CHICKEN HISTONE GENE ORGANISATION

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

by

RICHARD JAMES D'ANDREA, B.Sc. (Hons.)

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

February, 1985

CONTENTS

		Page
STATEMENT		i
SUMMARY		ii
ACKNOWLEDGEMENTS		v
ABBREVIATIONS		vi
CHAPTER 1.	**LITERATURE REVIEW**	1
1.1	INTRODUCTION	2
1.2	EVOLUTION OF GENE SYSTEMS	2
1.2.1	General Homologous Recombination	3
1.2.2	Gene Conversion	5
1.2.3	Concerted Evolution of Gene Families	6
1.2.4	Transposable Elements	8
1.2.5	Illegitimate Recombination	10
1.2.6	Transposition by Reverse-Transcription of mRNA	11
1.2.7	Horizontal Gene Transfer	11
1.2.8	Secondary Structure and DNA Rearrangement	12
1.2.9	Directed Gene Arrangements	13
1.2.10	Genome Alterations: Effects on Gene Expression	14
1.3	GENE EXPRESSION	15
1.3.1	Initiation of Transcription	16
1.3.2	Termination of Transcription	17
1.3.3	Control of Gene Expression	20
1.3.4	Factors Affecting Gene Expression	25
1.4	CHROMATIN	31
1.4.1	Nucleosome Structure	31
1.4.2	Active Chromatin	32
1.4.3	Chromatin Structure: Repression of Gene Activity	34
1.5	HISTONES	37
1.5.1	Histone Proteins	37
1.5.2	Histone Modifications	37
1.5.3	Histone Structural Variants	38
1.6	HISTONE GENES : ORGANISATION	40
1.6.1	Sea Urchin	40
1.6.2	*Drosophilia*	42
1.6.3	*Xenopus*	43
1.6.4	Newt	44
1.6.5	Yeast	44
1.6.6	Chicken	45
1.6.7	Mammals	46
1.6.8	Orphons	47
1.6.9	Histone Gene Organisation: Summary	47

			Page
1.7	HISTONE GENES : EXPRESSION		49
	1.7.1	Cell-Cycle Regulation	49
	1.7.2	Developmental Regulation	51
	1.7.3	Introns	54
CHAPTER 2.	**MATERIALS AND METHODS**		55
2.1	MATERIALS		56
	2.1.1	Chemicals and Reagents	56
	2.1.2	Enzymes	56
	2.1.3	Nucleic Acids	57
	2.1.4	Bacterial Strains	58
	2.1.5	Growth Media	58
	2.1.6	Antibiotics	59
	2.1.7	Isotopically Labelled Compounds	59
	2.1.8	Distilled Water	59
	2.1.9	Buffer Saturated Phenol	59
	2.1.10	Glassware and General Equipment	60
2.2	METHODS		60
	2.2.1	Restriction Enzyme Digestion	60
	2.2.2	Routine Gel Electrophoresis	61
	2.2.3	Transfer of DNA to Nitrocellulose and Hybridisation with a Labelled Probe	62
	2.2.4	Purification by Sucrose Gradient Velocity Centrifugation	63
	2.2.5	Isolation of Clones from a λ-Recombinant Genomic Library	63
	2.2.6	Construction of a Cosmid Recombinant Library	65
	2.2.7	Construction of Recombinant DNA	68
	2.2.8	Transformation of *E. coli*	70
	2.2.9	Detection and Examination of Recombinant Plasmid/Cosmid Clones	70
	2.2.10	Large-Scale Preparation of Recombinant Plasmid/Cosmid DNA	71
	2.2.11	Preparation of *In Vitro* Labelled DNA	73
	2.2.12	Sequence Analysis of DNA	74
	2.2.13	Containment Facilities	77
CHAPTER 3.	**CONSTRUCTION AND CLONING OF CORE HISTONE GENE SPECIFIC PROBES**		78
3.1	INTRODUCTION		79
3.2	RESULTS		80
	3.2.1	Isolation of Coding Specific Fragments	80
	3.2.2	Ligation of *Hin*dIII Recognition Site	81
	3.2.3	Ligation of Linkered Fragments into pBR322 Plasmid DNA	81
	3.2.4	Transformation, Selection and Screening of Recombinants	82
	3.2.5	Discussion	83

Page

CHAPTER 4. ISOLATION OF HISTONE GENE RECOMBINANTS FROM BACTERIOPHAGE LIBRARIES 84

4.1 INTRODUCTION 85

4.2 RESULTS 86
4.2.1 Screening of the Chicken Genomic Library 86
4.2.2 The 2.0 kb *Eco*RI-*Sma*I Fragment Contains an H1 gene 89
4.2.3 Isolation of Clones Overlapping λCHO1 90
4.2.4 Discussion 91

CHAPTER 5. CHARACTERISATION OF CHICKEN HISTONE GENOMIC CLONES 92

5.1 INTRODUCTION 93

5.2 RESULTS 93
5.2.1 Preparation of Bacteriophage DNA 93
5.2.2 Restriction Analysis of λCHO3 94
5.2.3 The Coding Potential of λCHO3 96
5.2.4 Isolation of λCHO3 Subclones 96
5.2.5 Restriction Analysis of λCHO3 Subclones 97
5.2.6 Restriction Analysis of λCHO5 100
5.2.7 The Coding Potential of λCHO5 101
5.2.8 Isolation of a λCHO5 Subclone 102
5.2.9 Restriction Analysis of pCH7.0E 102
5.2.10 Search for Clones Extending λCHO3 103
5.2.11 Search for Clones Extending λCHO5 103
5.2.12 Restriction Analysis of λCHO7 104
5.2.13 Coding Potential of λCHO7 104
5.2.14 Isolation of λCHO7 Subclones 105
5.2.15 Restriction Analysis of pCH3.5E 105
5.2.16 Search for Clones Extending λCHO7 106
5.2.17 Other Histone-Gene Containing Genomic Clones 106
5.2.18 Discussion 106

CHAPTER 6. CONSTRUCTION, ISOLATION AND CHARACTERISATION OF CHICKEN HISTONE COSMID RECOMBINANTS 108

6.1 INTRODUCTION 109

6.2 RESULTS 110
6.2.1 Preparation of High Molecular Weight Chicken DNA 110
6.2.2 Partial Digestion of Chicken DNA 110
6.2.3 Size Fractionation of *Sau*3AI Partially Digested Chicken DNA 111
6.2.4 Preparation of Cosmid Vector 112
6.2.5 Ligation of Chicken DNA to pHC79 Vector DNA 113

Page

6.2.6 *In Vitro* packaging and Cosmid Infection 113
6.2.7 Replication and Screening of Cosmid Recombinants 115
6.2.8 Preliminary Characterisation of Cosmid Recombinants 116
6.2.9 Probing for Cosmid Recombinants Overlapping λCHO3 and λCHO7 117
6.2.10 Restriction Analysis of Cosmid 6.3C 117
6.2.11 The Coding Potential of 6.3C 118
6.2.12 Construction of 6.3C Subclones 120
6.2.13 Restriction Analysis of pCH11.5E 120
6.2.14 Partial Sequencing of Isolated Restriction Fragments 121
6.2.15 Restriction Analysis of Cosmid Clones 2.1C and 5.1C 122
6.2.16 Discussion 123

CHAPTER 7. FINAL ANALYSIS OF HISTONE GENE RECOMBINANTS 125

7.1 INTRODUCTION 126

7.2 RESULTS 126
7.2.1 Relationship of Core and H1 Histone Gene Containing Regions to Each Other 126
7.2.2 Relationship of Core and H1 Histone Gene Containing Regions to 2.1C and 5.1C 127
7.2.3 Relationship of the Histone H5 Gene to Core and H1 Histone Genes 128
7.2.4 The Region Between the Histone Genes 129
7.2.5 Discussion 133

CHAPTER 8. FINAL DISCUSSION 136

8.1 ORGANISATION OF CHICKEN HISTONE GENES 137
8.1.1 Gene Number 137
8.1.2 Gene Organisation 138
8.1.3 Preferred Associations of Histone Genes 139
8.1.4 Restriction Enzyme Site Polymorphisms 141
8.1.5 Spacer DNA 142

8.2 EXPRESSION OF THE CHICKEN HISTONE GENES 143

8.3 EVOLUTION OF THE CHICKEN HISTONE GENES 145
8.3.1 Homogeneity 145
8.3.2 Symmetry Within Histone Gene Clusters 146
8.3.3 Secondary Structure - Induced Recombination 149
8.3.4 Transposition Associated Recombination 150
8.3.5 Cellular Processes and the Rearrangement of Histone Genes 151

8.4 FINAL SUMMARY 153

BIBLIOGRAPHY 155

APPENDIX

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge, it contains no material that has been previously published by any other person, except where due reference is made in the text.

RICHARD JAMES D'ANDREA

SUMMARY

The work discussed in this thesis involves analysis of the organisation of the core and H1 histone genes in the chicken.

1. DNA fragments specific for the coding regions of either the H_2A, H_2B, H3 or H4 genes were isolated from a previously characterised recombinant, λCHO1, and subcloned into pBR322. An H1 specific probe was provided by L.S. Coles (this laboratory).

2. A chicken genomic library was screened with ^{32}P-labelled core and H1 histone gene sequences and a bank of histone gene containing recombinants isolated.

Non-histone coding probes, from each end of the previously characterised recombinant λCHO1, were used to select overlapping clones λCHO3 and λCHO5.

3. Genomic clones, λCHO3 and λCHO5, and subclones derived from them were characterised by restriction enzyme and hybridisation analysis using the histone gene-specific probes. Fine restriction enzyme and Southern analysis allowed the number and position of each gene to be determined.

Further "chromosome crawling" with non-histone coding probes from each of these clones led to the isolation of λCHO7 (overlapping λCHO5) and confirmed the identification of other clones, already characterised, which extend λCHO3. This region, spanning 50 kb, contained thirteen histone genes and was covered by six λ-recombinants.

4. To aid in further analysis of the histone gene distribution in the chicken genome a cosmid genomic library (1 genome equivalent) was constructed in *E. coli* strain HB101. High molecular weight chicken erythrocyte DNA partially digested with *Sau*3AI, to an average size of 35-45 kb, was inserted into *Bam*HI digested cosmid arms prepared from

the plasmid pHC79. This library was screened using the histone gene-specific probes. Five different cosmid recombinants were isolated, three of which (6.3C, 2.1C and 5.1C) were fully characterised and are described in this thesis.

Cosmid 6.3C extended the region already characterised by 20 kb (eight further genes) giving a continuous genomic region of approximately 70 kb spanning twenty-one histone gene regions.

The genes in this region are arranged in clusters of 2-8. These clusters are separated by non-histone coding regions of 2.5-8 kb. It is clear from this data that no long-range histone gene repeat exists at this locus. However, the organisation of the chicken histone genes within this region is not totally random. H_2A genes are closely associated with H_2B genes and in these cases each gene pair is divergently transcribed. In addition, there is a non-random association of H1 genes with H_2A and H_2B genes. The fact that there are exceptions to these associations (i.e. H_2A occurs between H3 and H4 in λCH03) suggests that the expression of each individual gene is independent of the precise linkage.

Detailed restriction mapping of subclones containing individual gene clusters revealed regions containing symmetrical enzyme sites. In each case histone genes appear as an inverted duplication centered around H3 genes, viz. pCH8.4E, H4.H_2A.H3.H_2A.H4 and pCH11.5E, H1.H_2A.H_2B.H3.H3.H_2B.H_2A.H1. Analysis of sequence data from pCH8.4E indicated that in this case a 2.1 kb region containing H_2A and H4 genes is almost exactly duplicated in the reverse orientation. The boundaries of the duplication are characterised by a 10bp direct repeat at the H4 end and a closely related 10bp inverted repeat at the H_2A end. These repeats may represent the sites of recombination events which generated the symmetrical structure.

5. All the characterised recombinants with chromosomal, histone gene-containing inserts were probed for non-histone encoding regions using cDNA made from 5-day chick embryo poly (A) plus RNA as a probe. Only one cDNA positive region was detected and this corresponded to a region spanned by overlapping cosmid recombinants 2.1C and 5.1C. These recombinants also contain a single H3 gene (subsequently identified as a split H3 gene reported previously by Engel *et al*, 1982).

