CHICKEN AND HUMAN HISTONE GENES

A thesis submitted to the

University of Adelaide

for the degree of Doctor of Philosophy

by

Susan Joy Clark, B.Sc. (Hons.) A.N.U.

Department of Biochemistry,
University of Adelaide,
South Australia.

November, 1982

SUMMARY

The work presented in this thesis concerns elucidation of the organisation and structure of chicken and human histone genes.

1. A genomal clone, λCH.02, containing chicken histone genes was characterised by restriction enzyme and hybridisation analysis using both sea urchin and chicken histone gene probes. This clone contains two H3 genes, two H4 genes, and H2B and an H1 gene within its 14.4 kb insert; the H3 and H4 genes are organised in adjacent pairs with the two H3 genes being divergently transcribed and separated by 900 base pairs. A spacer region of 3.5 kb separates the H3/H4 gene pairs from the H2B and H1 genes. This arrangement of genes does not conform to a simple repeating unit and transcription is from both DNA strands.

2. Sequence data was obtained from regions of both H3 genes, the H4, H2B and H1 genes and these partial sequences allowed for their positive identification. These genes appear to be independently controlled, since each 5' untranslated region sequenced, contained putative polymerase II regulatory sequences. In addition, the 3' untranslated regions from these genes contained the conserved sequence of dyad symmetry common to all histone genes.

3. A genomal library was constructed from human placental DNA by ligating EcoR1* digested DNA fragments, 15-20 kb in length, into the EcoR1 site of Charon 4A phage vector.

4. Both the EcoR1* library and the Maniatis human library were screened using chicken histone DNA as probes, and a genomal clone from each containing human histone genes was isolated (λHH.02 and λHH.03) and characterised by restriction enzyme and hybridisation analysis.

5. λHH.02 contains a 14.8 kb DNA insert in which two H3 genes, an H2A and an H2B gene were located. The H2A, H2B and H3 genes are adjacent, and are separated by 7.2 kb of spacer from the other H3 gene.

λHH.03 contains an H2A/H2B pair of genes, separated by 3.7 kb from another H2B gene, within its 18.6 kb DNA insert.

6. DNA sequences were obtained from regions of the H2B and H2A genes in λHH.02 as well as from the H2B gene in λHH.03. Both H2B genes code for variant histone proteins not yet characterised, and the H2A gene predicted an amino acid sequence which represents the first region of protein sequence available for a human H2A histone.

Comparison of the human H2B and H2A nucleotide sequences with those of chicken has demonstrated a low frequency of base substitution, leading to amino acid changes, reflecting the conserved nature of the histone proteins. In contrast, a higher rate of base substitutions occurs at silent sites, reflecting more the evolutionary distance between chicken and human genes.

7. It was estimated by genomal blots that there are 16-18 copies of each of the core histone genes in the human genome. The human genomal blots, in conjunction with the gene organisation in the human histone

clones, indicates that the human histone genes are not organised in a simple repeating unit but appear to be arranged in a large, disperse cluster. This type of arrangement is similar to that found in chicken and also appears to be a consistant feature of higher vertebrate histone gene organisation.

CONTENTS

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.

Susan Joy Clark

ACKNOWLEDGEMENTS

My sincere thanks go to all members of the Biochemistry Department, for not only creating a stimulating atmosphere of research but also for their friendship and support throughout my stay in the Department. I would like to thank Professor W.H. Elliott, for permission to work in the Department of Biochemistry, University of Adelaide.

In particular, I would like to extend my gratitude and thanks to the following people who played a part in the production of this thesis:

Dr. Julian Wells for his supervision and enthusiasm.
Dr. Peter Molloy for his assistance, advice and critical discussions.
Members of my laboratory (David Bird, Leeanne Coles, Richard D'Andrea, Paul Krieg, Alan Robins, Richard Sturm and Jenny Whiting) for their many discussions, help and humour, and Jan Dinan for keeping the lab in running order.
Molecular and Cellular Biology Unit, C.S.I.R.O., Sydney for permission to use their facilities.
Jenny Brazier for her excellent assistance in preparing the diagrams.
Judy Geesing for her enthusiasm and skill in typing this thesis.
Finally, I would like to express my love and thanks to Peter, my husband, for his never ending support and encouragement and to my parents for their confidence and patience.

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

ABBREVIATIONS

A:	adenine
$A_{\underline{n}}$:	optical absorbance measured at wave length n over a pathlength of 1 cm
Ap:	Ampicillin
bp:	base pairs, kb: kilobase pairs
C:	cytosine
cDNA:	DNA complementary to RNA
Ci:	Curie
Cp:	Chloramphenicol
dATP:	adenine deoxyribose-5'-triphosphate
dCTP:	cytosine deoxyribose-5'-triphosphate
dGTP:	guanine deoxyribose-5'-triphosphate
dTTP:	thymine deoxyribose-5'-triphosphate
DNAse:	deoxyribonuclease I
EDTA:	ethylenediaminetetracetic acid
G:	guanine
mRNA:	messenger RNA
PEG:	polyethyleneglycol 6000
p.s.i.:	pounds per square inch
RNAse:	ribonuclease A
rRNA:	ribosomal RNA
S:	sedimentation coefficient
SAM:	S-adenosyl methionine
SDS:	sodium dodecyl sulphate
T:	thymine
tRNA:	transfer RNA
Tet:	tetracycline
TEMED:	N,N,N',N'-tetramethylethylenediamine
Tris:	Tris(hydroxymethyl) aminomethane
U.V.:	ultraviolet

CHAPTER 1

INTRODUCTION

INTRODUCTION

Work presented in this thesis focusses on one aspect of eukaryotic gene expression - the control and co-ordination of a multigene family. In particular it concerns the organisation and structure of chicken and human histone genes and their comparison with other histone multigene families. An important part of this analysis is the consideration of features of histone gene organisation which have been conserved through evolution and thus may be important in controlling the co-ordinate expression of this gene family. The introduction to this thesis considers the control of eukaryotic gene expression in a wider context, and detailed information about histone proteins and the genes coding for them is presented.

How gene expression is programmed to control the development of a single cell into a multicellular organism composed of many cell types, each expressing different specialised functions, is one of the fundamental questions of eukaryotic biology. There are a wide variety of genetic mechanisms by which the expression of eukaryotic genes may be regulated. These mechanisms will be considered in three convenient, if somewhat arbitrary sections.

Firstly, genome alteration, where gene expression may be effected by gene loss, amplification or DNA rearrangement.

Secondly, modulation of expression of individual genes, which may occur as transcriptional, post-transcriptional, or translational control.

Thirdly, how the organisation of genes, such as in multigene families, may influence control of their expression.

1.1 GENOME ALTERATION IN GENE EXPRESSION

1.1.1 Gene Loss

Somatic cells of some animals are altered extensively during development by elimination of certain chromosomes (Wilson, 1928). This event occurs in early cleavage in progenitors of somatic cells so that only the germ cells maintain their genome intact. Chromosome diminution has been observed in some nematodes, protozoa, crustaceans and insects, but not in other eukaryotes.

In general however, the DNA composition of all cells of an organism is identical. The best evidence for this is that *Xenopus* tadpoles can be produced from enucleated eggs injected with a nucleus derived from a differentiated tadpole intestinal cell (Gurdon, 1962). Experiments yielding similar conclusions have since been performed with *Drosophila*, (Illmensee, 1972; Okada *et al.*, 1974), and mouse, (Mintz and Illmensee, 1975). Therefore gene loss is now thought by most investigators not to be a mechanism of general significance for gene control in higher eukaryotes.

1.1.2 Gene Amplification

Gene amplification is one means by which a cell can produce large quantities of a specific gene product. Oocytes of some vertebrates and insects specifically amplify their genes for 18s and 28s ribosomal RNA (rRNA) (Brown and Dawid, 1968; Gall, 1968). These genes are present as extra chromosomal circular elements and comprise about half the total DNA of the cell at that stage. This increase in genes supports extremely active rRNA synthesis in oocytes, a synthetic rate so high that one oocyte of *Xenopus*

synthesises several thousand times more ribosomes per unit time than in a single somatic cell. The amplified rRNA genes only function during oogenesis: after meiosis these genes become degraded. When rRNA synthesis begins in embryogenesis, it is directed by the normal complement of rDNA in the chromosomes. Differential replication of the rRNA genes relative to other regions of the genome is also seen in the acellular slime mold (Zellweger, 1972), the tetrahymena macronucleus (Yao and Gall, 1977) and in larval cells of *Drosophila* during polytenisation (Endow and Gall, 1975).

Until recently, rRNA genes were the only genes that had been demonstrated to be amplified or differentially replicated at a specific time in the development of a cell. However, Spradling and Mahowald (1980), have found that genes for chorion proteins in *Drosophila* are amplified in ovarian follicle cells before their active expression in those cells. These are the first genes coding for proteins shown to be amplified as a normal process of development.

Cells in culture can be forced to amplify genes for certain proteins by selection techniques. For example, cells cultured with methotrexate, an inhibitor of dihydrofolate reductase, gradually become resistant to the drug by increasing the amounts of the dihydrofolate reductase gene (Schimke *et al.*, 1977).

Similarly the amplication of the CAD gene occurs in Syrian hamster cells when cultivated with PALA(N-(phosphonacetyl)-L-aspartate), a specific inhibitor of the aspartate transcarbamylase activity of the multifunctional protein CAD (Wahl *et al.*, 1982).

Analogous examples are well known for genes for antibiotic resistance in micro-roganisms (Davies and Rownd, 1972).

Thus, while it is clear that the mechanism of gene amplification exists in many higher eukaryotes, there is no evidence that "forced" gene amplification plays a role in any normal developmental process.

1.1.3 Nondirected Gene Rearrangement

Evidence for movable elements in eukaryotic chromosomes has been available for many years. McClintock, (1967) working with maize, showed that drastic alterations of chromosome organisation, including inversions, deletions and duplications, occurred at sites containing "controlling elements" and that the controlling elements were capable of transposition to new sites. While this particular system has not been characterised in molecular terms, evidence has recently been obtained for the presence of transposable elements in the DNA of *Drosophila melanogaster*, (Finnegan *et al.*, 1977) and yeast,(Cameron *et al.*, 1979). The "copia" sequences of *D. melanogaster* and the "Ty1" sequences of yeast are both moderately repeated, about 35 copies of each being dispersed throughout their respective genomes, and are terminated with short direct repeats, making their overall

structures extremely similar to those of some bacterial transposons, (Ptashne and Cohen, 1975). Both the copia and Tyl sequences are transcribed and represent a significant proportion of the non-ribosomal RNA present in a cell. Copia sequences, for example, can comprise up to 3 per cent of the poly-A-containing RNA from a number of *D. melanogaster* tissues. It has been suggested, (Cameron *et al.*, 1979) that the Tyl sequence may have an effect on the pattern of gene expression in yeast, and because it is movable, it may have the ability to alter this pattern.

However, the movement of these DNA elements has not been correlated with development nor with specific chromosomal locations. It appears that their influence on a neighbouring gene may be a fortuitous event caused by a chance rearrangement of the middle repetitive elements next to the gene, rather than an important means of controlling related genes in a coordinate fashion (Britten and Davidson, 1969).

1.1.4 Directed Gene Arrangement

The structural genes for immunoglobulins in specialised cells of the immune system provide the best known example of a precise genetic rearrangement programmed into the development of a cell for an important biological purpose.

Hozumi and Tonegawa, (1976), using restriction and hybridisation 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 synthesising immunoglobulin

as it is in the germline DNA. These observations were verified by cloning and sequencing experiments, (Bernard *et al.*, 1978), and have since been extended to the DNA coding for the 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 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 germline tissue, but close together in DNA from tissues synthesising 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 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 antibody-producing cell.

Rearrangement of DNA is also responsible for switching of mating types in yeast (Hicks *et al.*, 1979) and with change in surface antigens exhibited by trypanosomes (Williams *et al.*, 1979) and perhaps will explain the extraordinary variety of surface antigens of carcinogen-induced tumors in mice (Baldwin, 1970).

However, gene rearrangement does not appear to be a general mechanism responsible for controlling expression of specific structural genes. In a number of systems investigated, including

rabbit globin, (Jeffreys and Flavell, 1977), chicken ovalbumin (Breathnach *et al.*, 1977), fibroin genes (Manning and Gage, 1978), 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 synthesising the protein and in tissues in which the protein is not produced. In these cases, at least, expression must therefore involve the activation of pre-existing control regions rather than the insertion or rearrangement of control regions adjacent to the gene. These conclusions suggest that the DNA rearrangement observed for the immunoglobulin genes may have evolved specifically as a means of generating at least part of the remarkable variability required by the immune system and is not a feature common to most eukaryote structural genes.

1.2 MODULATION OF GENE EXPRESSION

1.2.1 Transcriptional Control

Evidence has been presented above, suggesting that in some cases, gene amplification, deletion or rearrangement are important factors in gene regulation. However, as in prokaryotes, the major mechanisms of gene control in eukaryotes are thought to operate at the level of transcription.

(a) RNA Polymerases

An important demonstration of transcriptional control is the finding that three distinct forms of RNA polymerase are present in eukaryotic cells and that these transcribe

different sets of genes (Roeder, 1976).

RNA polymerase type I transcribes genes for 18s and 28s ribosomal RNA, RNA polymerase type III transcribes tRNA and 5s ribosomal RNA sequences and RNA polymerase type II apparently transcribes most other DNA sequences, making it responsible, therefore, for mRNA synthesis. Whether additional controlling elements can bind to RNA polymerase type II and further affect the specificity of the sequence it transcribes, as observed during the bacteriophage T_4 infection of prokaryotic cells, (Schmidt *et al.*, 1970), is yet to be determined.

(b) mRNA Levels

The general principle is well established that expression of a protein product is normally directly reflected in the concentration of its mRNA, i.e. most genes are not controlled by activation of stored mRNAs. Further evidence for transcriptional control of gene expression can be obtained from the cytological examination of dipteran polytene chromosomes. The polytene chromosome puffs have been shown to be the sites of intense RNA synthesis, and these features can be localised to specific chromosomal regions, varying according to the state of differentiation, (Pelling, 1964; McKenzie *et al.*, 1975).

A number of developmental systems have been studied in which the amount of a certain protein rise from very low or

undetectable levels to the very high levels characteristic of the mature, differentiated cell. Measurements made using complementary DNA probes to determine the levels of mRNA for haemoglobin in the erythroid cell series, (Hunt, 1974), ovalbumin in the chicken oviduct, (Palmiter, 1973), and keratin in the developing chick feather, (Powell *et al.*, 1976), have shown that during the process of differentiation, the increase in specific protein synthesis is due to the increase in the concentrations of the specific mRNAs.

While the evidence presented above adequately demonstrates that, in general, the expression of a specific gene is controlled at the transcriptional level, the problem of how transcription is controlled now arises.

(c) Steroid Hormone Control of Transcription

A great deal of experimental evidence has shown that steroid hormones are capable of specifically altering the expression of certain genes. The molecular biology of steroid hormone action has been most thoroughly investigated in the chick oviduct system, where administration of oestrogen or progesterone immediately induces the synthesis of mRNAs coding for eggwhite proteins, of which ovalbumin is a major component, (Harris *et al.*, 1975). Upon entering the cell, steroid hormones are initially bound to specific cytoplasmic receptor proteins. The hormone-receptor complex moves from the cytoplasm into the nucleus where it binds to sites on the target cell chromatin. This is followed by activation of specific

genes, resulting in the appearance of new species of mRNA. The progesterone receptor of chicken oviduct cells has been purified (Vedeckis *et al.*, 1978) and has been shown to be a protein-dimer, one subunit locating the hormone-receptor complex to specific regions of the chromatin, while the other subunit appears to alter the structure or conformation of the specific chromatin-DNA sites so that initiation of new RNA synthesis can occur (O'Malley *et al.*, 1978).

This remains, as one of the few examples, where we can begin to understand the process by which transcription can be promoted.

(d) DNA Methylation and its Role in Transcription Control

The idea that methylation is involved in gene regulation is gathering an increasing amount of support. The major way that vertebrate DNA is modified is by methylation of cytosine at the 5'-position (Vanyushin, *et al.*, 1970). It has been shown, using restriction enzyme analysis, (MspI recognises the same sequence as HpaII but can cut methylated DNA), that certain CpG sites, in vicinity of a wide variety of genes, are under-methylated in a tissue in which a gene is active.

Recently, the HpaII methylase enzyme has been used to introduce methyl groups into defined DNA sequences (Stein *et al.* 1982). Clones transformed within unmethylated aprt sequences were found to express the aprt gene. Clones transformed with methylated aprt neither transcribed nor expressed the gene,

even though the methylated sequences were incorporated into the genome with an efficiency similar to that for unmethylated aprt sequences. When the clones carrying methylated aprt were subjected to conditions selecting for aprt+ phenotype, reversion was found to be associated with undermethylation of the aprt sequences. It was concluded that the observed effect of methylation on aprt activity arises from a direct influence of methylation on gene expression. The connection between DNA methylation and gene expression is also supported by studies of mouse retroviruses (Struhlmann *et al.*, 1981; Harbers *et al.*, 1981).

Levels of methylation can be deliberately altered by the use of the nucleoside analogue 5-azocytidine which inhibits DNA methylation *in vivo*. Jones and Taylor (1980), have shown that 5-aza C can induce differentiation in cultured mouse embryo cells, and that at the same time and dosage, the cellular DNA becomes demethylated.

While a general undermethylation of specific sites in and around genes, such as globins (McGhee *et al.*, 1981) and metallothionein (Compere and Palmiter, 1981) has been correlated with their expression, the pattern in these cases is far from simple. Some genes, however, do not display such obvious correlations. For example, the α (2) 1 collagen gene of chicken has a pattern of HpaII site methylation that is invariant (with some sites completely unmethylated) in

five cell types that have been studied, whether or not the cell types synthesise detectable amounts of collagen (McKeon *et al.*, 1982).

However, it does appear that DNA methylation may play an important role in gene expression during development, but its generality and mechanism still remain unknown.

1.2.2 Sequence and Structures Involved in the Control of Transcription

Regardless of the mechanism of transcriptional control; via hormone-receptor action, different polymerase specificities or DNA methylation, the crucial step appears to involve the recognition of specific DNA control sequences.

Comparison of several cellular and viral protein-coding genes has revealed the existence of an AT-rich region of homology centered about 25-30 bp upstream from the mRNA start sites (Gannon *et al.*, 1979). This sequence, known as the "TATA" box, was first noticed by Goldberg and Hogness, (Goldberg, 1979), and has now been found upstream from every mRNA start site of all the sequenced eukaryotic mRNA-coding genes transcribed by RNA polymerase type II.

The consensus "TATA" box sequence, 5'-G T A T $\frac{A}{T}$ A $\frac{A}{T}$ G-3', has been compiled from the 5' sequences of 60 eukaryotic genes and shows a striking similarity to the prokaryotic promoter sequence (5'-T A T A A T G - 3') or "Pribnow" box (Rosenberg and Court, 1979), which is located about 10 bp upstream from the mRNA start site in all bacterial genes.

However, the "TATA" box cannot be the direct functional equivalent of the Pribnow box, because deletion of the "Pribnow" box in a bacterial gene abolishes all transcription (Miller, 1978) whereas deletion of the "TATA" box, demonstrated in sea urchin H2A (Grosschedl and Birnstiel, 1980), SV40 early genes (Benoist and Chambon, 1980), and polyoma early genes (Bendig *et al.*, 1980), does not eliminate transcription *in vivo*, but results in new RNA species having heterogeneous 5' ends. In *in vitro* systems, mutations in the "TATA" box have caused a significant decrease of specific transcription (Wasylick and Chambon, 1981).

The major conclusion from these results is that the "TATA" box is necessary for correct transcription initiation, but is not the primary *in vivo* target for promotion of transcription; it rather serves to guide the RNA polymerase into a correct initiation frame. Thus, it may act as a selector element for the transcription of eukaryotic genes that specifies the correct 5' end of the mRNA.

A second region of homology exists around the mRNA start site. Rather than having a particular sequence, as suggested previously by comparison of β-globin and adenovirus major late genes (Konkel *et al.*, 1978; Ziff and Evans, 1978), the mRNA start site appears generally to consist of an A residue, surrounded by pyrimidines (5'-py C A T T C pu 3'). The efficiency of initiation has been shown in several systems (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1980; Bendig *et al.* 1980) to be influenced by the actual DNA sequence around the mRNA start site.

In addition, there is evidence that sequences far upstream from the mRNA start site are important for *in vivo* transcription of the histone H2A and SV40 early genes. The sequence C C A A T is one such candidate that has been found in all eukaryotic genes located about 77^{+}_{-} 10 bp upstream from the mRNA start site (Efstratiadis *et al.*, 1980). Mutants of SV40 early genes (Benoist and Chambon, 1981) and H2A (Grosschedl and Birnstiel, 1980) genes with intact sequences around the mRNA start sites and "TATA" boxes, but with deletions mapping more than 150 bp upstream, display no *in vivo* transcription. It is suggested that sequences located more than 100 bp upstream may correspond to binding sites for proteins involved in the positive control of transcription (Breathnach and Chambon, 1981). It is also possible that this region is involved in the generation of "open", "active" chromatin structure (Mathis *et al.*, 1980), since there is evidence that chromatin structure of both viral (Saragosti, 1980) and cellular (Wu, 1980; Stadler *et al.*, 1980) genes is modified in regions corresponding to their 5' ends. Furthermore, the 5' end of each of the histone genes in *Drosophila*, has been shown to be exposed (i.e. sensitive to DNA'ase I) in active chromatin, whereas the corresponding gene sequences are protected in the nucleosome in inactive chromatin (Samal *et al.*, 1981; Elgin, 1981). This nucleosome phasing has been suggested (Bryan *et al.*, 1981) to be important in the regulation of transcription, by operating at the level of defining higher order domains, which in turn either facilitates or impedes access of particular regions of the DNA to the transcriptional machinery.

Knowledge concerning the sequences specifying termination of transcription is still very limited. Sequence data from the 3' end of all the sea urchin histones (Busslinger *et al.*, 1979) and data from our laboratory for chicken (D'Andrea *et al.*, 1981) and human histones genes so far sequenced have revealed a conserved sequence which, if transcribed into mRNA, would form a stable hairpin loop structure and therefore may play a role in maturation or termination of histone RNA transcripts.

In other eukaryotic genes, some sequence homologies have been recognised around the poly(A) addition site,(5'-A A T A A A - 3') (Proudfoot and Brownlee, 1976), and this sequence has been shown to be necessary for correct polyadenylation of SV40 mRNA (Fitzgerald and Shenk, 1981). This sequence may not be involved in the termination of transcription but is probably a signal for cleavage and polyadenylation of a larger primary transcript.

1.2.3 Post-transcriptional Control

The separation of the genetic material in a nucleus from the translation machinery in the cytoplasm is at the heart of post-transcriptional control mechanisms that are unique to eukaryotes.

Translation and degradation of a prokaryotic mRNA molecule can begin before its synthesis has been completed. A eukaryotic gene is transcribed completely, and then the RNA is usually modified in several ways before it is finally translated. An enzymatic complex joins ("caps") the 5' end of the RNA transcript with an inverted GTP (guanosine triphosphate residue); it then

methylates one internal adenine (A) residue (or more) of the mRNA molecule (Darnell, 1979). Another early event, is the addition of poly(A) (polyadenylate) residues to the 3' end of the completed RNA molecule (Darnell, 1979). The majority of eukaryotic genes that have been isolated from genomic DNA, have one or more intervening sequences that are transcribed with the coding regions into a long precursor RNA molecule (Dawid and Wahli, 1979). This precursor RNA is then processed by excision of the extra transcripts and splicing of the coding pieces of RNA (Tilghman *et al.*, 1978).

There is accumulating evidence that these three post-transcriptional modifications play essential roles in gene expression. Recent studies suggest that capping of mRNA facilitates binding to eukaryotic ribosomes (Kozak, 1978) and enhances stability of mRNA (Furuichi *et al.*, 1977). While mRNAs can function without poly(A), it is thought that poly(A) can prolong the half-life of mRNA (Huez *et al.*, 1974). Alternatively it has been recently suggested by Bira *et al.* (1980), that poly(A) may facilitate the splicing of nuclear RNA precursors. Thus the controlled inhibition of capping and poly(A) addition could modulate gene expression.

The role of intervening sequences in genes remains one of the biggest enigmas of all. Studies on the mechanism of RNA splicing have shown that different genes follow different pathways. Individual intron transcripts are removed in several steps in some genes. Mouse β-globin mRNA precursor, for example, requires

at least two splicing events to excise the large intron transcript (Kinniburgh and Ross, 1979) and three distinct stepwise pathways have been identified for the removal of an intron transcript in the chick α-2 collagen mRNA precursor (Avvedimento *et al.*, 1980). Studies on the vitellogenin (Ryffel *et al.*, 1980), ovalbumin (Chambon *et al.*, 1979) and ovomucoid (Tsai *et al.*, 1980) putative mRNA precursors, visualized in the electron microscope, suggest that there is not a rigid order of removal of intron transcripts from these molecules. However, a specific order of splicing is observed by the yeast mitochondrial cob-box mRNA precursor. Individual intron mutations in this gene block processing of the precursor at defined stages of its maturation and lead to the accumulation of intermediates containing full intron transcripts (Van Ommen *et al.*, 1980; Halbreich *et al.*, 1980).

The specificity and accuracy of splicing appears to reside in the nucleotide sequence around the splice site. From a study of 90 splice junctions, it was found that every intron begins with the dinucleotide GT at its 5' end and ends with the dinucleotide AG at its 3' end. (Breathnach and Chambon, 1981).

An experiment where a pair of "splice-sites" from a mouse β-globin gene were cloned into an "intronless" SV40 16s RNA gene resulted in a stable splice RNA being formed, whereas no detectable stable RNA was produced when the splice-functions were removed (Gruss and Khoury, 1980). These and similar results (Hamer and Leder, 1979; Hamer *et al.*, 1979) suggest that a splicing event is required for the accumulation of stable RNA, and that splicing may

be linked to transport of the RNA from the nucleus to the cytoplasm.

Prime candidates for post-transcriptional control are the proteins that interact with RNA in the nucleus, even during transcription, forming ribonucleoprotein (RNP) particles. It has been suggested that the splicing enzyme complexes with small nuclear RNAs (SnRNAs) and that these "splicer" RNAs could hybridise near the splice points and "guide" the splicing enzyme (Murray and Holliday,1979). One of these RNAs, U_1 RNA, has been found to exhibit complementary nucleotide sequence to splice junctions (Lerner *et al.*, 1980) and could function by forming base pairs with a number of residues lying within the two ends of an intron, thus bringing the exon sequences into proximity.

This suggests a powerful post-transcription mechanism in which the presence of specific splicing enzymes or alternative splicing patterns could determine which mRNAs can reach the cytoplasm and function there. It is clear from the above discussion that RNA splicing is a complex process and may be used as a control mechanism in gene expression.

1.2.4 Translational Control

Modulation of translation by one or more factors involved in protein synthesis is a potentially powerful way to influence gene expression. One means of translational control that is important in eukaryotes is the stability of mRNA. A single fibroin gene from the posterior silk gland of *Bombyx mori* is responsible for the synthesis of 10^{10} fibroin protein molecules in several days (Suzuki *et al.*, 1972). This is accomplished by a higher rate of mRNA synthesis and efficient utilisation and stabilisation of mRNA formed. About 10^5 molecules are transcribed from one gene, and each serves as templates for the synthesis of about 10^5 protein molecules. At each larval molt, fibroin mRNA is degraded and is resynthesised during the next instar (Suzuki and Suzuki, 1974).

A change in the stability of mRNA for casein in mammary glands has been induced by a change in hormone concentrations (Guyette *et al.*, 1979). This has been shown to be due in part to the increased lability of the mRNA in the absence of hormone. An extreme, but poorly understood example, of stable and inactive mRNA is a phenomenon referred to as "masked" mRNA (Gross, 1967). Unfertilized sea urchin eggs store mRNA for months in an inactive state. Minutes after fertilisation, the rate of protein synthesis increases, encoded by this stored mRNA. The activation of other dormant states (seeds, encysted embryos) may occur by similar but still unknown mechanisms.

1.3 MULTIGENE FAMILIES

Many eukaryotic genes are arranged in multigene families. An examination of their diverse arrangement and abundance may provide an insight into the role of gene organisation, with respect to the control of gene expression.

1.3.1 Simple Multigene Families - Ribosomal and 5s RNA Genes

Some genes are present in many reiterated copies in tandem array along a chromosome. The first of these multigene familes that was studied was the 18s and 28s rRNA gene family (rRNA) of sea urchin (Wallace and Birnstiel, 1966), followed by 5s genes (5s DNA) (Brown *et al.*, 1971) in *Xenopus*. These genes could be purified and characterised from genomic DNA, before the development of recombinant DNA technology, because of their abundance in the genome, their high G/C content and the availability of their RNA products as hybridisation probes. Studies with purified high-molecular-weight genomic rDNA and 5s DNA, demonstrated that these genes are organised in tandem along chromosomes and separated from each other by "spacer" DNA.

In *E. coli*, the ribosomal and 5s genes comprise a single transcription unit, and the three RNAs are processed from one polycistronic RNA. Yeast (Rubin and Sulston, 1973) and the cellular slime mold (Maizels, 1976) have all three genes in each repeating unit, but the 5s RNA gene is transcribed separately from the 18s and 28s rRNA genes. Different forms of RNA polymerase transcribe the 5s RNA genes and the large rRNA genes.

The 18s and 28s rRNA of *Xenopus* are closely linked and separated by short transcribed spacer regions and longer non-transcribed spacer regions (Reeder, 1979). One large RNA transcript is processed to the 18s and 28s rRNAs. Each ribosome contains one molecule each of 5s, 18s and 28s DNA (or their equivalents in prokaryotes).

In multicellular eukaryotes, the 5s RNA genes are clustered in their own separate multigene families (Brown and Weber, 1968). It is not clear why the simple coordinate transcription and processing of these three genes has evolved to these more complex independently controlled arrangements. The dilemma is more puzzling when one considers the control mechanism of 5s RNA genes that has evolved in fish and amphibians. In the oocytes of these animals, there is a massive amplification of rDNA, and these cells are capable of thousands of times higher rate of ribosome synthesis than occurs in single somatic cells. However, the 5s genes are not amplified (Brown and Dawid, 1968); instead, the genome contains one or more large additional multigene families encoding 5s RNA (Brown and Fedoroff, 1978). The oocyte-specific 5s DNA is only expressed in growing oocytes; these genes are shut off in somatic cells so that only members of the smallest of the multigene families, termed somatic 5s DNA, function. The largest oocyte-specific 5s DNA multigene family studied (Brown *et al.*, 1971) has about 20,000 copies per haploid set of chromosomes, an amount that comprises about 0.7 per cent of the genomic DNA.

The function of simple multigene families appears to be to increase an organism's capacity to synthesise a product. However, defining a role for tandem organisation of these genes is still a major problem.

1.3.2 Linked, Related Genes, Under Developmental Control

Five different functional β-type globin genes have been described in humans (Fritsch *et al.*, 1980) and four α-type genes (Lauer *et al.* 1980). The β-globin-type genes are closely linked; all have been cloned and characterised (Fritsch *et al.*, 1980). The genes are related to each other and must have evolved from a common ancestral gene. The β-type genes probably diverged from the same ancestral gene as the α genes, but much earlier and these genes are not on the same chromosome as the β-type genes, thus illustrating that clustering is not a necessary feature for coordinate expression.

The genes for the β-type globins in humans are under developmental control. Human embryos, fetuses and adults make globins with different β-type subunits. So perhaps it is significant that β-type genes are arranged in the same order along the chromosome as they are expressed chronologically in development. One deletion that removes adult β genes alters the control of the fetal β-type genes (termed γ) that are linked to them (Fritsch *et al.*, 1979). This suggests that regions some distance from a gene may influence its expression.

Other structural genes are being identified as members of multigene families. There are several kinds of actin (Fyrberg *et al.*, 1980), vitellogenin (Wahli and Dawid, 1980), collagen (Eyre, 1980) and keratin (Fuchs and Green, 1979) proteins encoded by multiple genes. Batteries of genes for eggshell proteins in silkworm follicle cells are expressed at specific times in development of the egg chorion (Goldsmith and Basehoar, 1978). These genes are clustered in units of divergently oriented pairs in the silk worm genome in developmental order.

More complex multigene families group together genes that are related but not identical. In some cases, their products are needed at the same time by a cell and for related purposes. Also, neighbouring genes in some multigene families are expressed at different development stages to carry out related functions (globin and chorion multigene families).

In the examples cited above, it is seen that in some gene families where clustered, related genes (e.g. globin) are transcribed at different times, whereas in other cases (e.g. ovalbumin family) clustered genes are coordinately expressed. Clearly then no general rules can yet be deduced from the arrangement of gene families.

It was thought that the histone gene system would provide an excellent system for studying the role of gene clustering and order in gene expression since it was known at the commencement of this thesis that the five histones were organised into a single repeating unit in sea urchin and *Drosophila*. Also the expression of these genes

are tightly regulated with respect to the cell cycle, the developmental stage and the expression of the other histones. Since histones are highly conserved proteins and ubiquitous in eukaryotic cells, comparison of histone gene organisation in higher evolutionary systems may relate gene organisation to gene expression. The next section presents in more detail what is now known about histones and their genes.

1.4 THE HISTONE GENE FAMILY

1.4.1 The Histone Proteins

The histones are a set of five small basic proteins. They are fundamental structural proteins of chromatin and are found in all eukaryotic organisms. Bacteria also contain histone-like proteins whose functions appear, in some ways, to parallel those of histones in eukaryotes. (Lathe *et al.*, 1980; Hubscher *et al.*, 1980) but the evolutionary relationships between these proteins and the histones is unclear.

The histones, H2A, H2B, H3 and H4 interact with DNA and with each other to form a nucleosome core, the fundamental unit of chromatin structure. The fifth histone, H1, is thought to play a role in the higher order structure of chromatin and is present at about half the molar concentration of the other histones (Noll and Kornberg, 1977).

Histones may be modified postsynthetically in a variety of ways (Isenberg, 1979). Phosphorylation of histones has been clearly

correlated with chromosome condensation. Acetylation has been associated with regions of the genome that are structurally active. A modification of H2A, named A24, appears to be preferentially located in the inactive regions of the genome. Histones may also be methylated and ribosylated, but there is still only limited data correlating these modifications with any structural or functional aspects of chromatin.

The primary sequence of the histone proteins has been highly conserved during evolution (reviewed by Isenberg, 1979), reflecting directly the very fundamental character of functions performed by histones within the cell. H3 and H4 are the most stringently conserved of the histone proteins, as evidenced by the fact that histone H4 of pea plants differs from H4 of calf thymus by only 2 of 102 amino acid residues. H3 and H4 have two different, but conserved domains, a basic N-terminal region and a C-terminal globular domain, whereas H2B and probably H2A proteins are thought to be evolutionary hybrids, because only the C-terminal globular regions are conserved, while the rest of the protein is variable. Histone H1 is the most variable of the histones but like H2B has highly conserved stretches which have globular-like residue distributions.

Only recently has the concept of regarding the inner histones H2A, H2B, H3 and H4 as simply "highly conserved" changed. This has been due to the introduction of new and more sensitive techniques for resolving and characterising tissue-specific nonallelic histone protein structural variants (Alfageme *et al.*, 1974; Von Holt *et al.*, 1979), coupled with the ability to clone and characterise sea urchin

histone variant genes (Busslinger *et al.*, 1980).

The variations of histone structures that have been examined can be classified into two types. The first group comprises simple point mutations or deletions of one or several residues, and occurs in all histones. The second type of structural variations have only been identified in H2A and H2B and consist of extensive modifications of major areas of the molecules through reiteration, insertion, deletion and point mutations. Such changes have led, in some variants, to the establishment of new domains in the structure. These histone sub-types, or variants, have been identified in sea urchin to appear in a programmed fashion (Ruderman *et al.*, 1974; Cohen *et al.*, 1975) as both developmental and tissue-specific histones. This supports the notion that histones could perhaps perform a specific role during differentiation.

It has been suggested (von Holt *et al.*, 1979) that specific histone variants are responsible for the particular functional state of chromatin, presumably, by lending to the DNA a conformation which is either transcriptionally active or inactive.
H1 histones are probably involved in the close packing of the elementary nucleosome repeat structure into higher orders of chromosome condensation, to allow the chromatin to take on highly specific conformations in certain areas.

In the course of evolution the component polypeptide chains have been selected to enable chromatin to fold and unfold reproducibility for the replication cycle, a function which is based on

protein-protein recognition. The fundamental identity of that process in all eukaryotic cells probably accounts for the conservative structures of H3 and H4, and is reflected in the high degree of homology of the hydrophobic C-terminal half of the many H2B variants, and also in the homologies of the globular cores of the histone H1, whereas demands for suitable protein-dictated DNA conformations on which regulatory processes can manifest themselves have shaped the variable N-terminal regions in the histones H2A, H2B and H1 (von Holt *et al.*, 1979). Further work on the comparison of histone structures at different levels of evolutionary complexity, as well as during the process of differentiation, is needed before the function of histone variants is clear.

1.4.2 Histone mRNA

Histone RNAs were first identified during S phase of mammalian tissue culture cell cycle on the basis of size of the RNA, approximately 9s, and on the appearance of these RNAs during the cell cycle, when histone proteins are synthesised (Borun *et al.*, 1967; Gallwitz and Mueller, 1969). Subsequently, these criteria were used by Kedes and Gross (1969a, 1969b) and Nemer and Lindsay (1969), to identify histone mRNAs from rapidly cleaving sea urchin embryos, where 30 per cent of the total protein synthesis is histones. The identity of the sea urchin histone mRNA was later verified by *in vitro* template activity (Gross *et al.*, 1973), RNA sequence analysis (Grunstein *et al.*, 1976) and homology with histone genes (Childs *et al.*, 1979a).

The mature histone mRNAs are about 9s in size and in general are not polyadenylated (Adesnik and Darnell, 1972). Recently however, evidence has accumulated showing that at least a sub-fraction of histone mRNAs may be polyadenylated, (Browne *et al.*, 1977; Ruderman and Pardue, 1978), and it has been proposed that these differences may be related to stage specific switches of histone mRNA sequences.

Histone mRNA 5' termini are known to be capped in sea urchin (Surrey and Nemer, 1976) and in HeLa cells (Stein *et al.*, 1977) with structures characteristic of the 5' mRNA termini of a variety of eukaryotes (Moss *et al.*, 1977). Methylated nucleotides in internal sites are a rare but constant feature of eukaryotic mRNAs (Wei *et al.*, 1975), but examinations of histone mRNAs of both echinoderms and HeLa cells has failed to detect such bases (Surrey and Nemer, 1976; Stein *et al.*, 1977a).

Intensive searches for the presence of high-molecular-weight or polycistronic precursor molecules, especially in nuclear RNA, have been unsuccessful when sea urchin cleavage stage embryos were examined (Childs *et al.*, 1979a, 1979b; Kunkel and Weinberg, 1978). However, Kunkel *et al.*, (1978) detected high-molecular weight RNA hybridising to sea urchin histone genes in sea urchin gastrulae. That report suggested that early and late histone genes are not only different in sequence, but may be transcribed or processed in different ways.

Support for the presence of high-molecular-weight histone mRNA in HeLa cells has been provided by Melli *et al.*, (1977a), who

hybridised sea urchin DNA probe to pulse-labelled nuclear RNA. They detected nonpolyadenylated high-molecular-weight RNAs that disappeared from the nuclei when cells were treated with actinomycin D.

Using an entirely different approach, Hackett *et al.*, (1978) have also suggested that the HeLa cell gene transcriptional unit is larger than histone mRNA molecules, but that each mRNA is independently transcribed. These authors used the indirect method of UV inactivation mapping to determine that the decay rates of histone mRNA translational activity were similar for each of the five genes. The resolution of these conflicting results must await further experimentation at the level of gene structure and organisation.

However, the general opinion is that for most histone mRNAs, no precursor form exists. It is interesting to note though, that Engel *et al.*, (1982) have found intervening sequences in an H3 chicken gene sequence.

1.4.3 Regulation of Histone Gene Expression

The regulation of histone gene expression is known to occur at several levels. Histone protein synthesis is closely coupled to DNA replication in somatic cells of animals (Robbins and Borun, 1967), yeast (Moll and Wintersberger, 1976) and protozoans, (Prescott, 1966) and numerous lines of evidence have shown that there is a rapid loss of histone mRNA activity from the cytoplasm and polyribosomes when DNA replication is stopped, (Robbins and Borun, 1967 ; Gallwitz and

Mueller, 1969). Therefore, it was generally agreed the regulation of histone mRNA activity is controlled at the transcriptional level (Detke *et al.*, 1978, 1979; Hereford *et al.*, 1981).

Cases of apparently uncoupled synthesis have, however, also been described. Melli *et al.*, (1977b) find putative histone mRNA sequences in the nuclear RNA of HeLa cells throughout the cell cycle and so suggest that processing or transport may be more important than transcription in controlling protein synthesis. In addition, Groppi and Coffino (1980) find that histones in CHO cells are synthesised at equivalent rates throughout the cell cycle, but only assemble into chromatin during S phase. Past experiments attempted to measure the rate of histone synthesis during the cell cycle by pulse-labelling synchronised cells with radiolabelled amino acids and purifying the histone proteins by extraction from isolated nuclei of chromatin prior to analysis. The classic result from such experiments was that nuclei prepared from S phase cells contained newly synthesised histones, while nuclei from G1 phase cells did not. The tacit assumption inherent in these past experiments was that histones synthesised in all phases of the cycle would rapidly and quantitatively associate with chromatin. Groppi and Coffino (1980) question this assumption and suggest that if histones were made in the absence of DNA synthesis, they would not bind to chromatin and would be discarded during sample preparation.

However, the above results do not adequately explain the data suggesting that histone mRNA is absent in the cyotoplasm of G1 cells

(Stein *et al*, 1978; Stein *et al.*, 1980). Further studies with cloned histone probes will be required before this question is resolved.

Developmental regulation of histone gene expression of both a quantitative and a qualitative (variant specific) kind also occurs (Zweidler, 1980). The latter type is particularly evident for the H1 class of histones during terminal differentiation of specific tissues such as the nucleated erythroid cells of *Xenopus*, (Destree *et al.*, 1979; Risley and Eckhardt, 1981) and chickens (Molgaard *et al.*, 1980), or during spermotogenesis in several organisms (Geraci *et al.*, 1979; Seyedin and Kistler, 1979). A particularly interesting example is the switch from H1 to H5 histone gene expression in terminally differentiating erythroid cells of the chick (Appels and Wells, 1972). The protein sequences or partial sequences of many of these types of late variants are known (von Holt *et al.*, 1979; Brandt *et al.*, 1980), although little is understood about the control of their synthesis.

Early developmental events have been most thoroughly studied in the sea urchin (Grunstein *et al.*, 1981). A shift occurs in the utilisation of histone mRNAs during early embryogenesis. Maternal histone mRNAs are present in the unfertilised egg and can be utilised during early embryogenesis until the blastula stage. Soon after fertilisation of the egg, a family of several hundred (early) histone genes are activated (Schaffner *et al.*, 1978; Sures *et al.*, 1978). During blastrulation the mRNA products of these genes gradually disappear from the polysomes and are replaced by a new class of (late)

histone mRNAs that differ in size and sequence and code for a new group of histone proteins (Cohen *et al.*, 1975; Newrock *et al.*, 1978; Childs *et al.*, 1979a).

On the basis of mRNA lengths, their melting characteristics in hybrids formed within early histone DNA repeats (Kunkel and Weinberg, 1978; Childs *et al.*, 1979a; Spinelli *et al.*, 1979), and comparative finger prints (Grunstein, 1978; Grunstein *et al.*, 1981), these late mRNAs must derive from genes quite highly diverged from the "early" principal repeats. While the molecular events controlling this switch are poorly understood, transcriptional level regulation is clearly implicated in the switchoff of the early genes, since these events can be observed in isolated nuclei under conditions that reflect only the elongation of RNA chains initiated *in vivo* (Levy *et al.*, 1978). Furthermore, the switch is clearly "determined" as early as the 16-cell stage and occurs in disaggregated blastomeres from this stage even when cultured separately (Arceci and Gross, 1980).

Histone gene regulation during *Drosophila* embryogenesis bears marked parallels to the situation in the sea urchins. A key regulatory mode in this case, however, seemingly occurs at the level of mRNA turnover (Anderson and Lengyel, 1980). *Drosophila* also differs from sea urchins in that no clear "switch" to the predominant expression of late histone variants occurs, and in that no cloned coding variant genes have been detected (Goldberg, 1979).

In *Xenopus* the genes are only reiterated 20-50 fold (Jacob *et al.*, 1976) making zygotic transcription a relatively minor factor during early embryogenesis. Instead, the developing embryo utilises stored histone proteins to a greater extent and also activates a large pool of maternal mRNAs (Adamson and Woodland, 1974; Woodland and Adamson, 1977; Woodland, 1980).

Perhaps the major conclusion now to emerge from all these studies is that no universal developmental control mechanism of histone gene expression should be anticipated. Rather, each organism seems to pick from an array of potential regulatory mechanisms (storage of histones or masked mRNAs, transcriptional control, translational control, control of mRNA stabilities etc.) to supply adequately the specific, temporal demand of its growing differentiating cells for histones. How such multiple regulatory mechanisms are effected and coordinated in molecular terms is obscure. Further comparative data on the organisation of the histone genes and their transcripts may clarify the situation.

1.4.4 Histone Gene Organisation

As is clear from above, histone gene expression is known to be regulated at several levels. A prerequisite for a greater insight into perhaps the most fundamental level of regulation - the transcriptional level - is a knowledge of the organisation of the histone genes.

Progress in unravelling the organisation and mode of expression of various histone genes present in the eukaryotic genome has, however, been remarkably uneven. Until recently, almost all available information has been derived from examination of the sea urchin histone genes (Kedes and Birnstiel, 1971). The ability to physically enrich these genes (Birnstiel *et al.*, 1974) led to their early cloning and characterisation. In turn, the sea urchin histone genes have been used as a source of probes for the cloning of further histone genes from several species. Consequently, the organisation of the sea urchin histone genes has served as a prototype against which the histone genes of other species can be compared.

(a) Sea Urchin

The genes for the five histones H1, H2A, H2B, H3, H4 are clustered into a unit that is tandemly repeated several hundred times in the sea urchin genome. Restriction analysis of sea urchin genomal DNA of *Sammechinus miliaris*, showed the presence of an approximately 6 kilobase long repeat unit, in which genes coding for each of the five histone proteins were interspersed with one another, and that spacer DNA was located between each of the coding regions (Cohn *et al.*, 1976; Schaffner *et al.*, 1976; Kedes, 1976).

Gross *et al.*, (1976a) isolated the individual sea urchin mRNA species and then, using restriction enzymes, limited exonuclease digestion of cloned DNA, and hybridisation with purified mRNAs, were able to demonstrate that the order and polarity of the histone gene cluster in *P. miliaris* is -

5' - H1, H4, H2B, H3, H2A - 3'

with all mRNAs transcribed from the same DNA strand, (Gross *et al.*, 1976b)(Table 1.1). Similar studies on other sea urchin species, *S.* (*Strongylocentrotus purpuratus*), using restriction, (Cohn *et al.*, 1976) and electron microscopic (Wu *et al.*, 1976) techniques, demonstrated that the order and polarity of the histone repeat and the position of A - T rich spacer regions were conserved between these two species.

Direct DNA-sequencing of cloned histone gene repeats (Schaffner *et al.*, 1978; Sures *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. The A - T rich spacer regions are made up of relatively simple nucleotide arrangements, but do not show any evidence of internally repetitive sequences.

The topology exemplified by the major histone repeats of *P. miliaris* and *S. purpuratus* has been also remarkably conserved between the genes of several different sea urchin species that are separated by millions of years of evolution with respect to gene order and transcriptional polarity (Table 1.1).

However, except for these constraints, various forms of heterogeneity have been found. Southern blot analysis of restricted genomic DNA from *P. miliaris* (Schaffner *et al.*, 1976), and particularly from *S. purpuratus* has revealed microheterogeneity in the length of the repeats that is reflected in restriction-site and spacer sequence microheterogeneities in separately cloned

Table 1.1

Histone Gene Organisation

Summarised is the organisation and polarity of histone genes from a number of species (Hentschel and Birnstiel, 1981). The specific clones, repeat length or insert length* and histone gene repetition frequency is also shown. Arrows indicate direction of transcription. Further details on the organisation of specific chicken (λCH.02) and human histone clones is presented in the thesis.

Table 1.1

Species	Clone	Repetition Frequency	Repeat Length Kb	Organisation and Polarity
P. miliaris	h22	300-600	6.3	→ → → → → H1 H4 H2B H3 H2A
	h19	5-10	6.7	" " " " "
S. purpuratus	Sp17/2	300	6.5	" " " " "
D. melanogaster	500	100	4.8	← → ← → → H3 H4 H2A H2B H1
X. laevis	X1-hi-1	20-50	5.8*	→ → ← → H3 H4 H2A H2B
	X1h1		15*	H3 H4 H2A H2B H1 H3 H4
N.viridescens	NV51	600-800	9.0*	← ← → ← ← H4 H2A H2B H3 H1
Chicken	λCH.01	10-20	15*	← → ← ← → H3 H2A H4 H2A H2B
	λCH.02		14.8*	H4 ← H3 → H3 H4 H2B H1
	ch1a		>10*	H3 H2B H2A
	ch2e		>10*	H2A H4 H2A H2B
Yeast	TRT 1	1	6*	← → H2A H2B
	TRT 2/3	1	13*	← → H2A H2B

major DNA repeats (Overton and Weinberg, 1978). In *P.miliaris* some such microheterogeneity has been shown to be due in part to non-allelic histone-coding variants (Birnstiel *et al.*, 1979; Busslinger *et al.*, 1980).

In a further sea urchin species, *Lytechinus pictus*, two types of major non-allelic repeats are present in a non-intermingled fashion in the genome; both of these code for similar, if not identical, early embryonic mRNAs, but they have substantial regions of totally divergent spacers. (Cohn and Kedes, 1979a, 1979b).

Recently, possible examples of pseudogenes in the histone gene family have been discovered. Childs *et al.*, (1981) restricted *L. pictus* DNA with an enzyme (BamHl) that does not cut within highly reiterated repeats and on probing a Southern blot of the digest with an H3-gene-specific probe observed lower molecular weight bands containing H3 sequences. Analysis of a clone containing one of these fragments revealed that an H3 gene together with some spacer was present outside of its "parental" repeat, and was instead embedded in moderately repetitive non-histone sequences. Further examination of this phenomenon with other gene-specific probes revealed the general presence of a mobile class of "orphon" histone genes (or pseudogenes) represented 5-20 times per genome for each gene and in a highly idiosyncratic fashion in the individual sea urchins.

Notwithstanding the presence of such orphons, it is still the identical repeat organisations, despite a hundred million years of evolutionary divergence, that remains the most striking feature of sea urchin histone gene organisation. The preservation of the early repeat topology in all sea urchins presumably results from some selective advantages, and several inferences have been made as to what these might be. The "higher dosage" repetition of the major repeats is often assumed to reflect the high transcriptional demand for histone mRNA during early embryogenesis (Birnstiel *et al.*, 1979). Another rationalisation for tandem repetition has been that such organisation enables coordinate control of the genes (Finnegan *et al.*, 1977), while clustering of the individual genes into a repeat unit would ensure that the relative number of different histone genes remains balanced after an unequal-cross-over event. Such cross-over events may on the other hand, be responsible for maintaining the homogeneity of tandem repeats in general (Smith, 1976). Furthermore, the fact that the sea urchin histone genes are all coded by the same DNA strand has led to the suggestion that a single transcriptional unit could coordinately control their stoichiometric production.

These proposals will be analysed in more detail later in this thesis.

(b) *Drosophila* Histone Genes

The histone genes of *D. melanogaster* are present at about 100 copies per haploid genome. A number of clones of *D. melanogaster* DNA containing histone mRNA coding sequences have been isolated, (Lifton *et al.*, 1977). These were selected by their ability to hybridise to sea urchin histone mRNA.

The structure of the major *Drosophila* histone repeats resemble the sea urchin-type organisation in that they are highly reiterated in tandem clustered arrays of the five genes, but they differ in gene order,

5' H1 H2B H2A H4 H3 3'

and, most strikingly, in that there are two genes coded on one strand and three on the other (Table 1.1). This clearly means that *Drosophila* histone mRNAs cannot be derived from a single RNA transcript (Lifton *et al.*, 1977).

As with sea urchin repeats, microheterogeneity is observed superimposed on the basic pattern. Two noninterspersed predominant repeats are present that differ by a 208 bp insertion in the long H1-H3 spacer, as well as several different repeats, none of which have so far been shown to code for structural protein variants (Goldberg, 1979).

The two major repeat types appear to be located together in region 39D-E of *Drosophila* polytene chromosomes (Pardue *et al.*, 1977), however, *Drosophila* histone gene "orphons" exist as well, outside the major cluster (Childs *et al.*, 1981).

(c) Other Invertebrates

Very little data is currently available on the distribution and arrangement of the histone genes in species other than *D. melanogaster* and sea urchin. Restriction and hybridisation techniques have been used to determine the size of the histone gene repeat unit for a number of invertebrates (Freigan *et al.*, 1976), including the horseshoe crab, (4.1 kb), clam, (4.5 kb), oyster (6.3 kb) and worm (5.2 kb). Hybridisation with different histone probes suggests that, once again, the genes coding for each of the histone proteins are clustered within a repeat unit, in these species.

Recently recombinant DNA techniques have been used to isolate yeast DNA clones containing all the core histone genes, (Hereford *et al.*, 1979). The H2A and H2B genes are adjacent, are divergently transcribed and exist as two genetically unlinked copies that appear to have arisen from a duplication event about two hundred million years ago (Wallis *et al.*, 1980). Interestingly, the H2A and H2B gene pairs are not contiguous, within 8 kb, with the H3 and H4 genes (Hereford *et al.*, 1979) (Table 1.1). Moreover, it is questionable whether the structural H2B variants have unique functions in the yeast life cycle, because yeast cells constructed with frameshift mutations in either of the H2B variant genes appear to be substantially normal, despite having chromatin clearly depleted of the mutated histone (Hereford *et al.*, 1981).

(d) Vertebrates

Detailed analysis of the organisation of several vertebrate histone genes are presently emerging. *Xenopus* contains 20-50 copies of its histone genes (Jacob *et al.*, 1976). Moorman *et al.*, (1980) have isolated and partially characterised a clone containing four of the five histone genes (all except H1) clustered on a 5.8 kb EcoR1 fragment of *Xenopus* DNA. The gene order of these four genes (Table 1.1) is the same as the *Drosophila* order (Lifton *et al.*, 1977), but, in contrast to *Drosophila*, the H2A and the H3 are transcribed covergently (Moorman and Destree, cited by Hentschel and Birnstiel, 1981). The same gene order is observed on a 15 kb genomic clone, which contains seven linked histone genes (Zernik *et al.*, 1980), in what appears to be more than one repeat of a tandem motif (H3 - H4 - H2A - H2B - H1 - H3 - H4). More surprisingly, different genomic clones from the same library show at least one other gene order, each of the two gene orders being associated with a particular H1 structural variant (H1A or H1B) present in the repeats (Zernik *et al.*, 1980).

The organisation of the newt histone genes in the species *N. viridescens* has recently been elucidated and proven to be completely different from *Xenopus*. In this newt, the majority of the histone genes comprise homogeneous 9 kb clusters repeated 600-800 fold in a genome 15 times larger than that of *Xenopus*. Some of these genes are located in the sphere loci of chromosome 2 and 6 (Gall *et al.*, 1981).

Analysis of a cloned cluster has revealed a gene order H1 - H3 - H2B - H2A - H4 in which all the genes except for H2B are transcribed in the same direction (Stephenson *et al.*, 1981). Most surprisingly these clusters, although homogeneous with respect to several restriction sites, are not tandem repeats as in the sea urchin genes but rather are separated by up to 50 kb of a repeating 225 bp satellite DNA.

Chickens contain about 10 histone gene copies (Scott and Wells, 1976). Several genomic histone clones have now been isolated from the Engel and Dodgson chicken library by our laboratory (Harvey and Wells, 1979; Harvey *et al.*, 1981; D'Andrea and Coles, unpublished) and by Engel and Dodgson (1981). Restriction analysis and gene mapping of these clones have revealed a degree of clustering but no repeat unit, as first suggested by Crawford *et al.*, (1979). (Table 1.1). Chicken histone H5, which is a tissue-specific histone variant that significantly displaces H1 in mature erythrocyte chromatin (Neelin *et al.*, 1964; Champagne *et al.*, 1968, Miki and Neelin, 1975) has now been cloned from chicken histone H5 mRNA, by our laboratory (Kreig *et al.*, 1982) and by Ruiz-Vasques and Ruiz-Carrillo, (1982). A genomal clone has also been isolated containing the chicken H5 gene and no other histones have been identified within 7.5 kb on either side of this gene (Kreig, unpublished). It is therefore thought that H5 may not be included in the chicken histone gene cluster but further characterisation is required.

Similar organisational heterogeneity emerges from very recent examination of mammalian genomic clones. A mouse 5.2 kb fragment containing an H4 gene appears to lack other histones (Seiler-Tuyns and Birnstiel, 1981) while other fragments have diversely organised clusters of histone genes (Sittman *et al.*, 1981).

The human histone genes are reiterated 30-40 fold (Wilson and Melli, 1977) and have been tentatively located on the long arm of chromosome 7 (7q 32-36) by *in situ* hybridisation studies using H4 mRNA (Yu *et al.*, 1978) and cloned sea urchin histone genes (Chandler *et al.*, 1979) as probe.

Evidence suggesting that the human histone gene organisation also lacks any repeating unit is presented in this thesis as well as by recent results from Heintz *et al.*, (1981) and Sierra *et al.*, (1982).

It is now becoming possible to view the gross histone gene organisation in an evolutionary span greater than that possible for any other eukaryotic family. However, instead of a clear unifying theme, it is clear that there is no such thing as a "typical histone gene organisation".

1.5 Perspectives

Work presented in this thesis describes the isolation and characterisation of chicken and human histone gene clones. At the time when this work was undertaken genomal chicken histone clones had been isolated in our laboratory and preliminary characterisation of one of these (λCH.01) was in progress. Assuming sufficient homology between chicken and human histone gene sequences, it became possible to screen for human histone genes. The advantage of studying human histone genes was two fold. Firstly, since it had just recently been reported (Melli *et al.*, 1977a) that human histone genes may be transcribed into a large precursor molecule, it was of interest to establish the organisation of the human histone genes to determine if co-ordinate expression of the five histone genes could be achieved through processing of a polycistronic mRNA. Secondly, a study of the human and chicken histone gene systems provided an attractive model for an evolutionary comparison of gene organisation, structure and expression, since the sea urchin and *Drosophila* histone gene systems were fairly well characterised.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemicals and Reagents

All chemicals were of analytical reagent grade. The sources of the more important chemicals and reagents are listed.

Acrylamide: Sigma - twice crystallised from $CHCl_3$

Agarose: Sigma

Agarose (low melting point): B.D.H. Australia

Ampicillin: Sigma

Bromophenol Blue: B.D.H. (Australia)

B.C.I.G.: Sigma

Caesium chloride: Harshaw

Chloramphenicol: A gift from Parke-Davis

Diethyl Pyrocarbonate: Sigma

Dithiothreitol: Sigma

Ethidium Bromide: Aerosol Industries

Ficoll 400: Pharmacia

Formic Acid: Carlo Erba

Hydrazine Hydrate: Tokyo Kasei

I.P.T.G.: Sigma

M13 17 base primer: Collaborative Research

M13 15 base primer: P.L. Biochemicals

Nitrocellulose Filters, BA85: Schleicher and Schuell

N, N^1-methylene bisacrylamide: Sigma

Nucleoside and deoxynucleoside triphosphates (ATP, dATP, dCTP, dGTP, dTTP): Sigma

γ-{^{32}P}-ATP, α-{^{32}P}-dCTP, α-{^{32}P}-dGTP (1500, 500 and 500 Ci/m mole respectively): A gift from Dr. R.H. Symons.

Oligo $(dT)_{10}$: P.L. Biochemicals

Phenylhydrazine: Sigma

Piperidine: B.D.H. (Australia)

Polyvinyl pyrrollidone: May and Baker

Sephadex G-50 (Medium): Pharmacia

Sepharose-Blue: Pharmacia

Spermine tetrahydrochloride: Calbiochem

TEMED: Tokyo Kasei

Tetracycline: A gift from Commonwealth Serum Laboratories

Xylene Cyanol FF: Tokyo Kasei

2.1.2 Enzymes

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

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

E. coli DNA-polymerase I: Boehringer-Mannheim.

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

Ribonuclease A: Sigma

E. coli Deoxyribonuclease I: Sigma

Proteinase k: E. Merck, Darmstadt.

Bacterial alkaline phosphatase: Worthington

Polynucleotide kinase: Boehringer-Mannheim.

E. coli DNA ligase: Boehringer-Mannheim.

T4 DNA ligase: Was prepared by Dr. P. Molloy

Restriction Enzymes: Were purchased from New England Biolabs and Boehringer-Mannheim.

EcoR1 was prepared as described in (2.2.3).

2.1.3 Histone DNA Gene Probes

Sea urchin histone DNA gene probes were prepared from h19 genomal clone (Schaffner *et al.*, 1978) and were a gift from Dr. Jacob.

Chicken histone DNA gene probes (H3, H4, H2A and H2B) were isolated from λCH.01, and prepared as subclones in pBR322 by R. D'Andrea.

H3 Probe: 150 bp containing the coding region from amino acid 10 to amino acid 59.

H4 Probe: 338 bp containing 14 bp 5' noncoding, the entire coding region and 12 bp 3' noncoding.

H2A Probe: 709 bp containing 196 bp 5' noncoding, the entire coding region and 123 bp 3' noncoding.

H2B Probe: 537 bp containing 146 bp 5' noncoding, the entire coding region and 10 bp 3' noncoding.

H1 Probe: 1.3 kb containing 60 bp 5' noncoding the entire coding region and 680 bp 3' noncoding. Isolated from λCH.02 by A. Robins.

2.2 METHODS

2.2.1 Isolation of Chicken Histone mRNA

All procedures involving RNA were carried out at 0^{o} using sterile solutions and glassware.

RNA was isolated essentially by the method of Seeburg *et al.*, (1977). Approximately 30 five-day-old chick embryos were snap-frozen in liquid nitrogen, then homogenised in 7 M guanidinium-Cl, 20 mM Tris-Cl 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 at 27,000 rpm for 16 h at 15^{o} in a SW41 rotor. 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-Cl, pH 7.5,1 mM EDTA, 0.5% SDS, heated at 65^{o} for 5 minutes, chilled and centrifuged on a 10-40% sucrose gradient in 10 mM Tris-Cl pH 7.5, 1 mM EDTA, at 37,000 rpm for 16 hours at 4^{o}. The 7-11s RNA was collected and then refractionated on a second 10-40% sucrose gradient.

2.2.2 Preparation of High Molecular Weight DNA

Freshly prepared human placental tissue was snap-frozen in liquid nitrogen. A sample of the frozen tissue was fragmented and homogenised at 4^{o} in 50 ml of homogenising solution {11% (w/v) sucrose, 2 mM EDTA, 0.5 mM EGTA, 1 vol. 10 x Buffer A (0.6 M KCl, 0.15 M NaCl, 1.5 mM Spermine, 5 mM Spermidine, 0.15 M Tris pH 7.8)} with a B-dounce homogenising pestle until the tissue was dispersed into a gravy.

The suspension was poured through muslin gauze and overlayed on top of 5 mls of (50% (w/v) sucrose, 1 mM EDTA, 0.5 mM EGTA, 1 vol. of 10 x Buffer A) in a 50 ml centrifuge tube, and centrifuged for 20 min at 4^{o} in an SS34 rotor at 16,000 g.

The supernatant was discarded and the nuclear pellet resuspended in 10 ml of NK M (0.15 M NaCl, 5 mM KCl, 2 mM $MgCL_2$) and added dropwise over 5 h to 100 ml of gently stirring PKM (10 mM NaCl, 10 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.5% SDS) and 100 µg/ml proteinase K at 37^{o}. The suspension was incubated overnight and a further 100 µg/ml proteinase K was added.

