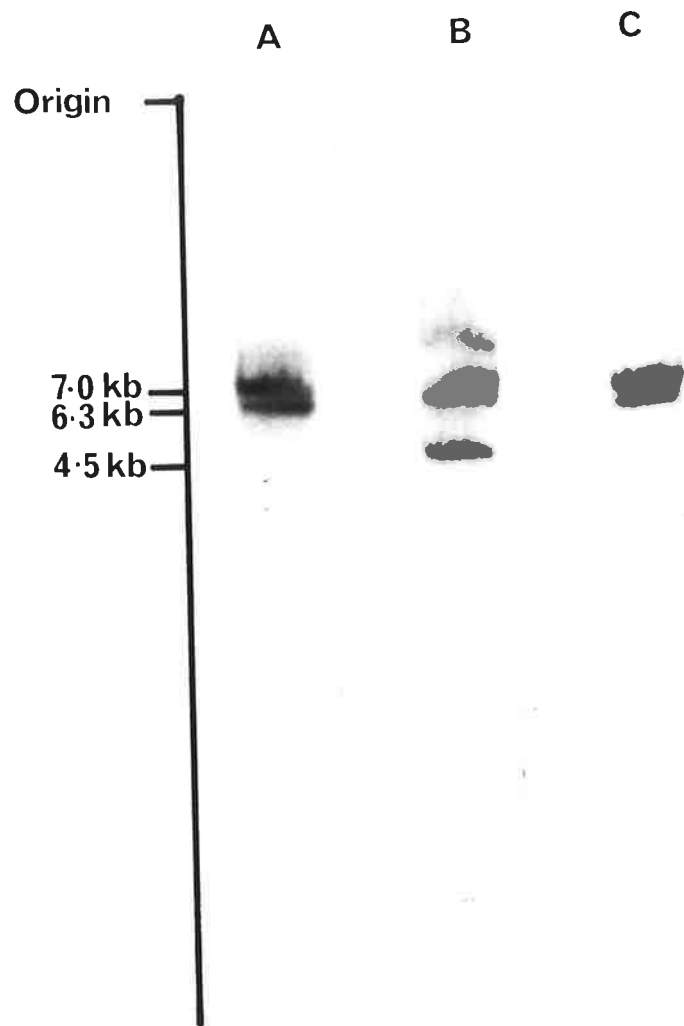
In order to obtain further evidence that $\lambda 7-4$ would behave as histone DNA, digests of *E. esculentus* DNA (EcoRI), and of chicken DNA (EcoRI, HindIII, BamHI, BglII) were electrophoresed on agarose gels and transferred to nitrocellulose. Separate sea urchin filters were probed with total $\lambda 7-4$ and total pSU55-E DNA, labelled by nick-translation. Chicken digests were hybridized with both embryo cDNA from which the ribosomal sequences had been removed (Crawford *et al.*, 1979) and nick-translated $\lambda 7-4$ DNA.

As expected, pSU55-E probe detected predominantly 6.3 and 7 Kb fragments in EcoRI digested sea urchin DNA (Fig. 4.3(a)). The $\lambda 7-4$ probe, when hybridized to an

FIGURE 4.3

Hybridization of λ 55 and λ 7.4 probes to sea urchin DNA. Total *E. esculentus* DNA was digested with EcoRI, electrophoresed in a 1% agarose gel and transferred to nitrocellulose (section 2.2.2).

- A. Filter hybridized to nick-translated (section 2.2.3) λ 55 insert.
- B. Filter hybridized to nick-translated λ 7.4 DNA.
- C. Filter hybridized to nick-translated 3.3 Kb EcoRI fragment derived from λ 7.4 DNA (section 4.2.3).



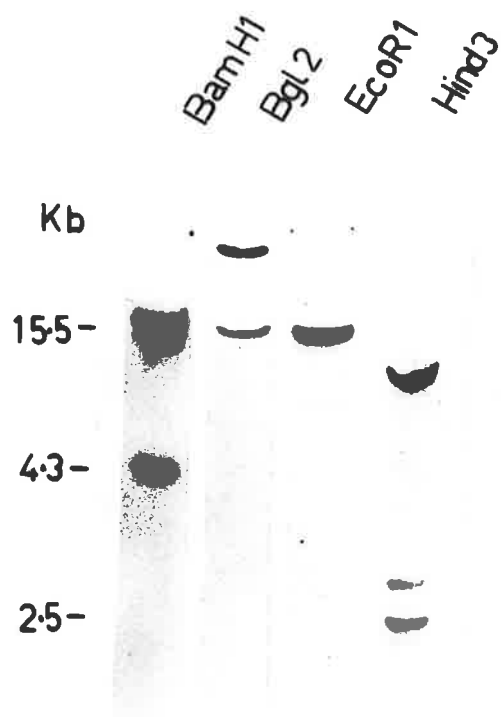
equivalent digest (Fig. 4.3(b)), also detected these bands, and in addition, another fragment of 4.5 Kb. It was possible that this additional fragment arose from non-coding sequences in λ 7-4 hybridizing to sequences in sea urchin DNA. λ 7-4 contains at least 6 Kb of insert DNA which does not hybridize to embryo cDNA (Fig. 4.2(b)). To limit this possibility, one of the coding domains within λ 7-4 (3.3 Kb EcoRI fragment, Fig. 4.2) was isolated by sucrose gradient centrifugation of EcoRI digested λ 7-4 DNA, and used to probe an EcoRI digest of sea urchin DNA (Fig. 4.3(c)). In this case only the fragments expected for a histone DNA probe were detected (6.3 and 7 Kb).

The hybridization patterns to the chick digests were contradictory (Fig. 4.4). In contrast to the relatively simple patterns obtained with embryo cDNA, patterns obtained with the λ 7-4 probe were complex. Diagnostically, no large BglIII fragment was detected. Although bands in the region of 15 Kb were seen in EcoRI, BamHI and HindIII digests, many other bands were also apparent. Again, it was possible that the complex patterns obtained were a result of the extensive non-coding sequences of λ 7-4 participating in the hybridizations. To see if the blot patterns were simplified by enriching for coding sequences within λ 7-4, the two coding domains indicated in Fig. 4.1(b) were isolated and used to probe EcoRI and HindIII digests of chicken DNA. As indicated in Figure 4.5, the use of these probes did not simplify the patterns obtained with total λ 7-4 probe.

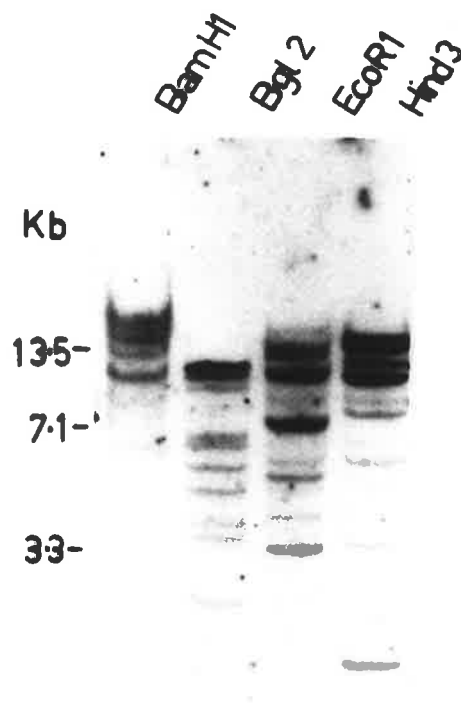
FIGURE 4.4

Hybridization of λ 7.4 and embryo cDNA probes to chicken genomic DNA. Chicken DNA was digested with BamHI, BglII, EcoRI and HindIII, electrophoresed on 1% agarose gels and transferred to nitrocellulose (section 2.2.2).

- (a) Filters hybridized to chicken 5-day embryo 7-11S cDNA from which the ribosomal sequences had been removed (Crawford *et al.*, 1979). Data of R. Crawford reproduced from Crawford *et al.*, (1979).
- (b) Filters hybridized to nick-translated (section 2.2.3) λ 7.4 DNA.



(a)

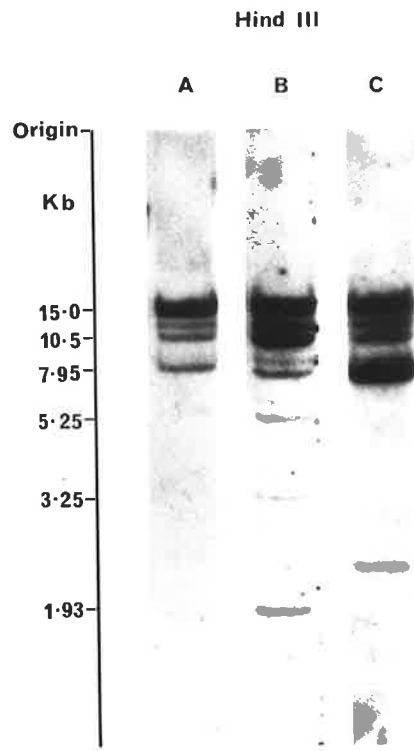
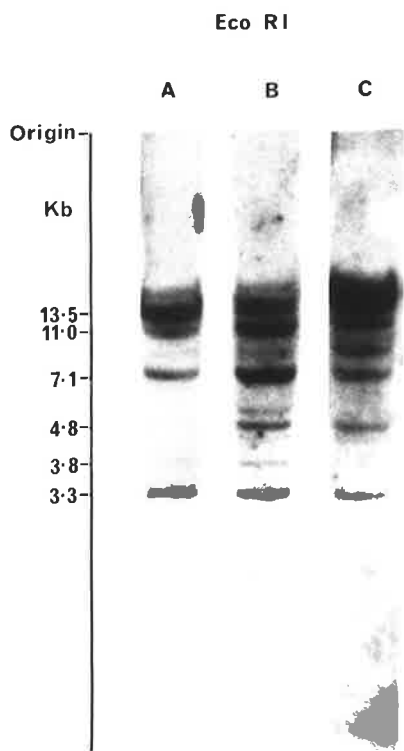


(b)

FIGURE 4.5

Hybridization of λ 7.4 probes to chicken genomic DNA. DNA was digested with EcoRI and HindIII, electrophoresed on a 1% agarose gel and transferred to nitrocellulose (section 2.2.2).

- A. Filters hybridized to nick-translated (section 2.2.3) λ 7.4 DNA.
- B. Filters hybridized to nick-translated 3.3 Kb coding domain of λ 7.4 (Fig. 4.2).
- C. Filters hybridized to nick-translated 3.75 Kb coding domain of λ 7.4 (Fig. 4.2).



These results questioned either the nature of the sequences in $\lambda 7-4$, or the validity of the Southern blots obtained with embryo cDNA. Independent of these findings however, $\lambda 7-4$ appeared to contain histone sequences because of its interaction with sea urchin DNA. It therefore seemed a good candidate on which to concentrate a search for chicken histone gene sequences.

4.2.4 Shotgun sequencing to identify histone genes

The coding regions in $\lambda 7-4$ had been localized to two domains in an EcoRI/HindIII digest of $\lambda 7-4$ DNA (Fig. 4.1(b)). The 3.3 Kb domain was the most readily isolated, as it could be resolved from other fragments on sucrose gradients. A random sequencing strategy was therefore applied to this 3.3 Kb fragment in an attempt to identify histone gene coding sequences. The strategy was to cut the 3.3 Kb fragment into a suitable array of subfragments and then systematically sequence these until a histone gene region was encountered. To produce the array of subfragments, the restriction enzyme HaeIII was chosen, as this recognizes the base sequence 5'GGCC3'. Sea urchin histone gene coding regions are rich in G/C and have a codon bias towards G and C in the third position (Kedes, 1979). If chicken histone genes retained this bias, HaeIII should cut frequently within histone genes.

An HaeIII digest of purified 3.3 Kb fragment was dephosphorylated with bacterial alkaline phosphatase and end-labelled with ^{32}P using polynucleotide kinase (section

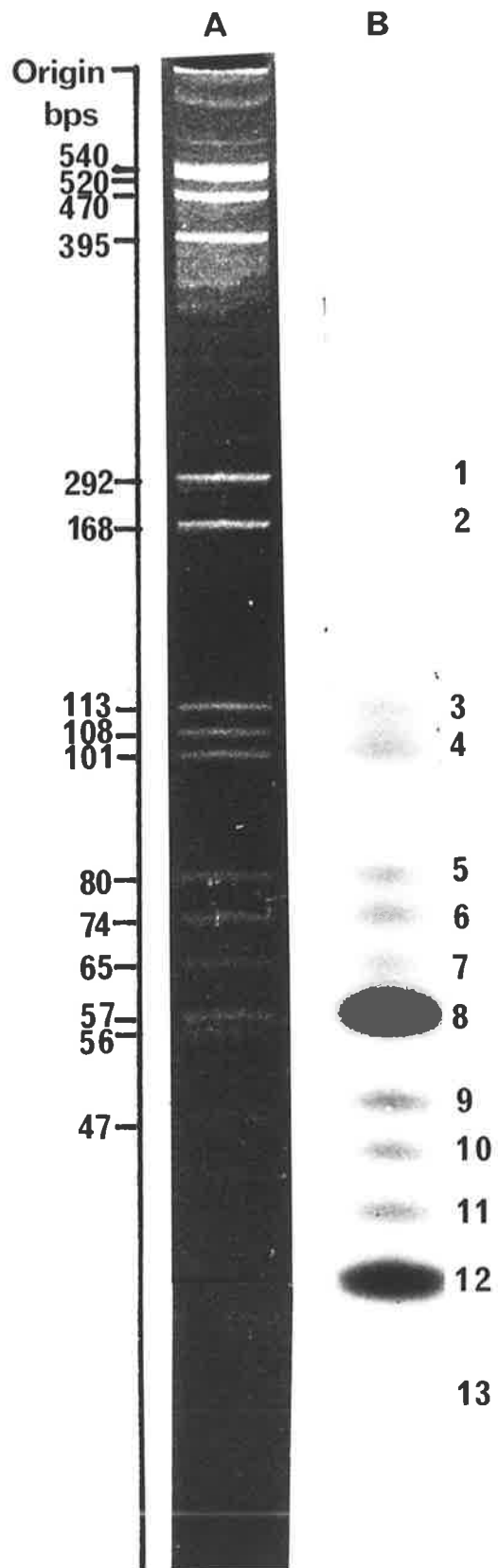
2.2.7). The labelled fragments were electrophoresed on a polyacrylamide gel and the gel was autoradiographed (Fig. 4.6). A gradient of labelling was observed with smallest bands labelling more efficiently. This may have been due to inefficient elimination of the alkaline phosphatase which becomes active in the gel in the higher M.W. regions and therefore removes the radionucleo-phosphate from larger fragments (Chaconas and van de Sande, 1980).

The two very intense bands in Figure 4.6 were probably due to labelling of the fragments containing the two EcoRI ends. These have overhanging 5'-termini and label much more efficiently than the blunt-ended, HaeIII generated termini (section 6.2.1).

Labelled bands were excised from the gel and the DNA electro-eluted and precipitated with ethanol. To subject fragments to the sequencing procedures of Maxam and Gilbert (1980), one labelled strand is required. This can be achieved by further restriction of a kinased fragment, or by strand separation. Eluted DNA was denatured in DMSO and electrophoresed through polyacrylamide gels suitable for resolution of single stranded DNA (section 2.2.7). Most fragments resolved into their respective single strands. After autoradiography, the single-stranded components were excised, electro-eluted and recovered by ethanol precipitation. Poor recovery precluded further analysis of some fragments. Those with sufficient radioactivity remaining, were sequenced using

FIGURE 4.6

Shotgun sequencing of λ 7.4 DNA. The 3.3 Kb coding domain of λ 7.4 DNA was isolated, digested with HaeIII and kinased (section 4.2.4). The kinased fragments were electrophoresed on a 6% polyacrylamide gel and autoradiographed. A HaeIII digest of 3.3 Kb DNA was electrophoresed in parallel and bands detected by EtBr staining.



the chemical degradation methods of Maxam and Gilbert (1980). Reactions were electrophoresed on thin denaturing polyacrylamide gels and autoradiographed. Short readable sequences were obtained from fragments 7,8,9,11 and 12 (Fig. 4.7). These sequences were translated into amino-acid sequences in all possible reading frames, and the amino-acid sequences examined for homology with known histone protein sequences from various species (Sures *et al.*, 1978; Elgin and Weintraub, 1975).

Fragments 9 and 11, when translated in one reading frame, predicted amino-acid sequences with perfect homology to the known sequence of chicken histone H2A from erythrocytes (Fig. 4.7). Also in one reading frame, fragment 8 predicted an amino-acid sequence with good homology to trout and rabbit H1 sequences. The sequence of chicken H1 proteins are not known.

These homologies identified λ 7-4 as a histone gene containing recombinant, and thus worthy of detailed characterization.

4.3 DISCUSSION

The recombinant λ 7-4 was positively identified as containing chicken histone genes on the basis of three criteria:

- a) by cross-hybridization with cloned sea urchin histone genes
- b) by cross-hybridization with histone genes in total sea urchin DNA
- c) by direct DNA sequence analysis.

FIGURE 4.7

DNA sequences obtained from shotgun sequencing of HaeIII fragments derived from the 3.3 Kb, EcoRI coding domain of λ 7.4 DNA. Fragment numbers are relative to bands excised from the acrylamide gel shown in Fig. 4.6.

FRAGMENT 7: 5' GCACCGGCCGCTGCGCAAGGGCAACACGCGGAAGGCGGG 3'

FRAGMENT 8: 5' GCT GAG CCC AAG GCT GCC AAG CCC AAG GCG ACC AAA
ala glu pro lys ala ala lys pro lys ala thr lys

TROUT H1 183 ALA LYS PRO LYS ALA ALA LYS PRO LYS ALA ALA LYS
RABBIT H1 ALA LYS PRO LYS ALA PRO LYS PRO LYS ALA ALA LYS

FRAGMENT 9: 5' C AAC AAG AAG ACG CGC ATC ATC CCC CGC CAC CTG CAG CTG
asn lys lys thr arg ile ile pro arg his leu gln leu

CHICKEN H2A 73 ASN LYS LYS THR ARG ILE ILE PRO ARG HIS LEU GLN LEU

FRAGMENT 11: 5' TG CTG CTG CCC AAG AAG ACC GAC AGC GA
gln
leu leu pro lys lys thr asp ser his

CHICKEN H2A LEU LEU PRO LYS LYS THR ASP SER HIS

FRAGMENT 12: 5' GCCGGCGCCTGGGGGCGGCGACGTGGCGGG

DNA sequencing identified one gene within the 3.3 Kb coding domain. The presence of two split coding domains (Fig. 4.1(b)), however, suggests that multiple genes are present in $\lambda 7-4$. This is also supported by the recognition of an H1-like sequence in the 3.3 Kb domain. The predicted H1 amino-acid sequence varies slightly from the trout and rabbit H1 sequences, however H1 is the most variable of all the histones (Kinkade, 1969), and perfect evolutionary conservation might not be expected. In addition, the region of the H1 from which the sequence is derived adopts a random-coil configuration, both when bound to DNA and when in solution (Hartman *et al.*, 1977). Because this region of the molecule is not subject to the same degree of structural constraints that apply for other histones, some amino-acid variability can be tolerated, and this is reflected in evolutionary comparisons (von Holt, 1979). It is possible then, that this sequence represents a true H1 gene and that the 3.3 Kb domain contains both H2A and H1 genes.

That the $\lambda 7-4$ -derived probes used in hybridizations to chicken genomal DNA (section 4.2.3) do in fact contain histone DNA, casts some doubt on the blot patterns obtained by Crawford *et al.*, (1979) with embryo cDNA. Using $\lambda 7-4$ probes there is no evidence of a 15 Kb repeating structure. Although these probes are somewhat undefined, they rather suggest that histone genes are arranged in the chicken genome in a complex manner.

Sittman *et al.*, (1981) have reported isolation of mouse histone genes using a mouse cDNA probe. In hybridization experiments to mouse genomic DNA using this probe, they also detected a 15 Kb repeating structure, but determined it to be a result of contaminating mitochondrial DNA hybridizing to mitochondrial ribosomal sequences present in the cDNA probe. This would appear also to be the source of indications of the 15 Kb repeating unit in chicken DNA, and would explain the contradiction between these results and those obtained using cloned histone probes.

The isolation of λ 7-4 enables new examination of both detailed and long-range organization of chicken histone genes. This clone (λ 7-4) was subsequently called λ CH-01.

CHAPTER 5

THE CODING REGIONS OF λCH-01

(N.B. Some of the data included
in this chapter has been published:
Harvey, R.P., Krieg, P.A., Robins, A.J.,
Coles, L.S. and Wells, J.R.E. (1981)
Nature 294, 49-53).

5.1 INTRODUCTION

Histone genes from a number of invertebrate species show certain structural similarities (Chapter 1). Genes are clustered into tandem repeating units each containing one copy of the 5 histone genes and inter-gene non-coding regions. The conservation of this format through evolution may reflect constancy in the mode of expression of histone genes in different species, and may imply that an intimate relationship exists between histone gene conformation and expression. It was recognized that this "unit structure" might be related to the need for stoichiometric amounts of core histones for nucleosome assembly. Transcription of histone DNA into polycistronic mRNAs was postulated (Melli *et al.*, 1977) but has not been proven (Childs *et al.*, 1979). Little, however, is known of how eukaryotic genes are controlled through development or the cell cycle. The structure of gene conformations which are conserved through evolution, therefore, may still shed light on aspects of control.

In the analysis of histone genes in vertebrates, several relevant questions can be asked in relation to this continuity of gene conformation. Are the genes clustered? Does a repeat unit exist? Is such a repeat unit, (if present), conserved? Before dissecting the microstructure of individual genes, knowledge of overall arrangement of histone genes is needed as this may have relevance to control of specific genes.

This chapter describes the use of homologous and heterologous gene coding probes to further delineate the coding potential of λ CH-01.

5.2 RESULTS

5.2.1 Location of the H2A Gene in the 3.3 Kb Coding Domain

To allow further analysis of the 3.3 Kb coding domain, this EcoRI fragment was subcloned into the plasmid vector pBR325 (S. Clarke, this laboratory). The resultant subclone was termed pCH3.3E. Single and double restriction digests were performed on pCH3.3E DNA to establish a restriction map. Figure 5.1 shows digests using the enzymes EcoRI, HindII, HindIII, PstI, PvuII, SmaI and XhoI which had been electrophoresed on agarose gels. The restriction map established from these digests is shown in Figure 5.2.

The sequences obtained from the shotgun sequencing of the 3.3 Kb coding domain (section 4.2.4) were scanned for restriction sites. Fragment 9 which was derived from H2A-like sequences, contained overlapping PstI and PvuII sites (Fig. 5.2). This fragment could then be positioned on the pCH3.3E map because it contained a single PvuII site mapping extremely close to a PstI site. The presumptive H2A gene was therefore assigned to this region (Fig. 5.2).

5.2.2 Other Coding Regions within pCH3.3E

To determine whether pCH3.3E contained other genes, the digests shown in Figure 5.1(a) were blotted to nitro-

FIGURE 5.1

Restriction analysis of pCH3.3E (section 5.2.1). pCH3.3E DNA was digested with EcoRI, HindII, HindIII, PstI, PvuII, SmaI and XhoI, and electrophoresed on 1% agarose gels. Bands were detected by EtBr staining. The restriction map of pCH3.3E deduced from these digestions is shown in Fig. 5.2.