DNA from the histone-containing recombinants was also hybridised to *in vitro* labelled total chicken DNA and fragments containing repeat-sequence DNA were detected by autoradiography. Further hybridisation analysis with specific probes identified two major unrelated repeats interspersed with the histone genes. A third unknown repeated DNA sequence is represented only once in the regions analysed.

6. Work described in this thesis together with other results from this laboratory has allowed an extensive analysis of the chicken histone gene arrangement. In all, there are forty-two core and H1 histone genes detected in four independent regions, and it is likely that these account for the majority of the core and H1 histone genes in the chicken genome.

ACKNOWLEDGEMENTS

I wish to thank Professor W.H. Elliot for the opportunity of undertaking this work in his department. I also extend my sincere thanks to Dr. Julian Wells for his friendly and enthusiastic supervision.

My grateful thanks also to the past and present members of labs 20 and 20A and in particular Leeanne Coles, Alan Robins, Rick Sturm and Peter Wigley for their critical discussion, useful technical advice and friendship throughout the course of my experimental work.

My sincere thanks to Lesley Crocker and Jenny Brazier for the excellent diagrams and to Jutta Waiblinger (Suzy's Secretarial Service) for her excellent typing of this manuscript.

Finally I am indebted to Sue Williams for her constructive advice, patience and encouragement during the preparation of this thesis and to my parents for their support and understanding.

During the course of this work I was supported by a Commonwealth Post-graduate Research Award.

ABBREVIATIONS

A_n	:	optical obsorbance measured at wavelength *n* over a pathlength of 1cm.
amp^r	:	resistance to ampicillin
BCIG	:	5-bromo-4-chloro-3-indolyl-β-D-galactoside
bis	:	N,N'-methyl-bisacrylamide
bp	:	nucleotide base pairs
cap^r	:	resistance to Chloramphenicol
cDNA	:	DNA complementary to RNA
Ci	:	Curie
DMSO	:	Dimethylsulphoxide
DNA	:	deoxyribonucleic acid
DNA'se	:	deoxyribonuclease
DTT	:	dithiothreitol
EDTA	:	ethylenediaminetetracetic acid
IPTG	:	isopropyl-β-D-thio-galactopyranoside
mRNA	:	messenger RNA
oligo(dT)	:	oligo (thymidylic acid)
PEG	:	polyethylene glycol
poly (A)	:	poly (adenylic acid)
RNA	:	ribonucleic acid
RNA'se	:	ribonuclease
S	:	sedimentation coefficient
SDS	:	sodium dodecyl sulphate
TEMED	:	N, N, N', N'-tetramethylethylenediamine
tet^r	:	resistance to tetracycline
Tris	:	tris (hydroxymethyl) aminomethane
UV	:	ultraviolet

CHAPTER 1

LITERATURE REVIEW

1.1 INTRODUCTION

The work in this thesis is concerned with the structural organisation of a eukaryotic multigene family; specifically the histone gene family in chicken. Characterisation of the organisation of this, or any, gene family is fundamental to studies of the mechanisms affecting its co-ordinate expression. The aim of this work was to determine the organisation of the chicken histone gene family and to investigate any aspects of gene organisation relevant to control of gene expression.

Although there is considerable information concerning factors influencing expression of individual genes, it is of great interest to extend this knowledge to include the co-ordinate regulation of gene families, and the programming of expression throughout the development of multicellular organisms. Recent progress in these fundamental areas of eukaryote biology is discussed in sections 1.3 and 1.4. The histone gene systems, their expression and regulation are given particular emphasis in sections 1.5, 1.6 and 1.7.

Once the spatial arrangement of the genes in a family is determined, one can speculate, not only on the contribution of gene arrangement to co-ordinate expression, but also on how the gene arrangement has evolved. It is now clear that there are many processes which lead to alterations of genome structure and consequently contribute to the rearrangement of genes. Novel gene arrangements are subject to natural selection, and it is therefore possible to speculate on the possible advantages of particular arrangements. In the following section the known mechanisms which contribute to alterations of genome structure are discussed.

1.2 EVOLUTION OF GENE SYSTEMS

Genome re-arrangements can occur in a variety of ways, all involving recombination events, i.e. the cleavage of DNA molecules and

their rejoining in different combinations. The different pathways for recombination are summarized in this section.

1.2.1 General Homologous Recombination

Two duplexes of DNA with extended base sequence homology can recombine within those regions to yield reciprocal recombinant duplexes. Joint molecules involving heteroduplex regions formed from strands of each parent molecule are intermediates in the recombination event (Meselson and Radding, 1975). The lower limits of homology required for this form of recombination are not well defined although Rubnizt and Subramani (1984) have shown that a low frequency of homologous recombination can be observed with as little as 14 bp homology. This form of recombination is stimulated in the vicinity of certain sequences which do not need to be located in the homologous segments. These include *Chi* sequences of *E. coli* (Stahl, 1979; Smith, 1983) and *cog* sequences of *Neurospora* (Catcheside, 1974; Catcheside, 1977; Fincham *et al*, 1979). These regions may represent binding sites for recombination proteins or sites of endonuclease nicking (nicking creates free ends which are capable of invading another duplex).

Recently the nature of some of the proteins and biochemical steps involved in general homologous recombination has been determined (see Dressler and Potter, 1982). The *rec*A gene product is required in *E. coli* for all general homologous recombination and has now been studied in considerable detail (Clarke, 1973; Howard-Flanders *et al*, 1984). Little is known of the processes involved in this type of recombination in eukaryotes.

The presence of dispersed repeat DNA in eukaryotic genomes indicates that chromosomes contain multiple substrates for homologous recombination. Recombination between related sequences at different locations in the genome can generate numerous rearrangements (Jeffreys and Harris, 1982). If intrachromosomal recombination occurs between

homologous sequences, which are in the same orientation on a single chromosome arm, a deletion will result. Short repeats in opposite orientations can lead to localised inversions. This type of event may have generated the lesion observed in one form of thalassaemia in man (Jones *et al*, 1981). A duplication can result from interchromosomal, unequal crossing-over between repeats on homologous chromosomes. With this type of event the size of the duplication is dependent upon the spacing of the repetitive element. Shen *et al* (1981) found a short direct repeat at each end of the duplicated segment of DNA containing the γ-globin genes. The two closely related fetal globin genes, γ^G and γ^A have clearly arisen from a recent duplication event. The simplest hypothesis for this duplication involves mispairing and unequal crossing-over between short repeated elements on homologous chromosomes (Jeffreys and Harris, 1982). Once a segment is duplicated in tandem by a process such as this it provides an excellent substrate for unequal crossing-over, and can rapidly amplify to form an array of many closely related genes, such as those found in ribosomal DNA clusters, tandemly repeated histone gene clusters and satellite DNA's. Unequal crossing-over between daughter DNA duplexes (or sister chromatids) after replication has also been proposed as a mechanism for generating tandem duplications (Weisberg and Adhya, 1977).

Recombination between homologous sequences on heterologous chromosome arms, oriented in the same direction can result in reciprocal translocation.

Thus, the prevelance of repeat-sequence DNA in the eukaryote genome suggests rearrangements of DNA mediated by general homologous recombination could be quite common. Of all the rearrangements generated, natural selection will eliminate those which are unfavourable and select for those which increase fitness. Any neutral genetic changes will be either fixed or lost by random genetic drift.

1.2.2 Gene Conversion

Studies in fungal genetic systems indicated that gene segregation was not always reciprocal (Catcheside, 1977; Fincham *et al*, 1979). Non-Mendelian meiotic segregation is sometimes observed such that in heterozygotes the alleles A and a are distributed 3A:1a or 1A:3a instead of the usual 2A:2a. The term "conversion" was initially used by Winkler (1930) to indicate this apparent conversion of one allele to another. Conversion has been shown to occur at meiosis and mitosis and in both cases approximately 50% of the conversion events are associated with reciprocal recombination in the immediate vicinity. (Jackson and Fink, 1981; Catcheside, 1977). Jackson and Fink (1981) have demonstrated intrachromosomal gene conversion in yeast and have shown that between a pair of closely linked related genes, conversion is more common than reciprocal recombination. Evidence for gene conversion in higher eukaryotes has been obtained from DNA sequence analysis for several families of repeated genes (Shen *et al*, 1981; Slightom *et al*, 1980; Baltimore, 1981; Weiss *et al*, 1983), including the late histone genes of the sea urchin (Roberts *et al*, 1984).

The most popular theories for general recombination take into account the observed association between gene conversion and reciprocal recombination and consider the two as consequences of one event. The models postulate the formation of heteroduplex DNA containing a single-stranded section of DNA from each parental chromatid. Models of this type were first proposed by Whitehouse and Hastings (1965) and Holliday (1964) and have since been extended (see Catcheside, 1977; Szostak *et al*, 1983; Whitehouse, 1983).

Recently however, there have been suggestions that gene conversion may not always be associated with reciprocal recombination. Fink and Petes (1984) suggested the possibility that conversion events occurring between homologous chromosomes were associated with reciprocal

recombination of flanking genes, while intrachromosomal conversion events were not. Alternatively Klar and Strathern (1984) argue that the extent of the homologous region determines whether a conversion event will lead to reciprocal exchange. It is now unclear what the relationship between the two events is, however it seems that conversion unassociated with reciprocal recombination can occur at mitosis and meiosis (Fink and Petes, 1984).

1.2.3 Concerted Evolution of Gene Families

Once gene families were discovered, it became clear that mechanisms were operating to keep sequences of related family members very similar. The term "concerted evolution", coined by Zimmer *et al* (1980), refers to the apparent interchange of sequences between related regions of the genome maintaining close sequence homology between two related loci. This form of evolution has now been documented in a variety of gene families and repetitive DNA sequences (reviewed in Dover, 1982; Baltimore, 1981; Weiner and Denison, 1982).

Gene conversion and unequal crossing-over are both forms of general recombination and both have been implicated in the maintenance of sequence homogeneity in gene families.

There is an important contrast between unequal crossing-over and gene conversion. Unequal crossing-over can only occur among tandemly arranged genes of a family and results in a reciprocal increase and decrease in gene number on different chromosomes. Gene conversion can homogenise dispersed genes (as well as those arranged in tandem) and is predicted to be the most frequent form of recombination affecting members of dispersed gene families (Baltimore, 1981; Jackson and Fink, 1981; Shen *et al*, 1981; Weiner and Denison, 1982). The tyrosine tRNA genes in yeast (see Weiner and Denison, 1982) and the late histone genes of sea urchin (Roberts *et al*, 1984) are examples of dispersed but homogeneous multigene families and provide good evidence for gene

conversion as an efficient homogenising mechanism.

In sea urchin, a number of distinct families of early histone genes exist (Holt and Childs, 1984), and it is of interest to know how these have been maintained independently. Holt and Childs (1984) propose that the homogenisation of genes within but not between arrays is a consequence of gene conversion processes because these occur more readily between closely linked sequences (see Roeder *et al*, 1984). Thus a mechanism for keeping the members of a tandem array of the same genes nearly identical while this group of genes remains independent of another related family of genes is to have the two gene families on separate chromosomes (Coen and Dover, 1983). In this situation recombination events between genes of the separate families will be rare.

Unequal Crossing Over

Smith (1976) proposed that multiple occurrences of unequal crossing-over within a tandem array could maintain homogeneity of repeating units. Any variant that arises within a tandem array may be either fixed or lost by this process. This form of recombination has been shown to be sufficient to maintain the sequence homogeneity of the ribosomal RNA repeat unit in yeast (Szostak and Wu, 1980) and has been invoked as an explanation for the high degree of homogeneity observed between histone gene tandem arrays (Coen *et al*, 1982; Dover, 1982; Hentschel and Birnstiel, 1981; Kedes, 1979). Recurrent unequal exchange within a tandem array has the effect of expanding and contracting the copy number of the family and may explain the striking differences in structure and repetition frequency of histone gene clusters observed between closely related species (Maxson *et al*, 1983a).

Without a mechanism to effect homogenisation, multigene families could probably not exist as functional units. Given the inevitability of mutation most genes within a multigene family would accumulate mutations and diverge independently. A mechanism for efficient

homogenisation between members of moderately and highly repeated gene families reduces such a multigene family to the equivalent of a single gene locus with respect to natural selection.