An equal volume of buffer saturated phenol/chloroform (1:1) was added, mixed and centrifuged for 10 min at 10,000 g. The aqueous phase was re-extracted with an equal volume of phenol/chloroform re-centrifuged and finally extracted with chloroform. The aqueous phase was dialysed against (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) with three changes of buffer in 2 litres for 24 h. The solution was then incubated with 20 µg/ml pancreatic RNA'ase (previously heat-treated at 80^{o} for 20 min. to destroy DNA'ase) at 37^{o} for 4 h, and phenol/chloroform extracted twice as before. The solution was then dialysed against (10 mM Tris-HCl pH 7.5, 1 mM EDTA) with 3 changes of buffer in 2 litres for 24 h.

2.2.3 Construction of a Genomal Library

(a) RI Methylase and EcoR1 Restriction Enzyme Preparation

Both enzymes were isolated from a 8 litre culture of *E. coli* 2012. The method of Greene *et al.*, (1974) was followed through the streptomycin sulphate and ammonium sulphate precipitation steps, except that cells were initially disrupted by two passages through a French pressure cell at 1200 p.s.i. Further purification was then as described below.

A 4 x 10 cm column was poured of phosphocellulose and equilibrated at 4^{o} with EB + 0.2 M NaCl (EB = 10 mM KH_2PO_4/K_2HPO_4, pH 7.0, 7 mM β-mercaptoethanol, 1mM EDTA) and washed at gravity pressure with 100 mls EB + 0.2 M NaCl. The dialysed ammonium sulphate precipitate was applied to the column and washed with 200 mls of EB + 0.2 M NaCl. A 400 ml gradient of 0.2 M NaCl to 0.8 M NaCl in EB was applied and 5 ml fractions collected at 1 ml/minute. Fractions were assayed and those with enzyme activity pooled.

EcoR1 restriction enzyme was further purified by blue sepharose chromatography, basically as described by Baski *et al.* (1978). Blue sepharose was washed with guanidinium-HCl and water and then equilibrated with buffer A (20 mM Tris-HCl, pH 7.6, 20 mM $MgCl_2$, 10 mM β-mercaptoethanol) and a 1.8 cm column prepared. The active EcoR1 fraction was loaded and washed with 100 mls of buffer A. A 100 ml gradient from 0 to 0.5 M NaCl in buffer A was applied, 2 ml fractions collected, and the active fractions pooled and stored at -20^{o} in 50% glycerol.

(b) R1 Methylase Assay

Assay reactions contained 1 µg of λDNA incubated in 100 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1.1 µM {^{3}H}-S-adenosyl-methionine (SAM) (250 Ci/m mole) with aliquots of methylase enzyme at 37^o for 30 minutes. Reactions were stopped with 1 ml of cold 7.5% perchloric acid and 80 µg of carrier coalfish DNA, left on ice for 30 minutes and the acid precipitates collected on GF/A glass fibre filters. After washing with 10 ml of cold perchloric acid, 8 ml of cold 2M HCl and 5 ml of 95% ethanol, the glass filters were dried and radioactivity counted in Triton scintillation fluid.

(c) R1 Methylase Incubation

DNA was incubated at a concentration of 100 µg/ml with 100 mM Tris HCl, pH 8.0, 10 mM EDTA, 120 µM SAM/100 µg and 100 units of R1 methylase/ml at 37^o for 1 hour.

(d) EcoR1* Incubation

DNA was incubated at a concentration of 100 µg/ml with 25 mM Tris-HCl, pH 8.5, 2 mM $MgCl_2$ and 1000 units of EcoR1 per ml at 37^o for 1 hour.

(e) Preparation of Charon 4A Arms

Charon 4A DNA was pre-annealed in 200 mM Tris-HCl, pH 8.0, 100 mM NaCl and 20 mM $MgCl_2$, by incubation at 45^o for 3 hours. The volume was increased with water to double and incubated with EcoR1 (1 unit/µg DNA) for 90 minutes at 37^o. The annealed Charon 4A arms (19.5 kb and 11.1 kb) were separated from the two EcoR1 internal fragments by 10-40% sucrose gradient centrifugation at 34,000 rpm for 16 hours at 4^o in an SW41 rotor.

(f) Ligation Conditions

The Charon 4A DNA and the EcoR1* human DNA were incubated in a 2:1 molar ratio in 65 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 0.1 mM ATP, 1 mM EDTA, 40 mM NaCl and T4 DNA ligase (1 unit/µg) at 8^o for 24 hours.

(g) Growth and Maintenance of Bacterial Strains for *In vitro* Packaging

The bacterial cultures used were:

(i) NS428:N205 (λAam11 b2 red 3 cIts 857 Sam7)

(ii) NS433:N205 (λEam4 b2 red 3 cIts 857 Sam7)

Extreme care was taken to ensure the purity and viability of the bacterial strains used for *in vitro* packaging. Upon receipt of the strains, single colonies were prepared by streaking the cultures out on TB plates (10 g Tryptone, 5 g NaCl, 11 g agar/1) and incubating them at 32^o for about 24 h or until colonies were visible. The incubation temperature was critical, because the strains are λ lysogens and carry a thermoinducible prophage. Growth at intermediate temperatures (e.g. 34^o - 36^o) can inevitably lead to loss of the prophage, cell death and selection of unwanted mutants. Glycerol stocks were prepared every 2 or 3 months. To verify that the cultures still contained a thermoinducible prophage, loopfuls of cells were streaked on two TB plates and incubated, one at 32^o and the other at 42^o. Thick growth appeared on the 32^o plate after overnight incubation but not on the 42^o plate.

(h) Preparation of Packaging Extracts

Packaging extracts were prepared from these strains basically by the method of Sternberg *et al.*, (1977).

Extract A

E. coli NS428 and NS433 were each inoculated into 5 ml of Sup 9 broth (6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl, 20 g casamino acids, pH 7.4/1,000mls, autoclaved then supplemented/100 mls with 1 ml 20% glucose 0.3ml 1M $MgSO_4$, 25 µl 0.4 M $CaCL_2$) and incubated overnight at 37°. Each *E. coli* strain was inoculated into 150 ml of Sup 9 to a final OD_{600} of 0.025 and incubated at 32° for 4 h to give a final OD_{600} of 0.25. At this absorbance, the cultures were induced at 42° for 30 min, then incubated at 37° for 70 min.

The cells were harvested after mixing both cultures by centrifuging at 5,000 g for 10 min in the JA-10 rotor. The supernatant was drawn off and the pellet resuspended in 0.6 ml buffer A (20 mM Tris-HCl, pH 8.0, 1 mM EDTA, 3 mM $MgCl_2$, 5 mM β-Mercaptoethanol) and sonicated using the microtip on the lowest setting for 12 x 2 sec bursts at 4°. Glycerol was added to 20% and stored in 20 µl aliquots at -80° .

Extract B

E. coli 428 was inoculated into 2 flasks of 300 mls Sup 9 to an OD_{600} of 0.025, and incubated at 32° for approximately 4 h (until the OD_{600} had reached 0.25). The culture was then induced at 42° for 30 min and incubated for a further 70 min at 37°. The cells were harvested by centrifuging at 5,000 g for 10 min in the JA-10 rotor.

The supernatant was drawn off and the pellet resuspended in 1.2 ml of 10% sucrose, 50 mM Tris-HCl, pH 7.4, and 90 µl samples were aliquoted into Eppendorf tubes and snap-frozen in liquid nitrogen. The tubes were thawed, refrozen and thawed again, then 5 µl of lysozyme (1 mg/ml in 0.25 M Tris-HCl, pH 7.4) was added per aliquot, and incubated at 4^{o} for 30 min. To each aliquot was added 10 µl of buffer B (6 mM Tris-HCl, pH 7.4, 15 mM ATP, 16 mM $MgCl_2$, 60 nM spermidine, 30 mM β-Mercaptoethanol) and 30 µl of glycerol prior to storage at -80^{o}.

(i) *In vitro* Packaging

At toom temperature in an Eppendorf tube 30 µl of buffer A, 2 µl of buffer B, 20 µl of freshly thawed extract A and 5 µl of DNA (5 µg/µl) were mixed in that order. After incubation at room temperature for 15 min the mix was added to a freshly thawed tube of extract B and stirred thoroughly with a 10 µl glass micropipette. After incubation for 60 min at 37^{o}, 150 µl of DNAse buffer (100 mM NaCl, 10 mM $MgSO_4$, 50 mM Tris-HCl, pH 7.5, 0.01% gelatin, 10 µg/ml DNAse I) was added and incubation continued for 15 min with occasional stirring. Debris was removed by a 5 min centrifugation in a microfuge and packaged DNA stored as a conventional phage lysate over 2 or 3 drops of chloroform.

2.2.4 Amplification of Phage

(a) The Plate Lysate Method

A phage plaque containing 10^5 to 10^6 phage was added to 1 ml of PSB (0.1 M NaCl, 10 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$), and 1 drop of a fresh saturated culture of LE392 grown overnight in L broth. After 5 min adsorption at 37^o, 2 ml of L broth followed by 2 ml of 0.7% top agar was added and the mix poured onto L plates. The overlay was allowed to solidify and the plates then incubated at 37^o until the bacterial lawn cleared, usually 5-7 hours. The top agar overlay was removed and mixed with 10 mls of PSB and 1 drop of chloroform to kill any surviving bacteria. The lysate was placed on ice for 10 minutes, and the remaining agar removed by centrifugation in an SS34 Sorvall rotor at 5,000 rpm for 10 minutes at 4^o. Supernatants contained 10^{10} to 10^{11} phage/ml.

(b) The Liquid Lysate Method

The procedure was the same for the plate lysate method through the phage adsorption step. After adsorption the infected cells were diluted 200 fold into L broth and vigorously aerated at 37^o until lysis occurred, usually 4-6 hours. Lysis was completed by shaking the culture an additional 5-10 minutes with several drops of chloroform. Debris was removed by centrifugation, e.g. 10,000 rpm for 10 minutes in the Sorvall SS34 rotor.

2.2.5 Detection of Recombinant Phage

Basically the method of Benton and Davis (1977) was employed. Chilled agar plates containing phage were overlayed with circles of nitrocellulose and the filters oriented by ^{32}P indian ink markers.

Each filter was removed and incubated with 0.5 M NaOH, 1.5 M NaCl for 1 minute, followed by 0.5 M Tris-HCl, pH 7.4, 1.5 M NaCl for 1 min and baked at 80° in a vacuum oven for 2 hours. Hybridization and washing conditions were as described for Southern blot experiments.

2.2.6 Preparation of Phage DNA

(a) The PEG Method

The phage lysate prepared by the plate or liquid method (500 ml) was added to 50 g PEG, 9.5 g NaCl and stirred overnight at 4°, then 50 µl 10 mg/ml RNAase A and 25 µl 5 mg/ml DNAase I were added and incubated at 4° for 5-12 hours. The precipitate was centrifuged at 10,000 rpm for 15 mintues in a SS34 rotor. The pellet was resuspended in 1 ml of PSB and adjusted to 0.5 g CsCl/ml. This was overlayed on a 4 ml step gradient of 1.6 g/ml and 1.4 g/ml of CsCl in PSB and centrifuged at 30,000 rpm for 120 min in a Ti50 rotor.

The phage band was removed with a needle and syringe, adjusted to 1.75 g/ml with CsCl and overlayed with 2 ml of 1.6 g/ml and 2 ml of 1.4 g/ml of CsCl in PSB. After centrifugation for 2 hours at 30,000 rpm in a Ti50 rotor, the phage was again removed and dialysed against 3 litres of 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, overnight with 3 changes.

The dialysed phage were then phenol/chloroform (1:1) extracted, and the aqueous phase ether extracted and ethanol precipitated. The λDNA was dissolved in water and stored at -20°.

(b) The High Speed Spin Method

The phage lysate was centrifuged at 48,000 rpm for 1 hour in a Ti50 rotor and the phage pellet dissolved in PSB, and made 0.5% SDS and 50 mM EDTA to lyse the phage. An equal volume of phenol/chloroform (1:1) was added and the aqueous phase extracted twice, then ethanol precipitated. The λ phage DNA was dissolved in water and stored at -20°.

(c) DNA Mini-preparation of Phage DNA

The phage lysate (2 ml) was added to 0.4 ml SDS mix (0.25 M EDTA, 0.5 M Tris-HCl pH 9.0, 2.5% SDS) and heated at 65° for 30 minutes. Then 0.5 ml of 8 M KAc was added, chilled at 4° for 15 minutes and centrifuged at 17,000 rpm for 10 minutes. The supernatant was removed, 5.6 ml of ethanol was added and after 20 minutes at -80° it was centrifuged again. The precipitate was dissolved in 0.4 ml 2 M Ammonium acetate and mixed with 0.8 ml ethanol. After 20 minutes at -80°, the solution was centrifuged and the precipitate was dissolved in 50 µl water and stored at -20°.

2.2.7 Construction of Recombinant Plasmids

(a) Phosphatase-treatment of the Vector DNA

In order to prevent self-ligation, the restricted (EcoR1 or HindIII) plasmid vector (10 µg) was incubated with 0.2 units of bacterial alkaline phosphatase (BAPF) (previously dialysed against 25 mM Tris-HCl, pH 8.0) at 65° in a 100 µl reaction mix containing 25 mM Tris-HCl, pH 8.0, 7 mM $MgCl_2$, for 30 minutes. The reaction mix was then made 5 mM with EDTA, phenol/chloroform extracted three times, ethanol precipitated, and redissolved in 20 µl of water.

(b) Ligation of Restriction Fragments to Plasmid DNA

DNA restriction fragments were ligated to the dephosphorylated, linearised plasmid vector DNA in a 1:1 molar ratio in 10 µl reaction containing 60 mM Tris-HCl, pH 7.5, 9 mM $MgCl_2$, 15 mM dithiothreitol, 1 mM ATP, and 0.5 units of T4 ligase at 10^{o} for 16 hours.

2.2.8 Transformation and Selection of Recombinant Plasmids

(a) Transformation

E. coli strain ED8654: r^-_k m^+_k *sup*E *sup*F *trp*R was grown overnight at 37^{o} 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 ½ 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 cells were kept on ice for at least one hour. 0.2 ml of these competent cells was added to 0.1 ml of the ligated DNA in 0.1 M Tris-HCl pH 7.4, and stored on ice for 30 minutes, with occasional stirring. The cells were heat-shocked at 42^{o} for 2 minutes, returned to ice for 30 minutes and then warmed slowly to room temperature. 0.5 ml of L broth was added to the transformed cells and they were then incubated at 37^{o} for 20 to 30 minutes. The transformed cells were mixed with 3 mls of 0.7% L-agar and plated on 1.5% L-agar plates containing 20 µg/ml ampicillin. These were incubated overnight at 37^{o}.

(b) Selection of Recombinants

Basically the method of Grunstein and Hogness, (1975). Colonies from a transformation were transferred by toothpick to a master plate and to a sheet of nitrocellulose that had been boiled

three times in distilled water and lain onto an L-agar plate with 20 µg/ml ampicillin. The colonies were grown overnight on the nitrocellulose at 37°C, and the colonies lysed by transferring the nitrocellulose sequentially onto 3 MM paper saturated with 0.5 N NaOH for 7 minutes, 1 M Tris-HCl, pH 7.4 twice for 2 minutes, 0.5 M Tris-HCl, pH 7.4, 1.5 M NaCl for 4 mintues. The nitrocellulose filter was washed in 95% ethanol and then baked at 80°C, under vacuum for 2 hours. Hybridisation and washing conditions were as described for Southern blot experiments.

2.2.9 Preparation of Plasmid DNA

(a) Large-scale CsCl Fractionation

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 µg/ml. The cells were incubated overnight at 37°C to allow amplification of the plasmid DNA (Clewell, 1972). The cells were pelleted by centrifugation, (10,000 g for 5 minutes), and resuspended in 10 mls of 15% sucrose, 50 mM Tris-HCl, pH 9.0, 50 mM EDTA.
2 mls of 6 mg/ml lysozyme solution was added and the cells were incubated at room temperature for 15 minutes, followed by 30 minutes on ice.
A 15 ml solution of 0.1% Triton X-100, 50 mM EDTA, 50 mM Tris-HCl, pH 8.0 was added, mixed gently and centrifuged at 30,000 g for 30 mins. The supernatant was removed and treated with DNAase-free RNAase A (20 µg/ml), for 30 minutes at 37°C and then proteinase K, (50 µg/ml), for 30 minutes at 37°C. The solution was extracted with an equal volume of phenol/chloroform and the aqueous phase ethanol precipitated.

The plasmid DNA was resuspended in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA (TE buffer) and one gram of CsCl was added per ml of DNA solution. One tenth the volume of ethidium bromide (10 mg/ml) was added and the mixture centrifuged at 200,000 g at 15^{o} for 48 hours. Both bands, identified by fluorescence in ultraviolet light, were removed by side puncture with a syringe, extracted five times with isoamyl alcohol (saturated with 1 g/ml CsCl solution) to remove ethidium bromide, then dialysed against TE (3 changes, overnight) then 0.1 x TE. The lower band contained supercoiled DNA. Plasmid DNA was stored at 4^{o} in 0.1 x TE.

(b) Large Scale PEG Precipitation

The procedure described in an adaption of the alkali lysis procedure of Birnboim and Doly (1979) was used. 500 ml cultures of cells as in part (a) were pelleted by centrifugation (10,000 g for 5 minutes) and resuspended in 4 mls of 25 mM Tris-HCl, pH 8.0, 10 mM EDTA, 15% sucrose and 2 mg/ml lysozyme per 500 ml culture, and incubated on ice for 20 to 40 minutes. 0.8 mls of 0.2 M NaOH, 1% SDS was added, gently mixed and incubated on ice for 10 min; then 5 ml of 3 M sodium acetate, pH 4.6, was added and incubated for 40 minutes on ice. The lysate was centrifuged at 30,000 g for 30 minutes, the supernatant removed and treated with DNAase-free RNAase A, (20 µg/ml) for 30 minutes and proteinase K (50 µg/ml for 30 minutes at 37^{o}. The solution was extracted with an equal volume of phenol/chloroform and the aqueous phase ethanol precipitated. The precipitate was resuspended in 1.6 mls of water and 0.4 ml of 4 M NaCl, 2 mls of 13% PEG were added and incubated on ice for 1 hour. After centrifugation at 10,000 g for 10 minutes the DNA pellet was washed in 70% (v/v) ethanol and resuspended in 0.1 x TE and stored at 4^{o}.

(c) Mini-plasmid DNA Preparation

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. The cells were pelleted by centrifugation for 30 seconds in an Eppendorf centrifuge, washed in 10 mM Tris-HCl, pH 9.0, 1 mM EDTA, pelleted again and resuspended in 150 µl of 15% sucrose, 50 mM Tris-HCl, pH 9.0, and 50 mM EDTA. 50 µl of freshly prepared 4 mg/ml lysozyme solution was added and the solution incubated at room temperature for 15 mintues and at 0^{o} for 30 minutes. 200 µl of ice-cold water was added and the solution heated at 72^{o} for 15 minutes. The solution was centrifuged at 30,000 g for 20 minutes and the supernatant removed from the white filamentous pellet and ethanol precipitated. The plasmid DNA was resuspended in water and 1/10th volume used for restriction enzyme analysis.

2.2.10 Restriction Endonuclease Digestion

All restriction endonuclease digestions were performed according to the conditions of the supplier (New England Biolabs or Boehringer-Mannheim).

DNA was dissolved in water and 0.1 volume of 10 x HaeIII buffer (66 mM Tris-HCl, pH 7.5, 66 mM $MgCl_2$, 66 mM β-mercaptoethanol) added. The concentrations of NaCl and KCl were adjusted to those recommended for the particular restriction endonuclease, as was the temperature of incubation (usually 37^{o}). Sufficient units of enzyme activity were added so as to give complete digestion, as judged by pilot experiments using λ or pBR322 DNA. All reactions were stopped by the addition of EDTA to a final concentration of 25 mM. The reaction mix

was extracted with an equal volume of phenol/chloroform, (1:1) and the aqueous phase ethanol precipitated.

2.2.11 Routine Gel Electrophoresis

(a) Agarose Gel Electrophoresis

Electrophoresis of DNA for analytical purposes or for transfer to nitrocellulose was carried out on 14 cm x 14 cm x 0.3 cm slab gels containing 0.8% or 1% agarose. Electrophoresis buffer consisted of 40 mM Tris-acetate pH 8.2, 20 mM sodium acetate, 1 mM EDTA and electrophoresis was carried out at 60 mA for about 3 hours. DNA was visualised by staining with 10 µg/ml ethidium bromide solution for 15 minutes and examination under ultraviolet light.

Samples were dissolved in 20 µl water, 5 µl of loading buffer (25% (v/v) glycerol, 5% (w/v) sodium dodecyl sulphate, 25 mM EDTA, 0.05% (w/v) bromophenol blue) added and loaded directly into gel slots (4 mm or 9 mm wide).

(b) Polyacrylamide Gel Electrophoresis

Electrophoresis of DNA species less than 1 kb in length was carried out on vertical 14 cm x 14 cm x 0.2 cm gels containing between 4% and 8% acrylamide, using a 19:1 ratio of acrylamide to N, N'-methylenebisacrylamide.

Samples were dissolved in 20 µl water, 5 µl of 5 x loading buffer (50% (w/v) sucrose, 4 mM EDTA, 0.05% (w/v) xylene cyanol FF, 0.05% (w/v) bromophenol blue; which included 0.5% Sarkosyl if protein was in the sample).

(TBE) Electrophoresis buffer was 90 mM Tris-borate pH 8.3, 2.5 mM EDTA and electrophoresis was performed at 150 V for 2 hours. DNA was visualised by ethidium bromide staining.

(c) Strand-separating Polyacrylamide Gels

DNA strands were separated by heating at 90°C for 2 minutes, in 40 µl of 30% (v/v) dimethylsulphoxide, 1 mM EDTA, pH 8.0, 0.05% xylene cyanol FF, 0.05% bromophenol blue and electrophoresis on a 5% polyacrylamide gel with a 50:1 acrylamide to bis-acrylamide ratio, run at 100 volts in TBE. Radioactively labelled DNA was cut out and eluted as described in (2.2.13).

2.2.12 Elution of DNA from Agarose Gels

(a) Freeze-thaw Method

Slices of agarose gel containing DNA fragments were crushed, frozen at -80°, thawed and centrifuged at 40,000 rpm for 1 h in a Ti50 rotor. The supernatant was ethanol precipitated twice and redissolved in water for use in ligation and restriction enzyme experiments.

(b) Low Melting Point Method

Slices of low melting point agarose containing DNA fragments were heated at 60° for 10 minutes and an equal volume of 100 mM Tris-HCl pH 8.0, 1 mM EDTA added. The DNA was recovered in the aqueous phase following three phenol extractions and two chloroform extractions and then ethanol precipitated twice.

2.2.13 Elution of DNA from Polyacrylamide Gels

(a) Electroelution

Slices of polyacrylamide gel containing DNA fragments were placed in 18/32 dialysis tubing (previously boilded in EDTA/$NaHCO_3$ then washed and boiled in water) with 0.8 ml of electroelution buffer (25 mM Tris-borate, pH 8.3, 0.5 mM EDTA), sealed by clamping and placed in an electroelution chamber containing 500 ml of ½ x TBE, between the electrodes. A current of 50 mA (~50 volts) was applied overnight to electrophorese the DNA out of the gel slice into the buffer.

The buffer was removed from the dialysis tubing and the DNA ethanol precipitated by addition of 0.05 volume of 5 M $(NH_4)_2SO_4$ and 2.5 volumes of ethanol and storage at -80^{o} for 1 hour.

(b) Ammonium Acetate Elution

Slices of polyacrylamide gel containing DNA fragments were immersed in 500 mM Ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% SDS, and incubated at 37^{o}, shaking, overnight. The supernatant was ethanol precipitated, redissolved in 300 µl of 300 mM sodium acetate and ethanol precipitated again. The DNA was resuspended in water and used in restriction or sequence analysis.

2.2.14 *In vitro* Synthesis of Labelled DNA

(a) Oligo-dT-primed Reverse Transcription

Oligo-dT-primed reverse transcription was carried out in a 20 µl reaction mix containing up to 2 µg of mRNA, 1 mM each of dATP, dGTP, dTTP, about 5 µM α-{^{32}P}-dCTP, (500 Ci/m mole), 50 mM Tris-HCl, pH 8.3, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol and 20 µg/ml of oligo-dT10. Reverse transcriptase (1 µl = 17 units) was added and the solution was incubated at 42^{o} for 30 minutes. The RNA template was removed by alkaline hydrolysis with 0.3 N NaOH for 30 minutes at 60^{o}C and the solution neutralised by the addition of HCl to 0.3 M and Tris-HCl pH 7.5 to 0.1 M. The mix was extracted with an equal volume of phenol:chloroform (1:1) and the aqueous phase loaded onto a 0.4 cm x 4 cm sephadex G-50 column and eluted with 10 mM Tris-HCl, pH 7.6, 1 mM EDTA, to remove the unincorporated nucleotides.

(b) Random-primed Reverse Transcription

Priming of RNA lacking a 3' polyA tract was achieved by the random hybridisation of oligo-nucleotides of salmon-sperm DNA, prepared as described by Taylor *et al.*, 1976). Conditions for the synthesis of this cDNA were exactly as described for the oligo-dT-primed reaction, except that oligo-dT_{10} was replaced by a 2 mg/ml final concentration of oligo-nucleotide, and the mix was incubated at 37^{o} for 60 minutes. The cDNA synthesised was isolated as described for oligo-dT-primed synthesis.

(c) 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 µl incubation mix contained 50 mM Tris-HCl pH 7.8, 5 mM $MgCl_2$

10 mM 2-mercaptoethanol, 50 µg/ml bovine serum albumin, 2 µM each of α-{^{32}P}-dCTP and α-{^{32}P}-dGTP (500 Ci/m mole) and unlabelled dATP and dTTP at 50 µM. The DNA (1 µg) was nicked by the addition of 20 pg of *E. coli* DNAase I and the reaction was started by the addition of 2 units of *E. coli* polymerase I. The solution was incubated at 15°C for 90 minutes, phenol/chloroform extracted and the unincorporated nucleotides removed as described for oligo-dT-primed reverse transcription. If the labelled DNA was to be used as a hybridisation probe, the DNA strands were separated by boiling the solution for 2 minutes and then snap-cooling.

(d) End-filling Extended 5' Termini

To label restricted DNA with 5' ends protruding, DNA was incubated in 60 mM NaCl, 60 mM Tris-HCl, pH 7.5, 6 mM $MgCl_2$ with 5 µM each of α-{^{32}P}-dCTP and α-{^{32}P}-dATP (500 Ci/m mole) and unlabelled dATP and dTTP at 50 µM (depending on the nucleotide requirements of the restricted fragments) and 0.1 unit of the Klenow fragment of DNA polymerase at 37° for 30 minutes. The solution was phenol/chloroform extracted and the unincorporated counts were either removed by agarose gel electrophoresis or by G-50 Sephadex chromatography.

(e) Labelling of Blunt or 3' Extended Termini

DNA was radioactively labelled at 3' ends and blunt ends using the 3' to 5' exonuclease, exonuclease III, and the Klenow fragment of DNA polymerase I with α-{^{32}P}-deoxynucleoside triphosphates. Incubation with 1 unit/µg of the exonuclease for 3 minutes at 25°C in 60 mM Tris-HCl, pH 7.5, 6 mM $MgCl_2$, 60 mM NaCl, results in the removal

of ~10 nucleotides from the 3' ends of the DNA. The reaction is then immediately stopped by heating at 60° for 5 minutes, snap-chilled, and labelled as described above for end-filling of extended 5' termini.

(f) Labelling 5' ends with γ-{^{32}P}-ATP and T4 Polynucleotide Kinase

The 5' terminal phosphate groups were removed from DNA (1-10 µg) in a 100 µl reaction mix containing 25 mM Tris-HCl, pH 8.0, 7 mM $MgCl_2$, and 0.2 units of bacterial alkaline phosphatase (BAPF) (previously dialysed against 25 mM Tris-HCl, pH 8.0) at 65°C for 30 minutes. EDTA, to 5 mM, was added and the reaction mix phenol extracted and ethanol precipitated.

20 µl of γ-{^{32}P}-ATP (1000-1500 Ci/m mole, about 200 µCi) was evaporated to dryness and dissolved in 60 mM Tris-HCl, pH 7.5, 9 mM $MgCl_2$, 15 mM dithiotheitol, 1 mM spermidine. Dephosphorylated DNA was then dissolved in this mixture, 0.5 µl of T4 polynucleotide kinase (5 units/µl) added and the reaction mixture incubated at 37°C for 45 minutes, phenol extracted and ethanol precipitated and resuspended in 20 µl of TE buffer.

2.2.15 DNA Blot Analysis

(a) Transfer of DNA from Gels

Restricted DNA fractionated on 1% agarose slab gels was transferred to nitrocellulose filter paper using the method of Southern, (1975), as modified by Wahl *et al.*, (1979). In this procedure the rapid transfer of DNA from the gel to the nitrocellulose is facilitated by a partial hydrolysis of the DNA in the gel with

0.25 N HCl, and the transfer is complete within about 2 hours.

Bidirectional transfer from both agarose and acrylamide gels basically followed the method of Smith and Summers (1980) in which the gel is neutralised in 1 M Ammonium acetate, 20 mM NaOH and nitrocellulose is placed directly above and below the gel; transfer is complete from agarose gels in 1 hour and from acrylamide gels in 24 hours.

(b) Hybridisation on Nitrocellulose Filters

After transfer the nitrocellulose filters were baked at 80°C, under vacuum, and prehybridised in 50% formamide, 5 x SSC (1 x SSC, 0.15 M NaCl, 15 mM sodium citrate, pH 7.6), 5 x Denhardt's solution (0.1% ficoll, 0.1% polyvinyl pyrrolidone, 0.1% bovine serum albumin), 1% glycine, 50 mM potassium phosphate, 200 µg/ml *E. coli* DNA (sonicated and heat-denatured) in plastic bags at 42°, for 1-2 hours. The filters were then hybridised with labelled DNA probe in 50% formamide, 5 x SSC, 1 x Denhardt's solution, 20 mM potassium phosphate, 10% dextran sulphate, 100 µg/ml *E. coli* DNA, sealed in plastic at 42° overnight. The filters were washed several times at room temperature in 2 x SSC, 0.1% SDS and then at 65° with several more washes. The washed, dried nitrocellulose filter was placed in contact with X-ray film and exposed at -80° in the presence of an intensifying screen.

2.2.16 Sequence Analysis of DNA

(a) Maxam and Gilbert - Sequencing of End-Labelled DNA with Base-specific Chemical Cleavages

All DNA fragment preparation and DNA-sequencing procedures were carried out exactly as described by Maxam and Gilbert (1977, 1979),

with the following exceptions. The (G+A)-specific chain cleavage was carried out by adding to 10 µl of water containing the labelled DNA, 25 µl of 100% formic acid and incubating at room temperature for two minutes. The reaction was stopped by addition of 250 µl of 0.3 M sodium acetate, pH 6, 0.1 mM EDTA, 25 µg/ml tRNA. The C and (C+T) stop reactions were carried out by adding 200 µl of 0.1 mM EDTA, 25 µg/ml tRNA and 200 µl of 0.3 M NaCl, 0.1 mM EDTA, 25 µg/ml tRNA respectively, (Busslinger *et al.*, 1979). Subsequent ethanol precipitations and piperidine cleavage reactions were done as described by Maxam and Gilbert (1979).

Products of the chemical degradation sequencing reactions were separated by electrophoresis on polyacrylamide gels which include 7M urea as a denaturant. Initially 1.5 mM thick gels were used, but these were later replaced by 0.5 mm thick gels giving increased resolution. For 20% gels a 100 ml mixture containing 90 mM Tris-borate, 2.5 mM EDTA, 19 g of acrylamide, 1 g of N,N'-methylene bisacrylamide, 42 g of urea, 1 ml of 10% (w/v) ammonium persulphate was filtered, 20 µl TEMED added and poured into a 30 x 40 cm gel mould, and allowed to polymerise for 4 hours. 10% and 8% gels were constructed in the same manner with correspondingly reduced amounts of acrylamide and bisacrylamide.

Samples, dissolved in deionized formamide containing 1 mM EDTA, 0.05% Bromophenol blue and 0.05% Xylene cyanol, were heated to 90^{o} for 2 minutes then chilled on ice before loading. Reservoir tanks contained approximately 2 litres of electrophoresis buffer and the gels were pre-electrophoresed for 1 to 4 hours. All gels were run at 1000-2000 volts. Gels were autoradiographed at -80^{o}C with an intensifying screen in an Ilford autoradiography cassette.

(b) DNA Sequence Analysis by Primed Synthesis

The chain termination sequencing procedure used was basically that of Sanger *et al.*, (1977), with modifications from G. Winter (personal communication).

Sequencing a region of DNA by the chain termination method requires both a single-stranded template and a primer complementary to the 3' end of the region. The DNA to be sequenced is digested with restriction enzymes and the fragments of DNA cloned into the unique restriction sites of the replicative form of M13 mp7, mp8 or mp9 (Messing *et al.*, 1981). The R.F. is then used to transform a male host, JM101, and the transformed cells are mixed with fresh cells and spread on an agar plate. The transformed cells excrete phage and form slow-growing plaques. To harvest the recombinant phage, each white plaque (complementation assay) is tooth picked into fresh medium and grown as a small culture. The phage are excreted into the medium and the single-stranded recombinant DNA can be used as template for sequencing.

(i) Host Strain

JM101: (lac, pro, supE, thi, F', traD36, proAB, $lacI^q$, z deltaM15) was used as the host strain. Because M13 is a male-specific phage and thus will not infect F^- cells, JM101 was grown on minimal media (+ glucose), so as to maintain the F episome.

(ii) M13 Phage

The phage strains used contain synthetic DNA sequences in a segment of the β-galactosidase gene such that introduction of a foreign DNA fragment will interrupt the protein coding sequences and

lack of enzyme activity can be detected as a clear (rather than blue) plaque on indicator plates. The restriction sites contained in the M13 strains are as follows:

M13 mp7 5' EcoRl,BamHl,AccI,PstI,AccI,BamHI,EcoRI, priming site 3'

mp8 5' EcoRl,SmaI,BamHI,AccI,PstI,HindIII, priming site 3'

mp9 5' HindIII,PstI,AccI,BamHI,SmaI,EcoRl, priming site 3'

(iii) Preparation of M13 Replicative Form DNA

A 5 ml culture of JM101 was grown to an A_{600} of 0.6 in 2 x YT (5 g NaCl, 10 g Tryptone, 8 g yeast in 1 litre) at 37^{o} by inoculating directly from a minimal plate. To 3 mls of molten 2 x YT agar, 30 µl of BCIG (5-bromo-4-chloro-3-indoyl-beta-galactoside) (20 mg/ml in dimethyl formamide) and IPTG (isopropyl-beta-D-thio-galactopyranoside) (20 mg/ml in water) was added, followed by 0.2 ml of JM101 cells and 0.1 ml of dilute M13 phage. This was poured onto a minimal (+ glucose) plate and incubated at 37^{o} overnight. A blue plaque was tooth picked into 1 ml of 2 x YT and grown with shaking for 6 hours. Meanwhile a 10 ml culture of JM101 was grown to an A_{600} of 0.5 from a single colony on a minimal glucose plate. This was added to 1 litre of 2 x YT and when A_{600} of 0.5 was reached, 1 ml of phage was added and grown for 4 hours. The culture was centrifuged at 5,000 rpm for 10 minutes and the cell pellet was used to isolate the replicative form as described in plasmid DNA procedures.

(iv) Ligation and Transformation

A three-fold molar excess of restricted DNA fragments over the linearised M13 vector was used in the following

10 µl ligation mix (2 µl of 10 ng/µl M13, 1 µl DNA fragments, 1 µl 10 mM ATP, 1 µl 10 x C (500 mM NaCl, 10 mM Tris-Cl, pH 8.0, 60 mM $MgCl_2$), 4 µl water, 1 µl ligase, 2.5 Weiss units). Sticky end ligations were incubated at 14° for 3-12 hours and blunt end ligations for 24 hours.

A colony of JM101 from a minimal glucose plate was grown up in 20 ml of 2 x YT for 4-5 hours at 37°C with shaking. When A_{600} reached 0.3, the cells were centrifuged at 5,000 rpm for 10 minutes and resuspended in 10 mls of cold 50 mM $CaCl_2$ and kept at 4°C for 1-24 hours.

The ligation mix was added to 0.2 ml of these competent cells and after 40 minutes on ice, the cells were heated at 42°C for 2 minutes, and added directly to 3 mls of 2 x YT molten agar at 42°C to which 30 µl BCIG (20 mg/ml), 20 µl IPTG (20 mg/ml) and 0.2 ml exponential JM101, in that order, had been added. This was plated onto minimal,(+ glucose) plates and incubated at 37°C overnight.

(v) Single-Stranded M13 DNA Preparation

White plaques were tooth-picked into 1 ml aliquots of 2 x YT (diluted 1/50 with a saturated culture of JM101) and incubated at 37°C, shaking, for 5-6 hours. After centrifugation in an Eppendorf tube and 200 µl of 2.5 M NaCl, 20% PEG6000 was added. After 15 minutes at room temperature, the PEG precipitate was centrifuged and the supernatant removed.

The pellet was washed in 70% ethanol and resuspended in 100 µl of 10 mM Tris-HCl, pH 7.6, 0.1 mM EDTA and 50 µl of phenol

added, vortexed for 10 seconds and after 5 minutes vortexed again and centrifuged for 1 minute. The aqueous layer was extracted with 0.5 ml of diethyl ether and 10 µl of 3 M sodium acetate, pH 5.5, and 250 µl ethanol added. After 30 minutes at -80°C the DNA was recovered by centrifugation and resuspended in 30 µl 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA and stored frozen ready for sequencing.

(vi) Chain Termination Reactions

The single stranded M13 DNA template and primer were added together in 10 µl reaction mix containing 5 µl M13 DNA, 1 µl 10 x Hin buffer (100 mM Tris-HCl, pH 7.4, 100 mM $MgCl_2$, 500 mM NaCl), 1 µl primer (0.2 p mole) and 3 µl water. This was sealed in a 50 µl capillary tube, boiled for 1 minute and allowed to reanneal slowly to room temperature, (15-30 minutes). Four 2 µl aliquots of α-{^{32}P}-dGTP (specific activity about 500 Ci/m mole) were dried down in four Eppendorf tubes and 2 µl aliquots of the hybridised clone mixture added to each tube. Working stocks of ddNTPs were prepared (0.5 mM ddTTP, 0.2 mM ddCTP, 0.125 mM ddGTP, 0.2 mM ddATP), and dNTP zero mizes were prepared (T^o:1 µl 0.5 mM dTTP, 20 µl 0.5 mM dCTP, 20 µl 0.5 mM dATP; C^o: 20 µl 0.5 mM dTTP, 1 µl 0.5 mM dCTP, 20 µl 0.5 mM dATP; G^o: 20 µl 0.5 mM dTTP, 20 µl 0.5 mM dCTP, 20 µl 0.5 mM dATP; A^o: 20 µl 0.5 mM dTTP, 20 µl 0.5 mM dCTP, 1 µl 0.5 mM dATP) and 1 µl of one of the ddNTP's and 1 µl of the corresponding zero mixture was added to each Eppendorf tube, followed by 1/16 µl of Klenow polymerase. After 15 minutes at room temperature 1 µl of 0.5 mM dATP chase was added and after a further 15 minutes, 5 µl of formamide dye (0.03% xylene cyanol FF, 0.03% bromophenol blue) 20 mM EDTA in dionised formamide) was added. The tubes were placed under vacuum for 10 minutes,

then boiled for 5 minutes and loaded onto a 6% polyacrylamide urea gel, as in (a), and electrophoresed for 1.5 and 4 hours respectively at 30 mAmps. After 1.5 hours, the bromophenol blue dye was at the bottom of the gel which corresponded to about 25 bases. A sequence of about 300-350 nucleotides was routinely read from a combined 1.5 hour and a 4 hour electrophoresis. After electrophoresis the gels were fixed in 10% acetic acid for 15 minutes, rinsed in water, dried at 60^{o} for 3 hours and exposed to X-ray film without a screen at room temperature.

2.2.17 Containment Facilities

All work involving recombinant DNA was carried out under C3 or C1 containment conditions for work involving viable organisms and C0 containment conditions for work not involving viable organisms, as defined and approved by the Australian Academy of Science Committee on Recombinant DNA and by the University Council of the University of Adelaide.

CHAPTER 3

CONSTRUCTION OF A HUMAN LIBRARY

3.1 INTRODUCTION

The term "genomic library" has been used to describe a set of recombinant DNA clones, derived from random fragments of a genome, sufficient that the number of clones should contain a representation of all sequences in the genome. Therefore, in principle, any gene can be isolated by screening the library with a specific hybridisation probe.

Construction of a library with subsequent screening provides many advantages over former techniques to clone individual genes. For example, all members of a family of evolutionary or developmentally-related genes can be isolated in a single step by screening a library with a mixed probe.

Furthermore, isolation of a set of overlapping clones permits the study of sequences extending many kilobases from the gene or set of genes, in both directions. The organisation of multigene families such as the human β-like globin cluster has recently been investigated in this way (Efstratiadis *et al.*, 1980).

Three recent advances in technology made the construction and screening of DNA libraries from complex genomes possible. First, a rapid *in situ* plaque hybridisation procedure developed by Benton and Davis (1977). Second, EK-2 certified lambda cloning vectors which were capable of incorporating 15 kb of foreign DNA were constructed (Blattner *et al.*, 1977). Third, *in vitro* packaging systems for λ DNA were developed for use in recombinant DNA experiments (Hohn and Murray, 1977; Sternberg *et al.*, 1977).

Libraries from DNA of a number of species such as *Drosophila*, silkworm, rabbit, chicken and human DNA have been constructed (Lawn *et al.*, 1978). The general strategy employed involved the partial digestion of high molecular weight genomal DNA with the restriction endonucleases HaeIII and AluI. The products were size-fractionated by sucrose gradient centrifugation and the large fragments (15-20 kb) were isolated and treated with EcoRl methylase, to render EcoRl sites within the genomal DNA resistant to cleavage with EcoRl. Synthetic dodecameric DNA molecules bearing an EcoRl cleavage site (EcoRl linkers) were ligated to the methylated DNA and digested with EcoRl to generated EcoRl cohesive ends. Following an additional size selection (15-20 kb), the genomal DNA was suitable for insertion into bacteriophage λ cloning vector, Charon 4A (Blattner *et al.*, 1977).

The strategy used here to construct a human library was adapted from the method described by Kemp *et al.*, (1980). This approach was preferred because it obviates the need for linkers, but still results in a random library of clones.

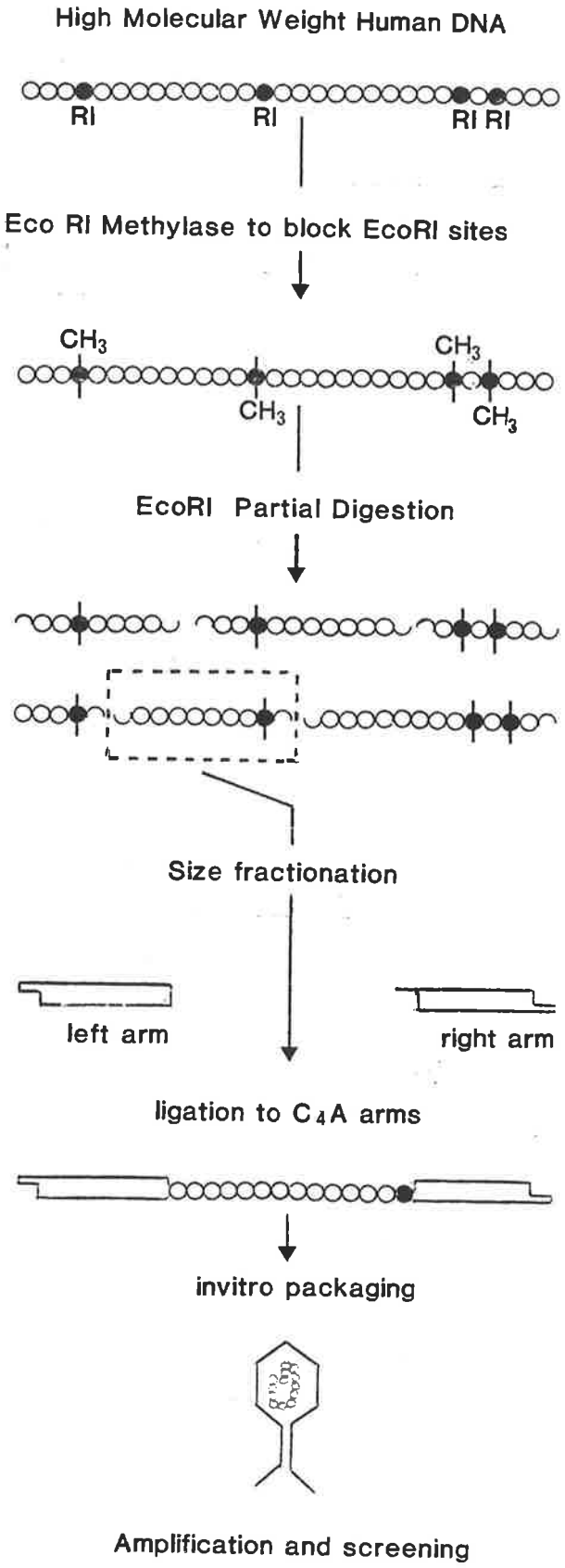
The method is based on two observations. First, that the specificity of EcoRl endonuclease for the sequence 5' G A A T T C 3' can be reduced to N_1 A A T T N_2 (EcoRl* site), where N_1 and N_2 are any base, when incubated under low salt conditions and second that EcoRl methylase blocks the standard EcoRl site but not the EcoRl* site (Hedgepeth, *et al.*, 1972, Tikchonenko *et al.*, 1978).

The approach to the construction of the human clone library outlined in Fig. (3.1) begins with methylation of high molecular weight DNA

Fig. 3.1

Construction of the EcoRI* Human Library

The strategy used to construct the EcoRI* library is outlined. The open circles represent DNA, the closed circles represent EcoRI sites and the dashed closed circles represent methylated sites. The left and right arms are the 11.1 and 19.5 kb fragments derived from λCharon 4A.


High Molecular Weight Human DNA
RI
RI
RI RI
Eco RI Methylase to block EcoRI sites
CH3
CH3
CH3
CH3
EcoRI Partial Digestion
Size fractionation
left arm
right arm
ligation to C4A arms
invitro packaging
Amplification and screening

at EcoR1 recognition sites to protect the DNA against subsequent cleavage at these sites by EcoR1. The methylated DNA is then partially digested using EcoR1* conditions to produce fragments averaging 12-20 kb in length. Protection of the EcoR1 sites by methylation is crucial since even under EcoR1* conditions these sites are cleaved far more rapidly than other sites (Polisky *et al.*, 1975).

An advantage of this approach is that EcoR1* fragments have cohesive termini (A A T T) identical to those of EcoR1 fragments and hence can be ligated directly into the EcoR1 cut Charon 4A vector without using linkers. A portion (¼) of the vector-insert joints should be cleavable by EcoR1 because ligation of EcoR1* ends to EcoR1 ends can produce an EcoR1 site (5' G A A T T C 3') if N_1 and N_2 contributed by the two ends are G and C respectively.

3.2 RESULTS

3.2.1 Purification of Human DNA

It was essential that the human DNA used was of high molecular weight to ensure that most fragments of 15-20 kb derived from partial digests contained cohesive ends.

The best method found to give consistently high molecular weight DNA is described in methods (2.2.2). Using this procedure 87 mgs of human genomal DNA was isolated from 100 g of frozen placenta. The DNA was approximately 80-100 kb in length as determined by 0.3% agarose gel electrophoresis.

3.2.2 Methylation of Human DNA

The initial step in construction of the library was methylation of the genomal DNA at the EcoR1 recognition sites. The RI methylase was isolated as described in methods (2.2.3a) and purified by elution from a phosphocellulose column Fig. (3.2). The active fractions (13-24) were pooled and the optimal concentration to fully protect against EcoR1 digestion was determined by incubating 1 µg of λ DNA with increasing volumes of methylase for 1 hour at 37°C. From Fig. (3.3) it can be seen that 0.2 µl of RI methylase per µg λ DNA (5 units/µl) was sufficient to prevent EcoR1 digestion (Track c). With increasing concentrations of methylase no endonuclease activity was observed, therefore the methylase appeared satisfactory for use within 10 fold excess without DNA degradation.

Human placental DNA (2 mg) was incubated with 2000 units of RI methylase in a 10 ml reaction volume. To monitor DNA methylation an aliquot of human DNA was digested with EcoR1 before and after incubation with RI methylase. As shown in Fig. (3.4), the human genomal DNA taken before the addition of RI methylase (Track b) was digested to completion, while the methylated human DNA (Track d) was totally resistant to EcoR1 cleavage. However, the RI methylated DNA was susceptible to cleavage under EcoR1* conditions (Track e). It was therefore concluded that the human genomal DNA was fully methylated at the EcoR1 restriction sites.

Fig. 3.2

Purification of RI Methylase

As described in Methods (2.2.3a), the ammonium sulphate precipitate from *E. coli* 2012 was applied to a phosphocellulose column and eluted with a salt gradient from 0.2 M - 0.8 M NaCl, and 5 ml fractions collected. Each fraction was assayed for methylase activity by incubation with λDNA and {^{3}H}-SAM (Methods 2.2.3b), and determining acid precipitable counts. The RI methylase activity was detected and pooled from fractions 13-24.

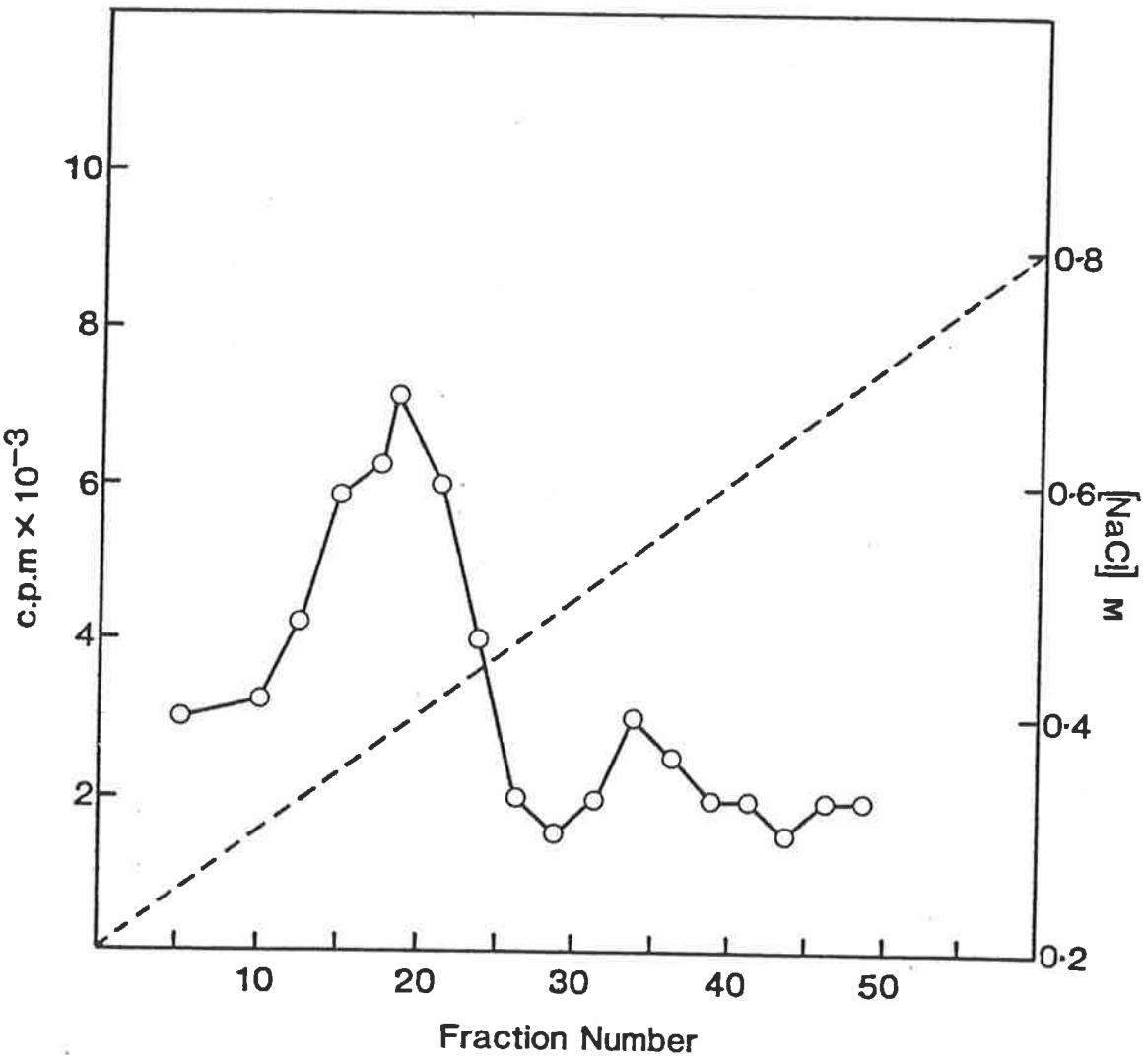

c.p.m x 10⁻³
10
8
6
4
2
0·8
0·6
0·4
0·2
[NaCl] M
10
20
30
40
50
Fraction Number

Fig. 3.3

Determination of RI Methylase Activity

The activity of RI methylase was determined by incubating increasing volumes of RI methylase with 1 µg of λDNA for 1 h at 37°, then incubating with 1 unit of EcoRI for 1 h at 37°. The samples were electrophoresed on a 1% agarose gel and stained with 10 µg/ml Ethidium Bromide for UV illumination of the DNA

The tracks are as follows:

A. 0.05 µl RI methylase
B. 0.1 µl RI methylase
C. 0.2 µl RI methylase
D. 0.5 µl RI methylase
E. 2 µl RI methylase

From these results it was deduced that 1 unit of methylase is equivalent to 0.2 µl, i.e. the minimum volume required to completely methylate 1 µg of DNA. Therefore the activity of the methylase was approximately 5 units/µl.

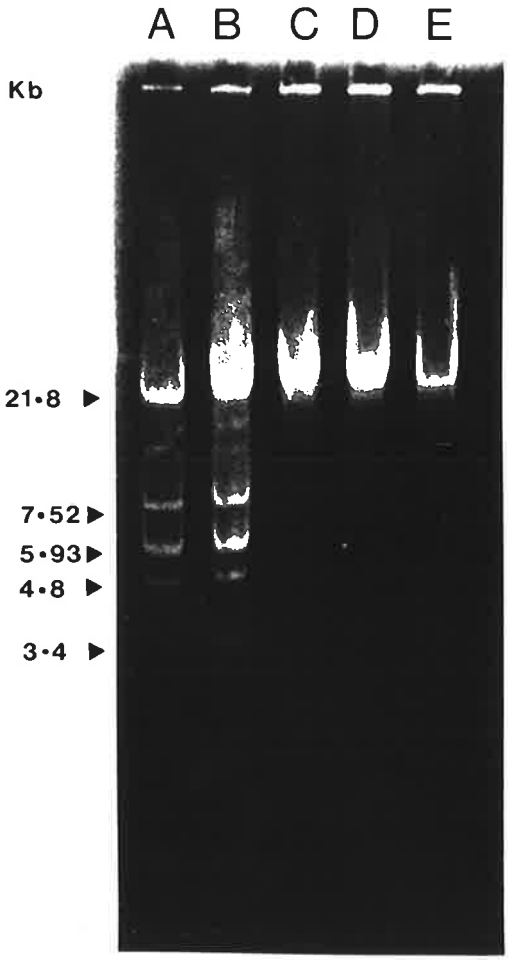
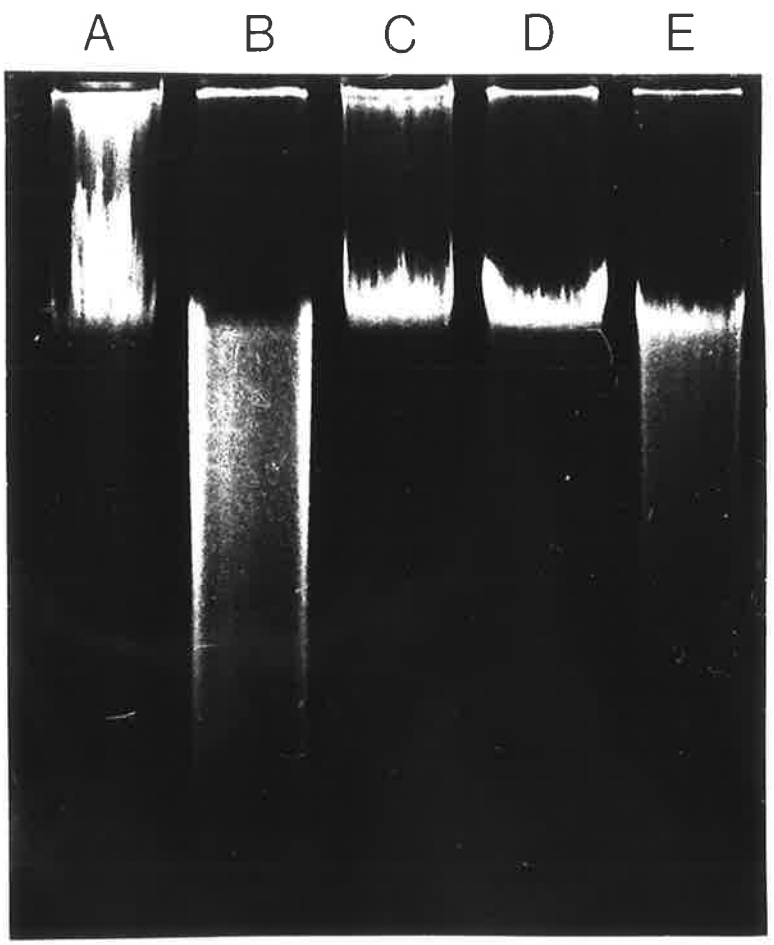

A B C D E
Kb
21·8
7·52
5·93
4·8
3·4

Fig. 3.4

Methylation of Human DNA

Aliquots (1 µg) were taken before and after RI methylation of the human DNA as described in the text and then digested with 1 unit of EcoRI. The samples were electrophoresed on a 1% agarose gel. The tracks are as follows

A. Undigested human DNA
B. EcoRI digested human DNA
C. Methylated human DNA
D. EcoRI digested methylated human DNA
E. EcoRI* digested methylated human DNA


A
B
C
D
E

3.2.3 Partial EcoRI* Digestion of Human DNA

EcoRI* recognises the sequence N_1 A A T T N_2, so potential EcoRI* sites should occur on average every 256 bp in DNA containing 50% G + C. Since only 1/80 of such sites need to be cleaved to produce 20 kb fragments, partial products of this length should approximate to a randomly fragmented population.

EcoRI* activity was initially examined by incubating 1 µg of λ DNA with increasing concentrations of EcoRI in low salt buffer. As can be seen from Fig. (3.5), 25 units of EcoRI (1 unit = amount of EcoRI required to digest 1 µg λ DNA in one hour) are needed to completely digest 1 µg λ DNA in one hour, under EcoRI* conditions. With such a high concentration of EcoRI required it was essential that the EcoRI was free of any contaminating exo- or endonuclease activity.

It was found that the EcoRI activity peak isolated from the phosphocellulose column was slightly contaminated with an endonuclease. Many repurification steps were tried and it was found that after elution from a blue sepharose column, as described in methods (2.2.3a), that up to 1000 fold excess of EcoRI was free of detectable non-specific nuclease activity.

To determine the optimal incubation conditions for partial EcoRI* cleavage of the human methylated DNA, increasing concentrations of EcoRI were incubated with DNA for 1 hour at 37°C.

Fig. 3.5

Determination of EcoRI* Activity

1 μg of λDNA was incubated with increasing amounts of EcoRI under EcoRI* conditions (25 mM Tris-HCl, pH 8.5, 2 mM $MgCl_2$) for 1 h at 37°. The tracks are as follows:

A: λdigested under EcoRI conditions using 1 unit of enzyme.

B: λDNA undigested

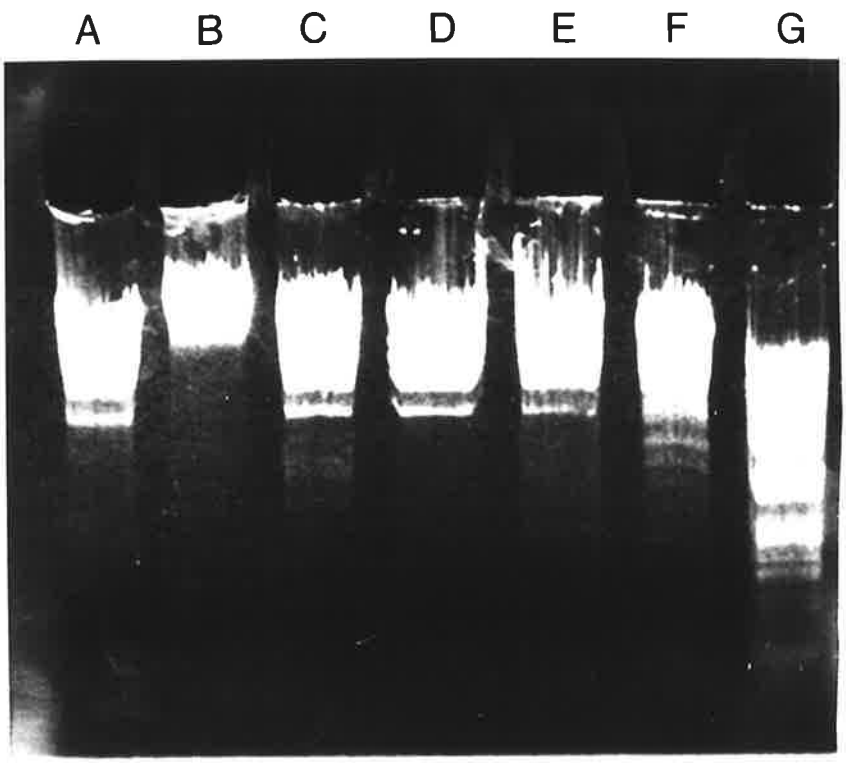
C: 1 unit EcoRI*

D: 2.5 units EcoRI*

E: 5 units EcoRI*

F: 10 units EcoRI*

G: 25 units EcoRI*


A
B
C
D
E
F
G

The results in Fig. (3.6) indicate that 15-20 units of EcoRl incubated for 1 hour/µg DNA resulted in partial digestion products from full length to 10 kb.

1 mg of human DNA was divided into three tubes and each digested with 15, 17, 19 units/µg/hour respectively and the partial digestion products were then pooled. A range of conditions were chosen to ensure randomness of digestion and to minimise the effect of possible preferred EcoRl* sites.

To ensure that the majority of EcoRl* fragments had two cohesive ends a sample of the digested human DNA was incubated with *E. coli* DNA ligase which cannot ligate blunt ends. Following ligation the DNA had been converted to high molecular weight molecules indicating that most of the DNA fragments have ligatable EcoRl* ends. (Fig. 3.7, Track a, b).

3.2.4 Size Fractionation of Human DNA

Before cloning the EcoRl* digested human DNA into λ Charon 4A it appeared advantageous to select fragments between 12-22 kb in length for a number of reasons.