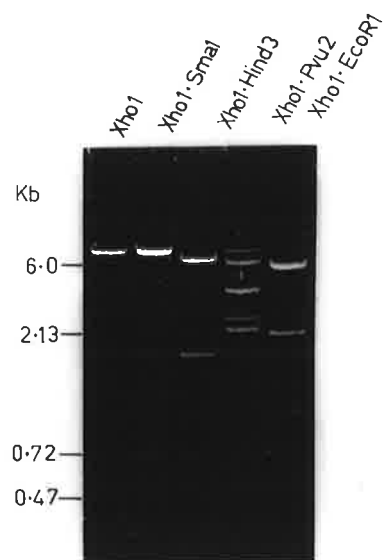
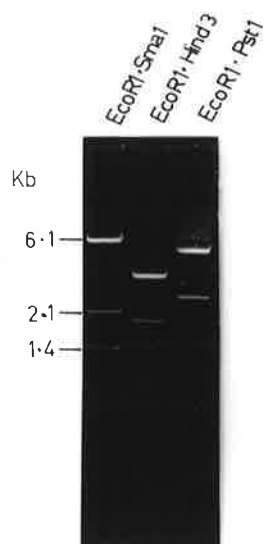
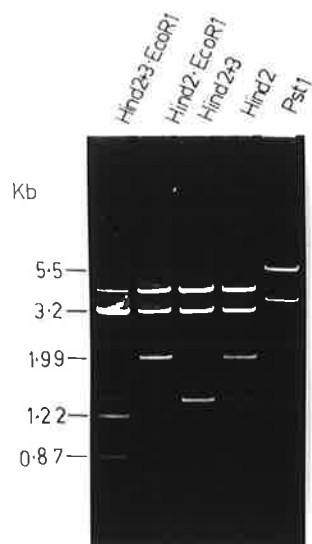
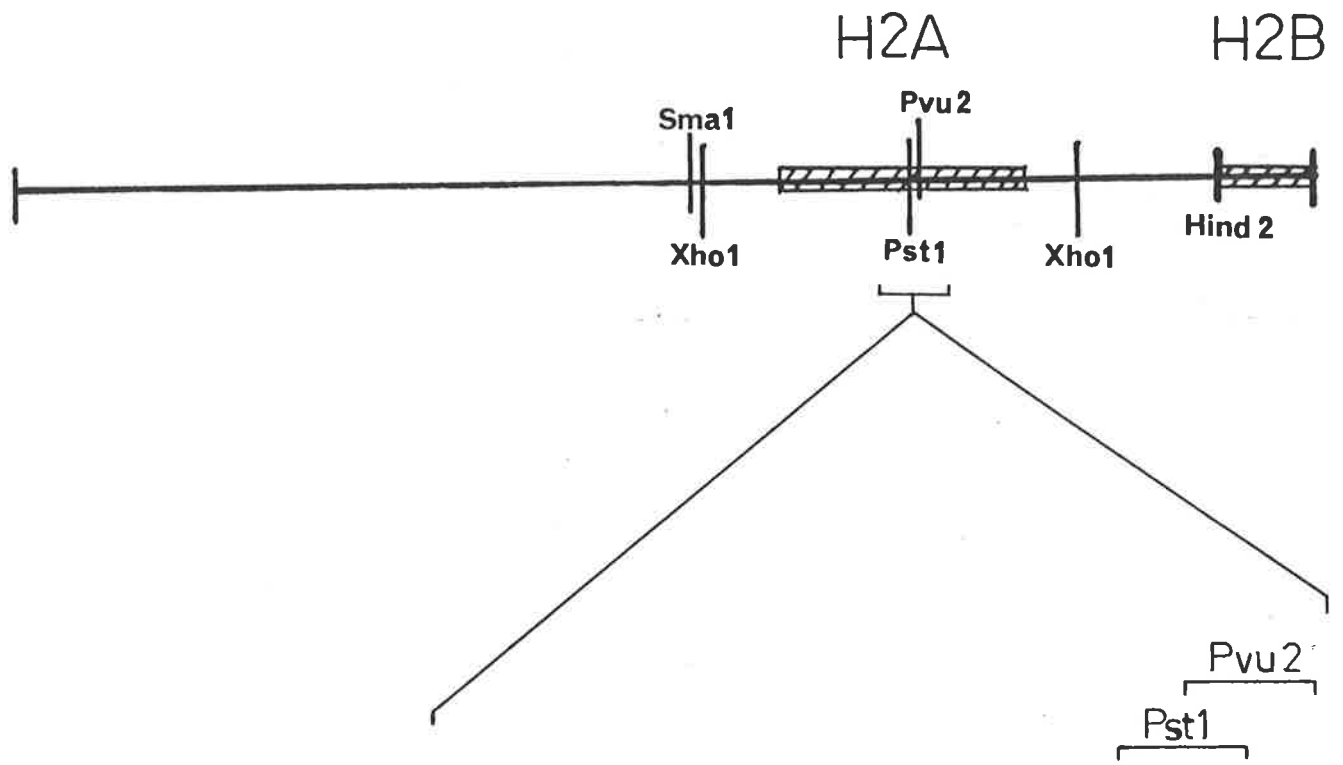


FIGURE 5.2

The restriction map of pCH3.3E insert. Digestions shown in Fig. 5.1 were used to determine the location of restriction sites within the pCH3.3E insert. The DNA sequence of a fragment derived from 3.3 Kb DNA (fragment 9, Fig. 4.7) predicted an H2A sequence and contained overlapping PstI and PvuII sites. This allowed positioning of the H2A coding region (blocked in). The HindII/EcoRI fragment which hybridized to embryo cDNA, and was shown to contain an H2B gene (section 5.2.2) is also blocked in.



fragment 9

CHICKEN H2a

C	AAC	AAG	AAG	ACG	CGC	ATC	ATC	CCC	CGC	CAC	CTG	CAG	CTG
	asn	lys	lys	thr	arg	ile	ile	pro	arg	his	leu	gln	leu
⁷³	ASN	LYS	LYS	THR	ARG	ILE	ILE	PRO	ARG	HIS	LEU	GLN	LEU

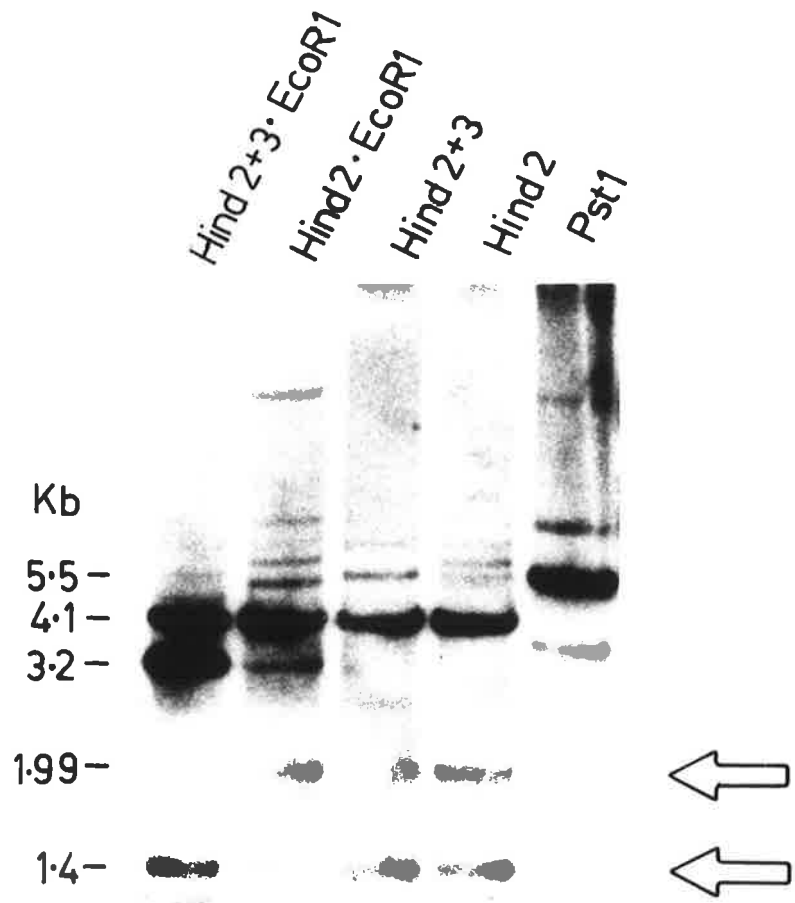
cellulose and hybridized with embryo cDNA (Fig. 5.3). In addition to fragments which obviously carried part of the H2A gene, the 1.4 Kb HindII/HindIII, or the 1.99 Kb HindII bands were also detected with the cDNA. This indicated that some coding sequences were present in the HindII/EcoRI region of the pCH3.3E insert (Fig. 5.2). Assuming modest untranslated regions, the H2A gene would not extend more than 300 bps either side of the PstI/PvuII sites provided there were no intervening sequences. Given that histone genes which contain intervening sequences in their coding regions have not been reported, it was likely that the HindII/EcoRI fragment contained a new coding region rather than part of a split H2A gene. Because of the H1-like sequence identified in the 3.3 Kb coding domain (section 4.2.4) it was possible that this fragment contained an H1 gene.

No cross species H1 probes were available, so the DNA sequence of this fragment was determined to identify the gene (see Chapter 8 for a description of this work). The sequence data indicated that this region contained an H2B gene. The complete sequence of the SmaI/EcoRI fragment (Fig. 5.2) was determined and this contained only the H2A and H2B genes.

In other hybridizations with embryo cDNA, only the SmaI/EcoRI fragment containing these genes was detected. This implied that either the H1-like sequence was encountered by chance or that a true H1 gene in pCH3.3E was not being detected with embryo cDNA. As H1 sub-types

FIGURE 5.3

Hybridization of pCH3.3E with 5-day embryo cDNA. The digestions of pCH3.3E in the gel shown in Fig. 5.1(a) were transferred to nitrocellulose (section 2.2.2). The filter was probed with cDNA made to 5-day embryo 7-11S RNA (section 2.2.3), washed in 0.5 X SSC and autoradiographed. The arrows indicate the 1.99 Kb HindII band and the 1.4 Kb HindII/HindIII band which hybridized to the embryo cDNA, yet did not contain fragments of the H2A gene (section 5.2.2).



are regulated through development, it is possible that embryo cDNA might not cross-hybridize with H1 genes expressed in adult tissues.

Hybridizations to λ CH-01 DNA with sea urchin H3 and H4 gene probes (L. Coles, this laboratory) indicated that pCH3.3E did not contain these genes.

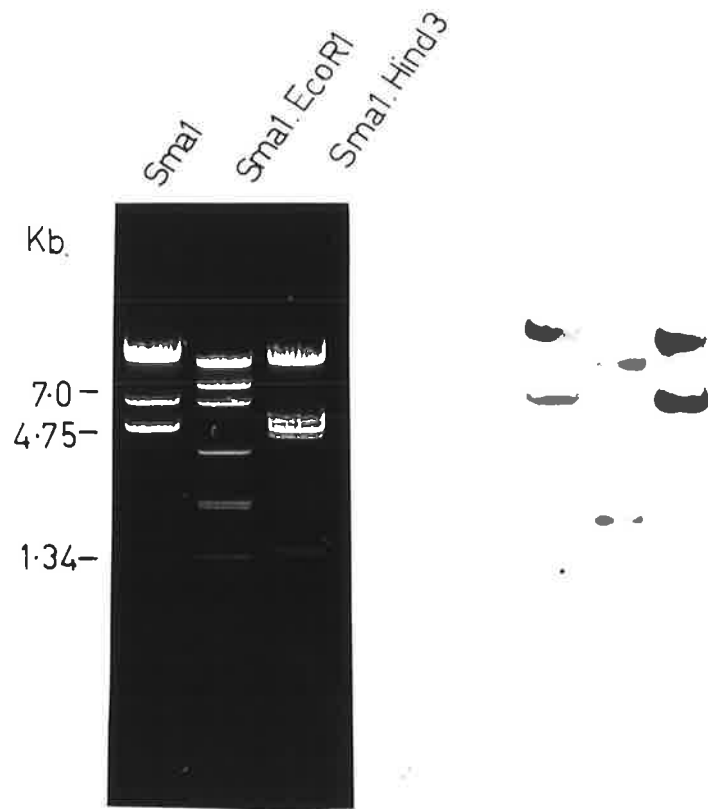
5.2.3 The Orientation of the 3.3 Kb Coding Domain within λ CH-01

In EcoRI digests of λ CH-01 DNA run on 2% agarose gels, a 0.7 Kb fragment was also detected. This fragment mapped to one side of the 3.3 Kb EcoRI fragment. To order these EcoRI fragments and to orientate the 3.3 Kb fragment within λ CH-01, the SmaI sites in λ CH-01 were mapped. The 3.3 Kb fragment could thus be orientated because it contained an asymmetric SmaI site. SmaI, SmaI/HindIII and SmaI/EcoRI digests of λ CH-01 DNA were electrophoresed on an agarose gel with size markers (Fig. 5.4(a)). These digests allowed the positioning and orientation of the EcoRI fragments as shown in Figure 5.4(c). To ensure the interpretation of these digests was correct, the gel shown in Figure 5.4(a) was blotted to nitrocellulose and probed with the 0.72 Kb XhoI fragment of pCH3.3E containing the entire H2A gene (D'Andrea *et al.*, 1981). The autoradiograph of this hybridization is shown in Figure 5.4(b). In addition to the fragments predicted to contain the H2A gene derived from the 3.3 Kb EcoRI fragment, an additional fragment in

FIGURE 5.4

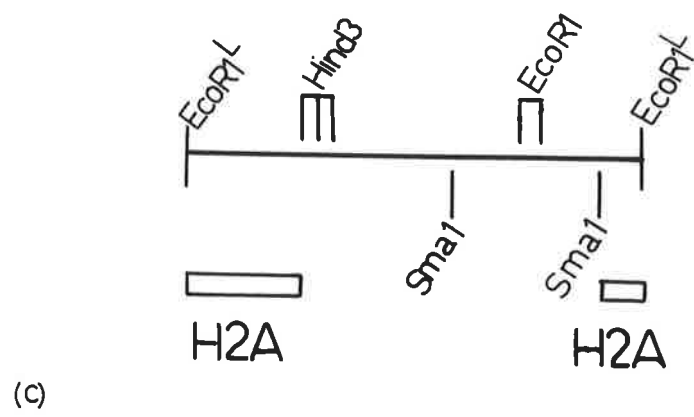
Orientation of the 3.3 Kb coding domain within λ CH-01: the H2A coding regions of λ CH-01.

- (a) λ CH-01 DNA was digested with SmaI, EcoRI and HindIII and electrophoresed on a 1% agarose gel (section 2.2.2). Bands were detected by EtBr staining.
- (b) The gel in Fig. 5.4(a) was transferred to nitrocellulose (section 2.2.2) and probed with H2A probe (the 0.72 Kb XhoI fragment of pCH3.3E, labelled by nick-translation (section 2.2.3)). The filter was washed in 0.5XSSC and autoradiographed.
- (c) Restriction map of λ CH-01 insert orientating the 3.3 Kb EcoRI coding domain, and indicating the regions of hybridization to the H2A probe (Fig. 5.4(b)).



(a)

(b)



(c)

each digest was also detected. The region of hybridization corresponded to the 3.75 Kb HindIII/EcoRI coding domain of λ CH-01 (Fig. 5.4(c)), suggesting that λ CH-01 contained two H2A genes.

5.2.4 The Coding Regions of the 3.75 Kb Coding Domain

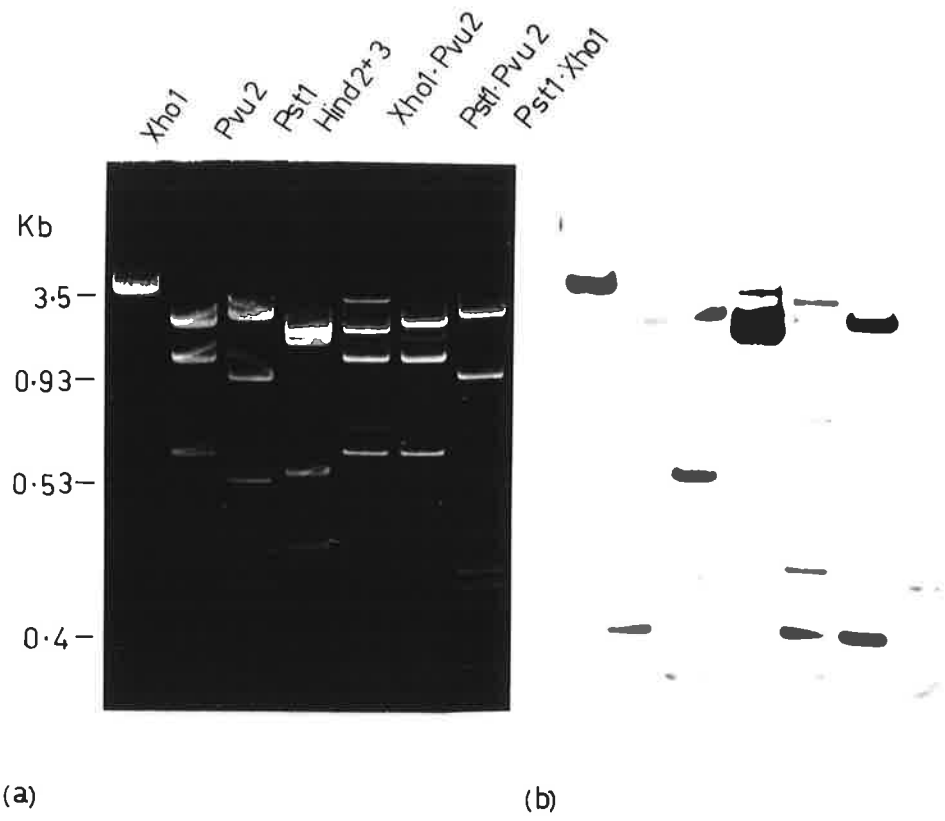
To further characterize the coding regions of the 3.75 Kb coding domain of λ CH-01, this fragment was subcloned into the plasmid vector pBR325 (L. Coles, this laboratory). The resultant clone was termed pCH3.75H/E. Plasmid DNA was digested with HindII+III, PstI, PvuII and XhoI to establish a restriction map. Single and double digests are shown in Figure 5.5(a) after electrophoresis on an agarose gel. Using this data and further data from L. Coles, a restriction map was constructed (Fig. 5.5(c)). To locate the H2A gene within pCH3.75H/E, the gel shown in Figure 5.5(a) was blotted to nitrocellulose and probed with the H2A containing XhoI fragment from pCH3.3E (Fig. 5.5(b)). The extent of the hybridization of this probe is shown on the restriction map in Figure 5.5(c).

The characterization of pCH3.75H/E was extended by P. Krieg and L. Coles in this laboratory. An H4 gene was located within the 1.2 Kb PstI/HindIII fragment of pCH3.75H/E by hybridization of λ CH-01 and pCH3.75H/E DNA with a sea urchin H4 gene probe. An H3 gene probe, also from sea urchin, failed to detect part of an H3 gene located next to the EcoRI site of pCH3.75H/E. The

FIGURE 5.5

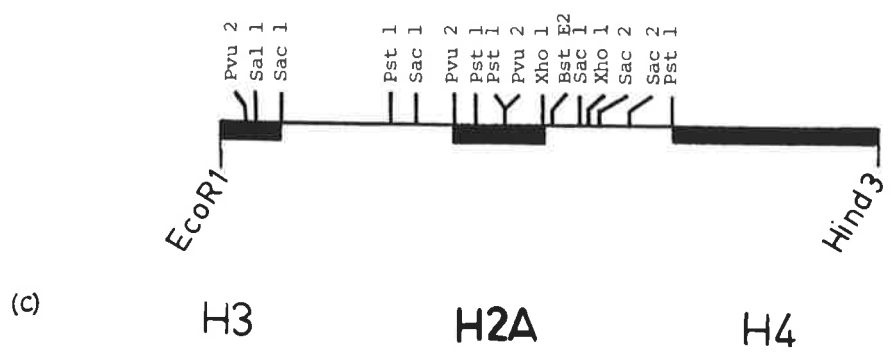
The coding domains of pCH3.75H/E.

- (a) pCH3.75 DNA was digested with HindII, HindIII, PstI, PvuII and XhoI, electrophoresed on a 1% agarose gel and stained with EtBr (section 2.2.2).
- (b) The DNA in the gel shown in (a) was transferred to nitrocellulose (section 2.2.2) and hybridized with H2A probe (the 0.72 XhoI fragment of pCH3.3E, labelled by nick-translation (section 2.2.3)). The filter was washed in 0.5 X SSC and autoradiographed.
- (c) Restriction map of pCH3.75H/E derived from the digestion shown in (a) and the data of L. Cole (this laboratory). The fragments hybridizing to gene specific probes or embryo cDNA are blocked in (H2A, Fig. 5.5(b); H4, data of L. Cole; H3, data of P. Krieg).



(a)

(b)



(c)

presence of this gene was suggested by hybridization of this region to embryo cDNA, and it was identified by DNA sequencing. DNA sequences were also obtained for the H2A and H4 genes in pCH3.75H/E to confirm their identification from hybridizations. The overall gene structure of pCH3.75H/E is shown in Figure 5.5(c).

5.2.5 The Orientation of Genes within λ CH-01

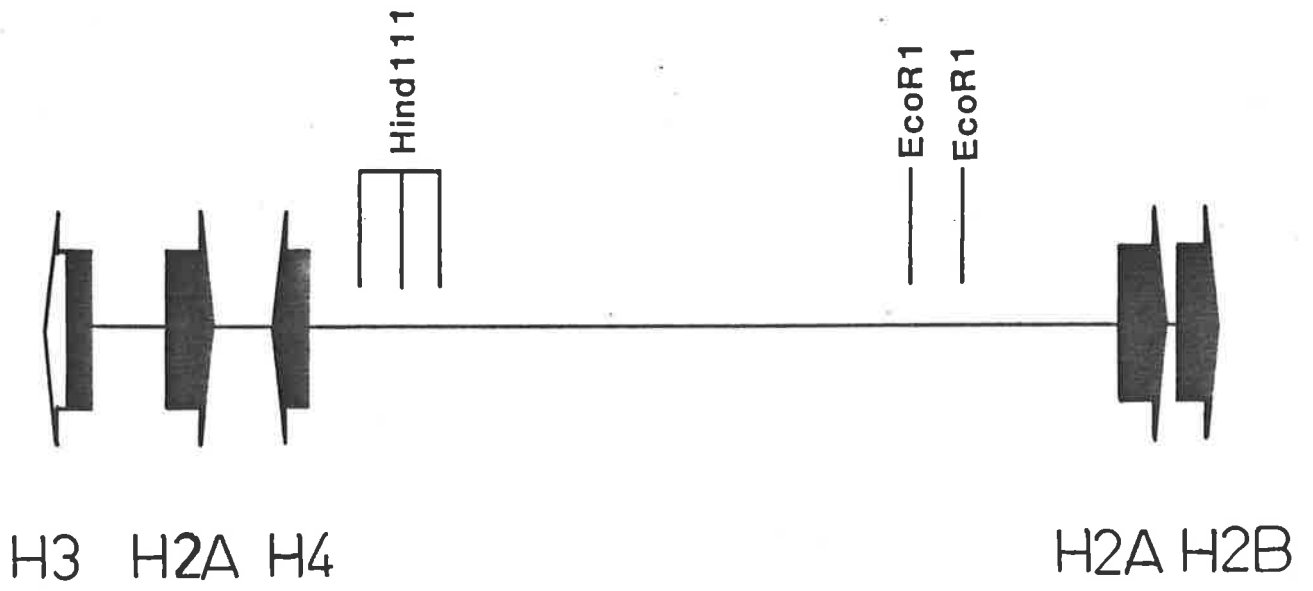
All genes in λ CH-01 were orientated by reference of DNA sequencing data to known restriction endonuclease sites. The overall gene arrangement within λ CH-01 is shown in Figure 5.6.

5.3 DISCUSSION

The impression of λ CH-01 in Figure 5.6 bears no resemblance to the histone gene conformations reported for sea urchins and *Drosophila*, and therefore represents a unique and apparently disordered structure. Before detailed discussion of this organization, another recombinant, λ 1-6, was characterized to strengthen these conclusions (see Chapter 6).

FIGURE 5.6

The arrangement of the genes within λ CH-01. The arrows represent the transcriptional orientation of each gene (5'-3').



CHAPTER 6

THE CODING REGIONS OF λCH-02

(N.B. Some of the data in this
chapter has been published;
Harvey, R.P., Krieg, P.A., Robins, A.J.,
Coles, L.S. and Wells, J.R.E. (1981)
Nature 294, 49-53)

6.1 INTRODUCTION

It seemed necessary to analyse other regions of chicken histone DNA in the event that λ CH-01 was in some way atypical of general organization, and also to see if there might be some conservation of arrangement, despite the apparent disorder. Another clone, λ 1-6, was chosen from those likely to contain histone genes (section 3.2.4). It was subsequently identified as containing histone gene sequences and termed λ CH-02. (It will be referred to as λ CH-02 in this chapter).

6.2 RESULTS

6.2.1 The Coding Domains of λ CH-02

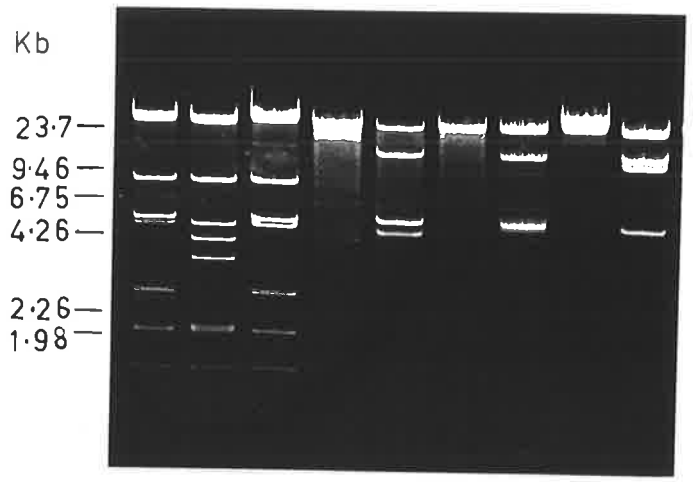
λ CH-02 DNA was prepared (section 2.2.6) and digested with the restriction enzymes EcoRI, HindIII, SalI and SmaI to establish a limited restriction map. These digests, electrophoresed on an agarose gel, are shown in Figure 6.1(a). To define the coding regions of the clone, the gel was blotted to nitrocellulose and probed with embryo cDNA (Fig. 6.1(b)). The limited restriction map deduced from these and other digests is represented in Figure 6.1(c) and the minimal areas hybridizing to embryo cDNA are blocked in. Two separate regions are apparent, one contained within each of the two EcoRI fragments of the λ CH-02 insert. From the limited restriction map, it did not seem likely that λ CH-02 overlapped λ CH-01.

FIGURE 6.1

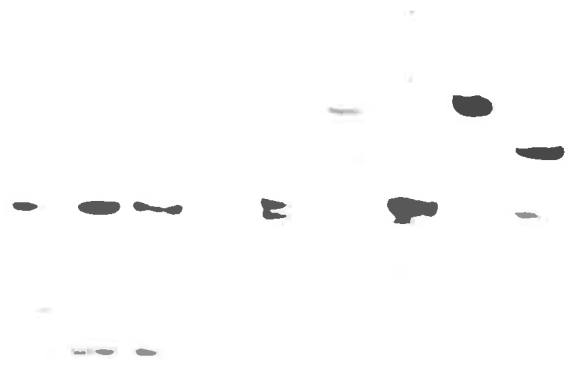
The coding domains of λ CH-02.

- (a) λ CH-02 DNA was digested with the enzymes, EcoRI, HindIII, SmaI and SmaI, electrophoresed on a 1% agarose gel and stained with EtBr (section 2.2.2).
- (b) The DNA in the gel shown in (a) was transferred to nitrocellulose and probed with 5-day embryo cDNA (section 2.2.2). The filter was washed in 0.5XSSC and autoradiographed.
- (c) Limited restriction map of λ CH-02 insert DNA derived from the data shown in (a) and from other digestions. Areas hybridizing to embryo cDNA (b) are blocked in.