Directed Homogenisation

It should be noted that persistent, non-random exchange of genetic information can lead to directed homogenisation of a gene family. Dover (1982) defined molecular drive as the fixation of variants in a population as a result of directional processes. Bias has been observed for conversion at mutant sites in fungi (see Catcheside, 1977; Fincham *et al*, 1979). It is possible that in some situations there is a very strong preference for the conversion of one allele by another. In fact there is suggestive evidence that the invading strand in a heteroduplex is preferentially favored during mismatch repair (Markham and Whitehouse, 1982). Replicative transposition (discussed in section 1.2.4) can also lead to directed homogenisation. A sequence which has an optimal transposition frequency can duplicate itself readily, and will have a high probability of spreading in a genotype and becoming fixed in a population (see Dover, 1982).

1.2.4 Transposable Elements

Early studies by McClintock (1950, 1951) demonstrated the presence of genetic determinants in maize which move from one chromosome locus to another, altering the regulation of gene activity and leading to drastic alterations of chromosome organisation, including inversions, duplications and deletions. Further findings, that most spontaneous mutations in bacteria result from insertion of DNA elements (Shapiro, 1969) and that drug resistance elements spread through bacterial populations by insertion into different plasmids (Hedges and Jacob, 1974), indicated clearly that mobile DNA insertion elements were significant in creating diversity in prokaryotes. Further studies in bacteria showed that these insertion elements also promote chromosome rearrangements

(see Bukhari *et al*, 1977). It now appears that transposable elements are ubiquitous and play a significant role in the evolution of chromosome primary structure. A number of insertion elements are now well characterised at the molecular level (see Shapiro, 1982; Bukhari *et al*, 1977; Cold Spring Harbour Laboratory, 1981). Some of the features of the unique mode of recombination proposed for transposition are given below.

Replicative Recombination

Recombination generally does not involve the net synthesis of DNA, although some single stranded degradation and resynthesis may occur. However when a region of DNA containing a transposable element recombines with a second target site, a short segment of DNA from the new target site and the element itself are duplicated (Bukhari *et al*, 1977; Cold Spring Harbour Laboratory, 1981; Shapiro, 1982; Cohen and Shapiro, 1980).

A model for this replicative recombination has been proposed (Cohen and Shapiro, 1980; Shapiro, 1979) and predicts the existence of recombinant intermediates in which two unrelated sequences are linked by a copy of the transposable element and the target sequence (see Cohen and Shapiro, 1980). Full transposition requires resolution of these intermediates so that parental sequences are now linked by the transposable element.

The intermediate structures which are predicted will explain all the known re-arrangements which are associated with transposition (Cohen and Shapiro, 1980). Any recombinant structures generated by this process will contain a copy of the transposable element and the target oligonucleotide at the junction between previously unconnected sequences.

Transposable elements, then, are specific segments of DNA capable of repeated insertion into many locations in a genome. These elements can cause the reorganisation of cellular DNA in a number of ways;

a) They specifically duplicate themselves and insert into another

region of the genome while remaining at their original location.

b) They serve as a specific site for recombination events and can lead to a variety of structural rearrangements.

c) They can enclose a unique sequence and form a larger transposon; which can then be duplicated, via the proposed replicative mechanism discussed above, to generate a dispersed repeat sequence.

Transposon associated genome rearrangements involving histone genes have not been reported as yet, however insertion of transposable elements into or near histone genes has been reported for *Drosophila* (Ikenaga and Saigo, 1982) and sea urchin (Liebermann *et al*, 1983). The *Drosophila copia297* transposable element has been located specifically in the TATA box of H3 histone genes and presumably will have a profound effect on the transcription of this gene.

1.2.5 Illegitimate Recombination

Abnormal provirus excision events leading to the incorporation of cellular genes into viral genomes are an example of recombination between DNA segments which have little or no sequence homology. Shapiro (1982) refers to these non-specific and non-homologous events as illegitimate recombination events, as distinct from those events which require limited (12-25bp) homology and specific recombinases (e.g. integration and excision of bacteriophage λ).

Ikeda *et al* (1981, 1982) have shown that inhibitors of DNA gyrase will affect the illegitimate insertion of small plasmid DNA into λ DNA. It appears that these inhibitors do not have their effects indirectly, by reducing supercoiling, but rather affect illegitimate recombination directly. These results suggest that, in prokaryotes, DNA gyrase has a direct role in illegitimate recombination, possibly catalysing recombination events by a modification of its normal cleavage/ligation mechanism.

DNA gyrase binds to double-stranded DNA as an A_2B_2 tetrameric

complex and transiently cleaves DNA, leaving four base single-strand termini with the A-subunits covalently attached to the 5' ends at the cleavage site. This gyrase - DNA complex then assembles with another gyrase - DNA complex to form an A_4B_4 structure (Cozzarrelli, 1980) and at this stage it is proposed (Ikeda *et al*, 1981; 1982) that recombination can occur by subunit exchange, release of DNA gyrase and the repair and resealing of DNA. A cascade of these DNA gyrase mediated recombination events has been postulated to explain multiple recombination events seen in prokaryotes (Marvo *et al*, 1983). Some of the illegitimate recombination events seen in prokaryotic systems resemble events observed in animal cells (Marvo *et al*, 1983), however DNA gyrase has not been isolated from eukaryotic sources. It may be possible that, in animal cells, some events of this type are catalysed by other eukaryotic Type II - topoisomerases.

1.2.6 Transposition by Reverse-Transcription of mRNA

Another process capable of duplicating and scattering gene sequences was indicated when an unusual α-globin pseudogene was found (Flavell, 1982; Lueders *et al*, 1982). This gene had been processed, to remove introns, and retrovirus-like sequences near the pseudogene suggested that a transposition, involving mechanisms similar to retroviral DNA insertion, may have occurred (Lewin, 1983). A number of pseudogenes of this type have now been identified and this form of transposition may be a common mechanism for duplication of gene sequences, leading to the production of pseudogenes and orphons. There is some evidence (discussed in Maxson *et al*, 1983a) that this mechanism of gene duplication has been involved in the generation of histone gene orphons.

1.2.7 Horizontal Gene Transfer

Horizontal gene transfer between species has also been postulated as a mechanism of gene duplication. Busslinger *et al* (1982) have observed that the gene and spacer DNA of a family of histone genes in *Psammechinus* (h19) and *Strongylocentrotus* (Nor5) are much more similar than expected

given the rate of base mutation in the major histone gene clusters, and the distant relationship between the species. The similarity observed cannot be explained satisfactorily by stringent selection or by homogenising mechanisms (see section 1.2.3) which cannot act between genes which are reproductively isolated (Busslinger *et al*, 1982 also discount the possibility of sexual transmission between the two species). Given that other histone gene systems from these two species have evolved as expected, it has been proposed (Busslinger *et al*, 1982) that horizontal gene transfer has occurred between these two species, and that retroviruses may have acted as vectors. These viruses can cross species boundaries and can carry host genes (Bishop, 1981; Bishop, 1983)

1.2.8 Secondary Structure & DNA Rearrangement

It may be significant that various spontaneous mutations and recombination events are associated with sequences which can form strong secondary structures (see Ripley and Glickman, 1982). Several models have been proposed which associate the initiation of recombination with palindromic sequences (Sobell, 1975; Wagner and Radman, 1975). It is possible that nicks or gaps occur preferentially in regions of secondary structure and promote recombination events. Alternatively, secondary structure may inhibit strand transfer during general recombination and be associated with the termination of recombination (see Todd and Glickman, 1982). Interestingly, transposable elements have palindromic sequences at their termini which are required for transposition and have the potential to form secondary structure. Whether these structures form recognition sites for enzymes involved in transposition or have some other role is unknown.

DNA secondary structures will presumably present a significant challenge to the DNA replication machinery, and Ripley and Glickman (1982) suggest a number of mechanisms whereby DNA secondary structures may

mediate the formation of point mutations, additions, deletions and gene fusions during DNA replication.

1.2.9 Directed Gene Arrangements

Studies in both prokaryotes and eukaryotes indicate that gene expression can be altered by precise genetic rearrangements which serve specific biological roles.

For example, regulatory DNA inversions are responsible for phase variations in *Salmonella* (Simon *et al*, 1980) and Mu host-range alterations (Van de Putte *et al*, 1980). These inversions are generated from reciprocal recombination events across short regions of homology and are catalysed by specific recombinases (*hin* protein for *Salmonella* and *gin* protein for Mu). The fact that these two recombinases are interchangeable suggests that these systems may have a common origin (Kutsukake and Iina, 1980).

Recombination events are also involved in switching of mating types in yeast (Nasmyth, 1982) and in the formation of functional immunoglobulin genes (reviewed in Adams and Cory, 1983; Robertson, 1981).

Immunoglobulin molecules are made up of four chains; one pair of heavy and one pair of light chains. Each of these chains is divided functionally into two regions viz; the variable (V) region, which mediates antigen recognition, and the constant (C) region, which, in the case of the heavy chains, mediates the effector function of the molecule.

The V and C portions of the immunoglobulin molecule are encoded seperately in germline DNA and are brought together during development of the immune system to form the complete functional immunoglobulin genes.

Formation of each light chain gene is achieved by the joining of two seperate regions. The constant region is linked via a joining (J) segment to the V-gene segment, which contains the promoter and the major portion of the variable region (Robertson, 1981). The formation of a complete heavy chain gene involves the joining of a V-segment to a D(diversity) gene segment and a J-C region (Adams and Cory, 1983; Gough;

1981).

The characteristics of the systems described above indicate that the events involved are highly specific and probably mediated by site specific recombinases. The DNA rearrangements observed for the immunoglobulin genes may have evolved specifically as a means of generating at least part of the remarkable diversity required by the immune system, and may not resemble other events in eukaryotic genomes.

1.2.10 Genome Alterations : Effects on Gene Expression

Clearly there are a variety of processes for duplicating and rearranging genes and gene sequences. These include well known modes of general recombination, such as unequal crossing-over and gene conversion, which have been proposed to play key roles in maintaining sequence homogeneity among members of gene families. Other mechanisms for genome rearrangement include the processes of transposition, illegitimate recombination, integration after reverse transcription and the unprecedented possibility of gene transfer between species mediated possibly by viral vectors.

Various aspects of gene expression are highly dependent on the relationship of gene coding sequences to other determinants, i.e. promoter and terminator signals (see section 1.3). The relationship of gene position and gene expression is not completely understood, but there are a number of conceivable ways in which genome alterations can affect gene activity.

Firstly, insertion of DNA into a coding sequence, as can occur with transposable elements, generally abolishes expression of that gene. If a transposon inserts between a promotor and its coding sequence it can have a number of effects. In certain situations it may lead to premature termination of transcription and thus abolish expression. Alternatively the insertion element itself may provide or create new promoter elements and thus increase or decrease expression or provide

an alternative mode of regulation. Integration of viral DNA into the genome can bring powerful regulatory functions, which are part of the virus e.g. enhancer sequences, into the proximity of host genes (see Bishop, 1983). Insertions into intervening sequences could possibly affect processing of primary transcripts by providing alternative splicing sites. This effect has been observed with the *in vitro* insertion into introns of foreign DNA (Gruss and Khoury, 1980).

The rearrangement of DNA can have at least three consequences. The rearrangement may remove the promoter element and/or part of the gene coding region and as a result abolish expression. Deletion can eliminate the DNA between coding regions and lead to gene fusion. Homologous recombination between related genes can lead to hybrid gene formation and can account for certain forms of thalassaemia (Weatherall and Clegg, 1979). Rearrangement may also separate a gene from its original promoter element and place it in proximity of alternative controlling sequences.

The position of a gene in the genome can also have profound effects on its expression. The variegating position effect observed in *Drosophila* is well documented and demonstrates clearly a change in gene expression with altered gene position (Spofford, 1976). When a chromosome rearrangement moves a gene closer to heterochromatin, gene expression is reduced. This inhibitory effect extends across large distances (several kilo-bases) and diminishes with increasing distance from the centromere. The mechanisms responsible for this effect are still obscure.

1.3 GENE EXPRESSION

In eukaryotes and prokaryotes the major mechanisms of gene control appear to operate at the level of transcription. The transcriptional unit in eukaryotes contains the appropriate signals for the generation of a primary transcript, including those specifying transcriptional initiation and termination. Other factors which can

modulate transcription include proteins (interacting with regulatory DNA sequences), chromatin and DNA structure, and methylation state. These facets of gene expression are discussed in this section. A number of recent experiments leading to a further understanding of the control of transcription are cited.

Expression of gene products can also be regulated post - transcriptionally by factors which affect the processing, stability, transport or translation of the transcript. These forms of regulation are not discussed here but have been extensively reviewed (Nevins, 1983; Darnell, 1982).