Firstly, DNA fragments greater than 22 kb cannot be incorporated into viable recombinant phage. Also the lower size limit was selected because it minimises the possibility of introducing more than one noncontiguous fragment of human DNA into a single recombinant phage. The selection of this size class also minimises

Fig. 3.6

Optimisation of EcoRI* Cleavage Conditions

Increasing concentrations of EcoRI were incubated under EcoRI* conditions (Methods 2.2.3d) with 1 µg human methylated DNA for 1h at 37°, and electrophoresed on a 0.5% agarose gel. EcoRI digested λDNA was used as markers. The tracks are as follows:

A: 10 units of EcoRI

B: 15 units of EcoRI

C: 20 units of EcoRI

D: 40 units of EcoRI

E: 80 units of EcoRI

F: 100 units of EcoRI

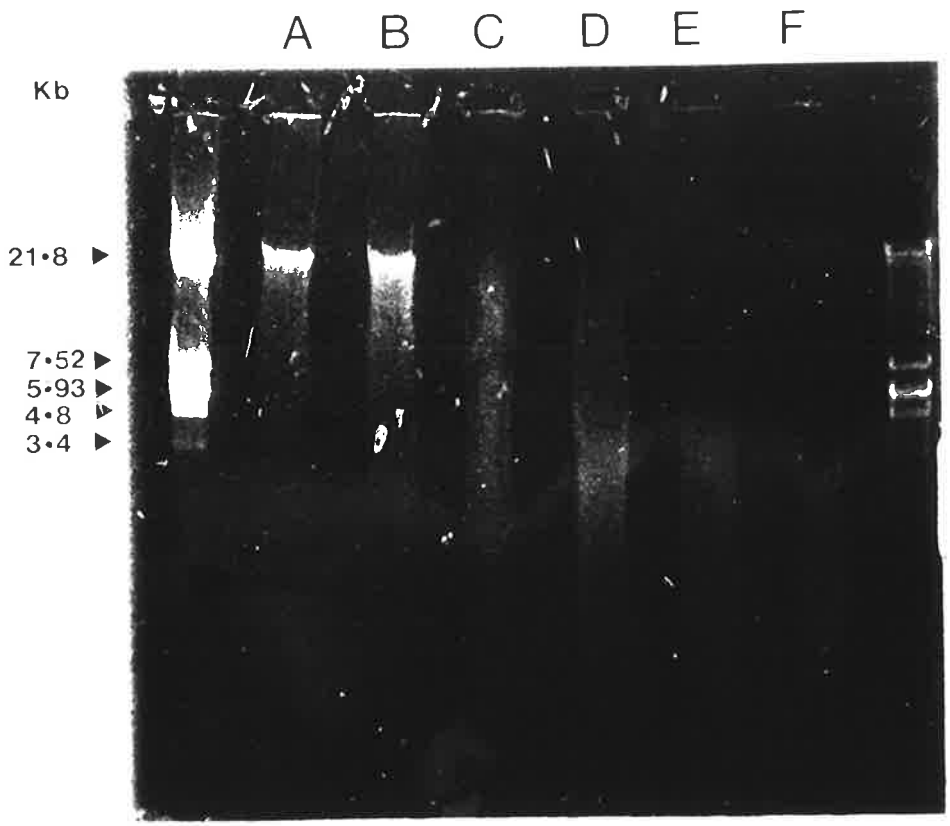

A B C D E F
Kb
21•8
7•52
5•93
4•8
3•4

Fig. 3.7

Ligation of EcoRI* DNA

Samples of EcoRI* digested human DNA (1 µg) before and after phosphorylation were incubated with 1 unit of *E. coli* DNA ligase at 4^{0} for 1 h and electrophoresed on a 0.8% agarose gel. EcoRI digested λDNA was used as size markers.

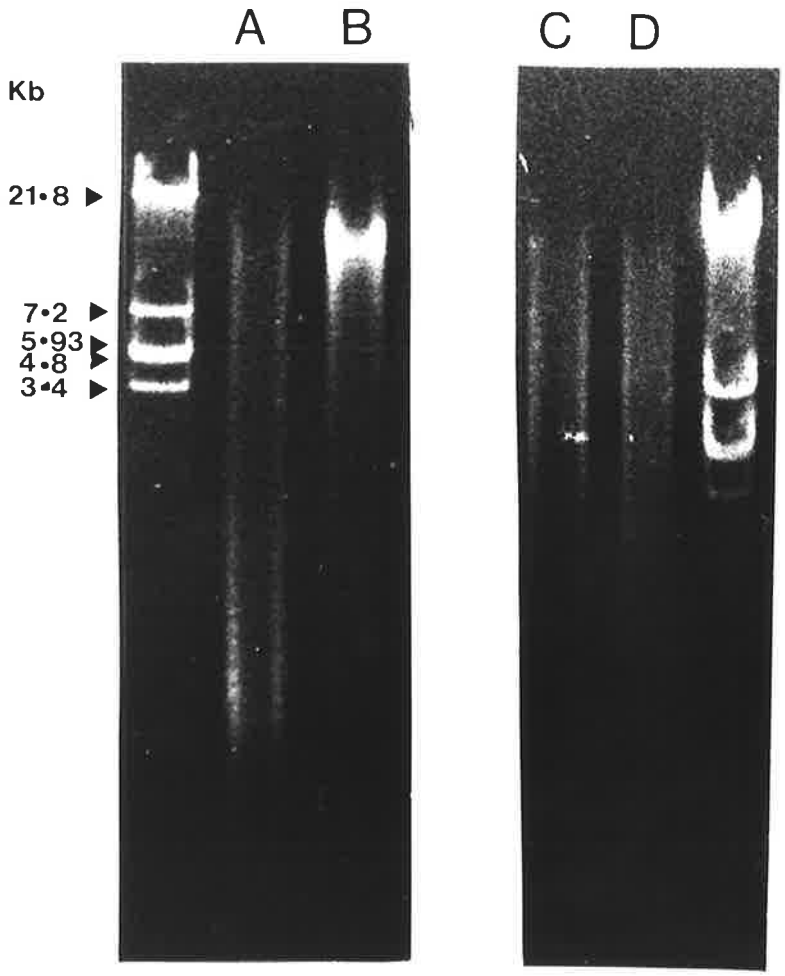
The tracks are as follows:

A: EcoRI* digested human DNA

B: Ligated EcoRI* human DNA

C: Phosphatased EcoRI* human DNA

D: Ligated, phosphatased EcoRI* human DNA


A
B
C
D
Kb
21•8
7•2
5•93
4•8
3•4

the number of recombinant phage required for representation of the entire human genome, while at the same time maximising the size of individual clones.

The selection of a size range rather than DNA fragments of close to one particular length (e.g. 20 kb) ensures that the sequence complexity of the DNA to be used is as representative as possible of the single copy gene distribution found in the intact genome.

Three methods were used to fractionate the partial EcoR1* fragments and the results obtained using these different DNA preparations are considered in Section 3.2.6.

(a) 10-40% sucrose density gradient centrifugation was used with EcoR1 digested λ DNA as size markers and fragments between 15-22 kb in length were collected as shown in Fig. (3.8). This technique enriched for the desired size class but was still contaminated by smaller DNA fragments.

(b) Preparative 1% agarose gel electrophoresis was used and the DNA between 15-22 kb was cut out of the gel and eluted. This method had the advantage of enabling accurate size fractions to be taken and also ensured that no smaller DNA fragments remained (Fig. 3.9).

(c) As an alternative to physical DNA fractionation, a third procedure was used to select the appropriate size class

Fig. 3.8

Sucrose Gradient Fractionation of EcoRI* Human DNA

Human DNA, digested with EcoRI*, was fractionated on a 10-40% sucrose gradient in NET (200 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl, pH 7.4), centrifuged at 34,000 rpm in an SW41 rotor for 16 h at 4^{o}.
EcoRI digested λDNA (labelled curve) was used as size markers.
The absorbance of A_{260} was recorded on an Isco fraction collector.

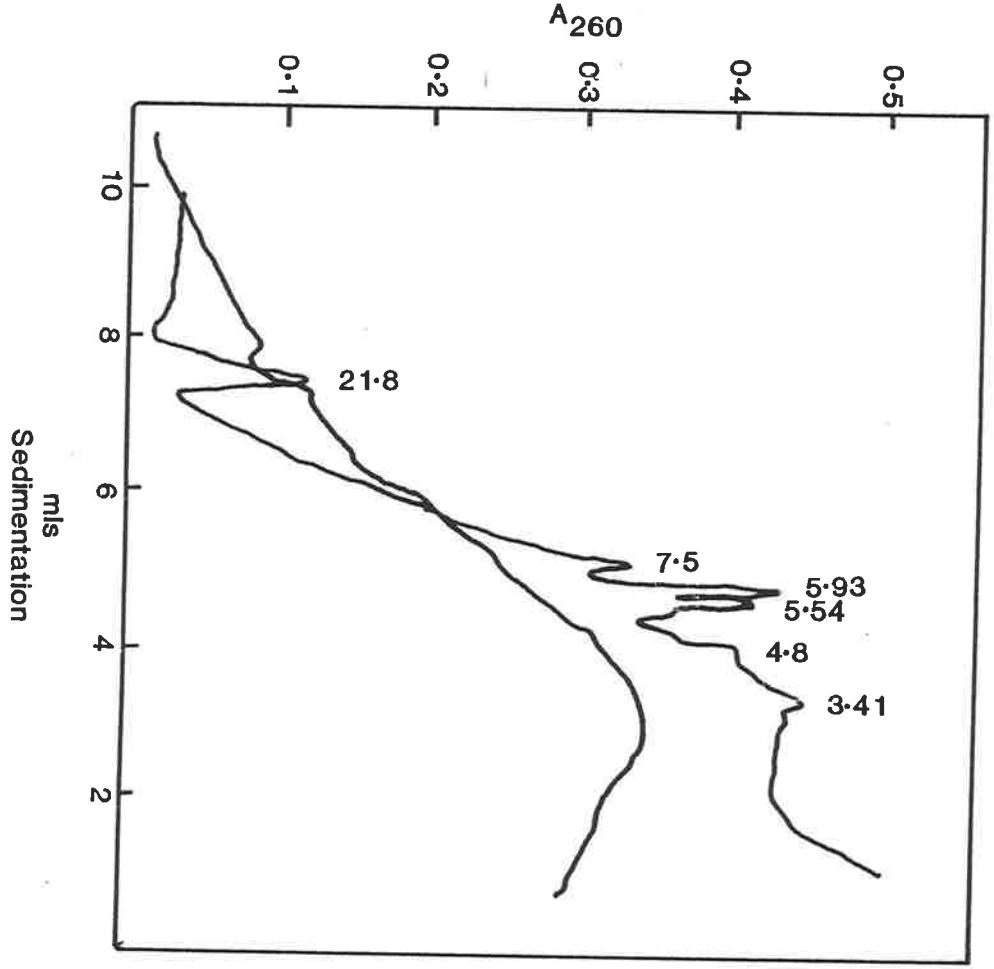
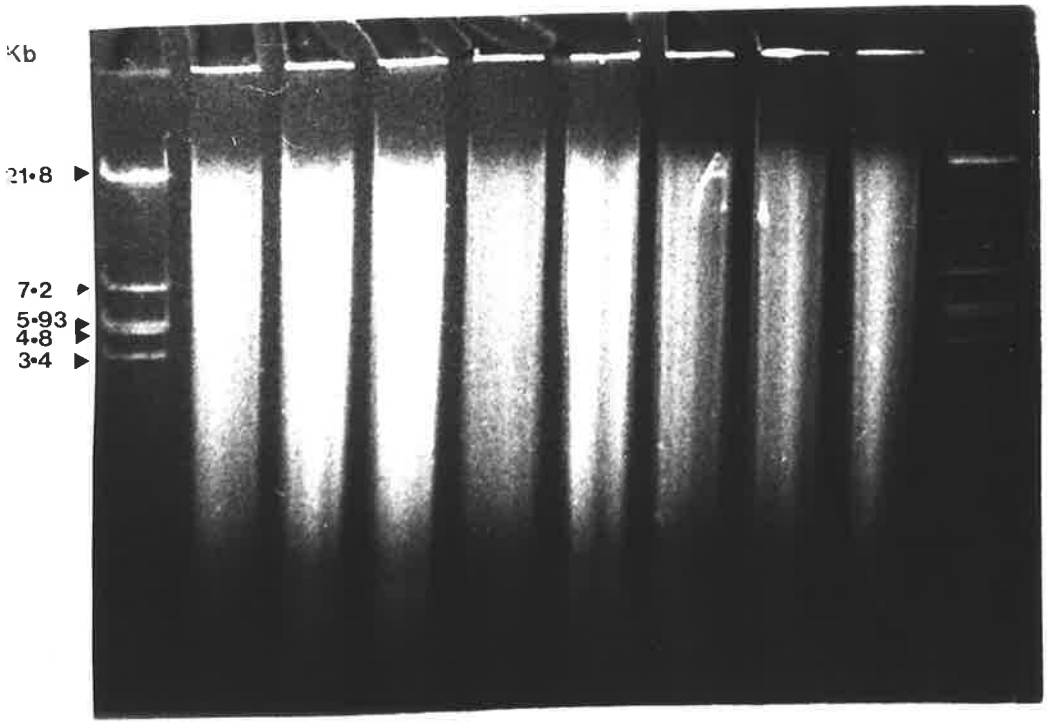

A260
0·1
0·2
0·3
0·4
0·5
10
8
6
4
2
mls Sedimentation
21·8
7·5
5·93
5·54
4·8
3·41

Fig. 3.9

Agarose Gel Fractionation of EcoRI* Human DNA

EcoRI* digested human DNA was electrophoresed on a 1% agarose gel, and DNA fragments ranging in size between 15-22 kb cut out and eluted as described in Methods (2.2.12a). EcoRI digested λDNA was used as size markers.


Kb
21·8
7·2
5·93
4·8
3·4

of DNA fragments. This involved dephosphorylating the 5' ends of the EcoR1* DNA fragments, thus preventing self ligation (Fig. 3.7, Track c and d). Therefore only DNA between 10-22 kb could be ligated into the Charon 4A arms. This method does, however, result in a decrease in the average size of insert obtained and a lower productive ligation efficiency.

3.2.5 Ligation of Human DNA to Charon 4A Vector DNA

The DNA of λ phage Charon 4A contains three EcoR1 sites (Blattner *et al.*, 1977) and yields four fragments when digested with EcoR1: two end fragments 11.1 and 19.8 kb and two internal fragments 6.8 and 8.0 kb. The two internal fragments contain genes which are not required for lytic infection and therefore can be replaced by foreign DNA. These fragments were removed by first annealing the cohesive ends of the λ DNA and then digesting with EcoR1. The two arms were separated from the two internal fragments by sucrose density gradient centrifugation (Fig. 3.10). Analysis of the purified Charon 4A arms by electrophoresis on 0.8% agarose gel indicated no visible contamination by internal fragments (Fig. 3.11, Track a).

The Charon 4A DNA arms were ligated to the 15-20 kb EcoR1* DNA fragments using approximately a 2-fold molar excess of the Charon 4A DNA, and at a high DNA concentration to minimise intramolecular joining and to maximise the formation of concatemeric

Fig. 3.10

Isolation of λCharon 4A Arms

λCharon 4A DNA was digested with EcoRI and the internal fragments were separated by centrifugation on a 10-40% sucrose gradient in NET at 34,000 rpm for 16 h at 4^{o}. The absorbance at A_{260} was recorded on an Isco fraction collector.

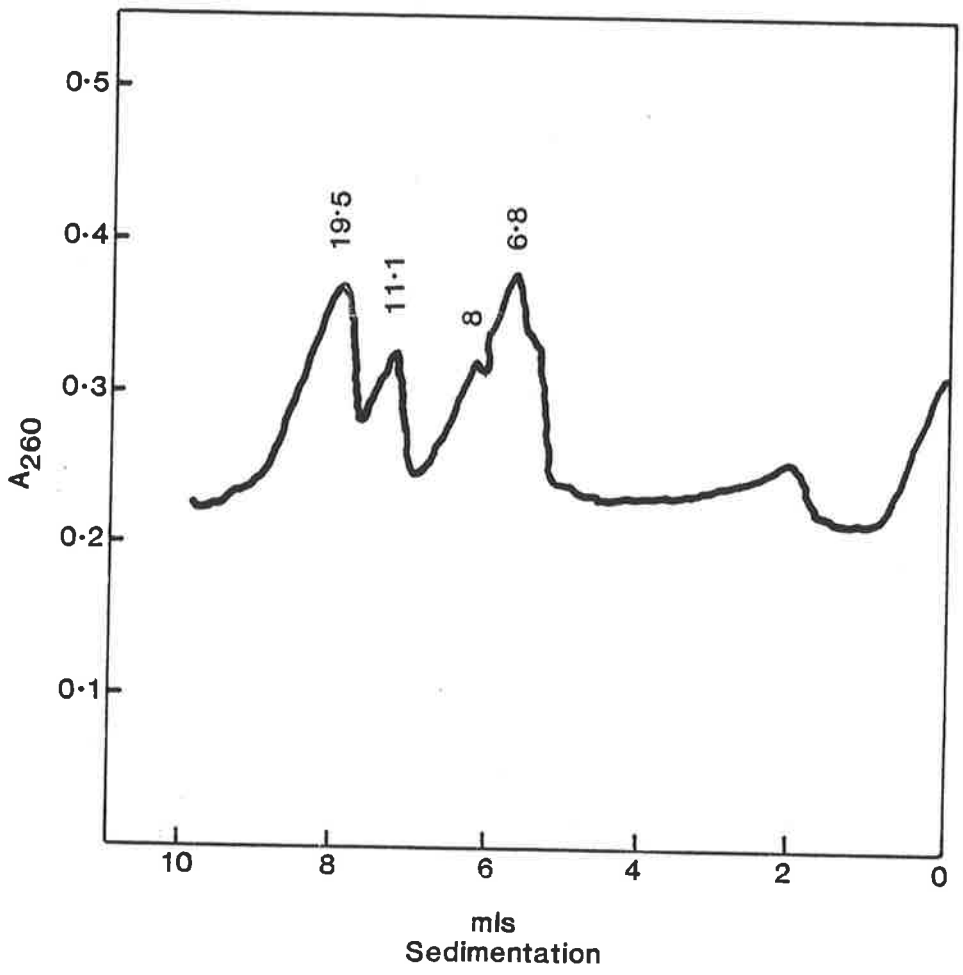

0·5
0·4
0·3
0·2
0·1
A_{260}
19·5
11·1
8
6·8
10
8
6
4
2
0
mls
Sedimentation

Fig. 3.11

Ligation of λCharon 4A Arms to EcoRI* Human DNA

An aliquot of the ligation mix containing the λCharon 4A arms and human DNA was electrophoresed on a 1% agarose gel before and after incubation with T4 ligase. EcoRI digested λDNA was used as size markers.

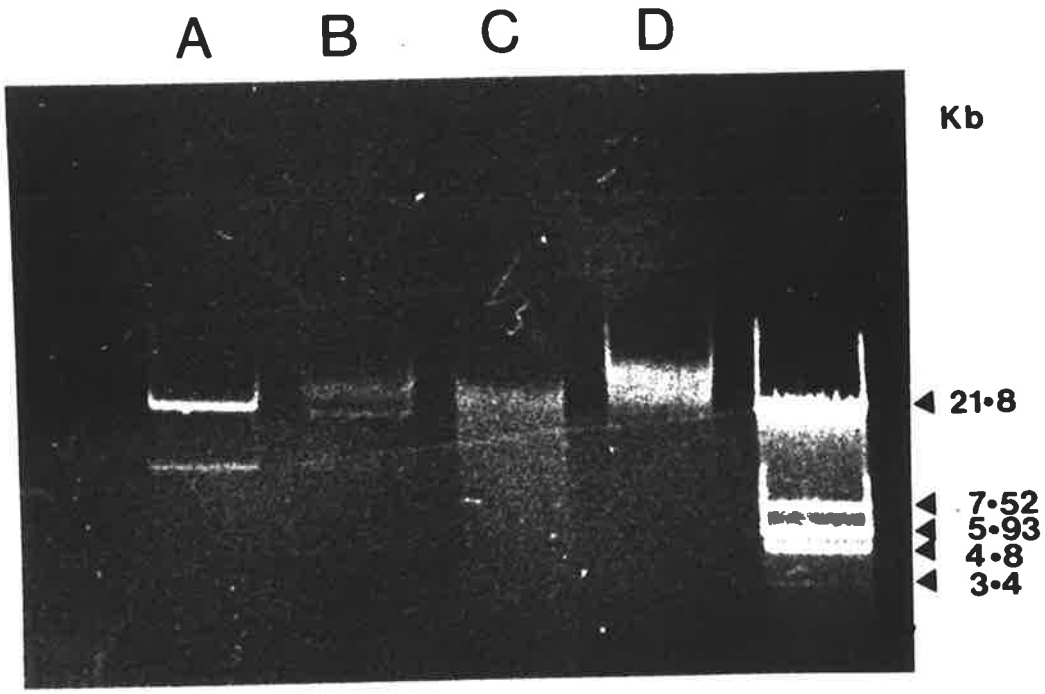
The tracks are as follows:

A: λCharon 4A arms

B: Ligation mix before ligation

C: EcoRI* human DNA

D: Ligation mix after ligation


A
B
C
D
Kb
21·8
7·52
5·93
4·8
3·4

DNA recombinants, the substrate for *in vitro* DNA packaging.

Under these conditions the human DNA appeared to ligate to the λ Charon 4A arms as determined by analysing a fraction of the ligation mix on a 0.8% agarose gel (Fig. 3.11).

3.2.6 *In vitro* Packaging of DNA into Phage Particles

Using $CaCl_2$ transfection procedure of Mandel and Higa (1970), approximately 10^6 pfu/μg of intact λ DNA can be obtained but efficiencies drop to 2-10 x 10^3 pfu/μg for cleaved and religated λ vector DNA. Thus approximately 100-400 μg of DNA fragments are needed to construct a complete human library, an amount difficult to obtain. Fortunately efficiencies one order of magnitude higher have been achieved using *in vitro* packaging procedures (Hohn and Murray, 1977, Sternberg *et al.*, 1977).

The principle of *in vitro* packaging is the following. In the presence of high concentrations of phage head precursor or pre-head, and the packaging proteins, which are the products of the genes A, Nul, D and F_1, λ DNA is packaged (Hohn, 1975). The full heads are then matured *in vitro* into plaque-forming units (PFU) in the presence of proteins from genes W and Fll and phage tails.

Practically, *in vitro* packaging is most efficiently performed in a very concentrated mixture of two induced lysogens, one of which is genetically blocked at the pre-head stage (by an

amber mutation in gene D) and therefore accumulates these precursors, while the other is inhibited in forming capsid protein. These two lysates provide all the necessary components to convert λ DNA into a plaque-forming particle.

Using the packaging procedure described by Sternberg *et al.*, (1977) and detailed in methods (2.2.3h), packaging extracts were obtained which varied between 0.2 - 2.0 x 10^7 pfu/μg of intact λ DNA depending on the preparation. The extracts were consistently less efficient for intact Charon 4A DNA (.1 - .5 x 10^6 pfu/μg) and even lower still using cleaved and religated Charon 4A DNA (2-10 x 10^4 pfu/μg). This lowering in efficiency is widely observed (Maniatis *et al.*, 1978, Robbins *et al.*, 1979).

In small scale packaging reactions ligated human DNA prepared by sucrose gradient centrigutation or agarose gel electrophoresis were compared with DNA that had been dephosphorylated.

Table (3.1) summarises the results. The most efficient packaging of the human DNA was obtained with the sucrose fractionated DNA giving 4.2 x 10^4 pfu/μg. This efficiency is similar to that obtained in the construction of libraries of rabbit,(3.8 x 10^4) pfu/μg (Maniatis *et al.*, 1978), human (2.5 x 10^4 pfu/μg) (Lawn *et al.*, 1978) and goat (5.3 x 10^4 pfu/μg)(Robbins *et al.*, 1979) all using linked DNA, however, is still lower than the efficiency obtained for the packaging of mouse EcoRI* DNA (1.3 x 10^5 pfu/μg) (Kemp *et al.*, 1980).

TABLE 3.1

Effect of DNA Source on Packaging Efficiencies

Source of DNA	Efficiency of Packaging/μg DNA
Sucrose Fractionated DNA	4.2×10^4
Agarose Fractionated DNA	5.2×10^2
Phosphatased DNA	1.5×10^4
λ wild-type	1.2×10^7
λ Charon 4A	2.6×10^6
λ Charon 4A cut and religated	9.6×10^4

The packaging efficiency of the human DNA prepared by agarose gel electrophoresis was very low at 5.2×10^2 pfu/µg. This is probably due to contaminating agarose remaining after elution. At this time no low melting point agarose was available, and its use may have given better results.

The dephosphorylated human DNA also showed a lower packaging efficiency than the sucrose-fractionated DNA of 1.5×10^4 pfu/µg. This suggests either that the sucrose centrifugation step aids in removing some inhibitor of packaging, that dephosphorylation reduces ligation efficiency, or that the larger size range of fragments results in a lower proportion of the DNA being of packageable size.

3.2.7 Construction of a Human Library

On the basis of this data 20 µg of λ Charon 4A arms were ligated with the sucrose fractionated EcoR1* human DNA. This was divided into 10 packaging mixes and a plate stock prepared from each, generating 10 pools. The efficiency of this large scale packaging was 3.8×10^4 pfu/µg or 7.6×10^4 independent plaques per pool. Thus a library of human genomic DNA was constructed consisting of 7.6×10^5 phage.

The characterisation of the library is described in Table (3.2). The background of non recombinant phage (i.e. intact λ Charon 4A) was estimated by plating the phage on a lawn of lac^- bacteria in the presence of an indicator for the β-galactosidase gene (5-bromo-4-chloro-3-indolyl-β-D-galactoside)(Blattner *et al.*, 1977). The recombinant phage are lac^- because the inserted DNA

Table 3.2

Characterisation of Human EcoRI* Library

Amount of Charon 4A arms to Human DNA in ligation reation	20 µg:16 µg
Packaging Efficiencies for Intact Charon 4A DNA	1.5×10^6 pfu/µg
Packaging Efficiency of Human DNA, Charon 4A ligation mix	3.8×10^4 pfu/µg
Background of non-recombinant phage packaged	~0.5%
Total Number of Independent Phage DNA recovered	7.6×10^5
Number of Recombinant Phage required for a "complete" library	8×10^5

fragment has split this gene and thus the phage produce colourless plaques in the presence of the indicator. λ Charon 4A is lac^+ and so non recombinant phage arising from contamination of the 31 kb fragment with the 7 kb and 8 kb internal fragments of Charon 4A will produce blue plaques. Data from this assay indicated that only 0.5% of the library consisted of non recombinant phage.

An essential feature of gene isolation is to establish a permanent library that can be screened repeatedly. To achieve this it is necessary to amplify *in vitro* packaged recombinant phage and store the library in the form of a plate lysate. There is of course a risk that a particular recombinant phage will exhibit a growth disadvantage and will be eliminated from the library during amplification. It is therefore important to minimise competitive growth. This was accomplished by plating the *in vitro* packaged phage on fresh agar plates at a low density (10,000 pfu/15 cm diameter plate), i.e. one plate per pool. In this manner the phage were amplified approximately 1×10^6 fold and recovered as a phage lysate and stored at 4^oC.

3.2.8 Screening Human Library for Ribosomal Genes

To test whether the library does contain a random distribution of genes, it was screened for ribosomal genes as these are a repeated gene family and should therefore act as a model system for the feasibility of screening for histone genes.

Approximately 20,000 phage from pool one were plated on a 15 cm agar plate. Using conditions described by Benton and Davis (1977), duplicate nitrocellulose filters were sequentially applied to the plates and then hybridised with cDNA made to 18S chicken ribosomal RNA. Two positive signals were detected. Both positive plaques were confirmed by their appearance on duplicate filters and by rescreening.

Two positive clones from 20,000 plaques screened would suggest that there are 76 different 18S ribosomal positive clones in the library constructed assuming it contains 760,000 independent clones. This number of ribosomal clones correlates well with the estimated number of ribosomal genes in the human genome (Southern and Bird, 1976).

Consequently, it was considered that the library provided a sufficient representation of the human genome to permit screening for clones containing histone genes.

3.3 DISCUSSION

By packaging the recombinant DNA into phage heads *in vitro*, 7.6×10^5 recombinant clones of human DNA were generated. Since the human genome is approximately 3×10^6 kb in length (Lawn *et al.*, 1978) and the average length of the cloned segments should be about 15 kb (as 12-22 kb DNA fragments were used in the ligation) the human library should contain 4 genome equivalents. Assuming randomness, the probability that it contains any human DNA segments is greater than 95% (Clarke and Carbon, 1976).

The library was segregated into 10 pools to simplify subsequent screening, since clones derived from separate pools represent independent cloning events. Since the individual phage generated by packaging were amplified by ~10^6 during plating, these pools represent a permanent library which can be repeatedly screened.

In order to demonstrate that a repeated gene can be isolated, a fraction of the library was screened for 18S ribosomal genes and two out of 20,000 plaques were positive.

Therefore it was considered that the library provided a sufficient representation of the human genome to permit screening for clones containing histone genes.

CHAPTER 4

CHARACTERISATION OF A CHICKEN HISTONE GENOMAL CLONE

4.1 INTRODUCTION

A prerequisite for the study of histone genes from any organism is the availability of a probe of sufficient purity and homology to allow unequivocal detection of the gene sequences. In the sea urchin system, advantage was taken of the fact that early cleavage embryos synthesise predominantly histone mRNA. Pulse-labelled RNA isolated from these cells, although contaminated with other RNA, proved to be perfectly usable because most of the radioactivity was contained in histone mRNA (Gross *et al.*, 1976a).

Histone gene sequences from sea urchin were also useful in studying *Drosophila* histone genes. Cross-reactivity between sea urchin and *Drosophila* histone-coding sequences led to the isolation and subsequent analysis of cloned *Drosophila* genes (Lifton *et al.*, 1977). However, the use of cloned sea urchin histone genes to detect histone genes from higher vertebrates has proved more elusive ~~elusive~~ difficult (K. Murray, K. Gross, personal communication).

Despite intensive efforts, isolation of histone mRNA of sufficient radiochemical purity from pulse-labelled *Drosophila* tissue culture cells (Burckhardt and Birnstiel, 1978) or Hela cells (Wilson and Melli, 1977) has been very difficult. By comparison, 5-day old chick embryos provided an excellent source of histone mRNA (Krieg, 1980; Crawford *et al.*, 1979). A heterogeneous 7-11S fraction could be resolved from the 5-day old chicken total RNA by sucrose fractionation and this RNA fraction contained histone mRNA for each of the five histone proteins, as identified by *in vitro* translation in the wheat embryo cell free system (Krieg, 1980).

Radio-labelled cDNA made to this RNA fraction was used to screen the chicken genomal library of Dodgson *et al.* (1979), and four recombinants were selected which were positive for the chicken histone cDNA probe, but negative to a "combined" cDNA probe made from RNAs considered likely to contaminate the histone mRNA preparations (28S, 18S, 5.8S and 5S rRNA, globin mRNA and 4S RNA; Harvey and Wells, 1979). One of these chicken genomal clones, λCH.01, was selected and characterised further (Harvey and Wells, 1979). Fig. (4.1) shows an EcoR1 and HindIII restriction map of λCH.01. Both the 3.3 kb EcoR1 fragment and the 3.75 kb EcoR1/HindIII fragment hybridise to cDNA made from the chicken histone mRNA. Limited DNA sequencing of HaeIII fragments indicated that the 3.3 kb EcoR1 fragment contained H2A sequences and an amino acid sequence with strong homology to an H1 histone sequence (Harvey and Wells, 1979).

Assuming sufficient homology between chicken and human histone gene sequences, it now became possible to screen for human histone genes. However, since the EcoR1* Library was still in preparation and the Maniatis human library (Lawn *et al.*, 1978) had become available, it was decided to screen the latter, using both the chicken histone mRNA and the 3.3 kb EcoR1 fragment from λCH.01 as hybridisation probes. Several recombinants were found to be positive to both hybridisation probes. The clone which gave the strongest hybridisation signal was selected for further characterisation on the assumption that it would be more likely to contain a genomal repeat of human histone genes.

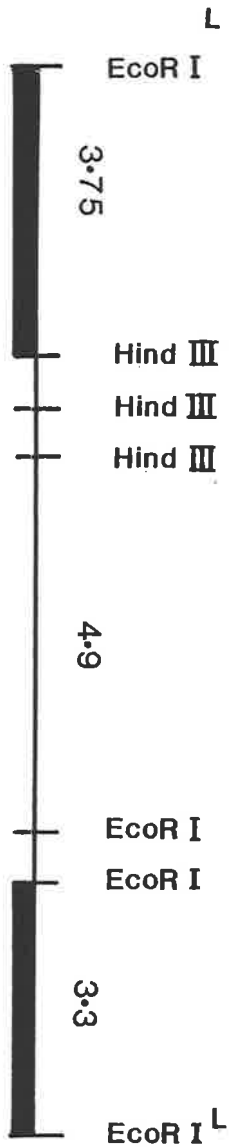
This clone, initially termed λHH.01* (Clark *et al*, 1981), was later identified by sequencing to be in fact a chicken histone recombinant clone, λCH.02, previously isolated in the same laboratory by

* The error in assigning λHH.01 as a human histone clone has been communicated to the Editor of the journal.

Fig 4.1

Restriction Endonuclease Map of λCH.01 Insert DNA

Blocked regions indicate the EcoRI/HindIII fragments which hybridise with chicken histone cDNA. Map distances are shown in kilobase pairs. $EcoRI^{L}$ denotes the boundaries of the insert.


L
EcoR I
3·75
Hind III
Hind III
Hind III
4·9
EcoR I
EcoR I
3·3
EcoR I
L

R. Harvey but not characterised until much later (Harvey *et al.*, 1981). The source of the contamination is not clear, though later screening of the human library suggested that the library stock itself was not contaminated (chapter 6).

The structure and organisation of this chicken clone will be presented in this and the following chapter, prior to consideration of two subsequently isolated genomal clones containing human histone genes.

4.2 RESULTS

4.2.1 Preparation of Probes

In order to identify clones from the genomal library containing histone gene sequences, a combination of two chicken histone probes were used. These included a probe enriched for histone mRNA sequences and a probe containing genomal histone coding sequences. This strategy excludes selection for other sequences in the mRNA population as well as for possible non-coding spacer cross-reacting sequences in the chicken genomal probe.

(a) Chicken Histone mRNA

A histone mRNA enriched fraction from 5-day old chick total RNA was prepared by the method described by Crawford *et al.*,1979). Two successive sucrose gradient centrifugations resolved a 7S to 11S RNA fraction (Fig. 4.2) which has been shown previously to contain histone mRNA sequences (Krieg, 1980).

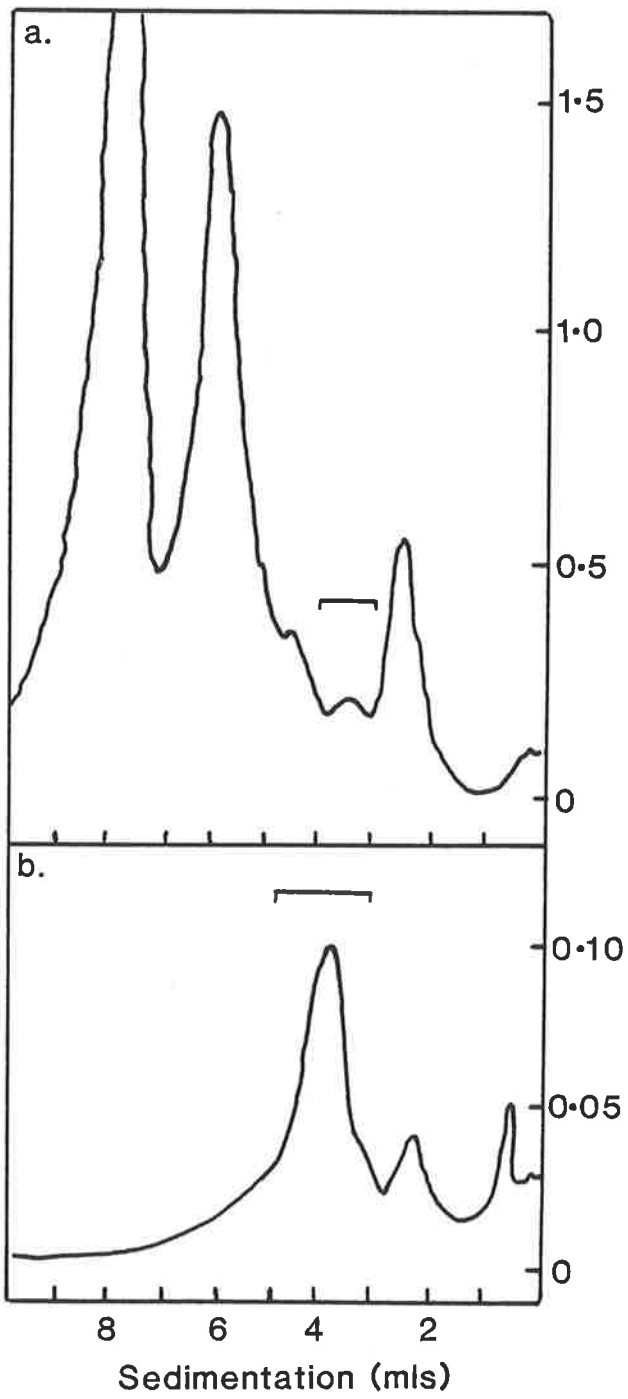
Fig. 4.2

Isolation of Chicken Histone mRNA

RNA was prepared from 5 day old chick embryos as described in Methods (2.2.1).

a. 10-40% sucrose gradient fractionation of total cellular RNA.

b. Rerun of 7-11S RNA fraction selected from the first gradient (bracketed region).


a.
1·5
1·0
0·5
0
b.
0·10
0·05
0
8
6
4
2
Sedimentation (mls)
A260

(b) Chicken Histone Genomal DNA

Total DNA from the chicken genomal clone λCH.O1 could not be used to screen the human library since the Charon 4A arms in λCH.O1 would cross-react with every recombinant in the library. The 3.3 kb EcoR1 fragment from λCH.O1 was sub-cloned into pBR325, since this EcoR1 fragment had previously been shown by sequencing to contain histone genes (Harvey and Wells, 1979).

DNA was prepared from λCH.O1 as described in methods (2.2.6a), and digested with EcoR1. This DNA was ligated with pBR325 DNA which had been digested with EcoR1 and then incubated with bacterial alkaline phosphatase, to remove the 5' phosphates so as to prevent self-ligation of the EcoR1 ends. A sample of the ligation mix was transformed into *E. coli* 392 and plated onto L+ tet plates. One hundred transformants were repicked onto L+ cap and L+ tet plates. A number of transformants which were chloramphenicol sensitve (i.e. they contained inserts in the EcoR1 site which is located in the chloramphenicol resistance gene of pBR325) were selected and plasmid DNA isolated from them by the procedure described in Methods (2.2.9a).

Track (b) Fig. (4.3) shows an EcoR1 digest of one of these plasmids which contained the 3.3 kb EcoR1 fragment, alongside an EcoR1 digest of λCH.O1 Track (a) Fig. (4.3). When this digest was hybridised by the Southern transfer technique to cDNA made to chicken histone mRNA, the 3.3 kb EcoR1 fragment in the pBR325 subclone was detected Track (c)

Fig. 4.3

Characterisation of pCH3.3E

DNA from the transformant pCH3.3E (derived from an EcoR1 digest of λCH.O1 ligated with pBR325) was digested with EcoR1, and the DNA transferred to nitrocellulose and hybridised with chicken histone cDNA. The tracks are as follows -

A: λCH.O1 digested with EcoR1
B: pCH3.3E digested with EcoR1
C: pCH3.3E hybridised with chicken histone cDNA

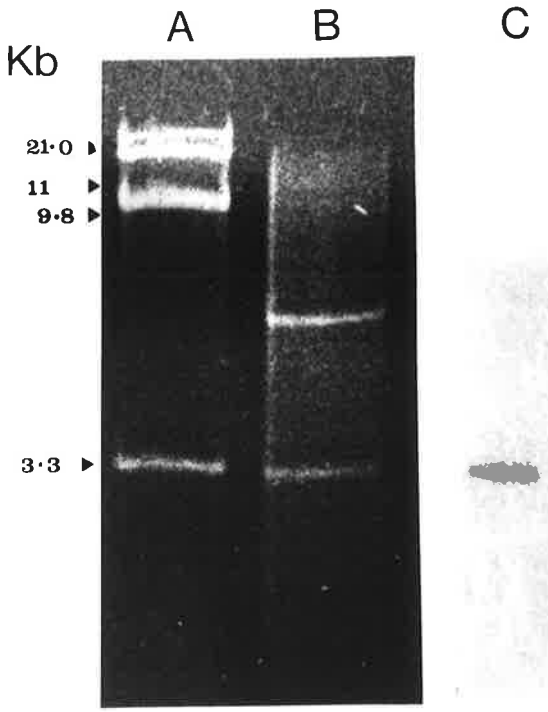

A
B
C
Kb
21·0
11
9·8
3·3

Fig. (4.3). This indicated that the 3.3 kb EcoRI fragment of λCH.01 was in fact subcloned and this plasmid was termed pCH3.3E.

4.2.2 Hybridisation of Chicken Histone Probes with Human DNA

Prior to screening the human genomic library, it was necessary to ensure that chicken histone gene sequences were sufficiently homologous to enable detection of human histone genes. To examine this, digests of human DNA (prepared as described in Methods (2.2.2) were transferred to nitrocellulose and hybridised with nicktranslated pCH3.3E DNA and random primed cDNA to the 7-11S RNA fraction.

Track (b) Fig. (4.4) shows that pCH3.3E detects a number of discrete bands in human DNA, whereas histone cDNA appears to detect only fragments corresponding to human ribosomal genes Track (c) Fig. (4.4) ~~(Arnheim and Bird,1977)~~. This is not surprising since the 7-11S RNA fraction is thought to be at least 80% ribosomal breakdown products (Krieg, personal communication). However, the chicken genomal probe which is enriched for histone genes appeared sufficiently homologous to detect specific fragments in human DNA. By using both probes in combination it appeared encouraging that clones containing human histone genes could be selected.

4.2.3 Screening of the Maniatis Human Library

The Maniatis human library consists of 1×10^6 independently derived bacteriophage λ clones containing large (15-20 kb) fragments of human DNA covalently joined to bacteriophage λ vector, as described by Maniatis *et al.*,(1978). The number of phage required

Fig. 4.4

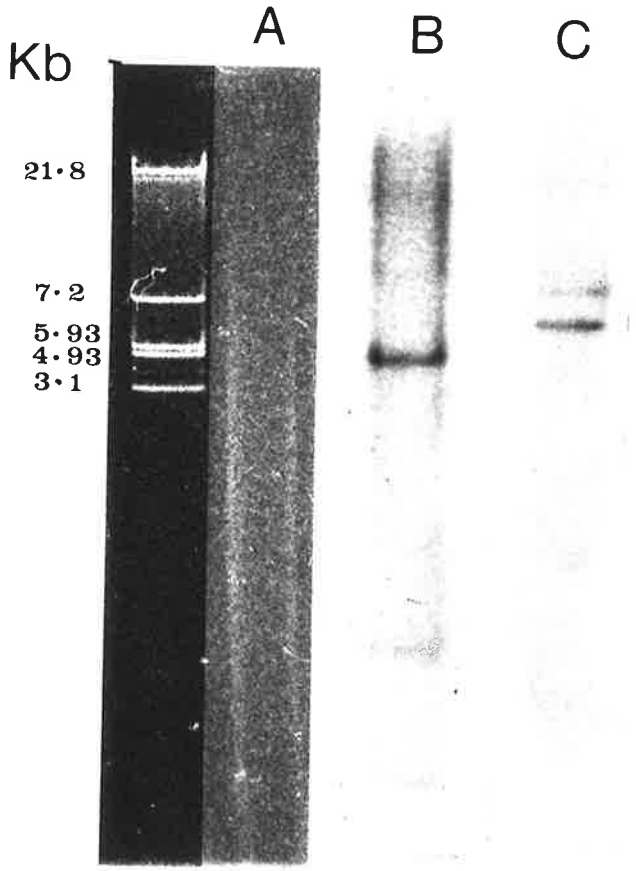
Cross-reaction of Chicken Histone DNA Probes with Human DNA

Human DNA was digested with EcoR1, electrophoresed on a 1% agarose gel, transferred to nitrocellulose and hybridised with pCH3.3E, chicken histone cDNA and 18 and 28 s rRNA. The tracks are as follows-

A: Human DNA digested with EcoR1

B: pCH3.3E hybridisation

C: Chicken histone cDNA hybridisation


Kb
A
B
C
21·8
7·2
5·93
4·93
3·1

for a "complete library" (i.e. a library having >99% probability of containing any sequence present in the genome) is 8×10^5. Since human histone genes are thought to be repeated 20-40 times per genome (Wilson and Melli, 1977), fewer recombinants are required to be screened in order to detect one containing histone sequences.

A total of 240,000 phage were screened by plating 8 (15 cm diameter) plates each containing 30,000 phage. Duplicate nitrocellulose filters were made from each plate using the procedure of Benton and Davis (1977). One set of filters was hybridised with 8×10^6 cpm of cDNA made to histone mRNA. The other set of filters was hybridised with 10×10^6 cpm of 3.3 kb EcoR1 fragment which had been previously prepared by digesting pCH3.3E with EcoR1, nicktranslating with $\{^{32}P\}$-dCTP and then electrophoresing on an 0.8% agarose gel. The 3.3 kb EcoR1 labelled band was excised, boiled and added directly to the hybridisation solution. It was necessary to separate the 3.3 kb fragment from the plasmid because pBR325 was found to cross-react weakly with the phage Charon 4A arms.

After washing the filters several times at 65°C in 1 x SSC, .1% SDS, they were autoradiographed with an intensifying screen at -80°C overnight. Histone cDNA detected at least 40 plaques per filter, as shown in Fig. (4.5b); the majority of these clones however, probably contained ribosomal or globin genes, since the RNAs for these genes are the major contaminants of histone mRNA preparations (Crawford *et al.*, 1979). However, the 3.3 kb EcoR1 probe detected

Fig. 4.5

Screening Human Library

The human library was screened as described in Methods (2.2.5).

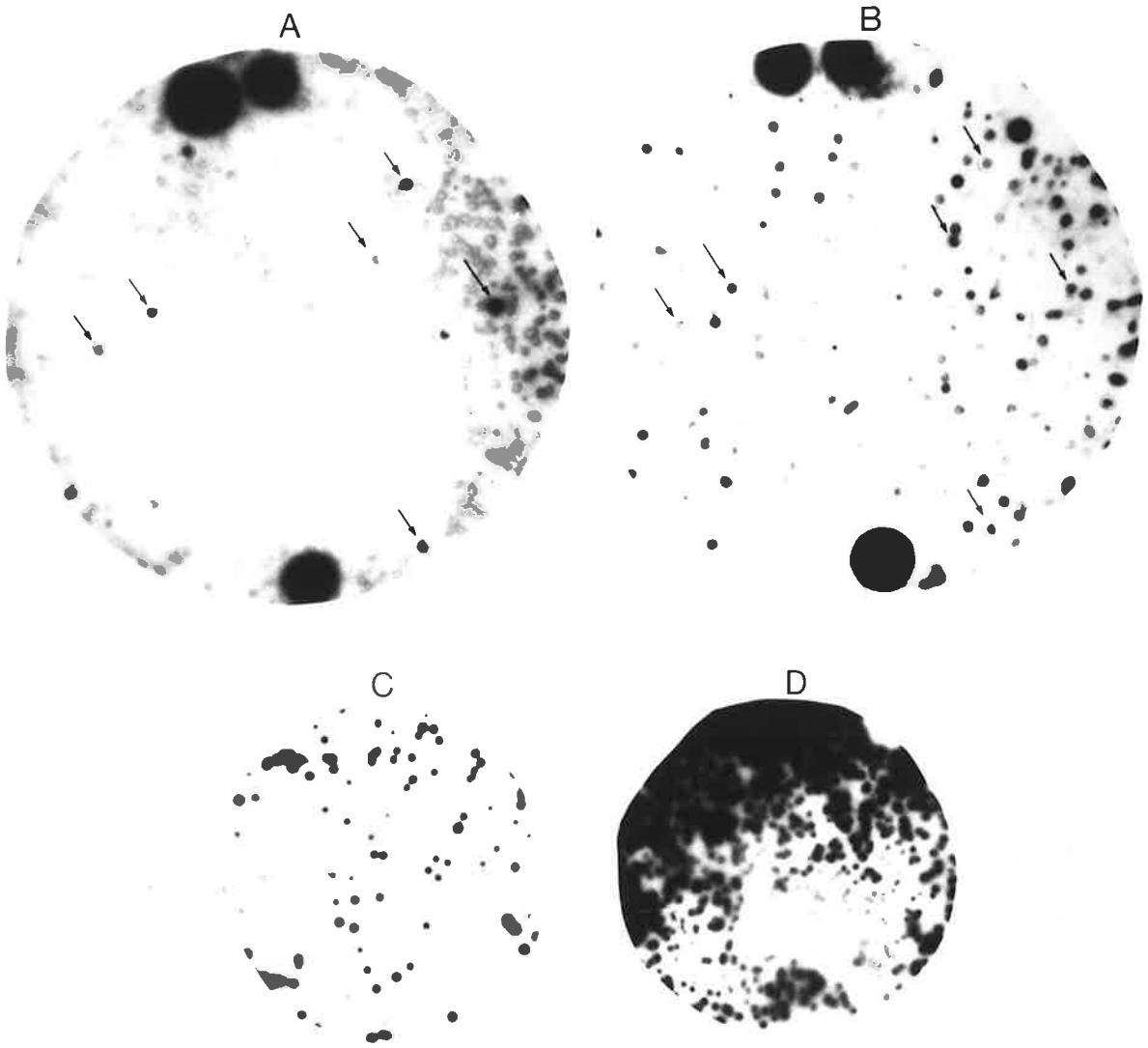
A: Duplicate filter hybridised with the 3.3 kb EcoRI fragment from pCH3.3E.

B: First round screening hybridised with chicken histone cDNA.

C: Second round screening.

D: Third round screening.

Arrows indicated plaques which hybridise with both the chicken histone cDNA and the 3.3 kb EcoRl DNA.


A
B
C
D

fewer recombinants Fig. (4.5a), demonstrating the higher specificity of this probe. The duplicate filters were aligned and approximately twenty phage gave positive signals with both hybridisation probes, and the arrows in Fig. (4.5) demonstrate some examples. Twelve of these recombinants were picked and approximately 100 phage were replated on each plate. Duplicate filters were again prepared and rescreened with both histone probes to not only verify the original positive hybridisation, but to also allow selection of the appropriate single plaque (Fig. 4.5c). A single plaque was then picked for each of the twelve positive recombinants and plated for third round screening to ensure the purity of each phage (Fig. 4.5d).

DNA was prepared from six isolates and as can be seen from Fig. (4.6), there are only three independent clones.

Since clone 5 gave the strongest positive signal with both the cDNA and genomal histone probe, it was decided to select this clone to investigate further, as it was likely to contain the highest representation of histone sequences. It was thought that if the human histone genes were organised in a 6-7 kb repeat unit similar to sea urchin and *Drosophila*, then this clone could potentially contain the entire histone repeating unit. This clone, which was initially termed λHH.01 (Clark *et al.*, 1981), will be now referred to as λCH.02 throughout the rest of the thesis, due to its subsequent characterisation.

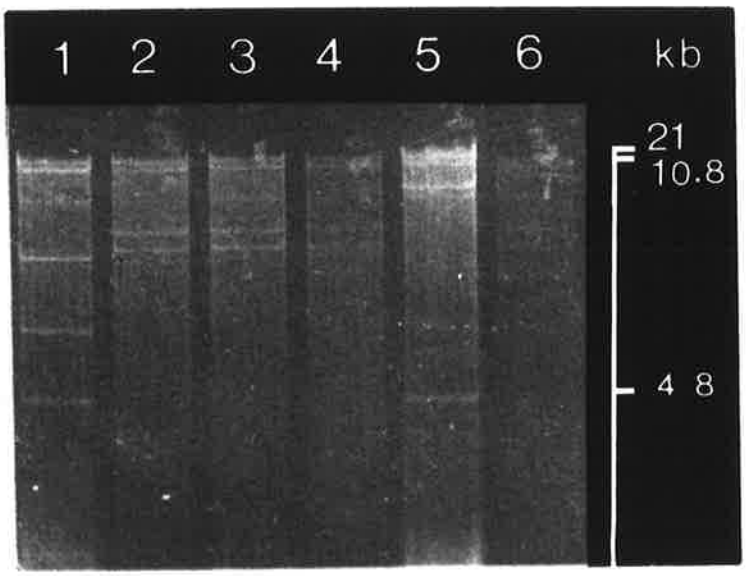
Fig. 4.6

DNA Digests from "Positive" Isolates

DNA was prepared from six of the positive phage isolates, digested with EcoR1 and electrophoresed on a 1% agarose gel.

The tracks are as follows -

Clone 1
Clone 2
Clone 3
Clone 4
Clone 5
Clone 6


1
2
3
4
5
6
kb
21
10.8
4 8

4.2.4 Preparation of Phage DNA

DNA was prepared from λCH.02 by a number of methods. Initially the liquid culture method was used as described in methods (2.2.4b) however, this method proved to be particularly unsuitable for recombinant clones in Charon 4A vectors, as the DNA routinely degraded during restriction digests. The formamide lysis method was also tried, but the resulting phage DNA failed to be digested with any restriction enzymes. Finally, a plate-stock method was developed as described in Methods (2.2.6b) followed by a CsCl step gradient and a sucrose gradient; this resulted in good yields of phage DNA that would digest with most restriction enzymes. However, HindIII restriction endonuclease would usually only digest after prolonged incubations or in double digestions with other enzymes. This effect could be reduced by dialysis of the phage DNA prior to digestion.

4.2.5 Restriction Analysis of λCH.02

An EcoRl/HindIII restriction map of λCH.02 was initially constructed. Digestion of λCH.02 with EcoRl resulted in two insert fragments, 4.8 kb and 9.6 kb in size (Track (b), Fig. 4.7). The 4.8 kb EcoRl fragment is not digested with HindIII, whereas the 9.6 kb EcoRl fragment digests with HindIII to give a 4.3 EcoRl/HindIII fragment, a 2.6 kb HindIII fragment, a 1.6 kb HindIII fragment and three smaller fragments less than 500 bp (Track (d) Fig. 4.7).

An EcoRl/HindIII digestion of λCH.02 was electrophoresed on a 1% agarose gel and the fragments transferred to nitrocellulose.

Fig. 4.7

Restriction Enzyme Digests of λCH.02

DNA from λCH.02 was digested with EcoRI and HindIII, electrophoresed on a 1% agarose gel. The EcoRI/HindIII digest of λCH.02 was transferred to nitrocellulose and hybridised with chicken histone cDNA and autoradiographed at -80°. The tracks are as follows:

A: λ digested with HindIII

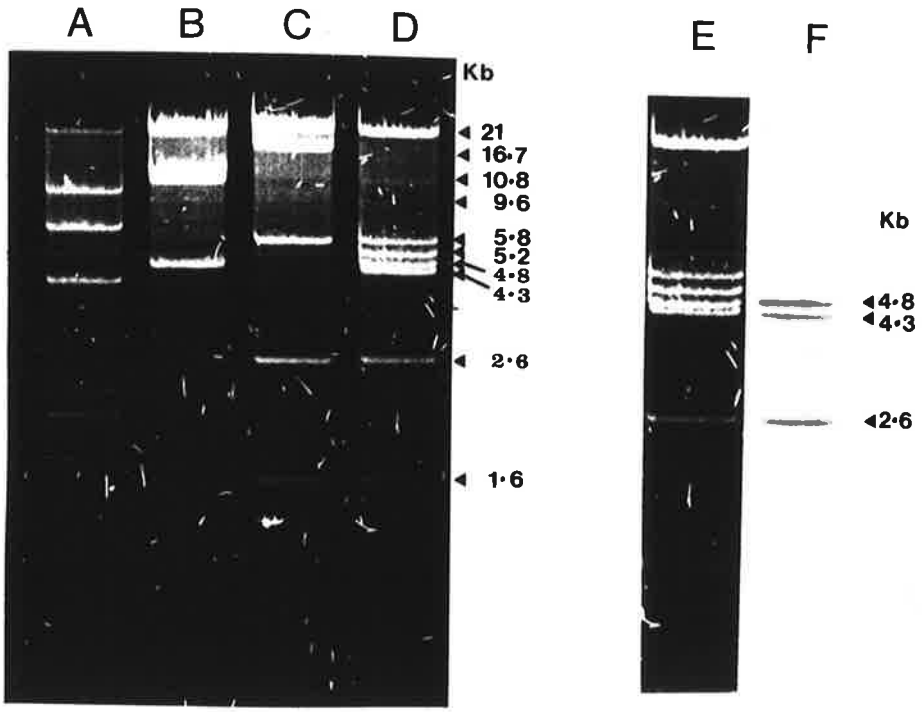
B: EcoRI λCH.02

C: HindIII λCH.02

D: EcoRI/HindIII λCH.02

E: EcoRI/HindIII λCH.02

F: Chicken histone cDNA hybridisation

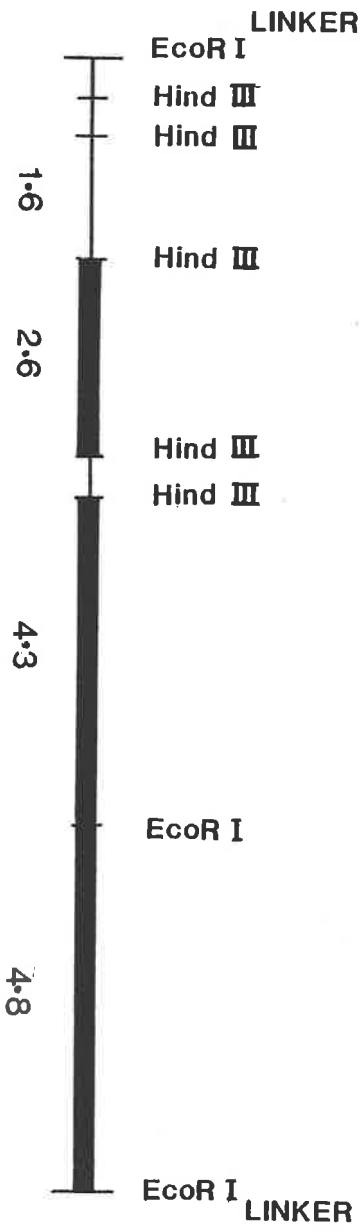

A
B
C
D
Kb
21
16·7
10·8
9·6
5·8
5·2
4·8
4·3
2·6
1·6
E
F
Kb
4·8
4·3
2·6

This filter was hybridised with chicken histone cDNA in order to locate which restriction fragments of λCH.02 contain histone coding sequences. This probe should detect all the histone genes in the clone as it contains sequences for the 5 chicken histone proteins (Crawford *et al.*, 1979). Track (f), Fig. (4.7) shows that the 4.8 kb EcoRl fragment, 4.3 kb EcoRl/HindIII fragment and the 2.6 kb HindIII fragment all hybridise to the histone cDNA probe. These three coding fragments are indicated in a simple EcoRl/HindIII restriction map of λCH.02 Fig. (4.8). On the basis of this data, the histone coding potential of λCH.02 appeared to be extensive, as it could span a maximum of 12 kb of the 14.9 kb clone.

Fig. 4.8

Restriction Endonuclease Map of λCH.02 Insert

Blocked regions indicate the EcoRI/HindIII restriction fragments which hybridise with chicken histone cDNA. Map distances are shown in kilobase pairs.


LINKER
EcoR I
Hind III
Hind III
1·6
Hind III
2·6
Hind III
Hind III
4·3
EcoR I
4·8
EcoR I
LINKER

4.2.6 Isolation of λCH.02 Subclones

In order to define in more detail the restriction map of λCH.02 and the location of individual histone genes, the EcoR1 and HindIII restriction fragments containing coding sequences were subcloned into plasmid vectors.

EcoR1, HindIII or EcoR1/HindIII double digests of λCH.02 DNA were ligated with plasmid vector DNA which had been linearised with the appropriate enzymes and incubated with bacterial alkaline phosphatase to prevent self-ligation. A sample of each ligation mixture was transformed into *E. coli* LE392 and plated onto L+ Ap plates. Recombinant plasmids were assayed either by colony hybridisation (Fig. 4.9), sizing the DNA insert on agarose gels or by antibiotic sensitivity. The nomenclature of these subclones, as well as the type of plasmid vector they are ligated into, is summarised in Table (4.1). Restriction enzyme mapping of these subclones is considered in some detail below.

4.2.7 Restriction Analysis of λCH.02 Subclones

(a) pCH9.6E

pCH9.6E contains the 9.6 kb EcoR1 fragment of λCH.02 subcloned into the EcoR1 site of pKC7 (by A. Robins). The 9.6 kb EcoR1 fragment contains all the HindIII sites found in λCH.02, consequently pCH9.6E was used in determining the size and organisation of the smaller HindIII fragments within λCH.02.

A HindIII digest of pCH9.6E was electrophoresed on a 5% acrylamide gel and three small (490 bp, 440 bp and 340 bp)

Fig. 4.9

Colony Hybridisation

100 colonies from a transformation of HindIII digested λCH.02 ligated with pBR322 were picked onto nitrocellulose and incubated at 37^{o} overnight. The nitrocellulose filter was treated as described in Methods (2.2.8b) and hybridised with chicken histone cDNA and exposed to X-ray film at -80^{o} overnight.

TABLE 4.1

Subclone	Size of Fragment (kb)	Vector	Size of Insert	Method of Detection
pCH2.6H	2.6	pBR322	HindIII	Insert size
pCH4.3EH	4.3	pBR325	EcoRl/HindIII	Cap^S Insert size
pCH4.8E	4.8	pBR322	EcoRl	Grunstein Insert size
pCH9.6E	9.6	pKC7	EcoRl	Cap^S Insert size

fragments were resolved Track (b) Fig. (4.10). The 490 bp fragment is restricted by EcoRl to give a 450 bp EcoRl/HindIII fragment and a 40 bp HindIII/EcoRl fragment from pKC7, Track (c) Fig. (4.10). In addition it was possible to locate the 340 bp HindIII fragment adjacent to the 450 bp EcoRl/HindIII fragment from a partial digest of pCH9.6E which gave a (490 + 340) = 830 bp fragment that was digested by EcoRl to give a (450 + 340) = 790 bp fragment (data not shown).

Furthermore, it was possible to locate the 440 bp HindIII fragment between the 2.6 kb HindIII fragment and the 4.3 bp EcoRl/HindIII fragment of λCH.02 by mapping the SalI sites in pCH9.6E. A SalI digest of pCH9.6E results in four SalI fragments (7.0 kb, 6.2 kb, 0.88 kb, 0.64 kb), Track (a), Fig. (4.11). The 0.64 kb SalI fragment is in pKC7. Therefore the 0.88 kb Sal fragment which is cut by HindIII cannot cover the HindIII site of pKC7 (as the nearest SalI site within pKC7 is 1.5 kb away); it therefore must be located within the 9.6 kb EcoRl insert overlapping the 2.6 kb HindIII fragment and the 4.3 kb EcoRl/HindIII fragment, as these two fragments are digested by SalI (Track (b), Fig. (4.11). In fact new fragments produced by the double digestion are a 300 bp SalI/HindIII fragment and a 200 bp SalI/HindIII fragment. This locates the 440 bp HindIII fragment between the 4.3 kb and 2.6 kb HindIII fragments to account for the SalI fragment which overlaps them. The location of all the HindIII fragments in pCH9.6E are summarised in Fig. (4.12). Other restriction digests and sequencing results are in agreement with this map.

Fig. 4.10

HindIII and EcoR1 Digests of pCH9.6E

HindIII and HindIII/EcoR1 digests of pCH9.6E were end-labelled as described in Methods (2.2.14d), electrophoresed on a 5% polyacrylamide gel and autoradiographed (the origin is indicated by O).