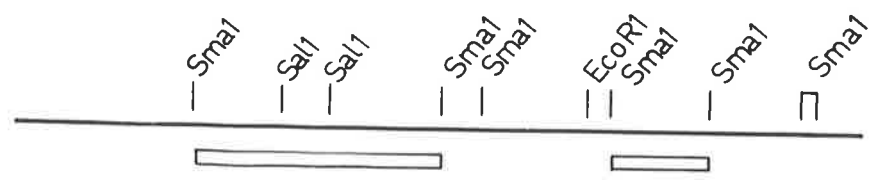
SmaI-Hind3
 SmaI-EcoR1
 SmaI
 SalI-Hind3
 SalI-EcoR1
 SalI
 Hind3-EcoR1
 Hind3
 EcoR1



(a)



(b)



(c)

6.2.2 The Coding Potential of λ CH-02

The coding domains of λ CH-02 could be confined to two SmaI fragments, 4.55 Kb and 1.5 Kb in length (Fig. 6.1). To define which genes were present within these domains, digests of λ CH-02 DNA were hybridized with individual histone gene probes. The H2A and H2B probes were derived from pCH3.3E, being the XhoI and the XhoI/EcoRI fragments respectively (Fig. 5.2). The H3 and H4 probes were fragments derived from the sea urchin histone gene clone h22 (Schaffner *et al.*, 1976) which had been subcloned into pBR322 (kindly supplied by E. Jacob). Probe fragments were resolved from vector or other insert DNA on either low melting point agarose gels with subsequent elution (section 2.2.2) or on sucrose velocity gradients with recovery by ethanol precipitation (section 2.2.1). Radiolabelling was by nick-translation in the presence of 32 P-precursors (section 2.2.3).

λ CH-02 DNA was doubly digested with SmaI and EcoRI. Aliquots were run on an agarose gel together with EcoRI digested λ 55 DNA (section 3.2.2) to serve as hybridization controls. Filters carrying the λ CH-02 and λ 55 digests were individually hybridized to the H3, H4 and H2B probes described.

The H2A probe was hybridized to filters carrying a SmaI digest of λ CH-02 and an EcoRI digest of pCH3.3E as a hybridization control. The results of these experiments are shown in Figure 6.2.

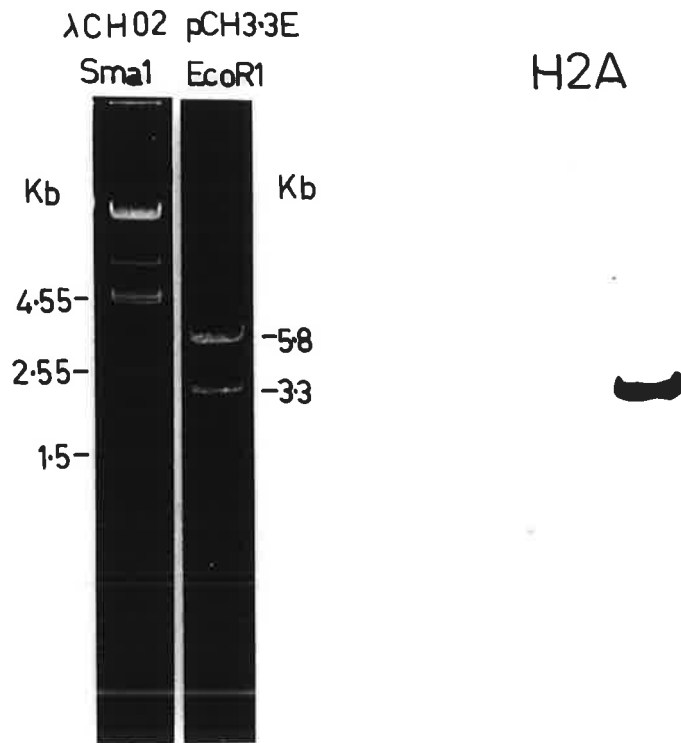
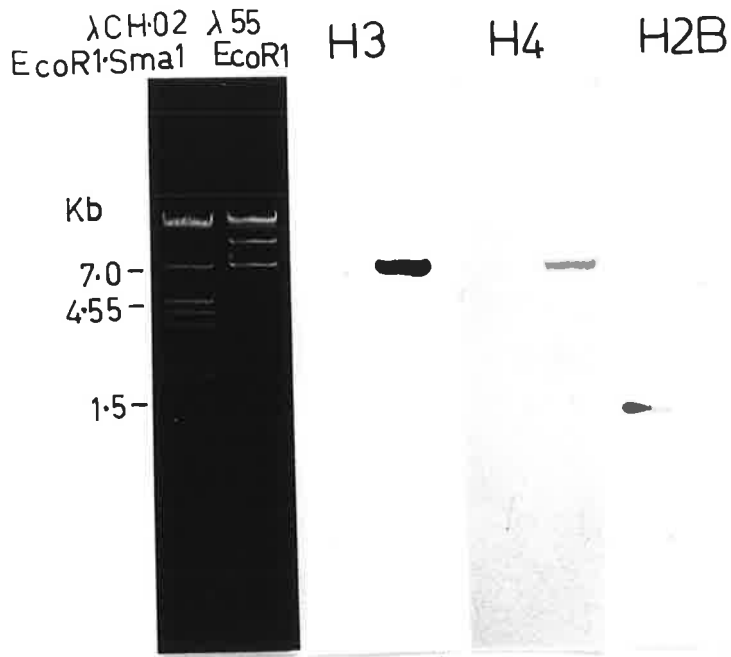
FIGURE 6.2

The coding capacity of λ CH-02.

λ CH-02 DNA was digested with EcoRI/SmaI and SmaI, electrophoresed on 1% agarose gels and transferred to nitrocellulose (section 2.2.2). As hybridization controls, digests of λ 55 DNA (EcoRI; Fig. 3.2) and pCH3.3E DNA (EcoRI; Fig. 5.2) were electrophoresed and transferred. Filters containing λ CH-02 DNA (SmaI/EcoRI) and λ 55 DNA (EcoRI) were hybridized with H3, H4 and H2B probes. Filters containing λ CH-02 DNA (SmaI) and pCH3.3E DNA (EcoRI) were hybridized to the H2A probe.

- PROBES: H3: Fragment of the sea urchin clone *h22* (Schaffner *et al.*, 1976) containing the H3 gene and a fragment of the H2B gene. Hybridization data (not shown) indicated that the sea urchin H2B gene did not hybridize appreciably to chicken H2B genes. Filters washed in 2XSSC.
- H4: Fragment of clone *h22* containing the H4 gene. Filters washed in 2XSSC.
- H2A: The 0.72 Kb XhoI fragment of pCH3.3E (Fig. 5.2). Filters washed in 0.5XSSC.
- H2B: The 0.58 Kb XhoI/EcoRI fragment of pCH3.3E (Fig. 5.2). Filters washed in 0.5XSSC.

Probes were labelled by nick-translation (section 2.2.3).



A positive response was observed with the H3 and H2B probes, H3 being detected in the 4.55 Kb coding domain and H2B in the 1.5 Kb coding domain. The H4 and H2A probes did not detect any fragments in λ CH-02, while their respective hybridization controls were positive.

6.2.3 The H3 Coding Domain of λ CH-02

The 10 Kb EcoRI fragment of λ CH-02 containing the 4.55 Kb SmaI fragment which hybridized to the H3 probe (section 6.2.2) was sub-cloned into the EcoRI site of the plasmid vector pKC-7 (Nagaraja Rao and Rogers, 1979), by A. Robins (this laboratory). This sub-clone was called pCH10E. Plasmid DNA was isolated (section 2.2.4) and digested with restriction enzymes to construct a restriction map. Digestions with the enzymes EcoRI, HindIII, SacII, SalI and SmaI, after electrophoresis on agarose gels are shown in Figure 6.3. The restriction map derived from these and other digestions is shown in Figure 6.5. To locate more accurately the H3 coding region, the gels shown in Figure 6.3 were blotted to nitrocellulose and probed with the sea urchin H3 probe (section 6.2.2). The autoradiograph of this hybridization is shown in Figure 6.4. Two distinct regions of homology to the H3 probe were observed, as indicated on the restriction map in Figure 6.5. These regions were separated by a 0.9 Kb SalI fragment which hybridized poorly to the probe. This suggested the presence of an H3 gene

FIGURE 6.3

The H3-coding region of λ CH-02: the mapping of pCH10E.

pCH10E DNA was digested with EcoRI, HindIII, SacII, Sali and SmaI, electrophoresed on 1% agarose gels and stained with EtBr. M represents size marker tracks containing EcoRI/HindIII cut λ DNA and HaeIII cut pBR322 DNA.

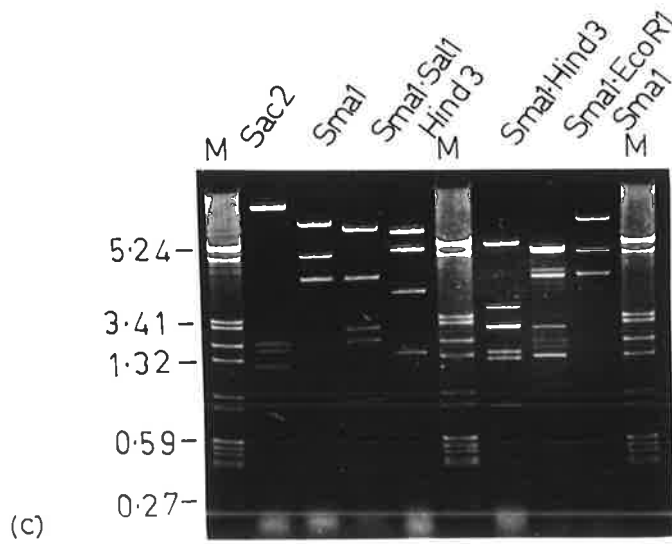
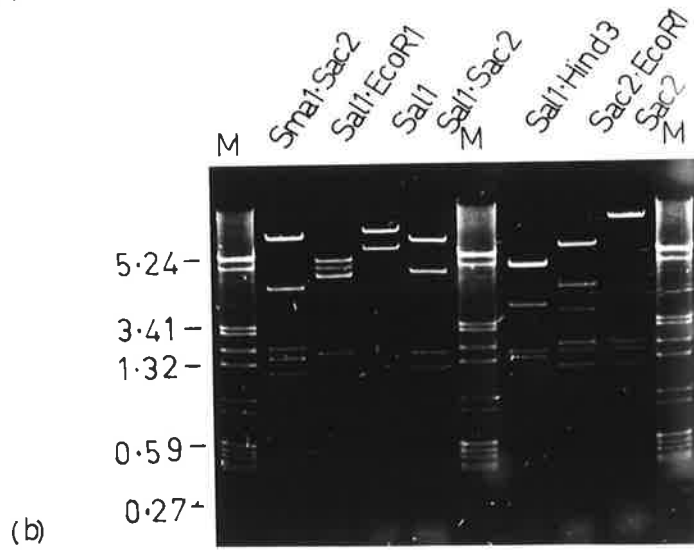
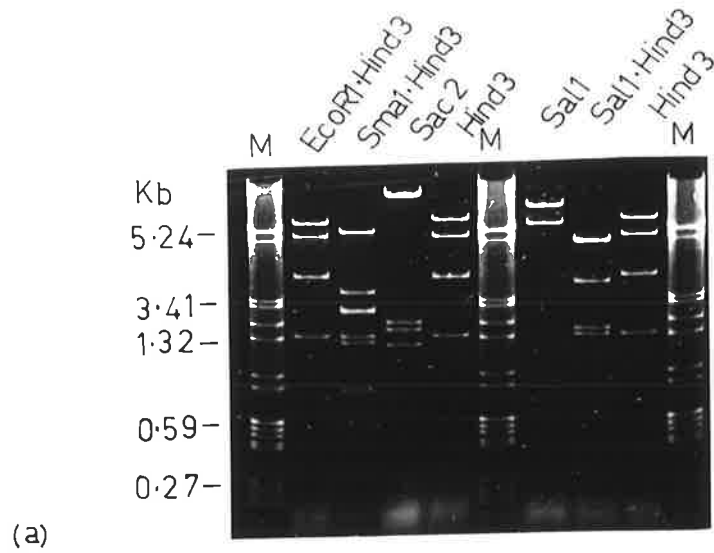
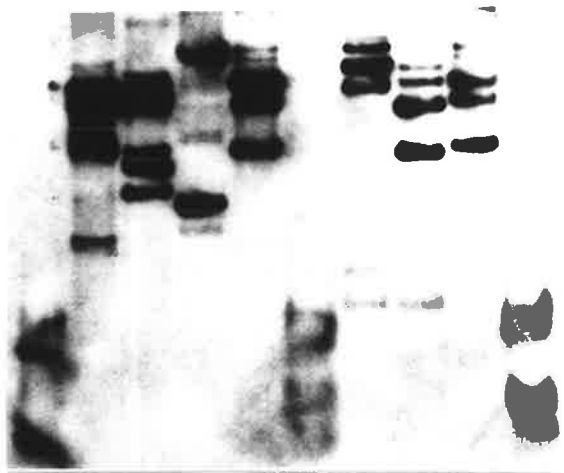


FIGURE 6.4

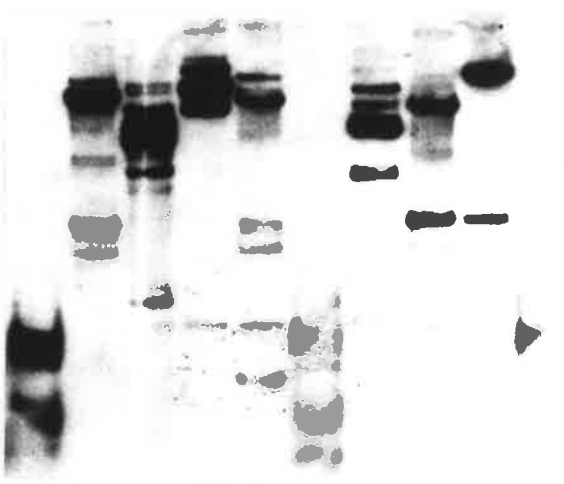
The H3-coding domain of λ CH-02: hybridization of pCH10E to H3 probe.

The gels shown in Fig. 6.3 were transferred to nitrocellulose (section 2.2.2) and hybridized to sea urchin H3 probe (see figure legend to Fig. 6.2).

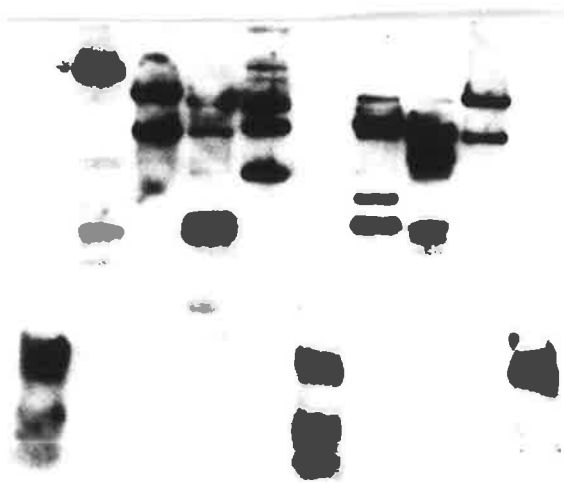
Filters were washed in 2XSSC and autoradiographed. Due to contaminating pBR322 in the H3 probe, some hybridization is seen to the vector bands of pCH10E (pKC-7, which is pBR322 derived). Interpretation of these results was from hybridization to band known not to contain vector DNA.



(a)



(b)



(c)

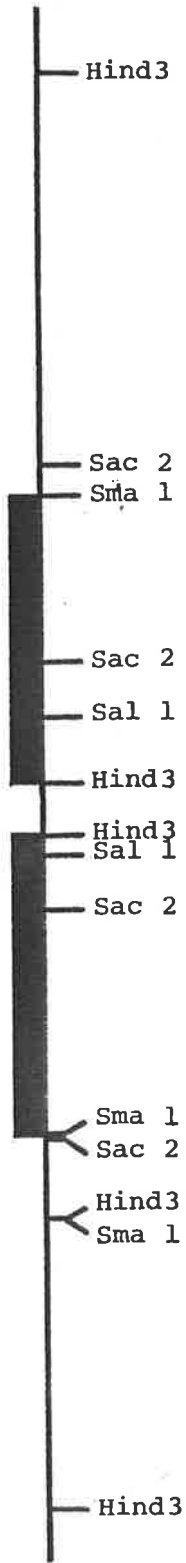
FIGURE 6.5

The H3-coding domain of λ CH-02: restriction map of pCH10E.

The restriction map of pCH10E derived from the digests shown in Fig. 6.3. Minimal areas hybridizing to the sea urchin H3 probe (Fig. 6.4) are blocked in.

H3

H3



interrupted by non-coding DNA. However, some elements of symmetry were noticed in the restriction map of pCH10E, with the 0.9 Kb SalI fragment centrally located. In fact, the two opposed SalI/SacII fragments flanking the 0.9 Kb SalI fragment ran as a doublet on 1% agarose gels (Fig. 6.3). This rather suggested a duplication of an H3 gene in opposed orientation.

To distinguish between these two possibilities and to confirm the assignment of the H3 gene to this region, the 0.9 Kb SalI fragment was isolated from polyacrylamide gels (section 2.2.7) for DNA sequencing (see Chapter 8 for a more detailed description of sequencing procedures). The eluted DNA was dephosphorylated with calf intestinal phosphatase and 5'-labelled with ^{32}P using polynucleotide kinase and ^{32}P ATP as precursor (section 2.2.7). The fragment was further cut with HindIII and the sub-fragments were resolved in a polyacrylamide gel. After autoradiography, labelled fragments were excised, eluted and sequenced using the Maxam and Gilbert procedures (section 2.2.7). The resultant sequences, effectively read in from each SalI terminus, are shown in Figure 6.6. Both nucleic acid sequences, in one reading frame, predict the protein sequence of the first 10 amino acids of the chicken H3 protein (Brandt and Von Holt, 1974), and include the initiation AUG codon, and a portion of the 5'-untranslated region. Within the protein coding region both nucleic acid sequences are identical, but 5' to the

FIGURE 6.6

H3 DNA sequences from pCH10E.

The 0.9 Kb SalI fragment of pCH10E was isolated and sequences determined in from the SalI termini by the Maxam and Gilbert method (section 6.2.3).

Sequences from each SalI terminus are shown aligned to display nucleic acid homologies (*).

A restriction map of pCH10E (Fig. 6.5) is shown, and the transcriptional orientation of the H3 genes indicated (large arrows). The small arrows represent the direction of the DNA sequencing.

initiation codon, the two sequences diverge, confirming the presence of two distinct H3 genes. Because the 5'-ends of these genes were encountered in proximity within the 0.9 Kb SmaI fragment, the orientation of the two H3 genes must be divergent (Fig. 6.6).

6.2.4 The H2B Coding Domain of λ CH-02

To allow further analysis of the 1.5 Kb SmaI fragment of λ CH-02 which hybridized to embryo cDNA (section 6.2.1) and to the H2B probe (section 6.2.2), the 4.6 Kb EcoRI fragment containing this region was sub-cloned into the plasmid vector pBR325. 4.6 Kb EcoRI DNA was isolated from an EcoRI digest of λ CH-02 DNA electrophoresed on a low melting temperature agarose gel. After elution (section 2.2.2), 4.6 Kb DNA was ligated into EcoRI-cut and dephosphorylated pBR325 in proportions according to the mathematical treatment of Dugaiczky *et al.*, (1975), to maximize insert formation. Ligated DNA was transformed into competent *E. coli* K12 (section 2.2.4) and transformants selected on tetracycline plates. Crude plasmid DNA was isolated from mini-cultures of 8 chloramphenicol sensitive transformants and aliquots were cut with EcoRI and electrophoresed on an agarose gel (Fig. 6.7(a)). Seven of the eight plasmids showed EcoRI released insertions of 4.6 Kb in length. Further aliquots of plasmid from four of these colonies were linearized with HindIII and electrophoresed on an agarose gel (Fig. 6.7(b)). The sizes of the linear

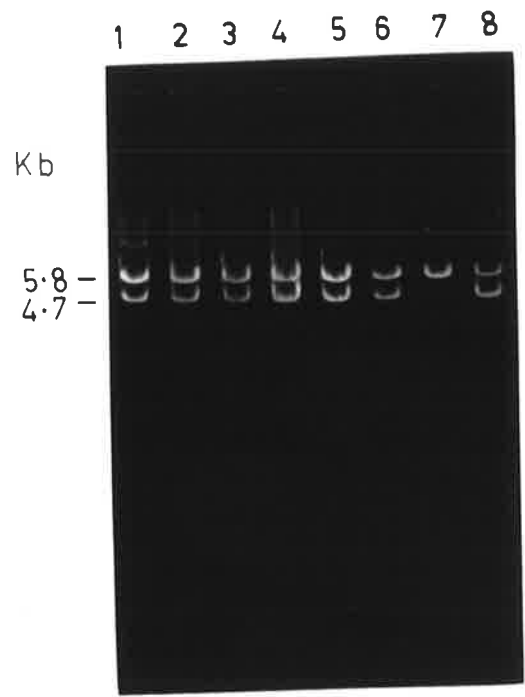
FIGURE 6.7

Subcloning of the H2B-coding domain of λ CH-02.

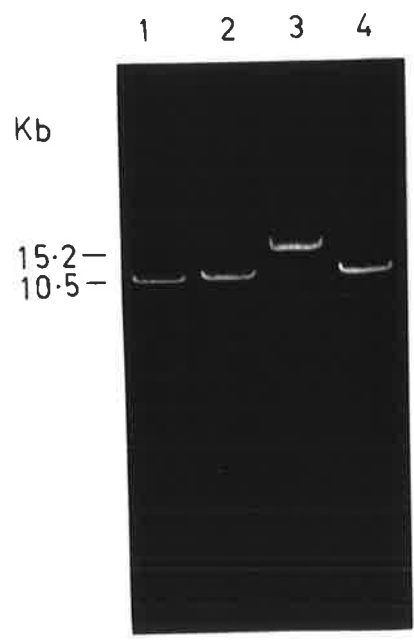
The 4.7 Kb EcoRI fragment of λ CH-02 was subcloned into the vector pBR325 (section 6.2.4).

- (a) Mini-plasmid preparations from 8 transformants displayed on a 1% agarose gel. Bands were detected by EtBr staining. Plasmid DNA was digested with EcoRI to release 4.7 Kb insert.

- (b) Mini-plasmid preparations from transformants 1-4(a), linearized with HindIII and electrophoresed on a 1% agarose gel. DNA was detected by EtBr staining.



(a)



(b)

plasmid forms indicated that three contained single 4.6 Kb insertions and one contained a double insert. One sub-clone containing a single 4.6 Kb insert was selected and plasmid DNA isolated in bulk for further study. This plasmid was called pCH4.6E.

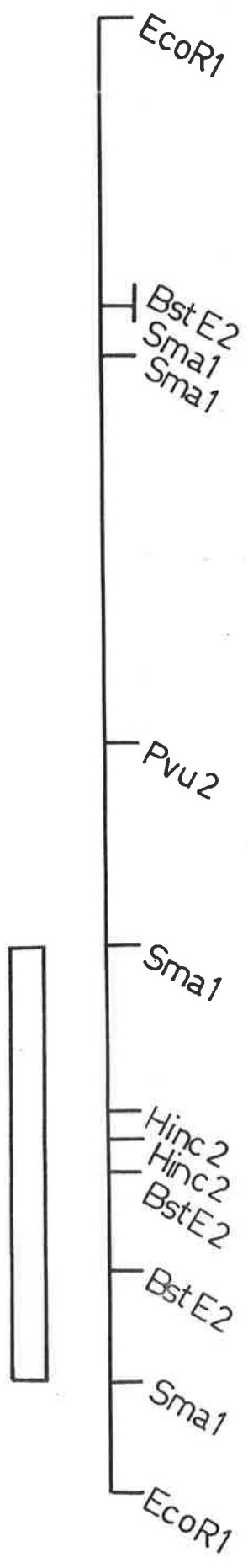
A limited restriction map was established for the pCH4.6E insert using the enzymes BstEII, HincII, PvuII and SmaI, and is shown in Figure 6.8. The 1.5 Kb SmaI fragment which hybridized to embryo cDNA (section 6.2.1) and the H2B probe (section 6.2.2) is blocked in. A 300 bp BstEII fragment is contained within this region. The presence of a 300 bp BstEII fragment recognized in the DNA sequence of the H2B gene from pCH3.3E (see Chapter 8) suggested that the H2B gene in pCH4.6E might also lie across the BstEII sites. DNA sequencing was undertaken to identify and locate the H2B gene, to orientate it relative to known restriction endonuclease sites, and to check for the presence of multiple H2B genes or other coding sequences within this domain.

A SmaI/BstEII digest of pCH4.6E was run on a preparative polyacrylamide gel and the 0.72 Kb BstEII/SmaI fragment excised and eluted (section 2.2.7). This fragment was treated with exonuclease III to produce overhanging 5'-termini (see section 8.2.2), then dephosphorylated with calf intestinal phosphatase, and ³²P end-labelled with polynucleotide kinase. Aliquots of the labelled fragment were digested with various restriction enzymes to find a digest which produced cleavage

FIGURE 6.8

Restriction map of pCH4.6E (section 6.2.4).

The region hybridizing to embryo cDNA
(section 6.2.1) and the H2B probe (section
6.2.2) is blocked in.



products of a suitable size for sequencing and which would resolve on an acrylamide gel. PstI was found to give suitable fragments. The bulk of the kinased fragment was digested with PstI and the sub-fragments resolved on a polyacrylamide gel. After elution, these sub-fragments were sequenced by the Maxam and Gilbert procedures (section 2.2.7). The sequences read from the BstEII end of the fragment are shown in Figure 6.9. They predict amino-acids with direct homology to the N-terminal seven amino-acids of the bovine H2B protein (Iwai *et al.*, 1972), and include an additional 104 bps of the H2B 3'-untranslated and flanking sequences (see Chapter 8). This identified the H2B gene and orientated it relative to the restriction endonuclease sites shown on the map of pCH4.6E in Figure 6.9.

DNA sequences of 230 bps were also read from the SmaI end of the sequenced fragment. No other histone gene segments could be recognized in these sequences. Because of this, and the spacial requirements for a further histone gene to be present in the 1.5 Kb SmaI coding domain, it was concluded that the H2B gene was the only histone gene present in this region.