1.3.1 Initiation of Transcription

In contrast to prokaryotes, eukaryotes utilise three RNA polymerases which transcribe different sets of genes (Roeder, 1976). RNA polymerase I transcribes genes for 18S and 28S ribosomal RNA's and RNA polymerase III transcribes tRNA and 5S ribosomal RNA sequences. Genes encoding histones and other proteins are transcribed by RNA polymerase II to give rise to messenger RNA (mRNA). Expression of these genes is discussed below.

In most cases, RNA polymerase II initiates transcription at a site located 25-30 bp downstream from a highly conserved 8-10 base-pair region. This sequence is known as the "TATA" or 'Goldberg-Hogness' box (Corden *et al*, 1980; Goldberg, 1978). This region is indispensible for transcription of polymerase II genes *in vitro* and its function appears to be related to accurate positioning of the start of transcription initiation. Deletion and mutation of this sequence does not abolish *in vivo* transcription but does result in heterogeneous start points (Breathnach and Chambon, 1981; Nevins, 1983).

There is good evidence that transcription begins at the sequence which specifies the 5' end of the message. This has been demonstrated

directly *in vitro* for late adenovirus and β-globin genes (Hagenbuchle and Schibler, 1981). The 5' ends of all polymerase II transcripts are modified post-transcriptionally by the addition of an ^{7}mG cap (Banerjee, 1980; Shatkin, 1976). This cap structure appears to function in translation initiation by interacting with certain initiation factors (Banerjee, 1980). It has also been speculated that the cap structure protects mRNA's from exonucleolytic degradation (Nevins, 1983).

Other sequences which modulate transcription have been located upstream of the transcription initiation site. Two regions upstream from the *Herpes simplex* thymidine kinase gene, both required for efficient transcription, were located using site directed mutagenesis (McKnight, 1982; McKnight and Kingsbury, 1982). A number of other controlling elements have been described and are discussed in later sections. The mechanisms by which these sequences promote transcription are generally unknown. Possibly they are involved in establishing active chromatin or DNA structures, or in interactions with non-histone regulatory proteins. These possibilities are discussed for specific examples in later sections.

Little is known of the protein factors (other than RNA polymerase) which are involved in polymerase II transcription. Matsui *et al* (1981) have defined at least four factors required for transcription initiation *in vitro* but only one of these has been purified. Recently, factors which bind specifically to the TATA box region and are necessary for accurate *in vitro* initiation of transcription have been identified (Parker and Topol, 1984a; Davison *et al*, 1983).

1.3.2 Termination of Transcription

a) Polyadenylated Messages

In every case examined in eukaryotes, transcription does not terminate at the poly (A) addition site but proceeds some distance downstream. This has been shown for adenovirus, SV40 and β-globin transcriptional units (Nevins and Darnell, 1978; Nevins *et al*, 1980;

Hofer and Darnell, 1981; Ford and Hsu, 1978). A clearly defined site for transcription termination has only been demonstrated so far for the mouse ß-major globin gene (Hofer *et al*, 1982; Salditt-Georgieff and Darnell, 1983). Using mouse erythroleukaemia (MEL) cells, induced with DMSO, newly synthesised ^{32}P-labelled RNA was purified and hybridised to probes specific for DNA segments 3' of the poly (A) addition site. The results of these experiments clearly indicated the presence of a discrete 3' termination point within a particular genomic region. Given this result it is possible that other eukaryote genes also have specific 3' termination sites.

It is clear that large segments of DNA sequence 3' of the poly (A) addition site of some eukaryotic genes are transcribed, however it is unclear at this stage whether these have a significant role. Proudfoot (1984) suggested that they may participate in the process of mRNA formation by forming a cleavage site recognised by a ribonucleoprotein. There is now good evidence that the small RNA species U1 is involved in intron splicing (Proudfoot, 1984) and that a small nuclear RNA is involved in processing of histone transcripts (discussed in next section). There may be a general mechanism involved in processing of RNA polymerase II transcripts, which involves specific sequence recognition between the gene transcript and a small nuclear RNA. Specific activities associated with these RNA's could possibly cleave the transcript and catalyse the subsequent processing, viz., splicing or polyadenylation.

The "Polyadenylation Signal"

Polyadenylated messages have the conserved sequence, AAUAAA, about 10-30 nucleotides upstream from the polyadenylation site (Proudfoot and Brownlee, 1976). The strong conservation of this sequence suggests that it plays a role in poly (A) addition, either in formation of a cleavage site, or directly, perhaps as a recognition site for poly (A) polymerase. However, some mRNA sequences have been found to contain

variant forms of this sequence (Nevins, 1983) and others, for example chicken histone H5 message (Krieg *et al*, 1983), are polyadenylated but lack this sequence. It seems likely therefore that there are alternative pathways by which polyadenylation can occur.

Site-directed mutagenesis experiments have shown directly that the sequence AAUAAA is involved in the formation of polyadenylated 3' ends. Deletion of this sequence from the late genes of SV40 (Fitzgerald and Shenk, 1981) abolished formation of processed late mRNA. Deletion of sequences between the AAUAAA and the poly (A) addition site, but not affecting the AAUAAA sequence, resulted in equally efficient addition of poly (A), but at a site downstream of the normal site, implying that there is a spatial requirement in the poly (A) addition mechanism. In another experiment, the AAUAAA downstream of the adenovirus E1A coding region was altered to AAGAAA (Montell *et al*, 1983). This change prevented the cleavage reaction so that most of the E1A transcripts obtained were found to extend beyond the cleavage site and into the next gene. However a small percentage of transcripts were processed and had the poly (A) tail, suggesting that the AAUAAA sequence is required for efficient cleavage but not for subsequent poly (A) addition.

b). Histone Messages

Mature histone messages are about 9S in size and are generally not polyadenylated (Adesnick and Darnell, 1972). Recent advances have led to an understanding of the unique processes involved in histone message 3' end formation.

Krieg and Melton (1984) demonstrated that long *in vitro* synthesised chicken histone H_2B transcripts were cleaved precisely in *Xenopus* oocytes, generating normal histone H_2B 3' termini. Results obtained by Price and Parker (1984) are consistent with this cleavage mechanism of 3' end formation. Studies by Birchmeier *et al* (1983) have shown that the conserved 3' dyad element found in most histone genes (Maxson *et al*,

1983a; Hentschel and Birnstiel, 1981) and capable of forming a hairpin structure at the end of the histone message, is required, along with an adjacent conserved sequence, for the 3' processing of histone mRNA.

Stunnenberg and Birnstiel (1982) found that the sea urchin H3 gene injected into *Xenopus* oocytes can only generate correctly processed mRNA in the presence of sea urchin cellular extracts. Purification has identified the complementing factor as a nuclear RNA - protein complex, containing a 60 nucleotide long RNA species. Price and Parker (1984) have also purified a factor from *Drosophila* cell nuclei which cleaves precurser H3 transcripts.

It appears that histone message formation involves termination and subsequent cleavage by a specific RNA - protein processing complex. It is still unknown whether termination of histone gene transcription occurs at a specific site.

1.3.3 Control of Gene Expression

Both positive and negative controls of transcription have been demonstrated for a number of eukaryotic gene systems. The most detailed information available for a negative control system is that of T-antigen repression of SV40 early genes (Nevins, 1983; Darnell, 1982), in which T-antigen binding at the early region promoter of SV40 prevents early SV40 transcription.

Dynan and Tjian (1983) have recently isolated a positive transcription factor from HeLa cell extracts which specifically activates a class of promoters, including the SV40 early promoter. Direct binding of this regulatory factor to a site located upstream of the transcription initiation site activates polymerase II catalysed transcription from the SV40 early promoter. Other examples of positive control of transcription of polymerase II genes have been documented, including the metallothionin genes and a number of hormonally controlled genes

(reviewed in Darnell, 1982; Nevins, 1983). Probably the best characterised positively regulated gene system is that encoding the 5S ribosomal RNA (reviewed in Miller, 1983; Brown, 1984).

Some very recent work concerning the heat shock and globin gene systems has contributed to the understanding of control in higher eukaryotes.

a) Heat Shock Genes

The heat shock proteins (hsp) are a set of proteins expressed when cells are subjected to elevated temperatures (reviewed in Ashburner and Bonner, 1979; Peterson and Mitchell, 1984). In *Drosophila*, heat shock leads to increased transcription of seven different *hsp* genes and a general repression of all other protein synthesis.

It has been shown that heat shock genes can be activated by raised temperatures in heterologous systems such as COS cells and *Xenopus* oocytes (Pelham, 1982; Mirault *et al*, 1982; Bienz and Pelham, 1982). The fact that heat-shock regulation operates in phylogenetically distinct organisms indicated that the heat-shock response has been highly conserved during evolution.

A small 10 bp conserved sequence motif is present between 47 and 66 bps upstream from the transcription initiation site of all sequenced heat-shock genes. This sequence is necessary and sufficient for induction of transcription by heat-shock *in vitro* (Pelham and Bienz, 1982). It is proposed that a factor, conserved in its essential features throughout evolution, interacts with this small sequence element and mediates the heat-shock response (Pelham and Bienz, 1982).

Wu (1984a) has identified two protein binding sites in the 5' end of the *Drosophila* heat-shock genes, one covering the TATA box and another the upstream control region. In a later paper (Wu, 1984b) the detection of heat-shock activator protein (HAP) activity is reported and the sequence specificity of this factor demonstrated.

Dudler and Travers (1984) recently provided the first analysis of the heat-shock promoter in transformed flies. Their study indicated that in this system, at least 97 bp 5' of the *hsp70* structural gene were required for heat inducibility, in contrast to results obtained for the *Drosophila hsp* genes in *Xenopus* oocytes and COS cells (discussed above) where the consensus element described by Pelham and Bienz (1982) was found to be sufficient to give optimal heat inducibility. The results of Dudler and Travers (1984) are consistent with the observation that a DNA-binding protein isolated from heat-shocked *Drosophila* cells protects a 55 bp region spanning from 37 to 92 bps upstream of the transcription initiation site (Parker and Topol, 1984b). This region includes the conserved heat shock element. It appears that additional sequences, perhaps constituting a second upstream sequence element, are necessary for efficient protein-binding and heat shock promotion in this system.

The factor isolated by Parker and Topol (1984b), which is specific for heat shock induction, may function in a manner analagous to the factor isolated by Dynan and Tjian (1983) which activates early gene transcription in SV40. The isolation of molecules which activate the heat-shock genes (Parker and Topol, 1984b) and the SV40 early genes (Dynan and Tjian, 1983) should make it possible to analyse the nature of these regulatory molecules, and the mechanisms involved in the gene activation processes.

The heat shock gene system fits the model of gene control proposed by Britten and Davidson (1969). In this model, repetitive sequences were originally proposed as *cis* regulation signals for co-ordinate expression and in this context these regulatory elements were designated *receptor* sequences. Trans-activating factors, which specifically bind to the *receptor* sequences, were proposed to be encoded by *integrator* genes. *Sensor* sequences were conceived as sites in the genome which, in response to external factors, induce the *integrator* genes to produce

activator. The heat-shock gene system, and the genes which are induced on amino acid starvation in yeast (see Davidson *et al*, 1983), provide examples of gene families which are co-ordinately regulated and contain homologous sequences in the vicinity of the genes, as predicted in the above model.

Two observations suggest that in the heat-shock gene system the *activator* and *sensor* properties are combined into a thermo-sensitive protein factor. Firstly, protein synthesis is not required for activation of heat-shock gene transcription *in vivo* (Ashburner and Bonner, 1979) and secondly the activating factor is present in non-shocked cells at about the same level as in heat-shocked cells (Parker and Topol, 1984b).

The histone gene system does not readily fit the Britten and Davidson model. As yet no sequences have been identified which are common to the 5' end of all coregulated histone genes. Whether the histone genes utilise a different form of control to that described above is uncertain. Possibly *receptor* sequences are present in the histone gene system but are in a form or position to make detection difficult.

b) The α- and β-globin Genes

The human α- and β-globin genes are part of the small globin gene family. They are expressed differently during fetal and adult development (reviewed in Weatherall and Clegg, 1981; Flavell and Grosveld, 1983) and provide a system for analysing differential expression of related eukaryotic genes. The α-globin genes are expressed throughout fetal and postnatal development while the adult β-globin gene is not fully active until after birth.