The tracks are as follows -

A: HinfI digested pBR322
B: HindIII digested pCH9.6E
C: EcoR1/HindIII digested pCH9.6E

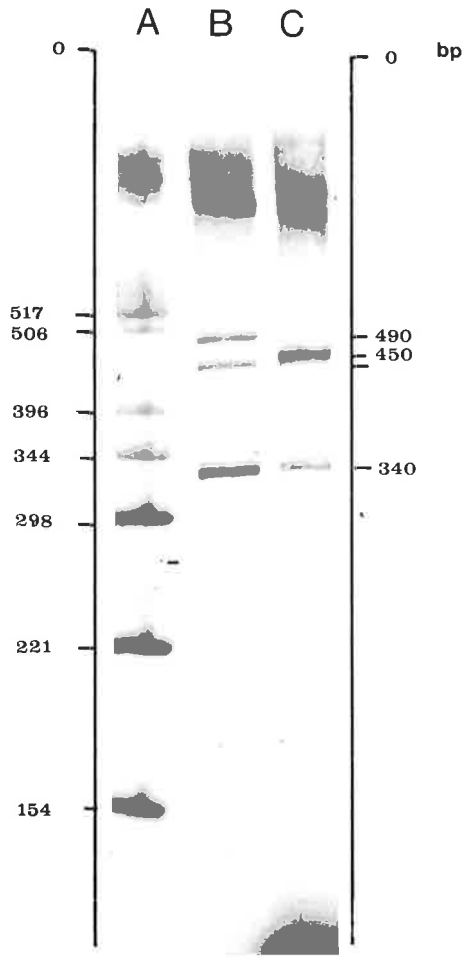

A
B
C
0
0
bp
517
506
396
344
298
221
154
490
450
340

Fig. 4.11

SalI Digests of pCH9.6E

DNA from pCH9.6E was digested with SalI and SalI/HindIII and electrophoresed on a 1.2% agarose gel. The tracks are as follows -

A: SalI digested pCH9.6E

B: SalI/HindIII digested pCH9.6E

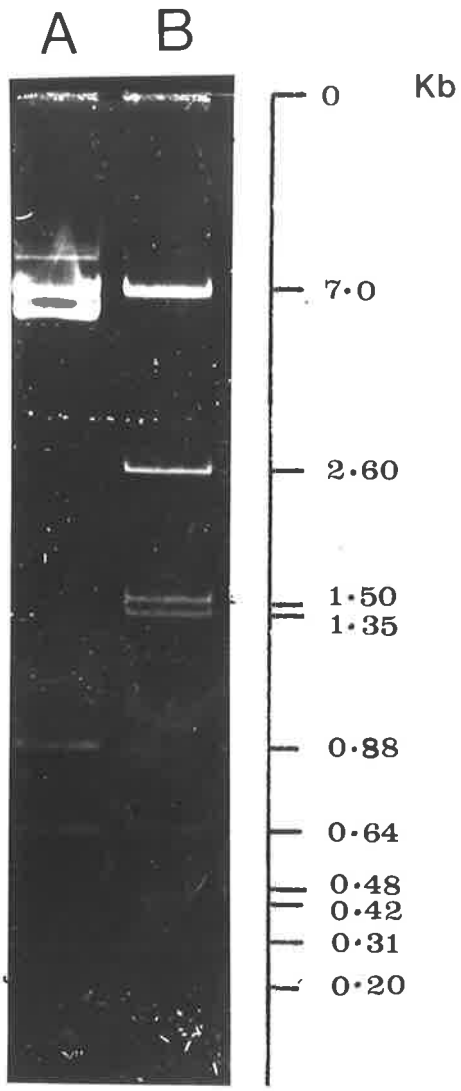
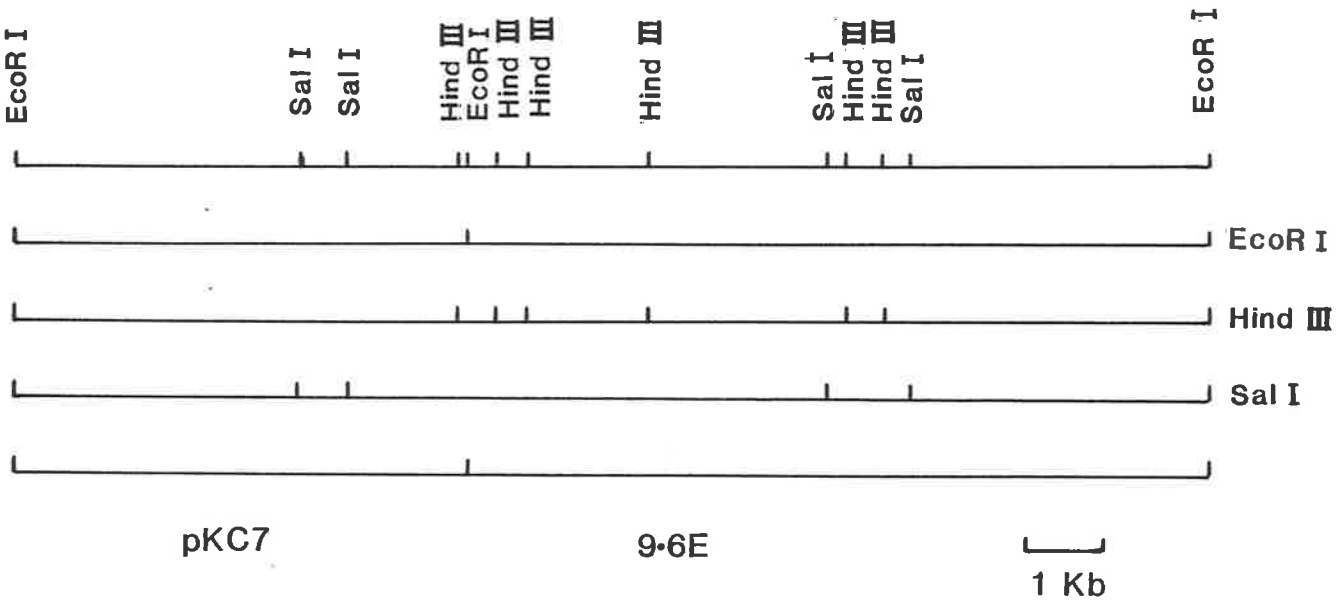

A
B
Kb
0
7·0
2·60
1·50
1·35
0·88
0·64
0·48
0·42
0·31
0·20

Fig. 4.12

Restriction Endonuclease Map of pCH9.6E

The left EcoRI restriction fragment is the vector pKC7 and the right fragment the 9.6 kb EcoRI insert.

pCH9•6E RESTRICTION MAP



(b) pCH2.6H

pCH2.6 contains the 2.6 kb HindIII histone coding fragment of λCH.02 subcloned into the HindIII site of pBR322. In order to define the location of the histone genes in more detail, DNA from pCH2.6H was mapped with SalI, SmaI, PvuII and SacII restriction endonucleases. Fig. (4.13) presents some of the relevant restriction digests which led to the construction of a restriction map for pCH2.6H.

A SalI digest of pCH2.6H gives a 3.8 kb and a 3.0 kb fragment, Track (b), Fig. (4.13), therefore there must be only one SalI site in the 2.6 kb HindIII insert, approximately 200 bp from the HindIII site, because the other SalI site is contained in pBR322, 600 bp from the opposite HindIII site. Fig. (4.14). The digest with SacII shows that there are two internal SacII sites in the 2.6 kb HindIII insert (pBR322 contains no SacII sites) which are 1.3 kb apart Track (h), Fig. (4.13). These can be located 750 bp and 570 bp respectively from the HindIII sites as determined by a SacII/HindIII double digestion of pCH2.6H (Track (i), Fig. 4.13).

The location of the 570 bp SacII/HindIII fragment is verified since it also contains the only SalI site in the 2.6 kb HindIII insert. This is shown by a SacII/SalI double digest of pCH2.6H which results in a 350 bp SacII/SalI fragment Track (g) Fig. (4.13). As there is only one PvuII site in pBR322 Fig. (4.14), there must be two PvuII sites in the 2.6 kb HindIII insert, 1.4 kb apart Track (e) Fig. (4.13).

Fig. 4.13

Restriction Endonuclease Digests of pCH2.6H

DNA samples of pCH2.6H were digested with restriction enzymes and electrophoresed on a 1% agarose gel. The tracks are as follows -

A: SalI/SmaI
B: SalI
C: PvuII
D: PvuII/HindIII
E: PvuII
F: SalI/PvuII
G: SalI/SacII
H: SacII
I: SacII/HindIII

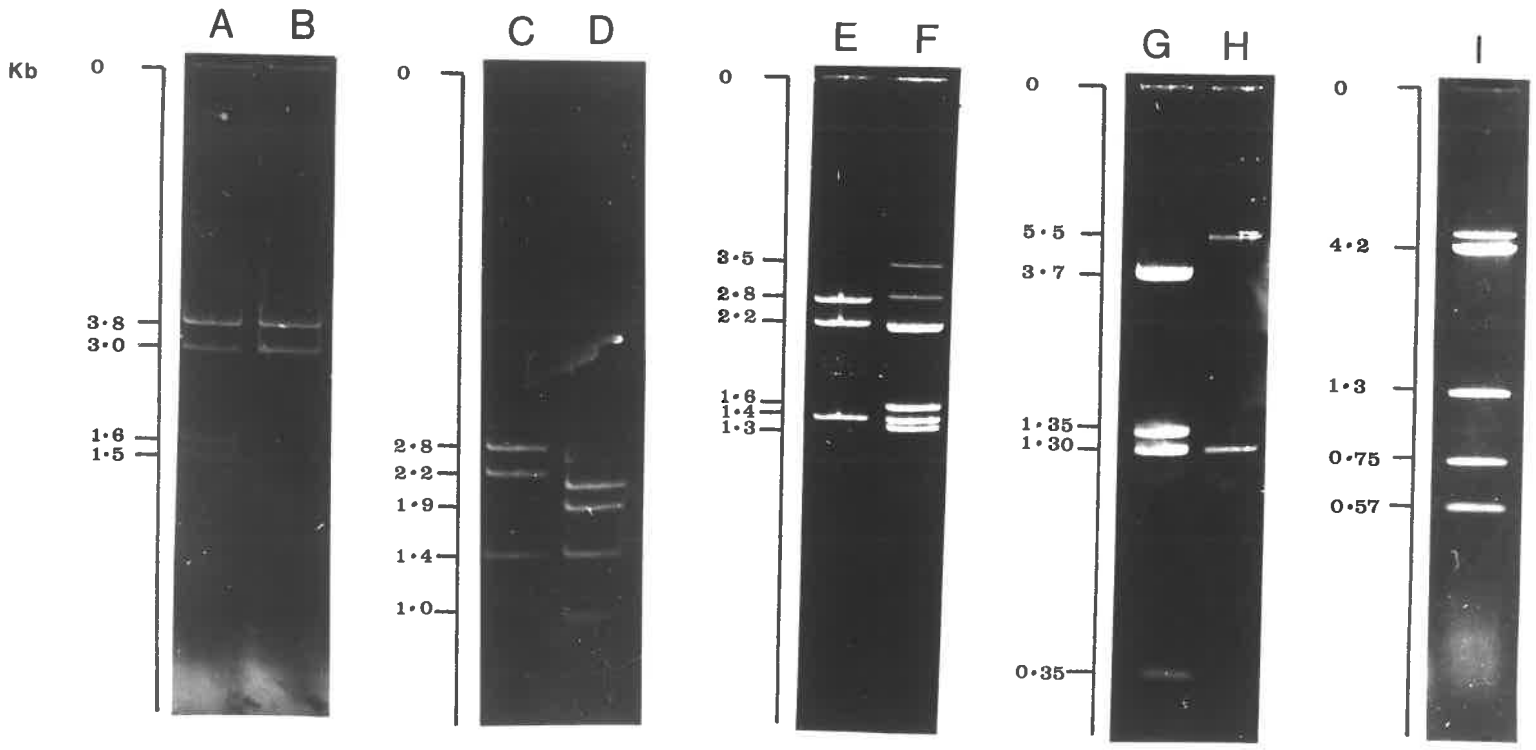
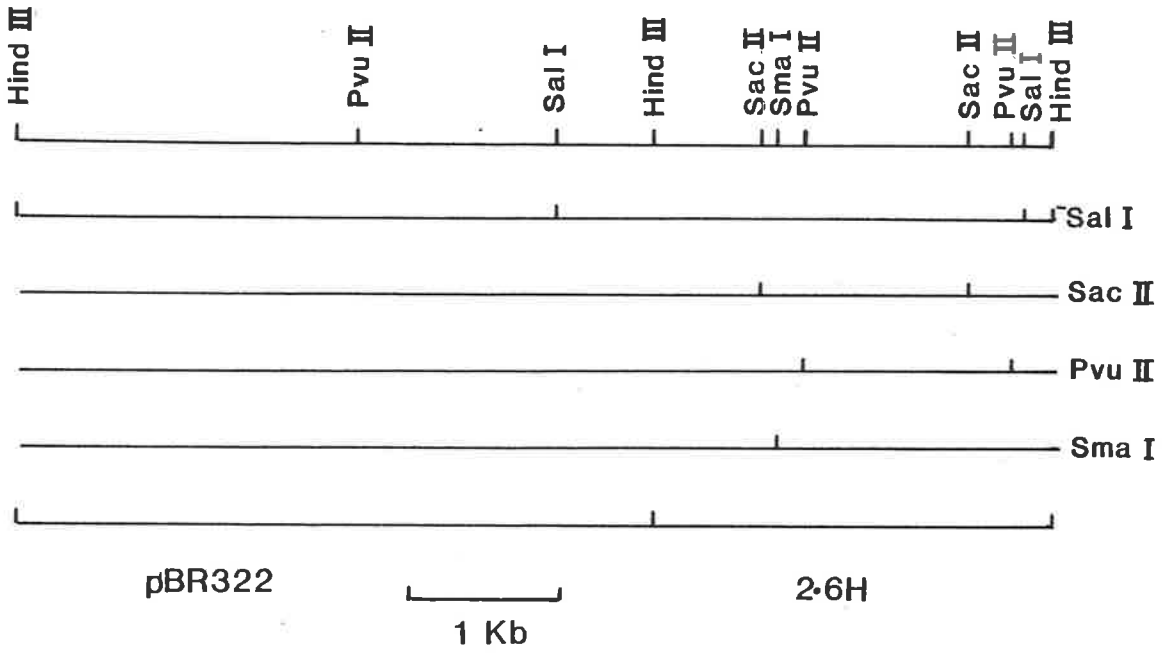

A
B
C
D
E
F
G
H
I
Kb
0
3·8
3·0
1·6
1·5
0
2·8
2·2
1·9
1·4
1·0
0
3·5
2·8
2·2
1·6
1·4
1·3
0
5·5
3·7
1·35
1·30
0·35
0
4·2
1·3
0·75
0·57

Fig. 4.14

Restriction Endonuclease Map of pCH2.6H

The left HindIII restriction fragment is the vector pBR322 and the right fragment is the 2.6 kb HindIII insert.

pCH2·6H RESTRICTION MAP



The location of these PvuII sites with respect to the HindIII and SalI sites are shown in Tracks (d) and (f), Fig. (4.13). There are no SmaI sites in pBR322, therefore there must be only one SmaI site in the 2.6 kb HindIII insert, 1.6 kb from the SalI site. Track (a) Fig. (4.13). These restriction sites have been verified by other double restriction digestions and the resulting restriction map of pCH2.6H is presented in Fig. (4.14).

(c) pCH4.3EH

The 4.3 kb EcoRl/HindIII histone coding fragment from λCH.02 was subcloned into the EcoRl/HindIII site of pBR325 to give pCH4.3EH. DNA from pCH4.3EH was mapped using SalI, SacII, SmaI and PvuII restriction endonucleases as shown in Fig. (4.15) and Fig. (4.16).

The single SalI site within the insert is most easily mapped. There is one SalI site in pBR325, 620 bp from the HindIII site, therefore there must be only one SalI site in the 4.3 kb EcoRl/HindIII insert 380 bp in from the HindIII site, since a SalI digest of pCH4.3EH results in a 8.35 kb and a 1.05 kb fragment track(a) Fig. (4.15). As there are no SacII sites in pBR322, there must be two internal SacII sites in 4.3 kb insert which are 1.43 kb apart Track (b) Fig. (4.15). One of these SacII sites is located 2.2 kb from the EcoRl site as determined by the SacII/EcoRl double digest Track (d), Fig. (4.15) and the other SacII site is located 660 bp from the HindIII site, as determined by the HindIII/EcoRl/SacII triple digest of pCH4.3EH (track c).

Fig. 4.15

Restriction Endonuclease Digestions of pCH4.3EH

DNA from pCH4.3E was digested with restriction enzymes and electrophoresed on a 1% agarose gel. The tracks are as follows -

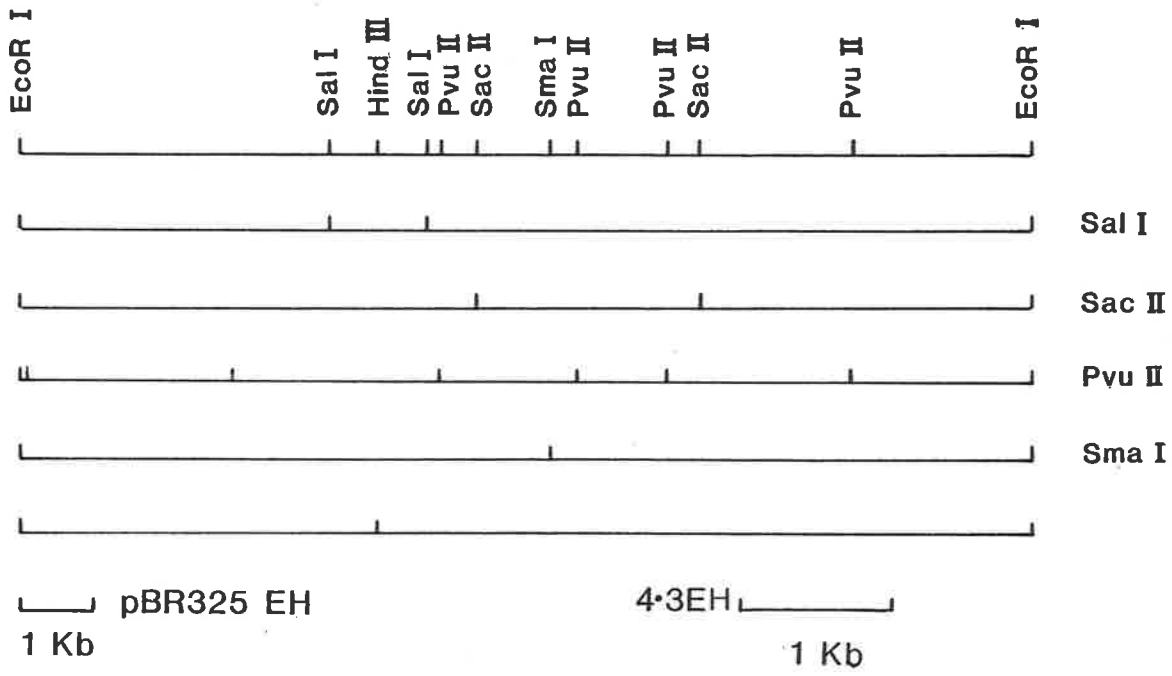
A: SalI
B: SacII
C: SacII/HindIII/EcoRI
D: SacII/EcoRI
E: PvuII/HindIII
F: PvuII/EcoRI
G: PvuII

Fig. 4.16

Restriction Endonuclease Map of pCH4.3EH

The left EcoRI/HindIII restriction fragment is the bector pBR325 and the right fragment is the 4.3 kb EcoRI/HindIII insert. Note that the vector and insert regions are drawn to different scales.

pCH 4·3 EH RESTRICTION MAP



Since there are two PvuII sites in pBR325 Fig. (4.16), there must be four PvuII sites in the 4.3 kb insert Track (g). An EcoRl/PvuII double digest indicates that one PvuII site is 1.25 kb from the EcoRl site, track (f),track (e) (Fig.4.15) indicates that there is a PvuII site 400 bp from the HindIII site. The placement of the other PvuII sites were verified by SacII/PvuII double digestions. The resulting restriction map of pCH4.3EH is shown in Fig. (4.16).

(d) pCH4.8E

pCH4.8E contains the 4.8 kb EcoRl fragment which contains histone coding sequences subcloned into the EcoRl site of pBR322. DNA from pCH4.8E was mapped using SmaI, SacII and PvuII restriction endonucleases by P.Krieg and Fig. (4.17) shows the resulting restriction map constructed for pCH4.8E.

4.2.8 Detailed Restriction Map of λCH.02

In order to complete the restriction map of λCH.02, it was necessary to orient the various EcoRl and HindIII subclones(Fig. 4.19), pCH2.6H can be oriented by the position of its single SalI site with respect to the SalI sites in pCH9.6E (Fig. 4.12). The orientation of pCH2.6H was also verified by analysis of a SmaI digest of λCH.02 Track (b), Fig.(4.18). The 3.4 kb SmaI fragment of λCH.02 is digested by HindIII to give a 1.8 kb SmaI/HindIII fragment, a 440 bp HindIII fragment and a 1.15 kb SmaI/HindIII fragment Track (a). Since the 440 bp HindIII fragment is known to lie between the 2.6 kb HindIII fragment and the 4.3 kb EcoRl/HindIII

Fig. 4.17

Restriction Endonuclease Map of pCH4.8E

The left EcoRI fragment is the vector pBR322 and the right is the 4.8 kb EcoRI insert.

pCH 4·8E RESTRICTION MAP

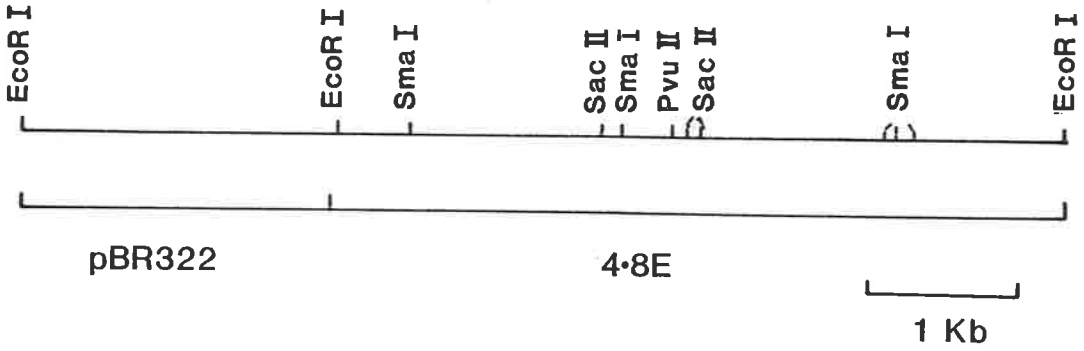
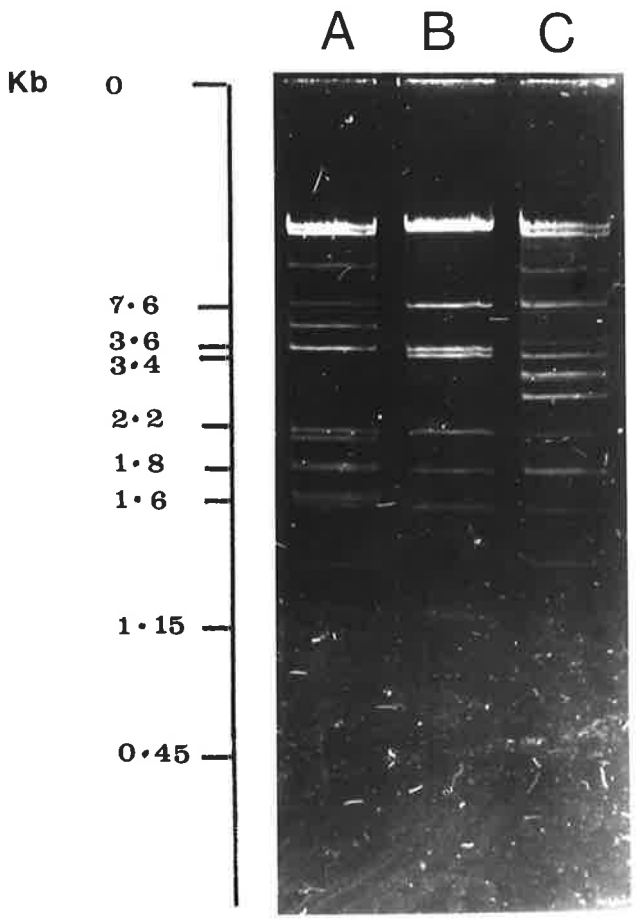


Fig. 4.18

SmaI Digests of λCH.02

DNA from λCH.02 was digested with SmaI, EcoRI and HindIII and electrophoresed on a 1% agarose gel. The tracks are as follows -

A: SmaI/HindIII
B: SmaI
C: SmaI/EcoRI


A
B
C
Kb
0
7·6
3·6
3·4
2·2
1·8
1·6
1·15
0·45

fragment in pCH9.6E (Fig. 4.12) and also it is known that pCH4.3EH contains a SmaI site 1.15 kb from the HindIII site, it was concluded that the single SmaI site in pCH2.6H (Fig. 4.14) must be 1.8 kb from the 440 bp HindIII fragment in λCH.02.

pCH4.8E can also be oriented by the SmaI digest of λCH.02 Track (b), Fig. (4.18). As can be seen from Track (c), Fig. (4.18) the 3.6 kb SmaI fragment of λCH.02 is digested with EcoR1 to give a 0.45 kb and 3.15 kb SmaI/EcoR1 fragment.

Since it is known that the single SmaI site in pCH4.3EH is 3.15 kb from the EcoR1 site Fig. (4.16), then the 4.8 kb EcoR1 fragment must be orientated so that the 0.45 kb SmaI/EcoR1 fragment is adjacent to the 4.3 kb EcoR1/HindIII fragment.

By combining the restriction maps constructed for the various λCH.02 subclones, a complete restriction map of λCH.02 can be prepared as shown in Fig. (4.19). The location of the histone genes could now be defined in more detail, by hybridising the appropriate fragments from the λCH.02 subclones with individual histone DNA probes.

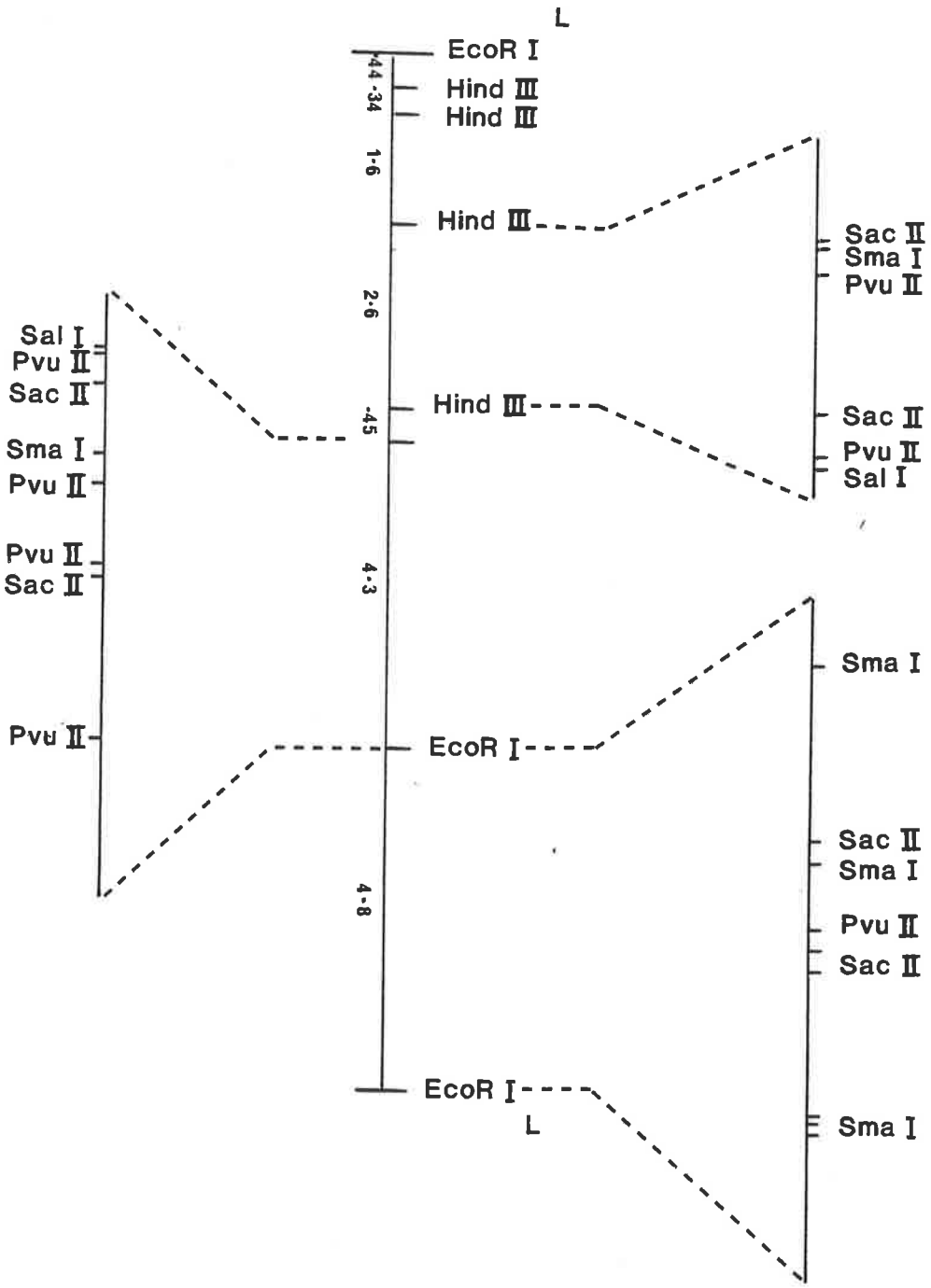
4.2.9 Hybridisation Analysis of λCH.02 Subclones

To locate the potential coding regions within λCH.02, a selection of digestions of the various subclones were electrophoresed on 1% agarose gels, and the DNA fragments were transferred by the Southern technique to nitrocellulose filters. These filters were hybridised with both chicken histone cDNA and the available hybridisation probes for individual histone genes.

Fig. 4.19

Restriction Endonuclease Map of λCH.02

The map distances are shown in kilobase pairs. $EcoRI^{L}$ denotes the linkered EcoRI ends of the λCH.02 insert.


L
EcoR I
Hind III
Hind III
·44
·34
1·6
Hind III
Sac II
Sma I
Pvu II
2·6
Sal I
Pvu II
Sac II
Hind III
·45
Sac II
Pvu II
Sal I
Sma I
Pvu II
Pvu II
Sac II
4·3
Sma I
Pvu II
EcoR I
Sac II
Sma I
4·8
Pvu II
Sac II
EcoR I
L
Sma I
λ CH·02 RESTRICTION MAP

(a) pCH2.6H

Digests of pCH2.6H with PvuII/HindIII and SacII/HindIII were hybridised with chicken histone cDNA. The 1.4 kb PvuII fragment (Track (b), Fig. (4.20), as well as the 1.3 kb SacII and 570 bp SacII/HindIII fragments (Track (e), Fig. (4.21), all hybridised to the probe. Therefore the potential histone genes must lie within the 1.4 kb PvuII fragment.

By using hybridisation probes for individual genes, it is possible to define the number and type of histone genes contained within this PvuII fragment. The only hybridisation probes initially available were two plasmids containing a sea urchin H3 gene and H4 gene respectively. These were derived from the sea urchin gene cluster in clone h22 (Schaffner *et al.*, 1978). Even though the total sea urchin histone gene probe did not hybridise significantly to total human DNA (data not shown), it was thought that since H3 and H4 are the most highly conserved proteins, it may be possible to detect these genes if they occur in pCH2.6H.

The PvuII/HindIII and SacII/HindIII digests of pCH2.6H were hybridised with nicktranslated sea urchin H3 and H4 probes. The 1.4 kb PvuII fragment hybridised with both probes Track (c and d), Fig. (4.20) whereas the 1.3 kb SacII fragment hybridised with the sea urchin H3 probe and the 570 bp SacII/HindIII fragment hybridised weakly to the H4 probe, tracks (b and c) Fig. (4.21).

Fig. 4.20

Histone Coding Regions in pCH2.6H

DNA from pCH2.6H was digested with PvuII/HindIII, electrophoresed on a 1% agarose gel, the DNA transferred to nitrocellulose, hybridised with histone DNA probes, washed in 0.5 x SSC at 65^{o} and autoradiographed at -80^{o}.
The tracks are as follows -

A: pCH2.6H digested with PvuII/HindIII
B: Chicken histone cDNA probe
C: Sea urchin H3 DNA probe
D: Sea urchin H4 DNA probe
E: Chicken histone H3 DNA probe
F: Chicken histone H4 DNA probe

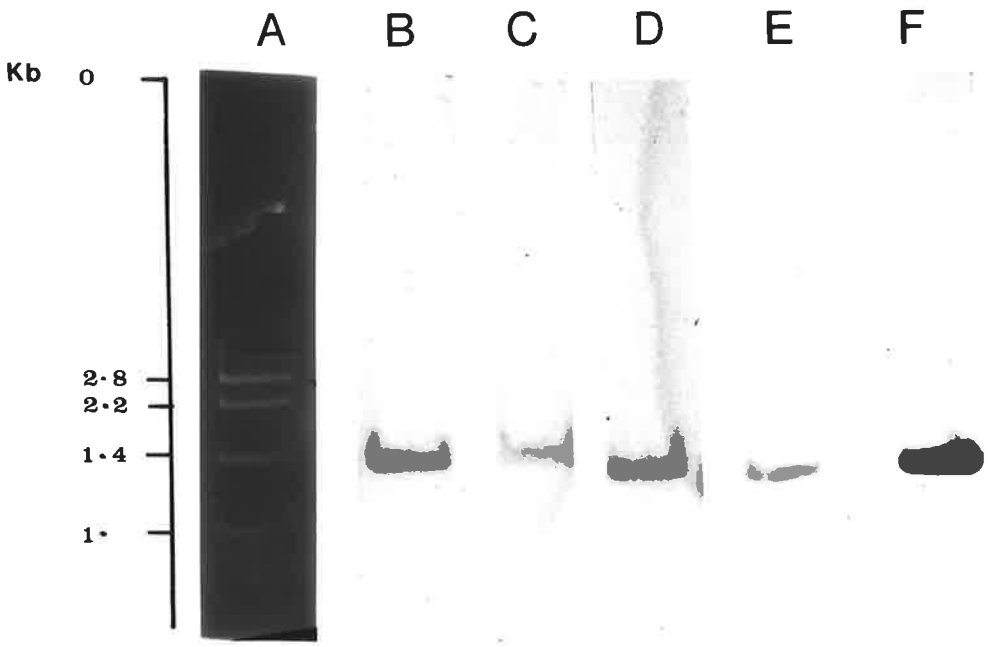

A
B
C
D
E
F
Kb
0
2·8
2·2
1·4
1·

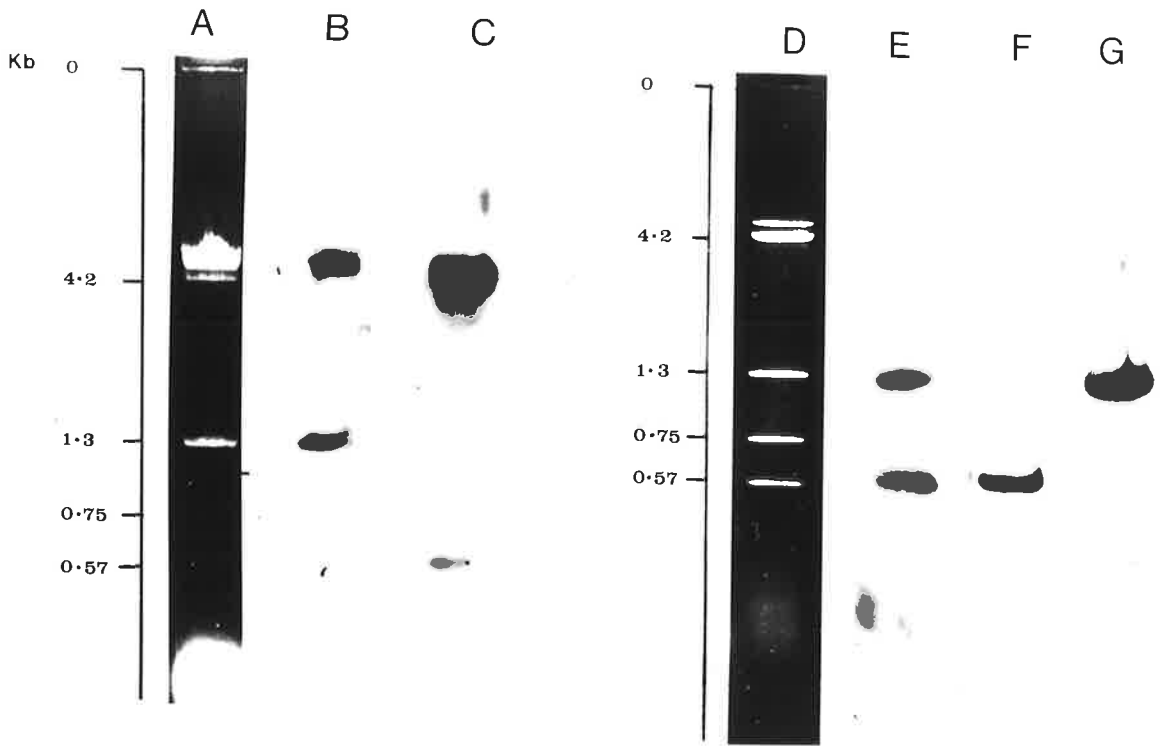
Fig. 4.21

Histone Coding Regions in pCH2.6H

DNA from pCH2.6H was digested with SacII/HindIII, electrophoresed on a 1% agarose gel, the DNA transferred to nitrocellulose filters which were then hybridised with histone DNA probes and washed in 0.5 x SSC at 65^{o}.
The tracks are as follows -

A: pCH2.6H digested with SacII/HindIII
B: Sea urchin histone H3 DNA probe
C: Sea urchin histone H4 DNA probe
D: pCH2.6H digested with SacII/HindIII
E: Chicken histone cDNA probe
F: Chicken histone H3 DNA probe
G: Chicken histone H4 DNA probe

Note that tracks A, B and C contain some linear pCH2.6H DNA which is detected by the probes.


A
B
C
D
E
F
G
Kb
0
4·2
1·3
0·75
0·57
0
4·2
1·3
0·75
0·57

Sequenced H3 and H4 chicken genes isolated from λCH.01 (Krieg, D'Andrea, unpublished data, 2.1.3) were later used as hybridisation probes to pCH2.6H. The 570 bp SacII/HindIII fragment hybridised strongly to the chicken H3 probe Track (f) Fig. (4.21), and the 1.3 kb SacII fragment hybridised strongly to chicken H4 probe Track (g), Fig. (4.21), and the 1.4 kb PvuII fragment hybridised to both, Fig. (4.20).

This obvious contradiction in hybridisation results observed between the sea urchin and chicken probes was resolved when the 570 bp SacII/HindIII fragment was later sequenced (Chapter 5) and was shown to contain an H3 histone gene, and similarly an H4 histone gene was located by sequencing in the 1.3 kb SacII fragment (Chapter 5). This was predicted only by the hybridisation data using chicken histone probes.

The anomoly observed using sea urchin hybridisation probes was investigated further by hybridising the 570 bp SacII/HindIII fragment which contains an H3 gene, back to DNA of the sea urchin H3 and H4 clones. Fig. (4.22) shows that both the sea urchin H4 DNA and H3 DNA hybridised similarly to the chicken H3 DNA probe. Since the H4 sea urchin probe contains all the BamHl spacer (Schaffner *et al.*, 1978), anomolous cross-reaction of these probes may only reflect spacer cross-reaction with pCH2.6H DNA, whereas the homology of the sea urchin histone genes to the histone sequence in pCH2.6H appears to be insufficient to favour hybridisation.

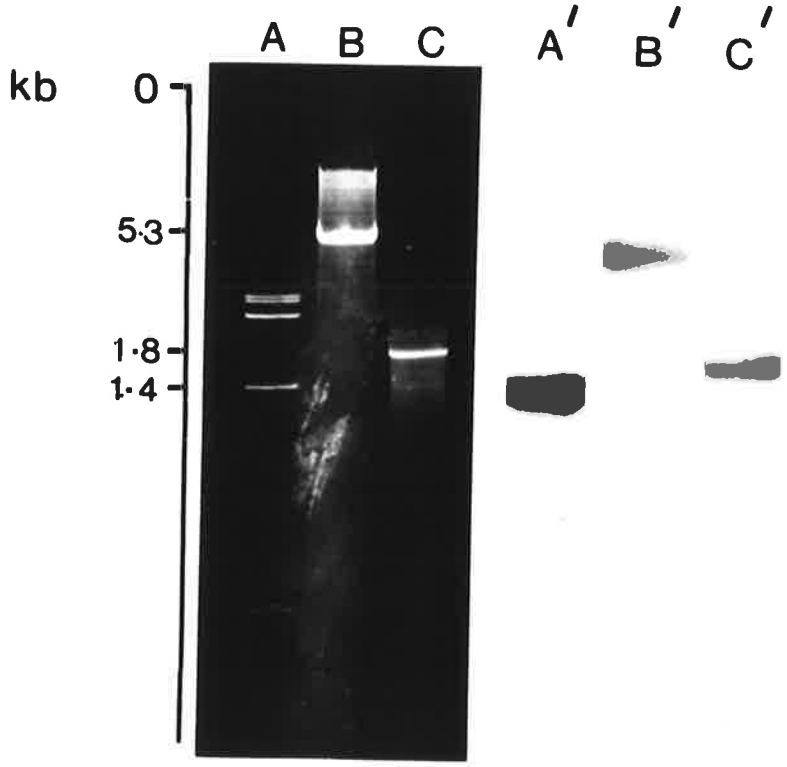
Fig. 4.22

Sea Urchin Histone Gene Cross Hybridisation

Sea urchin histone DNA coding for H3 and H4 genes was electrophoresed on a 1% agarose gel, transferred to nitrocellulose, hybridised with nick-translated 570 bp SacII/HindIII fragment from pCH2.6H and washed in 0.5 x SSC at 65^{o}.

The tracks are as follows -

A: p2.6H digested with PvuII
B: Sea urchin H4 EcoR1 digested
C: Sea urchin H3 DNA insert
A^{1}: Track A hybridised with 570 bp probe
B^{1}: Track B hybridised with 570 bp probe
C^{1}: Track C hybridised with 570 bp probe


A
B
C
A′
B′
C′
kb
0
5·3
1·8
1·4

Therefore previous results by other workers using sea urchin gene probes to analyse human and other higher vertebrate histone genes must be considered with caution.

Sequenced chicken DNA containing H2A, H2B and H1 genes were also hybridised to pCH2.6H DNA, however, no cross reaction was observed (data not shown). Therefore the 2.6 kb HindIII fragment from λCH.02 appears to contain an H3 gene in the 350 bp SacII/PvuII fragment and at least one H4 gene in the 980 bp SacII/PvuII fragment as summarised in Fig. (4.25a). Since these are the only fragments in pCH2.6H to hybridise with chicken histone cDNA is is likely that no other histone genes occur in the 2.6 kb HindIII fragment.

(b) pCH4.3EH

To locate the potential histone coding regions within the 4.3 kb EcoR1/HindIII fragment, a PvuII/EcoR1 double digest of pCH4.3EH DNA was hybridised with chicken histone cDNA. Track (b) Fig. (4.23) shows that both the 900 bp PvuII fragment and the 600 bp PvuII fragment hybridised with this probe.

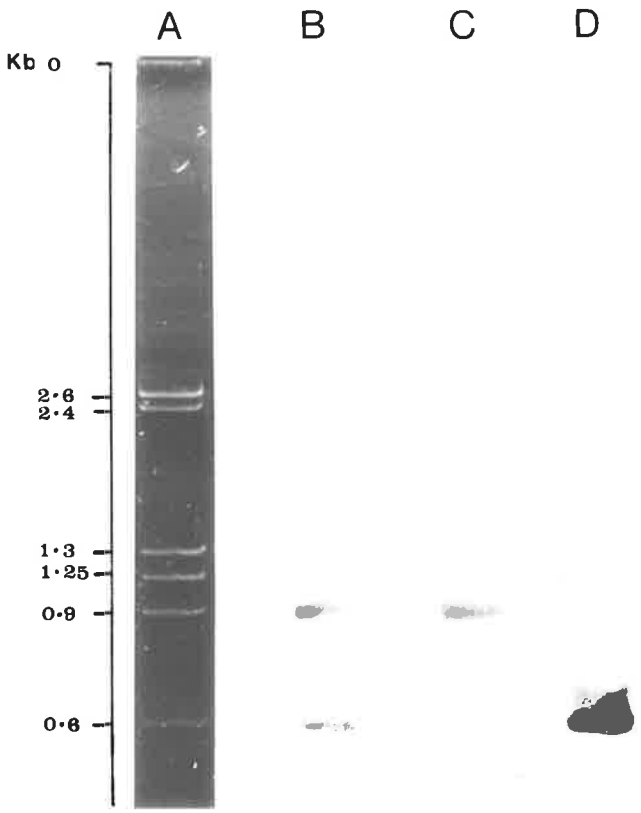
In order to determine which individual genes are contained in these fragments the PvuII/EcoR1 digest was hybridised independently with each of the sequenced chicken histone probes isolated from λCH.01 (H3, H4, H2A, H2B and H1) (2.1.3). As shown in Fig. (4.23) H3 hybridised to the

Fig. 4.23

Histone Coding Regions in pCH4.3EH

DNA from pCH4.3EH was digested with PvuII/EcoRI, electrophoresed on a 1% agarose gel, the DNA transferred to nitrocellulose filters, hybridised with chicken histone DNA probes and autoradiographed at -80^{o}.
The tracks are as follows -

A: pCH4.3E digested with PvuII/EcoRI
B: Chicken histone cDNA probe
C: Chicken histone H3 DNA probe
D: Chicken histone H4 DNA probe


A
B
C
D
Kb o
2·6
2·4
1·3
1·25
0·9
0·6

900 bp PvuII fragment Track (c), and H4 hybridised to the 600 bp PvuII fragment Track (d). H2A, H2B and H1 did not hybridise at all (data not shown). Since both coding fragments have histone genes assigned (Fig. (4.25b)) it was concluded that pCH4.3EH contained no other histone genes.

(c) pCH4.8E

The potential coding regions within the 4.8 kb EcoR1 fragment were located by a SmaI/EcoR1 double digest of pCH4.8E DNA which was hybridised with chicken histone cDNA. As shown in Track (b) Fig. (4.24) both the 1.8 kb SmaI fragment and the 1.4 kb SmaI fragment hybridised to this probe.

The same digest of pCH4.8E was then hybridised independently with the chicken histone sequences probes H3, H4, H2A and H2B, from λCH.01 (2.1.3). The histone genes H3 and H4 did not hybridise at all to pCH4.8E (data not shown), however, H2A hybridised weakly to the 1.8 kb SmaI fragment and H2B hybridised strongly to the 1.4 kb SmaI fragment. Tracks (c) and (d) Fig. (4.24).

From these results it was deduced that pCH4.8E contained no H3 or H4 genes and possibly contained an H2A gene in the 1.8 kb SmaI fragment and at least one H2B gene in the 1.4 kb SmaI fragment. However, when the 1.4 kb SmaI fragment was later sequenced (Chapter 5), it was found to completely contain an

Fig. 4.24

Histone Coding Regions in pCH4.8E

DNA from pCH4.8E was digested with SmaI/EcoRI, electrophoresed on a 1% agarose gel, the DNA transferred to nitrocellulose, hybridised with chicken histone DNA probes and autoradiographed at -80^{o}. The tracks are as follows -

A: pCH4.8E digested with SmaI/EcoRI
B: Chicken histone cDNA probe
C: Chicken histone H2B DNA probe
D: Chicken histone H2A DNA probe

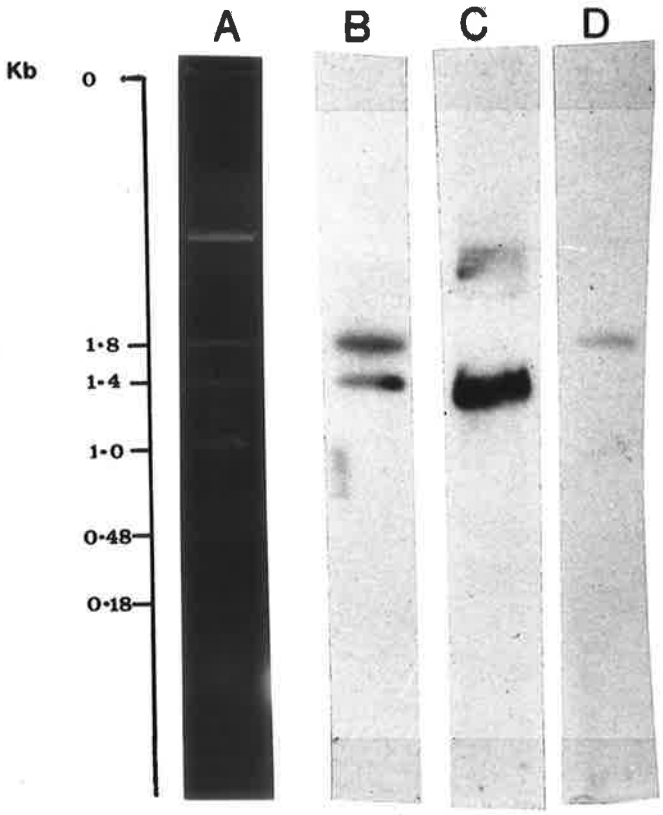
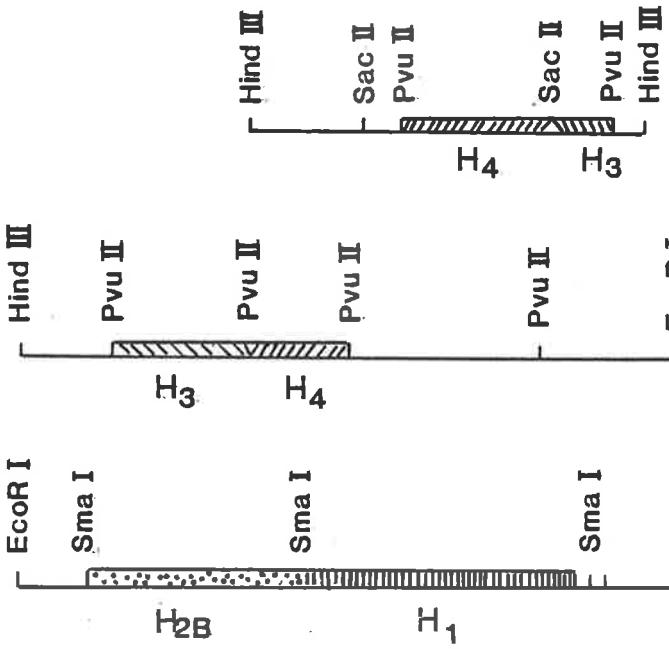

A
B
C
D
Kb
0
1·8
1·4
1·0
0·48
0·18

Fig. 4.25

Histone Gene Content of λCH.02 Subclones

The shaded regions in the maps delineate the restriction fragments to which individual histone gene probes hybridised.

The H1 probe is from the H1 gene characterised in Chapter 5.



(a) pCH 2·6H

(b) pCH 4·3EH

(c) pCH·4 8

H2B gene. Furthermore, when the 1.8 kb SmaI fragment was partially sequenced no H2A gene could be located, however, an H1 gene sequence was found.

~~Consequently~~ The hybridisation results were initially misleading and it became necessary to sequence each gene coding region to substantiate the identity of the histone genes. The location of the H1 and H2B genes in pCH4.8E is summarised in Fig. (4.25c).

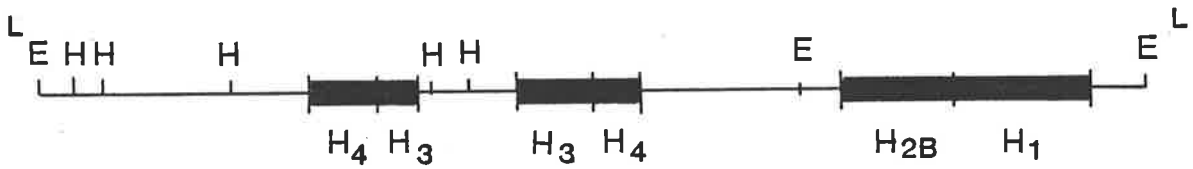
4.2.10 Location of Histone Genes in λCH.02

From the hybridisation data using both chicken histone cDNA and individual chicken histone gene probes, it was possible to locate six histone genes in λCH.02 (Fig. 4.26). Two H3/H4 gene pairs are found clustered together separated by at least 3.5 kb from an H1 and H2B gene; no H2A gene was located in λCH.02. The assignment of genes in λCH.02 appears to be complete because all the potential histone coding regions which hybridised with chicken histone cDNA have been identified as individual histone genes.

Fig. 4.26

Histone Gene Content of λCh.02

The blocked regions indicate the location of the histone genes from hybridisation data using individual chicken histone gene probes. E=EcoRI, H=HindIII and E^{L} denotes the boundaries of the λCH.02 insert.


L
E H H
H
H H
E
E
L
H4 H3
H3 H4
H2B
H1

4.3 DISCUSSION

During the isolation and initial characterisation of the clone λCH.02, following screening of the Maniatus human genomal library, its identification as a contaminant chicken clone was obscured for two reasons. Firstly, the HindIII restriction sites in λCH.02 had been initially incorrectly mapped (Harvey *et al.*, 1981) and the map did not correspond to that determined here and in Clark *et al.* (1981). Secondly, the identification of the genes in λCH.02 (Harvey *et al.*, 1981) had relied on hybridisation with the heterologous sea urchin H4 gene probe which failed to detect the H4 genes in the clone. In addition, it is also likely that λCH.1a, (a chicken recombinant histone clone recently isolated by Engel and Dodgson, (1981) from the same chicken library) is also identical to λCH.02. However, again only using sea urchin histone probes no H4 genes were detected and the H1 gene was identified as an H2A gene.

Because the clone was isolated by screening a human library, it was initially thought to be a human histone genomal clone, since its restriction map and gene content did not correspond to that determined for λCH.02 or any other isolated chicken genomal clone. Further restriction enzyme and hybridisation analysis by Harvey and D'Andrea, has now confirmed that the initially described λCH.02 (Harvey, *et al.*, 1981) is identical to the clone characterised in this chapter, as has direct sequencing data discussed in the following chapter.

One of the major features that is apparent from this work is that even though the histone amino acid sequence appears to have been conserved through evolution, and consequently sea urchin genes have proved useful in detecting histone genes in other species, *Drosophila* (Lifton *et al.*,

1978), *Xenopus* (Moorman *et al.*, 1980) and newt (Stephenson *et al.*, 1981), our findings suggest that these cross-species probes must be used with caution for higher vertebrates, since sea urchin H3 and H4 genes have failed to accurately detect the corresponding chicken genes.

There are a number of striking features about the organisation of chicken histone genes in λCH.02. Firstly, the genes are clustered; in fact there are six histone genes in λCH.02 within 11 kb of DNA. Clustering of histone genes has also been observed in other chicken genomal clones isolated. For example, λCH.01 has six histone genes within 14 kb (Harvey *et al.*, 1981; D'Andrea and Coles, unpublished data) and λCH3d has two histone genes within 4 kb of DNA (Engel and Dodgson, 1981). Therefore, clustering appears to be a conserved feature of histone gene organisation as it occurs in sea urchin (Kedes, 1979), *Drosophila* (Lifton *et al.*, 1979), *Xenopus* (Dongen *et al.*, 1981) and newt (Stephenson *et al.*, 1981).

Secondly, and perhaps the most unexpected feature of chicken histone gene organisation, is that the genes do not appear to be repeated in tandem arrays whereas histone genes organised in a repeating unit containing one each of the five genes is a common feature in sea urchin (Kedes, 1979), *Drosophila* (Lifton *et al.*, 1979), *Xenopus* (Dongen *et al.*, 1981) and Newt (Stephenson, *et al.*, 1981) systems. It is possible that the arrangement of genes in λCH.02 represents a long inverted repeat structure joining at the two H3 genes, however, the restriction map of λCH.02 shows no symmetry of sites which would be expected for such a repeating structure. Furthermore, other chicken histone gene clones characterised (Harvey *et al.*, 1981; Engel and Dodgson (1981); D'Andrea and Coles personal communication) display no apparent repeat arrangement of histone genes.

Large spacer regions between the histone genes also appear to be a common feature of the chicken histone genomal clones; for example, 9 kb of DNA separates the H4 gene from the H2A gene in λCH.01 (Harvey *et al.*, 1981) and an inter-gene distance of 3.5 kb is seen in λCH.02. Another interesting feature of the gene organisation in λCH.02 is the occurrence of two tightly linked H3/H4 gene pairs. H3/H4 gene pairs also occur in yeast (Hereford *et al.*, 1979), and this structure may play a role in controlling the expression of these genes.

The fact that chicken histone genes do not appear to be arranged in tandem repeating arrays adds little support to the suggestion by Melli *et al.*,(1977a) that histone genes are transcribed into a large polycistronic mRNA, thereby providing a means by which histone genes could be coordinately expressed. From the organisation of histone genes found in the chicken histone genomal clones, it appears more likely that each gene is independently controlled.

In order to locate potential control regions and verify the identity of the genes as well as determine the direction of transcription, a number of histone genes in λCH.02 were partially sequenced, as described in the next chapter.

CHAPTER 5

SEQUENCE ANALYSIS OF CHICKEN HISTONE GENES

5.1 INTRODUCTION

The aim of the work presented in this chapter was to obtain partial DNA sequence data from each region of λCH.02 which had previously been shown to hybridise with chicken histone cDNA as well as with individual chicken and sea urchin histone DNA probes (Chapter 4). Sequencing of these regions was important for a number of reasons.

Firstly, it would enable verification of the individual histone gene assignments which were deduced from the hybridisation results. This became essential, especially with respect to the H3 and H4 genes, where the sea urchin H3 and H4 DNA hybridisation results contradicted the hybridisation results obtained with the chicken H3 and H4 DNA probes. In addition, DNA sequence analysis of the restriction fragment in λCH.02 that hybridised weakly with chicken histone H2A DNA, would aid in identification of its coding potential.

Secondly, DNA sequence analysis of gene coding regions would aid in orienting the direction of transcription for each of the histone genes in λCH.02. Transcription of histone genes occurs from only one strand in the sea urchin (Kedes, 1979) and this has also been postulated for the human genes (Melli *et al.*, 1977a). Transcription from both strands however is observed in *Drosophila* (Lifton *et al.*, 1977), *Xenopus* (Moorman and Destree, 1980) and mouse (Sittman *et al.*, 1981). Orientation of the genes in λCH.02 will establish if transcription occurs from only one strand or from both as is suggested for this clone and λCH.01 from strand separation studies (Bruschi and Wells, 1981).

Thirdly, since the amino acid sequence for all the chicken histone proteins is not complete, DNA sequence analysis of the coding regions should enable accurate determination of the amino acid sequences coded for by the histone genes in λCH.02, as well as indicate if any of the genes in this clone code for histone variants.

Finally, in order to locate potential control regions surrounding the histone genes, it was necessary to obtain sequence data of the 5' and 3' untranslated regions. By comparison of the DNA sequence in these regions with other histone genes, it may be possible to identify specific DNA sequences which are conserved and therefore may play a role in control of expression.

In order to obtain DNA sequences, two different methodologies were employed. Initially, sequencing was done using the chemical cleavage procedures of Maxam and Gilbert (1977) on subfragments of the clones; for later sequencing, the chain termination method of Sanger (1977) was used on fragments subcloned into the bacteriophage M13. Before considering the sequence data obtained, the following section outlines the strategies using these techniques which are summarised in Fig. (5.1).

5.2 RESULTS AND DISCUSSION

5.2.1 DNA Sequencing by the Method of Maxam and Gilbert

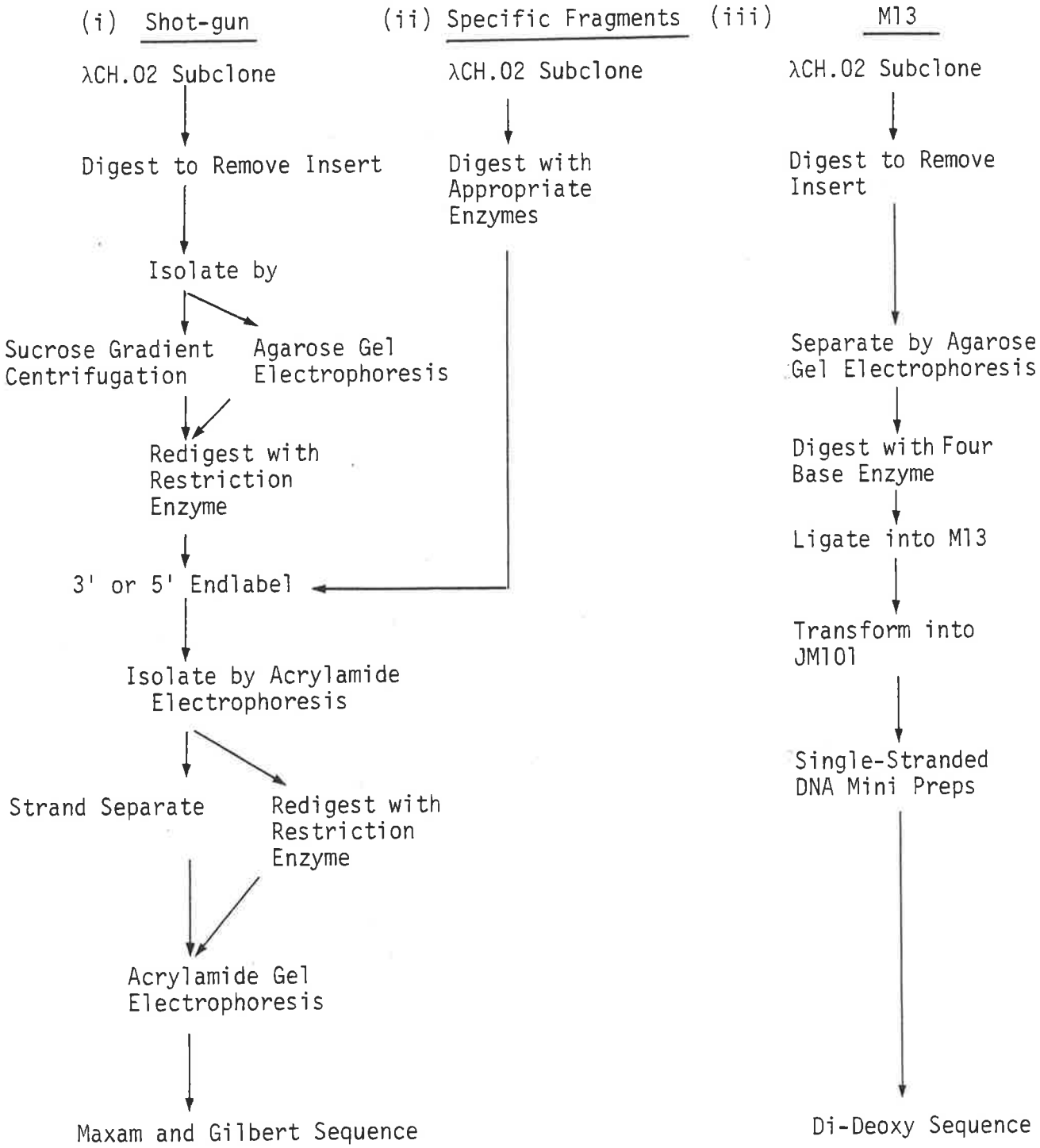
Sequencing by the chemical cleavage procedures described by Maxam and Gilbert (1977) necessitates DNA fragments which are labelled at only one end. For this purpose, different methods of both radioactive labelling of DNA and of separation of fragments with only one end labelled were used (described in Methods 2.2.14).

Fig. 5.1

Flow Diagram of Sequencing Strategies

Fig. 5.1

Sequencing Strategies



(a) End-Labelling of DNA

DNA fragments were radioactively labelled either by kinasing or end-filling.