6.2.5 The Arrangement of Genes in λ CH-02

The arrangement and orientation of the 10 Kb and the 4.6 Kb EcoRI fragments of λ CH-02 can be deduced from the SmaI maps of the individual sub-clones of these fragments (figures 6.5 and 6.7) and from the SmaI and SmaI/EcoRI

FIGURE 6.9

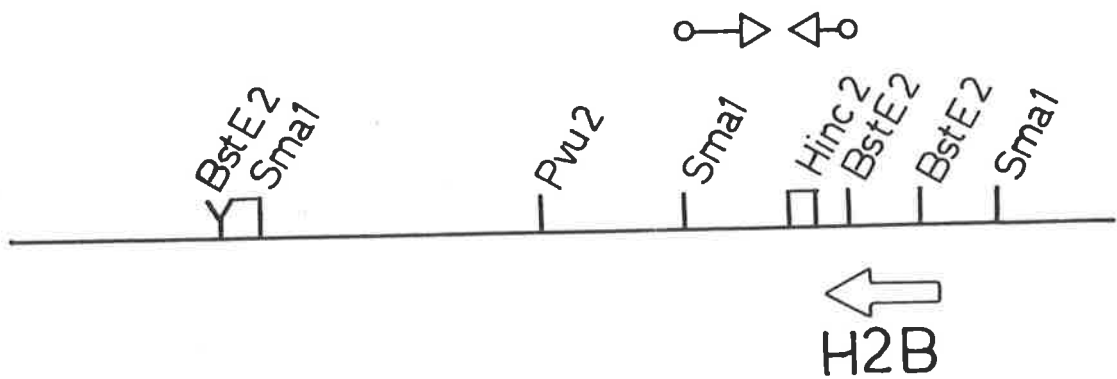
H2B DNA sequences from pCH4.6E.

The 0.72 Kb BstEII/SmaI fragment of pCH4.6E was isolated and DNA sequences were determined from both termini using Maxam and Gilbert procedures (section 6.2.4).

The sequences presented are from the BstEII terminus. A restriction map of pCH4.6E is shown. The large arrow represents the position and transcriptional orientation of the H2B gene in pCH4.6E. The small arrows represent the DNA sequences determined.

BstE2
 5' ↓ GTC ACC AAG TAC ACC AGC TCC AAG TAG
 VAL THR LYS TYR THR SER SER LYS STOP
 BOVINE VAL THR LYS TYR THR SER SER LYS
 H2B

AGCGGTGCGGATTACTCGATTTTAACCCAAAGGCT
 CTTTTCAGAGCCACCATTTGTTCTAATAAAAGGGC
 TGTATTACTTTTTTCTTTTTTTCCTGAGGGGTAT
 AGCGTGGGTAACTGAGTGAAATGGAAGCGAGTGC
 CTGAGGTATGTATATAATTGCTTAACTTCGCAGTT
 CGGAGGTCTCCGTTCCGAGTTAATTGAGCAGTAGC



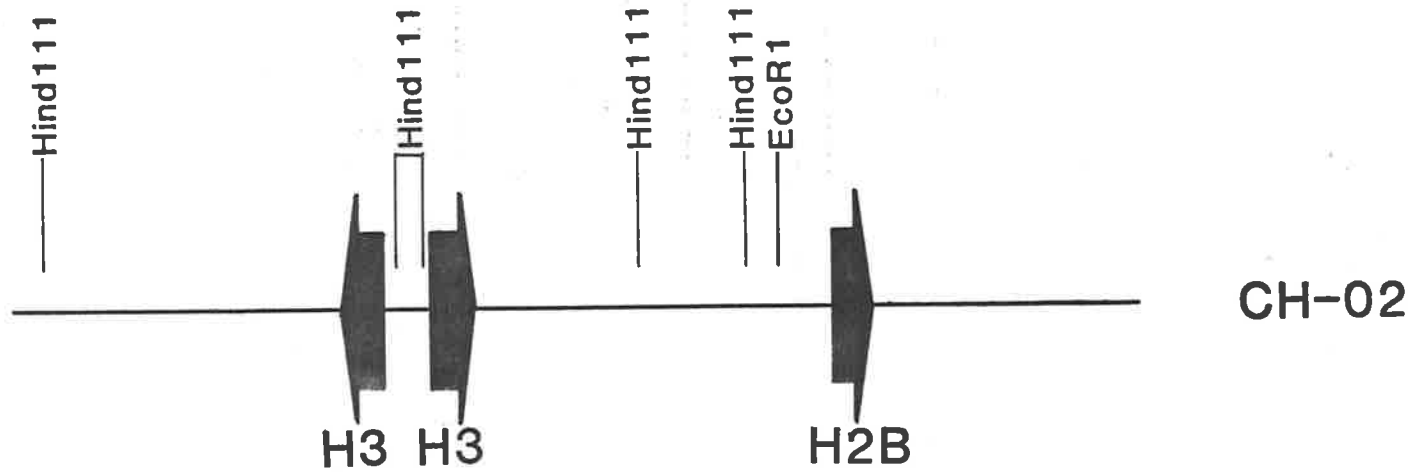
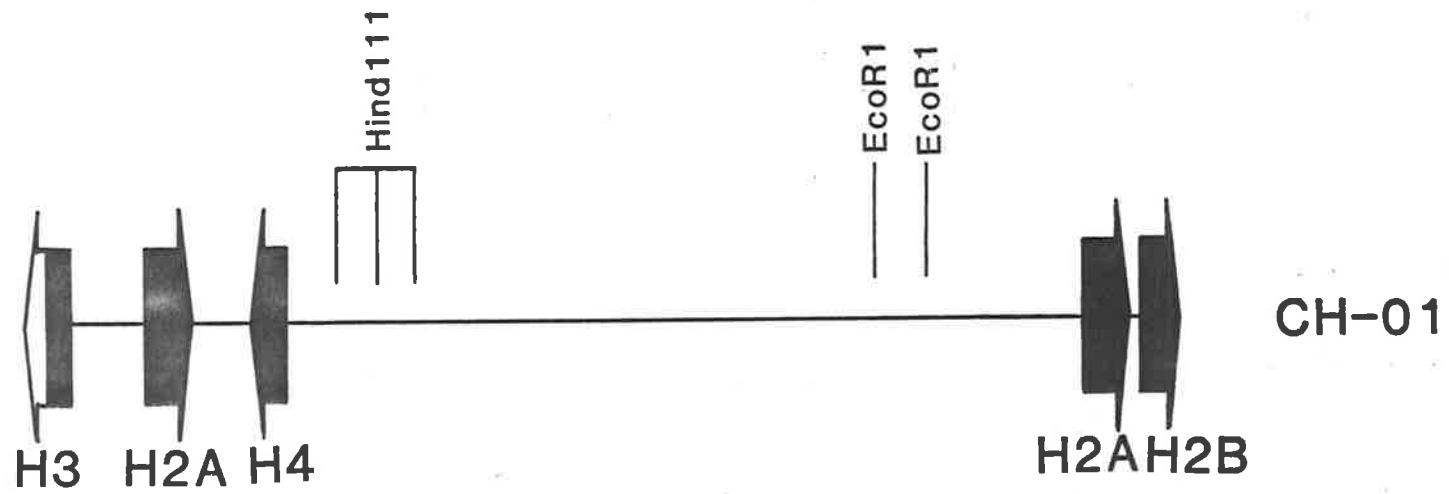
digests of λ CH-02 DNA (Figure 6.1). The overall organization of the genes within λ CH-02 is shown in Figure 6.10 together with (for convenient reference) the gene organization of λ CH-01.

6.3 Discussion

The organization of the histone genes in λ CH-01 and λ CH-02 (Figure 6.10) exhibit a number of unique features with respect to those well-characterized invertebrate histone gene systems (Kedes, 1979). The genes are certainly clustered into mixed groups at particular sub-loci, however large stretches of spacer DNA appears to separate these local clusters. For example, approximately 9 Kb of DNA separates the closely linked H2A/H2B pair from the H3/H2A/H4 group in λ CH-01. (The maximum size known for a complete histone gene cluster in the invertebrates is 9.5 Kb, for the sea urchin *M. tripneustes* (Kedes, 1979)). In addition, these somewhat isolated groups of genes do not resemble each other in gene content, number or arrangement. In fact, within the two inserts of these non-overlapping clones, no evidence is found for the existence of a repeating unit containing each of the histone genes. In λ CH-01, an H2A gene occurs twice before a full complement of genes is represented and these are separated only by an H4 gene. This represents a marked disorder in the arrangement of chicken histone genes. It might be tempting to postulate that these sub-groups of genes are clustered as such to

FIGURE 6.10

The arrangement of histone genes in λ CH-01 and λ CH-02. The arrows represent the transcriptional orientation of each gene (5'-3').



1KB



allow the co-ordinate expression of each group at different regulated stages of development. However, this may be more fruitfully discussed when more is known of the coding potential of these genes and the evolutionary origins of the observed disorder (see Chapter 9).

It is unlikely that a repeat unit larger than the insert size of these clones exists because where genes are represented twice, in no case do they occur in the same environment. H2A, H2B and H3 genes are duplicated within the two inserts, and no relationship can be seen between their respective macro-environments.

The orientation of the genes indicates that transcription occurs from both DNA strands. This is in agreement with a more general analysis of λ CH-01 and λ CH-02 by Bruschi and Wells (1981). Such an arrangement would seem to preclude co-ordinate expression of core histones being effected by a single polycistronic transcript as has been suggested by Melli *et al.*, (1977) for human histone genes. The results are more in agreement with the u.v. irradiation studies of P.B. Hackett *et al.*, (1978) who found that Hela cell histone genes are independently transcribed. However, this area is not entirely clear as newt lampbrush histone DNA appears to be transcribed into long multicistronic transcripts, despite the fact that within the newt histone gene cluster, genes occur in both orientations (Stephenson *et al.*, 1981).

One particularly notable feature is the divergently transcribed H3 gene pair in λ CH-02. The apparent restriction site symmetry about this region (Figure 6.5) might suggest that this arrangement originated by a relatively recent duplication of a single H3 gene. No other direct repetition of a histone gene has been observed in those systems studied.

Divergent arrangement of non-identical genes has been observed in other histone gene systems (Wallis *et al.*, 1980; Lifton *et al.*, 1977; Stephenson *et al.*, 1981) and in the chorion and heat shock genes of the silk moth (Jones and Kafatos, 1980). Although the functional implications of this arrangement are speculative at present, it has been suggested that such gene pairs may share controlling elements of DNA sequence to ensure co-ordinate expression (Lifton *et al.*, 1977). In support of this, all paired silk moth chorion genes belong to a specific developmental period despite the fact that developmentally distinct gene pairs are interdigitated along the DNA (Jones and Kafatos, 1980). Further DNA sequencing and surrogate genetics will be useful in providing some answers to these questions.

CHAPTER 7

SOUTHERN ANALYSIS OF CHICKEN DNA



7.1 INTRODUCTION

Southern blot analysis of genomic DNA has proved powerful in probing the structure of eukaryote genes. Although more readily applied to unique genes or highly reiterated, conserved genes, an analysis of chicken DNA using histone probes may give an indication of the overall arrangement of histone genes in the chicken genome.

7.2 RESULTS

Chicken DNA was digested with EcoRI and samples electrophoresed on agarose gels. After transfer to nitrocellulose (section 2.2.2), digests were hybridized with probes derived from λ CH-01 and λ CH-02, specific for the core histone genes. The H2A and H2B probes were the XhoI and XhoI/EcoRI fragments respectively, derived from pCH3.3E (Fig. 5.2). Both of these probes contained the whole protein coding portion of each gene, together with some untranslated and flanking sequences (see Chapter 8 for the complete sequences of both fragments). The H3 probe was the Sali/SacII fragments from the pCH10E H3 genes (Fig. 6.5). These fragments have not been sequenced. The H4 probe was an HinfI fragment from pCH3.75H/E (section 5.2.4) containing the complete protein coding portion of the H4 gene together with 25 bps of untranslated regions. This gene has been completely sequenced by P. Krieg (this laboratory). Probes were labelled by nick-translation (section 2.2.3). Filters were washed under stringent conditions (0.1 X SSC).

Figure 7.1 shows the results of these hybridizations.

FIGURE 7.1

Southern hybridization analysis of chicken DNA with gene-specific probes. Chicken genomic DNA was digested with EcoRI and electrophoresed on 0.8% agarose gels. The DNA was transferred to nitrocellulose (section 2.2.2) and filters were hybridized separately to the core histone probes.

H2A: the 0.72 Kb XhoI fragment of pCH3.3E (Fig. 5.2)

H2B: the 0.58 Kb XhoI/EcoRI fragment of pCH3.3E (Fig. 5.2)

H3: the SalI/SacII fragments of pCH10E (Fig. 6.5)

H4: an HinfI fragment from pCH3.75H/E (section 5.2.4) containing the complete H4 gene (supplied by P. Krieg).

Probes were labelled by nick-translation (section 2.2.3) and filters washed in 0.1 X SSC.

H2A H2B H4 H3

Kb

13.5 —

11.0 —

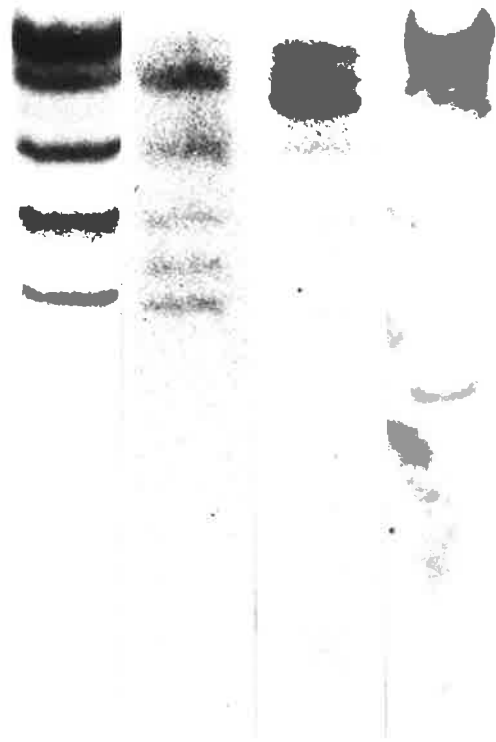
7.1 —

4.8 —

3.8 —

3.3 —

2.0 —



The patterns obtained do not differ significantly from the hybridizations using more extensive probes described in section 4.2.3 and as expected, fail to indicate the presence of a tandemly reiterated and conserved histone gene cluster that has been observed for other histone gene systems (Kedes, 1979). This further strengthens the argument that chicken histone genes are arranged in the genome in a distinctly disordered fashion.

7.3 DISCUSSION

Although the patterns observed in Figure 7.1 do not indicate a simple repeating structure for chicken histone genes, they are obviously related, and in fact relatively simple. This probably reflects the low copy number of chicken histone genes (approximately 10, Crawford *et al.*, 1979). Using the H2A pattern as an example, five distinct bands can be seen of approximate relative intensity 1 : 1 : 2 : 2 : 4. If the bands of faintest intensity are from single H2A genes, then this pattern would account for 10 genes. Controlled gene number blots would, however be difficult due to partial cross reaction between copies of variant genes.

A possible anomaly in these blots which is noteworthy, is the nature of the H3 pattern. Within λ CH-01 and λ CH-02 there are H3 genes on two EcoRI fragments. Because the chicken genomal library was constructed using EcoRI linkers (section 3.2.3), the EcoRI genomic fragments on which these H3 genes reside would be greater than 3.75 Kb and 10 Kb (Figs. 5.5 and 6.5). The H3 blot patterns are not consistent

with this unless the uppermost band represents a doublet. There are two other possible explanations. One is that there is some sequence heterogeneity in the histone DNA of chickens, the DNA used in these experiments showing a different H3 pattern to the DNA from which the library was constructed. Restriction site heterogeneity between individuals has been observed in other gene systems. The other possibility is that the H3 genes of pCH10E (Fig. 6.5) are variant, and do not cross react well with the pCH3.75H/E H3 gene (Fig. 5.5). The H3 probe in these blotting experiments is the only one which has not been completely sequenced.

One conclusion which can be drawn from this data is that, in general, the genes coding for the different histone types are intermingled on the DNA, as is the case in λ CH-01 and λ CH-02. This is suggested by the detection of similarly sized bands with most probes. The intensity of the band specific to the H4 track however, indicates that there may be a cluster of H4 genes which is not associated with the other histone genes. These are in addition to these H4 genes which are intermingled with other genes. The significance of this in the light of the observed "randomness" of chicken histone gene order, is unknown.

The H2B and H3 tracks also contain specific bands, but these are not predominant in intensity.

It is likely that most of the chicken histone genes exist at one locus (albeit disperse) on the DNA. This is suggested by the fact that, except for the H4 specific band,

the longer EcoRI fragments are darkest in intensity, and a gradient of intensity exists from shortest to longest fragments. This indicates that the largest fragments contain progressively more copies of each particular gene. If this is indeed the case, the size of the chicken histone cluster may be estimated to be about 65 Kb. Based on the type of organization seen in λ CH-01 and λ CH-02, this would be enough DNA to code for 40-50 histone genes. Again, this is in good agreement with the estimation of chicken histone gene number (approximately 10 copies each) by Crawford *et al.*, (1979).

CHAPTER 8

THE MICROSTRUCTURE OF CHICKEN HISTONE GENES

(N.B. Some data presented in
this chapter has been published:
D'Andrea, R., Harvey, R.P., and
Wells, J.R.E. (1981). Nuc. Acids
Res. 9, 3119-3128)

8.1 INTRODUCTION

Analysis of DNA sequences of genes and their surrounding regions may indicate how genes are regulated. Comparative DNA sequence data within and between species has allowed development of notions of gene evolution.

The relative disorder of chicken histone genes has been described (Chapters 6 and 7). A logical progression was to investigate the microstructure of the individual genes in order to answer further structural questions, and to search for possible links between organization, evolution and function of the histone genes. This chapter contains a description of sequencing techniques, and strategies for the determination of the DNA sequences of the H2B gene in pCH3.3E (section 5.2.2) and the H2B gene in pCH4.6E (section 6.2.4). The 5'-leader region of the H2A gene of pCH3.3E was also determined for appropriate comparisons.

The sequence of the H2B gene of pCH3.3E (section 5.2.2) was determined using the chemical degradation techniques of Maxam and Gilbert (1979). The H2B gene of pCH4.6E (section 6.2.4) was sequenced in collaboration with A. Robins (this laboratory), using the chain termination method of Sanger *et al.*, (1977).

8.2 RESULTS

8.2.1 Sequencing Procedures Using the Chemical Degradation Method

The sequencing procedure of Maxam and Gilbert (1977)

relies on limited, but base specific, chemical cleavage of aDNA fragment with a single ^{32}P -radiolabelled terminus. A series of separate cleavage reactions can be performed, each specific for one or two of the DNA bases. The reactions do not proceed to completion, and so a pool of partially cleaved products is generated. When run on an appropriate high resolution poly-acrylamide gel, only those partial products containing the labelled terminus will be detected by autoradiography. In effect, all of these fragments detected will have one common (labelled) terminus, so a ladder of fragments is observed representing the distance from the labelled terminus, of each successive base which is the target of a specific cleavage reaction. When cleavage reactions specific for different bases are run on adjacent gel tracks, the DNA sequence of the fragment can be read from the ladders.

The chemistry of the cleavage reactions proceeds in three steps: modification of a DNA base, removal of the modified base from its sugar and DNA scission at that sugar. The specificity of the technique is ensured by each of the above steps being contingent upon the previous step. In this way a partial modification of specific bases will result in a pool of partial cleavage products. Modification reagents are either specific for purine (the G+A reaction), or pyrimidine (the C+T reaction) rings, or specific for particular bases (the G and C reactions). A combination of this information

is sufficient to determine the DNA sequence of a fragment.

Considerable variation is possible in the selection of a sequencing strategy and in the choice of labelling procedures (Maxam and Gilbert, 1980). A description of the methods found most suitable and convenient to this work will be given here. The choice of strategy often depends on economics and the quality of reagents.

In developing the technique, considerable problems were encountered in the ability to restrict DNA after passage through polyacrylamide gels, and in the quality and specificity of the sequencing reactions. These problems were alleviated to some degree by the use of acrylamide stocks which had been filtered and exhaustively deionized using mixed bed resins, although failure of restriction enzymes to digest eluted DNA still persisted. This was less apparent using restriction enzymes whose buffers contained high concentrations of mono-valent cations. Because of this problem, the strategies used had to maximise restriction enzyme handling before passage through acrylamide gels. In general, plasmid DNA was restricted with the appropriate enzymes and passaged through a polyacrylamide gel. Fragments to be sequenced were eluted, end-labelled and cut with a second restriction enzyme to separate the two labelled termini. A further polyacrylamide gel separated labelled fragments before sequencing.

Restriction endonuclease digested plasmid DNA (50-100µg) was dephosphorylated using calf intestinal or

bacterial alkaline phosphatase (section 2.2.7) prior to electrophoresis on a polyacrylamide gel. This was adopted to assist in the removal of the phosphatase before the kinasing reaction. The phosphatases are not totally inactivated by heating and efficiently removed by phenol extraction. Persistence of phosphatase into subsequent steps resulted in loss of the radioactivity incorporated by polynucleotide kinase.

Restricted and dephosphorylated DNA was passaged through thin (0.6 mm) preparative polyacrylamide slab gels, polymerized from deionized acrylamide stocks. Thin gels were important in reducing acrylamide contaminants which accumulated during elution of fragments from the gel, and subsequently caused inhibition of restriction enzymes. An 8 cm well on a 0.5 mm gel could accommodate a 100 μ g digest of plasmid DNA with adequate resolution of smaller fragments.

DNA fragments in the gel were detected by staining with ethidium bromide and bands were excised under UV illumination. To elute the DNA, the gel slice was placed in 1 ml of elution buffer (section 2.2.7) in an Eppendorf tube and incubated at 37°C for a few hours. Using thin preparative gels it was not found necessary to crush the gel slice nor use the long elution times as described by Maxam and Gilbert (1980). This was also found to decrease acrylamide contaminants. Routinely, 90% of DNA could be recovered using this technique, although this percentage rapidly declined when eluting

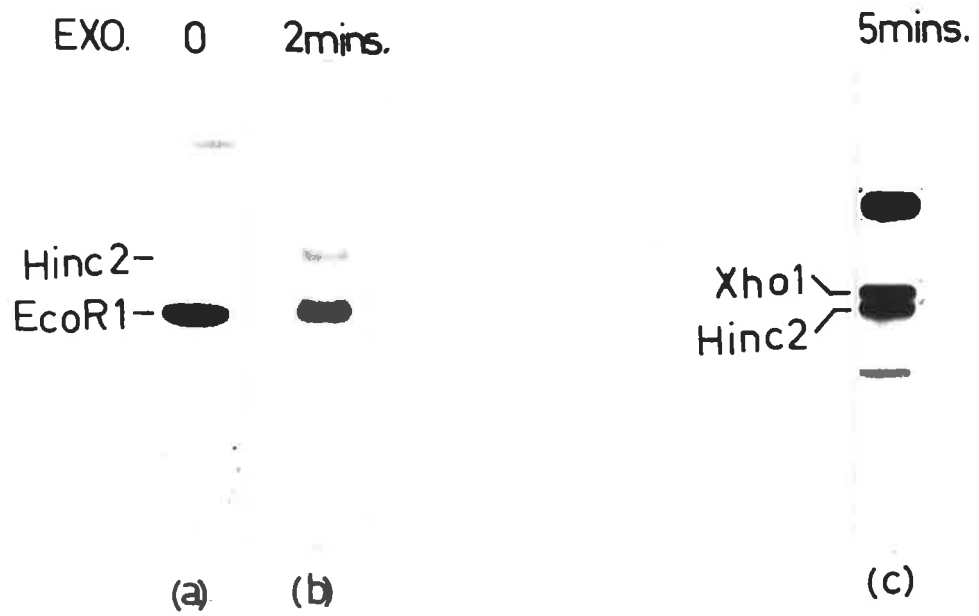
fragments over 800 bps in length. After elution, the buffer was clarified by phenol extraction and DNA recovered by ethanol precipitation.

The addition of ^{32}P -phosphate to the 5' termini of DNA fragments using polynucleotide kinase has been used to create the labelled terminus necessary for DNA sequencing. Combined with the dephosphorylation procedures described, the forward reaction (section 2.2.7) has been used for both 5' protruding and flush ended fragments in preference to the exchange reaction (Maxam and Gilbert, 1980). Using this procedure it was found that 5' protruding ends kinased well, but flush and 3' protruding ends kinased very poorly. It was often convenient to sequence from flush ends. To improve the labelling efficiency of these ends, a method was devised to convert flush ends to 5' protruding ends prior to labelling. Fragments were treated with exonuclease III for 5 minutes on ice (section 2.2.7) generally before dephosphorylation. This enzyme is a 3' progressive exonuclease on double stranded DNA. When the digestion is performed at 4°C, only a limited number of nucleotides (6-12) are digested (Donelson and Wu, 1972), and so can be used to convert flush ends to 5' protruding ends without serious degradation of the DNA. Figure 8.1 shows the secondary cleavage products of fragments containing one 5' protruding end and one flush end, with and without treatment with exonuclease III prior to kinsing.

FIGURE 8.1

ExonucleaseIII treatment of blunt ended restriction fragments to create 5' over-hanging termini (section 6.2.1). An EcoRI/HincII fragment was treated with exonucleaseIII (section 2.2.7) and then end-labelled with polynucleotide kinase and $\alpha^{32}\text{P}$ ATP. To separate the labelled ends the fragment was further digested with HaeIII ((a) and (b)) or MspI ((c)) and subfragments were resolved on a 6% polyacrylamide gel (section 2.2.2).

- (a) The EcoRI/HindIII fragment kinased without prior exonucleaseIII treatment.
- (b) Treatment with exonucleaseIII for 2 mins. on ice prior to kinasing.
- (c) Treatment with exonucleaseIII for 5 mins. on ice prior to kinasing.



After kinasing, the two labelled termini were separated by either strand separation on low cross-linked polyacrylamide gels (section 2.2.7), or by secondary cleavage of the fragment with a restriction enzyme (Maxam and Gilbert, 1980). Enzyme cleavage was used preferentially because strand separation is dependent on an assymmetric base composition in each strand of the duplex, and this assymetry is generally maximal in very small fragments. In all restriction enzyme manipulations after kinasing, 10 mM ATP was included in the reaction buffer to inhibit phosphatases which contaminate some commercial restriction endonuclease preparations. Serious losses of radioactivity had been encountered in digestions without inhibitors.