The expression of the human α- and β-globin genes during erythroid cell differentiation has been analysed by introduction into mouse erythroleukaemia (MEL) cells. When introduced into these cells on intact chromosomes (by cell fusion) the human α- and β-globin genes are regulated in the same manner as the endogenous mouse globin genes (Diesseroth

and Hendrick, 1978; Diesseroth *et al*, 1980; Willing *et al*, 1979). However, when introduced into MEL cells by transformation, the α-globin genes are expressed constitutively while the β-globin genes are only expressed after induction of erythroid cell differentiation by dimethylsulphoxide (DMSO) (Charnay *et al*, 1984; Wright *et al*, 1983; Chao *et al*, 1983). To explain these observations, Charnay *et al*, (1984) propose that the α-globin gene is under negative control while the β-globin gene is under combined negative and positive control. They suggest that derepression, possibly involving a change in chromatin structure (see section 1.4.3) is all that is required for activation of endogenous α-globin genes. Globin genes introduced by transformation presumably take up this active, derepressed state (perhaps integration preferentially occurs in active regions) so that α-globin genes introduced in this fashion are expressed constitutively. Further, they suggest that β-globin gene expression requires derepression and positive activation by an erythroid cell specific activator which is present after DMSO induction.

Using DMSO induced differentiating MEL cells as a system, Charney *et al*, (1984) have been able to localise the sequences responsible for the differential control of human α- and β-globin gene expression. Surprisingly, they found that constitutive expression of the human α-globin gene, and regulated expression of the human β-globin gene, in this system, were clearly conferred by sequences within the transcribed region of the genes (Charney *et al*, 1984; Wright *et al*, 1984).

There are other examples of internal gene sequences influencing gene transcription. The *Xenopus* 5S rRNA genes have been shown to have internal promoter sequences which are required for transcription (Bogenhagen *et al*, 1980; Sakonju *et al*, 1980; Brown, 1984; see section 1.4.3). In addition, a tissue specific "enhancer" sequence has been located in an intron splitting the variable and constant regions of the immunoglobulin genes. This enhancer sequence is responsible for

activation of the variable region promoter in the lymphocyte (Banerji *et al*, 1983; Queen and Baltimore, 1983; see section 1.3.4). Furthermore, recent work suggests that a region contained within the *Xenopus* histone H4 gene is involved in the regulation of transcription of this gene (Gargiulo *et al*, 1984).

1.3.4 Factors Affecting Gene Expression

a) Enhancers

Extensive analysis of polymerase II genes has revealed a complex set of gene control elements. The delineation of eukaryotic promoters into essential and regulatory elements is beginning to become clear, but knowledge of the mechanisms involved in modulating gene expression, differential and co-ordinate regulation is still preliminary. The discovery of eukaryotic enhancers has uncovered a system with the potential for controlling differential gene activity and has led to widespread studies concerning these controlling elements and the cellular factors which interact with them.

Enhancers have been identified, mostly, in viruses and are a class of controlling elements which can dramatically increase transcriptional activity of nearby genes. They are generally DNA sequences which can have their effect irrespective of orientation or distance from a promoter (reviewed in Khoury and Gruss, 1983). The enhancer elements of different viruses vary, but there is a conserved core sequence (Weiher *et al*, 1983). Viral enhancers have been shown to activate expression of a number of eukaryotic genes including rabbit ß-globin, conalbumin, chick lysozyme and mouse metallothionin (see Khoury and Gruss, 1983).

A number of viral enhancers exhibit host-specificity, showing different activities in cells of various species (Laimins *et al*, 1982; de Villiers *et al*, 1982; Kriegler and Botchan, 1983). This host specificity may be a function of the sequence of enhancers and may reflect their interaction with specific host factors.

The first non-viral enhancers demonstrated were those located adjacent to the immunoglobulin heavy chain J-region (Gillies *et al*, 1983; Banerji *et al*, 1983). These sequences are responsible for activation of the promoters of the immunoglobulin heavy chain variable regions. The gene for each chain of an immunoglobulin molecule is assembled by the translocation of a variable region, of low transcriptional activity, to the site of the constant region gene. The V-gene promoter is only fully active after translocation, even though it is unaltered (see Gough, 1981; Adams and Cory, 1983 for reviews). The increased activity of the promoter of the transposed V-gene is due to the influence of the enhancer which is in close proximity after the heavy chain gene has been assembled. A sequence with homology to the core region of many viral enhancers is present in the immunoglobulin enhancer region. Similar sequences have also been found in the J-C intron of the κ and λ light chain genes (Queen and Baltimore, 1983) and there is evidence that this region acts to enhance transcription from the promoter of the rearranged V_k - gene promoter (see Boss, 1983).

Gillies *et al*, (1983) and Banerji *et al*, (1983) have demonstrated that the immunoglobulin heavy chain enhancer operates much less efficiently in fibroblasts than in lymphoid cells, suggesting that there may be factors present in different cell types which recognise and interact with specific enhancer sequences (Dunn and Gough, 1984).

Scholer and Gruss (1984) have used an *in vitro* competition assay to demonstrate the interaction between cellular components and viral enhancer DNA. The results of these competition studies indicated a preference of viral enhancers for cellular factors from the virus' natural host cells and indicated clearly that cellular molecules are required for enhancer DNA to function. This was also demonstrated with a hormone dependent enhancer from mouse mammary tumour virus, which acts to increase the level of transcription only in the presence of

bound glucocorticoid receptor (Chandler *et al*, 1983; Yamamoto, 1983). It is possible that all enhancer elements function in concert with specific, regulatory cellular proteins.

The mechanism by which enhancer sequences increase transcription is unknown but it has been speculated that they act as entry sites for RNA polymerase II and other components involved in transcription (Moreau *et al*, 1981; Wasylyk *et al*, 1983). The fact that the regions of chromatin containing glucocorticoid dependent enhancer sequences and immunoglobulin gene enhancer sequences are hypersensitive to DNA'se I digestion (section 1.4.2b) (Zaret and Yamamoto, 1984; see Boss, 1983), is consistent with active enhancers being regions of altered chromatin or DNA structure, however the nature and properties of any proteins which interact with enhancers is still unknown.

The effects of enhancer sequences may be relevant to the evolution of gene arrangement. Clearly the arrangement of genes in a cluster, relative to enhancer sequences could affect their expression.

b) DNA Supercoiling

Supercoiling is the coiling of the DNA helix axis. In prokaryotes it is clear that the extent of supercoiling differentially activates promoters and affects recombination and replication (Cozzarrelli, 1980; Smith, 1981; Gellert, 1981; Fisher, 1984).

The enzymes which alter the topological state of DNA are called topo-isomerases and are ubiquitous. These enzymes catalyse the transient breakage of DNA and the passage of another DNA helix through the breaks (Gellert, 1981). Type I topo-isomerases break and rejoin one strand at a time and catalyse the passing of one strand through another. Type II enzymes break and rejoin both strands simultaneously and catalyse the passage of a double-stranded DNA segment through the double stranded break. The enzyme, DNA gyrase, is a type II topo-isomerase isolated from prokaryotes, and is the only topo-isomerase known which catalyses

the negative supercoiling of DNA (Cozzarrelli, 1980). Type II topo-isomerases have been isolated from *Xenopus* and *Drosophila* (see Cozzarrelli, 1980; Gellert, 1981) and a number of eukaryotic type I topo-isomerases are known (reviewed in Gellert, 1981). So far the only other gyrase-type enzyme known is a "reverse-gyrase" isolated from archaebacterium. This enzyme can positively supercoil DNA at temperatures greater than 50^{o}C (Kikuchi and Asai, 1984; Wang, 1984). No eukaryotic DNA gyrase-type enzymes are known.

Circular DNA isolated from eukaryotic sources is negatively supercoiled (Bauer, 1978), however the role of supercoiling in eukaryotes is unclear. It has been suggested that supercoiling has a role in the regulation of eukaryotic genes (Weisbrod, 1982; Luchnik *et al*, 1981). Luchnik *et al* (1981) showed that the transcriptionally active fraction of SV40 minichromosomes was in the supercoiled form. This is consistent with evidence that eukaryotic RNA polymerases prefer to transcribe tortionally strained DNA (Mandel and Chambon, 1974; Hossenlopp *et al*, 1974; Lescure *et al*, 1978; Lilley and Houghton, 1979; Chandler and Gralla, 1981).

In prokaryotes, there are a number of promoters for which expression is clearly dependent on DNA conformation (Smith, 1981; Fisher, 1984). The mechanism by which supercoiling enhances transcription is unclear, but, as negatively supercoiled DNA is partially unwound, it is assumed that this facilitates the unwinding of DNA by RNA polymerase (Fisher, 1984).

The fact that DNA in eukaryotic cells is organised into closed loops, or domains, has led to the proposal that supercoiling of these closed loops may serve a regulatory function in eukaryotes (Luchnik *et al*, 1982). Smith (1981) has proposed the existence of special sites, separate from promoters, at which topo-isomerases act to affect gene expression.

Another feature of DNA supercoiling is that it stabilises the formation of the left-handed form of the DNA helix, Z-DNA (Wang *et al*, 1979), which can form in sequences containing alternating purine and pyrimidine residues (see Nordheim and Rich, 1983). Whether this helical conformation is involved in gene regulation is still uncertain, but it is interesting that there is a clustering of sequences, capable on supercoiling of forming Z-DNA, observed in SV40 and other viral enhancer elements. Furthermore, the region of the SV40 genome which contains the Z-DNA forming segments is free of nucleosomes (see Nordheim and Rich, 1983). These findings suggest that Z-DNA and DNA supercoiling may be involved in transcriptional activation.

c) DNA Methylation State

While it is apparent that methylation of the dinucleotide CpG to mCpG inhibits transcription of several eukaryotic genes (Felsenfeld and McGee, 1982; Bird, 1983) it is difficult to assess whether reduced methylation is necessary for gene expression i.e. whether methylation is involved in regulating gene expression. Most CpG dinucleotides in vertebrate DNA are methylated, but a number of studies indicate that methylation is inhibitory only at certain CpG sequences. Busslinger *et al*, (1983) found that methylation of CpG sequences in the 5' region, but not in the transcribed region, of the human γ-globin gene abolished its transcription in mouse cells.

In the chicken vitillogenin gene, transcription after oestrogen treatment is associated with the loss of methylation at a CpG sequence 600 bp upstream of the transcription initiation site (Burch and Weintraub, 1983; Folger *et al*, 1983; Meijlink *et al*, 1983; Wilks *et al*, 1982). Other CpG sites in this region do not change their methylation state and the loss of methylation at this site appears to follow the initiation of transcription (see Bird, 1984). Interpretation of these results is complicated because of further methylation studies with the *Xenopus*

vitillogenin gene (Gerber-Huber *et al*, 1983). In this case none of the CpG sequences tested in this gene region changed their methylation state after hormonal induction. However the relevant sites may not have been analysed in this study.

Indirect evidence that methylation of particular CpG sequences is significant is presented in Max (1984) and is as follows. The sequence CpG is present in vertebrate DNA at a significantly lower frequency than expected given random distribution. This observation is termed CpG suppression and it has been suggested that it is due to deamination of 5-methylcytidine generating thymidine and thus leading to a replacement of mCpG by TpG through evolution. Max (1984) has noted that localised CpG rich regions have escaped this mechanism and he postulates that these conserved CpG rich regions are maintained in an unmethylated state. Studies using sperm DNA (discussed in Max, 1984) confirmed that these sequences are undermethylated in some eukaryotic genes. What features of these regions are responsible for maintaining the undermethylated state is unknown. These sequences may be conserved because they are important to gene function and if this is the case, a mechanism may have evolved to prevent methylation and subsequent deamination and mutation of these sequences. Alternatively these sequences may be maintained in an undermethylated state in order to facilitate gene expression, and consequently these sequences would escape deamination and mutation.

Methylation has so far only been detected in vertebrates and at a low level in invertebrates (Bird, 1984; Bird, 1983) and thus appears to have evolved late in animal history. It is therefore unlikely that it is involved in a general way in differentiation and gene control. Consistent with this is the fact that some CpG sequences, which, if methylated, prevent transcription of genes *in vitro*, are found unmethylated *in vivo* in cells which are not expressing these genes (see Bird, 1984). Thus the absence of methylation, in itself, is not enough to allow

transcription. It may however be a necessary precondition.

Methylation, therefore, still represents a phenomenon of unknown significance, and further gene methylation and expression studies are required for a more complete understanding of its role.