(i) Kinasing involved initially de-phosphorylating the DNA fragment to remove the 5' phosphate, and then re-phosphorylating with polynucleotide kinase and γ-{^{32}P}-ATP. DNA fragments which were digested with restriction enzymes that made staggered cuts with 5'-extended termini (e.g. HindIII, TaqI and HpaII), were found to kinase more easily and therefore incorporated more radioactivity than DNA fragments which had blunt ends (e.g. PvuII, SmaI, HaeIII) or 3' ends protruding (e.g. SacII). However, kinasing efficiencies could be improved for these latter fragments by heat-denaturing prior to phosphorylation with polynucleotide kinase and γ-{^{32}P}-ATP (Chaconas and van de Sande, 1980).

ii) 3'-end-filling involved incubating DNA which had 5'-extended termini, with α-{^{32}P} deoxynucleoside triphosphates and the Klenow fragment of *E. coli* DNA polymerase I. DNA which was blunt-ended or had 3'-extended termini could also be end-filled by initially incubating with exonuclease III under conditions which removed approximately 10 nucleotides from the 3' end (as described in Methods 2.2.14e).

(b) Separation of End-labelled DNA Fragments

To isolate 5'-labelled DNA fragments for sequencing, two approaches were employed: "shotgun" isolation of random fragments and isolation of specific restriction fragments with hybridised to histone mRNA.

(i) Isolation of Random End-labelled DNA Fragments

Initially a "shotgun" approach was used to isolate 5'-labelled DNA from the λCH.02 subclones for sequencing. This approach consisted of digesting a λCH.02 subclone with an appropriate enzyme to excise the DNA insert and either purifying by sedimentation through a 10-40% sucrose density gradient or by electrophoresis on an agarose gel. The purified DNA insert was then digested with a four base restriction enzyme which preferably left a 5'-extended termini for radioactive labelling. These fragments were electrophoresed on an acrylamide gel ~~with~~ which was autoradiographed and each fragment eluted. The 5'-labelled DNA was then either strand separated or redigested with an appropriate restriction enzyme to give two DNA fragments each with only one labelled 5' phosphate. These fragments were then electrophoresed on an acrylamide gel, eluted, and subjected to the chemical degradation reactions of Maxam and Gilbert.

Although this procedure resulted in sequence data, it had a number of technical disadvantages. For example, isolating DNA initially from a 1% agarose gel (as no low melting point agarose was then available), resulted in poor yields and often the DNA would not redigest with restriction enzymes. Sucrose density gradient centrifugation gave better yields, however, more vector contamination occurred and similar restriction enzyme digestion problems were also encountered.

In order to overcome the problem of redigesting the DNA fragments after elution from agarose gels or from sucrose gradients, the λCH.02 subclone DNA was initially digested with a four base enzyme,

then end-filled (since end-filling was more efficient when a large number of fragments were present) and electrophoresed on an acrylamide gel with digested vector DNA as markers (Fig. 5.2). All the insert DNA fragments were eluted from the acrylamide gel and strand separated or recut with another restriction enzyme, then the fragments with only one end labelled, isolated by acrylamide gel electrophoresis for sequencing by the Maxam and Gilbert procedure (e.g. Fig. 5.3).

This procedure obviated the initial insert purification step; however, restricted insert DNA which had the same molecular weight as the restricted vector was difficult to separate and very small digested DNA fragments of less than 80 bp could not be resolved.

Rather than sequence all possible fragments, a selection of histone coding fragments from a total digest of subclones was sometimes made (Fig. 5.2) by probing a Southern blot of the digested DNA with chicken histone cDNA or a specific histone coding probe. This procedure ensured the sequencing of specific coding fragments.

(ii) Isolation of Specific 5'-end labelled DNA Fragments

This strategy relied on the restriction and hybridisation data of the λCH.02 subclones. The λCH.02 subclones were digested with a combination of restriction enzymes so as to produce conveniently sized DNA fragments which were known to hybridise with chicken histone DNA. These fragments were then isolated by agarose or acrylamide gel electrophoresis. For example, Fig. (5.4) shows a SacII/HindIII preparative restriction enzyme digest of pCH2.6H, in which the 570 bp SacII/HindIII fragment was eluted, 5'-end-labelled by kinasing and

Fig. 5.2

Preparation of Restriction Fragments for Sequencing

An example of end-filling and separation of HpaII digested pCH2.6H on a 6% polyacrylamide gel, for sequencing both random and coding specific fragments.
The tracks are as follows:

A: End-labelled HpaII digested pCH2.6H

B: End-labelled HpaII digested pBR322

C: HpaII digested pBR322

D: HpaII digested pCH2.6H

E: Autoradiograph of Southern blot of HpaII pCH2.6H hybridised with chicken histone H4 probe.

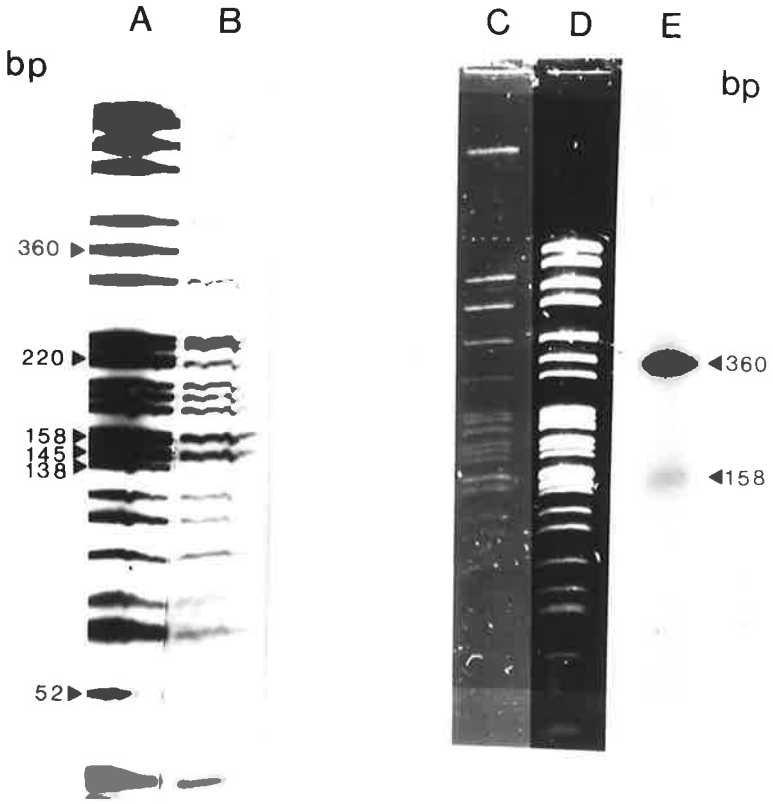
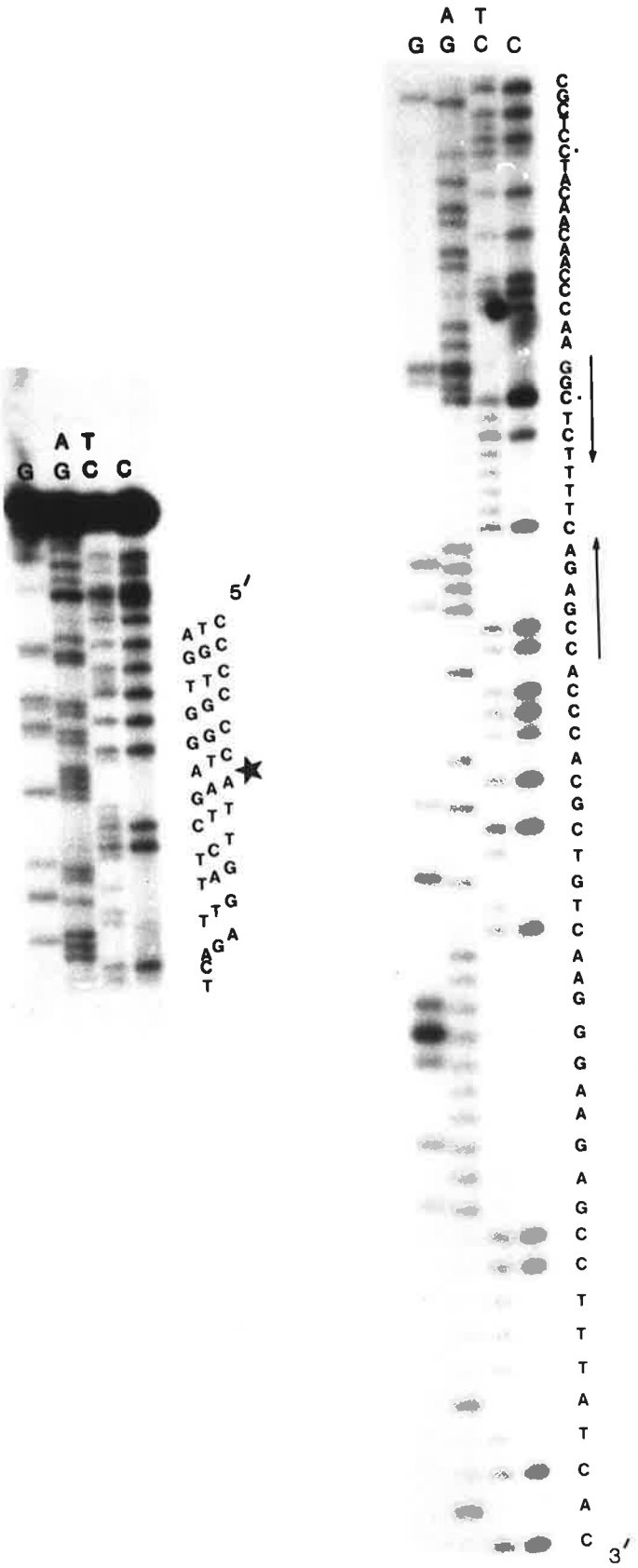

A
B
C
D
E
bp
360
220
158
145
138
52
bp
360
158

Fig. 5.3

DNA Sequencing Gel of 3' End of H4 Gene

DNA sequence determination of the 360 bp HpaII fragment (Fig. 5.2) which was 3'-labelled and re-cut with BstNI to yield a 320 bp fragment. It was sequenced using the base-specific cleavage reactions of Maxam and Gilbert (1977) and electrophoresed on an 8% acrylamide DNA-sequencing gel. The DNA sequence corresponds to the 3' untranslated region of the H4 gene, as indicated in Fig. (5-7c), and reads 3' to 5' up the gel.

The asterisk indicates the termination codon and the arrows show the region of dyad symmetry.


G A T C
5′
3′

then recut with PvuII (as predicted by the restriction map, Fig. 4.14). The DNA fragments labelled at only one end, after PvuII digestion, were isolated on an acrylamide gel and prepared for sequencing (e.g. Fig. 5.5).

Isolation of DNA fragments by this strategy was advantageous because firstly, DNA was isolated from gels in high yield and routinely recut with restriction enzymes and secondly, this strategy ensured that the DNA fragment sequenced would contain coding sequences. The disadvantage of this strategy was that most of the appropriate restriction enzymes sites which were identified as surrounding coding fragments produced blunt or 3'-extended termini, and these fragments did not end-label very efficiently either by kinasing or end-filling. (Fig. (5.4) demonstrates the relative incorporation of radioactivity into a SacII end versus a HindIII end.)

In addition, some of the large coding fragments did not contain any appropriate restriction enzyme sites and therefore, due to the limited distance that can be sequenced from one end (200-300 bp), only partial sequences of some genes could be obtained by this technique.

5.2.2. DNA Sequencing by the Dideoxy-method

The strategies described so far to prepare labelled DNA for sequencing all involve an initial input of large amounts of DNA, followed by many steps where losses occur and thus the sequencing output is reduced. However, with the ability to clone DNA into M13, a much simplified approach for preparing and sequencing DNA fragments was provided.

Fig. 5.4

Preparation of an H3-Containing Fragment for Sequencing

pCH2.6H was digested with SacII/HindIII and electrophoresed on a 0.7% low melting point agarose gel, at 50 mA for 2 h, eluted as described in Methods (2.2.12b), 5'end-labelled and prepared for sequencing.

The tracks are as follows:

A: SacII/HindIII digested pCH2.6H

B: The 570 bp SacII/HindIII fragment, 5'end-labelled and electrophoresed on a 6% polyacrylamide gel.

C: PvuII digestion of the 570 bp SacII/HindIII fragment

(Note: The 260 bp HindIII/PvuII fragment is labelled to a greater extent than the 320 bp SacII/PvuII fragment.)

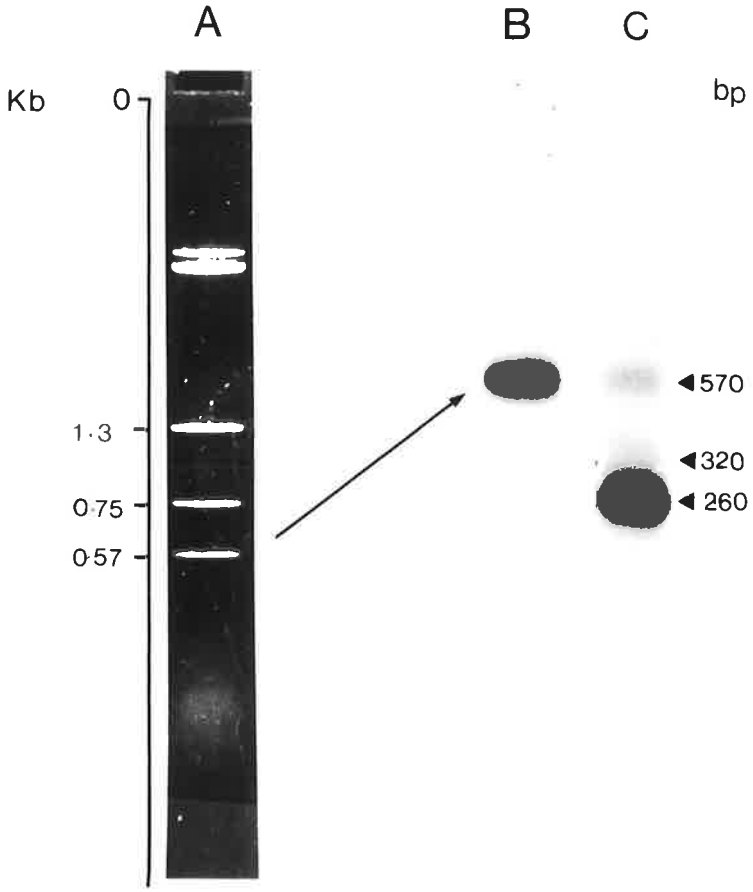
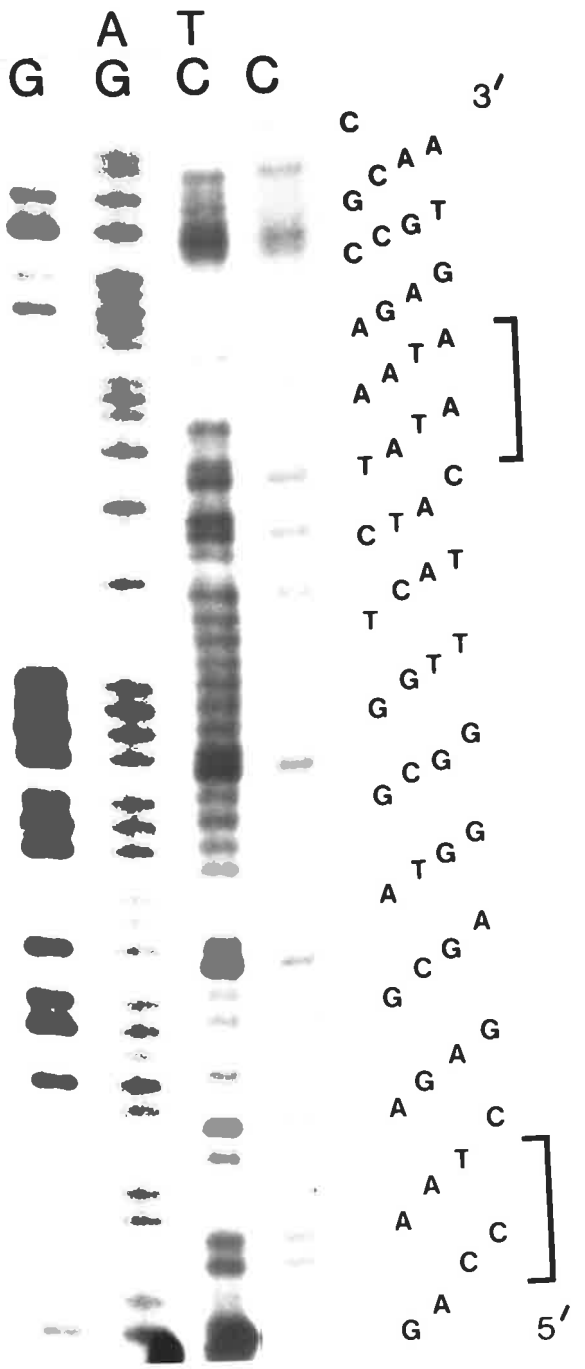

A
B
C
Kb
0
bp
1·3
0·75
0·57
570
320
260

Fig. 5.5

Maxam and Gilbert Sequencing Gel of 260 bp HindIII/PvuII Fragment of pCH2.6H

The 260 bp 5' end-labelled fragment (Fig. 5.4) was subjected to sequencing reactions and electrophoresed on an 8% polyacrylamide DNA sequencing gel. The DNA sequence corresponds to the 5' untranslated region of the H3 gene, as indicated in Fig. (5.7a) and reads 5' to 3' up the gel.

The brackets correspond to the "CAT" box and the"TATA" box.


G
A G
T C
C
3′
C
G C A A
G C C G T
A G A G
A A T A
T A T A
C
C T A
T C A T
G G T T
G C G G
A T G G
G C G A
A G A G
C
A T
A C
C
G A
5′

The strategy employed to sequence the histone genes by the dideoxy sequencing method (Sanger, 1977), involved isolating the large coding restriction fragments of the λCH.02 subclones by low melting point agarose gel electrophoresis and re-digesting with a four base enzyme, HpaII. The resulting fragments were then isolated by acrylamide gel electrophoresis and ligated into the AccI site of the phage M13 mp7. Each ligation mix was transformed into JM101 and single-stranded DNA was prepared from the clear plaques for sequencing. By sequencing M13 DNA subclones of each orientation up to 300 bases per fragment could be read.

5.2.3 Sequence Analysis of the Genes in λCH.02

The DNA regions sequenced in the λCH.02 subclones and the sequencing strategies which were employed are summarised in Fig. (5.6 a,b,c). Analysis of these sequences can be divided into a number of sections, which include verification of histone gene assignments, determination of direction of transcription, amino acid conservation, codon usage and analysis of regulatory sequences.

5.2.4 Histone Gene Assignments

Each DNA restriction fragment which hybridised with chicken histone cDNA (Fig. 4.26) was partially sequenced and each was found to code for a single histone gene, the sequences of which are summarised in Fig. (5.7 a,b,c,d,e).

The H3, H4 and H2B genes were located in DNA fragments predicted by the hybridisation results obtained with H3, H4 and H2B chicken histone DNA probes, but not with the sea urchin probes (Fig. 4.25). It can therefore be concluded that the hybridisation

Fig. 5.6

DNA Fragments Sequenced in λCH.02

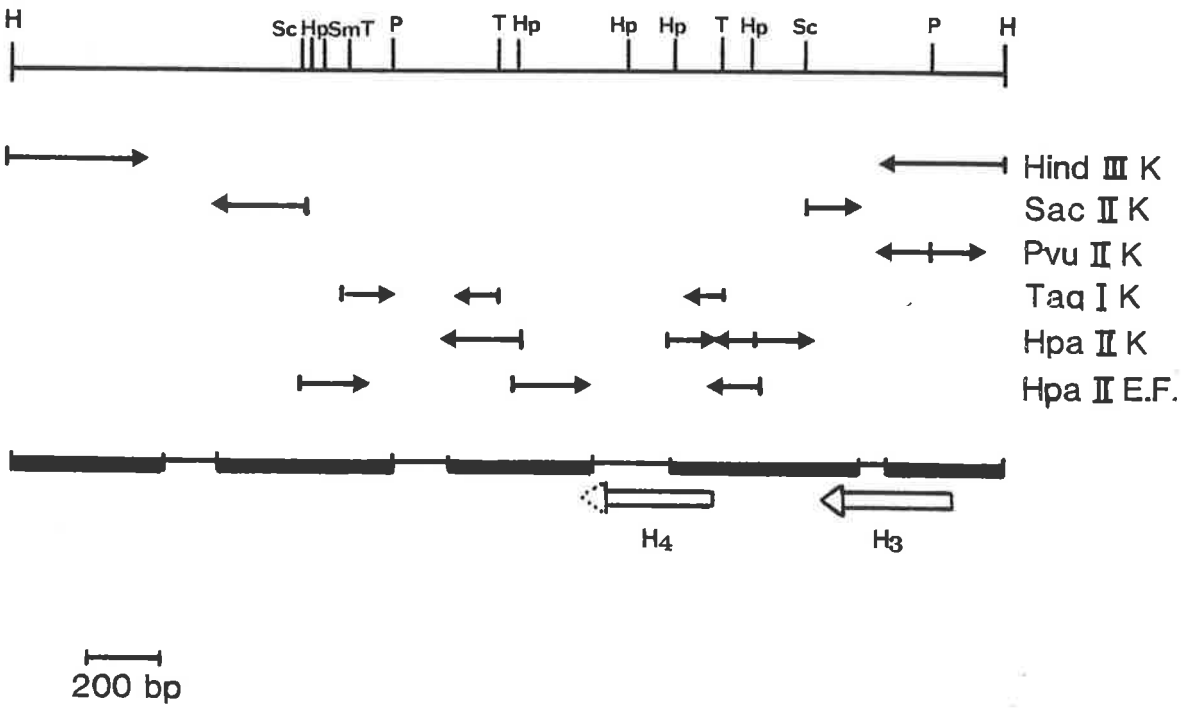
The various restriction fragments sequenced in the three λCH.02 subclones are shown. The arrowed lines represent the direction and extent of sequencing. The restriction sites and means of end-labelling (K = kinasing, E.F. = end-filling) are shown for sequences obtained by the method of Maxam and Gilbert 1977) and fragments sequencing in M13 by the dideoxy method indicated.
The blocked regions summarise the extent of DNA sequenced and the location of the histone genes are indicated by the open boxes.

a. pCH2.6H

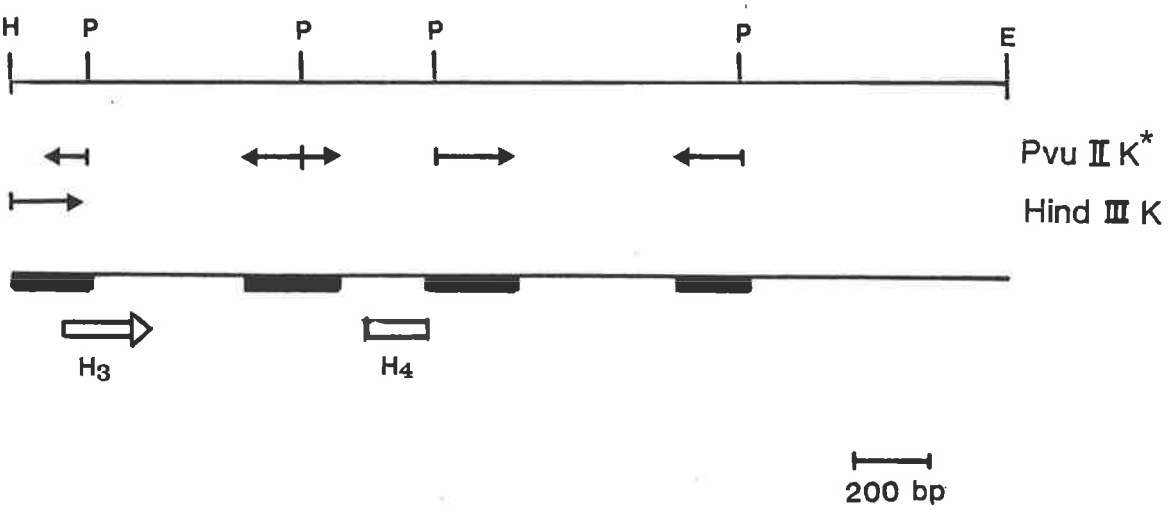
b. pCH4.3EH

c. pCH4.8E

(a)


H
Sc Hp Sm T P
T Hp
Hp Hp
T Hp
Sc
P
H
Hind Ⅲ K
Sac Ⅱ K
Pvu Ⅱ K
Taq Ⅰ K
Hpa Ⅱ K
Hpa Ⅱ E.F.
H4
H3
200 bp

(b)



(c)

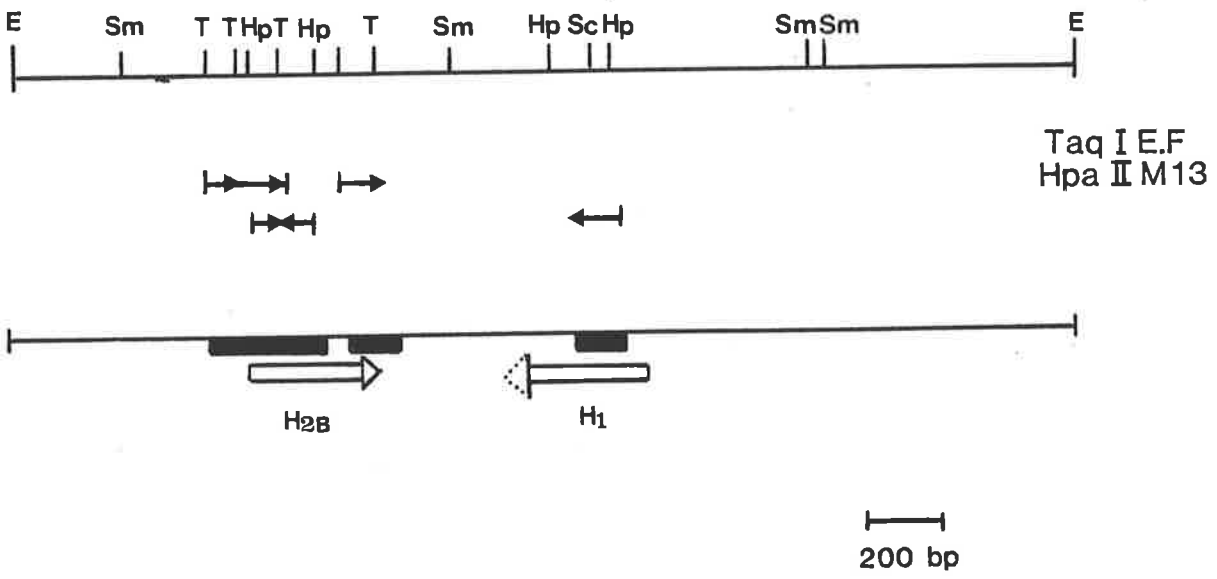


Fig. 5.7

Histone Gene Sequences from λCH.02

The histone gene sequences obtained and the corresponding amino acid sequences are shown. Conserved sequences, the "CAT" boxes, "TATA" boxes and regions of dyad symmetry are underlined.

a) H3 Gene, pCH2.6H

The combined DNA sequence determined from the 570 bp HindIII/SacII fragment in pCH2.6H as summarised in Fig. (5.6a) is shown. It contains 165 bp of 5' non-coding, the first 123 bp and the last 12 nucleotides of the H3 coding region and 83 bp of 3' non-coding sequence.

b) H3 Gene, pCH4.3EH

The DNA sequence was determined from the HindIII/PvuII fragment in pCH4.3EH, as summarised in Fig. (5.6b). It contains 41 bp of 5' non-coding and the first 30 nucleotides of the H3 coding region.

c) H4 Gene

The DNA sequence determined in pCH2.6H, as shown in Fig. (5.6a), contains the last 15 nucleotides of the H4 coding region and 127 bp of 3' non-coding sequence.

d) H2B Gene

Shown as the combined DNA sequence determined from the HpaII and TaqI fragments of pCH4.8E (as outlined in Fig. (5.6c)), containing 87 bp of 5' non-coding, the first 218 and last 66 nucleotides of the H2B coding region and 70 bp of the 3' non-coding region is indicated in upper case. The complete H2B sequence was determined independently by A. Robins and R. Harvey and the remaining sequences are shown in lower case. Asterisks indicate nucleotide changes when compared with the chicken H2B gene coded in λCH.01 (R. Harvey, personal communication).

e) H1 Gene

The DNA sequence determined from the HpaII fragment in pCH4.8E (Fig. 5.6c), containing 180 bases of H1 coding region indicated in upper case. A further 129 nucleotides of the H1 coding region sequenced by A. Robins are shown in lower case.

(a) pCH2.6H Partial H3 Gene Sequence

```
5' AAGCTTCTTTGCAACCTGGGACAGGACAGGCAGAAGGC

   TTGGAGTTAGCCGGTTAATTCATTGTTTTGTTGACCAA

   TCAGAGGCGAATGGGCGGGGTTTCATCTACTATAAATA

   AGAGCCGTGCAACGAGACCGTCGACTTTCGGTTGCAGA

                    1                   5
   GAGTTTTGGAATG GCG CGT ACG AAG CAG ACG
                 Ala Arg Thr Lys Gln Thr

                10                      15
   GCG CGT AAG TCG ACG GGC GGC AAG GCG
   Ala Arg Lys Ser Thr Gly Gly Lys Ala

   CCC CGC AAG CAG CTG GCC ACC AAG GCG
   Pro Arg Lys Gln Leu Ala Thr Lys Ala

    25                     30
   GCC CGG AAG AGC GCG CCC GCC ACG GGG
   Ala Arg Lys Ser Ala Pro Ala Thr Gly

        35                     40
   GGG GTG AAG AAG CCG CAC CGC TAC 3'
   Gly Val Lys Lys Pro His Arg Tyr


                   135
5' GGC GAG CGC GCG TGACGTTAGACTTCGCTCTCTT
   Gly Glu Arg Ala

   ACACACAACAACCCAAGCTCTTTTCAGAGCCACCCACG

   TGTCGAGGGAAGAGCCTTTATCA 3'
```

(b) Partial H3 Gene Sequence pCH4.3EH

```
5' ACGGGCCATTTGTGTAGTTGTAAGGGAAA

                  1                   5
                 Ala Arg Thr Lys Gln
GGCTGTGTG ATG GCG CGT ACG AAG CAG

                     10
Thr Ala Arg Lys Ser
ACG GCG CGT AAG TCG
```

(c) pCH2.6H Partial H4 Gene Sequence

Tyr Gly Phe Gly Gly
5' TAC GGC TTC GGC GGC TAAAGTTCTCTGATGTTAGACTCGCTCCT

ACAACAACCCAAGGCTCTTTTCAGAGCCACCCACGCTGTCAAGGGAAGA

GCCTTTATCACTGTTGATTATGCAATAACAATCTTTCACTTAATTGCCG 3'

(d) H2B Gene Sequence

5' ctgttatccaatcagagaccagatacagaaggcactcgaTTTGCATACTGCCCC

TATAAATAGGCGAGCAGTGCTCGCAGCCGGCACTCCGCTGCGCCGAAGGGATCGTGGAGAGTTCGAC

```
     1*           *  5        *         10*                *  15
ATG CCT GAG CCG GCC AAG TCC GCA CCC GCC CCC AAG AAG GGC TCC AAG AAG
      C            T           G           G                T
    pro glu pro ala lys ser ala pro ala pro lys lys gly ser lys lys

              20                  25                  30
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

     35                  40                  45                  50
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

                 55                  60                  65
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

          70                  75                  80
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

 85                  90                  95   *              100
aag cgc tcg acc atc acg tcg cgg cag atc cag axg gcc gtg cgg ctg ctg
lys arg ser thr ile thr ser arg glu ile gln thr ala val arg leu leu
                                            a

              105                 110       *         115
ctg ccc GGC GAG CTG GCC AAG CAC GCG CTC TCC GAG GGC ACC AAG GCG GTC
leu pro gly glu leu ala lys his ala val ser glu gly thr lys ala val
                                          G

     120                 125
ACC AAG TAC ACC AGC TCC AAG TAG AGCggtgcggattacTCGATTTTAACCCAAAGGCT
thr lys tyr thr ser ser lys
```

CTTTTCAGAGCCACCATTTGTTCTAATAAAAGGGCTGTATTACTTTTTTTC 3'

(e) Partial H1 Gene Sequence

```
   103     105                 110                 115
5' GCC GGC GCC TCG GGC TCT TTC AAG CTA ATA AAA AAG GCG CTG GAG AAA
   Ala Gly Ala Ser Gly Ser Phe Lys Leu Ile Lys Lys Ala Val Glu Lys

    120                 125                 130
   AAA GAG AAA AGG ACT AAG AAG AAG CCC GCG GCG AAG CCC AAG AAG GCG
   Lys Lys Lys Arg Thr Lys Lys Lys Pro Ala Ala Lys Pro Lys Lys Ala

   135                 140                 145                 150
   GCG AAG AAG CCT GCG GCT GCT GCC AAG AAG CCC AAG AAG GCA GCA GCG
   Ala Lys Lys Pro Ala Ala Ala Ala Lys Lys Pro Lys Lys Ala Ala Ala

                   155                 160                 165
   GTG AAG AAG AGC CCC AAG AAA GCC AAG AAG CCG GCg agc gct gcc acc
   Val Lys Lys Ser Pro Lys Lys Ala Lys Lys Pro Ala Ser Ala Ala Thr

             170                 175                 180
   aag aag gcg cgg aag agc ccc aag aag gct acc aag gcg gtg aag cca
   Lys Lys Ala Arg Lys Ser Pro Lys Lys Ala Thr Lys Ala Val Lys Pro

         185                 190                 195
   aag gct gcc aag tca aag gcg gcc aaa tcc aag gcg gcc aag gca aag
   Lys Ala Ala Lys Ser Lys Ala Ala Lys Ser Lys Ala Ala Lys Ala Lys

    200                 205
   aag gca gcg acc aaa aag aag taa
   Lys Ala Ala Thr Lys Lys Lys ——
```

results obtained with the sea urchin H3 and H4 DNA probes were indeed misleading, as discussed in Chapter 4.

The DNA restriction fragment in pCH4.8E (Fig. 4.24) which hybridised weakly with chicken H2A was found by sequencing (initially by A. Robins) to code for an H1 histone protein. The location of the H1 gene in pCH4.8E was also independently verified by hybrid-arrest translation studies using pCH4.8E DNA (G. Stein, personal communication). They demonstrated that pCH4.8E specifically hybridised with only H2B and H1 RNA.

DNA sequence data therefore proved to be vital in positive identification of the histone genes in λCH.02. The most important finding however which came from the DNA sequence analysis was the identification of this clone as the previously isolated chicken histone λCH.02. When the H2B gene sequence in pCH4.8E (Fig. 5.7d) was compared to the H2B gene sequenced in λCH.02 by A. Robins and R. Harvey, the coding, non-coding and spacer regions were found to be identical. This led to the re-mapping and hybridisation of the initially characterised λCH.02 clone (Harvey *et al.*, 1981) and to the final conclusion that the clones were indeed identical.

5.2.5 Direction of Transcription

The orientation of some histone genes in λCH.02 was deduced from the restriction enzyme mapping and sequence analysis (Figs. 5.6, 5.7). The direction of transcription was positively determined for each of the two H3 genes in λCH.02; the H3 gene located in pCH2.6H was oriented by sequencing from a HindIII site through a SalI and PvuII site predicted from the restriction map (Fig. 5.6a); the H3 gene in pCH4.3EH was oriented similarly by sequencing from a HindIII

site through a SalI and PvuII site (Fig. 5.6b). The H2B gene was oriented by comparison of the restriction sites found in the complete H2B gene sequence (A. Robins and R. Harvey) with the restriction map of pCH4.8E (Fig. 5.6c). The H4 gene in pCH2.6H was tentatively oriented by the restriction map location of the HpaII fragment from which the gene sequence was obtained (Fig. 5.6a). Similarly, H1 was tentatively oriented with respect to the SacII site in the gene sequence from pCH4.8E, (Fig. 5.6c).

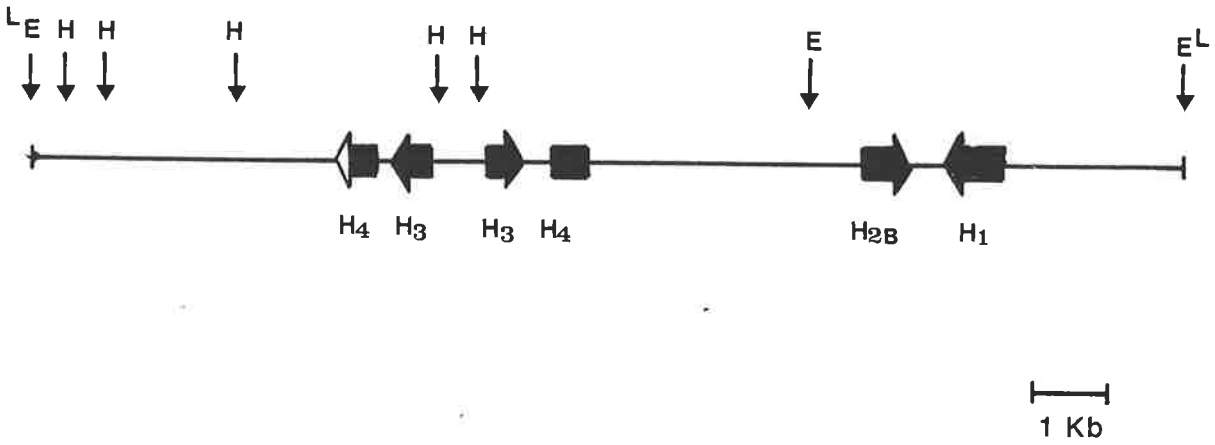
A summary of the direction of transcription of the histone genes coded for in λCH.02 is shown in Fig. (5.8). A most remarkable feature of the orientation of these genes is the presence of adjacent, divergently transcribed H3 genes separated by about 900 bp of DNA. Although the H2A/H2B and H3/H4 gene pairs of yeast (Hereford *et al.*, 1979) and *Drosophila* (Lifton *et al.*, 1977) are divergently transcribed, this kind of organisation for a histone gene pair coding for the same protein has not been described previously. A similarly unusual arrangement is seen for the silk moth chorion genes (Jones and Kafatos, 1980). Further investigation will be required to determine if such an arrangement has biological significance.

The divergent transcription of the two H3 genes as well as the probable opposite orientation of the H2B and H1 genes indicates that transcription must occur from both DNA strands in λCH.02 (as predicted by hybridisation with strand-separated DNA (Bruschi and Wells, 1981)). This is also the case in the chicken histone clone λCH.01 (Bruschi and Wells, 1981), as well as in the *Drosophila* (Lifton *et al.*, 1977), *Xenopus* (Zernick *et al.*, 1980) and mouse (Sittman *et al*, 1981) histone

Fig. 5.8

Histone Gene Content in λCH.02

Summary restriction map of λCH.02 showing location and orientation (arrows point 5' to 3') of the histone genes.


LE H H
H
H H
E
EL
H4 H3 H3 H4
H2B H1
1 Kb

clones that have been studied. Thus histone gene transcription from only one DNA strand as typified in sea urchin histone gene organisation is not a conserved feature of the histone genes.

Since it also appears that the overall organisation of the histone genes has not been conserved in higher eukaryotes in the simple tandem repeat arrangement seen in sea urchin, it was of interest to determine what features of histone gene structure had been conserved at the DNA sequence level.

5.2.6 Protein Sequences of Histone Genes in λCH.02

The primary sequence of the core histone proteins are generally highly conserved (Isenberg, 1979) directly reflecting the fundamental character of the functions performed by the major histones within the cell. However, histone variants have been identified in sea urchin (Cohen *et al.*, 1975) as both developmental and tissue-specific histones. At this stage there is still little data available to establish the general occurrence of tissue specific and developmental variants in higher eukaryotes.

In order to study the nature of the chicken histone proteins encoded in λCH.02, the amino acid sequences predicted from the H3, H2B and H1 gene regions were analysed and compared to the protein sequences of chicken and other species. As only the C-terminal five amino acids were deduced for the H4 gene it is not considered below.

(a) H3 Histone Protein Sequence

From the nucleotide sequence data shown in Fig.(5.7 a,b) the N-terminal 41 amino acids and the last 4 amino acids of the C-terminus of the H3 gene in pCH2.6H, and the N-terminal 10 amino acids of H3 from pCH4.3EH have been deduced. Within the common regions sequenced, the amino acids of both are identical as well as identical to the chicken H3 protein sequence encoded in λCH.01 (P. Krieg, personal communication) and the major H3 amino acid sequence determined for chicken (Brandt, W.F. and Von Holt, C., 1974).

In addition, the complete protein homology observed in the chicken H3 genes also extends to comparisons with sequenced H3 proteins of shark, carp, sea urchin, limpet, mouse and calf (Von Holt *et al.*, 1979). There is a single amino acid change at position 41 observed in pea and cycad H3 protein sequence, where a tyrosine residue is replaced by a phenylalanine.

The remarkable amino acid conservation observed in the H3 protein sequence through evolution however, is not reflected in codon third base conservation. Numerous silent base changes are observed. Fig. (5.9) shows that 5 third base differences exist between the 41 codons determined for the chicken H3 genes in λCH.02 and λCH.01, whereas 27 nucleotides out of the 45 codons are different between the chicken H3 gene in λCH.02 and the sea urchin H3 gene in h19 (Busslinger *et al.*, 1980).

It therefore appears probably that silent base changes reflect evolutionary distance, while lack of replacement substitutions demonstrates the strong selective pressure on the protein sequence. (Grunstein et al, 1976).

Fig. 5.9

Comparative Coding Regions of the Chicken H3 Genes

The first 41 and the last 4 amino acids of the H3 gene is pCH2.6H and the first 10 amino acids of the H3 gene in pCH4.3EH are compared to the overlapping coding regions in sea urchin *P. miliaris* H3 gene in h19 and h22 (Busslinger *et al.*, 1979). The asterisks above sequences indicate nucleotide changes with respect to the H3 gene in pCH2.6H, and the asterisk over amino acid 41 indicates an amino acid change with respect to the pea and cyad protein sequence (Von Holt *et al.*, 1979).

Figure 5.9

Coding H3 Comparisons

```
                                    5                   10                  15
                    Ala Arg Thr Lys Gln Thr Ala Arg Lys Ser Thr Gly Gly Lys Ala
H3 pCH2.6H      AUG GCG CGT ACG AAG CAG ACG GCG CGT AAG TCG ACG GGC GGC AAG GCG
                                                                  *
H3 λCH.01           GCG CGT ACG AAG CAG ACG GCG CGT AAG TCG ACG GGC GGG AAG GCG

H3 pCH4.3EH         GCG CGT ACG AAG CAG ACG GCG CGT AAG TCG
                      *   *   *           *   *   *   *   *   *   *   *       *
H3 h19              GCA CGC ACC AAG CAG ACC GCT CGC AAA TCT ACA GGA GGG AAG GCT
                      *   *   *               *   *   *   *   *   *   *   *   *
H3 h22              GCA CGA ACC AAG CAG ACG GCT CGC AAA TCT ACA GGA GGT AAA GCC


                                    20                  25                  30
                    Pro Arg Lys Gln Leu Ala Thr Lys Ala Ala Arg Lys Ser Ala Pro
   pCH2.6H          CCC CGC AAG CAG CTG GCC ACC AAG GCG GCC CGG AAG AGC GCG CCC
                                                          *               *
   λCH.01           CCC CGC AAG CAG CTG GCC ACC AAG GCG GCC CGC AAG AGC GCG CCG
                                          *       *   *   *       *   *
   h19              CCC CGC AAG CAG CTG GCA ACC AAA GCA GCC AGA AAG AGT GCC CCC
                                          *           *   *       *   *   *
   h22              CCC CGC AAG CAG CTG GCA ACC AAG GCT GCC AGA AAG AGT GCC CCA


                                    35                  40  *
                    Ala Thr Gly Gly Val Lys Lys Pro His Arg Tyr
   pCH2.6H          GCC ACG GGG GGG GTG AAG AAG CCG CAC CGC TAC
                              *   *
   λCH.01           GCC ACG GGC GGC GTG AAG AAG CCG CAC CGC TAC
                          *   *   *   *           *   *   *
   h19              GCC ACT GGA GGA GTC AAG AAG CCT CAT CGA TAC
                          *   *   *   *           *   *
   h22              GCC ACC GGA GGA GTA AAG AAG CCT CAT CGC TAC
                                                        Phe (pea, cycad)

                    132 133 134 135
                    Gly Glu Arg Arg
   pCH2.6H          GGC GAG CGC GCG TGA
                      *   *       *
   h19              GGA GAA CGC GCC
                      *   *   *   *
   h22              GGA GAA CGT GCC
```

(b) H2B Protein Sequence

The chicken H2B histone protein sequence coded for in λCH.02 was predicted from the DNA sequence analysis as outlined in Fig. 5.7d). The protein sequence is identical to that determined for chicken H2B amino acid sequence (Van Helden *et al.*, 1978), and is also identical to the chicken H2B protein sequence coded for in λCH.01 (R. Harvey).

When the nucleotide coding sequence is compared between the two H2B chicken genes (Fig. 5.7d) only 7 nucleotide changes occur out of 375, all of which are silent third base changes. These two chicken H2B genes therefore appear to be very similar at a nucleotide level which suggests that the coding portions of the genes have only recently diverged. Using the calculations described by Perler *et al.*, (1980), the percentage corrected divergence between the two chicken H2B gene sequences is 0% for replacement and 8.3% for silent sites. This may be compared with the mouse β major and β minor genes which diverged less than 30 million years ago (Perler *et al.*, 1980) and have a corrected divergence in replacement sites of 3.1% and in silent sites of 10.6%. Comparison of silent base changes therefore indicates that the two chicken H2B genes in λCH.02 and λCH.01 may have also diverged less than 30 million years ago.

In order to determine the extent of amino acid conservation of the H2B protein, the λCH.02 amino acid sequence was compared to sea urchin and calf H2B protein sequences (Von Holt *et al.*, 1979) (Fig. 5.10). Six amino acid changes, all conservative, were found between calf and chicken H2B amino acid sequences at positions (21) *Thr* → *Ala*, (25) *Gly* → *Asp*, (26) *Asp* → *Gly* (25 and 26 are reciprocal changes), (31) *Lys* → *Arg*, (39) *Ile* → *Val*, (79) *Arg* → *Ala*.

Fig. 5.10

Comparison of the Chicken H2B Protein Sequences with Calf and Sea Urchin H2B Sequences

The amino acid sequence for the H2B gene in λCH.02 is compared to the amino acid sequence of the H2B gene in Sea Urchin h19, (Busslinger *et al.*, 1979) and to the H2B protein sequence from calf (Von Holt *et al.*, 1979). Asterisks indicate amino acid changes and dashes indicate deletions with respect to the sea urchin H2B protein; any changes from the calf H2B protein are indicated at the appropriate amino acid.

H2B Protein Comparisons

	1																			20
H2B pCH4.8E	PRO	GLU	PRO	ALA	LYS	SER	ALA	PRO	ALA	PRO	LYS	LYS	GLY	SER	LYS	LYS	ALA	VAL	THR*	LYS*
H2B h22	PRO	ALA	—	—	—	—	—	—	—	—	LYS	LYS	GLY	SER	LYS	LYS	ALA	VAL	LYS	ALA

	21																			40
H2B pCH4.8E	THR*	GLN*	LYS*	LYS*	GLY	ASP*	LYS	LYS	ARG	LYS*	LYS	SER	ARG	LYS	GLU	SER	TYR	GLY*	ILE	TYR
H2B h22	PRO	ARG	PRO	SER	GLY	GLY	LYS	LYS	ARG	ASN	—	—	ARG	LYS	GLU	SER	TYR	SER	ILE	TYR
Calf	ALA				ASP	GLY					ARG								VAL	

	41																			60
H2B pCH4.8E	VAL*	TYR	LYS	VAL	LEU	LYS	GLN	VAL	HIS	PRO	ASP	THR	GLY	ILE	SER	SER	LYS*	ALA	MET	GLY*
H2B h22	ILE	TYR	LYS	VAL	LEU	LYS	GLN	VAL	HIS	PRO	ASP	THR	GLY	ILE	SER	SER	ARG	ALA	MET	ILE

	61																			80
H2B pCH4.8E	ILE	MET	ASN	SER	PHE	VAL	ASN	ASP	ILE	PHE	GLU	ARG	ILE	ALA	GLY	GLU	ALA*	SER	ARG	LEU
H2B h22	ILE	MET	ASN	SER	PHE	VAL	ASN	ASP	ILE	PHE	GLU	ARG	ILE	ALA	GLY	—	GLU	SER	ARG	LEU
Calf																			ALA	

	81																			100
H2B pCH4.8E	ALA	HIS*	TYR	ASN	LYS	ARG*	SER	THR	ILE	THR*	SER	ARG	GLU	ILE	GLN	THR	ALA	VAL	ARG	LEU
H2B h22	ALA	GLN	TYR	ASN	LYS	LYS	SER	THR	ILE	SER	SER	ARG	GLU	ILE	GLN	THR	ALA	VAL	ARG	LEU

	101																			120
H2B pCH4.8E	LEU*	LEU	PRO	GLY	GLU	LEU	ALA	LYS	HIS	ALA	VAL	SER	GLU	GLY	THR	LYS	ALA	VAL	THR	LYS
H2B h22	ILE	LEU	PRO	GLY	GLU	LEU	ALA	LYS	HIS	ALA	VAL	SER	GLU	GLY	THR	LYS	ALA	VAL	THR	LYS

	121				125
H2B pCH4.8E	TYR	THR	SER*	SER	LYS
H2B h22	TYR	THR	THR	SER	LYS

However, more extensive amino acid differences were observed when comparing sea urchin and chicken H2B protein sequences. Where the carboxy-terminal of the H2B protein appears to have been fairly well conserved, the amino-terminal region has been extensively modified. The modifications are due primarily to insertions and deletions in the first 33 amino acids. Therefore, the H2B histone protein, unlike H3, has not been totally conserved but still retains relative constancy in the C-terminal hydrophobic part of the molecule which is thought to be necessary in histone-histone interactions (Von Holt *et al.*, 1979).

(c) H1 Histone Protein Sequence

H1 histone protein is renowned for being the most variable of the histones, but like H2B has highly conserved stretches which have globular protein-like residue distributions. Sequence analysis of the H1 gene in λCH.02 has allowed for the determination of the last 103 amino acids of this protein as shown in Fig. (5.7e). No amino acid sequence data is available for chicken H1 protein in this region to allow for comparison of this H1 protein with other chicken H1 histones. However, the amino acid sequence predicted for the H1 gene coded by λCH.02 was compared with the same H1 protein region sequenced in rabbit and trout. It can be seen from Fig.(5.11) that regions of homology occur for stretches of 6-13 amino acids but these are studded by small deletions amid insertions. These extensive modifications in regions between the conserved globular domains are responsible for the higher interspecies divergence of H1 proteins.

Fig. 5.11

Chicken H1 Protein Comparisons

The last 124 amino acids from the H1 gene in λCH.02 are compared to the corresponding protein regions in trout and rabbit H1 proteins (Von Holt *et al.*, 1979). Asterisks indicate amino acid differences and dashes to amino acid deletions required for alignment. The amino acids are numbered with respect to the chicken H1 protein sequence.

H1 Protein Comparisons

103–121	
H1 pCH4.8E	ALA GLY *ALA SER GLY SER PHE LYS LEU *ILE LYS LYS ALA — *VAL *GLU *LYS *LYS *GLU *LYS
H1 Rabbit	ALA GLY THR SER GLY SER PHE LYS LEU ASN LYS LYS ALA ALA SER GLY GLU ALA LYS PRO
H1 Trout	ALA GLY THR SER GLY SER PHE LYS LEU ASN LYS LYS ALA — VAL GLU ALA LYS —

122–140	
H1 pCH4.8E	*ARG *SER *LYS LYS LYS *PRO *ALA ALA *LYS *PRO *LYS LYS *ALA *ALA — *LYS LYS *PRO *ALA ALA
H1 Rabbit	LYS PRO — LYS LYS ALA GLY ALA ALA LYS PRO LYS LYS PRO ALA GLY — ALA
H1 Trout	LYS PRO ALA LYS LYS ALA ALA ALA PRO LYS ALA LYS LYS VAL ALA ALA LYS LYS PRO ALA

141–154	
H1pCH4.8E	*ALA *ALA LYS LYS PRO LYS LYS *ALA ALA — ALA VAL — LYS LYS *SER
H1 Rabbit	THR PRO LYS LYS PRO LYS LYS ALA ALA GLY ALA LYS LYS ALA VAL — LYS LYS THR
H1 Trout	ALA ALA LYS LYS PRO LYS LYS VAL ALA ALA LYS LYS PRO ALA VAL ALA ALA LYS LYS SER

155–165	
H1 pCH4.8E	PRO LYS LYS ALA — LYS — LYS PRO — ALA *SER *ALA — *ALA
H1 Rabbit	PRO LYS LYS ALA PRO LYS PRO LYS ALA ALA ALA LYS PRO LYS VAL ALA LYS PRO LYS SER
H1 Trout	PRO LYS LYS ALA — LYS — LYS PRO — ALA — THR

166–185	
H1 pCH4.8E	*THR *LYS LYS *ALA ALA LYS SER PRO LYS LYS ALA THR LYS ALA *VAL LYS PRO LYS ALA ALA
H1 Rabbit	PRO ALA LYS VAL ALA LYS SER PRO LYS LYS ALA — LYS ALA VAL LYS PRO LYS ALA ALA
H1 Trout	PRO LYS LYS ALA ALA LYS SER PRO LYS LYS ALA THR LYS ALA ALA LYS PRO LYS ALA ALA

186–203	
H1 pCH4.8E	LYS *SER LYS — ALA *ALA LYS SER — LYS *ALA *ALA LYS ALA LYS *LYS *ALA ALA *THR LYS
H1 Rabbit	LYS PRO LYS — ALA PRO LYS — PRO LYS ALA ALA LYS ALA LYS LYS THR ALA ALA LYS
H1 Trout	LYS PRO LYS ALA ALA ALA LYS SER PRO LYS LYS VAL LYS — LYS PRO ALA ALA ALA LYS

204	
H1 pCH4.8E	LYS LYS
H1 Rabbit	LYS LYS LYS
H1 Trout	LYS

The most characteristic feature of the H1 protein sequence is the high lysine and alanine content and interspaced proline residues. Repeated duplication of resideis is also common, e.g. *Lys-Lys-Ala* at positions 113-115, 132-134, 146-148, 156-158, 167-169, 174-176 and 198-200 and *Lys-Lys-Pro* at positions 125-127, 136-138, 143-145 and 159-161.

The H1 histones are thought to be involved in close packing of the elementary nucleosome repeat structure into higher orders of chromatin condensation.

Variations that occur between H1 histone proteins possibly might allow the chromatin to take on highly specific conformations in certain areas, and the repeated duplication of amino acid sequences are thought to have been selected during the course of evolution to enable optimal interaction of histone H1 with the DNA (Von Holt *et al.* 1979).

In summary, it appears that the H3 protein sequence has remained highly conserved, whereas only the globular regions of H2B and H1 histone proteins are conserved. It has been suggested (Von Holt *et al.*, 1979) that during the course of evolution, the histone protein molecule (in particular H2B) has been subjected to two entirely different selection pressures focused onto different parts of the molecule. Histone-histone interaction, necessary for the chromatin structure, has favoured a relative constancy in the C-terminal hydrophobic part of the molecule, whereas the changing demands of histone-DNA interaction may have selected a variety of N-terminal sequences. The latter may thus provide a variety of chromatin structures suitable for specific regulatory processes.

5.2.7 Codon Usage

It was interesting to investigate whether the codon usage in the conserved domains of the histone proteins was also under the same stringent level of selection as the amino acid sequence. That is, is codon usage dictated by the protein type or is it characteristic of the species type in which the gene is expressed.

Table (5.1) compares the codons used by the chicken H3, H2B and H1 genes in λCH.02, with the codon usage of chicken H3 and H2B genes in λCH.01 (P. Krieg and R. Harvey, personal communication) and with the sea urchin H3, H2B and H1 genes from *P. miliaris* (Grantham *et al.*, 1980).

The H3 and H2B genes from both chicken clones have virtually identical codon preferences for each amino acid. This differs markedly from the codon preferences used by the sea urchin H3 and H2B genes. For example, in the chicken H3 and H2B genes, leucine is encoded only by GTG, lysine only by AAG, glutamic acid only by GAG and histidine only by CAC; whereas there is a more general selection of codons used by the sea urchin histone genes, e.g. leucine is encoded by CTA, CTC, CTG and CTT, lysine is encoded by both AAG and AAA and histidine by both CAC and CAT.

The codons utilised by the chicken H1 gene are similar for most amino acids to the other chicken genes (Table 5.1). Major exceptions however occur for proline, serine, lysine and alanine where a range of codons are used compared with the restricted usage in chicken H3 and H2B genes. This may primarily be a function of the higher frequency of these amino acids in the chicken H1 protein

Table 5.1

Codon Usage of Histone Genes

The frequency of use of individual codons is compiled for the chick H3, H2B and H1 genes sequences in λCH.02 compared with the genes of λCH.01, the sea urchin (SU), *P. miliaris* (PSM) genes (Grantham *et al.*, 1980) and the chick globin and ovalbumin genes (Grantham *et al.*, 1980).

Species		Chick λCH.02	Chick λCH.01	SU PSM	Chick λCH.02	Chick λCH.01	SU PSM	Chick λCH.02	SU PSM	Chick	Chick
Gene		H3	H3	H3	H2B	H2B	H2B	H1	H1	Oval	Globin
Arg	CGA	0	0	22	0	0	0	0	0	0	0
	CGC	3	5	44	4	5	10	0	12	5	21
	CGG	1	2	0	2	2	29	1	0	0	0
	CGT	2	28	37	0	0	10	0	12	0	0
	AGA	0	0	22	0	0	0	0	12	26	0
	AGG	0	0	7	0	0	29	1	0	8	21
Leu	CTA	0	0	15	0	0	0	1	0	3	0
	CTC	0	0	15	0	0	29	0	12	16	41
	CTG	1	2	37	6	6	0	0	24	36	75
	CTT	0	0	15	0	0	20	0	0	21	7
	TTA	0	0	0	0	0	0	0	0	8	0
	TTG	0	0	7	0	0	10	0	12	10	0
Ser	TCA	0	0	0	0	0	10	1	0	18	0
	TCC	0	1	7	5	3	20	1	12	13	34
	TCG	1	1	0	7	7	0	1	12	3	0
	TCT	0	0	7	0	1	0	1	35	26	0
	AGC	1	1	22	3	3	49	3	0	23	14
	AGT	0	0	7	0	0	20	0	0	16	0
Thr	ACA	0	0	15	1	1	10	0	12	18	0
	ACC	2	1	52	7	6	69	4	12	10	27
	ACG	3	6	7	2	2	0	0	35	0	0
	ACT	0	0	0	0	0	10	1	35	10	21
Pro	CCA	0	0	7	0	0	10	1	12	18	0
	CCC	2	2	22	5	4	0	5	0	3	27
	CCG	1	2	0	1	2	0	1	12	0	0
	CCT	0	0	15	1	0	10	1	12	16	7
Ala	GCA	0	0	30	1	0	10	3	35	42	0
	GCC	3	3	67	7	5	59	4	47	23	75
	GCG	5	6	0	4	6	0	11	35	0	7
	GCT	0	0	37	0	1	0	10	24	26	27
Gly	GGA	0	0	30	0	0	10	0	12	16	0
	GGC	3	4	0	7	7	20	3	24	21	27
	GGG	3	1	0	0	0	10	0	35	5	21
	GGT	0	0	22	1	0	29	0	0	8	7
Val	GTA	0	0	22	0	0	0	0	24	13	0
	GTC	0	0	15	4	4	39	0	24	18	34
	GTG	1	2	0	4	4	29	3	12	26	48
	GTT	0	0	7	0	0	10	0	12	23	0
Lys	AAA	0	0	15	0	0	29	7	35	29	7
	AAG	8	9	81	22	21	88	36	118	23	62
Asn	AAC	0	0	0	3	3	29	0	12	18	41
	AAT	0	0	7	0	0	0	0	0	26	7
Gln	CAA	0	0	15	1	0	0	0	24	21	7
	CAG	2	3	44	2	3	29	0	35	18	27
His	CAC	1	1	7	3	3	20	0	12	8	27
	CAT	0	0	7	0	0	10	0	12	10	21
Glu	GAA	0	0	15	0	0	10	0	24	47	27
	GAG	1	2	44	7	7	49	2	24	39	21
Asp	GAC	0	0	7	3	3	10	0	12	16	34
	GAT	0	0	15	0	0	10	0	24	21	7
Tyr	TAC	1	2	22	5	5	49	0	12	18	14
	TAT	0	0	0	0	0	0	0	0	8	0
Cys	TGC	0	0	7	0	0	0	0	0	3	7
	TGT	0	0	0	0	0	0	0	0	13	14
Phe	TTC	0	0	30	3	2	10	1	24	23	34
	TTT	0	0	0	0	0	10	0	0	29	21
Ile	ATA	0	0	0	0	0	0	1	12	8	0
	ATC	0	1	44	4	7	59	0	59	42	41
	ATT	0	0	7	0	0	10	0	0	16	7
Met	ATG	0	0	15	2	2	20	0	12	42	7
Trp	TGG	0	0	0	0	0	0	0	0	8	27

sequence, and therefore the more general codon usage for these amino acids may simply reflect the limited supply of specific tRNA's.

In order to determine if non-histone chicken genes share a similar codon usage to that of chicken histone genes, the third base codon preferences of chicken ovalbumin and chicken β globin genes were compared to that displayed by the chicken histone genes.

Table 5.2 shows that for chicken histone H3 and H2B genes, in both λCH.02 and λCH.01, there is a remarkably strong bias for G and C residues in the third base, whereas A and T residues comprised only 0-4% of the third bases. This is in contrast with sea urchin where a higher proportion of codons have A and T residues in third base position, thus resulting in the more general codon selection (Table 5.2). The chicken H1 histone gene also displays a higher proportion of A and T residues in the third base, but still retains the high G content (56%) as is characteristic of the other chicken histone genes.

However, when the codon third base preferences for chicken ovalbumin and chicken β globin are compared to the chicken histone genes, it is apparent that codon usage is not simply a function of species type, since chicken ovalbumin displayed an average use of all four nucleotides while chicken histones and chicken β globin displayed a general G and C preference.

In order to determine if correlations existed between third base preferences of another gene family in different species, the β globins were compared from chicken, mouse, rabbit and human, (Grantham *et al.*, 1980). Table 5.2 shows that in all these

Table 5.2

Nucleotide Use in Third Position of Codons

The percentage use of each of the four nucleotides is calculated for the third base position of codons in the H3, H2B and H1 genes in chicken λCH.02, and is compared to the chicken λCH.01 genes, the sea urchin h19 (Busslinger *et al.*, 1980) and PSM (Grantham *et al.*, 1979) genes and to the chick ovalbumin, chick β globin, mouse β globin, rabbit β globin and human β globin genes.

GENES	G	A	T	C	SPECIES
H3	62%	0	4%	33%	λCH.02
	61%	0	3%	35%	λCH.01
	22%	28%	15%	33%	h19
H2B	45%	3%	2%	50%	λCH.02
	49%	2%	2%	47%	λCH.01
	31%	11%	16%	42%	h19
	33%	9%	12%	46%	PSM
H1	56%	10%	8%	27%	λCH.02
	36%	20%	25%	24%	PSM
Chick Ovalbumin	20%	24%	27%	27%	Chick
Chick β Globin	31%	4%	14%	50%	Chick
β Globin	30%	8%	27%	35%	Mouse
β Globin	38%	6%	28%	27%	Rabbit
β Globin	34%	7%	27%	32%	Human

species third base nucleotide preferences have been conserved. Therefore, unlike the histone gene codon usage which appears to be both species and gene specific, the β-globin gene codon usage appears to be only gene specific. Consequently from these examples it is not possible to generalise about what governs the codon usage of a particular gene.