After secondary cleavage, subfragments were resolved by electrophoresis on thin polyacrylamide slab gels and detected by autoradiography. Fragments with a single labelled terminus were excised, eluted and the DNA recovered by ethanol precipitation.

The chemical cleavage reactions were as described by Maxam and Gilbert (1980) except that 100% formic acid was used for the G+A reaction instead of pyridinium formate (R. Richards, personal communication). After cleavage, the DNA was recovered by lyophilization, resuspended in formamide gel loading buffer and electrophoresed on long (0.5 mm x 30 mm x 40 mm) high resolution polyacrylamide-urea sequencing gels (Sanger and Coulson, 1978). The acrylamide percentage of the gels and the time of electrophoresis could be varied to maximize

resolution of different areas of the sequence.

Normally to read up to 200 bps of sequence, a single loading of sequencing reactions would be run on a 20% gel (bps 1-60), and two staggered loadings on an 8% gel (bps 40-200). Gels were run warm (25 mA) to maximize the denaturation effect of the urea. Bad compressions in sequencing ladders, due to inefficient denaturation of regions of DNA were sometimes seen when gels were run cold. After electrophoresis gels were transferred from the glass gel plates to an exposed sheet of X-ray film, covered in plastic wrap and autoradiographed at -80°C in the presence of an intensifying screen.

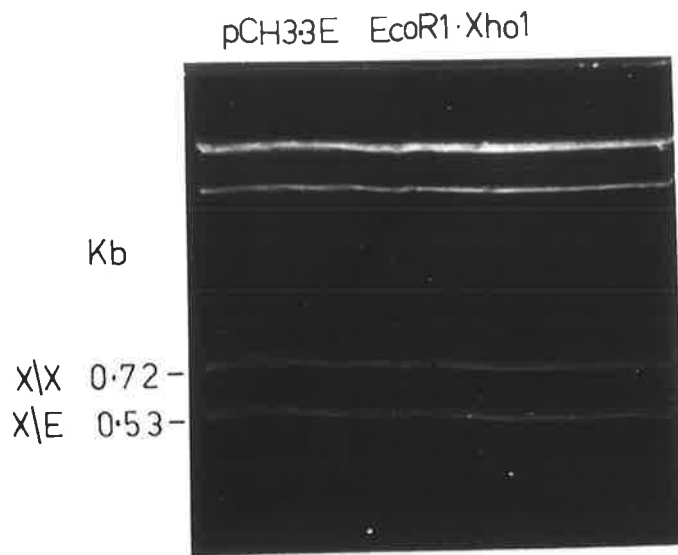
8.2.2 The H2B Gene of pCH3.3E

As discussed in section 5.2.2, the H2B gene of pCH3.3E was either wholly or partially contained within the 0.53 Kb XhoI/EcoRI fragment of the pCH3.3E insert (Fig. 5.2). Because XhoI and EcoRI generate 5' protruding termini, a suitable initial strategy in the sequencing of the fragment was to sequence in from its ends. The XhoI/EcoRI fragment was isolated (section 8.2.1) from a XhoI/EcoRI digest of pCH3.3E after electrophoresis on a thin polyacrylamide gel (Fig. 8.2(a)). The fragment was kinased and further digested with HincII to separate the labelled termini (Fig. 8.2(b)). Resolved and eluted fragments were subjected to Maxam and Gilbert sequencing reactions (section 8.2.1). Sequencing ladders produced

FIGURE 8.2

Isolation and end-labelling of the XhoI/EcoRI fragment of pCH3.3E containing the H2B gene.

- (a) pCH3.3E DNA (50 μ g) was digested with XhoI and EcoRI and electrophoresed on a 6% polyacrylamide gel. Bands were detected by EtBr staining, and the 0.53 Kb XhoI/EcoRI fragment excised and eluted (section 2.2.7).
- (b) The XhoI/EcoRI fragment was end-labelled with polynucleotide kinase and α ³²P ATP (section 2.2.7) and then digested with HincII to separate the labelled ends. DNA was then electrophoresed on a 6% polyacrylamide gel and bands were detected by autoradiography.



(a)



(b)

from these fragments are shown in Figure 8.3. From these ladders, 294 bps were determined from the XhoI end and 153 bps from the EcoRI end. The sequences (Fig. 8.4) indicated that this gene coded for an H2B protein (Elgin and Weintraub, 1975), and that the coding region was wholly contained within the XhoI/EcoRI fragment. The 3' untranslated region was cut short by the original cloning procedure since only 3 bps of untranslated region were present before the EcoRI linker (section 3.2.3).

To obtain further restriction information on the H2B fragment which might be useful in further sequencing strategy, an analytical XhoI/EcoRI/HincII digest was electrophoresed on a polyacrylamide gel and the H2B XhoI/HincII and EcoRI/HincII fragments (see Fig. 8.3) excised and eluted. The eluted DNA was kinased and aliquots were digested with the restriction endonucleases HaeIII, HhaI, HinfI, MboII and MspI (Fig. 8.5). The two ends of the fragments could be distinguished in each digestion because the 5' protruding XhoI and EcoRI ends labelled much more efficiently than the flush HincII end. (section 8.2.1). This gel yielded information on the distance in from known termini, of various enzyme sites.

The HincII site within the H2B fragment represented a single site, centrally located within the unsequenced region. When it became possible to label flush ends efficiently by exonuclease III treatment (section 8.2.1), a strategy was devised whereby sequence information would

FIGURE 8.3

DNA sequencing ladders produced by Maxam and Gilbert sequencing of the XhoI/EcoRI fragment from pCH3.3E containing the H2B gene. After separation of labelled ends by digestion with HincII and resolution on a polyacrylamide gel (Fig. 8.2), the EcoRI/HincII and XhoI/HincII sub-fragments were eluted and sequenced using the Maxam and Gilbert procedures (section 2.2.7). A double loading of sequencing reactions were electrophoresed on a 6% polyacrylamide-urea sequencing gel (section 6.2.1). Ladders were detected by autoradiography.

GATC



Hinc 2·Xho2

Hinc2·EcoR1

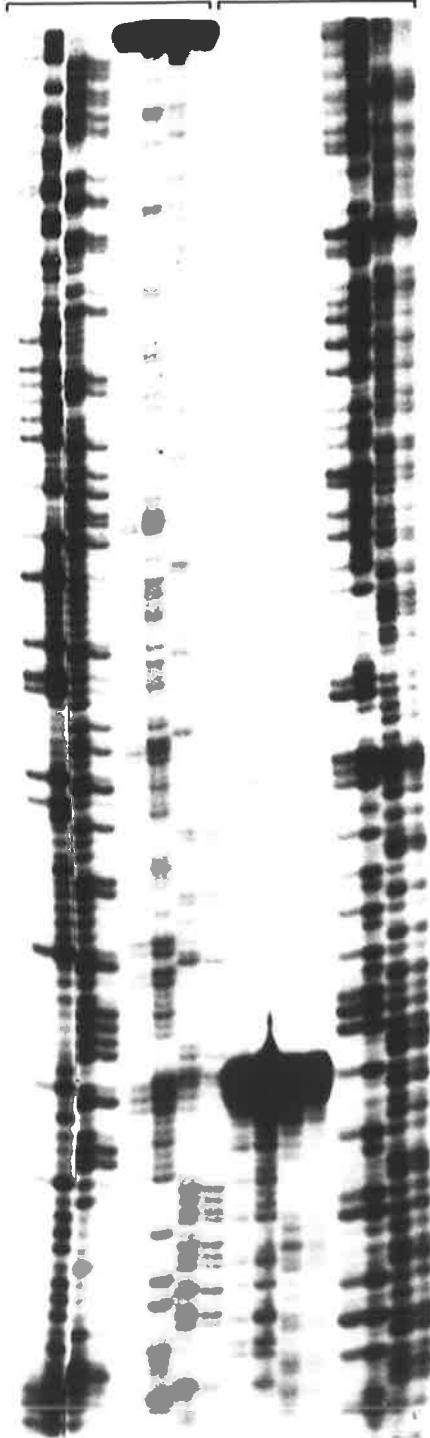
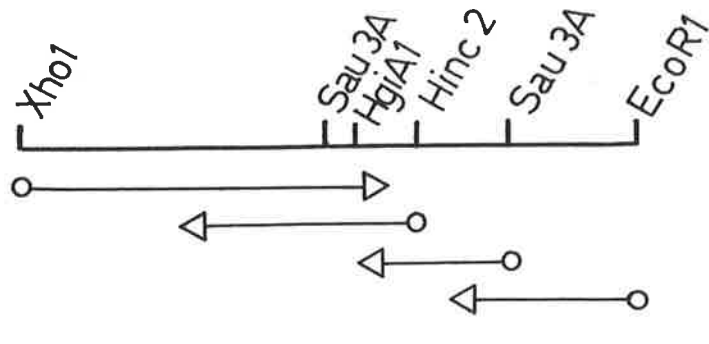


FIGURE 8.4

The DNA sequence of the H2B gene of pCH3.3E. The restriction map indicates the restriction sites from which the sequence was determined. The insert of λ CH-01 ends within the 3' untranslated region of this H2B gene, and so the 3' sequence has not been determined.



-134 -68
 AATGAAAGAGTGCGAAAGGAATGCTTCTCATTTCGCATAGAGGGGCTATAAAATAAAATGCCCTACGACCC

-67 -1
 CTTTCGTTTCATTCAGCGTCTCCTGGTCGTTTTTGTTCGCCTCGCTTCGTGAGCGCGTTGTGGCCACT

1 51
 ATG CCC GAG CCG GCT AAG TCC GCG CCC GCC CCG AAG AAG GGC TCT AAG AAG
pro glu pro ala lys ser ala pro ala pro lys lys gly ser lys lys

52 102
 GCG GTC ACC AAG ACC CAG AAG AAG GGC GAC AAG AAG CGC AAG AAG AGC CGC
ala val thr lys thr gln lys lys gly asp lys lys arg lys lys ser arg

103 153
 AAG GAG AGC TAC TCG ATC TAC GTG TAC AAG GTG CTG AAG CAG GTG CAC CCC
lys glu ser tyr ser ile tyr val tyr lys val leu lys gln val his pro

154 204
 GAC ACG GGC ATC TCG TCC AAG GCC ATG GGC ATC ATG AAC TCG TTC GTC AAC
asp thr gly ile ser ser lys ala met gly ile met asn ser phe val asn

205 255
 GAC ATC TTC GAG CGC ATC GCC GGC GAG GCG TCG CGC CTG GCG CAC TAC AAC
asp ile phe glu arg ile ala gly glu ala ser arg leu ala his tyr asn

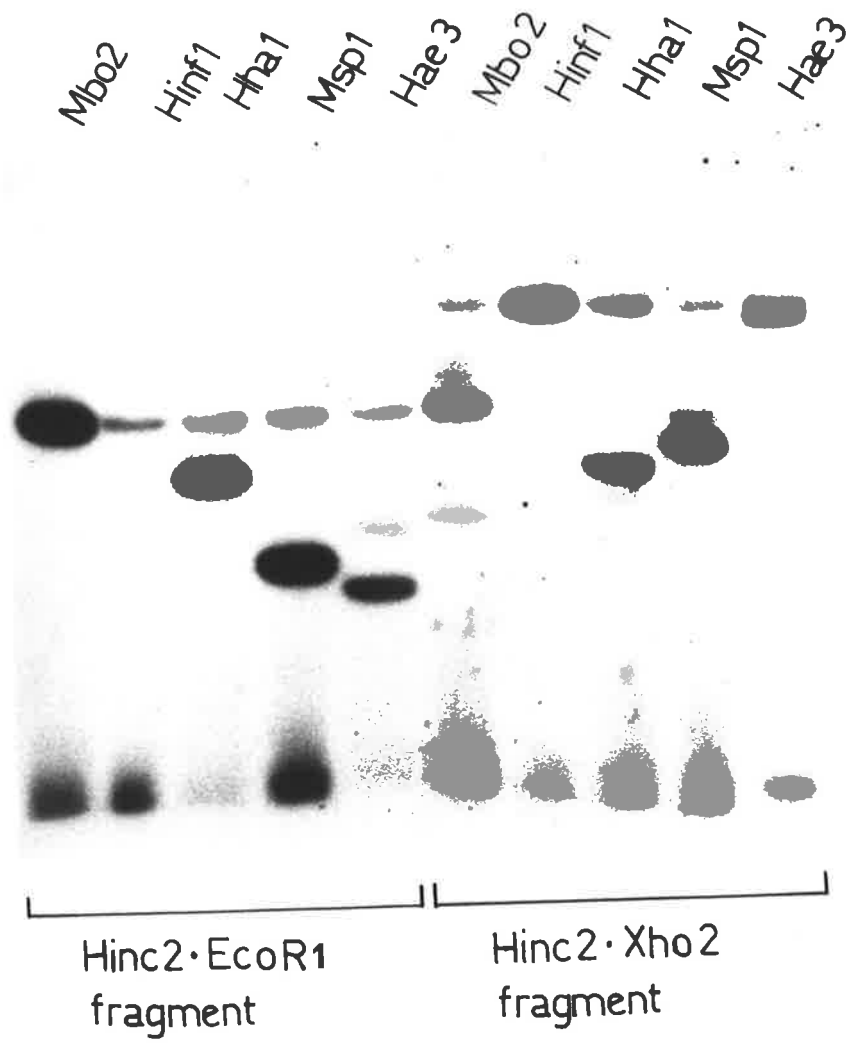
256 306
 AAG CGC TCG ACC ATC ACG TCG CGG GAG ATC CAG ACA GCC GTG CGG CTG CTG
lys arg ser thr ile thr ser arg glu ile gln thr ala val arg leu leu

307 357
 CTG CCC GGC GAG CTG GCC AAG CAC GCG GTC TCG GAG GGC ACC AAG GCG GTC
leu pro gly glu leu ala lys his ala val ser glu gly thr lys ala val

358 391
 ACC AAG TAC ACC AGC TCC AAG TAA GCATGAATT
thr lys tyr thr ser ser lys

FIGURE 8.5

Restriction analysis of the pCH3.3E H2B gene for use in sequencing procedures. The XhoI/HincII and EcoRI/HincII fragments containing the H2B gene were kinased (section 2.2.7) and then further digested with the restriction enzymes MboII, HinfI, HhaI, MspI and HaeIII. These digests were electrophoresed on a 6% polyacrylamide gel and bands containing the labelled termini were detected by autoradiography.



be obtained outwards from this HincII site.

The XhoI/EcoRI H2B fragment isolated as described above, was further digested with HincII, treated with exonuclease III, dephosphorylated and kinased. This material was secondarily cut with MspI because the data in Figure 8.5 indicated that this enzyme cut both the Xho/HincII and the EcoRI/HincII fragments once. The digestion products were resolved on a polyacrylamide gel and fragments eluted and sequenced. Sequences from the HincII site towards the EcoRI site were not obtained as MspI cut this fragment centrally and so sub-fragments ran as a doublet on the gel. Sequences obtained from the HincII site towards the XhoI site (138 bps) overlapped the sequences obtained in from the XhoI site (Fig. 8.4).

To finish the H2B fragment a strategy was required which spanned the unsequenced region and overlapped areas already sequenced. Sequencing in from the HincII site towards the EcoRI site using a different enzyme for secondary cleavage would leave a single base within the HincII site uncertain. The HincII site contains a redundant base and it is often difficult to determine the first base of a sequencing ladder.

The sequences already determined allowed positioning of Sau3A sites close to the borders of the unsequenced region (Fig. 8.4). From the calf H2B protein sequence it could be predicted that no Sau3A sites would occur within the unsequenced region. This region corresponded

to the highly conserved portion of the H2B protein. Also working from the protein sequence, the Sau3A fragment could be deduced to be 165 bps.

A Sau3A digest of pCH3.3E was dephosphorylated and electrophoresed on a preparative polyacrylamide gel (Fig. 8.6(a)). The 165 bp Sau3A fragment was eluted and kinased. An HgiAI site within the already sequenced region (Fig. 8.4) was found to be suitable for secondary cleavage. After digestion with HgiAI the sub-fragments were resolved on a polyacrylamide gel (Fig. 8.6(b)). The larger of the sub-fragments was eluted and sequenced. The sequencing ladder generated from this fragment is shown in Figure 8.7. The sequence determined from this ladder (90 bps) spanned the unsequenced region and overlapped sequenced regions at both ends.

The complete sequence of the XhoI/EcoRI H2B fragment is shown in Figure 8.4 together with a summary diagram of the strategies used.

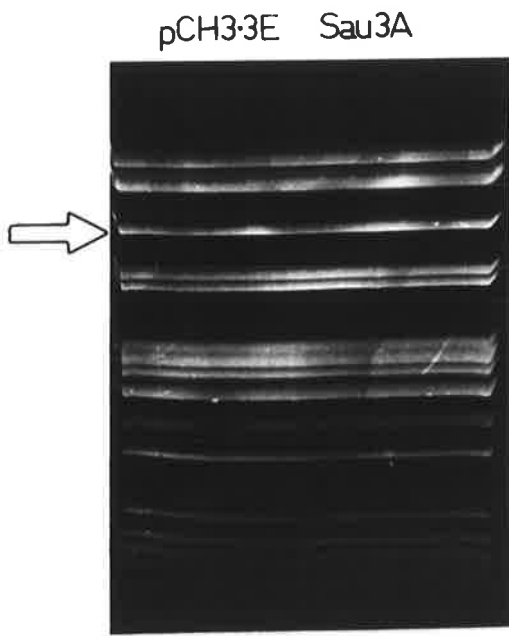
8.2.3 The 5'-Leader Region of the pCH3.3E H2A Gene

The DNA sequence of the H2A gene contained within the 0.72Kb XhoI fragment adjacent to the XhoI/EcoRI H2B fragment of pCH3.3E (Fig. 5.2) was determined by R. D'Andrea (this laboratory). The sequence included the sequence TATA, first observed by Goldberg (1979) and conserved in most eukaryotic genes. Busslinger *et al.*, (1980) had observed a region of sequence 5' to the

FIGURE 8.6

Isolation and end-labelling of a 165 bp Sau3A, H2B gene fragment of pCH3.3E for sequencing.

- (a) pCH3.3E DNA (50 μ g) was digested with Sau3A and electrophoresed on a 6% polyacrylamide gel. Bands were detected by EtBr staining. The 165 bp Sau3A fragment (arrowed) was excised and eluted.
- (b) The 165 bp Sau3A fragment was end-labelled with polynucleotide kinase and $\gamma^{32}\text{P}$ ATP (section 2.2.7), further digested with HgiAI and electrophoresed on a 6% polyacrylamide gel. Sub-fragments were detected by autoradiography. The larger band was excised and eluted for DNA sequencing.



(a)



(b)

FIGURE 8.7

DNA sequencing ladders generated from the Sau3A fragment (Fig. 8.6) of pCH3.3E containing part of the H2B gene. Eluted, end-labelled Sau3A/HgiAI DNA (Fig. 8.6) was sequenced using the Maxam and Gilbert procedures (section 2.2.7). Reactions were electrophoresed on a 6% polyacrylamide-urea sequencing gel (section 6.2.1) and ladders detected by autoradiography.

"TATA box" which was conserved between two sea urchin H2A gene variants. To see whether this region was ubiquitous to all H2A genes and to allow comparisons with the 5' flanking regions of the adjacent H2B gene, the sequences of R. D'Andrea were extended in the 5'-direction by 139 bps to the SmaI site (Fig. 5.2). The SmaI/XhoI fragment adjacent to the H2A gene was isolated, end labelled and strand separated (section 2.2.7). Both strands were sequenced using Maxam and Gilbert procedures. The sequence of this region together with the sequence of the XhoI H2A fragment determined by R. D'Andrea is shown in Figure 8.8.

8.2.4 DNA Sequencing Procedures Using the Enzymatic Chain-Termination Method

The "dideoxy" chain termination sequencing procedure of Sanger *et al.*, (1977), relies on the enzymatic generation of a series of randomly terminated fragments with a common 5' terminus. The termination reactions are base specific and collectively the 3' variable terminus is representative of every base in the stretch of DNA to be sequenced.

E. coli or viral DNA Polymerase is used to faithfully synthesize a (radioactive) copy of a single stranded DNA template in the presence of deoxynucleoside triphosphates, one or more of which are labelled. A short DNA fragment complimentary to the 3' end of the template is used as polymerase primer. Random base-specific termination in chain synthesis is achieved by the addition of analogues

FIGURE 8.8

The DNA sequence of the 5' flanking region of the H2A gene within pCH3.3E, together with the sequence of the H2A gene determined by R. D'Andrea (section 8.2.3).

-333 -326
5' GGGAGAAA

-325 -261
CTTCTTTTGGGAGAACGCTTTGGGCACGACTTTGTTAACGGAAGCATGGAAAGCGTTGCTATTAT

-260 -196
TACCCACCAAATAATACTGATAATAAATATGAGAAAAAAAAAAGAAAAAAAAAAGCACGGCTCG

-195 -131
AGAACACGCCTTTCCTCCCTTATAACTGCTTTTCATTGGTTCAAATTCGATTCGCTTTCTCATTG

-130 -66
GCTGCCAGAGCGGACCCAGACGTCAGCCATCAGCGCAGAGGGCAGGAGCCAAAGCGAGAGCGT

-65 -1
AGCGAGCCCGTAGGTTGCGCGCTGCGTTCTTGGCTTGTTGCTCTCTGAGTGTTCAAGTCGCTGCG

1 51
ATG TCG GGG CGC GGA AAG CAG GGC GGG AAG GCG CGC GCC AAG GCC AAG TCG
met ser gly arg gly lys gln gly gly lys ala arg ala lys ala lys ser

52 102
CGC TCG TCG CGG GCC GGG CTG CAG TTC CCC GTG GGC CGC GTG CAC CGG CTG
arg ser ser arg ala gly leu gln phe pro val gly arg val his arg leu

103 153
CTG CGC AAG GGC AAC TAC GCG GAG CGG GTG GGC GCC GGC GCC CCG GTG TAC
leu arg lys gly asn tyr ala glu arg val gly ala gly ala pro val tyr

154 204
CTA GCG GCC GTG CTG GAG TAC CTG ACG GCC GAG ATC CTG GAG CTA GCG GGC
leu ala ala val leu glu tyr leu thr ala glu ile leu glu leu ala gly

205 255
AAC GCG GCC CGC GAC AAC AAG AAG ACG CGC ATC ATC CCC CGC CAC CTG CAG
asn ala ala arg asp asn lys lys thr arg ile ile pro arg his leu gln

256 306
CTG GCC ATC CGC AAC GAC GAG GAG CTC AAC AAG CTG CTG GGC AAG GTG ACC
leu ala ile arg asn asp glu glu leu asn lys leu leu gly lys val thr

307 357
ATC GCG CAG GGC GGG GTG CTG CCC AAC ATC CAG GCC GTG CTG CTG CCC AAG
ile ala gln gly gly val leu pro asn ile gln ala val leu leu pro lys

358 412
AAG ACC GAC AGC CAC AAG GCC AAG GCC AAG TGA GCACCGGCAGGCAGCGCTGTC
lys thr asp ser his lys ala lys ala lys STOP

413 478
TGAGAGAACAGTCCAAAGCTCTTTTCAGAGCCACCCACAGCATCGCAGGAGAGCTCAGAAATCCGC

479 510
AATACAGTCGTGCAGGTCTATGAATTACTCGA 3'

of the deoxynucleotides, 2', 3'-dideoxynucleoside triphosphates, to four separate polymerase reactions. When incorporated, these analogues no longer serve as templates for chain elongation and the growing DNA chain is terminated. This results in four populations of partially synthesized DNA molecules, each radioactive, each having a common 5' terminus (the primer) and each having a variable but base-specific 3' terminus. Resolution of these populations of fragments on adjacent tracks of a polyacrylamide-urea sequencing gel (Sanger and Coulson, 1978), results in ladders from which the DNA sequence may be read.

This method offers considerable advantages over other rapid sequencing procedures. The polymerase reaction is easily controlled and offers a high degree of specificity. In addition, less radioactivity is handled and less hazardous chemicals are required.

Recent genetic manipulation of the bacteriophage M13 into a cloning vector, and the construction of a "universal" primer have overcome the two main limitations of the technique, i.e. the need for the template to be in single stranded form and the tedious preparation of restriction fragment primers complimentary to the 3' end of every 200-300 bps to be sequenced.

M13 is an *E. coli*, male specific, single stranded, filamentous phage (Dehardt *et al.*, 1978) which has been specifically engineered (Schreier and Cortese, 1979) to allow cloning of restriction fragments.

It is a non-lethal virus which is extruded into the *E. coli* growth medium without cell lysis. There appears to be no constraints associated with the capsid-filling mechanism of M13, so inserts of a variety of sizes can be tolerated.

The resultant vectors, M13mp7, M13mp8 and M13mp9, exhibit features of an excellent double stranded vector system, while also having a single stranded counterpart.

- (1) The intracellular double stranded replicative form (RF) is small, represented in high copy number, and contains a concentration of restriction enzyme sites into which a diversity of fragments may be cloned.
- (2) Inactivation of an engineered β -galactosidase gene occurs upon insertion of fragments within cloning sites and caused a change from blue to clear plaques in appropriate media.
- (3) The preparation of the single stranded form of the virus from the culture medium is rapid and straightforward (5-10 μ g per ml of culture).
- (4) Both complimentary strands of the insert can be obtained independently, each from separate clones with the insert in opposite orientations.
- (5) The clustering of cloning restriction sites in these vectors means that a single "universal" primer may be used for all sequencing reactions. This primer is complimentary to the M13 region adjacent to the cloning sites.

The combination of the M13 vector system and the

"dideoxy" sequencing technique has resulted in an extremely powerful sequencing procedure. Extensive regions of DNA can be sequenced using "shot-gun" sequencing approaches (Sanger *et al.*, 1980), where large DNA fragments are randomly sheared into smaller pieces which are cloned into M13 and sequenced. Sequence data is then overlapped with the aid of a computer programme into one large final sequence. The complete genomes of (50 Kb) human and bovine mitochondria (Anderson *et al.*, 1981), and bacteriophages ϕ x174 (Sanger *et al.*, 1978), and λ (unpublished) have been sequenced using these methods.