1.4 CHROMATIN

The DNA in the nucleus of eukaryotic cells is compacted tightly in the form of chromatin. Chromatin containing actively transcribed genes is in an active conformation which is different in form to the bulk genomic DNA. A number of facets of chromatin structure have been analysed and it has recently been proposed that histone H1 has a key role in gene repression by maintaining an inactive chromatin conformation. A summary of the distinguishing features of active and inactive chromatin is considered below, with an emphasis on recent developments in this field.

1.4.1 Nucleosome Structure

The nucleosome is the primary repeating unit of chromatin (reviewed in Igo-Kemenes *et al*, 1982). The nucleosome core particle can be generated by extensive digestion of chromatin with micrococcal nuclease and consists of 146 bp of DNA wrapped around a globular histone octamer, containing two each of the histones H_2A, H_2B, H3 and H4. The histone H1 is located in the region where the DNA enters and leaves the nucleosome (the linker DNA region) and holds two turns of DNA around the core. H1 protein is in close contact with the core particle and evidence suggests that it is involved in generating higher order chromatin structures (reviewed in Thomas, 1983).

Recently, the three dimensional structure of the nucleosome core particle was described at 7A resolution (Richmond *et al*, 1984). This study revealed regions of sharp bending in the DNA double helix in its path around the nucleosome. It has been suggested (see Widom, 1984) that certain regions of DNA (depending on nucleotide sequence)

can bend easily, and Richmond *et al*, (1984) suggest that such sequences may accomodate sharp bends in the nucleosome core while other sequences are excluded. This could possibly explain the observation that nucleosomes are associated non-randomly along DNA with respect to nucleotide sequence (see Zachau and Igo-Kemenes, 1981; Kornberg, 1981).

1.4.2 Active Chromatin

(a) Interaction with Non-Histone Proteins

A number of non-histone proteins, including RNA polymerases, are known to interact with DNA and chromatin. However, apart from the RNA polymerases, little is known of the role or significance of most of these non-histone proteins.

It has been shown that RNA polymerase II binds specifically to nucleosome core particles which contain actively transcribed DNA and that these core particles appear to lack an H_2A-H_2B dimer (Baer and Rhodes, 1983). Possibly RNA polymerase II binding is associated with the loss of an H_2A-H_2B dimer.

In addition to RNA polymerase II, some members of a group of non-histone proteins (the high mobility group, HMG proteins) also appear to selectively bind nucleosomes containing actively transcribed DNA sequences (reviewed in Weisbrod, 1982; Stein *et al*, 1983; Johns, 1982). These HMG proteins have been implicated in the formation of active chromatin from studies with DNA'se I. It is known that transcribed sequences reversibly lose their hypersensitivity to DNA'se I (see following section) in the absence of HMG 14 and HMG 17; and that this sensitivity can be restored by subsequent reconstitutions (see reviews, cited above). Competition studies have shown that the HMG 14 and HMG 17 proteins compete specifically with an H_2A-H_2B dimer for binding sites on nucleosomes from active chromatin (Stein *et al*, 1983), suggesting that these HMG proteins may bind to the site which normally interacts with the H_2A-H_2B

dimer. These studies suggest that formation of active chromatin may involve replacement of an H_2A-H_2B dimer with HMG proteins.

The recognition of a specific sequence by a regulatory protein has been analysed in nucleosomes reconstituted from *E. coli lac* operator DNA and calf thymus histone proteins (Chao *et al*, 1980). It was found that sequence specific binding of the *lac* repressor could only be detected when the *lac* operator sequence was associated with the H_2A-H_2B dimer. Thus there is evidence that RNA polymerase II and other non-histone proteins which interact with the nucleosome core particles of active chromatin possibly either replace an H_2A-H_2B dimer or interact with this dimer in some way. Possibly sequence specific recognition in eukaryotes is dependent on interactions of the DNA with the nucleosome core particle. It is not unreasonable to suggest that the relationship of DNA sequences to the H_2A-H_2B dimers in the nucleosome core could affect their transcription and regulation. In this context the nature of the association of histone H_2A variants, such as $H_2A.F$ (Harvey *et al*, 1983), with the other components of the nucleosome core particles is of great interest.

(b) Hypersensitivity to DNA'se I

Changes in chromatin structure associated with gene activity can be measured by analysing the accessibility of DNA to endonucleolytic cleavage. It has been known for some time that DNA'se I cleavage can distinguish active from inactive genes. DNA'se I hypersensitive regions of the genome were located within the immediate 5' regions of eukaryotic genes that are active or have the potential to be active (reviewed in Elgin, 1981; Elgin, 1984; Igo-Kemenes *et al*, 1982).

In the rat preproinsulin gene the DNA'se I hypersensitive region was found to be present in a rat insulinoma, but absent in non-expressing tissues, illustrating a tissue specificity which correlates with the expression pattern of the gene. Some DNA'se I hypersensitive sites have

also been found to appear only on hormonal stimulation. For example, a DNA'se I hypersensitive site appears transiently in the glucocorticoid dependent enhancer region of mouse mammary tumour virus (Zaret and Yamamoto, 1984; see section 1.3.4a), and a hypersensitive site in the chicken lysozyme gene appears with the presence of steroid hormones (Fritton *et al*, 1984; Reudelhuber, 1984). Other active enhancer elements also induce regions of DNA'se hypersensitivity (Jongstra *et al*, 1984). Some hypersensitive sites persist after genes are no longer transcribed (Stalder *et al*, 1980; Groudine and Weintraub, 1981; Igo-Kemenes *et al*, 1982).

The actual function and significance of the sites exhibiting this hypersensitivity to DNA'se I is still unknown. There is good evidence that these sites represent regions of DNA that are free of nucleosomes and are possibly interacting with, or available to interact with, non-histone regulatory proteins (Elgin, 1984; Reudelhuber, 1984; Thomas, 1983).

1.4.3 Chromatin Structure : Repression of Gene Activity

It appears that nucleosomes are absent from at least some types of active transcription units (see Thomas, 1983). Karpov *et al* (1984) used cross linking studies to assay for the presence of histones in the heat-shock gene regions during activation. With increasing transcription the coding regions of the genes were depleted first of H1 and then of all histones. They did not find histones in the 5' region of the genes at all and they suggest that this region remains free of nucleosomes - hence its hypersensitivity to DNA'se I (see previous section). This is consistent with electron-microscopic evidence which reveals a nucleosome free region, several hundreds of base pairs long, upstream from a number of transcribed sequences (Lamb and Daneholt, 1979; Saragosti *et al*, 1980). The results of Karpov *et al* (1984) suggest that one of the initial events in heat-shock gene activation may be

the removal of histone H1 from chromatin, followed by sequence specific binding of activating factors. Removal of other histones may accompany transcription or binding of RNA polymerases. A number of other recent studies have also implicated histone H1 in the general repression of gene activity, and these are discussed later in this section.

A recent study of active nucleolar chromatin from *Xenopus* oocytes is also consistent with active genes being free of histones. This study suggests that ribosomal genes are free of nucleosomes and other structures except for RNA polymerase molecules (Labhart and Kollar, 1982).

Further evidence is still required before any general statements can be made concerning the organisation of transcribed sequences in nucleosomes, but the above evidence, taken together with detailed studies of the micrococcal nuclease digestion patterns of the heat-shock genes (Keene and Elgin, 1981; Levy and Noll, 1981; Wu *et al*, 1979) and the ovalbumin gene (Bellard *et al*, 1982), suggest that transcribed sequences in many cases are not packaged into a regular nucleosomal array.

Xenopus 5S rRNA Genes

Xenopus has two types of 5S rRNA genes. The most abundant group are only active in *Xenopus* oocytes while the somatic type 5S rRNA genes are active in both oocytes and somatic cells (reviewed in Brown, 1982; Korn, 1982; Miller, 1983). Three factors, in addition to RNA polymerase III, are required for transcription of these two gene families (Segall *et al*, 1980). One of these, a positive transcription factor, TFIIIA (Engelke *et al*, 1980) has a four fold greater affinity for the intragene control region of the somatic type genes and this accounts for their slightly greater activity *in vitro* (Wormington *et al*, 1981).

When somatic and oocyte 5S rRNA genes are coinjected into *Xenopus* oocytes the oocyte type rRNA genes are preferentially inactivated. This inactivation is presumably due to the somatic type genes titrating out the necessary transcription factors (Gargiulo *et al*, 1984). A possible

explanation is that the specific transcription factors bind to the injected DNA in a competitive fashion and prevent histone binding which may be responsible for the formation of inactive chromatin. Hence the somatic rRNA genes, having greater affinity for these transcription factors, escape histone binding repression.

In other experiments, Schlissel and Brown (1984) found that only the somatic 5S rRNA genes were transcribed *in vitro* when purified somatic cell chromatin was used as template. Furthermore, they showed that the somatic genes are packaged in stable transcription complexes in somatic cell chromatin, while the inactive oocyte 5S rRNA genes are packaged such that transcription factors cannot bind. The repressed oocyte 5S rRNA genes could, however, be reversibly derepressed by removal of histone H1. Once H1 had been removed transcription of the oocyte 5S rRNA genes was possible by the addition of transcription factors.

In summary, activation of oocyte specific 5S rRNA gene transcription in somatic cell chromatin requires derepression, by removal of H1, and activation by specific transcription factors. This two step regulation (derepression and positive activation) is similar to that proposed for control of ß-globin gene expression (see section 1.3.3b). In that case the α-globin genes required only derepression, to give constitutive expression, while the ß-globin genes required derepression and activation by an erythroid cell specific, positive activation factor (Charnay *et al*, 1984).

Weintraub (1984) has shown a further association between H1 binding and gene activity. He suggests that higher order chromatin structure, mediated by H1 protein interactions, is responsible for overall gene repression. Several mechanisms for overcoming this repression have been conceived and include the binding of non-histone proteins (i.e. HMG 14 and HMG 17 see section 1.4.2), histone modifications (see section 1.5.2), interactions of histone variants (see section 1.5.3) or the

presence of sequence - specific transactivating factors. However a more detailed knowledge of H1 binding is required before an assessment of this speculation is possible.

1.5 HISTONES

1.5.1 Histone Proteins

The histones comprise a set of five small basic proteins. They are found in all eukaryotic cells and are the fundamental structural proteins of chromatin. The histones H_2A, H_2B, H3 and H4 interact with DNA to form the nucleosome core (discussed in section 1.4.1). H1 binds to the DNA between successive core particles (see section 1.4.1) and is present at approximately half the molar concentration of the other histones (Kornberg, 1974; Noll and Kornberg, 1977).

The primary structure of each histone has been highly conserved throughout evolution (see Isenberg, 1979), reflecting their fundamental role in nucleosome and chromatin structure. H3 and H4 are very stringently conserved (the H4 amino acid sequences of pea and cow differ by only two of 102 amino acids). The histone proteins, in general, appear to be comprised of two regions: a basic N-terminal domain and a C-terminal globular domain. In H_2A, and H_2B only the C-terminal globular regions are highly conserved while the rest of the proteins vary considerably. Histone H1 is the most variable, but like H_2A and H_2B, has highly conserved stretches. These distinct evolutionary variations probably reflect the functional importance of particular regions of the histone molecule with respect to interactions with DNA and other histone and non-histone proteins.

1.5.2 Histone Modifications

The histone proteins can be modified post-synthetically in a variety of ways (see Isenberg, 1979). Phosphorylation of histones has been clearly associated with chromosome condensation before mitosis (Bradbury *et al*, 1974), and histone acetylation has been associated

with transcriptionally active chromatin regions (Vidali *et al*, 1978). Other modifications include methylation and poly (ADP) ribosylation. Histone H_2A can combine with a small highly conserved protein, ubiquitin (Goldstein, *et al*, 1975; Hershko, 1983) to form protein A24 or uH_2A. Ubiquitin is linked to H_2A via an isopeptide bond (Goldknopf and Busch, 1977) and all H_2A variants so far characterised have ubiquinated counterparts (Wu and Bonner, 1981). Recent evidence suggests that the binding of ubiquitin to non-histone proteins is the initial event in intracellular protein degradation, although the ubiquination of H_2A may serve a different purpose (Hershko, 1983) as it has been observed that active genes in *Drosophila* cells are greatly enriched in ubiquinated nucleosomes (Levinger and Varshavsky, 1982). The incorporation of uH_2A into nucleosomes of actively transcribed regions may be involved in chromatin structural changes, but further evidence is required before assigning a regulatory role to ubiquitin.