The different codon preferences that are observed between the chicken histone genes and the chicken ovalbumin genes may only reflect some metabolic discrimination between bases or different codon and anticodon populations in different cell types. It may also be possible that mRNA secondary structure optimisation by third base choice may explain the similarities of codon usage observed for the β globin genes between species as well as the differences in codon usage observed for the chicken H1 gene with the other chicken histones.

5.2.8 Regulatory Sequences

The theory underlining sequence comparisons between genes has been that regions involved in gene regulation will show greater specific sequence conservation than other non-transcribed sequences, because of their basic requirement for interaction with other molecules. This approach has been successfully used to identify important eukaryotic sequence elements (Busslinger *et al.*, 1980) and in some cases enabled the design of experiments to demonstrate their function (e.g. Wasylik and Chambon, 1981).

In the sections below the non-translated sequences of the histone genes of λCH.02 determined in this study are compared with "consensus" sequences of other histone genes.

(a) 5' Untranslated Sequences

From the organisation and direction of transcription of the histone genes in λCH.02, it appears that each gene is individually transcribed. Therefore the 5' untranslated region preceding each of the genes should contain signals for the control of transcription initiation and gene "turn on".

The 5' untranslated regions, upstream from the structural genes have been determined for the two H3 genes (Fig. 5.7 a,b) and the H2B gene (Fig. 5.7 d) sequenced in λCH.02. The 5' untranslated sequences of the two H3 genes in λCH.02 have been compared in Fig. (5.12), to the 5' sequences upstream from the chicken H3 gene in λCH.01 (P. Krieg, personal communication), the *Xenopus* H3 gene in XI-hi-I (Moorman and Destree, 1980), the *Drosophila* H3 gene in DM500 (Lifton *et al.*, 1977) and the H3 genes in sea urchin clones h19 and h22 (Busslinger *et al.*, 1980). The 5' untranslated sequences upstream from the H2B gene in λCH.02 have been compared in Fig. (5.12) to the 5' untranslated sequences of the chicken H2B gene in λCH.01 (R. Harvey, personal communication) as well as to the 5' regions preceding H2B genes sequenced from *Drosophila* DM500 (Lifton *et al.*, 1977) and sea urchin h19 and h22 (Busslinger *et al.*, 1980).

Comparison of these prelude sequences reveals that the 5' untranslated regions for each of the histone genes have totally diverged except for a few regions of homology.

The first region of strong homology is an A-T rich sequence containing the "TATA" box, located 20-21 bp upstream from

Fig. 5.12

Histone Gene 5' Homology Blocks

The 5' homology sequences are shown for the λCH.02 chicken H3 genes in pCH2.6H and pCH4.3EH, and the λCH.02 chicken H4 gene in pCH4.8E. These are compared to the 5' sequences of the H3 and H2B genes in chicken λCH.01 (P. Kreig, R. Harvey, personal communication) *Xenopus*, XI-hi-1, *Drosophila*, DM500, (as summarised by Hentschel and Birnstiel, 1981) and sea urchin h19 and h22 (Busslinger *et al.*, 1980). The numbers indicate the number of bases between homology blocks. Where "Cap" box sequences are underlined, there is evidence that these nucleotides correspond to the 5' end of mRNA's (Hentschel and Birnstiel, 1981).

Histone Gene 5' Homology Blocks

Gene	"CCAAT Box"		"TATA Box"	"Cap Box"
H3				
pCH2.6H	GACCAATCA-	39	-CTATAAATAA-20-	CGACTTT-16-TGGAATG
pCH4.3EH				GCAGTTG-14-TGTGATG
λCH.01	GGCCAATCA-	39	-GTATAAATAG-21-	CTATTTC-31-AGCGATG
X1-hi-1	ATCCAAGGG		CTATAAGGTG-21-	ACAGAGG-83-CACTATG
Dm500	CGTCAAATG-42-CGTCC-9-		ATATAAGTAG-20-	TTATTGT-52-GGAGATG
h19	GACCAATCA-34-GATCC-10-		ATATAAATAG-20-	TCATTCA-47-AACTATG
h22	GACCAATCA-34-CATCC-10-		GTATAAATAG-21-	TCATTCA-46-AACCATG
H2B				
pCH4.8E	ATCCAATCA-25-GATTTGCATA-6-		CTATAAATAG-20-	GCACGTC-28-CGACATG
λCH.01	GACCAATGA-23-CATTTGCATA-6-		CTATAAATAA-20-	CCATTCA-48-CACTATG
Dm500			GTATAAACGT-13-	CCATTCG-37-AACAATG
h19	GCCCAATGA-45-GATCC		-10-GTATAAAAAG-16-	CCATTCA-70-CATAATG
h22	AGCCAATCA-50-GA CC		-9-GTATAAAGAG-20-	GCACTCA-67-CACCATG
CONSENSUS	$G\frac{G}{A}CCAAT-\frac{A}{G}$	GATCC	GTATAAATAG	$\frac{T}{C}CATTC\frac{G}{A}$ CA--ATG

the putative mRNA start or "cap box". The "TATA" box sequence was first recognised in the 5'-untranslated region of the *Drosophila* histone genes (Goldberg, 1979) and has subsequently been identified in many other eukaryotic genes (Efstratiadis *et al.*, 1980). The conserved location of the "TATA" box and its similarity to a sequence characteristic of all prokaryotic promoters (Pribnow, 1979), led many investigators to speculate that it might play a role in the initiation of transcription in eukaryotes. Indeed, recent experiments have shown that deletion or mutation within the "TATA" box sequence has drastic effects on the level and specificity of transcription (Breathnach and Chambon, 1981).

It is clear that for the chicken histone genes like almost all other RNA II polymerase genes, the "TATA" box sequence has been conserved because of its importance in controlling initiation of transcription.

A second major region of homology in the 5' untranslated region occurs at a variable distance upstream from the "TATA" box, and this region includes the sequence CCAAT ("CAT" box). This sequence is present not only in all histone genes compared, but also has been found in a similar position relative to the "TATA" box in other (but not all) eukaryotic genes (Efstratiadis, *et al.*, 1980). Its significance is not known, but recent experiments have shown that, *in vivo*, sequences isolated in this region are essential for transcription (Mellon *et al.*, 1981).

A third region of homology which is characteristic of all sea urchin histone genes, located 9-10 bp upstream from the "TATA" box is a pentameric sequence 5' GATCC 3'. However, this sequence does not appear to have been conserved in the prelude sequences of the H3 histone genes in chicken or *Xenopus*. The role of this sequence in sea urchin histone genes is not known; it is therefore not possible to determine whether this function is altered or no longer required in the expression of the chicken histone genes or whether the sequences involved in the function have diverged.

However, it is of interest that R. Harvey has observed a conserved sequence (5' CATTTGCATA 3') that occurs in the same position, 6 bp upstream from the "TATA" box in both the chicken H2B gene in λCH.02 and the H2B gene in λCH.01. This sequence appears to be unique to chicken H2B genes, as it has not been observed in other chicken histone genes sequenced (P. Krieg, R. Harvey and A. Robins, unpublished data) in λCH.01. It will therefore be interesting to compare 5' untranslated regions of other chicken and higher eukaryotic H2B genes in order to determine whether this homology block has also been conserved in these genes or whether it plays a specific role in a subset of H2B genes.

A fourth region of homology is found at the site of transcription of initiation of the sea urchin histone genes, 20-21 bp downstream from the "TATA" box, and there is evidence that the adenosine nucleotides underlined in Fig. (5.12) correspond to the 5' end of the mRNAs (Grunstein *et al.*, 1981). Similar putative "cap site" sequences are found 20-21 bp downstream

from the "TATA" box in the chicken H2B and H3 genes and it is likely that this sequence defines the end of the mRNA's. However, it is obvious that there is an absence of strong sequence conservation at this site, and this is consistent with the wide variation in cap sites reported for eukaryotic genes (Benoist *et al.*, 1980).

Both chicken H3 genes of λCH.02 probably have short (23 and 25 bases) untranslated regions, compared with 40 bp of the other chicken H3 gene of λCH.01 and 55 to 90 bases for other species. For the H2B genes, the postulated lengths of the chicken 5' non-translated regions of the mRNAs (37 and 57 bases) are not markedly different from those of other species (46 to 79 bases). Furthermore, there is essentially no sequence homology within the 5' untranslated regions of any of the genes and there remains little clue as to their function significance.

It is clear that whereas the coding potential of the chicken histone genes has been remarkably conserved, the 5' untranslated region has diverged except for a few homology blocks which are thought to play a role in controlling the expression of these genes.

(b) 3' Untranslated Sequences

Histone genes differ from other gene families in that large blocks of sequence homology are located downstream from the termination codons and that the mRNAs for most histones are not polyadenylated (Hentschel and Birnstiel,1981).

Fig. (5.13) compares the 3' untranslated region of the chicken H3, H4 and H2B genes of λCH.02, sequenced in this study, with the 3' homology blocks found in the chicken H3 gene of λCH.01, a mouse H4 gene, a *Xenopus* H3 gene, a *Xenopus* H4 cDNA clone sequence, a *Drosophila* H3 gene and sea urchin H3, H2B and H4 genes.

The 3' homology block, first noted in sea urchin histone genes by Sures *et al.*, (1978) is about 40 bases long, and has now been found in the 3' regions of all histone genes sequenced (Fig. 5.13), as well as in the chicken histone H3, H4 and H2B genes of λCH.02. This region is characterised by two highly homologous blocks (Busslinger *et al.*, 1979).

The first block, 23-47 bases 3' to the termination codon is 23 bases in length and contains a central GC-rich dyad symmetry motif hyphenated by TTTC. The hair-pin loop structure form is shown in Fig. (5.14) for the H3, H4 and H2B genes of λCH.02. Whereas a six base bair stem is seen in both the H2B and H4 chicken genes (as well as most other histone genes) a single base deletion in the H3 chicken sequence of pCH2.6H has resulted in a 5 base pair stem structure. A 5 base pair stem structure also occurs for the chicken H2A gene of λCH.01 (R. D'Andrea, personal communication), Fig. 5.13, also as a result of the same single base deletion. The functional significance however of a 5 base pair stem-loop versus a six base pair stem-loop is unknown.

The hyphenated dyad symmetry structure resembles in many respects the prokaryotic promoter or attenuator sequences

Fig. 5.13

Histone Gene 3' Homology Blocks

The 3' homology sequences are shown for the λCH.02 chicken histone H3 gene in pCH2.6H, the H2B gene in pCH4.8E and the H4 gene in pCH2.6H. These are compared to the 3' homology sequences in the H3, H2B, H2A and H4 genes in chicken, λCH.01 (P. Krieg, D'Andrea personal communication), *Xenopus*, X-hi-I, pXIch4, *Drosophila*, DM500 (as summarised in Hentschel and Birnstiel, 1981) and sea urchin h19, h20 (Busslinger *et al.*, 1979). Gene sequences are aligned for maximum homology.

Histone Gene 3' Homology Blocks

Gene	Hyphenated Dyad Symmetry
H3	
pCH2.6H	TGA-32- CAA GCTCTTTTCAGAGCCACCC- 8b -AGGGAAGAGC
X1-hi-1	AAAGGCTCTTTTCAGAGCCACAAcattccCAGTCAAAT
Dm500	ATCGGTCCTTTTCAGGACCACAA- 8b -CAATGAGAT
h19	AACGGCTCTTTTCAGAGCCACCAcaacccCAAGAAAGAA
h22	AACGGCTCTTTTCAGAGCCACCAcaccccCAAGAAAGAA
H2B	
pCH4.8E	TAG-28 - AAAGGCTCTTTTCAGAGCCACCA- 7b -TAATAAAAGG
h19	AACGGCCCTTTTCAGGGCCACCAaataatCAAGAAAGAA
h22	AACGGCCCTTTTCAGGGCCACCAaacatcCAAGAAAGAA
H2A	
λCH.01	AAA GCTCTTTTCAGAGCCACCC- 7b -CGCAGGAGAG
H4	
Mushi-1	AACGGCCCTTTTTAGGGCCACCA- 10b -CAGGAGAGC
pCH2.6H	TAA-35- CAAGGCTCTTTTCAGAGCCACCCA- 8b-AAGGGAAGAG
λCH.01	AAAGGCTCTTTTCAGAGCCGCCCA- 8b-CAACAAAGGG
pX1ch4	ATAAGCCCTTTTAAGGGCCACCA-polyA
h19	AACGGCTCTTTTCAGAG CACCAaataatCAAGAAAGAG
CONSENSUS	AACGGC$\frac{C}{T}$CTTTTCAG$\frac{G}{A}$GCCACCA 3' CAAGAAAGA$\frac{A}{G}$

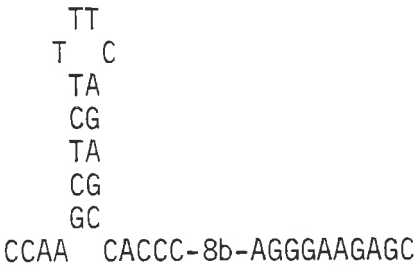
Fig. 5.14

3' Hypenated Dyad Symmetry

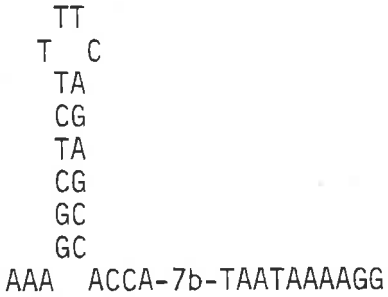
The hairpin loop structure is shown for the 3'-hyphenated dyad symmetry blocks of the H3, H2B and H4 chicken histone genes of λCH.02.

3' Dyad Symmetry

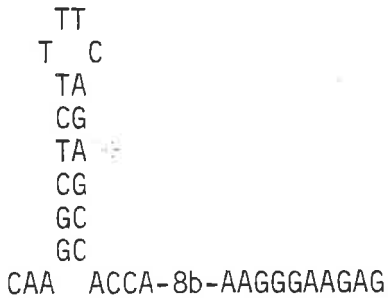
H3



H2B



H4



(Pribnow, 1979), Rosenberg and Court, 1979) as well as putative polymerase III terminators (Korn and Brown, 1978). In the case of the histone genes however, it is clearly the sequence *per se* that is being conserved and not just the potential of an RNA with a sequence to form a hairpin loop. This suggests that whatever the exact function of this conserved sequence, a specific protein or nucleic acid recognition event may be involved. In fact, recent experiments by Birchmeier *et al.*, (1982) have shown that this region is essential for the production of correct 3' termini in sea urchin histones.

The second homology block, 6-8 bases further downstream, is of the form 5' CAAGAAAG 3' in sea urchins; while less conserved in other species the region is still very rich in A and G residues. The functional significance of this region is unknown, but it appears this function is either altered or no longer required in chicken histone genes since the sequences have diverged considerably.

Evidence as to the exact location of the histone mRNA 3' termini relative to these homology blocks have been somewhat inconsistent, possibly reflecting real biological variability in this respect. SI-mapping experiments with *P. miliaris* mRNAs (Hentschel *et al.*, 1980) and the mouse H4 mRNA (Seiler-Tuyns and Birnstiel, 1981) located all 3' termini close to the end of the first homology block and within or just after the 5' ACCA 3' motif (Fig. 5.13), implying that these mRNAs share with many prokaryotic mRNAs the feature of a 3'-terminal hairpin loop. In contrast, Grunstein *et al.*, (1981)

failed to detect oligonucleotides corresponding to this region in a finger print of *S. purpuratus* H4 mRNA (the last detectable oligonucleotide mapping before the dyad symmetry region and only 13 nucleotides 3' to the termination codon).

Perhaps the most pertinent data in this regard stems from the 3'-terminal sequencing of DNA clones derived from *Xenopus* oocyte poly(A)+ H4 mRNAs (Zernick *et al.*, 1980). Here the mRNA clearly contains the dyad symmetry region as well as part of the 5' ACCA 3' motif followed by the poly(A) tail as shown in Fig. 5.14. It is also worth noting that these mRNA molecules lack the putative poly-adenylation signal sequence 5' AATAAA 3' (Proudfoot and Brownlee, 1976), although they are polyadenylated. Conversely, the yeast H4 gene (M. Smith, cited in Hentschel and Birnstiel, 1981) has the AATAAA sequence about 43 bases downstream from the termination codon, and the mRNA from it is polyadenylated (as are other yeast histone mRNAs (Fahrner *et al.*, 1980)). It will be interesting to see if the yeast gene also has the conserved homology block in addition to the polyadenylation signal.

The lack of polyadenylation signal however appears to be a general feature of all histone gene sequences (Busslinger *et al.*, 1979; Sures *et al.*, 1980) including the chicken histone genes; this is in contrast to most other eukaryotic polymerase II genes sequenced (Breathnach and Chambon, 1981). It is therefore interesting to note that in the chicken H2B gene of pCH4.8E, there is a 5'AATAAA 3' sequence 60 bases downstream from the termination

codon, but located within the second homology block region. It is therefore possible that the mRNA transcribed from this H2B gene may be polyadenylated.

However, it is quite clear that for most histone mRNA's the region of dyad symmetry defines the end point of the mRNA although it remains to be determined whether this site acts as a signal for transcription termination or for cleavage of a longer primary transcript. In addition, the presence of a hair pin loop structure correlates well with the lack of polyadenylation observed for the majority of histone mRNA's.

CHAPTER 6

ISOLATION AND CHARACTERISATION OF HUMAN HISTONE GENOMAL CLONES

6.1 INTRODUCTION

It was apparent from the organisation of the chicken histone genes in λCH.02 as well as the organisation of histone genes in other chicken genomal clones isolated (Harvey *et al.*, 1981; Engel and Dodgson, 1981), that the chicken histone genes do not conform to the 'typical' tandem repeat organisation observed in sea urchin (Kedes, 1979) and *Drosophila* (Lifton *et al.*, 1977). Therefore, it is of interest to determine the organisation and structure of the human histone genes in order to extend the comparison of histone gene organisation through evolution, with the aim of determining what features of histone genes have been conserved and therefore may be important for their expression.

It had already been established that the chicken histone genes could detect specific sequences in human DNA (4.2.2). Therefore, advantage was taken of the purified chicken histone coding probes isolated and subcloned from λCH.01 (2.1.3), to screen both the EcoR1* Human Library prepared earlier (as described in Chapter 3) and the Maniatis Human Library.

The first section of this chapter describes the isolation and characterisation of two genomal clones containing human histone genes and the second section describes an analysis of some of the histone gene sequences.

PART A: ISOLATION AND ORGANISATION OF HUMAN HISTONE GENES

6.2 RESULTS

6.2.1 Preparation of Probes

In order to identify clones containing human histone gene sequences from the two genomal libraries, a combination of chicken histone gene probes were used. Coding regions from the H4, H3, H2A and H2B genes from λCH.01 had been cloned into pBR322 (D'Andrea) as described in (2.1.3). To avoid pBR322 cross-hybridisation with the Charon 4A arms of the recombinant clones, the histone gene inserts were excised with the appropriate restriction enzymes and separated by sucrose density centrifugation followed by acrylamide gel electrophoresis.

6.2.2 Screening of the EcoR1* Human Library

The EcoR1* Human Library (described in Chapter 3), when constructed, consisted of 7.6×10^5 independently derived bacteriophage λ clones consisting of large (15-20 kb) fragments of human placental DNA covalently joined to a Charon 4A λ vector. The library had been segregated into 10 separate pools (7.6×10^4 plaques/pool) and 40,000 pfu from each pool had been amplified, approximately 1×10^6 fold, recovered as a phage lysate (4×10^{10} pfu/pool), and stored at 4^oC for later screening.

When the library was screened (18 months later) the titre of the amplified pools had reduced to 10^6 pfu/pool.

A total of 500,000 phage were screened by plating 50,000 phage from each pool onto ten 15 cm diameter plates. Duplicate nitrocellulose filters were prepared from each plate using the procedure of Benton and Davis (1977). These filters were hybridised with a combination of nick-translated H3, H4, H2A and H2B chicken histone DNA inserts (1×10^6 cpm of each per duplicate filter Fig. 6.1).

After washing the filters several times at 65°C in 1 x SSC, 0.1 % SDS they were autoradiographed with an intensifying screen overnight at -80°C. The autoradiographs were aligned and analysed for positive signals. Only one plaque was detected and this clone was picked and replated (approximately 100 phage/plate) and again duplicate filters were prepared and screened with the four chicken histone gene probes. This allowed verification of the original positive signal, as well as selection of a single plaque. The single plaque was picked and plated for third round screening to ensure purity of the phage. This clone was termed λHH.02 and its characterisation will be presented in Section 6.2.6.

6.2.3 Screening of the Maniatis Human Library

When the Maniatis Human Library (Lawn *et al.*, 1978) was previously screened (Chapter 4), twelve recombinants were selected which hybridised to both chicken histone cDNA and to a genomal fragment containing chicken histone genes. However, these isolates were no longer viable when tested 18 months later. Therefore the Maniatis library was rescreened for human histone-containing clones.

Fig. 6.1

Screening the EcoRI* Human Library

An example of the first, second and third round screening of the human EcoRI* library. Duplicate nitrocellulose filters from each plate were hybridised with labelled H3, H4, H2A and H2B histone DNA probe and autoradiographed at -80°. The arrow indicates the positive plaque picked for further screening (λHH.02).

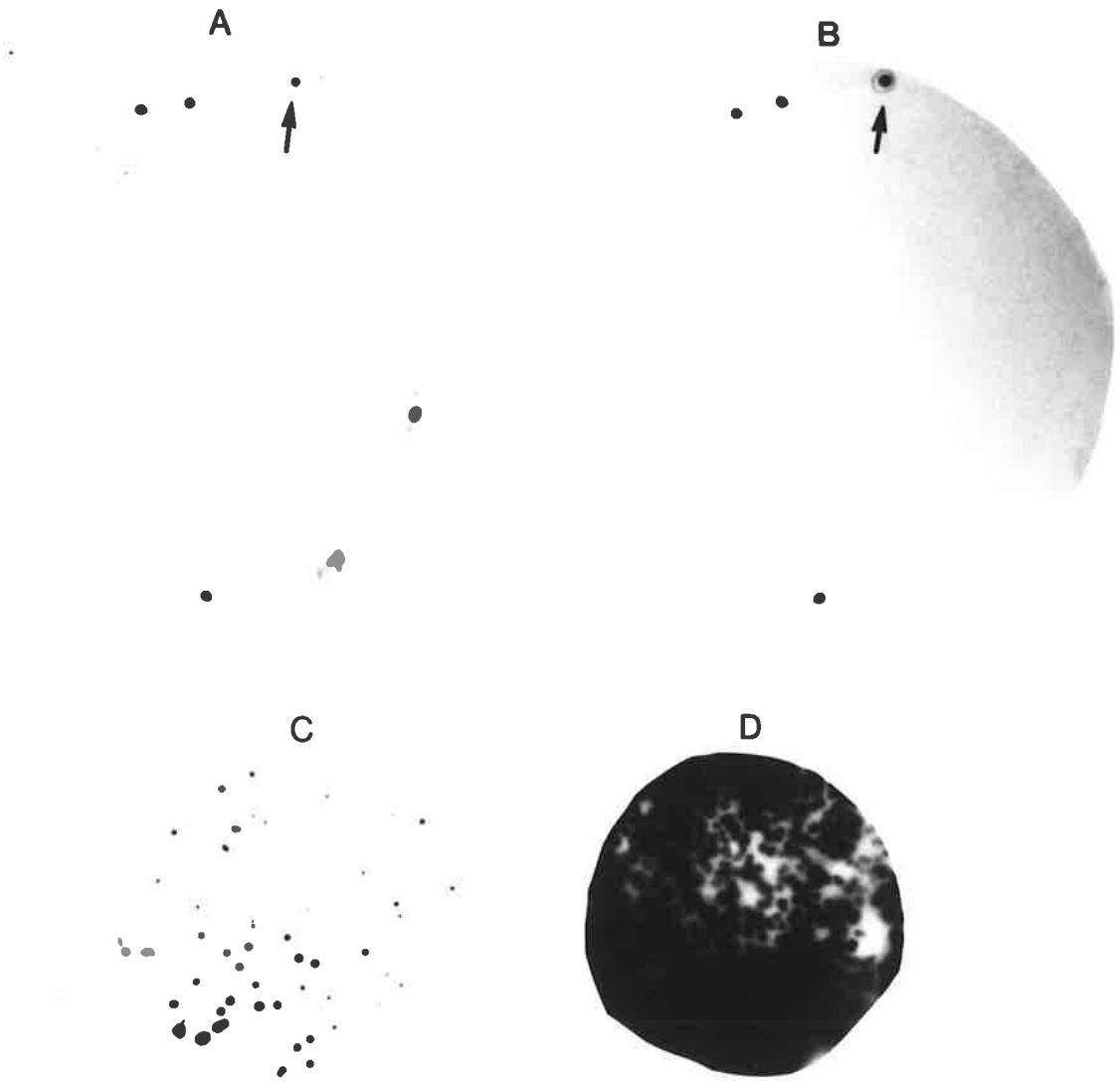
A: First round screening
50,000 plaques per plate

B: Duplicate filter as shown in A

C: Second round screening

D: Third round screening

Note: Marker dots to aid alignment are shown in A and B.


A
B
C
D

When this library was originally constructed (Lawn *et al.*, 1978) it consisted of 1 x 10^6 independently derived bacteriophage λ clones. However, to ensure storage, it became necessary to passage it through a second amplification in addition to the one before it was received. Subsequently, a total of 600,000 phage from the amplified library were screened by plating 10, 15 cm diameter plates, each containing 60,000 phage and hybridising the corresponding duplicate nitrocellulose filters with nick-translated ^{32}P- H3, H4, H2B and H2A chicken histone DNA probes (1 x 10^6 cpm of each per set of filters), as described in the previous section.

From the 600,000 phage screened, two duplicate positive clones were detected and these were picked for rescreening (~100 phage per plate). On the second round of hybridisation with the four chicken histone DNA probes, a single positive plaque was picked for each of the two clones and plated for third round screening to ensure the purity of each phage.

DNA was prepared from each of the clones, digested with EcoR1 and electrophoresed on a 1% agarose gel. Fig. (6.2) shows that both these clones are identical, but different to λCH.02. The clone was termed λHH.03 and its characterisation is presented in Section (6.2.8).

For both the EcoR1* and Maniatis Human Libraries it appeared that the lowering of titre during storage coupled with amplification contributed to the low yield of histone-containing clones detected in these libraries.

Fig. 6.2

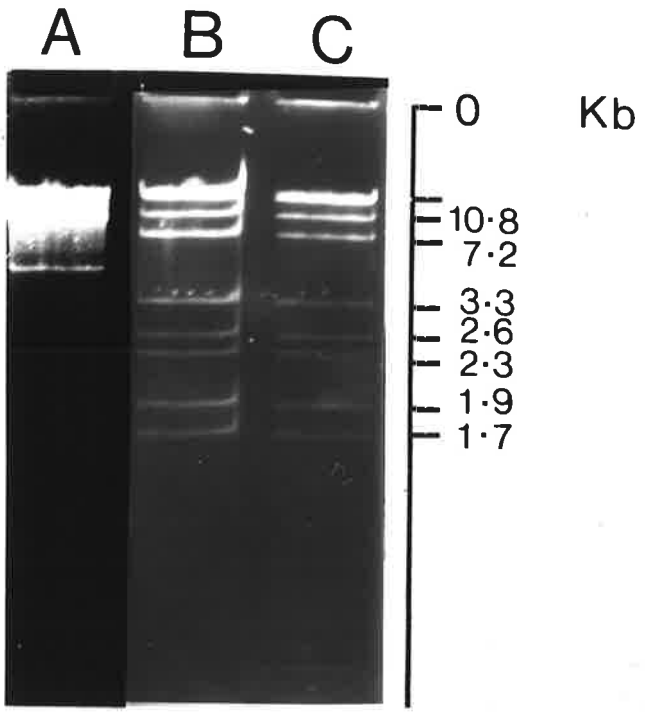
Restriction Enzyme Digests of Clones Selected from the Maniatis Human Library

DNA from the two clones, isolated from the Maniatis Human Library which hybridised with the chicken histone probes, were digested with EcoRI and electrophoresed on a 1% agarose gel. The tracks are as follows:

A: λCH.02 digested with EcoRI

B: Clone 1 digested with EcoRI

C: Clone 2 digested with EcoRI


A
B
C
0
Kb
10·8
7·2
3·3
2·6
2·3
1·9
1·7

6.2.4 Verification of λHH.02 and λHH.03 as Human Genomal Clones

In order to establish that the clones λHH.02 and λHH.03 were indeed of human origin, advantage was taken of the presence in eukaryotic genomes of repeated sequences approximately 300 nucleotides long which have been found interspersed with single copy sequences throughout the genome (Davidson *et al.*, 1975). In humans, the *Alu* DNA family is predominant among these reiterated sequences, its repetition number being 300,000 per haploid genome (Houck *et al.*, 1979). In addition, Tashima *et al.* (1981) have recently reported that over 95% of the recombinants present in the Maniatis Human Library hybridise to a probe containing *Alu* DNA sequences. Other similar families of interspersed repetitive sequences have also been identified in human DNA (Houck *et al.*, 1979). Recent evidence suggests that these interspersed repeat families show considerable species specificity. Therefore, if λHH.02 and λHH.03 do in fact contain human genomic DNA, they should have a high probability of also containing repetitive sequences and thereby hybridising strongly to total nick-translated human DNA.

Thus, in order to confirm that λHH.02 and λHH.03 both represented human recombinant clones (and not contaminating chicken histone genomal clones), DNA from each was spotted onto nitrocellulose and hybridised with nick-translated human genomic and chicken genomic DNA (1×10^6 cpm/filter). In addition λCH.02 and its subclones (pCH2.6H, pCH9.6E and pCH4.8E) as well as wildtype λ phage DNA and pBR322 DNA were spotted onto the same filters as controls.

Fig. (6.3) shows that the human DNA hydridises strongly to itself and to λHH.02 and λHH.03 but not to λCH.02 and its subclones. Conversely, the chicken DNA hybridises to λCH.02 and two of its subclones (pCH9.6E and pCH2.6H) as well as to itself. Both the λ and pBR322 DNA were not detected by either probe.

It is clear that under these conditions of hybridisation neither the total human DNA nor the chicken DNA were detecting histone gene sequences present in these clones, since a subclone from λHH.02, pHH1.9BE, which contains three histone genes (described in Section 6.2.7) was not detected; similarly pCH4.8E which is known to contain two chicken histone genes was also undetected.

Therefore the hybridisation observed corresponds to more highly repeated sequences in the spacer regions, specific to either the human DNA or to the chick DNA. From these results it can be concluded that λHH.02 and λHH.03 contain repeated DNA sequences homologous to human DNA and therefore must represent human genomic recombinant clones.

It is interesting to note that some of the human histone clones isolated by Sierra *et al.* (1982a) have been shown to be interspersed with members of the *Alu* DNA family as well as with other transcribed sequences (Sierra *et al.*, 1982b).

Fig. 6.3

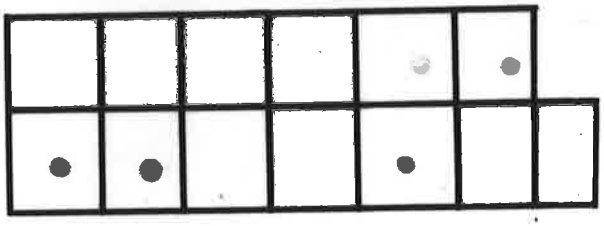
Human DNA Dot Blots

Both human and chicken genomal DNAs were nick-translated and hybridised to DNAs that had been denatured in 0.2N NaOH and spotted onto duplicate nitrocellulose filters in 10 x SSC (approximately 0.5 µg per spot). The order of the DNA spots is shown in the lower box. p4.8, p9.6, p2.6 are the pCH4.8E, pCH9.6E, pCH2.6H subclones of λCH.02 respectively. p1.9BE is the pHH1.9BE subclone of λHH.02 described in (6.4.1). pBR322 and λ DNAs were included as controls.

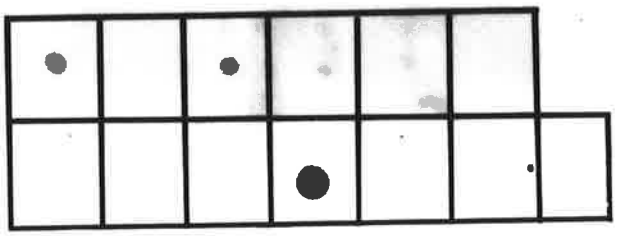
Box A: Probe: Nick-translated total human genomal DNA

Box B: Probe: Nick-translated total chicken genomal DNA

A



B



CH02	p4.8	p9.6	p2.6	HH03	HH03	
HH02	HH02	p1.9BE	CHICK	HUMAN	pBR	λ

6.2.5 Histone Gene Content of λHH.02 and λHH.03

To analyse which individual histone genes λHH.02 and λHH.03 contained, phage patches from each clone were transferred to nitrocellulose and filters hybridised independently with each of the five chicken histone gene probes (H3, H4, H2A, H2B from λCH.01 and H1 from λCH.02).

Fig. (6.4) shows that λHH.02 hybridises with the H3, H2A and H2B chicken histone genes and not with H1 or H4 genes. Whereas λHH.03 hybridises with the H2A and H2B chicken histone probes and not with H3, H4 and H1 gene probes. λCH.02 was used as a control since this clone contains all histone genes except H2A.

It therefore appears that neither λHH.02 nor λHH.03 contains a full complement of the 5 histone genes and that the H4 and H1 histone genes are not represented in either clone. In order to determine the location and order of the H3, H2A and H2B human histone genes in these two clones, a restriction map for each was constructed.

6.2.6 Restriction Enzyme Ananysis of λHH.02

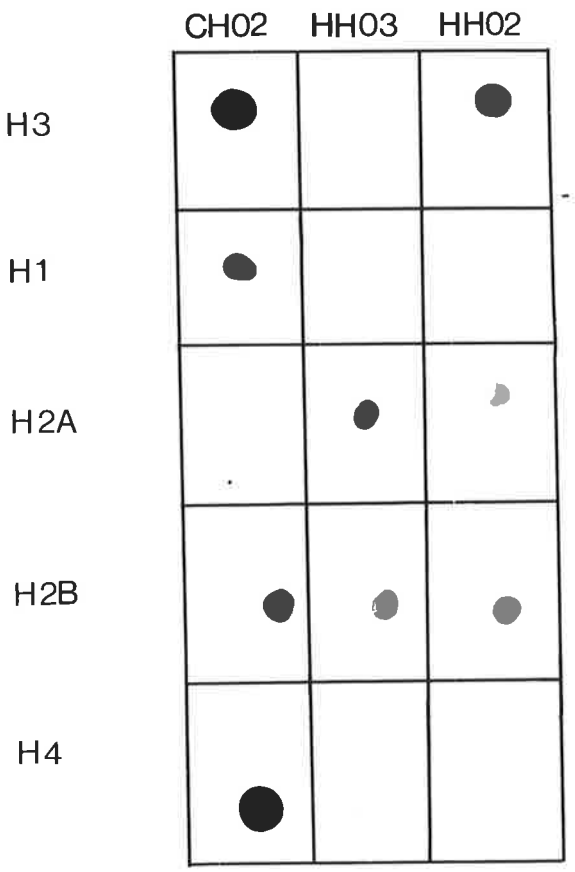
A restriction map of λHH.02 was constructed using the restriction enzymes EcoR1, HindIII, BamH1 and SalI.

Digestion of λHH.02 with EcoR1 resulted in three insert fragments, 9.2 kb, 3.9 kb and 1.5 kb in size (Track f), Fig. (6.5). The order of these fragments could be determined

Fig. 6.4

Histone Gene Content of λHH.02 and λHH.03

Approximately 0.5 μg of DNA from λCH.02, λHH.02 and λHH.03 was denatured on 0.2N NaOH, spotted onto nitrocellulose filters in 10 x SSC and hybridised separately with each of H3, H1, H2A, H2B and H4 labelled chicken histone probes (2.1.3) and autoradiographed at -80^{o}.


CH02
HH03
HH02
H3
H1
H2A
H2B
H4

from a BamHI/EcoRI double digest of λHH.02 and a SalI/EcoRI double digest, since both SalI and BamHI were found to digest the λHH.02 insert only at a single site each. BamHII digested the 9.2 kb EcoRI fragment to give a 1.9 kb BamII/EcoRI fragment and a 7.2 kb BamHI/EcoRI fragment (Track a, Fig. (6.5)) and SalI digested the 3.9 kb EcoRI fragment to give a 2.55 kb SalI/EcoRI fragment and a 1.25 kb EcoRI/SalI fragment (Track g, Fig. 6.5). A BamHI/SalI double digest indicated that these two sites were 11 kb apart, the SalI site being 6.0 kb from the BamHI site in the 10.8 kb Charon 4A arm (Track k, Fig. 6.5).

A HindIII digest of λHH.02 resulted in three large internal fragments 5.8 kb, 4.5 kb and 2.3 kb in size as well as smaller HindIII fragments 640, 355, 100 bp (Tracks D & M, Fig. 6.5, and Section 6.4.1).

The location of these HindIII fragments has been determined from an analysis of the products of double digestions with BamHI, SalI and EcoRI. Firstly, the unique sites of BamHI and SalI can be located in the 5.8 kb and 4.5 kb HindIII fragments respectively. Their orientations relative to the BamHI and SalI sites have been defined with reference to the products of an EcoRI/HindIII double digest. Two of the four small HindIII fragments (355 bp and 100 bp) have been located as shown in Fig. (6.6) from digests of the pHHI.9BE subclone described in Section (6.4.1). The summary map of this restriction data as shown in Fig. (6.6) is further supported by the hybridisation data with individual histone gene probes discussed in the next section.

Fig. 6.5

Restriction Enzyme Digests of λHH.02

DNA (1 μg) from λHH.02 was digested from EcoRI, HindIII, BamHI and SalI and electrophoresed on a 1% agarose gel. λDNA cut with EcoRI and HindIII was used as size markers. The tracks are as follows:

A:	BamHI/EcoRI
B:	BamHI
C:	BamHI/HindIII
D:	HindIII
E:	EcoRI/HindIII
F:	EcoRI
G:	SalI/EcoRI
H:	SalI
I:	Sal/HindIII
J:	BamHI
K:	BamHI/SalI
L:	SalI
M:	EcoRI/HindIII

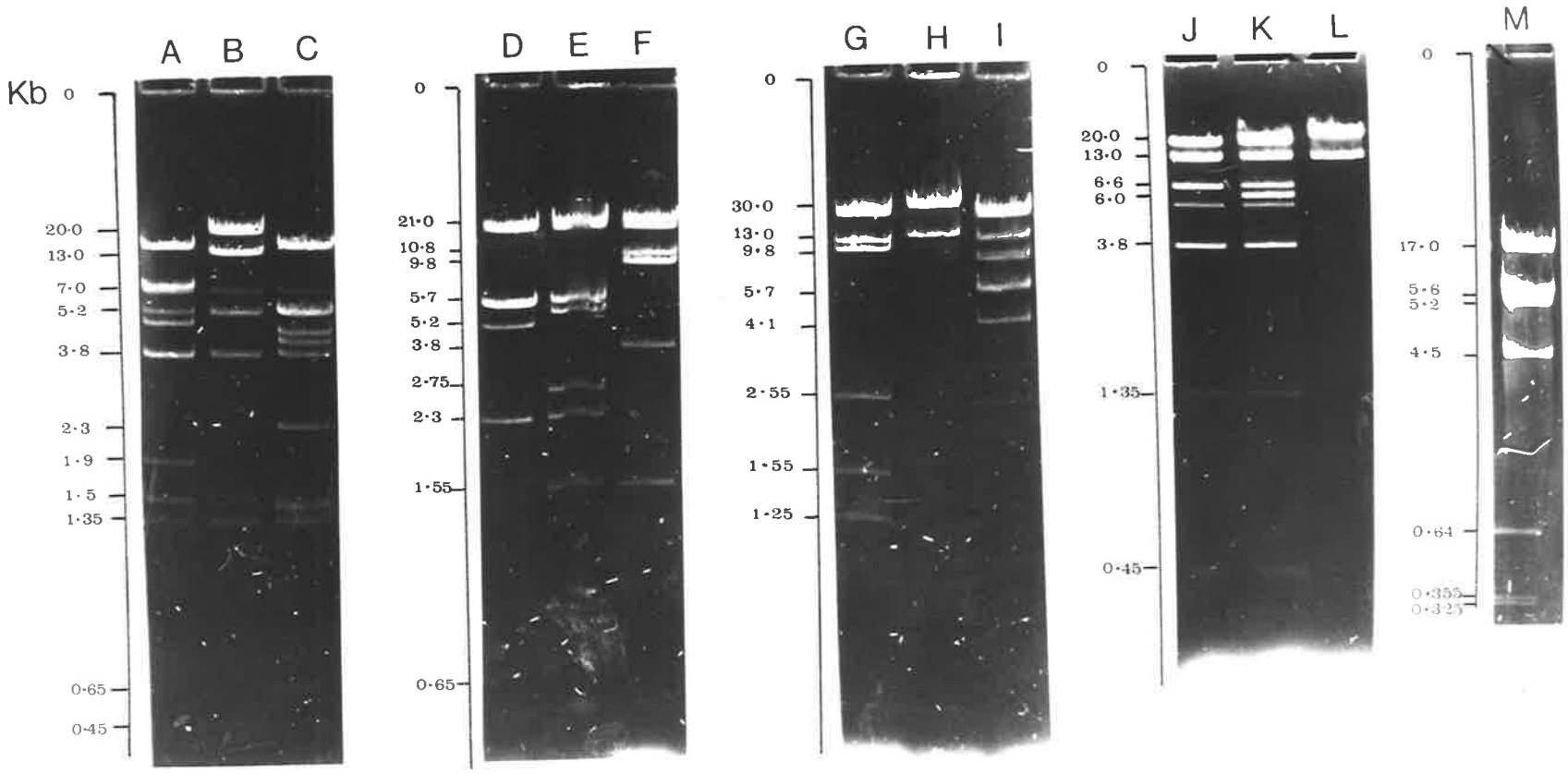
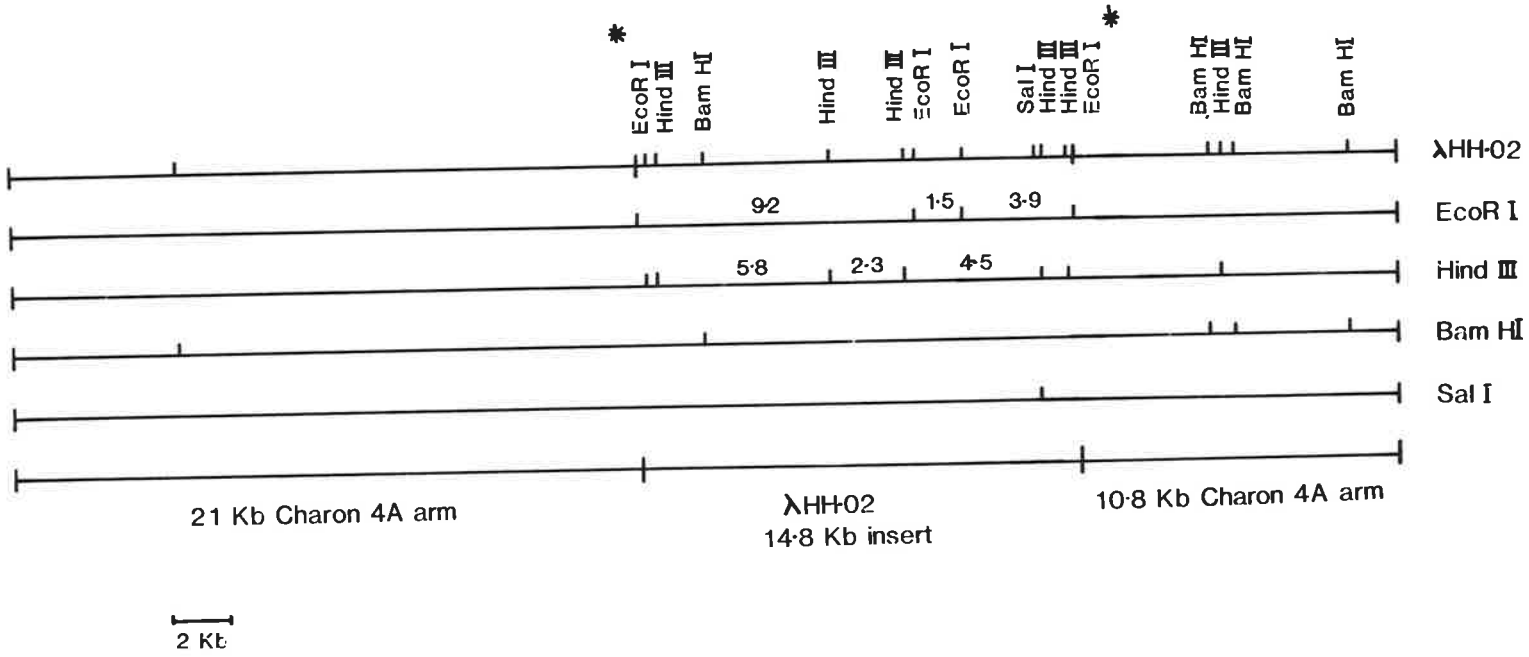

A B C
Kb 0
20·0
13·0
7·0
5·2
3·8
2·3
1·9
1·5
1·35
0·65
0·45
D E F
0
21·0
10·8
9·8
5·7
5·2
3·8
2·75
2·3
1·55
0·65
G H I
0
30·0
13·0
9·8
5·7
4·1
2·55
1·55
1·25
J K L
0
20·0
13·0
6·6
6·0
3·8
1·35
0·45
M
0
17·0
5·6
5·2
4·5
0·64
0·355
0·325

Fig. 6.6

Restriction Endonuclease Map of λHH.02

A restriction map is shown for λHH.02 insert and charon 4A vector arms. The distances are in kilobases.


EcoR I
Hind III
Bam HI
Hind III
Hind III
EcoR I
EcoR I
Sal I
Hind III
EcoR I
Bam HI
Hind III
Bam HI
Bam HI
λHH-02
9·2
1·5
3·9
EcoR I
5·8
2·3
4·5
Hind III
Bam HI
Sal I
21 Kb Charon 4A arm
λHH-02
14·8 Kb insert
10·8 Kb Charon 4A arm
2 Kb

6.2.7 Hybridisation Analysis of λHH.02

In order to locate the histone genes in λHH.02, a selection of restriction digestions of this clone were electrophoresed on a 1% agarose gel and the DNA fragments transferred by the Southern technique to nitrocellulose filters. These filters were hybridised with the individual chicken histone DNA probes. In agreement with the phage screening blots described in Section (6.2.5) only H3, H2A and H2B probes hybridised, and H4 and H1 histone genes were not detected.

Fig. (6.7) shows that chicken histone H2B probe hybridises to the 9.2 kb EcoR1 fragment, the 1.9 kb BamH1/EcoR1 fragment and to the 1.4 BamH1/HindIII fragment. Chicken histone H2A probe hybridises to these fragments also but in addition hybridises to a small HindIII fragment. Chicken histone H3, also hybridises to the same fragments as H2B and in addition hybridises to the 1.5 kb EcoR1 fragment.

These hybridisation results are summarised in Fig. (6.8) and demonstrate the presence of two H3 genes and one each of an H2A and H2B gene. The H2A, H2B and one H3 are clustered into a region of 1.9 kb separated by at least 7.2 kb from the other H3 gene. From the hybridisation data the H2A gene spans the 100 bp HindIII fragment at the end of the clone and the order of the adjacent H2B and H3 genes can be deduced from the hybridisation data to RsaI and HaeIII restriction fragments of p1.9BE described in Section (6.4.1).

Fig. 6.7

Hybridisation Analysis of λHH.02

λHH.02 DNA was restricted with the restriction enzymes shown (E=EcoRI, H-HindIII, BE=BamHI/EcoRI and BH=BamHI/HindIII) and electrophoresed on a 1% agarose gel. The DNA was transferred to nitrocellulose, hybridised with labelled chicken histone DNA probes (H3, H2B and H2A) (2.1.3) and autoradiographed at -80^{o}.

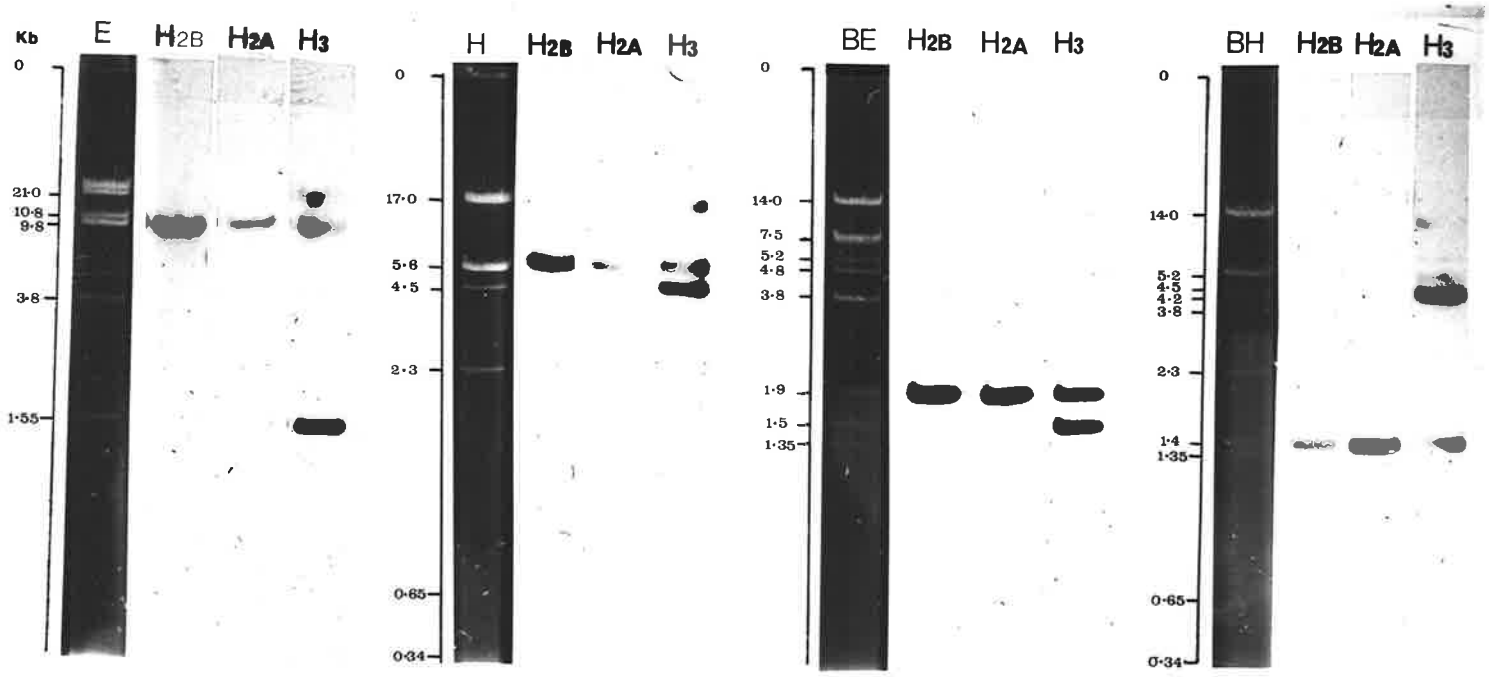
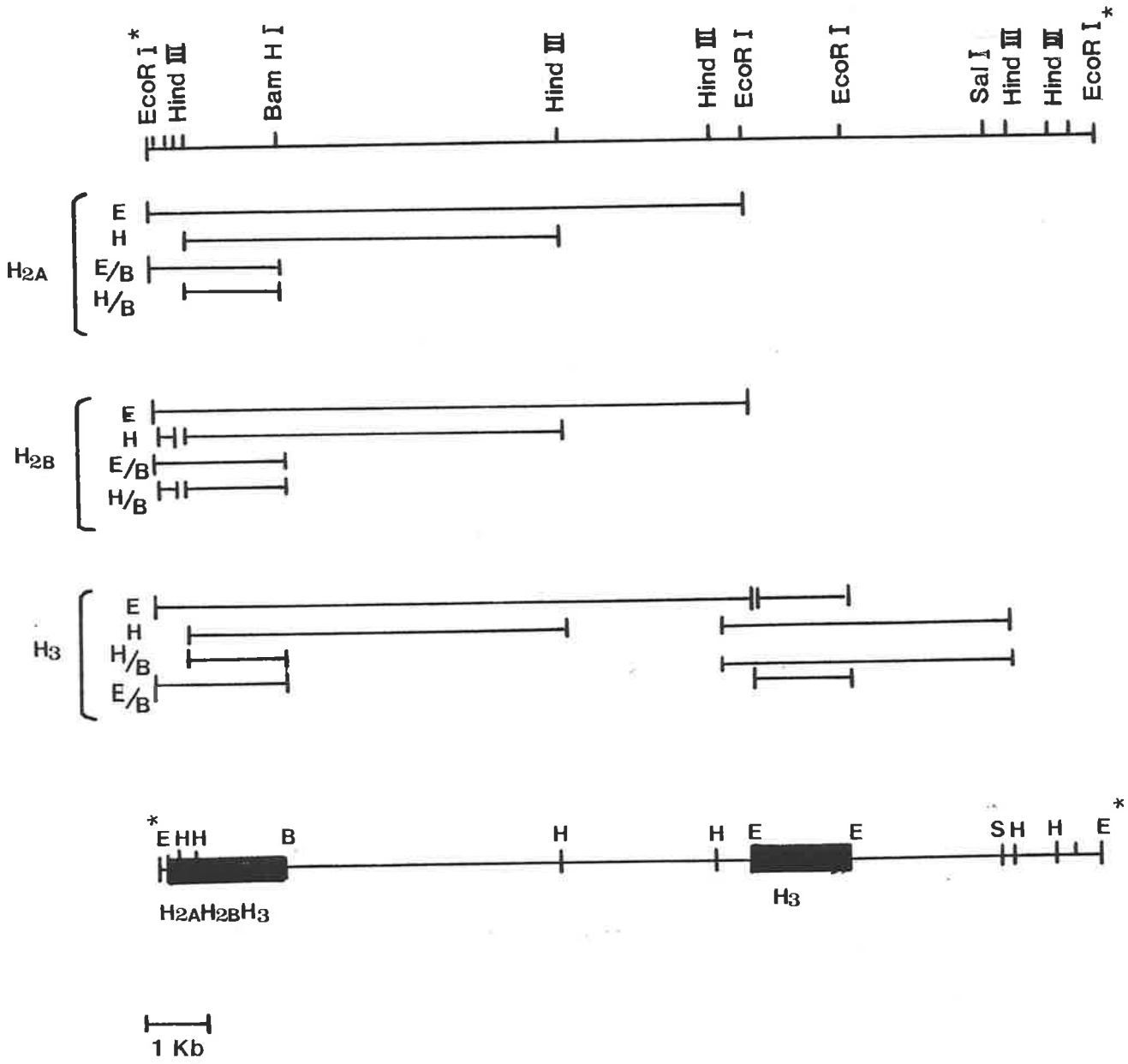

Kb
E
H2B
H2A
H3
0
21·0
10·8
9·8
3·8
1·55
H
H2B
H2A
H3
0
17·0
5·6
4·5
2·3
0·65
0·34
BE
H2B
H2A
H3
0
14·0
7·5
5·2
4·8
3·8
1·9
1·5
1·35
BH
H2B
H2A
H3
0
14·0
5·2
4·5
4·2
3·8
2·3
1·4
1·35
0·65
0·34

Fig. 6.8

Hybridisation Summary of λHH.02

A restriction enzyme map of the λHH.02 insert is shown. The restriction fragments which hybridised to the H2B, H2A and H3 chicken histone probes (Fig. 6.7) are indicated by the bars. The location of each histone gene is indicated by the blocked regions. E=EcoRI, H=HindIII, B-BamHI, S=SalI.


EcoR I*
Hind III
Bam H I
Hind III
Hind III
EcoR I
EcoR I
Sal I
Hind III
Hind III
EcoR I*
H2A
E
H
E/B
H/B
H2B
E
H
E/B
H/B
H3
E
H
H/B
E/B
*
E H H
B
H
H
E
E
S H
H
E
*
H2A H2B H3
H3
1 Kb

6.2.8 Restriction Analysis of λHH.03

A restriction map of λHH.03 was constructed using the enzymes EcoR1, HindIII and BamH1.

Digestion of λHH.03 with EcoR1 resulted in six insert fragments 6.8 kb, 3.3 kb, 2.6 kb, 2.35 kb, 1.9 kb and 1.7 kb (Track c, Fig. (6.9)). To order these fragments, λHH.03 was digested with BamH1/EcoR1. BamH1 cleaves the insert DNA three times, and the presence of six new double digestion products (Fig. (6.9), Track b) indicates that three of the EcoR1 fragments are cleaved once each by BamH1 as follows: The 6.8 kb EcoR1 fragment to 3.8 kb and 3.0 kbs; the 3.3 kb EcoR1 fragment to 1.8 kb and 1.6 kbs and 1.9 kb EcoR1 to 1.4 kb and 600 bps.

Digestion of λHH.03 with HindIII also demonstrated the presence of three sites within the insert. The 2.6 kb HindIII fragment is cut by neither BamH1 nor EcoR1 allowing its placement within the 6.8 kb EcoR1 fragment and the 3.3 kb BamH1 fragment. The other HindIII site is placed by the double digestions within the 3.3 kb EcoR1 fragment and the 2.0 kb BamH1 fragment.

Determination of the relative positions of these fragments within λHH.03 is completed by consideration of the fragments overlapping the insert sites giving the restriction map shown in Fig. (6.10).

Fig. 6.9

Restriction Enzyme Digests of λHH.03

DNA (1 μg) from λHH.03 was digested with EcoRI, BamHI and HindIII and electrophoresed on a 1% agarose gel. λDNA cut with EcoRI and HindIII was used as markers. The tracks are as follows:

A: BamHI

B: BamHI/EcoRI

C: EcoRI

D: EcoRI/HindIII

E: HindIII

F: HindIII/BamHI

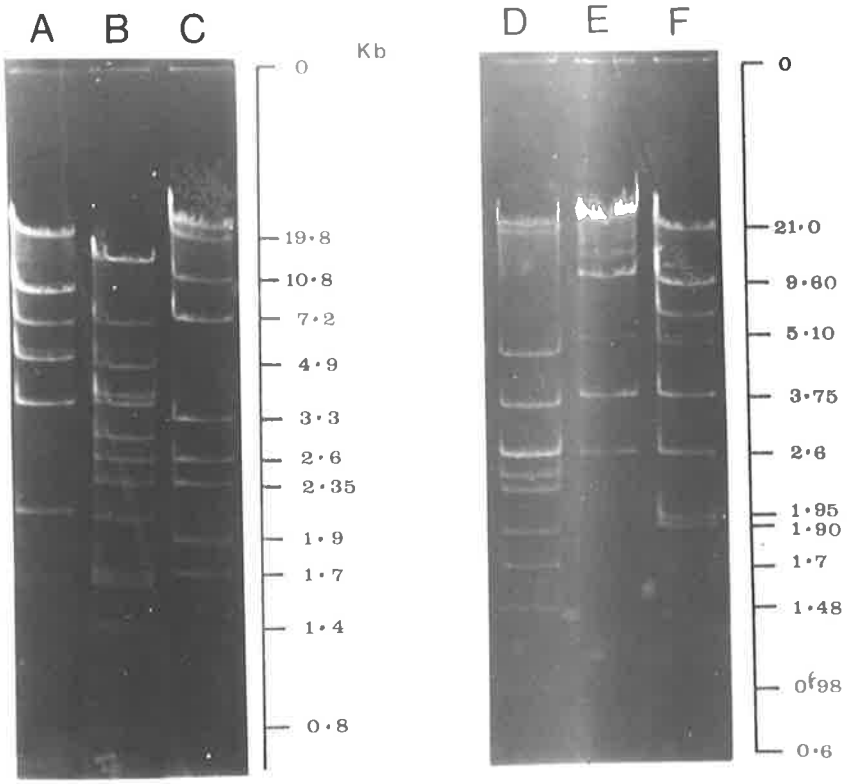
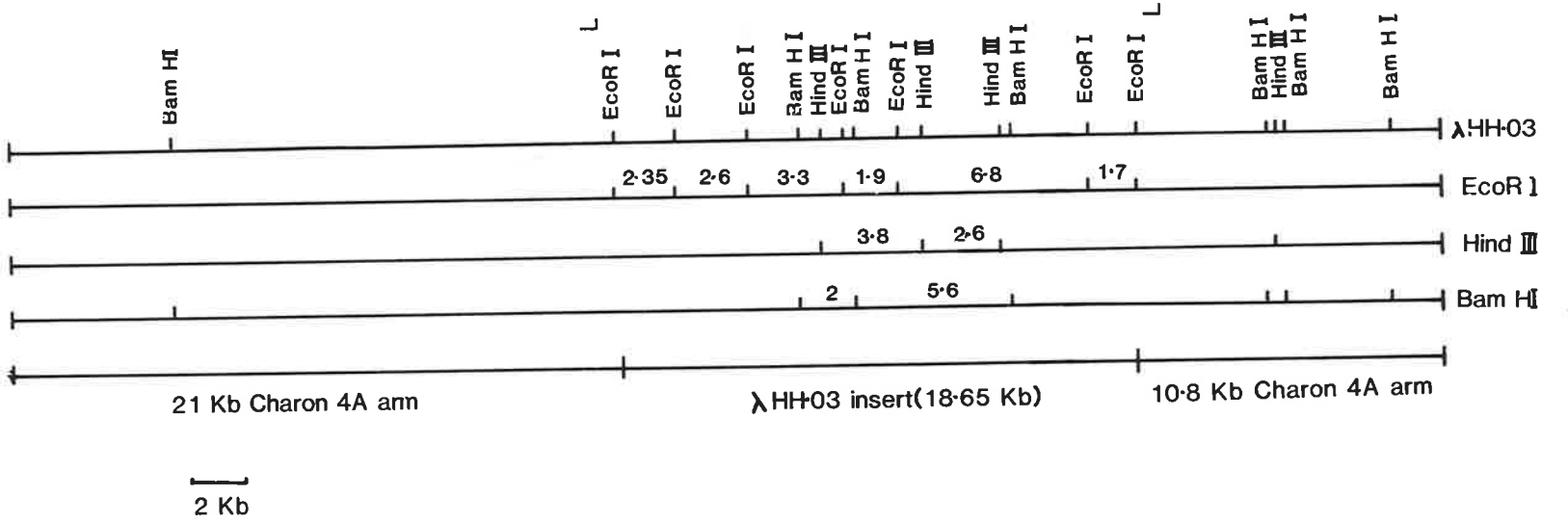

A B C
Kb
0
19·8
10·8
7·2
4·9
3·3
2·6
2·35
1·9
1·7
1·4
0·8
D E F
0
21·0
9·60
5·10
3·75
2·6
1·95
1·90
1·7
1·48
0·98
0·6

Fig. 6.10

Restriction Endonuclease Map of λHH.03

A restriction map of λHH.03, including the insert and the λ Charon 4A arms is shown. The map distances are in kilobases.


Bam HI
EcoR I
EcoR I
EcoR I
Bam HI
Hind III
EcoR I
Bam HI
EcoR I
Hind III
Hind III
Bam HI
EcoR I
EcoR I
Bam HI
Hind III
Bam HI
Bam HI
λHH-03
2·35
2·6
3·3
1·9
6·8
1·7
EcoR I
3·8
2·6
Hind III
2
5·6
Bam HI
21 Kb Charon 4A arm
λHH-03 insert(18·65 Kb)
10·8 Kb Charon 4A arm
2 Kb

6.2.9 Hybridisation Analysis of λHH.03

In order to locate the histone genes in λHH.03, restriction digests of this clone were electrophoresed on a 1% agarose gel and the DNA fragments were transferred to nitrocellulose filters. These filters were hybridised with individual chicken histone DNA probes, however, only H2A and H2B probes were found to hybridise, as with the phage patch hybridisations. Therefore λHH.03 does not contain genes for H3, H4 or H1 histones.