8.2.5 The H2B Gene of pCH4.6E

Only a limited description of the cloning of fragments generated from the 4.6 Kb EcoRI fragment of λ CH-02 containing an H2B gene (section 6.2.4) will be given here as these manipulations were carried out by A. Robins (this laboratory).

Preparative amounts of the 4.6 Kb EcoRI insert of pCH4.6E were isolated from a low melting temperature agarose gel. 4.6 Kb DNA was restricted with AluI and RsaI and ligated directly into the HincII site of M13mp7 by blunt end ligation (section 2.2.4).

A random library of clones was also generated by sonication. Sonication conditions were established which gave a concentration of shear fragments between 0.5-1.5 Kb. These fragments were fractionated on a 1.5% agarose gel

and DNA greater than 500 bps was eluted by electrophoresis into a well cut in the gel and filled with gel buffer. When cloning restriction fragments, double insertion events can be recognized by the regeneration of that particular restriction site. Double insertion of sonicated fragments cannot be recognized in this way, so fragments are cloned which are just longer than the maximum length which can be read from a sequencing gel (about 300 bps). If two fragments are cloned, sequences are generated from only one of these.

The ragged ends generated by sonication were repaired with DNA polymerase and fragments were ligated into the HincII site of M13mp7.

Ligated DNAs were transformed into competent *E. coli* bacteria. Clear plaques, indicating the presence of recombinants, were toothpicked into 1 ml vials of broth and grown for 6 hours. Bacteria were removed from cultures by centrifugation and M13 phage precipitated from the broth with polyethylene glycol. Phage precipitates were pelleted, then phenol extracted to remove the M13 coat protein. DNA was recovered by ethanol precipitation and resuspended in preparation for DNA sequencing.

Sequencing of the M13 single stranded recombinant templates was carried out in collaboration with A. Robins. It was convenient to sequence up to eight templates at one time. Single-stranded circular phage DNA and "universal primer" were mixed and sealed in capillary tubes. The mixture was denatured by boiling and annealed

at room temperature. Each annealing mix was split into four aliquots and synthesis from the primer was catalysed with the large fragment of DNA polymerase I (Klenow *et al.*, 1971) in the presence of the four deoxynucleotides and in each reaction, one of the four dideoxynucleotide analogues (Atkinson *et al.*, 1969). Radioactivity was incorporated into the synthesized fragments by using α - ^{32}P dGTP as one precursor triphosphate. After synthesis the reaction was "chased" with unlabelled dGTP to eliminate any non-specific chain termination due to limiting concentrations of ^{32}P dGTP. Further reaction was stopped by the addition of polyacrylamide gel, formamide loading buffer.

Prior to electrophoresis, samples were boiled to denature templates and products. Parallel size fractionation of the four DNA synthesis reaction products was carried out on 0.3 mm denaturing polyacrylamide-urea sequencing gels (Sanger and Coulson, 1978). Sequencing ladders were of the desired intensity after 12 hours autoradiography without an intensifying screen. The use of intensifying screens caused some loss of band resolution.

The sequences generated from these M13 sub-clones of the 4.6 Kb EcoRI insert of pCH4.6E were scanned for sequences that would predict an H2B protein sequence. Sufficient overlapping data was found to assemble the complete H2B gene sequence and some 5' and 3' flanking sequences. Figure 8.9 shows sequencing ladders of M13

FIGURE 8.9

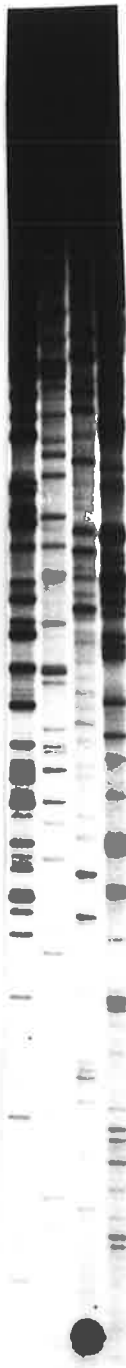
DNA sequencing ladders generated from
M13 clones of pCH4.6E DNA (section 8.2.5)
by the "dideoxy" sequencing methods
(section 8.2.4).

TCGA
→

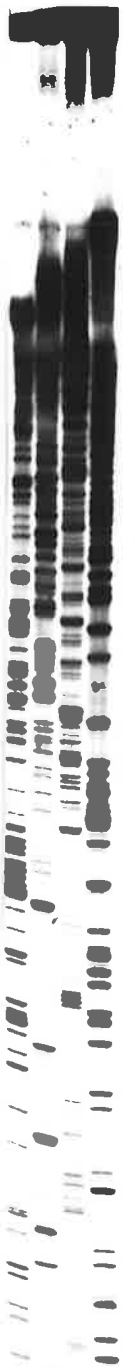
CHE·3R



CHE·11R



CHE·12R



clones containing H2B gene inserts. The DNA sequences determined for the H2B gene and its flanking sequences is shown in Figure 8.10.

Assembly of further regions of the pCH4.6E insert from the generated sequence data awaits the development of computer assistance.

8.3 DISCUSSION

The topology of the H2A and H2B genes described in this chapter, conform to the present notions of histone gene structure. No intervening sequences are apparent although further analysis will be required to see if they exist in non-translated regions. Lack of intervening sequences seems to be characteristic of, but not unique to histone genes. Other eukaryote and viral genes, notably the human interferon genes (Ohno *et al.*, 1981; Lawn *et al.*, 1981) also lack intervening sequences.

The genes possess putative regulatory elements (see below) and no frameshift mutations or deletions occur within the structural genes. Although this does not directly prove that the genes are expressed, it appears likely that they are since non-functional genes would accumulate mutations at a high rate. They certainly do not have the characteristics of "pseudogenes" observed in other systems (Vanin *et al.*, 1980; Childs *et al.*, 1981). The apparent structural integrity of these genes strengthens the idea that the genes observed in λ CH-01

FIGURE 8.10

The DNA sequence of the H2B gene of pCH4.6E.
This sequence was assembled from sequences
generated from M13 clones of the whole
pCH4.6E insert (section 8.2.5).

CTGTTATCCAATCAGAGAGCAGATACAGAAGGCACTCGATTTGCATACTGCCCC

-67

-1

TATAAATAGGCGAGCAGTGTCTCGCAGCCGGCACTCCGCTGCGCCGAAGGGATCGTGGAGAGTTTCGAC

1 ATG CCT GAG CCG GCC AAG TCC GCA CCC GCC OCC AAG AAG GGC TCC AAG AAG 51
pro glu pro ala lys ser ala pro ala pro lys lys gly ser lys lys

52 GCG GTC ACC AAG ACC CAG AAG AAG GGC GAC AAG AAG CGC AAG AAG AGC CGC 102
ala val thr lys thr gln lys lys gly asp lys lys arg lys lys ser arg

103 AAG GAG AGC TAC TCG ATC TAC GTG TAC AAG GTG CTG AAG CAG GTG CAC CCC 153
lys glu ser tyr ser ile tyr val tyr lys val leu lys gln val his pro

154 GAC ACG GGC ATC TCG TCC AAG GCC ATG GGC ATC ATG AAC TCG TTC GTC AAC 204
asp thr gly ile ser ser lys ala met gly ile met asn ser phe val asn

205 GAC ATC TTC GAG CGC ATC GCC GGC GAG GCG TCG CGC CTG GCG CAC TAC AAC 255
asp ile phe glu arg ile ala gly glu ala ser arg leu ala his tyr asn

256 AAG CGC TCG ACC ATC ACG TCG CGG GAG ATC CAG AXG GCC GTG CGG CTG CTG 306
lys arg ser thr ile thr ser arg glu ile gln thr ala tyr arg leu leu

307 CTG CCC GGC GAG CTG GGC AAG CAC GCG GTC TCC GAG GGC ACC AAG GCG GTC 357
leu pro gly glu leu ala lys his ala val ser glu gly thr lys ala val

358 ACC AAG TAC ACC AGC TCC AAG TAG AGCGGTGCGGATTTACTCGATTTTAAACCCAAAGGCT 416
thr lys tyr thr ser ser lys

417 CTTTTTCAGAGCCACCATTTGTTC TAATAAAAGGGCTGTATTACTTTTTTTTCTTTTTTTCTGAGGGG 483

484 TATAGCGTGGGTTAACTGAGTGAAATGGAAGCGAGTGCCCTGAGGTATGTATATAATTGCTTAACTTC 550

551 GCAGTTCGGAGGTTCTCCGTTCCGAGTTAATTGAGCAGTAGCAACTCCCGGAGGTTAACTGGGTTGTTC 617

618 645
GGTAGCCGTGCTATTACTGCAGCACGGT

and λ CH-02 are genuine representatives of the "disordered" chicken histone system, and not a minority population akin to the sea urchin "orphons" (Childs *et al.*, 1981).

The amino-acid sequences coded for by the two H2B genes are identical (Figs. 8.4 and 8.10). This is the first report of a chicken histone H2B sequence and it is likely to correspond to the chicken erythrocyte H2B(1) variant, isolated and peptide-mapped by Urban *et al.*, (1979). A detailed comparison with the H2B sequences of calf thymus, human spleen, trout testis, patella testis, *Drosophila* and one sea urchin sperm subtype (Busslinger *et al.*, 1980), is shown in Figure 8.11.

The H2B protein is an evolutionary hybrid, with a hypervariable N-terminal region and an extremely conserved C-terminal region. The latter region is involved in contact with other histones in the nucleosome (McGee and Felsenfeld, 1980). The N-terminal portions of H2B proteins of distantly related species, e.g. calf and *Drosophila*, are extremely dissimilar. More closely related species show fewer differences. This region of the chicken H2B sequence is closely homologous to both the trout and calf sequences, reflecting the evolutionary relationship of these species. The only amino-acid change in a position which is shown to be invariant in the H2B sequences of other species is the lysine at amino-acid 31 (Fig. 8.11). All other species have

FIGURE 8.11

Comparison of the chicken H2B amino acid sequence with the H2B sequences of other species (tabulated in Busslinger *et al.*, 1980). The amino acid sequence of the chicken H2B is written in full (one letter abbreviations, see Busslinger *et al.*, 1980). For the other species, only those amino acids varying from the chicken sequence are indicated. The star (*) indicates the lysine of the chicken sequence which is an invariant arganine in the H2B sequences of other species.

H2B

(1) CHICK	PEPAKSAPAPKKGSKK
(2) CALF THYMUS AND HUMAN SPLEEN	
(3) TROUT TESTIS	--
(4) PATELLA TESTIS	PPKVSSKGAKK
(5) DROSOPHILA	PPKTSGKAAKKA
(6) SEA URCHIN VARIANT B	PRSPAKTSPRKGSPRKGSPRKGSPSRKASPKRGGKGAKR

(1) AVTKTQKKGD [*] KRRKKS	RKESYSIYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEASRLAHYNKRST
(2) A DG R V	S
(3) AG G R A	V
(4) AGKAKAARS RR	S A
(5) GKAQKNITKT K RK A	S A
(6) AGKGRRRRVV RRRR G I	R SV V S TSA R

(1) ITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK	
(2)	
(3)	
(4)	
(5)	
(6) V	T R

arginine in this position. However it is probably a coincidence that no other H2Bs have been described with lysine in this position, because the change is conservative, and is embedded in an extremely basic region of 8 amino acids with almost total arganine/lysine content.

The DNA sequences of the H2B structural genes and their 5' non-translated regions are shown aligned in Figure 8.12. The pCH3.3E H2B sequence shows a well defined "CAP box" (see below) recognized as containing the start point of transcription (Hentschel and Birnstiel, 1981). The exact initiation point is not known, so Figure 8.12 includes the 5' untranslated region and the complete "CAP box". The same sequence cannot be recognized in the pCH4.6E gene, however, the distance from the "TATA box" to the "CAP box" is well conserved (Hentschel and Birnstiel, 1981), so sequences are shown to a point 30 bps from the first T in the "TATA box", equivalent to the pCH4.6E H2B gene.

As discussed above, the structural genes code for identical H2B sub-types. Between the DNA sequences of the genes, there are 7 base pair differences but all of these are silent changes to iso-coding triplets.

Similar to many other eukaryotic genes, there is a bias towards G/C in the third triplet position. In both genes only three of 125 triplets do not end in G or C. It is commonly suggested that this bias is related to secondary structure stabilities in mRNA molecules, however

FIGURE 8.12

The DNA sequences of H2B genes from pCH3.3E (section 8.2.2) and pCH4.6E (section 8.2.5) shown aligned for sequence comparison. The stars (*) indicate the differences between the two gene sequences.

H2B pCH3.3E CCAATTCAGCGTCTCCTGGTCGTTTTGTTTCGCCTCGCTTCGTG
H2B pCH4.6E GCACTCCGCTGCGCCGAAGGGA

AGCGCGTTGTGCCACTATG CCC* GAG CCG GCT* AAG TCC GCG* CCC GCC CC*
TCGTGGAGAGTTCGACATG CCT GAG CCG GCC AAG TCC GCA CCC GCC CCC

AAG AAG GGC TCT* AAG AAG GCG GTC ACC AAG ACC CAG AAG AAG GGC
AAG AAG GGC TCC AAG AAG GCG GTC ACC AAG ACC CAG AAG AAG GGC

GAC AAG AAG CGC AAG AAG AGC CGC AAG GAG AGC TAC TCG ATC TAC
GAC AAG AAG CGC AAG AAG AGC CGC AAG GAG AGC TAC TCG ATC TAC

GTG TAC AAG GTG CTG AAG CAG GTG CAC CCC GAC ACG GGC ATC TCG
GTG TAC AAG GTG CTG AAG CAG GTG CAC CCC GAC ACG GGC ATC TCG

TCC AAG GCC ATG GGC ATC ATG AAC TCG TTC GTC AAC GAC ATC TTC
TCC AAG GCC ATG GGC ATC ATG AAC TCG TTC GTC AAC GAC ATC TTC

GAG CGC ATC GCC GGC GAG GCG TCG CGC CTG GCG CAC TAC AAC AAG
GAG CGC ATC GCC GGC GAG GCG TCG CGC CTG GCG CAC TAC AAC AAG

CGC TCG ACC ATC ACG TCG CGG GAG ATC CAG ACA* GCC GTG CGG CTG
CGC TCG ACC ATC ACG TCG CGG GAG ATC CAG AXG GCC GTG CGG CTG

CTG CTG CCC GGC GAG CTG GCC AAG CAC GCG GTC TCG* GAG GGC ACC
CTG CTG CCC GGC GAG CTG GCC AAG CAC GCG GTC TCC GAG GGC ACC

AAG GCG GTC ACC AAG TAC ACC AGC TCC AAG TAA*
AAG GCG GTC ACC AAG TAC ACC AGC TCC AAG TAG

it may also be a reflection of biased codon preferences, also observed in most eukaryote genes. This may have evolved to satisfy the available translation machinery of particular species. The histone genes of sea urchins exhibit extremely consistent codon-preference ratios within a single species, but these ratios differ significantly between species (Kedes, 1979). It is also feasible that protein synthesis rates may be regulated by gene sequences possessing uncommon codon triplets.

Although the DNA sequences of the two H2B structural genes have in fact diverged very little, the 5'-untranslated regions are totally dissimilar. The transcripts may also be of different length. If the point of transcription initiation of the pCH4.6E H2B gene lies at the same distance from the "TATA box" as in the pCH3.3E gene (30 bps), or other histone genes (23-32 bps; Hentschel and Birnstiel, 1981), then the two 5'-untranslated regions differ in length by at least 14 bps. This seems to indicate that the 5'-untranslated regions of these genes have no distinct sequence or length requirements. The divergence also indicates that the two genes are very distantly related. Homologies in the untranslated regions of various globin genes, reflect the evolutionary time since their origin (Richards *et al.*, 1979; Richards and Wells, 1980). Because histones are ubiquitous in eukaryotes, it is not surprising that these two chicken H2B genes have evolved

from an ancient duplication or amplification of a primitive H2B gene.

The search for regulatory elements in DNA sequences has, at least in eukaryotes, relied almost entirely on the comparison of the sequences of different genes and gene families, often across an evolutionary spectrum. In this way, elements common to the expression of all or most genes have been recognized and these constitute part of the "eukaryote promoter". Information more specific to selective expression has been obtained by comparison of DNA sequences from gene copies within a reiterated gene family (Jones and Kafatos, 1980), however this is often difficult due to the concerted evolution of reiterated genes (Brown *et al.*, 1972). Busslinger *et al.*, (1980), have made comparisons between sea urchin histone genes belonging to distinct developmentally related groups and have recognized sequences specific to gene types.

The H2B sequencing data suggests that these two genes are independently evolving. The disordered nature of the chicken histone genes also implies that the mechanisms which enforce the concerted evolution of sea urchin and *Drosophila* histone genes, do not apply in the chicken system. This situation is extremely favourable for recognition of regulatory sequences by making direct sequence comparisons. Independently evolving genes, coding for the same product, should exhibit common features important for expression or regulation.

Figure 8.13 tabulates the homology blocks which are found in the H2B and H2A genes. Comparisons are made with sea urchin variant H2B and H2A genes from clones h22 (Portman *et al.*, 1976) and h19 (Busslinger *et al.*, 1980), and also with a consensus sequence derived from the histone genes of several species, (Hentschel and Birnstiel, 1981).

The 5'-flanking sequences of the two chicken H2B genes show remarkable similarity, in that homology blocks are spaced exactly and are almost perfectly conserved in sequence. Three regions can be recognized. The "CAT box" (Benoist *et al.*, 1980) and the "TATA box" (Goldberg, 1979) are sequence motifs conserved in almost all eukaryote genes. Using *in vitro* mutagenesis and transcription systems, both of these sequence blocks have been implicated in correct quantitative and qualitative initiation of transcription. Several studies have shown that the "TATA box" is required for generation of faithful 5' termini (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981; Feye *et al.*, 1981; Dierks *et al.*, 1981) and also has influence on the rate of mRNA synthesis (Grosschedl and Birnstiel, 1980; Feye *et al.*, 1981; Dierks *et al.*, 1981; Wasylyk *et al.*, 1980). A single base change in the "TATA box" however (TATA to TAGA; Grosschedl *et al.*, 1981), only reduces the rate of synthesis. The most complete study has focused on systematic 5' and 3' leader sequence deletions of the Herpes Simplex Virus thymidine kinase gene (McKnight *et*

FIGURE 8.13

Common regions in the 5' leader sequences of the chicken H2A and H2B genes, and the H2A and H2B genes from sea urchin species (section 8.3). Consensus sequences (Hentschel and Birnstiel, 1981) are indicated.

FIGURE 8.13

5'-UNTRANSLATED AND FLANKING SEQUENCES

pCH3.3E H2B	GACCAATGA	...	25bps	...	CATTTGCATA	...	7bps	...	TATAAATA	...	22bps	...	CCATTCA	...	52bps	...	AUG
pCH4.6E H2B	ATCCAATCA	...	25bps	...	GATTTGCATA	...	7bps	...	TATAAATA	...	60bps	AUG
h22 H2B	AGCCAATCA	...	29bps	...	CATTTGCATA	...	26bps	...	TATAAAGA	...	21bps	...	GCACTCA	...	70bps	...	AUG
h19 H2B	GCCAATGA	...	20bps	...	CATTTGCATA	...	32bps	...	TATAAAAA	...	17bps	...	CCATTCA	...	74bps	...	AUG

HISTONE CONSENSUS	$\frac{G}{A}CCAAT-\frac{G}{A}$				TATAAATA				$\frac{T}{C}CATT\frac{G}{A}$								AUG
----------------------	--------------------------------	--	--	--	----------	--	--	--	------------------------------	--	--	--	--	--	--	--	-----

pCH3.3E H2A	AAAAAAAAAAGAAAAAAAAA	...	30bps	...	TATAACTG	...	19bps	...	CGATTG	...	142bps	...	AUG
h22 H2A	GGACAATTG	TATAAATA	...	25bps	...	CCATTCA	...	65bps	...	AUG
h19 H2A	GGACAATAG	TATAAATG	...	22bps	...	CCATTCA	...	65bps	...	AUG

"CAT BOX"

"TATA BOX"

"CAP BOX"

3'-UNTRANSLATED SEQUENCES

pCH3.3E H2A	TGA	36bps	AAAGGCTCTTTTCAGAGCCACC	6bps	CAGGAGAGC
pCH4.6E H2B	TAG	28bps	AAA-GCTCTTTTCAGAGCCACC	8bps	CTAATAAAA
h22 H2B	TAA	AACGGCCCTTTTCAGGGCCACC	6bps	CAAGAAAGA
h19 H2B	TAG	AACGGCCCTTTTCAGGGCCACC	6bps	CAAGAAAGA

HISTONE
CONSENSUS

AACGGC $\frac{C}{T}$ CTTTTCAG $\frac{A}{G}$ CCACC

CAAGAAAGA

al., 1981). A large region (60 bps) with the "CAT box" centrally located, is required for quantitative levels of transcription, although residual transcription does occur if it is deleted. When part of the "TATA box" is deleted from the 3' direction, a small reduction in transcription rate is noticed together with a decrease in the fidelity of the initiation point. However when all of the "TATA box" is deleted, transcription rate and fidelity are restored. Collectively, these results support what has been observed in independent studies - that regions containing the "TATA box" and the "CAT box" both influence the rate and fidelity of transcription.

Sequences further upstream have also been found to modulate the rate of transcription (Grosschedl *et al.*, 1980; Benoist and Chambon, 1981; Faye *et al.*, 1981; Dierks *et al.*, 1981; Gruss *et al.*, 1981; Guarente and Ptashne, 1981).

In addition to these "universal" sequence motifs, a region of 9 bps is seen in the two chicken H2B genes, conserved exactly in position and sequence. Remarkably, the same sequence precedes the H2B genes of the two *P. miliaris* clones (h19 and h22, Fig. 8.13), and also an H2B gene from the sea urchin *S. purpuratus* (Kedes, 1979).

Conservation of a gene specific sequence across such a long evolutionary time period is a powerful argument for its involvement in H2B gene expression.

Other gene specific sequences are seen in the prelude regions of all of the sea urchin histone genes of clone h22 and h19. The significance of some of these sequences are tenuous however, as these two clones are closely related in their spacer sequences (Busslinger *et al.*, 1980).

The positioning of the H2B specific sequence within the recognized "promotor region" of eukaryote genes, may be significant to the formation of an RNA polymeraseII "initiation complex" specific for transcription of H2B genes. Conservation of position suggests that promotor elements are stereochemically related and that a "complex" is indeed formed in the initiation event. Manipulation of this region in surrogate genetics systems may provide further clues to its role in H2B gene expression. Further sequencing upstream of these conserved elements should easily locate other promotor elements.

A comparison of the 5' flanking sequences of the pCH3.3E H2A gene (Fig. 8.13) shows some sequence elements common to histone, and other eukaryote genes (Hentschel and Birnstiel, 1981) and some unusual features. Examination of the flanking sequences in Figure 8.8 shows three possible "CAP-like" structures downstream from a "TATA box". All three "CAP boxes" have a six out of seven base fit with the consensus sequence (Fig. 8.13). One is centred 27 nucleotides from the first T in the "TATA box" and so is likely to be the true initiation

point (Fig. 8.13). However, in the sea urchin histone genes the position between the "TATA" and "CAP" boxes varies by up to 8 nucleotides, so it is possible that another, or all three "CAP boxes" are utilized as transcription initiation points.

In the H2A genes of h22 and h19, Busslinger *et al.*, (1980) have observed an "H2A-specific" sequence, 30 bps in length, 5' to the "TATA box". The sequence contains a region of dyad-symmetry and if organized into a stem-loop formation, a "CAT box" occurs in the loop. A similar structure can be drawn about the "CAT box" of the Herpes Simplex thymidine kinase gene. In transcription systems, fine deletion mapping of the thymidine kinase gene suggests that the maintenance of the dyad symmetry element is more critical to quantitative transcription than the "CAT box" itself (S. McKnight, personal communication).

Neither the sea urchin "H2A-specific" sequence, nor the "CAT box" structure is seen in the chicken H2A sequence. In the position where these elements would be expected, there is a run of 21 A residues, interrupted centrally by one G residue (Fig. 8.13). The position of this sequence and its particular asymmetry suggests that it may be important in regulation of the H2A synthesis. It will be instructive to compare these flanking sequences with those of other chicken histone H2A genes.

Sequencing of the 3' untranslated and flanking

regions of several sea urchin genes has revealed an impressive 23 bp and an adjacent 10 bp conserved sequence. The longer sequence contains a central region of dyad symmetry and the 3' termini of sea urchin mRNAs were found to map within a few nucleotides from the dyad element (Hentschel and Birnstiel, 1981). The 3'-terminal oligonucleotides of *L. pictus* H4 mRNA are ACCA-OH3' (Gurmstbein and Schedl, 1976), and this sequence can be found in the conserved region at a point where the 3' termini of the sea urchin mRNAs map. It has been suggested therefore, that this region is involved in termination or maturation of sea urchin histone gene transcription.

The 3' sequences of the pCH3.3E H2A gene and the pCH4.6E H2B gene show that the larger sea urchin homology block has been almost exactly conserved in the chicken genes. The smaller homology block appears to be poorly formed in both chicken genes. The extreme evolutionary conservation of the large homology block indicates that it has a specific function unique to histone genes.

The sequence AAUAAA, present at high frequency in eukaryotic mRNA trailer regions and proposed to be the signal for polyadenylation (Kedes, 1979), is not present in the chick or the sea urchin histone sequences. This is consistent with the observation that generally, histone mRNAs are not polyadenylated.

CHAPTER 9

FINAL DISCUSSION

9.1 INTRODUCTION

The aim of this thesis has been to understand some aspects of the control of histone gene expression. This problem has been approached from a structural standpoint - analysis of gene structure allows structure/function correlations to be made. The use of recombinant DNA technology in investigating gene structure, has resulted in an extremely rapid accumulation of knowledge on many gene systems.