1.5.3 Histone Structural Variants

Although the histones have been extremely well conserved through evolution, new techniques for resolving histone proteins and advanced cloning technology have led to the identification of several structural subtypes for each histone. The histone variants identified fall into two groups : those differing in only one or a few amino-acid positions and those with extensive differences in protein structure (only H_2A, H_2B and H1 variants fall into this class). The histone subtypes have been divided into four classes, based on the relationship of their expression to DNA replication (Zweidler, 1984; see Old & Woodland, 1984).

(a) **Replication dependent** subtypes are those which are expressed at the beginning of DNA replication and repressed at the end.

(b) **Partially replication dependent** subtypes are those histones which are induced on DNA replication but are not completely repressed outside S-phase.

(c) **Replication independent** subtypes are expressed throughout the cell-cycle and in non-dividing cells e.g. $H1^{o}$ (Zweidler, 1984), H5 from birds (S. Dalton, personal communication).

(d) **The non-competing minor histones** are those occurring in constant small amounts in somatic cells, regardless of changes in other histone variants. These are not dependent on DNA replication. Representatives of this class include the histones identified by Wu and Bonner (1981) which are associated with the Chinese Hamster ovary cell line. These were referred to as "Basal" histones as distinct from the predominant histone subtypes which are linked to S phase.

Tissue specific histones have also been described. In birds, amphibia and reptiles, histone H5 gradually (but not completely) replaces H1 in erythrocytes (Billet and Hindley, 1972; Ruiz-Carrillo *et al*, 1974; Appels *et al*, 1972). Work from this laboratory has shown that H5 expression is not linked to DNA replication in dividing pre-erythroblast cells (S. Dalton, personal communication).

H5 has considerable homology with a subset of H1 proteins from a number of sources (including mammalian cells) and designated $H1^{o}$ (Pehrson and Cole, 1981). $H1^{o}$ is replication independent (Zlatanova, 1980) and considered as a member of class (c) above (Zweidler, 1984). Sperm specific histones have also been described (Old and Woodland, 1984).

As yet there are no indications as to the function of any of these histone variants, although it has been speculated that erythrocyte and sperm specific proteins could function in compacting chromatin (Old and Woodland, 1984). Those variants exhibiting only minor alterations in protein sequence may not be functionally different from the major histone subtypes, and may have become fixed in the population by genetic drift. Mutation experiments suggest this may be the case. Yeast contains two copies of each core histone gene (see section 1.6.5). The two H_2B genes encode proteins differing by four amino-acids (Wallis *et al*, 1980).

Inactivation of either H_2B gene, leaving only a single functional protein does not affect viability, while inactivation of both H_2B genes is lethal (Rykowski *et al*, 1981). This result suggests that in this case, neither variant has a unique functional role.

However, the existence of specific histone variants which have been conserved throughout evolution, e.g. $H_2A.Z$ and $H1^o$, and developmentally regulated variant subtypes, may indicate the existence of heterogeneous nucleosome particles which have a role in development and gene expression. Work in this laboratory is aimed at determining functional roles for the chicken histone variants $H_2A.F$ (Harvey *et al*, 1983) and H5.

1.6 HISTONE GENES : ORGANISATION

Considerable progress has been made in studying the organisation and mode of expression of the histone genes of lower eukaryotes. The sea urchin histone gene system is the best characterised and has served as a prototype against which other gene systems have been compared. The histone gene systems of higher eukaryotes are less well studied and, although conclusions can be made regarding organisation, no system has been extensively characterised. A brief review of the current knowledge of histone gene organisation in different species is given in this section.

1.6.1 Sea Urchin

The organisation of the sea urchin histone genes has been reviewed, in detail, in Kedes (1979) and Hentschel and Birnstiel (1981). The genes for the five early α-subtype histones are clustered into a unit that is tandemly repeated several hundred times in the sea urchin genome. These genes are present in the same order; H1, H_2A, H_2B, H3, H4 and all the genes are transcribed from the same DNA strand. The length of this repeat varies between sea urchin species (6.6 kb in *P. miliaris* and 6.5 kb in *S. purpuratis*) and the five histone genes are separated by AT rich spacer DNA.

More than one family of histone gene repeat can exist in a given sea urchin species. An example of this occurs in *L. pictus* (Cohn and Kedes, 1979a & 1979b). In this species there are two major clusters, represented approximately equally in the genome. The clusters have identical gene orders but differ in spacer sequence. Both clusters encode α-type histones and are probably co-regulated. Recently, a third non-allelic, tandemly arrayed cluster was reported for this species (Holt and Childs, 1984). The order of polarity of genes in this cluster is identical to the other early sea urchin gene clusters. The presence of multiple heterogeneous histone gene clusters has also been reported in the sea urchins, *S. drobachiensis*, *P. miliaris*, and *T. gratilla* (Busslinger *et al*, 1980; Busslinger *et al*, 1982; Cohn *et al*, 1976) and possibly occurs in members of the *Xenopus* family (Turner and Woodland, 1983).

Busslinger *et al* (1980) have examined the coding regions of the major (h22) and minor (h19) early histone gene repeat units in *P. miliaris* . The two repeats differ by 12.4%. In contrast, the minor repeat of *P. milaris* and the major repeat of *S. purpuratis* differ by only 1.74%, suggesting that the *P. miliaris* minor repeat and the *S. purpuratis* major repeat have a common origin. Possibly these two repeats were derived from a common ancestral gene cluster that was present before the divergence of the two species. This ancestral repeat may have been amplified during evolution, probably by unequal crossing over, to become the major repeat in *S. purpuratis*, while a different ancestral repeat (h22) became dominant in *P. miliaris*. A similar explanation will account for the co-existence of the three early histone gene repeats in *L. pictus*. An alternative explanation for the origin of the h19 minor repeat in *P. miliaris* has been proposed by Busslinger *et al* (1982), and is discussed in section 1.2.7.

The sea urchin genes discussed above all encode the early α-subtype histone proteins. As yet little is known of the sperm-specific or cleavage-stage specific histone genes but the late γ-type histone genes have now been cloned (see below).

Sea Urchin Late Gene Family

Unlike the tandemly organised early genes discussed above, the sea urchin late genes are loosely clustered and heterogeneous in organisation (Maxson *et al*, 1983b). Analysis in *S. purpuratis* indicates there are about 20 copies of the H_2A and H_2B late genes in the genome, and the organisation and number of genes in the *L. pictus* genome appears to be similar (Childs *et al*, 1982; see also Maxson *et al*, 1983c).

Restriction enzyme mapping and DNA sequence analysis has revealed few similarities between the late genes outside of the coding regions, suggesting considerable microheterogeneity (Maxson *et al*, 1983b; Roberts *et al*, 1984). While there is also considerable variation between the protein coding regions of late genes (see Maxson *et al*, 1983c) considerable homology (95%) has been observed in the coding regions of the late histone genes of *L. pictus* (Roberts *et al*, 1984). Possibly gene conversion acts separately on late histone gene coding sequences to maintain homogeneity in these regions while allowing spacer DNA to diverge. A different mechanism, such as unequal crossing over (Smith, 1976; Stephenson *et al*, 1981), may be responsible for the conservation of genes and spacer DNA in the tandemly repeated histone gene families (Roberts *et al*, 1984).

1.6.2 *Drosophila*

The histone genes of *Drosophila* are present at approximately 100 copies per haploid genome. The organisation of the *Drosophila* histone genes resembles the sea urchin organisation in that the five genes are contained in a highly reiterated tandem array. The gene order in *Drosophila* (5' H1, H_2B, H_2A, H4, H3 3') is different to that in sea urchin and in *Drosophila* the histone genes are transcribed from both DNA strands

(Lifton *et al*, 1977).

Two major types of repeating unit were found and, as in sea urchin, the genes in both repeats are separated by AT rich spacer DNA. The two repeat types are 4.8 and 5.0 kb long and differ by the presence of a 208 bp insertion in the H1-H3 spacer region of the longer repeat. The 4.8 kb unit constitutes about 25% of the histone genes in the *Drosphila* genome (Lifton *et al*, 1977).

These two major repeats are located in the same region of chromosome 2 (region 39D-39E; Pardue *et al*, 1979) and appear to be grouped into 14 units (average 5-6 repeats per unit), separated by non-histone DNA sequences (Saigo *et al*, 1981). It is not clear whether this arrangement has a role in the expression of these genes (discussed in Anderson and Lengyel, 1984). There is no evidence for histone genes encoding stage-specific or tissue specific histones (Zweidler, 1980)

1.6.3 *Xenopus*

In the frog, *Xenopus borealis*, most of the histone genes (60-70%) appear to be contained in a single histone gene repeat (Turner and Woodland, 1983). This repeat has the structure H1, H_2B, H_2A, H4 and H3, and the minimum length of the cluster is 16 kb. The histone genes are transcribed from both DNA strands. The related species *X. laevis* has at least three different clusters, each showing a different gene arrangement (Destrée *et al*, 1984; Zernick *et al*, 1980). Minor histone gene arrangements have been observed in both species but it is clear that the majority of the histone genes are organised into quintets which are repeated throughout the genome, but possibly not in a tandem fashion (Van Dongen *et al*, 1984).

Xenopus laevis, which contains three major histone gene repeats, also has three major H1 protein subtypes (Flynn and Woodland, 1980; Risley and Eckhardt, 1981). *Xenopus borealis* which has a single major histone repeat has only one major histone H1 subtype (Flynn and Woodland,

1980). The observation by Zernick *et al* (1980) and Destrée *et al* (1984), that the genes for the different H1 subtypes in *X. laevis* are associated with different gene orders, has led to speculation that the three major H1 subtypes in this species may be derived from independently evolving clusters (Turner *et al*, 1983; Destrée *et al*, 1984).

The mechanisms responsible for the different gene arrangements of these two closely related species are obscure. Whatever the mechanism, it has generated the observed differences over a short evolutionary period as *X. laevis* and *X. borealis* diverged only 8-10 million years ago.

1.6.4 Newt

The histone genes of the Newt (*Notophthalmus viridescens*) are reiterated 600-800 times in the genome. The genes are organised in a highly conserved repeat, but the structure is different to that described for other species (above). The gene order is H1, H3, H_2B, H_2A, H4; and all genes, except H_2B, are transcribed from the same DNA strand. The repeats are not tandemly arranged but are separated by up to 50 kb of satellite DNA composed of a small 225 bp repeat sequence (Stephenson *et al*, 1981). As in the frog, minor repeat structures also exist.

Thus amphibians (*Xenopus* and *Newt*) show variation on the histone gene repeat structure seen in sea urchin and *Drosophila*. The ordered repeat is observed but is not fixed in a tandemly repeating structure. In yeast, mammals and birds the histone genes are not clustered into a quintet structure, but exist in small dispersed clusters with no resemblence to a repeat unit.

1.6.5 Yeast

The histone genes of the yeast, *Saccharomyces cerevisiae*, have been studied in considerable detail (Hereford *et al*, 1979; Smith, 1983). Yeast has a total of eight core histone genes and does not appear to contain histone H1. The H_2A and H_2B genes are adjacent and transcribed

as two genetically unlinked gene pairs (Hereford *et al*, 1979) that appear to have arisen from a duplication event approximately 200 million years ago (Wallis *et al*, 1980). The H3 and H4 genes are arranged in a similar fashion and are not linked to each other or to the H_2A-H_2B gene pairs (Smith and Murray, 1983). All the yeast histone genes lack introns and each gene pair is divergently transcribed (Smith and Murray, 1983; Wallis *et al*, 1980). The significance of the unusual arrangement in yeast is unclear. Sequence differences between copies of each histone gene pair do not result in any amino-acid differences for the H3 and H4 duplicates, but result in two amino-acid differences between the two H_2A proteins and four amino-acid changes between the two H_2B proteins. However as discussed in section 1.5.3, inactivation of either H_2B gene does not affect the viability of the organism - suggesting these differences in structure may be the result of genetic drift following an ancient gene duplication.