Fig. (6.11) shows that the H2A gene probe hybridised to the 2.6 kb EcoR1 fragment, the 2.8 kb HindIII fragment and the 20 kb BamH1 fragment. The chicken H2B gene probe also hybridises to these fragments, as well as the 1.9 kb EcoR1 fragment, the 3.3 kb HindIII fragment, the 1.95 kb Bam/HindIII fragment, the 5.0 kb BamH1 fragment and the 1.4 kb Bam/EcoR1 fragment.

As summarised in Fig. (6.12), there are thus two H2B histone genes in λHH.03 separated by at least 3.8 kb, with an H2A gene adjacent to one of the H2B genes.

6.2.10 Genomal Organisation of Human Histone Genes

Since both λHH.02 and λHH.03 contained a different restriction pattern and a different arrangement of histone genes, it appeared that either these two clones were not representative of the main cluster of human histone genes or more likely that the general organisation of the human histone genes is not as a tandem repeating unit.

Fig. 6.11

Hybridisation Analysis of λHH.03

Restriction enzyme digests of λHH.03 (E=EcoRI, BE=BamHI plus EcoRI, B=BamHI, H=HindIII and HB=HindIII plus BamHI) were electrophoresed on a 1% agarose gel. The DNA was transferred to nitrocellulose, hybridised with chicken histone DNA probes (H2A, H2B) and autoradiographed at -80^{o}.

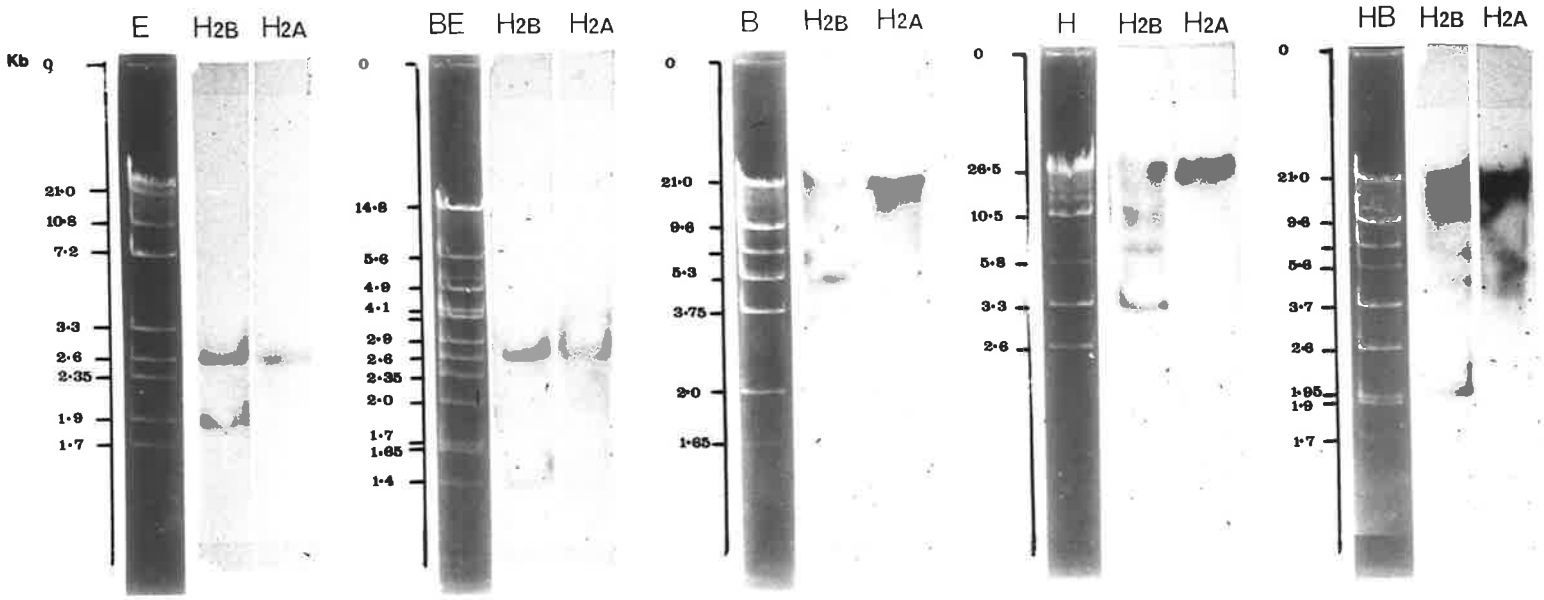
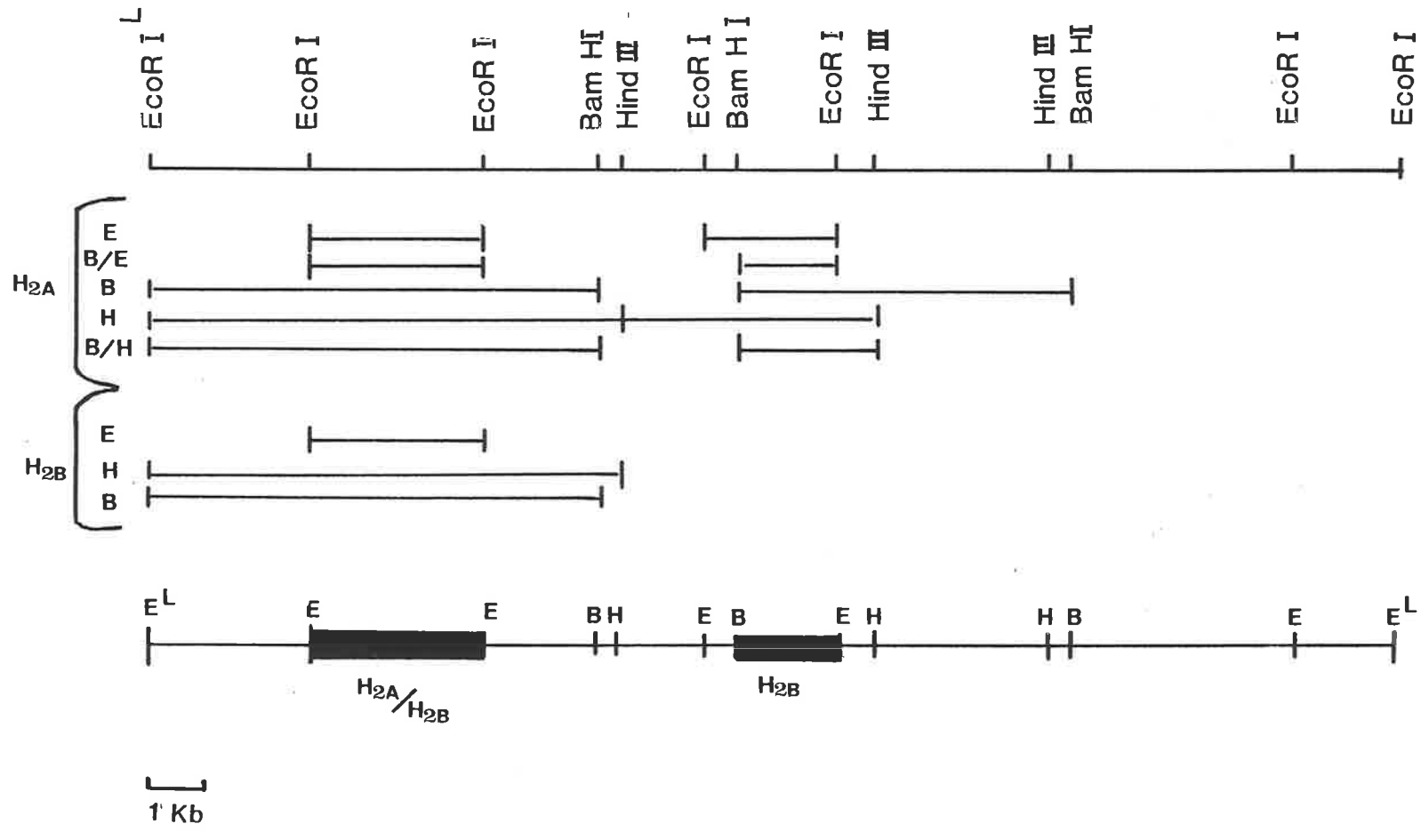

E
H2B
H2A
Kb
0
21·0
10·8
7·2
3·3
2·6
2·35
1·9
1·7
BE
H2B
H2A
0
14·8
5·6
4·9
4·1
2·9
2·6
2·35
2·0
1·7
1·65
1·4
B
H2B
H2A
0
21·0
9·6
5·3
3·75
2·0
1·65
H
H2B
H2A
0
26·5
10·5
5·8
3·3
2·6
HB
H2B
H2A
0
21·0
9·6
5·6
3·7
2·6
1·95
1·9
1·7

Fig. 6.12

Hybridisation Summary of λHH.03

The restriction enzyme map of λHH.03 insert is shown. The restriction fragments which hybridised to the H2B and H2A chicken histone DNA probes (Fig. 6.11) are indicated by the bars. The location of the histone genes is shown by the blocked regions. E=EcoRI, H=HindIII, B=BamHI.


EcoR I
EcoR I
EcoR I
Bam HI
Hind III
EcoR I
Bam H I
EcoR I
Hind III
Hind III
Bam HI
EcoR I
EcoR I
H2A
E
B/E
B
H
B/H
H2B
E
H
B
E B/H E B H B/H E B H B E
H2A/H2B
H2B
1 Kb

To study the gross organisation of the human histone genes, human genomal DNA was digested with EcoR1, electrophoresed on a 1% agarose gel, transferred to nitrocellulose, and hybridised with two chicken genomal clones pCH4.8E and pCH2.6H from λCH.02 which contain H2B and H1 histone genes and H3 and H4 histone genes respectively. These probes Fig. (6.13) shows that numerous EcoR1 fragments are detected with both common and unique fragments hybridising to the H3/H4 and H1/H2B histone gene probes.

To ensure that the hybridisation pattern observed was not due to spacer cross-reaction in the genomal probes, EcoR1 digested human DNA was also hybridised with chicken histone DNA containing only coding sequences, as described in (2.1.3). The blots in Fig. (6.13) show human EcoR1 restriction fragments which hybridise to H2A, H2B and H3 histone gene probes. Again, hybridisation to a number of restriction fragments was observed for each individual histone gene, demonstrating that the human histone genes are not arrayed in a repeating pattern with respect to EcoR1 restriction sites.

A number of fragments appear to be common for some of the histones; this is consistent with clustering, however, the appearance of common size fragments does not necessarily mean that these histone genes are on the same fragment. Approximately 12-15 fragments hybridise with each histone gene probe, one or two of which appear to contain more than one gene. It is therefore likely that there are 16-18 copies of each of the core histone genes in the human genome. This is approximately

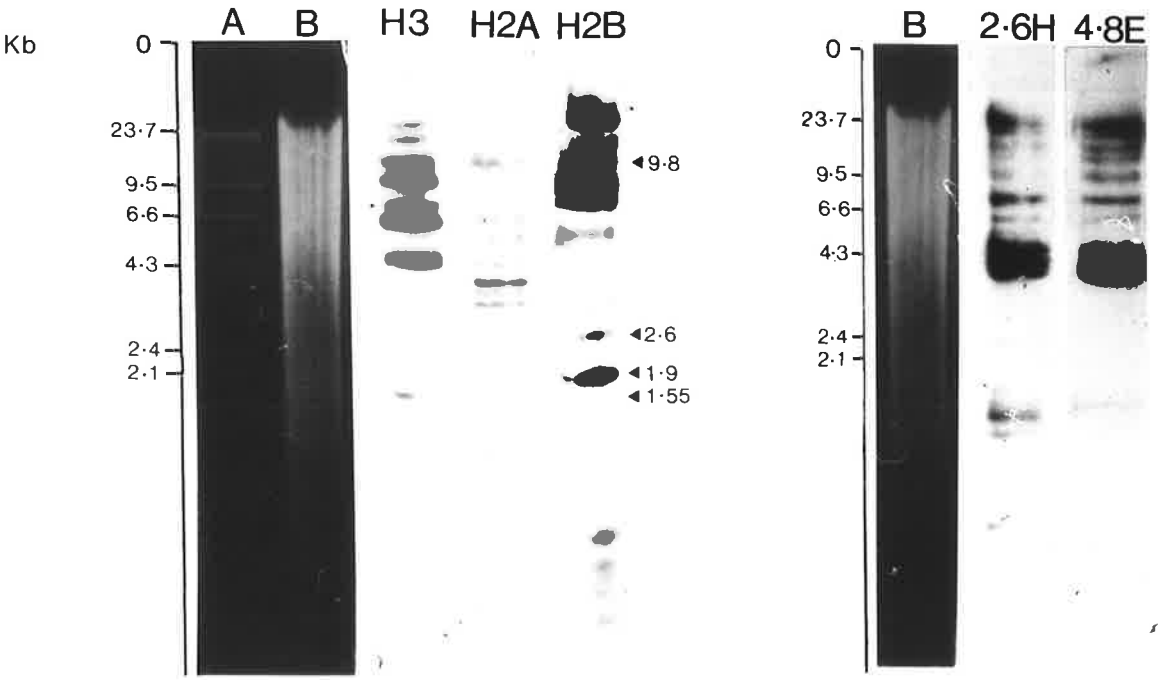
Fig. 6.13

Human Genomal Blots

Human genomal DNA (10 µg) was digested with EcoR1, electrophoresed on a 1% agarose gel, transferred to nitrocellulose, hybridised with individual chicken histone DNA probes, and autoradiographed. EcoR1/HindIII digested DNA was used as markers. The tracks are as follows:

A: λ DNA, HindIII cut
B: Human DNA cut with EcoR1

H3, H2A, H2B are tracks probed with individual gene probes (Section 2.1.3) and 2.6H and 4.8E tracks probed with the λCH.02 subclones pCH2.6H and pCH4.8E.


Kb
A
B
H3
H2A
H2B
0
23·7
9·5
6·6
4·3
2·4
2·1
9·8
2·6
1·9
1·55
B
2·6H
4·8E
0
23·7
9·5
6·6
4·3
2·4
2·1

half the reiteration frequency of 30-40 fold estimated by Wilson and Melli (1977) using Cot analysis and sea urchin histone genes as probes.

Three coding EcoR1 fragments are internal to the clones λHH.02 and λHH.03 and should therefore be present in the genomal blots. The 1.5 kb EcoR1 fragment detected in the genomal H3 blot corresponds to the 1.5 kb EcoR1 fragment in λHH.02, which contains a single H3 gene. The 2.6 kb EcoR1 fragment, which is common to both the H2A and H2B genomal blots corresponds to the 2.6 kb EcoR fragment of λHH.03 and the 1.9 kb EcoR1 fragment from λHH.03, containing only a H2B gene, also probably corresponds to the EcoR1 fragment shown in Fig. (6.13).

6.3 DISCUSSION

Two human genomal clones containing histone genes have been isolated. λHH.02 was selected from the EcoR1* human library which was prepared from human placental DNA and λHH.03 was selected from the Maniatis Human Library which was constructed from human fetal liver DNA (Lawn *et al.*, 1978).

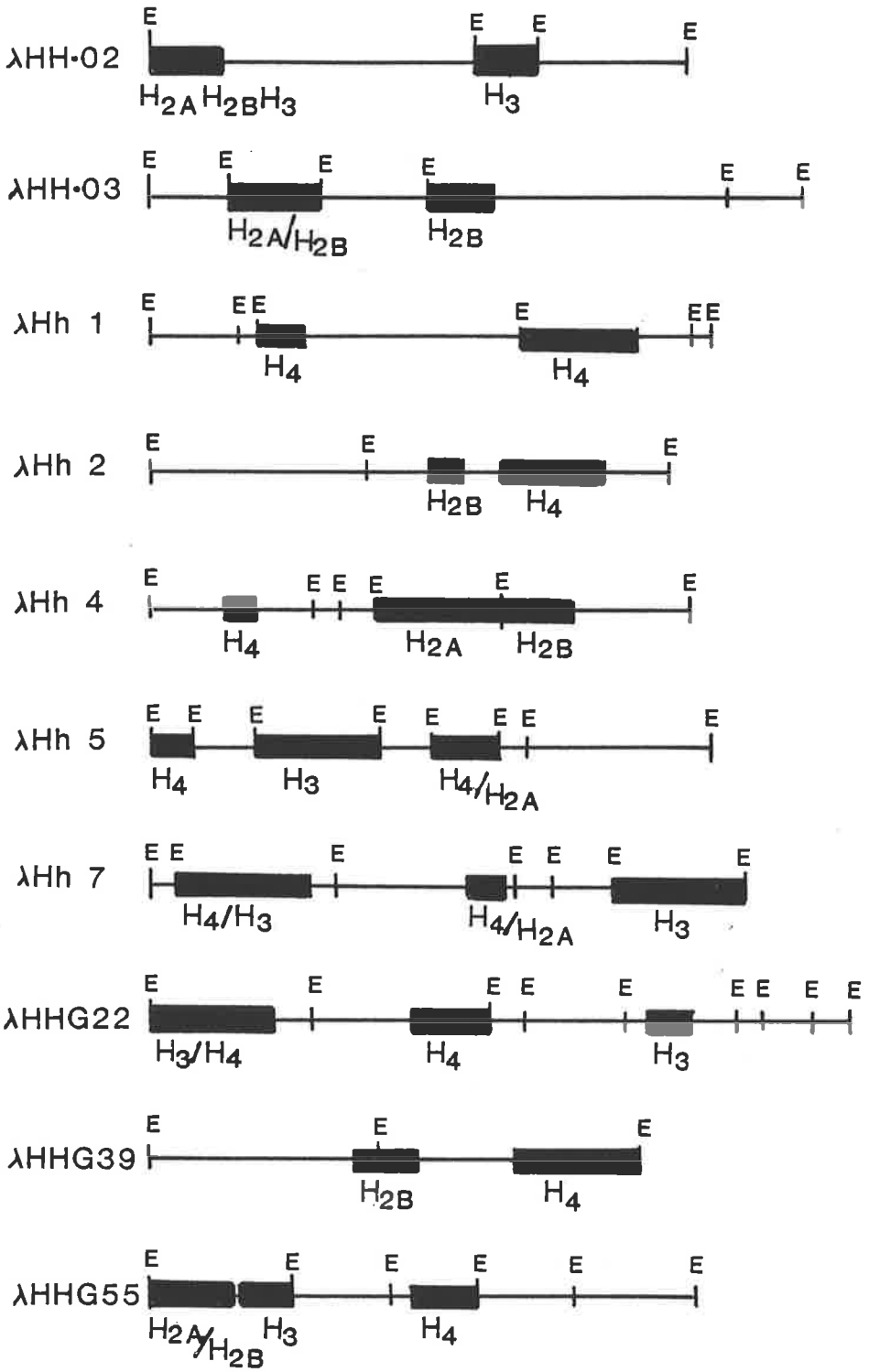
λHH.02 and λHH.03 have different restriction enzyme maps and therefore represent different regions of the human genome that contain histone genes. They are also different from the other human histone clones isolated from the Maniatis Library by Heintz *et al.*, (1981) and Sierra *et al.* (1982a) and all of these clones are compared in Fig. (6.14). λHH.02 contains two H3 genes, an H2A gene and an H2B gene. One of the H3 genes is clustered into 1.9 kb with the H2A and H2B genes, and these are separated

Fig. 6.14

Human Histone Clones

Summarised are EcoRI(E) restriction maps and histone gene locations of human histone genomal clones: λHH.02, λHH.03 (Fig. 6.8 , 6.12) λHh1, λHh2, λHh4, λHh5, λHh7 (Heintz *et al.*, 1981) λHHG22, λHHG39, λHHG55 (Sierra *et al.*,1982a). There are no HI genes in λHH.02 and λHH.03, however the location of HI genes in the other human clones is not known.

HUMAN HISTONE CLONES



by 7.2 kb from the other H3 gene. A further 4 kb of spacer extends from this isolated H3 gene to the end of the clone. λHH.03 contains two H2B genes separated at least 3.7 kb and an H2A gene. This clone spans a further 8.0 kb which contains no histone genes.

When the organisation of the human histone genes from other recombinant clones isolated from the Maniatis Human Library are compared with λHH.02 and λHH.03 (Fig. 6.14), a number of features are observed.

Firstly, there is no conservation of restriction sites surrounding the genes in any of the clones and in addition there appears to be no recognisable repeat structure in the ordering of the histone genes. For example, the 12 H4 genes occurring in all these clones are surrounded by different genes in every case. The lack of a repeating unit is also apparent from the genomal blots.

Secondly, the spacing between the histone genes may vary by several kilobases, e.g. in λHH.02 there are three genes (H2A, H2B, H3) within 1.9 kb, separated by 7 kb to an isolated H3 gene, and in λHh1 there are two single H4 genes separated by 6 kb from each other and at least 3 kb from any other histone gene.

However, there is more than one histone gene on each genomic clone which indicates that these genes are adjacent to each other in the human genome. Moreover, the apparent lack of an entire complement of histone genes in any of these 10 clones (average length of 15 kb, i.e. spanning 150 kb of genomal DNA) suggests that the histone genes are organised into a very disperse cluster and appear to lack a simple repeating unit.

There is as yet no evidence that any of these 10 clones shown in Fig. (6.15) are overlapping, and the number of genes which they contain (8 H3, 6 H2A, 7 H2B and 12 H4 genes) indicates they represent about half of the complement of human histone genes, as determined by the genomal blots. It is possible however that some of these clones could represent isolated orphon genes separated from the main cluster, as reported in sea urchin and *Drosophilia* (Childs *et al.*, 1981) although this does seem unlikely as there are much fewer copies of histone genes in the human genome.

Another notable feature of the histone gene organisation is the predominance of H2A/H2B gene pairs found in λHH.02, λHH.03, λHh4, λHHG55 as well as H3/H4 gene pairs found in λHh7 and λHHG22. This feature is similar to the H3/H4 and H2A/H2B gene pairing found in yeast histone gene organisation (Hereford *et al.*, 1979) and observed in chicken histone genes, e.g. λCH.01 and λCH.02. However, all these genes are also found either unpaired or paired with a different gene, e.g. H4/H2A in λHh5 and λHh7. It is possible however that the pairing of histone genes, in particular H2A/H2B and H3/H4 combinations, may be significant in the regulation of their expression during development.

It is unlikely that the heterogeneity exhibited by the organisation of the human histone genes results from alterations due to cloning or amplication in bacteria. It is also unlikely that the heterogeneity arises from polymorphisms in the human population, because all clones except λHH.02 are derived from the one individual and these display a wide degree of variation. The lack of repeating units within these clones therefore probably reflects the *in vivo* structure of the human histone genes accurately, and this conclusion is also supported by the genomal DNA blot data.

Consequently it appears that the human histone genes, as well as other higher eukaryotes, e.g. chicken (Chapter 4; Harvey *et al.*, 1981; Engel and Dodgson, 1981) and mouse (Sittman *et al.*, 1981), do not fit the simple pattern or organisation observed in sea urchin and *Drosophila* in which tandemly repeated clusters contain one each of the five histone genes.

PART B STRUCTURE OF THE HUMAN HISTONE GENES

To date only limited protein sequence data is available for human histones, only the sequence of the H2B protein having been published (Ohe *et al.*, 1979). Sequence data on human histone genes is therefore of considerable interest with respect to the proteins themselves, as well as for consideration of their control regions.

6.4 RESULTS

6.4.1 Sequencing Strategy for λHH.02

The 1.9 kb BamH1/EcoR1 fragment in λHH.02 was shown by hybridisation, using chicken histone DNA probes, to contain H3, H2A and H2B histone genes. In order to study these genes in more detail, this fragment, after elution from low melting point agarose, was subcloned into pBR325 digested with BamH1 and EcoR1. The subclone was termed pHH1.9BE.

To locate the coding regions within pHH1.9BE, the subclone was digested with RsaI and HaeIII, insert fragments (as shown in Fig. (6.15) were excised, eluted, denatured and spotted onto nitrocellulose. The nitrocellulose was then hybridised with chicken histone H2B, H2A and H3 gene probes. Fig. (6.15) shows which RsaI and HaeIII fragments from pHH1.9BE contain H2B and H2A gene sequences. None of the fragments excised contained H3 histone gene complementary sequences (on further investigation the H3 gene was located on the hybrid insert-vector fragment overlapping the BamH1 site).

Fig. 6.15

Coding Fragments of pHH1.9BE

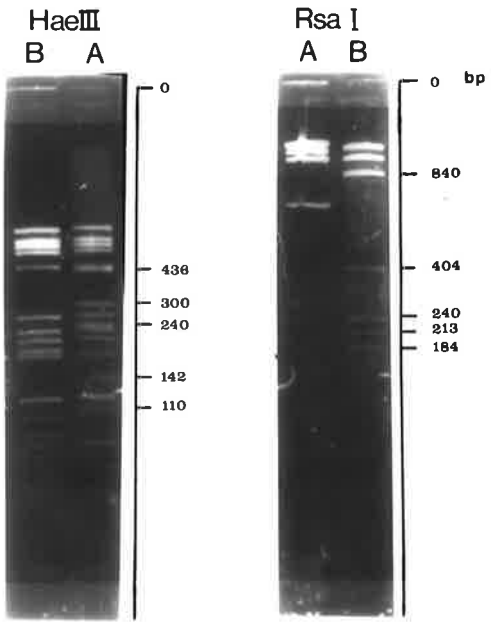
DNA from pHH1.9BE was digested with HaeIII and RsaI and electrophoresed on a 6% acrylamide gel. The insert fragments indicated, were eluted, denatured, and spotted onto nitrocellulose filters which were hybridised with the H2B, H3 and H2A chicken histone probes. λCH.01 was used as a control, since it contains all three histone genes.

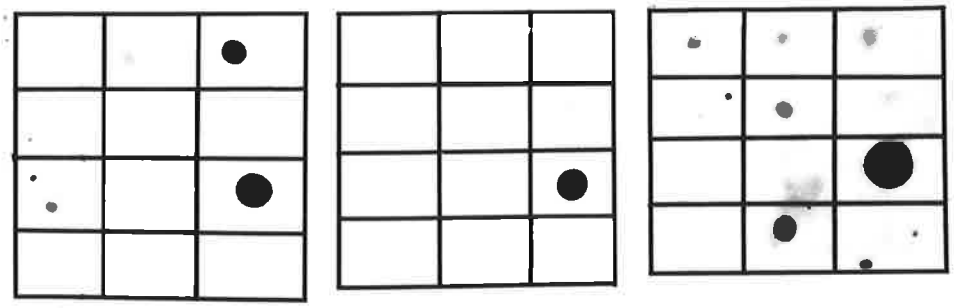
Track A: pBR325, Track B: pHH1.9BE

The DNA spots are arranged as follows:

840 bp RsaI	184 bp RsaI	140 bp HaeIII
404 bp RsaI	436 bp HaeIII	110 bp HaeIII
240 bp RsaI	300 bp HaeIII	λCH.01
213 bp RsaI	240 bp HaeIII	

Hybridisation of both H2A and H2B probes were observed to some fragments and this appeared to be due to some sequence cross reaction between the two genes or their untranslated regions.





H2B H3 H2A

The RsaI and HaeIII fragments containing coding regions were ligated into the SmaI site of the phage M13 vector, mp83, and transformed into JM101. Single stranded DNA was prepared from the resulting clear plaques and used for sequencing by the dideoxy-nucleotide method of Sanger.

In addition the small HindIII fragments located in pHH1.9BE, which had been previously shown to hybridise with chicken histone H2A probe, were end-labelled using α-{^{32}P}-dATP and strand separated on an acrylamide gel. The {^{32}P}-labelled single-stranded DNA was used in Maxam and Gilbert sequencing reactions. The small HindIII fragments were also cloned into the HindIII site of M13, mp83, and used for sequencing by the dideoxynucleotide method (Fig. 6.20a).

Consideration of the digestion products of pHH1.9BE and the results of hybridisation with gene specific probes allows the ordering of the genes shown in Fig. (6.16). Thus, the H2A gene covers the region of the HindIII sites at the EcoR1 end of the subclone. The H2B gene and the H3 gene are located between the HindIII site and the BamH1 site, with the H3 gene adjacent to the BamH1 site since it is present in the large RsaI and HaeIII fragments containing pBR322 sequences from pHH1.9BE.

6.4.2 Sequencing Strategy for λHH.03

The 1.9 kb EcoR1 fragment in λHH.03 had been shown by hybridisation to contain an H2B gene. Therefore, to obtain sequence data for this gene, the EcoR1 fragment was isolated from a low melting point agarose gel and cloned into pBR325. This sub-clone was called pHH1.9E.

Fig. 6.16

Histone Gene Analysis of pHH1.9BE and pHH1.9E

Restriction maps are shown for the DNA inserts of pHH1.9BE (λHH.02) and pHH1.9E (λHH.03). The DNA fragments sequenced are indicated by the lines. The crossed regions indication the location of the histone genes from both hybridisation and sequence data and the open region indicates the location of the histone H3 gene deduced from hybridisation data. The arrows indicate direction of transcription.

EcoR I
BamH I
FnuD II
Rsa I
FnuD II
FnuD II
Bal I
Rsa I
EcoR I
H2B
200 bp
pHH 1·9E

EcoR I
Hind III
Rsa I
Hind III
Rsa I
Hind III
Hae III
Hae III
Bam HI
H2A
H2B
H3
200 bp
pHH 1·9BE

To locate the H2B gene within pHH1.9E, the subclone was digested with RsaI, BstNI and FnuDII and electrophoresed on a 1% agarose gel, Fig. (6.17). The insert fragments were excised, eluted and spotted onto nitrocellulose, which was then hybridised with chicken histone H2B probe. Fig. (6.17) shows which RsaI, BstNI and FnuDII fragments contain H2B coding sequences.

The restriction fragments containing H2B coding sequences were ligated into the SmaI site of the phage M13 mp83, and transformed into JM101. Single-stranded DNA was prepared from the clear plaques for sequencing.

6.4.3 Sequence Analysis of the Human Histone Genes in λHH.02 and λHH.03

The DNA regions sequenced in the λHH.02 subclone pHH1.9BE and the λHH.03 subclone pHH1.9E, are summarised in Fig. (6.16). Sequence data was obtained for the H2A and H2B genes located in λHH.02 and for one of the H2B genes in λHH.03 Fig. (6.19, 6.20). Due to the method employed in selecting DNA fragments for sequencing, by their cross-reaction with chicken histone gene probes, the fragments sequenced were biased in that they contained only protein coding regions. Further mapping and sequencing of individual fragments will be required to obtain flanking regions of the genes.

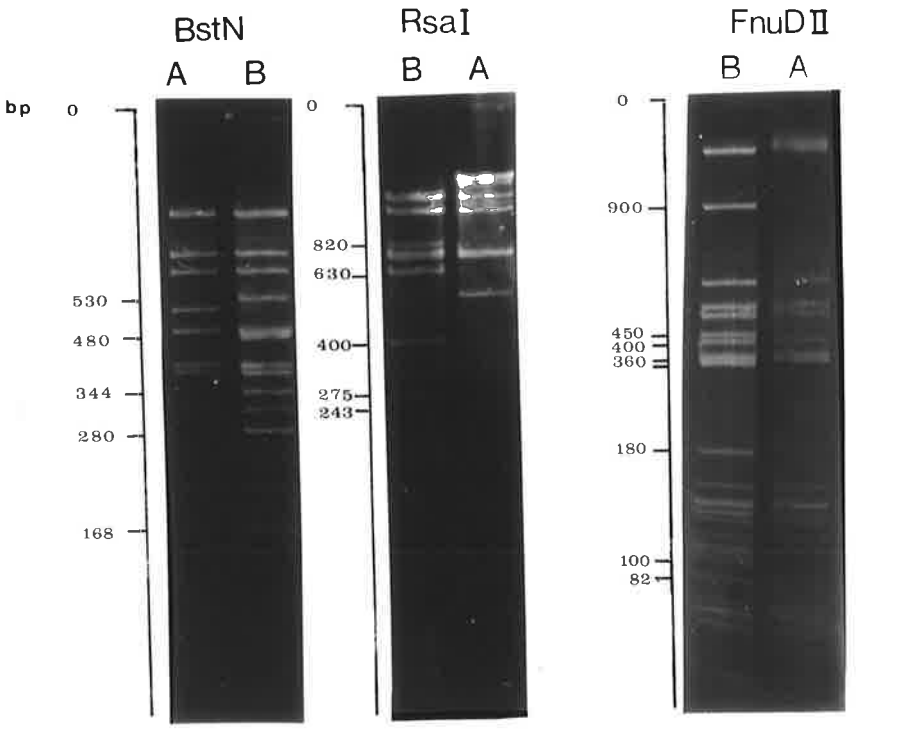
Fig. 6.17

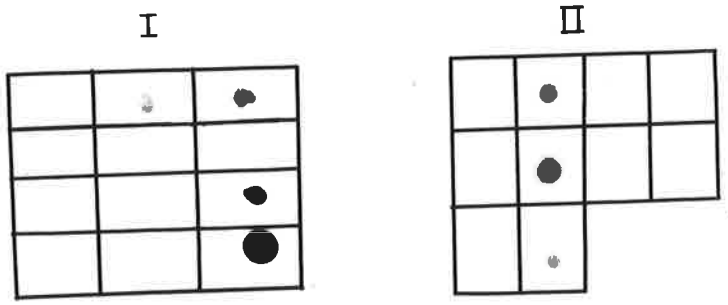
Coding Fragments of pHH1.9E

DNA from pHH1.9E was digested with BstNI, RsaI and FnuDII and electrophoresed on a 6% acrylamide gel. The insert fragments shown were eluted, denatured and spotted onto nitrocellulose filters which were hybridised with chicken histone H2B DNA probe. λCH.02 and pHH1.9BE DNA were used as controls.

The order of the spots are as follows:

I	530 bp BstNI	168 bp BstNI	275 RsaI	
	480 bp BstNI	820 bp RsaI	243 RsaI	
	344 bp BstNI	630 bp RsaI	pHH1.9BE	
	280 bp BstNI	400 bp RsaI	λCH.02	
II	900 bp FnuDII	450 bp FnuDII	400 bp FnuDII	360 bp FnuDII
	180 bp FnuDII	100 bp FnuDII	82 bp FnuDII	
		pHH1.9BE		


BstN
A B
bp 0
530
480
344
280
168
RsaI
B A
0
820
630
400
275
243
FnuDII
B A
0
900
450
400
360
180
100
82


I
II

6.4.4. Histone Gene Assignments

Since the histone gene assignments for λHH.02 and λHH.03 relied on hybridisation with heterologous chicken histone gene probes, it was important to ensure that this hybridisation, (especially for the H2A and H2B genes, which are less conserved) was indeed specific, since the cross-hybridisation between sea urchin and chicken histone genes had been misleading (discussed in Chapter 4).

The homology however between chicken and human H2B genes was found to be sufficient to allow specific gene detection, since the 140 bp HaeIII fragment from pHH1.9BE and the 243 bp RsaI fragment from pHH1.9E, which both hybridised strongly to chicken histone H2B (Fig. 6.15, 17), contained H2B gene sequences (Fig. 6.19) and the 184 bp RsaI fragment from pHH1.9BE which hybridised to chicken histone H2A (Fig. 6.15) was found to code for a human H2A gene (Fig. 6.20). Therefore the histone gene assignments in λHH.02 and λHH.03 would appear to be accurate.

6.4.5 Direction of Transcription

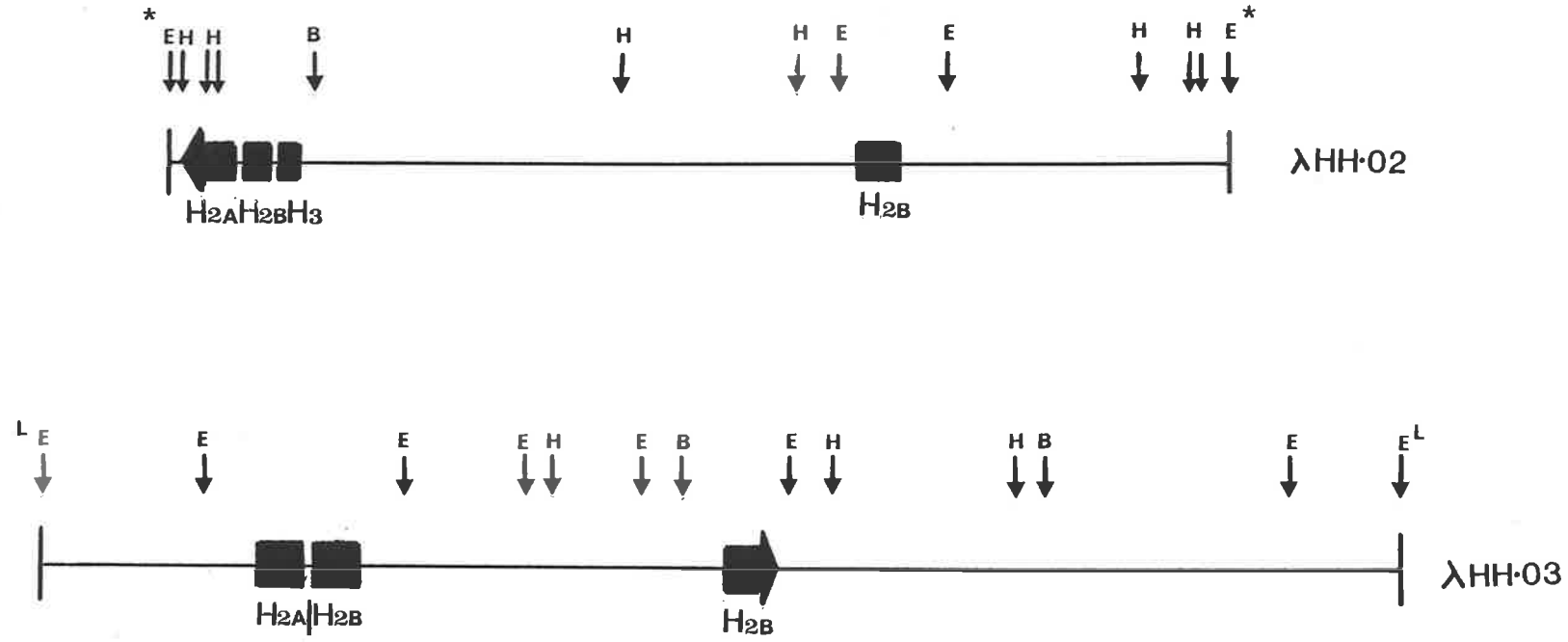
The orientation of the H2A gene in λHH.02 was deduced from the H2A gene sequence which spanned two HindIII sites in pHH1.9BE, since the amino terminal end of the protein was located in the HindIII/BamHI fragment the H2A gene must be transcribed away from this fragment (Fig. 6.16).

The orientation of the H2B gene in λHH.03 was determined by sequencing through the single BalI site which was mapped in pHH1.9E Fig. (6.16). Fig. (6.18) summarises the location and direction of these genes in λHH.02 and λHH.03.

Fig. 6.18

Histone Gene Content of λHH.02 and λHH.03

The restriction maps for the λHH.02 and λHH.03 insert DNAs are shown. E=EcoRI, H=HindIII, B=BamHI, E^{*} and E^{L} indicate the boundaries of the insert. The closed regions indicate the positions of the histone genes and the arrows indicate the direction of transcription.


*
EH H
B
H
H E
E
H H E*
λHH·02
H2A H2B H3
H2B
L E
E
E
E H
E B
E H
H B
E
E L
λHH·03
H2A H2B
H2B

It would be interesting to determine the direction of the other genes in these clones to see if their direction of transcription is in the same or opposite orientation, especially the H2A/H2B pairs which occur in both clones. In the chicken λCH.01 genomal clone, an H2A/H2B gene pair is divergently transcribed (Harvey, unpublished data) and a similar organisation is seen in yeast (Hereford *et al.*, 1979) and *Drosophila* (Lifton *et al.*, 1977). The direction of transcription of the H2A/H2B gene pairs found in other human genomal clones (Heintz *et al.*, 1981; Sierra *et al.*, 1982a) has not yet been determined. However, the role of histone gene pairs in expression and the significance of their orientation is still unclear.

6.4.6 Protein Sequences Encoded by the Human Histone Genes of λHH.02 and λHH.03

From the nucleotide sequences determined for the H2A gene of λHH.02 and the H2B genes of λHH.02 and λHH.03, it is possible to examine the deduced sequences of the histone proteins for which they code and to compare these with known protein sequences of chicken and other species.

(a) H2B Histone Protein Sequence from λHH.02

From the 140 bp HaeIII fragments of pHH1.9BE, a nucleotide sequence was obtained, which corresponded to the amino acid sequence from residues 22-57 of the human histone H2B protein sequence determined from spleen. (Ohe *et al.*, 1979) Fig. (6.19c). However, one amino acid change was detected from that of human spleen H2B protein at residue 39. Microheterogeneity is found at

Fig. 6.19(a)

Sequencing Gel of a H2B Gene of λHH.02

The sequence is shown of part of the 140 bp HaeIII fragment of pHH1.9BE which had been cloned into the SmaI site of M13 mp83 and sequenced by the dideoxynucleotide technique (Section 2.2.16b). The sequence of the opposite strand corresponds to the mRNA sequence (Fig. 6.19(c)). The arrow indicates the vector/insert junction. The underlined A's are residues where cross-banding was apparent in these sequencing reactions. A few "ghost" bands are also evident in the T track.

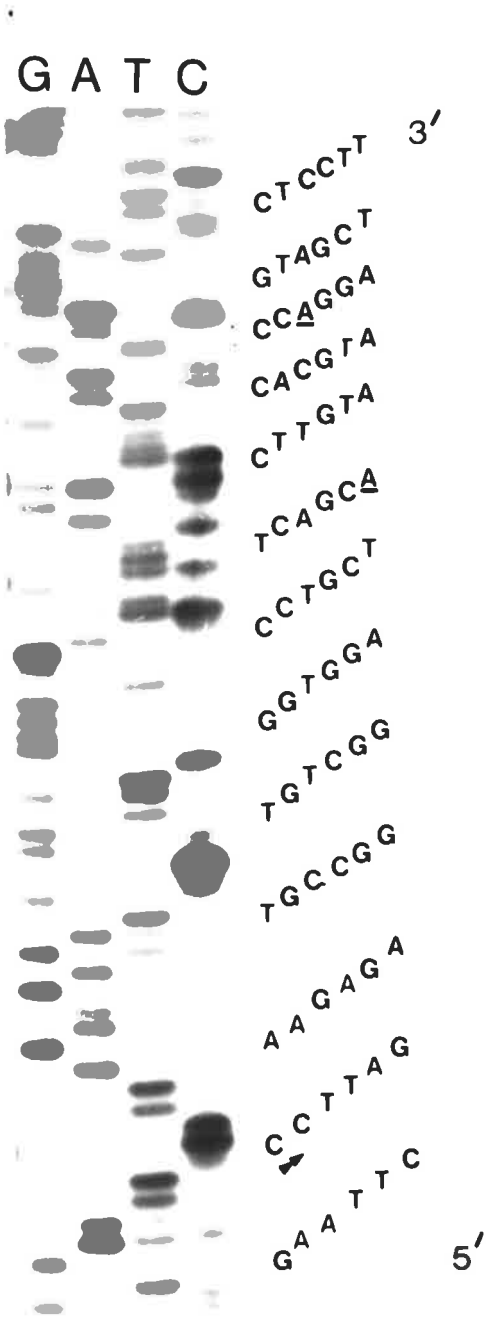

G A T C
3′
5′

Fig. 6.19(b)

Sequencing Gel of an H2B Gene from λHH.03

The sequence is shown of the 243 bp RsaI fragment of pHH1.9E which had been cloned into the SmaI site of M13 mp83 and sequenced by the dideoxynucleotide method (Section 2.2.16b). The sequence reads directly as the mRNA strand shown in Fig. (6.19(c)).

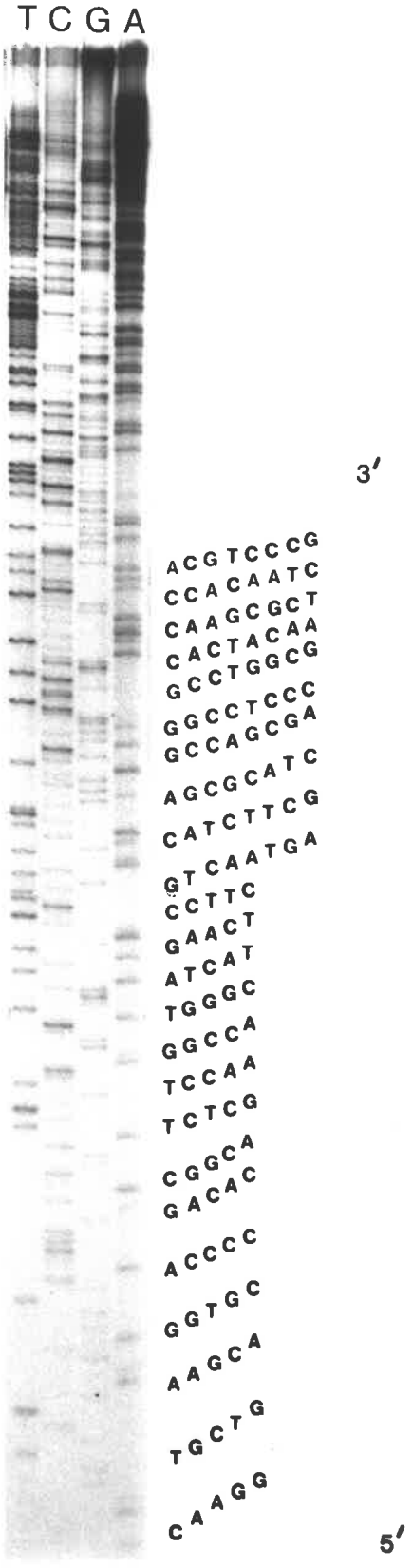

T C G A
3′
ACGTCCCG
CCACAATC
CAAGCGCT
CACTACAA
GCCTGGCG
GGCCTCCC
GCCAGCGA
AGCGCATC
CATCTTCG
GTCAATGA
CCTTC
GAACT
ATCAT
TGGGC
GGCCA
TCCAA
TCTCG
CGGCA
GACAC
ACCCC
GGTGC
AAGCA
TGCTG
CAAGG
5′

Fig. 6.19 (c)

Comparison of Human Histone H2B Gene Sequences from λHH.02 and λHH.03

The H2B gene sequence from λHH.02 was derived from the 140 bp HaeIII fragment of pHH1.9BE and the sequence corresponds to amino acid 22 to 57 of the human H2B protein.

The H2B gene sequence from λHH.03 was derived from the 243 bp RsaI fragment of pHH1.9E and the sequence corresponds to the amino acid sequence from residues 43-120 of the human histone H2B protein. These sequences are compared to the chicken H2B sequence from λCH.02 and the sea urchin H2B gene sequence in h19 (Busslinger *et al.*, 1980). The asterisks indicate nucleotide differences. The lines indicate amino acid changes. Changes from the calf (Von Holt *et al.*, 1979) and human H2B protein sequence (Ohe *et al.*, 1979) are shown.

H2B Gene Sequence Comparisons

	22			25					30			
	Gln	*Lys*	*Lys*	*Asp*	*Gly*	*Lys*	*Lys*	*Arg*	*Lys*	*Arg*	*Ser*	*Arg*
Human λHH.02	CAG	AAG	AAG	GAT	GGC	AAG	AAG	CGG	AAG	CGC	AGC	CGC
Human λHH.03												
Chick λCH.02	CAG	AAG	AAG	GGC	GAC	AAG	AAG	CGC	AAG	AAG	AGC	CGC
Sea Urchin h19	CGG	CCC	AGC	GGC	GGT	AGG	AAG	AGG	AAC	AGG	AAA	AGG

		35					40					45
	Lys	*Glu*	*Ser*	*Tyr*	*Ser*	*Ser*	*Tyr*	*Val*	*Tyr*	*Lys*	*Val*	*Leu*
Human λHH.02	AAG	GAG	AGC	TAC	TCC	TCG	TAC	GTG	TAC	AAG	GTG	CTG
Human λHH.03										AAG	GTG	CTG
Chick λCH.02	AAG	CAG	AGC	TAC	TCG	ATC	TAC	GTG	TAC	AAG	GTG	CTG
Sea Urchin h19	AAG	GAG	AGT	TAT	GGA	ATC	TAC	ATC	TAC	AAA	GTC	CTC
Human Spleen						*Val/Ile*						
Calf						*Val*						

					50					55		
	Lys	*Gln*	*Val*	*His*	*Pro*	*Asp*	*Thr*	*Gly*	*Ile*	*Ser*	*Ser*	*Lys*
Human λHH.02	AAG	CAG	GTC	CAC	CCC	GAC	ACC	GGC	ATC	TCT	TCT	AAG
Human λHH.03	AAG	CAG	GTG	CAC	CCC	GAC	ACC	GGC	ATC	TCG	TCC	AAG
Chick λCH.02	AAG	CAG	GTG	CAC	CCC	GAC	ACC	GGC	ATC	TCG	TCC	AAG
Sea Urchin h19	AAG	CAG	GTT	CAT	CCA	GAT	ACC	GGC	ATC	TCC	AGT	CGG

			60					65				
	Ala	*Met*	*Gly*	*Ile*	*Met*	*Asn*	*Ser*	*Phe*	*Val*	*Asn*	*Asp*	*Ile*
Human λHH.03	GCC	ATG	GGC	ATC	ATG	AAC	TCC	TTC	GTC	AAT	GAC	ATC
Chick λCH.02	GCC	ATG	GGC	ATC	ATG	AAC	TC*G*	TTC	GTC	AA*C*	GAC	ATC
Sea Urchin h19	GCC	ATG	**ATC	ATC	ATG	AAC	***AGC	TTC	GTC	AA*C*	GAC	ATC

	70					75					80	
	Phe	*Glu*	*Arg*	*Ile*	*Ala*	*Ser*	*Glu*	*Ala*	*Ser*	*Arg*	*Leu*	*Ala*
Human λHH.03	TTC	GAG	CGC	ATC	GCC	AGC	GAG	GCC	TCC	CGC	CTG	GCG
Chick λCH.02	TTC	GAG	CGC	ATC	GCC	*GGC	GAG	**GGT	TCC	**GCC	CTG	GCG
Sea Urchin h19	TTC	GAG	CG*A*	AT*T*	GCC	*GGC	GA*A*	**TCT	TCC	CGC	CT*C*	GC*T*
Human Spleen Calf						*Gly* *Gly*						

				85					90			
	His	*Tyr*	*Asn*	*Lys*	*Arg*	*Ser*	*Thr*	*Ile*	*Thr*	*Ser*	*Arg*	*Glu*
Human λHH.03	CAC	TAC	AAC	AAG	CGC	TCA	ACA	ATC	ACG	TCC	CGC	GAA
Chick λCH.02	CAC	TAC	AAC	AAG	CGC	TC*G*	AC*C*	ATC	ACG	TC*G*	CG*G*	GA*G*
Sea Urchin h19	CA*G*	TAC	AAC	AA*A*	***AAG	TCA	AC*C*	ATC	**AGC	***AGT	CGC	GA*G*

		95					100					105
	Val	*Gln*	*Thr*	*Ala*	*Val*	*Arg*	*Leu*	*Leu*	*Leu*	*Pro*	*Gly*	*Glu*
Human λHH.03	GTG	CAG	ACG	GCC	GTG	CGC	CTG	CTG	CTG	CCC	GGC	GAG
Chick λCH.02	**ATC	CA*A*	AC*A*	GCC	GTG	CG*G*	CTG	CTG	CTG	CCC	GGC	GAG
Sea Urchin h19	* *AAT	CAG	AC*C*	GCC	GT*C*	CGC	CT*C*	* *ATT	CT*C*	CCC	GG*A*	GAG
Human Spleen	*Ile*											

					110					115		
	Leu	*Ala*	*Lys*	*His*	*Ala*	*Leu*	*Ser*	*Glu*	*Gly*	*Thr*	*Lys*	*Ala*
Human λHH.03	CTG	GCC	AAG	CAC	GCG	CTG	TCC	GAG	GGC	ACT	AAG	GCT
Chick λCH.02	CTG	GCC	AAG	CAC	GCG	GTC	TCG	GAG	GGC	ACC	AAG	GCG
Sea Urchin h19	CTG	GCA	AAG	CAC	GCT	GTG	AGC	GAG	GGT	ACC	AAG	GCA
Human Spleen						*Val*						
Calf						*Val*						

			120							
	Val	*Thr*	*Lys*							
Human λHH.03	GTC	ACA	AAG	——	——	——	——	——	——	TAG
Chick λCH.02	GTC	ACC	AAG	——	——	——	——	——	——	TAA
Sea Urchin h19	GTG	ACG	AAA	TAC	ACT	ACC	TCC	AAG	TAG	Stop

this residue in H2B protein isolated from spleen with 75% of the residues coding for valine and 25% for isoleucine. However, the H2B gene in λHH.02 codes for serine at this residue. It therefore appears that this gene is not expressed in human spleen tissue and therefore must correspond to a H2B histone protein not yet characterised.

The amino acid sequence of calf histone H2B (Von Holt *et al.*, 1979) is identical to that of human spleen and thus differs from that of the λHH.02 protein only at position 39 (Valine → Serine). However, when the λHH.02 H2B protein sequence is compared to the same region of the chicken H2B gene in λCH.02, four amino acid changes are observed. These are at residues (25) Aspartate → Glycine, (26) Glycine → Aspartate, (31) Arginine → Lysine and (39) Serine → Isoleucine.

In contrast, however, to the relative conservation in the H2B sequence between human, calf and chicken, it is seen from Fig. (6.19) that there are 10 amino acid changes between the sea urchin H2B protein sequence from *P. miliaris* in h19, and the 36 amino acids of the human H2B gene in λHH.02. This region however covers the amino third of the H2B protein which is known to be poorly conserved (Isenberg, 1979).

(b) H2B Histone Protein Sequence From λHH.03

From the 243 bp RsaI fragment of pHH1.9E, a nucleotide sequence was obtained which corresponded to the amino acid sequence from residues 43 - 120 of the human histone protein sequence determined from spleen (Ohe *et al.*, 1979) Fig. (6.19c).

As before, most divergence is observed when the H2B protein sequence in λHH.03 is compared with that of sea urchin H2B (h19) where 9 changes occur in 78 amino acids. In contrast, when the two H2B protein sequences from λHH.03 and λHH.02 are compared, Fig. (6.19c), no amino acid changes are observed from residue 43 - 57, which is the only region of overlap. However, three amino acid changes are observed between the human spleen and calf H2B protein sequence and the λHH.03 protein sequence; these are at residues (75) Glycine → Serine, (94) Isoleucine → Valine and (111) Valine → Leucine. Therefore, the H2B gene coded in λHH.03 must also express a variant H2B protein to that isolated from human spleen.

Extension of this comparison to the chicken H2B gene of λCH.02 is particularly interesting, as at residues 75, 94 and 111, the sequence of the chicken protein is identical to that of human spleen and calf H2B, but divergent from the H2B gene sequence contained in λHH03

The implication of these amino acid changes through evolution is of great interest because these changes observed in λHH.02 are more likely to be important functionally than to have arisen through evolutionary drift, and therefore may reflect conformational changes of the nucleosome, altered interactions of the nucleosome, or both.

(c) H2A Histone Protein Sequence

A total of 330 nucleotides from the H2A gene of λHH.02 (Fig.6.20b) was sequenced from the 100 bp HindIII fragment; the

Fig. 6.20(a)

Sequencing Gels of an H2A Gene of λHH.02

The left tracks show the sequence determined from the HindIII/BamHl fragment of pHHl.9BE which had been cloned into HindIII/BamHl cut Ml3 mp83 and sequenced by the dideoxynucleotide method (Section 2.2.16b). The sequence reading from the HindIII site is the opposite strand to the mRNA sequence shown in Fig. 6.20(b).

The right tracks show the Maxam-Gilbert sequence reactions (Section 2.2.16a) of one strand of the end-filled 100 bp HindIII fragment of pHHl.9BE. The sequence reads 5' to 3' down the gel and is the opposite strand to the mRNA sequence (Fig. 6.20(b)).

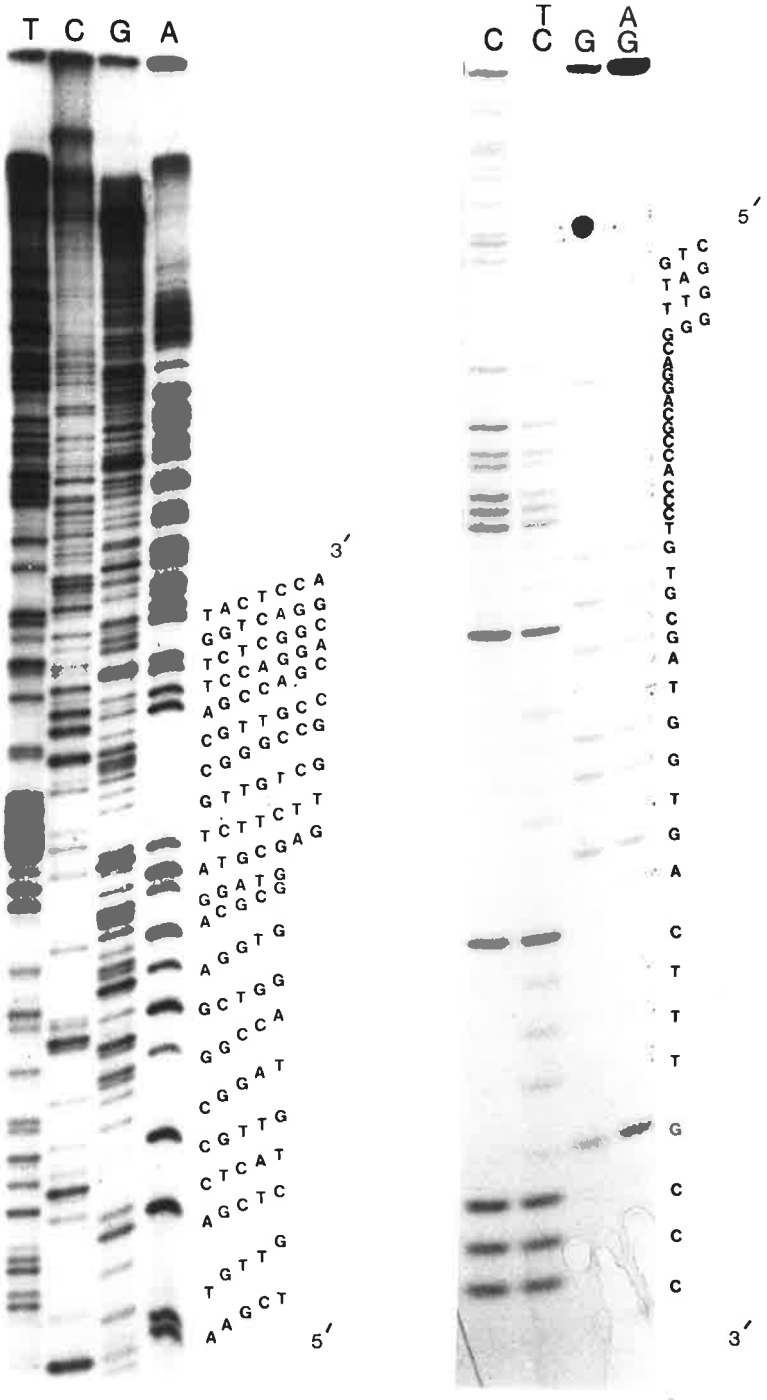

T C G A
3′
5′
C T G A
C G G G
5′
3′

Fig. 6.20(b)

Comparison of Human H2A Gene Sequence

Nucleotide sequence of the H2A gene from λHH.02 is shown which corresponds to amino acids 21 to the end of the gene. This is compared to the chicken H2A gene sequence from λCH.01.(D'Andrea *et al.*, 1981) and the sea urchin H2A gene sequence from h19 (Busslinger *et al.*, 1980). The astersiks indicate nucleotide changes and the underlines indicate amino acid changes. Amino changes are also shown compared to the H2A sequence from trout, rat, and calf (Von Holt *et al.*, 1979). The dashes indicate amino acid deletions.

H2A Gene Comparisons

```
                21                  25                  30
                Ala Gly Leu Gln Phe Pro Val Gly Arg Ser His Arg
λHH.02          GCC GGG CTT CAG TTT CCC GTA GGC CGG AGT CAT CGC
λCH.01          GCC GGG CTG CAG TTC CCC GTG GGC CGC GTG CAC CGG
h19             GCA GGG CTC CAG TTT CCA GTG GGA CGT GTT CAT CGG

                Leu     Arg Lys Gly Asn Tyr Ala Glu Arg Val Gly
λHH.02          CTG xTC CGC AAA GGC AAC TAT GCG GAG CGG GTC GGT
λCH.01          CTG CTG CGC AAG GGC AAC TAC GCG GAG CGG GTG GGC
h19             TTT CTC CGA AAG GGC AAC TAT GCA AAG AGG GTC GGC

                45                  50                  55
                Ala Gly Ala Pro Val Tyr Leu Ala Ala Val Leu Glu
λHH.02          GCT GGA GCG CCG GTG TAC CTG GCG GCG GTG CTG GAG
λCH.01          GCC GGC GCC CCG CTG TAC CTA GCG GCC GTG CTG GAG
h19             GGT GGA GCT CCT GTC TAC ATG GCT GCC GTC CTA GAG

                            60                  65
                Tyr Leu Thr Ala Glu Ile Leu Glu Leu Ala Gly Asn
λHH.02          TAC CTG ACC GCC GAG ATC CTG GAG CTG GCT GGC AAC
λCH.01          TAC CTG ACG GCC GAG ATC CTG GAG CTA GCG GGC AAC
h19             TAC CTC ACT GCC GAA ATC TTG GAA CTC GCA GGC AAC

                    70                  75                  80
                Ala Ala Arg Asp Asn Lys Lys Thr Arg Ile Ile Pro
λHH.02          GCG GCC CGC GAC AAC AAG AAG ACT CGT ATC ATC CCG
λCH.01          GCG GCC CGC GAC AAC AAG AAG ACG CGC ATC ATC CCC
h19             GCT GCC CGC GAC AAC AAG AAA TCT AGG ATC ATC CCA
Trout                                                   (——)
```

```
                                        85                                  90
                  Arg  His  Leu  Gln  Leu  Ala  Ile  Arg  Asn  Asp  Glu  Glu
λHH.02            CGT  CAC  CTC  CAG  CTG  GCC  ATC  CGC  AAC  GAT  GAG  GAG

λCH.01            CGC  CAC  CTG  CAG  CTG  GCC  ATC  CGC  AAC  GAC  GAG  GAG

h19               CGC  CAC  CTT  CAA  CTC  GCT  GTG  CGT  AAT  GAT  GAA  GAA

Trout                                       Val

                       95
                    HindIII                          100
                  Leu  Asn  Lys  Leu  Leu  Gly  Lys  Val  Thr  Ile  Ala  Gln
λHH.02            CTC  AAC  AAG  CTT  CTG  GGC  AAA  GTC  ACC  ATC  GCA  CAG

λCH.01            CTC  AAC  AAG  CTG  CTG  GGC  AAG  GTG  ACC  ATC  GCG  CAG

h19               CTC  AAC  AAG  CTT  CTG  GGT  GGG  GTG  ACG  ATT  GCG  CAG

Trout                                       Gly
Rat                                         Arg

                  105                             110
                  Gly  Gly  Val  Leu  Pro  Asn  Ile  Gln  Ala  Val  Leu  Leu
λHH.02            GGT  GGC  GTC  CTG  CCC  AAC  ATC  CAG  GCC  GTG  CTA  CTG

λCH.01            GGC  GGG  GTG  CTG  CCC  AAC  ATC  CAG  GCC  GTG  CTG  CTG

h19               GGT  GGT  GTT  CTG  CCC  AAC  ATC  CAA  GCC  GTG  CTG  CTT

                            120                                 125
                  Pro  Lys  Lys  Thr  Glu  Ser  His  His  Lys  Ala  Thr  Gly
λHH.02            CCC  AAG  AAG  ACC  GAG  AGC  CAC  CAC  AAG  GCG  ACA  GGG

λCH.01            CCC  AAG  AAG  ACC  GAC  AGC  CAC  ---  AAG  GCC  AAG  GCC

h19               CCC  AAG  AAA  ACT  GCT  AAA  TCA  AGC  ---  ---  ---  ---

Calf                                                                 Lys
Trout                                                                Val  Ala

                  Lys      HindIII
λHH.02            AAG TAG  AAGCTT

λCH.01            AAG TGA  GCACCG

h19               --- TGA
```

1.4 kb HindIII/BamHI fragment; and the 184 bp RsaI fragment in pHH1.9BE. This sequence corresponded to the H2A protein sequence from residues 21-129 of calf (Von Holt *et al.*, 1979). No human histone H2A protein sequence has yet been reported, although Ohe *et al.* (1979) have indicated that spleen H2A protein may be identical in sequence to that determined for calf. One amino change however, is observed between that of calf H2A protein and the λHH.02 human H2A protein at residue (127) *Lys* → *Thr*, which suggests that at least one of the human H2A proteins has diverged from the calf H2A protein.