Adopting this approach for chicken histone genes, recombinants have been isolated which contain chicken histone DNA (Harvey and Wells, 1979). These were the first vertebrate histone gene isolates to be reported. Subsequently, histone genes from *Xenopus* (Moorman *et al.*, 1980; Zernik *et al.*, 1980), newt (Stephenson *et al.*, 1981), chickens (Engel and Dodgson, 1981), mice (Sittman *et al.*, 1981) and humans (Heintz *et al.*, 1981; Sierra *et al.*, 1982a) have been isolated and partially characterized. It is obvious from this diverse data that the organization of histone genes has changed radically through evolution. This is despite the observed conservation of histone proteins across wide species barriers. The possibility therefore arises, that the mode of expression of histone genes has also drastically changed through evolution, as reflected by the changes in gene organization.

This chapter addresses this question, and attempts to place the observed structure for chicken histone genes in a functional and evolutionary context.

9.2 CHICKEN HISTONE GENES

Prior to the isolation of vertebrate histone genes, the sea urchin and *Drosophila* genes had been characterized in considerable detail. The dominant feature of both of these systems is that their histone genes exist as a cluster which is tandemly reiterated in a very conserved fashion. The dogma was therefore established that all histone genes were likely to be arranged in this way, and that that particular structure was intimately related to their expression.

The most significant conclusion to be made from this analysis of chicken histone genes is that a conserved tandemly repeating organization for histone genes is not absolute. In fact, the chicken organization displays a considerable degree of disorder. The genes are clustered at least to some degree, but no evidence for a repetitive unit has been found. No conservation of gene order or intergene DNA sequences has been observed. This conclusion has been substantiated by Engel and Dodgson (1981), who have reported preliminary analysis of three chicken genomal recombinants. Their clones exhibit similar features of disorder.

9.3 HISTONE VARIANT GENES

An important aspect of histone gene arrangement to be considered is the nature of the genes coding for variant proteins. Many histone variant sub-types have been observed in a diversity of species.

The developmental regulation of such sub-types is

best characterized in the sea urchin. Distinct variants are present in sea urchin sperm (Strickland *et al.*, 1978). Maternal mRNA synthesized during oogenesis, and zygotic transcripts during the exponential growth phase up to blastulation, are primarily encoded by the major, highly reiterated class of gene repeats, of which clone *h22* is a representative (Gross *et al.*, 1976). Interestingly, this is not the only class of repeat active during this period. Clone *h19*, a minor gene class, appears to code for variants which are also expressed in the embryo (Busslinger *et al.*, 1980). After blastulation, when the rate of cell division declines, "late" or "larval-type" genes are switched on (Childs *et al.*, 1979). On the basis of mRNA lengths and melting characteristics, the "late" mRNAs appear to be highly diverged from the embryonic genes. (Kunkel and Weinberg, 1978; Childs *et al.*, 1979; Spinelli *et al.*, 1979).

Apart from the embryonically expressed genes represented by clones *h22* and *h19* (Busslinger *et al.*, 1980), the localization and organization of the "sperm specific" and "late" histone variant genes is unknown. However, the distinction between "late" and "early" genes appears to be a real one, at least in the sea urchin, and must be considered in any evolutionary discussion.

In higher eukaryotes, similar arrays of variants have been observed. For instance, mouse L-cells contain four H2A variants and their ubiquitinated counterparts (West and Bonner, 1980). These variants are also found in normal chicken and human cells. In *Xenopus*, no

striking developmental changes in the synthesis of core histones are found, but changes in the ratios of H1 variants have been reported (Koster *et al.*, 1979; Flynn and Woodland, 1980; Risley and Eckhardt, 1981). Interestingly, these variant H1 genes are encoded in distinct gene clusters which have different gene orders (Zeinik *et al.*, 1980). The genes for all of the developmental variants must eventually be accounted for.

Another distinction between classes of histone genes has recently become evident. Wu and Bonner (1982) have characterized the variants from a Chinese hamster ovary cell line and determined that a specific class of histone variants is synthesized throughout the cell cycle (basal level synthesis). Most synthesis of histone proteins is S-phase specific. This introduces the possibility of another distinct histone gene class, based on "S-phase" and "basal" regulation of variant synthesis.

9.4 CHICKEN VARIANT-HISTONE GENES

Perhaps the first real indication of the existence of "classes" of histone genes which have unique organization or structure, is in the isolation of a variant chicken H3 gene by Engel *et al.*, (1982). This gene is not (over a 15 Kb range) intermingled with other histone genes and unlike all other histone genes characterized, contains two intervening sequences. It may be the first representative of the "late" class of histone gene (Engel *et al.*, 1982).

The major class of gene isolated from sea urchins

(clone *h22*, Gross *et al.*, 1976), is certainly of the "early" class. The genes present in λ CH-01 and λ CH-02 are most likely to be of the "early" class also, if such a clear distinction exists in vertebrates. This is suggested by the fact that the chicken genomic clones isolated by Engel and Dodgson (1981), of which one is identical to λ CH-01, were selected using sea urchin "early" gene probes. "Late" sea urchin mRNAs hybridize poorly to the "early" genes. In addition, an H3 gene derived from one of the chicken clones, hybridizes to embryonic, but not adult mRNAs (Engel *et al.*, 1982).

No indication has yet been obtained that λ CH-01 and λ CH-02 encode different variant genes. The H2B genes on λ CH-01 and λ CH-02 code for the same sub-type (Chapter 8). This also appears to be true of the two H2A genes of λ CH-01 (Chapter 8, and data from P. Krieg, this laboratory). The first 59 amino acids of the protein specified by the H3 gene of λ CH-01 (data of P. Krieg, this laboratory), are identical to the H3 protein encoded in the chicken recombinant, λ CH3d (Engel and Dodgson, 1981). It is likely, therefore, that the genes in λ CH-01 and λ CH-02, and the recombinants isolated by Engel and Dodgson (1981), encode a specific class of embryonic histone sub-types.

It would appear that the histone genes isolated from other species also belong to this class. *Drosophila* (Lifton *et al.*, 1977), *Xenopus* (Moorman *et al.*, 1980), newt (Stephenson *et al.*, 1981) and mouse (Sittman *et al.*, 1981) histone genes have all been isolated or characterized

using sea urchin "early" histone probes. It is therefore relevant to discuss the evolution of these histone genes as a class, without reference to the as yet, uncharacterized histone genes which encode developmental and cell-cycle variants.

9.5 THE EVOLUTION OF "EARLY" HISTONE GENES

The reiterated histone gene clusters of sea urchins and *Drosophila* are highly conserved. That is, they exhibit concerted evolution. The question arises as to whether this is a necessity for expression of histone genes, or a consequence of the fact that they are highly reiterated. The "randomization" of histone genes through their evolution to vertebrates, argues that histone genes are independently controlled and that arrangement into conserved arrays is not a requirement for expression. An explanation is therefore needed to account for the evolution of histone genes from a highly ordered to a considerably disordered disposition.

9.5.1 Homogenization of Reiterated Genes

Ever since the discovery of gene families, the homogeneity of their sequences has posed an evolutionary problem (Hood *et al.*, 1975). In the case of the histone genes, where the scope for evolutionary divergence in the protein coding regions is minimal, maintenance of gene clusters in conserved arrays would be a convenient mechanism to suppress evolutionary variability. However, given the inevitability of mutation, a specific mechanism

of gene correction must exist.

Recently, evidence has accumulated suggesting that the genomes of eukaryotes are much more mobile than once imagined. Two mechanisms have been implicated in the homogenization of reiterated sequences - unequal crossing-over and gene conversion (Baltimore, 1981). In one sense, this homogenization is effected by genome plasticity, rather than rigid invariability.

Direct evidence has been obtained for the role of both unequal crossing-over (Tartof, 1974; Szostak and Wu, 1980; Petes, 1980) and gene conversion (Klein and Petes, 1981) in the homogenization of reiterated genes. In the ribosomal genes of *Drosophila* and *Xenopus*, the homogenization has been shown to be extremely rapid. In *Xenopus*, the non-transcribed spacer of the ribosomal gene repeat, while homogeneous within one species (*Xenopus laevis*), is divergent in a very closely related species (*Xenopus borealis*) (Long and Dawid, 1980). A similar invasion of spacer-variant repeats has occurred in the ribosomal cistrons of *Drosophila*, on both the X and Y chromosomes (Dover and Coen, 1981).

Gene conversion has also been implicated in recombination between a single gene duplication (Slightom *et al.*, 1980). Two adjacent β -like human globin genes were found to be more similar to each other in a specific region of the gene, than to the respective allele of one of the genes. A similar situation has recently been recognized in the mouse immunoglobulin constant region genes (Schreier *et al.*, 1981).

In practice, an important distinction must be made between unequal crossing-over and gene conversion. This is that unequal crossing-over is reciprocal in action and will cause fluctuation in the gene number whereas gene conversion is non-reciprocal and maintains the gene dosage. Which mechanism is most active in concerted evolution is not known, however studies on an engineered gene duplication in yeast suggests that gene conversion is by far the most prevalent mechanism (Jackson and Fink, 1981). Nevertheless, the occurrence of unequal crossing-over may be significant. The ribosomal gene number varies from ca200 to 600 in wild type *Xenopus* individuals (D.D. Brown, cited in Hood *et al.*, 1975), and this could be attributed to the reciprocal nature of unequal crossing-over.

9.5.2 Histone Gene Number

The minimum gene dosage in any particular organism is presumably dictated by the demand for expression of that gene in obtaining maximum fitness. Some fluctuation in gene number due to the mechanisms of gene duplication and loss described above, are obviously tolerated in reiterated gene clusters (section 9.5.2). For unique genes, a duplication, if endowing no selective advantage, absolves one of the progeny genes from selective pressure. It has been suggested that pseudo-genes have arisen in this way, being non-functional relics of excessive gene duplication (Proudfoot, 1980).

For the ribosomal genes, gene number in general,

corresponds to genome size (Long and Dawid, 1980). For the histone genes, gene number should depend on both genome size and rate of cell division. Yeast has a small genome (200X smaller than mouse) and contains only two copies of each of the core histone genes (Hereford *et al.*, 1979). Sea urchins, however, contain 300-1000 copies of their histone genes, and *Drosophila* has 100 copies. Both sea urchins and *Drosophila* exhibit extremely rapid cell division during early development (Rabinowitz, 1941; Hinegardner, 1967).

This trend has been somewhat circumvented in *Xenopus*. Although early division is rapid, *Xenopus* only contains 20-50 histone genes (Jacob *et al.*, 1976). Histone gene transcription is a relatively minor factor during embryogenesis, because the developing embryo utilizes histone proteins stored in the oocyte and also activates a pool of maternal mRNAs (Adamson and Woodland, 1974). Another amphibian, the newt, contains 600-800 histone genes, yet has a similar developmental rate to *Xenopus* (Stephenson *et al.*, 1981). However, its genome is 15 times larger than that of *Xenopus*, and so requires 15 times more histone protein for cell division.

As a further trend, histone gene number drops in other vertebrates where the rate of cell division is more moderate (about 12 hours). Chickens have a reiteration frequency of about 10 (Crawford *et al.*, 1979), mouse 10-20 (Jacob, 1976), and Hela Cells 10-20 (Wilson *et al.*, 1974).

It would appear, at least in yeast, that the gene

number reflects exactly the transcriptional requirements. A strain containing an engineered duplication of an H2A/H2B pair shows double the transcriptional rate, but unaffected steady-state mRNA levels (Osley and Hereford, 1981).

9.5.3 Histone Gene Organization and Evolution

Compelling evidence argues that reiterated genes evolve in a concerted fashion as a result of correctional mechanisms such as unequal crossing-over and gene conversion (section 9.5.1). To be sure that the reverse is not true, it must be established that histone genes are independently controlled and do not rely on a conserved reiterated configuration for expression. The fact that histone genes have become dispersed through evolution would support this premise. It also seems highly unlikely that a conserved gene cluster is maintained to allow poly-cistronic transcription. No such transcript has been identified in sea urchins, and in all other systems, transcription is from both DNA strands.

DNA sequencing data has recognized sequences flanking each sea urchin histone gene which are typical of eukaryote promoter sequences (Busslinger *et al.*, 1980). In addition, the sea urchin histone genes have been exploited extensively in *in vivo* transcription studies. All five genes of the sea urchin repeat are faithfully transcribed in *Xenopus* oocytes but with strikingly different efficiencies (Hentschel *et al.*, 1980). Authentic histone mRNAs are also synthesized following oocyte

injection of one or a few individual histone genes. Taken together, it seems likely that each histone gene is under independent control, and that the concerted evolution of histone genes seen in sea urchins and *Drosophila* is a consequence of gene correction mechanisms. In a limited view, such mechanisms can be seen to have evolved as a general correctional function to maintain the homogeneity of highly reiterated genes. In the case of histone genes and ribosomal genes, relative gene stoichiometry would also be maintained - at least a partial prerequisite for coordinate expression.

The evolution of histone genes now follows a predictable trend. As copy number drops, dependent on the organisms requirements for histone synthesis, the frequency of correctional events becomes less until they are no longer sufficient to maintain homogeneity of gene clusters.

In *Xenopus*, the organization of histone genes appears to be in a transitional position. Although some evidence for a conserved repeat has been found (Zernick *et al.*, 1980), the order of this cluster is seen to vary and a substantial degree of dispersal of gene order has occurred (van Dongen *et al.*, 1981). It is significant that in the newt, where reiteration frequency is high, histone genes are again in conserved arrays (Stephenson *et al.*, 1981). That they are embedded in tracts of repetitive satellite would facilitate gene conversion events.

In higher vertebrates (chicken, mouse and human) where gene number is lower than in *Xenopus*, no semblance

of a conserved histone gene cluster is found.

In general, the evolution of histone genes from a highly ordered, to a considerably disordered disposition, is probably a function of the diminishing capacity of correctional mechanisms to maintain gene homogeneity as copy number falls. It is unlikely that the mode of histone gene expression has changed drastically through evolution.

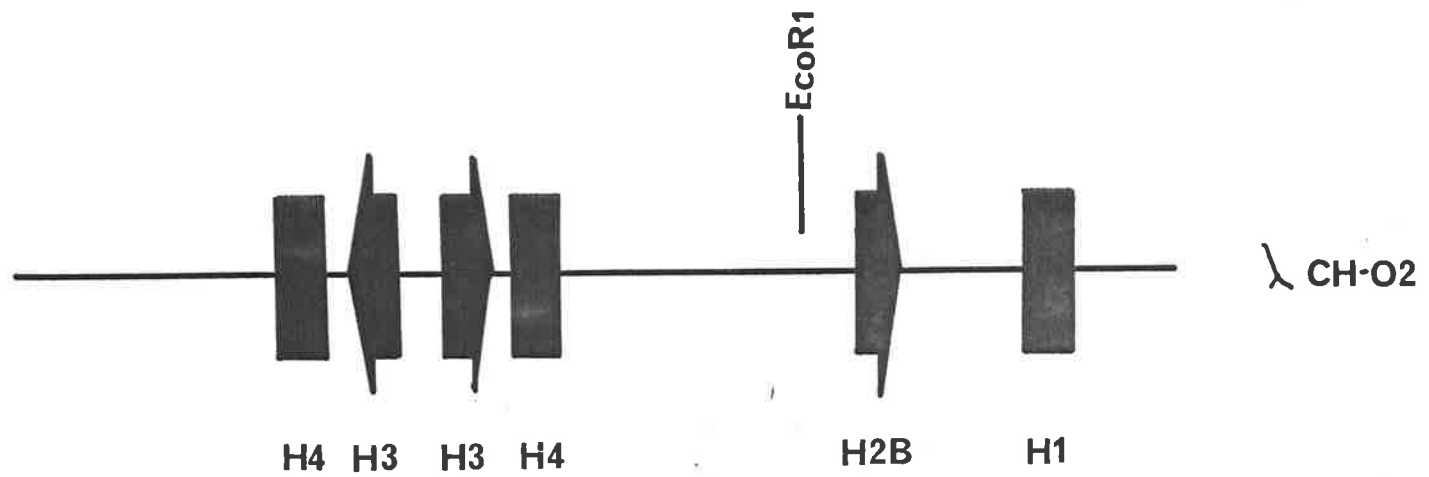
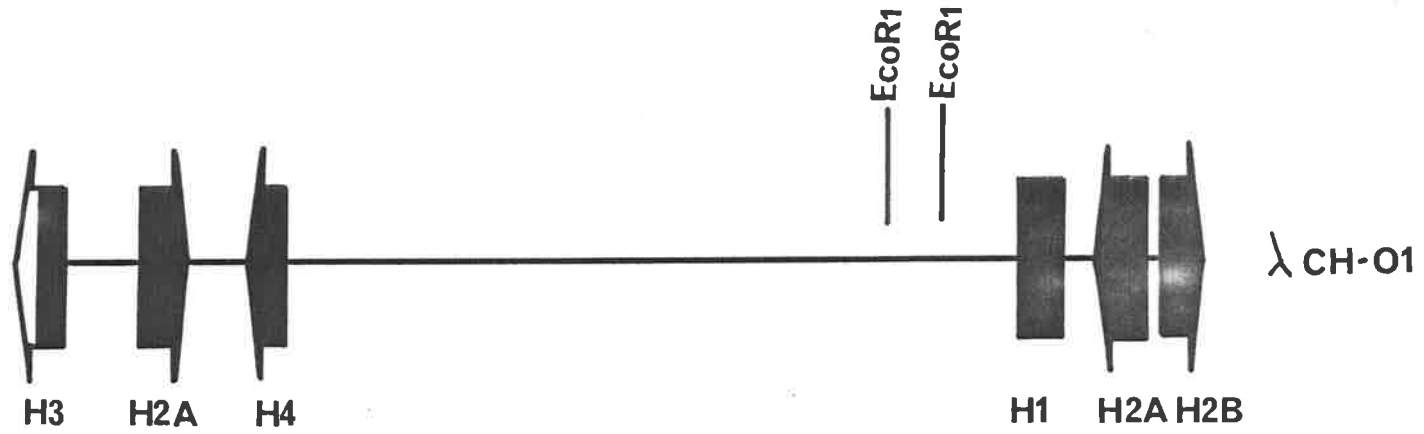
POSTSCRIPT

Since the work described in this thesis has been completed, additions to the gene structure of λ CH-01 and λ CH-02 have been recognised by other workers in this laboratory. Firstly, DNA sequencing data on pCH10E (section 6.2.3) has revealed the presence of H4 genes, one on either side of the divergent H3 gene pair (Fig. 6.5). These H4 genes were not detected with sea urchin H4 probe (section 6.2.2), but were subsequently detected with a homologous H4 probe derived from pCH3.75H/E (Fig. 5.5). It is not known whether the orientation of the H4 genes is also divergent, however, their relationship to the H3 genes strengthens the apparent symmetry in this area and makes the possibility that the region arose from a relatively recent duplication event, more striking (section 6.3).

Secondly, the orientation of the XhoI fragment of pCH3.3E containing the H2A gene (Fig. 5.2) had been incorrectly determined (from restriction data). In compiling the sequences of the XhoI, H2A fragment and the adjacent XhoI/SmaI fragment (section 8.2.3), the XhoI junction was not sequenced through. This should have been done to confirm the orientation assigned from restriction mapping. When it was subsequently sequenced, the XhoI fragment was found to be in the reverse orientation to that indicated in Fig. 8.8. The orientation of that H2A gene within λ CH-01 (Figs. 5.6 and 6.10) is also reversed. A corrected gene organization of λ CH-01 and λ CH-02 is shown in Fig. 10.

FIGURE 10

Additional information on the gene organization of λ CH-01 and λ CH-02 (see POSTSCRIPT).



1KB

The implications of this finding to the general conclusion of "chicken histone gene disorder" are slight. Significantly though, the H2A/H2B gene pair of pCH3.3E is divergently orientated and the sequences in the XhoI/SmaI fragment (Fig. 5.2) are 3' to the H2A gene. The tight linkage of these two genes warrants some discussion.

Divergently orientated, tightly linked gene pairs are becoming a common feature of histone gene systems (Hentschel and Birnstiel, 1981), and are also observed in *Drosophila* heat shock genes and silk moth chorion genes (Jones and Kafatos, 1980). It has been suggested that such an arrangement may facilitate coordinate expression of genes by the existence of shared promotor sequences (Lifton *et al.*, 1977; Jones and Kafatos, 1980), however, there is no evidence for this claim. Silk moth chorion gene pairs, for instance, appear to have their own distinct promotor elements (Jones and Kafatos, 1980).

The H2A/H2B genes of pCH3.3E are very tightly linked indeed. Approximately 135 bps separates the proposed transcription initiation sites ("CAP boxes") of the two genes (chapter 8). Both genes have an easily recognised "TATA box", but only the H2B gene has a true "CAT box" sequence. It is interesting though, that this "CAT box" falls almost midway between the "TATA boxes" of the two divergent genes. No "CAT box" can be recognised for the H2A gene. Given this symmetry in promotor elements, it is possible that the H2B "CAT box" serves both the H2A and H2B genes. It is not necessarily implied however,

that the genes would be transcribed at the same rate (the "CAT box" has a modulator function on gene transcription; Hentschel and Birnstiel, 1981). Deletion or mutation of the "CAT box" region of both a sea urchin H2A gene (Grosschedl and Birnstiel, 1980) and the Herpes Simplex virus thymidine kinase gene (S. McKnight, personal communication) result in an increase in transcription initiation efficiency *in vivo*. More significantly, inversion of the "CAT box" region of the sea urchin H2A gene also results in a stimulated initiation rate (Grosschedl and Birnstiel, 1980). It will be interesting to examine and manipulate the expression of the chicken H2B/H2A gene pair in transcription systems.

H1 Genes:

The position of H1 genes has not been considered in this thesis because no suitable probe was available for their detection. Recent sequencing of the random clones generated from pCH4.6E (section 8.2.5) by A. Robins, revealed an H1 gene sequence. This clone has been used as probe and an H1 gene has also been detected in λ CH-01. The approximate positions of these genes are indicated in Fig. 10. The environments of the two H1 genes are different, as is the case for the core histone genes. This indicates that in the chicken histone gene system, all five histone genes are clustered, but arranged in an apparently disordered fashion.

BIBLIOGRAPHY

- Abelson, J. (1979) *Ann. Rev. Biochem.* 48, 1035-1069.
- Adamson, E.D. and Woodland, H.R. (1974) *J. Mol. Biol.* 88, 263-285.
- Adesnick, M. and Darnell, J.E. (1972) *J. Mol. Biol.* 67, 397-406.
- Alt, F., Kellems, R.E., Bertino, J.R. and Schimke, R.T. (1978) *J. Biol. Chem.* 253, 1357-1370.
- Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) *Nature* 290, 457-464.
- Atkinson, M.R., Deutscher, M.P., Kornberg, A., Russell, A.F. and Moffat, J.G. (1969) *Biochemistry* 8, 4897.
- Axel, R., Feigelson, P. and Schutz, G. (1976) *Cell* 7, 247.
- Ballal, N.R., Kang, Y.J., Olson, M.O.J. and Bursch, H. (1975) *J. Biol. Chem.* 250, 5921-5925.
- Baltimore, D. (1981) *Cell* 24, 592-594.
- Bard, E., Efron, D., Marcus, A. and Perry, R.P. (1974) *Cell* 1, 101.
- Beach, L.R. and Palmiter, R.D. (1981) *Proc. Natl. Acad. Sci. USA.* 78, 2110-2114.
- Benoist, C. and Chambon, P. (1981) *Nature* 290, 304-310.
- Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) *Nuc. Acids Res.* 8, 127-142.
- Benton, W.D. and Davis, R.W. (1977) *Science* 196, 180-182.
- Bernard, O., Hozumi, N. and Tonegawa, S. (1978) *Cell* 15, 1133.
- Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977) *Science* 196, 161-169.
- Bogenhagen, D.F., Sakouju, S. and Brown, D.D. (1980) *Cell* 19, 27-35.
- Bonner, J., Chalkley, G.R., Dahmus, M., Fambrough, D., Fujimura, F., Huang, R.C., Huberman, J., Jensen, R., Marushige, K., Ohlenbusch, H., Olivera, B. and Widholm, J. (1968) *Methods Enz.* 12B, 3.
- Borst, P. and Grivell, L.A. (1981) *Nature* 280, 439-440.
- Borun, T.W., Ajiro, K., Zweidler, A., Dolby, T.W., and Stephens, R.E. (1977) *J. Biol. Chem.* 252, 173-180.
- Brack, C. and Tonegawa, S. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5652-5656.
- Bradbury, E.M., Inglis, R.J. and Matthews, H.R. (1974) *Nature* 247, 257-261.

- Brandt, W.F., Strickland, W.N., Strickland, M., Carlisle, L., Woods, D. and Von Holt, C. (1979) *Eur. J. Biochem.* 94, 1-10.
- Brandt, W.F. and Von Holt, C. (1974) *Eur. J. Biochem.* 46, 419-429.
- Brandt, W.F. and Von Holt, C. (1978) *Biochim. Biophys. Acta* 263, 351-367.
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* 75, 4853-4857.
- Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
- Breathnach, R., Mandel, J.L. and Chambon, P. (1977) *Nature* 270, 314.
- Britten, R.J. and Davidson, E.H. (1969) *Science* 165, 349-357.
- Britten, R.J. and Davidson, E.H. (1971) *Q. Rev. Biol.* 46, 111.
- Brown, D.D. and Blackler, A.W. (1972) *J. Mol. Biol.* 63, 75-83.
- Brown, D.D. and Littna, E. (1966) *J. Mol. Biol.* 20, 95.
- Brown, D.D., Wensink, P.C. and Jordon, E. (1972) *J. molec. Biol.* 63, 57.
- Brown, I.R. and Church, R.B. (1972) *Dev. Biol.* 29, 73.
- Bruschi, S. and Wells, J.R.E. (1981) *Nuc. Acids. Res.* 9, 1591-1597.
- Burckhardt, J. and Birnstiel, M.L. (1978) *J. molec. Biol.* 118, 61-79.
- Busslinger, M., Portmann, R., Irminger, J.C. and Birnstiel, M.L. (1980) *Nuc. Acids Res.* 8, 957-977.
- Cameron, J.R., Loh, E.Y. and Davis, R.W. (1979) *Cell* 16, 739-751.
- Childs, G., Levy, S. and Kedes, L.H. (1979) *Biochem.* 18, 208-213.
- Childs, G., Maxon, R. and Kedes, L. (1979) *Dev. Biol.* 73, 153-173.
- Childs, G., Maxon, R., Cohn, R.H. and Kedes, L. (1981) *Cell* 23, 651-664.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* 24, 5294-5299.
- Clarkson, S., Smith, H., Schaffner, W., Gross, K. and Birnstiel, M. (1976) *Nucleic Acids Res.* 3, 2617-2632.
- Cleary, M.L., Haynes, J.R., Schon, E.A. and Lingrel, J.B. (1980) *Nuc. Acids Res.* 8, 4791-4802.
- Cleary, M.L., Schon, E.A. and Lingrel, J.B. (1981) *Cell* 26, 181-190.

- Cohn, R.H. and Kedes, L.H. (1979a) *Cell* 18, 843-853.
- Cohn, R.H. and Kedes, L.H. (1979b) *Cell* 18, 855-864.
- Compere, S.J. and Palmiter, R.D. (1981) *Cell* 25, 233-240.
- Corces, V., Pellicer, A., Axel, R. and Meselson, M. (1981) *Proc. Natl. Acad. Sci. USA* 78, 7038-7042.
- Cox, R.A. (1968) *Methods Enzymol.* 12B, 120-129.
- Crawford, R.J., Krieg, P., Harvey, R.P., Hewish, D.A. and Wells, J.R.E. (1979) *Nature* 279, 132-136.
- Crawford, R.J., Scott, A.C. and Wells, J.R.E. (1977) *Eur. J. Biochem.* 72, 291.
- Danehold, B. and Hosick, H. (1974) *Cold Spring Harbor Symp. Quant. Biol.* 38, 629.
- Davidson, E.H., Haslett, G.W., Finney, R.J., Allfrey, V.G. and Mirsky, A.E. (1965) *Proc. Natl. Acad. Sci. USA* 54, 696.
- Davidson, E.H. and Hough, B.R. (1971) *J. Mol. Biol.* 56, 491.
- Davidson, E.H., Hough, B.R., Amenson, C.S. and Britten, R.J. (1973) *J. Mol. Biol.* 77, 1.
- Dehardt, D.T., Dressler, D.H. and Ray, D.S. eds. (1978) in "Single-Stranded DNA Phages". Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Denhardt, D. (1966) *Biochem. Biophys. Res. Comm.* 23, 641-646.
- Desrosiers, R.C., Mulder, C. and Fleckenstein, B. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3839-3843.
- Diaz, M.O., Barsacchi, Pilone, G., Mahon, K.A. and Gall, J.G. (1981) *Cell* 24, 649-659.
- Dierks, P., van Ooyen, A., Nantei, N. and Weissmann, C. (1981) *Proc. Natl. Acad. Sci. USA* 78, 1411-1415.
- Dodgson, J.B., Strommer, J. and Engel, J.D. (1979) *Cell* 17, 879-888.
- Donelson, J.E. and Wu, R. (1972) *J. Biol. Chem.* 247, 4661.
- Dover, G. and Coen, E. (1981) *Nature* 290, 731-732.
- Dugaiczyk, A., Boyer, H. and Goodman, H. (1975) *J. Mol. Biol.* 96, 171-184.
- Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) *Cell* 20, 313-319.
- Ehrenberg, L., Fedorcsak, I. and Solymosy, F. (1976). *Prog. Nucl. Acid. Res. Mol. Biol.* 16, 189-262.
- Elgin, S.C.R. and Weintraub, H. (1975) *Ann. Rev. Biochem.* 22, 725-774.
- Engel, J.D. and Dodgson, J.B. (1981) *Proc. Natl. Acad. Sci. USA* 78, 2856-2860.
- Engel, J.D., Sugarman, B.J. and Dodgson, J.B. (1982) *Proc. Natl. Acad. Sci. USA*. In press.

- Faye, G., Leuna, D.W., Tatchell, K., Hall, B.D. and Smith, M. (1981) Proc. Natl. Acad. Sci. USA 78, 2258-2268.
- Finnegan, D.J., Robin, G.M., Young, M.W. and Hogness, D.S. (1978) Cold Spring Harbor Symp. Quant. Biol. 42, 1053-1063.
- Flynn, J.M. and Woodland, H.R. (1980) Dev. Biol. 75, 222-230.
- Freigan, N., Marchionni, M. and Kedes, L.H. (1976) Bio. Bull. 151, 407 (Abstract).
- Gall, J.G. (1968) Proc. Natl. Acad. Sci. USA 60, 553.
- Gallwitz, D. (1975) Nature 259, 242.
- Gallwitz, D. and Mueller, G.C. (1969) J. Biol. Chem. 244, 5947-5952.
- Gallwitz, D. and Sures, I. (1980) Proc. Natl. Acad. Sci. 77, 2546-2550.
- Gambino, R., Metafera, S., Felicetti, L. and Raisman, J. (1973) Biochim. Biophys. Acta. 312, 377-391.
- Gannon, F., O'Hare, K., Perrin, F., Le Penne, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) Nature 278, 428-434.
- Gardner, R.L. (1971) Adv. Biosci. 6, 279.
- Gingle, A.R. (1977) Dev. Biol. 58, 394-401.
- Glover, C.L.C. and Gorovsky, M.A. (1978) Biochemistry 17, 5705-5713.
- Glover, D.M. and Hogness, D.S. (1977) Cell 10, 167-176.
- Goeddel, D.V., Leung, D.W., Dull, T.J., Gross, M., Lawn, R.M., McCandliss, R., Seeburg, P.H., Ullrich, A., Yelverton, E. and Gray, P.W. (1981) Nature 290, 20-26.
- Goldberg, M.L. (1979) Ph.D. Thesis, Stanford University, Stanford, California.
- Goldknopf, I.L. and Busch, H. (1977) Proc. Natl. Acad. Sci. USA 74, 864-868.
- Goldstein, G., Scheid, H., Hammerling, U., Boyse, E.A., Schlesinger, D.H. and Niall, H.D. (1975) Proc. Natl. Acad. Sci. USA 72, 11-15.
- Groppi, V.C. and Coffino, P. (1980) Cell 21, 195-204.
- Gross, K., Schaffner, W., Telford, J. and Birnstiel, M. (1976) Cell 8, 479-484.
- Gross, P.R., Gross, K.W., Skoultchi, A.I. and Rudderman, J.V. (1973) In "Protein Synthesis in Reproductive Tissue, Karolinska Symp." 6, 244-262. Stockholm, Karolinska Institutet.
- Grossbach, V. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 619.
- Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 1432-1436.

- Grosschedl, R., Wasylyk, B., Chambon, P. and Birnstiel, M.L. (1981) *Nature* 294, 178-180.
- Groudine, M., Eisenman, R. and Weintraub, H. (1981) *Nature* 292, 311-317.
- Groudine, M. and Weintraub, H. (1975) *Proc. Natl. Acad. Sci. USA* 72, 4464.
- Grouse, L., Chilton, M.D. and McCarthy, B.J. (1972) *Biochemistry* 11, 798.
- Grunstein, M. (1978) *Proc. Natl. Acad. Sci. USA* 75, 4135-4139.
- Gruss, P., Dhar, R. and Khoury, G. (1981) *Proc. Natl. Acad. Sci. USA* 78, 943-947.
- Guarente, L. and Ptashne, M. (1981) *Proc. Natl. Acad. Sci. USA* 78, 2199-2203.
- Gurdon, J.B. (1962) *Dev. Biol.* 4, 256.
- Gurmstbein, M. and Schedl, P.J. (1976) *J. molec. Biol.* 104, 325-349.
- Hackett, P.B., Traub, P. and Gallwitz, D. (1978) *J. Mol. Biol.* 126, 649.
- Hadorn, E. (1965) *Brookhaven Symp. Biol.* 18, 148-161.
- Hagenbüchle, O., Tosi, M., Schibler, U., Bovey, R., Wellaver, P.K. and Young, R.A. (1981) *Nature* 289, 643-646.
- Hahn, W.E. and Laird, C.D. (1971) *Science* 173, 158.
- Harding, J.D., MacDonald, R.J., Przybyla, A.E., Chirgwin, J.M., Pictet, R.L. and Rutler, W.J. (1977) *J. Biol. Chem.* 252, 7391-7397.
- Harris, S.E., Rosen, J.M., Means, A.R. and O'Malley, B.W. (1975) *Biochemistry* 14, 2072.
- Harrison, P.R., Birnie, G.D., Hell, A., Humphries, S., Young, B.D. and Paul, J. (1974) *J. Mol. Biol.* 84, 539.
- Heintz, N., Zernik, M. and Roeder, R.G. (1981) *Cell* 24, 661-668.
- Hensketh, T.R., Smith, G.A., Houslay, M.D., Warren, G.B. and Metcalfe, J.C. (1977) *Nature* 267, 490-494.
- Hentschel, C.C., Irminger, J.C., Bucher, P. and Birnstiel, M.L. (1980) *Nature* 285, 147-151.
- Hentschel, C.C., Probst, E. and Birnstiel, M.L. (1980) *Nature* 288, 100-102.
- Hereford, L.M., Fahrner, K., Woodford, J. and Rosbash, M. (1979) *Cell* 18, 1261-1271.
- Hille, M.B. (1974) *Nature* 249, 556-558.
- Hillman, N., Sherman, M.I. and Graham, C. (1972) *J. Embryol. Exp. Morphol.* 28, 263.
- Hinegardner, R.T. (1967) in *Methods in Developmental Biology*, eds. Wilt, F.H. and Wessells, N.K. (Crowell-Collier, New York), 139-155.

- Hirsh, D. and Vanderslice, R. (1976) *Dev. Biol.* 49, 220-235.
- Holliday, R. and Pugh, J.E. (1975) *Science* 187, 226-232.
- Holmes, D.S. and Bonner, J. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1108.
- Hood, L., Campbell, J.H. and Elgin, S.C.R. (1975) *Ann. Rev. Genet.* 9, 305-353.
- Hourcade, D., Dressler, D. and Wolfson, J. (1974) *Cold Spring Harbor Symp. Quant. Biol.* 38, 537.
- Hozumi, N. and Tonegawa, S. (1976) *Proc. Natl. Acad. Sci. USA* 73, 3628-3632.
- Hunt, J.A. (1974) *Biochem. J.* 138, 499.
- Ingram, V.M. (1974) *Annals of New York Academy of Science* 241, 93-98.
- Iwai, K., Hayashi, H. and Ishikawa, K. (1972) *J. Biochem. (Tokyo)* 72, 357-367.
- Jacob, E. (1976) *Eur. J. Biochem.* 65, 275-282.
- Jacob, E., Malacinski, G. and Birnstiel, M.L. (1976). *Eur. J. Biochem.* 69, 45-54.
- Jacob, F. and Monod, J. (1961) *J. Mol. Biol.* 3, 318-356.
- Jeffreys, A.J. and Flavel, R.A. (1977) *Cell* 12, 1097-1108.
- Jencks, W.P. (1969) in *Catalysis in Chemistry and Enzymology*, McGraw Hill, N.Y., Chapter 7.
- Jones, C.W. and Kafatos, F.C. (1980) *Cell* 22, 855-867.
- Kandler-Singer, I. and Kalthoff, K. (1976) *Proc. Natl. Acad. Sci. USA* 73, 3739-3743.
- Kedes, L.H. (1979) *Ann. Rev. Biochem.* 48, 837-870.
- Kedes, L.H., Chang, A.C.Y., Housman, D. and Cohen, S.N. (1975) *Nature* 255, 533-538.
- Kedes, L.H. and Gross, P.R. (1969) *J. Mol. Biol.* 42, 559-575.
- Kirby, K.S. (1956) *Biochem. J.* 3, 687-691.
- Klein, H.L. and Petes, T.D. (1981) *Nature* 289, 144-148.
- Kleinschmidt, A.M. and Martinson, H.G. (1981) *Nuc. Acids Res.* 9, 2423-2431.
- Klenow, H., Overgaard-Hanesen, K. and Patkar, S.A. (1971) *Eur. J. Biochem.* 22, 371.
- Kornberg, R.D. (1974) *Science* 184, 868-871.
- Koster, J.G., Kasinsky, H.E. and Destree, O.H.J. (1979) *Cell Diff.* 8, 93-104.
- Kunkel, N.S., Hemminiki, K. and Weinberg, E.S. (1978) *Biochemistry* 17, 2591-2598.
- Kunkel, N.S. and Weinberg, E.S. (1978) *Cell* 14, 313-326.
- Kurtz, D. (1981) *Nature* 291, 629-631.

- Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977) Cold Spring Harbor Symposia on Quantitative Biology 42, 1047-1052.
- Lawn, R.M., Gross, M., Hocuk, C.M., Franke, A.E., Gray, P.V. and Goeddel, D.V. (1981) Proc. Natl. Acad. Sci. USA 78, 5435-5439.
- Laskey, L.A. and Gurdon, J.B. (1970) Nature 228, 1332.
- Levenson, R. and Housman, D. (1980) Cell 25, 5-6.
- Levenson, R., Housman, D. and Cantley, L. (1980) Proc. Natl. Acad. Sci. USA 77, 5948-4952.
- Levinger, L. and Varshausky, A. (1982) Cell 28, 375-385.
- Levy, C.C., Schmukler, M., Frank, J., Karpetsky, T., Jewett, B., Hieter, P., Le Gendre, S. and Dorr, R. (1975) Nature 256, 340.
- Lewis, E.B. (1978) Nature 276, 565-570.
- Lilley, D.M.J. and Pardon, J.F.A. (1979) A. Rev. Genet. 13, 197-233.
- Lin, T.P. (1969) Nature 222, 480.
- Lodish, H.F. (1976) Ann. Rev. Biochem. 45, 39-72.
- Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979) Cell 18, 545-558.
- Long, E.O. and Dawid, I.B. (1980) Ann. Rev. Biochem. 49, 727-764.
- Mandel, J.L. and Chambon, P. (1979) Nuc. Acids Res. 7, 2081-2103.
- Maniatis, T., Fritsch, E.F., Laver, J. and Lawn, R.M. (1980) Ann. Rev. Genet. 14, 145-178.
- Maniatis, T., Hardison, R.C., Lacy, E., Laver, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell 15, 687-701.
- Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) Proc. Natl. Acad. Sci. USA 72, 1184-1188.
- Marian, B. and Wintersberger, U. (1980) PEBS Lett. 117, 63-67.
- Max, E., Seidman, J. and Leder, P. (1979) Proc. Natl. Acad. Sci. USA 76, 3450.
- Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
- Maxam, A.M. and Gilbert, W. (1980) Methods Enzym. 65, 499-560.
- McClements, W. and Skalka, A.M. (1977) Science 196, 195-197.
- McClintock, B. (1956) Cold Spring Harbor Symp. Quant. Biol. 21, 197-216.
- McGhee, J.D. and Felsenfeld, G. (1980) Ann. Rev. Biochem. 49, 1115-1156.

- McKenzie, S.L., Henikoff, S. and Meselson, M. (1975) Proc. Natl. Acad. Sci. USA 72, 1117.
- McKnight, S.L., Gavis, E.R. and Kingsbury, R. (1981) Cell 25, 385-398.
- Melli, M., Spinelli, G., Wyssling, H. and Arnold, E. (1977) Cell 11, 651-661.
- Milcarek, C., Price, R. and Penman, S. (1974) Cell 3, 1.
- Milmer, J. (1978) Nature 275, 660-661.
- Mintz, B. (1970) Symposia of the International Society for Cell Biology 9, 15.
- Moll, R. and Winterberger, E. (1976) Proc. Natl. Acad. Sci. USA 73, 1863-1867.
- Moorman, A.F.M., de Laaf, R.T.M., Destree, O.H.J., Telford, J. and Birnstiel, M.L. (1980) Gene 10, 185-193.
- Nagaraja Rao, R. and Rogers, S.G. (1979) Gene 7, 79-82.
- Nemer, M., Graham, M. and Dubroff, L.M. (1974) J. Mol. Biol. 89, 435.
- Nigau, V., Guerrier, P. and Monin, H. (1960) Bull. Biol. Fr. Belg. 94, 131-202.
- Noll, M. and Kornberg, R.D. (1977) J. Mol. Biol. 109, 393-404.
- Nunberg, J.H., Kaufman, R.J., Chang, A.C.Y., Cohn, S.N. and Schimke, R.T. (1980) Cell 19, 355-364.
- Ochao, S. and de Haro, C. (1979) Ann. Rev. Biochem. 48, 549-580.
- Ohkubo, H., Vogeli, G., Mudryj, M., Avvedimento, V.E., Sullivan, M., Pastan, I. and de Crombrughe, B. (1980) Proc. Natl. Acad. Sci. USA 77, 7059-7063.
- Ohno, S. and Taniguchi, T. (1981) Proc. Natl. Acad. Sci. USA 78, 5305-5309.
- O'Malley, B.W., Roop, D.R., Lai, E.C., Nordstrom, J.L., Catterall, J.F., Swaneck, G.E., Colbert, D.A., Tsai, M.J., Dugaiczky, A. and Woo, S.L.C. (1978) Presented at: Laurentian Hormone Conference, August 1978.
- O'Malley, B.W., Schwartz, R.J. and Schrader, W.T. (1976) Journal of Steroid Biochemistry 7, 1151-1159.
- Osley, M.A. and Hereford, L.M. (1981) Cell 24, 377-384.
- Parish, R.W. (1979) Biochim. Biophys. Acta. 553, 179-182.
- Pelling, C. (1964) Chromosoma 15, 71.
- Perry, R.P. and Kelley, D.E. (1973) J. Mol. Biol. 71, 681-696.
- Portmann, R., Schaffner, W. and Birnstiel, M. (1976) Nature 264, 31-34.
- Prescott, D.M. (1966) J. Cell Biol. 31, 1-9.
- Proudfoot, N.J. (1980) Nature 286, 840.

- Rabinowtiz, M. (1941) *J. Morphol.* 69, 1-21.
- Revel, M. and Groner, Y. (1978) *Ann. Rev. Biochem.* 47, 1079-1126.
- Rhoads, R.E., McKnight, G.S. and Schimke, R.T. (1973) *J. Biol. Chem.* 248, 178.
- Richards, R.I., Shine, J., Ullrich, A., Wells, J.R.E. and Goodman, H.M. (1979) *Nuc. Acids Res.* 7, 1137-1142.
- Richards, R.I. and Wells, J.R.E. (1980) *J. Biol. Chem.* 255, 9306-9311.
- Rickles, R., Marashi, F., Sierra, F., Wells, J., Stein, J. and Stein, G. (1982) *Proc. Natl. Acad. Sci. USA* 79, 749-753.
- Riggs, A.D. (1975) *Cytogen. Cell Genet.* 14, 9-25.
- Risley, M.S. and Eckhardt, R.A. (1981) *Dev. Biol.* 84, 79-87.
- Ritossa, F.M., Atwood, K.C., Lindsley, D.L. and Spiegelman, S. (1966) *Natl. Cancer Inst. Monogr.* 23, 449.
- Robbins, E. and Borun, T.W. (1967) *Proc. Natl. Acad. Sci. USA* 57, 409-416.
- Robertson, A.D.J. and Grutsch, J.F. (1981) *Cell* 24, 603-611.
- Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L. and Wall, R. (1980) *Cell* 20, 303-312.
- Ruderman, J.V. and Pardue, M.L. (1978) *J. Biol. Chem.* 253, 2018-2025.
- Rykowski, M.C., Wallis, J.W., Choe, J.C. and Grunstein, M. (1981) *Cell* 25, 477-487.
- Saint, R.B. (1979) Ph.D. Thesis; University of Adelaide.
- Sakana, H., Rogers, J.H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. and Tonegawa, S. (1979) *Nature* 277, 627.
- Samal, B. and Worcel, A. (1981) *Cell* 23, 401-409.
- Sanders, M.M. (1981) *J. Cell Biol.* 91, 579-583.
- Sanger, F. and Coulson, A.R. (1975) *J. Mol. Biol.* 94, 441.
- Sanger, F. and Coulson, A.R. (1978) *FEBS Lett.* 87, 107.
- Sanger, F., Coulson, A.R., Barrel, B.G., Smith, A.J.H. and Roe, B.A. (1980) *J. Mol. Biol.* 143, 161-178.
- Sanger, F., Coulson, A.R., Friedmann, T., Air, G.M., Barrell, B.G., Brown, N.L., Fiddes, J.C., Hutchinson III, C.A., Slocombe, P.M. and Smith, M. (1978) *J. mol. Biol.* 125, 225.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
- Schaffner, W., Gross, K., Telford, J. and Birnstiel, M. (1976) *Cell* 8, 471-478.

- Schmid, C.W. and Deninger, P.L. (1975) *Cell* 6, 345.
- Schmidt, G. (1957) *Methods Enzymol.* 3, 687-691.
- Schreier, P.H., Bothwell, A.L.M., Mueller-Hill, B. and Baltimore, D. (1981) *Proc. Natl. Acad. Sci. USA*
- Schreier, P.H. and Cortese, R. (1979) *J. Mol. Biol.* 129, 169-172.
- Sciler-Tuyns, A. and Birnstiel, M.L. (1981) *J. Mol. Biol.* 151, 607-625.
- Seeberg, P.H., Shine, J., Martial, J., Baxter, J.D. and Goodman, H.M. (1977) *Cell* 12, 157-165.
- Sierra, F., Leza, A., Marashi, F., Plumb, M., Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G. and Stein, J. (1982b) In press.
- Sierra, F., Lichtler, F., Marashi, R., Rickles, T., Van Dyke, T., Clark, S., Wells, J., Stein, G. and Stein, J. (1982a) In press.
- Sittman, D.B., Chiu, I., Pan, C., Cohn, R.H., Kedes, L.H. and Marzluff, W.F. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4078-4082.
- Slightom, J.L., Blechl, A.E. and Smithies, O. (1980) *Cell* 21, 627-638.
- Smith, L.D. (1966) *Dev. Biol.* 14, 330.
- Southern, E.M. (1975) *J. mol. Biol.* 98, 503-517.
- Spinelli, G., Gianguzza, F., Casano, C., Acierno, P. and Burckhardt, J. (1979) *Nucl. Acids Res.* 6, 454-560.
- Stephenson, E.C., Erba, H.P. and Gall, J.G. (1981) *Cell* 24, 639-647.
- Stern, M.S. (1973) *Differentiation* 1, 407.
- Stone, L.S. (1950) *Anat. Rec.* 106, 89.
- Strickland, M., Strickland, W.N., Brandt, W.F., Von Holt, C., Lehmann, A. and Wittmann-Liebhold, B. (1978) *Eur. J. Biochem.* 89, 443-452.
- Struhl, G. (1981) *Nature* 293, 36-41.
- Sullivan, D., Palacios, R., Staunezer, J., Taylor, J.M., Faras, A.J., Kiely, M.L., Summers, N.M., Bishop, J.M. and Schimke, R.T. (1973) *J. Biol. Chem.* 248, 7530.
- Sures, I., Lowry, J. and Kedes, L.H. (1978) *Cell* 15, 1033-1044.
- Szostak, J.W. and Wu, R. (1980) *Nature* 284, 426-431.
- Tartof, K.D. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1272-1276.
- Tashima, M., Calabretta, B., Torelli, G., Scofield, M., Maizel, A. and Saunders, G.F. (1981) *Proc. Natl. Acad. Sci. USA* 78, 1508-1512.
- Taylor, J.M. (1979) *Ann. Rev. Biochem.* 48, 681-717.

- Taylor, J.M., Illmensee, R. and Summers, J. (1976)
Biochim. Biophys. Acta 442, 324.
- Urban, M.K., Franklin, S.G. and Zweidler, A. (1979)
Biochem. 18, 3952-3960.
- van Dongen, W., de Laaf, L., Zaal, R., Moorman, A. and
Destree, O. (1981) Nuc. Acids Res. 9, 2297-2311.
- Vanin, E.F., Goldberg, G.I., Tucker, P.W. and Smithies, O.
(1980) Nature 286, 222-226.
- Vedeckis, W.V., Schrader, W.T. and O'Malley, B.W. (1978)
in "Biochemical Action of Hormones" ed. J. Litwack,
Academic Press, New York, 5, 321.
- Vidali, G., Boffa, L.C., Bradbury, E.M. and Allfrey, V.G.
(1978) Proc. Natl. Acad. Sci. USA 75, 2239-2243.
- Wahl, G.M., Padgett, R.A. and Stark, G.R. (1979) J. Biol.
Chem. 254, 8679-8689.
- Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Natl.
Acad. Sci. 76, 3683-3687.
- Wasylyk, B., Keding, C., Corden, J., Brison, O. and
Chambon, P. (1980) Nature 285, 267-373.
- Weinstein, I.B. and Wigler, M. (1977) Nature 270, 659-660.
- Weintraub, H., Larsen, A. and Groudine, M. (1981) Cell 24,
333-344.
- Weisbrod, S., Groudine, M. and Weintraub, H. (1980) Cell
19, 289-391.
- West, M.H.P. and Bonner, W.M. (1980) Biochemistry 19,
3238-3245.
- White, R.L. and Hogness, D.S. (1977) Cell 10, 177-192.
- Williamson, R., Crossley, J. and Humphries, S. (1974)
Biochemistry 13, 703.
- Wilson, E.B. (1925) in "The Cell in Development and
Heredity", 3rd edition, Macmillan, New York.
- Wilson, M.C. and Melli, K.J. (1977) J. molec. Biol. 110,
511-535.
- Wilson, M.C., Melli, M. and Birnsteil, M.L. (1974)
Biochem. Biophys. Res. Commun. 61, 404-410.
- Wozney, J., Hanahan, D., Morimoto, R., Boedtker, H. and
Doty, P. (1981) Proc. Natl. Acad. Sci. USA 78,
712-716.
- Wu, R.S. and Bonner, W.M. (1981) Cell 27, 321-330.
- Yajima, H. (1960) J. Embryol. Exp. Morphol. 8, 198.
- Zachav, H.G. and Igo-Kemenes, T. (1981) Cell 24, 597-598.
- Zernik, M., Heintz, N., Boime, I. and Roeder, R.G. (1980)
Cell 22, 807-815.
- Zlatanowa, J. and Swetly, P. (1978) Nature 276, 276-277.