Fig.(6.20b) compares the human H2A gene sequence from λHH.02 to the chicken H2A sequence from λCH.01 (D'Andrea *et al.*, 1981), and it is observed that at least three amino acid changes occur. These are at residues (121) *Glu* → *Asp*, (127) *Thr* → *Lys*, (128) *Gly* → *Ala*, and one amino acid deletion occurs in the chicken sequence at residue 124. In addition, it can be seen from Fig. (6.20b) that there are at least 5 amino acid changes between the human H2A protein and that of trout and one amino acid difference from rat chloroleukemia H2A.

Major amino acid changes however occur between the human H2A protein sequence and that for sea urchin in h19. As can be seen from Fig.(6.20b) 12 amino acid differences occur and 5 amino acid deletions occur at the carboxy end of the molecule. Little is yet known about the variable and constant regions of H2A since only protein sequence data from calf, trout, sea urchin, chicken

erythrocytes and rat chloroleukemia cells have been determined (Isenberg, 1979). However, it appears that the H2A protein sequence displays a high degree of conservation and the major variability occurs at the carboxy terminal end of the molecule.

6.4.7 Codon Usage and Silent Base Changes

It was of interest to determine if there was any obvious correlation between the codon usage displayed by the three human histone genes and the codon usage observed in the same chicken and sea urchin histone genes versus the codon preference seen in non-histone human genes. That is, is there any further evidence to suggest that codon usage may be dictated by protein type or species type.

Table (6.1) shows that both the human H2B genes display virtually identical codon third base preferences, with a strong bias for G and C residues. This pattern is also observed in the chicken H2B gene as well as the chicken H2A gene. However, the human H2A gene appears to display a slightly higher usage of A and T residues as is also observed for the β globin and pre-growth hormone human genes (Efstratiadis *et al.*, 1980; Perler *et al.*, 1980).

In contrast, the sea urchin histone genes display a relatively more even selection of third base residues. Therefore the human and chicken third base codon preferences are more related than to that observed for the sea urchin histone genes.

Table 6.1

H2A and H2B Codon Usage Comparisons

Summarised are the percentage codon third base preferences for the H2A and H2B human genes in λHH.02 and λHH.03, the chicken λCH.01 H2A gene, λCH.02 gene and sea urchin h19 H2A and H2B genes. These are compared to the human β globin and pre-growth hormone third base preferences. (Efstratiadis *et al.*, 1980; Perler *et al.*, 1980).

	G	A	T	C	
H2B	47%	0	8%	44%	λHH.02
	40%	5%	4%	50%	λHH.03
	45%	3%	2%	50%	λCH.02
	31%	11%	16%	42%	h19
H2A	35%	7%	10%	47%	λHH.02
	47%	0%	0%	53%	λCH.01
	19%	17%	24%	33%	h19
β Globin	34%	7%	27%	32%	Human
Pre-growth Hormone	35%	10%	9%	45%	Human

However, as discussed in Chapter 5, there is still insufficient data to generalise about what governs the codon usage of a particular gene.

Comparison at the level of individual nucleotide changes (silent and replacement) shows a strong correlation with evolutionary divergence. For example, in the H2B gene in λHH.03 there are 23 base changes in 78 codons observed when the nucleotide sequence is compared to the H2B chicken gene in λCH.02 and 59 base changes (of which 15 are silent) when compared to the sea urchin H2B gene in h19, as shown in Fig. (6.19). The difference in these figures reflects the time of divergence between sea urchin, chicken and human.

The next section examines the extent of divergence between the human and chicken histone genes.

6.4.8 Divergence of Human and Chicken Genes

Comparison of the protein sequences of histones of many species has shown them to be among the most highly conserved proteins. It is now possible to consider this conservation in terms of the nucleotide sequence of the genes. That is, how many base changes result in replacement of one amino acid by another, as well as compare changes at the silent sites (those at which nucleotide changes do not replace amino acids). While changes at replacement sites can easily be acted upon by selection, the changes in silent sites are putatively neutral.

By comparing the human histone genes with the chicken histone genes, an evolutionary time scale of 250-300 million years is spanned, since that is when bird/mammalian divergence has been estimated as occurred (Wilson *et al.*, 1977). Table (6.2) compares the human H2A and H2B histone genes in λHH.02 and λHH.03 with the chicken H2A and H2B genes in λCH.01 and λCH.02 respectively, using the percentage corrected divergence calculation described by Perler *et al.*, (1980). These figures are also compared to percentage divergence between human and chicken pre-proinsulin (A + B chains) and α + β globin genes.

It is apparent that the percentage divergence in replacement sites between the λHH.02 and λCH.01 H2A gene is very low (2.4%), in comparison to the percentage divergence in replacement sites observed between human and chicken pre-proinsulin and globin genes. A similar low divergence (2.5%) is seen in comparing the human H2B gene in λHH.03 with the chicken H2B gene in λCH.02. However, the human H2B gene from λHH.02 shows a somewhat higher divergence (8.2%), approximately that for pre-proinsulin chicken and human genes. These data point to the strong selective pressure on the histone protein sequence, in comparison, for example, with the α and β globin genes.

In contrast, the percentage divergence in silent sites of the human histone H2A gene is much higher (56.4%) and close to that observed for globin genes, reflecting the relative lack of constraint in silent substitutions. However, it is apparent by comparison with the percentage divergence of silent sites between other chicken and human genes (Table 6.2), that the silent sites exhibit widespread

Table 6.2

Divergence of Human and Chicken Genes

Comparison of the human and chicken H2A DNA sequences from λHH.02 and λCH.01 and H2B DNA sequences from λHH.02, λHH.03 and λCH.02 are compiled. These are compared to the percentage divergence between human and chicken pre-proinsulin A and B chains and α and β globin genes. The percentage figures represent percentage corrected divergence calculated as described by Perler *et al.*, (1980); this corrects divergence for codon type and multiple mutational events.

Human and Chicken Genes	Replacement Sites	Silent Sites
H2A Histone		
λHH.02 + λCH.01	2.4%	56.4%
H2B Histone		
λHH.02 + λCH.02	8.2%	48.4%
λHH.03 + λCH.02	2.5%	27.1%
Pre-proinsulin A + B Chains	8%	122%
α-Globin	20.9%	74.6%
β-Globin	22.9%	70.1%

differences in divergence rates. The immediate implication is that there is selective pressure on some fraction of the silent sites. This is apparent in the relatively low percentage divergenece of silent sites (27.1%) observed between the human and chicken H2B genes in λHH.03 and λCH.02. These constraints could be due to the imposition of required secondary structure at the mRNA level or could reflect specific anticodon requirements, such as the strong bias towards G and C as the codon third base in both chicken and human genes.

6.5 DISCUSSION

DNA sequences were obtained for the H2A and H2B genes in λHH.02 as well as for one of the H2B genes in λHH.03, thus verifying the hybridisation data obtained with heterologous chicken histone probes.

Both H2B genes were similar to each in the common amino acids sequenced, however, both were variants with respect to the H2B protein(s) expressed in human spleen tissue. The amino acid sequence predicted from the H2A gene in λHH.02, represents the first region of protein sequence for a human H2A histone; it is therefore not possible to determine in which tissue, or at what stage in development this protein is expressed.

A comparison of the human H2B and H2A protein sequence with those of other species, has demonstrated the remarkable conservation of amino acids in particular regions, as well as pointing to the lesser conserved regions of the histone molecule. For example, the amino-portion of the H2B protein from residues (22-33) is less conserved with respect to

sea urchin and chicken, than the carboxy-terminal of the H2B molecule. The significance, however, of the single amino acid changes observed between H2B genes in the clones and the calf and human spleen H2B at residues (39), (75), (111) is unclear. However, these changes could confer properties onto the nucleosome which are important at a particular stage in development or in a particular tissue.

The H2A protein sequence is strongly conserved, except for the last few amino acids at the carboxy-terminus, which vary between human, calf, chicken, trout and sea urchin. Again, there are a few single amino acid differences throughout the molecule but the significance of these remains to be determined.

A more accurate estimate of selection pressures on evolutionary divergence between the chicken and human histone genes resides in the percentage divergence of the replacement and silent sites. It is clear from these figures that the histone protein sequence is highly conserved, however, it is not clear to what degree the silent changes are influenced by the factors responsible for the observed bias in codon selection in human and chicken histone genes for G and C as codon third bases.

CHAPTER 7

GENERAL DISCUSSION

7. FINAL DISCUSSION

The histone proteins are collectively responsible for the maintenance of the primary eukaryotic chromosome structure (McGhee and Felsenfeld, 1980). The physical organisation, replication and expression of the eukaryotic genome is influenced to a greater or lesser degree by histone gene expression and histone organisation in chromatin in concert with non-histone chromosomal proteins (Mathis *et al.*, 1980). Histone gene expression may influence the expression of many other genes and thus an understanding of the control of histone genes is of considerable importance to an overall view of eukaryotic gene regulation.

7.1 Gross Histone Gene Organisation

It is now possible to discern the gross structural features of the histone genes over a large evolutionary span, as well as to describe the detailed molecular anatomy of the genes.

In fact, the single most significant result of the work described in this thesis, and that of other recent work in the field, is the demonstration that both the chicken and human histone genes are clustered, but not repeated in tandem arrays.

That the sea urchin type histone gene organisation was not universal, at least in all respects, was first suggested by the observation that vertebrate histone genes were only reiterated to a limited extent (10-50 fold). Nevertheless, the entrenched model of the gross histone gene topology was only clearly refuted by the elucidation of the structure of the major *Drosophila* histone repeats (Lifton *et al.*, 1977). These resemble the sea urchin type structure in that they are highly reiterated in tandem clustered arrays of five

genes, but differ in gene order and have two genes coded on one strand, and three on the other. This clearly means that in *Drosophila* histone mRNAs cannot be derived from a single RNA transcript.

Detailed analysis of the organisation of several vertebrate histone families emerged during the course of this work, suggesting a highly idiosynchratic organisational mode. For example, in *Xenopus*, the histone genes are clustered but different gene orders have been detected (Moorman *et al.*, 1980; Zernick *et al.*, 1980). The organisation of newt histone genes in the species *Nonopthalmus* has recently been elucidated and appears to be completely different from *Xenopus* (Gall *et al.*, 1981). In this newt, the majority of the histone genes comprise homogeneous 9 kb clusters repeated 600-800 fold in a genome fifteen times larger than that of *Xenopus*. However, these clusters are not tandem repeats as in the sea urchin genes, but rather separated by up to 50 kb of a repeating 225 bp satellite DNA.

The initial suggestion that the chicken histone genes may be present in a tandem array (Crawford *et al.*, 1979) appears now to be incorrect, since gene mapping and restriction analysis not only of λCH.02, but also several other genomic histone clones (Harvey *et al.*, 1981; Engel and Dodgson, 1981), indicates that a degree of clustering occurs but that there is no repeat unit.

A similarly complex picture emerges from examination of mammalian genomic clones. The organisation of the human histone genes in both λHH.02 and λHH.03, as well as in the other human genomic clones isolated

(Heintz *et al.*, 1981; Sierra *et al.*,1982a) also indicates a degree of clustering, but again no simple repeating unit is found. This also appears to be the case for mouse histone genes (Seiler-Tuyns and Birnstiel, 1981; Sittman *et al.*, 1981).

It is not yet possible to distingush whether the histone genes of higher vertebrates are contained in a single disperse cluster, or in a number of smaller ones. For this purpose it will be necessary to "chromosome walk", to relate the positions of the various histone genes. Preliminary results from such "walking" in the chicken genome (R. D'Andrea, personal communication) demonstrate the presence of clusters but no recognisable pattern of gene order.

All these results suggest that the mechanism for maintenance of highly conserved, repeated histone gene clusters is not operative in higher vertebrates and that the selective advantage of tandem repetition is not the major influence on vertebrate histone gene structure.

7.1.1 Evolution or Dispersion of Histone Gene Families

(a) Tandem and Non-Tandem Repetition

There are several important evolutionary considerations that must be examined to understand the differences in the organisation of higher vertebrate and previously characterised histone genes.

Tandemly reiterated multigene families are thought to arise and be maintained because of strong selection for expression of their gene products (Federoff, 1979). In cell culture, it can result in amplification of structural genes (Alt *et al.*, 1978). Presumably, the requirement for equimolar amounts of the RNAs encoded by a particular gene family is responsible for the evolution of a repeating unit

containing a single copy of each gene.

The mechanisms for tandem reiteration of gene families *in vivo* or for gene amplification *in vitro* are not understood, and may not be identical. Smith (1976), has proposed that unequal crossing-over at homologous sequences may provide a mechanism for rapidly expanding and contracting tandemly reiterated sequences.

If unequal crossing-over is a major component of the mechanism responsible for maintaining tandemly repeated gene families, then homogeneity between individual repeats maximises the efficiency of this process and provides the genomic flexibility necessary for this sort of gene correlation/amplification model. Both the copy number and the organisation of a particular cluster of repeating units therefore reflect the requirement for a high rate of expression of a particular gene family.

The high number of histone genes in sea urchin, reiterated 200-300 fold, could simply reflect the need for rapid synthesis of histone proteins during sea urchin development. In fact, 30% of the total protein synthesis in rapidly cleaving sea urchin embryos is histone (Nemer and Lindsay, 1969), and these histones are encoded in the highly reiterated "early" principal repeats. During blastulation the "early" histone proteins are replaced by a new class of "late" histones. The organisation of these "late" genes and the molecular events controlling this switch are poorly understood.

A similarly high copy number of histone genes occurs in *Drosophila* and a marked parallel exists between sea urchin and *Drosophila*

histone gene regulation during embryogenesis. However, no clear "switch" to the predominant expression of "late" histone variants has been detected in *Drosophila*.

Since the rate of cell division in mammalian embryos is much slower (about 12 hours) than in invertebrates, the selection pressure for high rates of expression of mammalian histone genes should be greatly reduced. It is therefore significant that the histone genes in higher vertebrates are only reiterated 10-20 fold, as exemplified by the chicken, mouse (Hentschel and Birnstiel, 1981) and human genomes. This number of histone genes could represent the minimum number of different genes required for both developmental and tissue specificity; however, the number of variant histone genes has not yet been determined.

The lack of a repeating unit containing the human and chicken histone genes may therefore reflect genetic drift from a tandemly reiterated precursor. One consequence of this divergence may be an increase in the stability of individual copies of human and chicken histone genes as a result of the decreased frequency of unequal crossing-over and gene conversion. This could result in the stabilisation of histone genes as separate entities able to evolve independently. Such independent evolution is essential if individual histone variants have biologically distinct and significant properties.

(b) Repeated DNA Sequences

In addition to the clonal expansion or contraction of a particular repeating unit by unequal crossing-over, mobile genetic elements, such as "copia" sequences of *Drosophila* (Finnegan *et al.*, 1977)

or as yet uncharacterised mechanisms for genome re-organisation, may be involved in the evolution or dispersion of gene families. The existence of these mechanisms suggests that the rate of evolution of gene families may be rapid and that particular structures of the histone gene clusters may have arisen recently during evolution.

In most eukaryotes, repeated sequences approximately 300 nucleotides long have been found interspersed with single copy and dispersed multigene families (Jelinek *et al.*, 1980) throughout the genome. In humans, the Alu DNA family is predominant among these reiterated sequences. Alu sequences have been found in some of the human histone clones (Sierra *et al.*, 1982b) and middle repetitive DNA in λHH.02 and λHH.03, as well as in the chicken histone clone λCH.02.

Two possibilities are apparent in considering the presence of interspersed middle repetitive DNA. One is that they are somehow involved in the regulation of expression of the histone genes. For example, it is known that members of the Alu family and other middle repetitive sequences are transcribed (Weiner, 1980; Haynes and Jelinek 1981; Davidson and Posakony, 1982) and that their expression can be correlated with developmental changes (Flavell *et al.*, 1980; Scherer *et al.*, 1981). There is as yet, however, no evidence for their having a role in the expression of other genes.

A second possibility related to the mobility of middle repetitive sequences and their dispersion in the genome, is that their presence in the histone gene clusters is a reflection of past rearrangement events such as could have been important in generating the disorganisation of the repeating histone gene unit.

7.1.2 Gene Organisation in Relation to Expression

As discussed in the previous section, it is likely that the organisation of histone genes into highly reiterated tandemly repeating units, each containing one copy each of the five histone genes, in sea urchin and *Drosophila*, as well as in the other invertebrates (Freigan *et al.*, 1976), is required to, not only maintain a high level of expression of the five histone genes during embryogenesis, but also to ensure gene homogeneity by gene correction mechanisms.

Therefore, what role, if any, does the organisation of histone genes from higher vertebrates into seemingly "disorganised arrays" play in regulating the co-ordinate expression of these genes? By comparing the organisation of the human histone genes in λHH.02 and λHH.03 with the arrangement of human histone genes from other non-overlapping clones (Fig. 6.14), it is apparent that the genes are clustered but appear to be scattered randomly within this cluster. For example, 10 H4 genes have been mapped, but are mostly flanked by different genes and separated from each gene by varying lengths of spacer DNA.

However, what is noticeable in organisation of the histone genes is the predominance of gene pairs, in particular H3/H4 and H2A/H2B. Histone gene pairs are also found in the two chicken histone genomal clones λCH.02 and λCH.01. The role of such organisation is not clear because there are examples of gene pairs oriented in opposite directions (e.g. two H3 genes in λCH.02 and H2A/H2B gene pair in λCH.01) and gene pairs oriented in the same direction (e.g. H3/H4 gene pair in λCH.02). Perhaps the most likely explanation is that because each gene has its own independent 5' and 3' control regions (see Section 7.3), it doesn't matter whether each gene is transcribed from the same or different

strands or in fact where genes are placed relatively to each other. However other possibilities perhaps could be considered for the role of divergently transcribed gene pairs. For instance, the arrangement could allow for simultaneous expression of the genes through a shared control region or conversely transcriptional activity of one gene may block activation of others. Determining which genes are expressed co-ordinately in relation to their organisation will help to resolve this question.

The larger spacer regions found between some of the histone genes in both chicken and human genomes could code for other proteins which may or may not influence the expression of various histone genes. In fact, Rickles *et al.* (1982) have detected a sequence that strongly hybridises to an RNA species of about 330 nucleotides in one of their human histone genomic clones and this RNA is present predominantly during the G1 phase of the HeLa cell cycle. It would, therefore be interesting to screen all the human genomic clones for the presence of other dispersed genes to determine if these genes code for proteins or RNA involved during histone gene expression or involved in chromatin formation.

Finally, the seemingly "disorganised array" of histone genes in higher vertebrates may have evolved simply to ensure the stability of functional variant proteins, by limiting gene conversion/correction mechanisms. In fact, the opposite orientation of many of the histone genes also would result in limiting unequal crossing-over and therefore help maintain the variety of histone variants.

It is therefore possible that the organisation of histone genes in higher vertebrates merely reflects the need to maintain a number of variants and that each gene is independently regulated and thus contains its own promoter. How the genes are co-ordinately expressed however is not obvious from their organisation and is therefore likely to involve, as yet unidentified, control sequences.

7.2 Structure of the Histone Genes

In contrast to the diverse organisation of the histone gene family, the primary structure of the histone proteins has been extremely conserved from a variety of species and thus provides a striking demonstration of the differences between evolution of a structural gene and that of the corresponding gene family.

7.2.1 Histone Proteins

The various regions sequenced from the chicken and human histone genes in λCH.02, λHH.02 and λHH.03, further demonstrated the highly conserved nature of the histone protein, directly reflecting the fundamental character of functions performed by histones within the cell. The H3 and H4 are the most stringently conserved of the histone proteins, while H2A and H2B proteins display a variable N-terminal region but a highly conserved C-terminal globular region. Histone H1 is the most variable of the histones but does contain a highly conserved central region.

However histone protein structural variants have been characterised (Von Holt *et al.*, 1979) (e.g. embryonic and adult histones of sea urchin, tissue specific, e.g. H5, basal histones synthesised throughout the cell cycle and S-phase-specific histones) and these appear to be

either tissue-specific or specific to a particular stage in development. DNA sequencing now permits a more detailed search for histone variants. In fact partial sequencing data from both the human histone H2B genes encoded in λHH.02 and λHH.03 indicates that each of these genes codes for a different H2B protein to that expressed in human spleen.

It is thought that specific histone variants are responsible for particular functional states of chromatin, presumably, by lending to the DNA a conformation which is either transcriptionally active or inactive. It would therefore be extremely interesting to determine the time of expression of the individual histone genes. The sites of expression of these histone variants could be ascertained by stringent hybridisation or by S1-nuclease mapping with mRNA isolated from either different tissues or at different stages in development. The predicted histone proteins could also be detected from specific tissues or during development by peptide mapping. Once the time of expression of the histone variants is determined, it might then be possible to relate the structure of the variants with their specific function. For example, it is known that in terminally differentiating erythroid cells, chick H5 protein replaces H1 protein and the H5 protein leads to a tighter higher-order chromatin structure. Perhaps the best way to determine the effect of variants on chromatin structure would be to introduce the gene into recipient cells, via e.g. SV-40 vectors,and look at its effect on the minichromosome vector itself and/or the chromatin of the host cell (that is if the variant gene was expressed in sufficient quantity to modify the host chromosome).

7.2.2 Histone Genes

A comparison of the histone genes sequenced from the chicken and human genomic clones ~~to~~ with the histone genes of other species has demonstrated several features which are common to all the histone genes.

(i) The coding regions ~~do not~~ seldom contain introns. (Recently however, an interesting exception of a chicken H3 gene containing an intron has been described by Engel *et al.*, (1982)).

(ii) The coding regions are G,C rich; this is even more pronounced for the chick and human genes than for sea urchin, as the third base preference for codons is almost exclusively G or C.

(iii) Downstream of the coding regions however histone genes appear to differ from other polymerase II-transcribed genes. In place of the normally highly variable 3' trailer, a region of exceptionally high sequence homology characterised by a GC-rich hyphenated dyad symmetry motif is found, while the normally ubiquitous AATAAA polyadenylation signal is almost invariably missing.

The lack of a polyadenylation signal correlates with the majority of the histone mRNAs being poly(A)$^-$. Since one of the roles purported for poly(A) tails is in mRNA stability, it is possible that this absence on most histone mRNAs allows a high turnover rate. Conversely, the histone mRNAs which have been identified as poly(A)$^+$ may be more stable. Thus it is possible that mRNA stability may play a role in regulation of histone gene expression.

Sequence analysis presented in this thesis has allowed a start to be made in evaluating the evolution of histone genes at the nucleotide

level. The rate of evolution of histone gene sequences is predictably lower than for other genes. This is true particularly of amino acid replacement substitutions, but also to some extent for silent substitutions. An interesting observation is that the H2B protein of human spleen has an identical sequence to that of the chick gene in λCH.02, but that the human gene of λHH.03 differs from both. This indicates that its divergence may have taken place prior to bird/mammalian radiation. (Support for this will need to come from sequencing of the gene for the human spleen-type H2B protein.) The accumulation and comparison of further sequences from histone genes of various species, will no doubt clearly provide many interesting insights into the way in which these families have evolved.

7.3 Regulatory Sequences

Examination of the gross organisation of the histone genes in organisms from several genera has not itself clarified the regulatory mechanism controlling their expression, although in most organisms regulatory models involving polycistronic operons comprising the five histone genes are now clearly excluded. More detailed structure versus function correlations have resulted from DNA sequence analysis.

Histone gene sequence organisation in the higher vertebrates is like that of the invertebrates as well as other genes transcribed by polymerase II. The 5' leader sequences contain a "TATA" box as well as promoter type sequences upstream from the 5' mRNA terminal (e.g. "CAT" box). This suggests that each histone gene is individually transcribed. However, no striking 5' sequence homologies unique to the histone genes, and hence suggestive of a transcription co-ordination function, are found with the possible exceptions of the motif 8-9 nucleotides

upstream from the "TATA" box in sea urchin histone genes and the larger homology block found in the same position in the two chicken H2B genes from λCH.01 and λCH.02 (R. Harvey, personal communication).

A possible function of the upstream "modulator" sequences has been suggested for *Drosophila* histone genes (Samal *et al.*, 1981) in that they may help to determine nucleosome phasing in and around the 5' region of the mRNA coding genes. The structure of *Drosophila* histone gene chomatin has recently been examined and there is evidence that the nucleosomes may not be randomly aligned relative to the 5' histone gene sequences. Obviously such concepts can also be involved in models for the developmental regulation of gene expression.

Equally, and despite increasing evidence that DNA methylation may be a key element in regulating gene expression (Razin and Riggs, 1980), little is known about the specific pattern of histone gene methylation save that the histone genes belong to the poorly (or unmethylated) fraction of the sea urchin DNA (Bird *et al.*, 1979).

These arguments however do not negate the probable primacy of transcriptional regulation in the dramatic development or cell cycle changes in histone gene expression. Indeed such changes have been monitored in isolated nuclei (Levy *et al.*, 1978; Detke *et al.*, 1978). Rather, they suggest that either the control events occur in larger sections of DNA than the individual gene transcription units or that we still cannot percieve important sequence commonalities, or modulators, for example, perhaps because they are highly hyphenated.

APPENDIX

APPENDIX

One of the major problems in examining the control of synthesis of histone mRNAs in higher vertebrates has been the lack of sufficiently homologous probes for the mRNAs. Consequently DNA from the subclones of λCH.02, pCH2.6H containing H3 and H4 chicken histone genes and pCH4.8E containing H2B and H1 histone genes were sent to Drs. G. and J. Stein and co-workers who have used them to examine a number of questions. The major points which they have concluded from these studies are as follows.

1. Although it is generally agreed that histone protein synthesis is restricted to the S phase of the cell cycle, and therefore parallels DNA replication, both transcriptional and post-transcriptional levels of control have been suggested. To gain more definite information about the regulation of histone gene expression during the cell cycle, the H3/H4 and H2B/H1 genomic subclones were used as probes in blot hybridisation experiments to re-examine the representation of histone mRNAs in the nucleus and cytoplasm of G1 and S phase synchronised HeLa S3 cells (Rickles *et al.*, 1982). The representation of histone mRNA sequences of G1 compared with S phase cells was <1% in the cytoplasm and ~1% in the nucleus. These data are consistent with transcriptional control, although it is still possible that regulation of histone gene expression is, to some extent, mediated post-transcriptionally. If histone gene transcription does occur in G1, the RNAs must either be rapidly degraded or be transcribed to a limited extent compared with S phase.

2. At least seven different species of H4 histone mRNA were isolated from the polysomes of S phase HeLa S3 cells (Lichtler *et al.*, 1982).

The presence of different mRNAs coding for the same protein could be explained either by the presence of different genes or by differential processing of the same primary transcript. By hybridising a series of cloned human genomic H4 histone sequences (λHHG41, λHHG39 and λHHG17) to the histone H4 RNAs and digesting with S1 nuclease, Lichtler *et al.*, (1982) have shown that different H4 mRNA species are protected from nuclease digestion by different H4 histone genes, suggesting at least three of the HeLa H4 histone mRNAs are the products of distinct genes.

3. Seven independent λCharon 4A recombinant phages (λHHG) which contain human histone genes were isolated from the Maniatis human library using pCH2.6H DNA as hybridisation probe (Sierra *et al.*, 1982a). Histone genes were localised in these clones by using the pCH2.6H and pCH4.8E DNA as probes, as well as by hybrid selection and *in vitro* translation. The λHHG clones fall into at least three distinct types of arrangement, but display no apparent repeat (as discussed in Chapter 6,2.10).

At least one clone, λHHG41, contains, in addition to the histone genes, a region that hybridises with a cytoplasmic RNA approximately 330 nucleotides in length. This RNA is not similar in size to known histone-encoding RNAs and is present in the cytoplasm of HeLa cells predominantly in the G1 phase of the cell cycle. In addition, some of the λHHG clones were found to contain Alu sequences (Sierra *et al.*, 1982 b).

LIST OF PUBLICATIONS

1. Published Papers

Clark, S.J., Krieg, P.L., and Wells, J.R.E. (1981) 'Isolation of a Human Genomal Clone Containing Histone Genes' *Nucleic Acids Res.* 9, 1583-1590

Rickles, R., Marashi, F., Sierra, F., Clark, S., Wells, J., Stein, J. and Stein, G. (1982) 'Analysis of Human Histone Gene Expression During The Cell Cycle in HeLa Cells using Cloned Human Histone Genes' *Proc. Natl. Acad. Sci.* U.S.A. 79, 749-753

Sierra, F., Lichtler, A., Marashi, F., Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G., and Stein, J. (1982) 'Organisation of Human Histone Genes' *Proc. Natl. Acad. Sci.* U.S.A. 79, 1795-1799

Sierra, F., Leza, A., Marashi, F., Plumb, M., Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G. and Stein, J. (1982) 'Human Histone Genes are Interspersed with Members of the Alu Family and with Other Transcribed Sequences' *Biochem. Biophys. Res. Commun.* 104, 785-792

Lichtler, A.C., Sierra, F., Clark, S., Wells, J., Stein, J. and Stein, G. (1982) 'The Multiple H4 mRNA's of HeLa Cells are Encoded in Different Genes' *Nature* 298, 195-198

2. Papers Presented at Meetings

Clark, S.J., and Wells, J.R.E. (1981) 'Human Histone Gene Organisation' *Proc. Aust. Biochem. Soc.* 14.

REFERENCES

Abelson, J. (1979) *Ann. Rev. Biochem.* 48, 1035-1069

Adamson, E.D. and Woodland, H.J.R. (1974) *J. Mol. Biol.* 88, 236-285

Adensnik, M. and Darnell, J.E. (1972) *J. Mol. Biol.* 67, 397-406

Alfageme, C.R., Zweidler, A., Mahowald, A. and Cohen, L.H. (1974) *J. Biol. Chem.* 249, 3729-3736

Alt, F.W., Kellems, R.E., Bertino, J.R. and Schimke, R.T. (1978) *J. Biol. Chem.* 253, 1357-1370

Anderson, K.V. and Lengyel, J.A. (1980) *Cell* 21, 717-727

Appels, R. and Wells, J.R.E. (1972) *J. Mol. Biol.* 70, 425-434

Arceci, R.J. and Gross, P.R. (1980) *Science* 209, 607-609

Arnheim, N., and Southern, E.M. (1977) *Cell* 11, 363-370

Avvedimento, V.E., Vogeli, G., Yamada, Y., Maizel, J.V. Jr., Pastan, I., and de Grombrugghe, B. (1980) *Cell* 21, 689-696

Baksi, K., Rogerson, D.L. and Rushizky, G.W. (1978) *Biochemistry* 17, 4136-4139

Baldwin, R.W. (1970) *Rev. Eur. Etud. Clin. Biol.* 15, 593

Bendig, M.M., Thomas, T. and Folk, W.R. (1980) *Cell* 20, 401-409

Benoist, C. and Chambon, P. (1980) *Proc. Natl. Acad. Sci.* U.S.A. 77, 3865-3869

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

Benoist, C. and Chambon, P. (1981) *Nature* 290, 304-315

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

Bernard, O., Hozumi, N. and Tonegawa, S. (1978) *Cell* 15, 1133-1144

Bina, M., Feldmann, R.J., Deeley, R.G. (1980) *Proc. Natl. Acad. Sci.* U.S.A. 77, 1278-1282

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

Bird, A.P., Taggart, M.H. and Smith, B.A. (1979) *Cell*, 17, 889-891

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

Birnstiel, M.L., Telford, J., Weinberg, E.S. and Stafford, D. (1974) *Proc. Natl. Acad. Sci.* U.S.A. 71, 2900-2904

Birnstiel, M.L., Schaffner, W. and Smith, H.O. (1977) *Nature* 266, 603-607

Birnstiel, M.L., Portmann, R., Busslinger, M., Schaffner, W., Probst, E. and Kressmann, A. (1979) *Alfred Benson Symp.* 13, 117-129

Blattner, F.R., Williams, B.G., Blechl, A.E., Thompson, K.D., Faber, H.E., Furlong, L.A. Gnunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1971) *Science* 196, 161-169

Borun, T.W. (1975) Histones Differentiation and the Cell Cycle in Cell Cycle and Cell Differentiation, J. Reinert and H. Holtzer, eds. (Springer-Verlag, Berlin) p 249

Borun, T.W., Scharff, M.D. and Robbins, E. (1967) *Proc. Natl. Acad. Sci.* U.S.A. 58, 1977-1983

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.* U.S.A. 74, 5652-5656

Brandt, W.F. and Von Holt, C. (1974) *Eur. J. Biochem.* 46, 407-417

Brandt, W.F., Patterson, K. and Von Holt, C. (1980) *Eur. J. Biochem.* 110, 67-76

Breathnach, R., Mandel, J.L. and Chambon, P. (1977) *Nature* 270, 314-319

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

Britten, R.J. and Davidson, E.H. (1969) *Science* 165, 349-357

Brown, D.D. and Dawid, I.B. (1968) *Science* 160, 272-280

Brown, D.D. and Weber, C.S. (1968) *J. Mol. Biol.* 34, 661-680

Brown, D.D., Wensink, P.C. and Jordan, E. (1971) *Proc. Natl. Acad. Sci.* U.S.A. 68, 3175-3179

Brown, D.D. and Fedoroff, N.V. (1978) in *Cell Differentiation and Neoplasia* G.F. Saunders Ed. (Raven, New York, 1978)

Browne, M.J., Turnbull, J.F., McKay, E.L., Adams, R.L.P. and Burdon, R.H. (1977) *Nucleic Acids Res.* 4, 1039-1044

Bruschi, S. and Wells, J.R.E. (1981) *Nucleic Acids Res.* 9, 1591-1597

Bryan, P.N., Holfstetter, H. and Birnstiel, M.L. (1981) *Cell* 27, 459-466

Burckhardt, J. and Birnstiel, M.L. (1978) *J. Mol. Biol.* 118, 61-79

Busslinger, M., Portmann, R. and Birnstiel, M. (1979) *Nucleic Acids Res.* 6, 2997-3008

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

Cameron, J.R., Loh, E.Y. and Davis, R.W. (1979) *Cell* 16, 739-751

Chaconas, G. and van de Sande, J.H. (1980) *Methods in Enzymol.* 65, 75-85

Chambon, P., Benoist, C., Breathnach, R., Cochel, M., Gannon, R., Gerlinger, P., Krost, A., Le Meur, M., Le Pennec, J.P., Mandel, J.L., O'Hare, K. and Perrin, F. (1979) pp 55-81 in "From Gene to Protein: Information Transfer in Normal and Abnormal Cells" Eds. T.R. Russell, K. Brew, H. Faber and J. Schultz (New York Academic Press)

Champagne, M., Mazen, A. and Wilhelm, C. (1968) *Bull. Soc. Chim. Biol.* 50, 1261-1272

Chandler, M., Kedes, L.H., Cohen, R.H. and Yunis, J.J. (1979) *Science* 205, 908-910

Childs, G., Levy, S. and Kedes, L.H. (1979a) *Biochemistry* 18, 208-213

Childs, G.,Maxson, R. and Kedes, L. (1979b) *Develop. Biol.* 73, 153-173

Childs, G., Maxson, R., Cohn, R.H., and Kedes, L. (1981) *Cell* 23, 651-663

Christman, J.K., Price, P., Pedrinan, L. and Acs, G. (1977) *Eur. J. Biochem.* 81, 53-62.

Clark, S.J., Krieg, P.A. and Wells, J.R. (1981) *Nucleic Acids Res.* 9, 1583-1590

Clarke, L. and Carbon, J. (1976) *Cell* 9, 91-99

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

Cohen, L.H., Newrock, K.M. and Zweidler, A. (1975) *Science* 190, 994-997

Cohn, R.H., Lowry, J.C. and Kedes, L.H. (1976) *Cell* 9, 147-161

Cohn, R.H. and Kedes, L.H. (1979a) *Cell* 18, 843-854

Cohn, R.H. and Kedes, L.H. (1979b) *Cell* 18, 855-864

Compere, S.J. and Palmiter, R.D. (1981) *Cell* 25, 233-240

Crawford, R.J., Krieg, P., Harvey, R.P., Hewish, D.A. and Wells, J.R.E. (1979) *Nature* 279, 132-136

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

Darnell, J.E. Jr. (1979) *Miami Winter Symp.* 16, 207

Davidson, E.H., Galou, G.A., Angerer, R.C. and Britten, R.J. (1975) *Chromosoma* 51, 253-259

Davidson, E.H. and Posakony, J.W. (1982) *Nature* 297, 633-635

Davies, J.E. and Rownd, R. (1972) *Science* 176, 758-768

Dawid, I.B., Brown, D.D. and Reeder, R.H. (1970) *J. Mol. Biol.* 51, 341-360

Dawid, I.B. and Wahli, W. (1979) *Develop. Biol.* 69, 305-328

De Lange, R.J., Williams, L.C. and Searcy, D.G. (1981) *J. Biol. Chem.* 256, 905-911

Desrosiers, R.C., Mulder, C. and Fleckenstein, B. (1979) *Proc. Natl. Acad. Sci.* U.S.A. 76, 3839-3843

Destree, O.H.S., Hoenders, H.J., Moorman, A.F.M. and Charles, R. (1979) *Biochim. Biophys. Acta* 577, 61-70

Detke, S., Stein, J.L. and Stein, G.S. (1978) *Nucleic Acids Res.* 5, 1515-1529

Detke, S.,Lichtler, A., Phillips, I., Stein, J. and Stein, G. (1979) *Proc. Natl. Acad. Sci.* U.S.A. 76, 4995-4999

Diaz, M.O. (1981) *Cell* 24, 649-659

Dodgson, J.B., Strommer, J. and Engel, J.D. (1979) *Cell* 17, 659-668

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., De Riel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L. Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) *Cell* 21, 653-668

Elgin, S.C.R. (1981) *Cell* 27, 413-415

Elgin, S.C.R. and Weinstraub, H. (1975) *Ann. Rev. Biochem.* 22, 725-774

Endow, S.A. and Gall, J.G. (1975) *Chromosoma* 50, 175-192

Engel, J.D. and Dodgson, J.B. (1981) *Proc. Natl. Acad. Sci.* U.S.A. 78, 2856-2860

Engel, J.D., Sugarman, B.J. and Dodgson, J.B. *Nature* 297, 434-436

Eyre, D.R. (1980) *Science* 207, 1315-1322

Fahrner, K., Yarger, J. and Hereford, L. (1980) *Nucleic Acids Res.* 8, 5725-5738

Federoff, N.V. (1979) *Cell* 16, 697-710

Finnegan, D.J., Rubin, G.M., Young, M.W. and Hogness, D.S. (1977) *Cold Spring Harbor Symp. Quant. Biol.* 42, 1053-1063

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

Flavell, A.J., Ruby, S.W., Toole, J.J., Roberts, B.E. and Rubin, G.M. (1980) *Proc. Natl. Acad. Sci.* U.S.A. 77, 7107-7111

Freigan, N., Marchionni, M. and Kedes, L.H. (1976) *Bio. Bull.* 151, 407 (Abstract)

Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1979) *Nature* 279, 598-603

Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1980) *Cell* 19, 959-972

Fuchs, E. and Green, H. (1979) *Cell* 17, 573-582

Furuichi, Y., La Fiandra, A. and Shatkin, A.J. (1977) *Nature* 266, 235-239

Fryberg, E.A., Kindle, K.L., Davidson, N. and Sodja, A. (1980) *Cell* 19, 365-378

Gall, J.G. (1968) *Proc. Natl. Acad. Sci.* U.S.A. 60, 553-560

Gall, J.G., Stephenson, E.C., Erba, H.P., Diaz, M.O. and Bersacchi-Peilone, G. (1981) *Chromosoma* 84, 159-172

Gallwitz, D. and Mueller, G.C. (1969) *J. Biol. Chem.* 244, 5947-5952

Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J.P., Benoist, I.C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, G. and Chambon, P. (1979) *Nature* 278, 428-434

Geraci, G., Lanciere, M., Marchi, P. and Noviello, L. (1979) *Cellular Differentiation* 81, 187-194

Goldberg, M.L. (1979) Ph.D. Thesis, Stanford University

Goldsmith, M.R. and Basehoar, G. (1978) *Genetics* 90, 291-310

Grantham, R., Gautier, C. and Gouy, M. (1980) *Nucleic Acids Res.* 8, 1893-1909

Greene, P.M., Beslach, M., Goodman, H.M. and Boyer, H.W. (1974) *Methods in Mol. Biol.* 7, 87-111

Grippo, P., Iaccarino, M., Parisi, E. and Scarano, E. (1968) *J. Mol. Biol.* 36, 195-208

Groppi, V.E. and Coffino, P. (1980) *Cell* 21, 195-204

Gross, K.W., Jacobs-Lorena, M., Baglioni, C. and Gross, P.R. (1973) *Proc. Natl. Acad. Sci.* U.S.A. 70, 2614-2618

Gross, K., Probst, E., Schaffner, W., and Birnstiel, M.L. (1976a) *Cell* 8, 455-469

Gross, K., Schaffner, W., Telford, J. and Birnstiel, M.L. (1976b) *Cell* 8, 479-484

Gross, P.R. (1967) *Curr. Topics Develop. Biol.* 2, 1-46

Grosschedl, R. and Birnstiel, M.L. (1980) *Proc. Natl. Acad. Sci.* U.S.A. 77, 1432-1436

Grunstein, M. (1978) *Proc. Natl. Acad. Sci.* U.S.A. 75, 4135-4139

Grunstein, M.L. and Hogness, D.S. (1975) *Proc. Natl. Acad. Sci.* U.S.A. 72, 3961-3965

Grunstein, M., Schedl, P. and Kedes, L.H. (1976) *J. Mol. Biol.* 104, 351-369

Grunstein, M., Diamond, K.E., Knoppel, E. and Grunstein, J. (1981) *Biochem.* 20, 1216-1223

Gruss, P. and Khoury, G. (1980) *Nature* 286, 634-637

Grudon, J.B. (1962) *Develop. Biol.* 4, 256-273

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

Hackett, P.B., Traub, P. and Gallwitz, D. (1978) *J. Mol. Biol.* 126, 619-635

Halbreich, A., Pajot, P., Foucher, M., Grandchamp, C. and Slonimski, P. (1980) *Cell* 19, 321-329

Hamer, D.H. and Leder, P. (1979) *Cell* 17, 737-747

Hamer, D.H., Smith, K.D., Boyer, S.H. and Leder, P. (1979) *Cell* 17, 725-735

Harbers, K., Schneike, A., Struhlmann, H., Jahner, D. and Jaenisch, R. (1981) *Proc. Natl. Acad. Sci.* U.S.A. 78, 7609-7613

Harris, S.E., Rosen, J.M., Mears, A.R. and O'Malley, B.W. (1975) *Biochem.* 14, 2072-2081

Harvey, R.P. and Wells, J.R.E. (1979) *Nucleic Acids Res.* 7, 1787-1798

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

Haynes, S.R. and Jelinek, W.R. (1981) *Proc. Natl. Acad. Sci.* U.S.A. 78, 6130-6134

Hedgepeth, J., Goodman, H.M. and Boyer, H.W. (1972) *Proc. Natl. Acad. Sci.* U.S.A. 69, 3448-3452

Heintz, N., Zernik, M. and Roeder, R.G. (1981) *Cell* 24, 661-668

Hentschel, C., Irminger, J-C., Bucher, R. and Birnstiel, M.L. (1980) *Nature* 285, 147-151

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

Hereford, L.M., Fahrner, K., Woodford, J. and Rosbash, M. (1979) *Cell* 18, 1261-1271

Hereford, L.M., Osley, M.A., Ludwig, J.R. III and McLaughlin, C.S. (1981) *Cell* 24, 367-375

Hicks, J., Strathern, J.N. and Klar, A.J.S. (1979) *Nature* 282, 478-483

Hohn, B. (1975) *J. Mol. Biol.* 98, 93-106

Hohn, B. and Murray, K. (1977) *Proc. Natl. Acad. Sci.* U.S.A. 74, 3259-3263

Hood, L., Campbell, J.A. and Elgin, S.C.R. (1975) *Ann. Rev. Genet.* 9, 305-353

Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) *J. Mol. Biol.* 132, 289-306

Hozumi, N., and Tonegawa, S. (1976) *Proc. Natl. Acad. Sci.* U.S.A. 73, 3628-3632

Hubscher, U., Lutz, H. and Kornberg, A. (1980) *Proc. Natl. Acad. Sci.* 77, 5097-5101

Huez, G., Marbaix, G., Hubert, E., Leclercq, M., Nudel, U., Soreq, H., Salmon, R., Lebleu, B., Revel, M. and Littauer, U.Z. (1974) *Proc. Natl. Acad. Sci.* U.S.A. 71, 3143-3146

Hunt, J.A. (1974) *Biochem. J.* 138, 499-510

Illmensee, K. (1972) *Wilhelm Roux Archiv.* 170, 267

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

Jacob, E. (1976) *Eur. J. Biochem.* 65, 275-284

Jacob, E., Malacinski, G. and Birnstiel, M.L. (1976) *Eur. J. Biochem.* 69, 45-54

Jeffreys, A.J. and Flavell, R.A. (1977) *Cell* 12, 1097-1108

Jones, C.W. and Kafatos, F.C. (1980) *Cell* 22, 855-867

Jones, P.A. and Taylor, S.M. (1980) *Cell* 20, 85-93

Kedes, L.H. (1976) *Cell* 8, 321-331

Kedes, L.H. (1979) *Ann. Rev. Biochem.* 48, 837-870

Kedes, L.H. and Gross, P. (1969a) *J. Mol. Biol.* 42, 559-575

Kedes, L.H. and Gross, P. (1969b) *Nature* 223, 1335-1339

Kedes, L.H. and Birnstiel, M.L. (1971) *Nature New Biol.* 230, 165-169

Kemp, D.J., Cory, S. and Adams, J.M. (1979) *Proc. Natl. Acad. Sci.* U.S.A. 76, 4627-4631

Kinniburgh, A.J. and Ross, J. (1979) *Cell* 17, 915-921

Konkel, D.A., Tilghman, S.M. and Leder P. (1978) *Cell* 15, 1125-1132

Korn, L. and Brown, D. (1978) *Cell* 15, 1145-1156

Kozak, M. (1978) *Cell* 15, 1109-1123

Krieg, P.A. (1980) Ph.D. Thesis, Adelaide University

Krieg, P.A., Robins, A.J., Gait, M.J., Titmas, R.C. and Wells, J.R.E. (1982) *Nucleic Acids Res.* 10, 1495-1503

Kunkel, N.S. and Weinberg, E.S. (1978) *Cell* 14, 313-326

Kunkel, N.S., Hemminiki, K. and Weinberg, E.S. (1978) *Biochem.* 17, 2591-2598

Lathe, R., Buc, H., Lecocq, J.P. and Bautz, E.K.F. (1980) *Proc. Natl. Acad. Sci.* U.S.A. 77, 3548-3553

Lauer, J., Shen, C-K.J. and Maniatis, T. (1980) *Cell* 20, 119-130

Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) *Cell* 15, 1157-1174

Leder, P., Tiemeier, D. and Enquist, L. (1977) *Science* 196, 175-177

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980) *Nature* 283, 220-224

Levy, S., Childs, G. and Kedes, L.H. (1978) *Cell* 15, 157-162

Lichtler, A.C., Detke, S., Phillips, I.R., Stein, G.S. and Stein, J.L. (1980) *Proc. Natl. Acad. Sci.* U.S.A. 77, 1942-1946

Lichtler, A.C., Sierra, F., Clark, S., Wells, J.R.E., Stein, J.L. and Stein, G.S. (1982) *Nature* 298, 195-198

Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977) *Cold Spring Harb. Symp. Quant. Biol.* 42, 1047-1051

Lowry, D.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951) *J. Biol. Chem.* 193, 265-275

Maden, B.E.H. (1978) *Nature* 272, 580-581

Maizels, N.(1976) *Cell* 9, 431-438

Mandel, M. and Higa, A. (1970) *J. Mol. Biol.* 53, 159-162

Mandel, J.L. and Chambon, P. (1979) *Nucleic Acids Res.* 7, 2081-2103

Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) *Proc. Natl. Acad. Sci.* U.S.A. 72, 1184-1188

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C. and Quon, D. (1978) *Cell* 15, 687-701

Manning, R.F. and Gage, L.P. (1978) *J. Biol. Chem.* 253, 2044-2052

Mathis, D., Oudet, P. and Chambon, P. (1980) *Prog. Nucleic Acids Res. and Mol. Biol.* 24, 1-54

Max, E.E., Seidman, J.G. and Leder, P. (1979) *Proc. Natl. Acad. Sci.* U.S.A. 76, 3450-3454

Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci.* U.S.A. 74, 560-564

Maxam, A.M. and Gilbert, W. (1979) *Methods in Enzymology* 65, 499-560

McClintock, B. (1967) *Develop. Biol. Suppl.* 1, 84

McGhee, J.D. and Ginder, G.D. (1979) *Nature* 280, 419-421

McGhee, J.D. and Felsenfeld, G. (1980) *Ann. Rev. Biochem.* 49, 1115-1156

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

McKenzie, S.L., Henikoff, S. and Meselson, M. (1975) *Proc. Natl. Acad. Sci.* U.S.A. 72, 1117-1121

McKeon, C., Ohkubo, H., Pastan, I. and de Crombrugghe, B. (1982) *Cell* 29, 203-210

Melli, M., Spinelli, G., Wussling, H. and Arnold, E. (1977a) *Cell* 11, 651-661

Melli, M., Spinelli, G. and Arnold, E. (1977b) *Cell* 12, 167-174

Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. (1981) *Cell* 27, 279-288

Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nucleic Acids Res.* 9, 309-321

Miki, B.L.A. and Neelin, J.M. (1975) *Can. J. Biochem.* 53, 1158-1169

Miller, J.H. (1978) p66, in The Operon, Eds. Miller, J.H. and Reznikoff, W.S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York)

Mintz, B. and Illmensee, K. (1975) *Proc. Natl. Acad. Sci.* U.S.A. 72, 3585-3589

Molgaard, H.V., Perucho, M. and Ruiz-Carrillo, A. (1980) *Nature* 283, 502-505

Moll, R. and Wintersberger, E. (1976) *Proc. Natl. Acad. Sci.* U.S.A. 73, 1863-1867

Moore, G.D., Procunier, J.D., Gross, D.P. and Grigliatti, T.A. (1979) *Nature* 282, 312-314

Moorman, A.F.M., de Laaf, R.T.M., Destree, O.H.J., Telford, J. and Birnstiel, M.L. (1980) *Gene* 10, 185-193

Moss, B., Gershowitz, A., Weber, L.A. and Baglioni, C. (1977) *Cell* 10, 113-130

Murray, V. and Holliday, R. (1979) *Genet. Res.* 34, 173-188

Neelin, J.M. Callahan, P.X., Lamb, D.C. and Murray, K. (1964) *Can. J. Biochem.* 42, 1743-1752

Nemer, M. and Lindsay, D.T. (1969) *Biochem. Biophys. Res. Commun.* 35, 156-160

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

Noll, M. and Kornberg, R.D. (1977) *J. Mol. Biol.* 109, 393-404

Ohe, Y., Hayashi, H. and Iwai, K. (1979) *J. Biochem.* 85, 615-624

Okada, M., Kleinman, I.A. and Schneiderman, H.A. (1974) *Develop. Biol.* 39, 286-294

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., Dugaiczyk, A. and Woo, S.L.C. (1978) Presented at: Laurentian Hormone Conference, August, 1978

Overton, G.C. and Weinberg, E.S. (1978) *Cell* 14, 247-254

Palmiter, R.D. (1973) *J. Biol. Chem.* 248, 8260-8270

Pardue, M.L., Kedes, L.H., Weinberg, E.S. and Birsntiel, M.L. (1977) *Chromosoma* 63, 135-151

Pelling, C. (1964) *Chromosoma* 15, 71-122

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) *Cell* 20, 555-566

Polisky, B., Green, P., Garfin, D.E., McCarthy, B.J., Goodman, H.M. and Boyer, H.W. (1975) *Proc. Natl. Acad. Sci.* U.S.A. 72, 3310-3314

Pollack, Y., Stein, R., Razin, A. and Cedar, H. (1980) *Proc. Natl. Acad. Sci.* U.S.A. 77, 6463-6467

Portmann, R., Schaffner, W. and Birnstiel, M. (1976) *Nature* 264, 31-34

Powell, B.C., Kemp, D.J., Partington, G.A., Gibbs, P.E.M. and Rogers, G.E. (1976) *Biochem. Biophys. Res. Commun.* 68, 1263-1271

Prescott, D.M. (1966) *J. Cell Biol.* 31, 1-9

Pribnow, D. (1979) in *Biological Regulation and Development* Vol. 1, 250-277 (Ed. Goldberger, R.F., Plenum, New York)

Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214

Ptashne, K. and Cohen, S.N. (1975) *J. Bacteriol.* 122, 776-781

Rae, P.M.M. and Steele, R.E. (1979) *Nucleic Acids Res.* 6, 2987-2995

Razin, A. and Riggs, A.D. (1980) *Science* 210, 604-610

Reeder, R.H. (1979) p.93 in *Genetic Engineering* , J.K. Setlow and A. Hollaender, Eds. (Plenum, New York)

Rickles, R., Marashi, F., Sierra, F., Clark, S., Wells, J., Stein, J. and Stein, G. (1982) *Proc. Natl. Acad. Sci.* U.S.A. 79, 749-753

Risley, M.S. and Eckhardt, R.A. (1981) *Develop. Biol.* 84, 79-87

Roberts, R. (1980) *Nature* 283, 132-133

Robbins, E. and Borun, T.W. (1967) *Proc. Natl. Acad. Sci.* U.S.A. 57, 409-416

Robbins, J., Rosteck, P., Jr., Haynes, J.R., Freyer, G., Cleary, M.L., Kalter, H.D., Smith, K. and Lingrel, J.B. (1979) *J. Biol. Chem.* 254, 6187-6195

Roeder, R.G. (1976) in *RNA Polymerase* R. Losick and M. Chamberlin, Eds. (Cold Spring Harbor Press, Cold Spring Harbor, New York). pp285-329

Rosenberg, M. and Court, D. (1979) *Ann. Rev. Genet.* 13, 319-353

Rubin, G.M. and Sulston, J.E. (1973) *J. Mol. Biol.* 79, 521-530

Ruderman, J.V. and Pardue, M.L. (1978) *J. Biol. Chem.* 253, 2018-2025

Ruderman, J.V., Baglioni, C. and Gross, P.R. (1974) *Nature* 247, 36-38

Ruiz-Vazquez, R. and Ruiz-Carrillo, A. (1982) *Nucleic Acids Res.* 10, 2093-2108

Ryffel, G.V., Wyler, T., Muellener, D.B. and Weber, R. (1980) *Cell* 19, 53-61

Saint, R.B. (1979) Ph.D. Thesis, Adelaide University

Sakano, H., Rogers, J.H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. and Tonegawa, S. (1979) *Nature* 277, 627-633

Salomon, R., Kaye, A.M. and Herzberg, M. (1969) *J. Mol. Biol.* 43, 581-592

Samal, B., Worcel, A., Louis, C. and Schedl, P. (1981) *Cell* 23, 401-409

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci.* U.S.A. 74, 5463-5467

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

Schaffner, W., Gross, K., Telford, J. and Birnstiel, M. (1976) *Cell* 8 471-478

Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978) *Cell* 14, 655-671

Schaffner, W., Topp, W. and Botchan, M. (1979) *Alfred Benson Symp.* 13, 192-204

Scherer, G., Telford, J., Baldari, C. and Pirotta, V. (1981) *Develop. Biol.* 86, 438-447

Schimke, R.T., Alt, F.W., Kellems, R.E., Kaufman, R.J. and Bertino, J.R. (1977) *Cold Spring Harbor Symp. Quant. Biol.* 42, 649

Schmidt, O.A., Mazaitis, A.J., Kasai, R. and Bautz, E.K.F. (1970) *Nature* 225, 1012-1016

Scott, A.C. and Wells, J.R.E. (1976) *Nature* 259, 635-638

Seeburg, P.H., Shine, J., Martial, J., Baxter, J.D. and Goodman, H.M. (1977) *Cell* 12, 157-165

Seiler-Tuyns, A. and Birnstiel, M.L. (1981) *J. Mol. Biol.* 151, 607-625

Seyedin, S.M. and Kistler, W.S. (1979) *Biochem.* 18, 1376-1379

Sharp, P.A. (1981) *Cell* 23, 643-646

Sierra, F., Lichtler, A., Marashi, F., Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G. and Stein J. (1982a) *Proc. Natl. Acad. Sci.* U.S.A. 79, 1795-1799

Sierra, F., Leza, A., Marashi, F., Plumb, M. Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G.S. and Stein, J.L. (1982b) *Biochem. Biophys. Res. Commun.* 104, 785-792

Sittman, D.B., Chiu, I.M., Pan, C.J., Cohn, R.H., Kedes, L.H. and Marzluff, W.F. (1981) *Proc. Natl. Acad. Sci.* U.S.A. 78, 4078-4082

Smith, G.E. and Summers, M.D. (1980) *Anal. Biochem.* 109, 123-129

Smith, G.P. (1976) *Science* 191, 528-535

Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517

Spinelli, G., Gianguzza, F., Casano, C., Acierno, P. and Burckhardt, J. (1979) *Nucleic Acids Res.* 6, 454-460

Spradling, A.C. and Mahowald, A.P. (1980) *Proc. Natl. Acad. Sci.* U.S.A., 77, 1096-1100

Stadler, J., Larsen, A., Engel, J.D., Dolan, M., Groudine, M., and Weintraub, H. (1980) *Cell* 20, 451-460

Stein, R., Razin, A. and Cedar, H. (1982) *Proc. Natl. Acad. Sci.* U.S.A. 79, 3418-3422

Stein, J.L., Stein, G.S. and McGuire, P.M. (1977a) *Biochem.* 16, 2207-2213

Stein, G., Stein, J., Shephard, E., Park, W. and Phillips, I. (1977b) *Biochem. Biophys. Res. Commun.* 77, 245-252

Stein, G.S., Stein, J.L., Park, W.D., Detke, S., Lichtler, A.C., Shephard, E.A., Jansing, R.L. and Phillips, I.R. (1978) *Cold Spring Harbor Symp. Quant. Biol.* 42, 1107-1120

Stephenson, E.C., Erba, H.P. and Gall, J.G. (1981) *Cell* 24, 639-647

Sternberg, N., Tiemeier, D. and Enquist, L. (1977) *Gene* 1, 255-280

Struhlmann, H., Jahner, D. and Jaenisch, R. (1981) *Cell* 26, 221-232

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

Sures, I., Levy, S. and Kedes, L.H. (1980) *Proc. Natl. Acad. Sci.* U.S.A. 77, 1265-1269

Surrey, S. and Nemer, M. (1976) *Cell* 9, 589-595

Sutter, D. and Doerfler, W. (1980) *Proc. Natl. Acad. Sci.* U.S.A. 77, 253-256

Suzuki, Y., Gage, L.P. and Brown, D.D. (1972) *J. Mol. Biol.* 70, 637-649

Suzuki, Y. and Suzuki, E. (1974) *J. Mol. Biol.* 88, 393-407

Szabo, P., Yu, L.C., Borun, T., Varricchio, F., Siniscalco, M. and Prensky, W. (1978) *Cytogenet. Cell. Genet.* 22, 359-363

Tashima, M., Calabretta, B., Torelli, G., Scofield, M., Maizel, A. and Saunders, G.F. (1981) *Proc. Natl. Acad. Sci.* U.S.A. 78, 1508-1512

Taylor, J.M., Illmensee, R. and Summers, J. (1976) *Biochim. Biophys. Acta.* 442, 324-330

Tikchonenko, T.I., Karamov, E.V., Zavizion, B.A. and Naroditsky, B.S. (1978) *Gene* 4, 195-212

Tilghman, S.M. Tiemeier, D.C., Seidman, J.G., Peterlin, B.M., Sullivan, M., Maizel, J.V. and Leder, P. (1978) *Proc. Natl. Acad. Sci.* U.S.A. 75, 725-729

Tsai, M-J., Ting, A.C., Nordstrom, J.L., Zimmer, W., and O'Malley, B.W. (1980) *Cell* 22, 219-230

van Dongen, W., de Laaf, L., Zaal, R., Moorman, A. and Destree, O. (1981) *Nucleic Acids Res.* 9, 2297-2306

Van Helden, P., Strickland, W.N., Brand, W.F. and Von Holt, C. (1978) *Biochim. Biophys. Acta.* 533, 278-281

Van Ommen, G-J.B., Boer, P.H., Groot, G.S.P., De Haan, M., Roosendall, E., Griyell, L.A., Haid, A. and Schweyen, R.J. (1980) *Cell* 20, 173-183

Vanyushin, B.F., Tkacheva, S.G. and Belozersky, A.M. (1970) *Nature* 225, 948-949

Vedeckis, W.V., Schrader, W.T. and O'Malley, B.W. (1978) *Biochemical Action of Hormones* 5, 321

Von Holt, C., Strickland, W.L., Brandt, W.F. and Strickland, M.S. (1979) FEBS *Lett.* 100, 201-218

Waalwijk, C. and Flavell, R.A. (1978) *Nucleic Acids Res.* 5, 4631-4641

Wahl, G.M., Stern, M. and Stark, G.R. (1979) *Proc. Natl. Acad. Sci.* U.S.A. 76, 3683-3687

Wahl, G., Vitto, L., Padgett, R.A. and Stark, G.R. (1982) *Mol. and Cell. Biol.* (Submitted for publication)

Wahli, W. and Dawid, I.B. (1980) *Proc. Natl. Acad. Sci.* U.S.A. 77, 1437-1441

Wallace, H. and Birnstiel, M.L. (1966) *Biochim. Biophys. Acta.* 114, 296-310

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

Wasylyk, B. and Chambon, P. (1981) *Nucleic Acids Res.* 9, 1813-1824

Wei, C.M., Gershowitz, A. and Moss B. (1975)*Cell* 4, 379-386

Weiner, A.M. (1980) *Cell* 22, 209-218

Williams, R.O., Young, J.R. and Majiwa, P.A.O. (1979) *Nature* 282, 847-849

Wilson, A.C., Carlson, S.S. and White, T.J. (1977) *Ann. Rev. Biochem.* 46, 573-639

Wilson, E.B. (1928) *The Cell in Development and Hereditary* (Macmillan, New York, Edn. 3)

Wilson, M.C. and Melli, M.J. (1977) *J. Mol. Biol.* 110, 511-535

Woodland, H.R. (1980) FEBS *Lett.* 121, 1-7

Woodland, H.R. and Adamson, E.D. (1977) *Develop. Biol.* 57, 118-135

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

Wu, M., Holmes, D.S., Davidson, N., Cohen, R.H. and Kedes, L.H. (1976) *Cell* 9, 163-169

Yao, M-C. and Gall, J.G. (1977) *Cell* 12, 121-132

Yu, L.C., Szabo, P., Borun, T.W. and Prensky, W. (1978) *Cold Spring Harbor Symp. Quant. Biol.* 42, 1101-1105

Zellweger, A., Ryser, U. and Braun, R. (1972) *J. Mol. Biol.* 64, 681-691

Zernik, M., Heintz, N., Boime, I. and Roeder, R.G. (1980) *Cell* 22, 807-815

Ziff, E.B. and Evans, R.M. (1978) *Cell* 15, 1463-1475

Zweilder, A.E. (1980) in Gene Families of Collagen and other Structural Proteins, D.J. Prock op and P.C. Champe, Eds. (Amsterdam: Elsevier North Holland)