1.6.6 Chicken

Considerable variation is apparent in the arrangement of the histone genes in the chicken, *Gallus domesticus*. It is estimated that there are about 10 copies of each histone gene in the chicken genome (Crawford *et al*, 1979; Sugarman *et al*, 1983). So far a number of independent genomic clones have been analysed in this and other laboratories (Harvey *et al*, 1981; Engel and Dodgson, 1981; Sugarman *et al*, 1983) and it is clear from these data that the genes are clustered, however, the gene order, the relative direction of transcription and the spacing of the histone genes are not conserved, thus ruling out a regular arrangement. Consistent with this, Southern blot analysis of genomic DNA shows that a number of genomic fragments of different size react with each individual labelled histone gene probe (Ruiz-Carrillo *et al*, 1983).

There is little conservation of sequences in the intergene

regions, although some histone subtype specific sequence elements have been identified (Harvey *et al*, 1982; Sugarman *et al*, 1983; Coles and Wells, 1985). Histone coding regions exhibit extensive divergence in third base positions and in some cases in amino-acid sequence (D'Andrea *et al*, 1981; Harvey *et al*, 1982; Sugarman *et al*, 1983).

Probes made from individual chicken histone genes hybridise to RNA species expressed in early embryo and adult (Sugarman *et al*, 1983; Engel, 1984). As most of these probes hybridised more strongly to a single form of embryonic RNA, the histone genes in this major set of clones have been referred to as "primarily embryonic". This includes the majority of chicken histone genes so far studied. Recent studies in this laboratory indicate that these "embryonic" histone genes are replication dependent (A. Col man and S. Dalton; personal communication).

Three chicken histone genes have been isolated which differ from those described above. The histone, H5, is related to H1 and almost completely replaces H1 in erythroid cell nuclei (Moss *et al*, 1973; see section 1.5.3). This is discussed further in Chapter 7. Engel *et al* (1982) have isolated an H3 gene, with introns, which encodes a protein very similar to a minor histone variant found in adult somatic tissue. An extremely variant H_2A gene has also been isolated (Harvey *et al*, 1983) and is currently being investigated in this laboratory.

1.6.7 Mammals

The histone genes of mammals are repeated 10-50 times per genome, and analysis of histone gene containing recombinants from mouse (Sittman *et al*, 1981; Marzluff and Graves, 1984) and human (Sierra *et al*, 1982; Heintz *et al*, 1981; Stein *et al*, 1984) libraries suggests that they are organised into small variable clusters, with no obvious repeat evident. In mouse, the replication dependent histone genes are present on chromosomes 13 and 3 (Marzluff and Graves, 1984) and in man on chromosome 1 (Chandler *et al*, 1979).

Other features of the mammalian histone gene system were revealed from characterisation of the human clones. Sierra *et al* (1982) observed that one of the histone gene containing clones also encoded an RNA of approximately 330 nucleotides in length which was only synthesised during G1 of the HeLa cell cycle. These same workers found that human histone genes were interspersed with multiple copies of the *Alu* DNA family (for review on *Alu* type repeats; see Jelenik and Schmid, 1982).

Mammalian histone genes are extremely heterogeneous in sequence. At least seven distinct species of H4 mRNA have been detected in cultured human cells (Lichtler *et al*, 1980; Lichtler *et al*, 1982) and a number of histone variants have been described in mouse (Zweidler, 1984) and in human (Wu *et al*, 1984).

1.6.8 Orphons

Single, isolated histone genes, derived from the tandemly repeated family of histone genes, were discovered by Childs *et al* (1981) and subsequently referred to as "orphons". These genes are presumed to result from a recombination event between the locus containing the tandemly repeated histone genes and a remote genomic locus, however, the exact mechanism of orphon generation is unclear. In some cases, *translocation* may have occurred via an RNA intermediate (see Maxson *et al*, 1983a; section 1.2.6), although it appears more than one mechanism of "orphon" generation probably operate in the genome. "Orphons" are removed from the strict homogenising mechanisms associated with the tandem structure and are therefore relatively free to diverge in sequence and function. The relocation of these genes may lead to a different expression mode for the "orphon".

1.6.9 Histone Gene Organisation : Summary

It appears that during the course of evolution of some higher organisms, single histone gene clusters have been selectively amplified, and subsequently maintained in tandem as nearly identical, highly repeated

units. Such repetitive clusters are present in fruit fly, sea urchin, newt and possibly *Xenopus* and were discussed in the previous sections. The repeating units of the various species differ in structure but have several common features. In each case the repeated unit is a quintet of genes, coding for the five histone proteins (H1, H_2A, H_2B, H3 and H4), in different orders, and interspersed with segments of non-coding spacer DNA. The repeated quintets are extremely homogeneous, exhibiting only minor differences in nucleotide sequence (Maxson *et al*, 1983a and 1983c; Hentschel and Birnsteil, 1981).

Distinct differences relating to histone gene order, relative directions of transcription, and sequence and length of spacer DNA are also seen between the tandem repeats of different species, suggesting that in each case the tandem array arose by amplification of a unique cluster.

The tandem arrangement of histone gene quintets occurs when the reiteration frequency of the histone genes is high (150-1000 copies), suggesting that the high copy number is needed to maintain this arrangement. Possibly the reiteration frequency of the repeats must be above a threshold for the tandem array to exist, otherwise the dispersed organisation, as seen in mammals and birds, is favored.

The mechanisms responsible for maintaining such a highly ordered tandem structure are still unclear. The existence of orphons (see previous section) suggests that it is possible for histone genes to become removed from the tandem array of genes, however this appears to happen only rarely. Possibly there is an unknown selective pressure maintaining the tandem array over long evolutionary periods (Old and Woodland, 1984). Perhaps the array is important for regulation of expression of the genes. Alternatively, it is argued by Maxson *et al* (1983c) that the tandem array of histone genes need not be advantageous for co-ordinate expression, stoichiometry, or high histone demand and suggested that the genes are

arranged in tandem and maintained by recombinational events which operate given the high gene copy number (see Maxson *et al*, 1983c; Old and Woodland, 1984). Once the number of quintets reaches a threshold value, unequal crossing over between repeats (see section 1.2.3) can maintain sequence homogeneity, giving rise to a regular and stable array of genes (Maxson *et al*, 1983c; Stephenson *et al*, 1981).

It is possible that during the evolution of mammals and birds an ancestral histone gene quintet was divided, before duplication and subsequent amplification (by unequal crossing-over) could achieve the threshold which would allow self maintenance.

It is conceivable that amplification of a quintet of histone genes would be more favourable than the separation and amplification of individual histone genes, because it would ensure a balanced histone gene copy number. However, it is clear that in histone gene families, with low gene reiteration frequencies, that equal gene dosage and homogeneity can be maintained despite a dispersed organisation. Roberts *et al* (1984) have described gene conversion events which maintain homogeneity between the coding regions of the dispersed sea urchin "late genes", and various dispersed arrangements of histone genes are obviously capable of supplying stoichiometric amounts of histone proteins throughout development.

1.7 HISTONE GENES : EXPRESSION

1.7.1 Cell-Cycle Regulation

Coupling of histone gene transcription with DNA replication is well documented in many cell types and appears to be controlled by the regulation of steady state mRNA levels (reviewed in Maxson *et al*, 1983a and 1983c). The rapid increase of histone mRNA concentration seen at the beginning of S-phase in HeLa cells results from an increase in transcription and an increase in mRNA half-life (Heintz *et al*, 1983; Stein *et al*, 1984). At the end of DNA synthesis the rate of histone

mRNA synthesis falls and histone mRNA is rapidly degraded (Heintz *et al*, 1983). The change in stability of the mRNA is dependent on protein synthesis. In mouse lymphoma cell lines, histone mRNA synthesis and DNA replication are less tightly coupled. Histone mRNA and histone protein synthesis can be detected in the G1 phase as well as during S-phase (Sittman *et al*, 1983; Groppi and Coffino, 1980; Coffino *et al*, 1984). Although replication dependent histone gene transcription is predominant in cultured animal cells, replication independent variants have also been found at a low level throughout the cell-cycle (Wu and Bonner, 1981).

Cell-cycle regulated histone gene expression also occurs in yeast. Hereford *et al* (1982) have shown that activation of histone gene transcription in yeast occurs in late G1 phase, just prior to initiation of DNA replication. Transcription ceases in early S phase. They suggest a simple model for control of histone expression, in which changes in chromatin structure precede the initiation of DNA replication and also activate histone mRNA synthesis.

Histone gene expression is reduced by DNA synthesis inhibitors which affect both the level of transcription and the stability of histone mRNA (see Maxson, 1983a). The effects of these inhibitors can be overcome by protein synthesis inhibitors which prevent destabilisation of histone mRNA (Stahl and Gallwitz, 1977; Butler and Mueller, 1973) and increase the rate of histone gene transcription (Graves and Marzluff, 1984). Taken together, these results have led to the tentative proposal that there exist unstable proteins which are required for histone mRNA degradation and negative repression of histone genes (see Old and Woodland, 1984). The recent work described by Graves and Marzluff (1984) implies that both histone gene transcription and mRNA degradation are linked to deoxynucleotide metabolism, but the mechanism for this is not known. Old and Woodland (1984) speculate that unique features of the histone

mRNA, such as the dyad symmetry 3' element, may be responsible for its specific sensitivity to destabilisation. The necessity for temporal destabilisation may, in part, explain the lack of poly(A) from replication dependent histone mRNA, while replication independent histone mRNA's (i.e. H5, H3.3) are found in the poly (A) plus fraction (see Old and Woodland, 1984).

A nucleotide sequence in the 3' flanking region of a yeast H_2B gene has been shown to be necessary for S-phase regulation of the adjacent H_2A gene (Osley and Hereford, 1982). This DNA sequence is either part of, or closely linked to, an autonomously replicating DNA sequence (*ars*) in yeast (Osley and Hereford, 1982) which may indicate that, at least in yeast, there is a connection between sequences required for replication and histone gene transcription. We are presently trying to determine whether *ars* sequences are linked to chicken histone genes, to determine whether there is a general association of replication sequences and histone genes.

1.7.2 Developmental Regulation

The best characterised example of developmental regulation and histone subtype switching occurs during sea urchin embryogenesis (reviewed in Kedes, 1979; Maxson *et al*, 1983a). In the sea urchin genome there are three developmental classes of histone genes, viz; the cleavage-stage (CS) histones, the early (α) subtypes and the late (γ) subtypes. The unfertilised sea urchin egg contains a store of histone mRNA present as ribonucleoprotein (see Maxson *et al*, 1983). This stored maternal histone mRNA remains within the egg nucleus until fertilisation and cleavage, and encodes a mixture of α and CS histone subtypes (Childs *et al*, 1979). The unfertilised egg also synthesises some labile histone transcripts (Ruderman and Schmidt, 1981; Brandhorst, 1980). Upon fertilisation the histone mRNP particles are mobilised and translated.

The major products of this translation are initially CS histones, however within 3 hours the α-subtype histones become predominant. The CS histones of the egg are utilised in pronucleus formation and replace the sperm specific histones associated with sperm DNA (Maxson *et al*, 1983a). Translation of stored maternal histone mRNA is predominant in the first six hours of development and continues until the early blastula stage (10 hours of development), by which time less than 5% of the total histone mRNA remains in the mRNP pool (Baker and Infante, 1982). After the 16 cell stage there is a rapid increase in cell division, S-phase becomes much shorter, DNA synthesis increases and there is an eightfold increase in histone synthesis due to activation of the highly repeated α-subtype histone genes (Maxson *et al*, 1983a). These α-subtype histones are expressed until the late blastula stage, with peak mRNA synthesis occurring at about the 100 cell stage (8-10 hours post-fertilisation) and then declining rapidly (Maxson and Wilt, 1982; Weinberg *et al*, 1983; Wilt, 1970). At later stages, larval chromatin contains the late γ-subtype histones (Weinberg *et al*, 1978) which are comprised of electrophoretically distinct forms of H1, H_2A and H_2B (Newrock *et al*, 1978; Cohen *et al*, 1975). The H3 and H4 histone proteins are indistinguishable from the early H3 and H4 counterparts, but are derived from distinct late mRNA's (Childs *et al*, 1979; Grunstein, 1978).

This developmentally regulated histone gene expression occurs in widely diverged echinoderm species and may play an important role in sea urchin development. The synthesis of the late histones is co-incident with significant changes in embryonic chromatin, including increased condensation and nuclease sensitivity (Maxson *et al*, 1983a).

Summary

In early sea urchin embryogenesis then, the expression of the histone genes is temporarily uncoupled from DNA synthesis. Pre-existing, stored maternal mRNA is utilised initially and transcription of early histone mRNA is not coupled to S-phase (Arceci and Gross, 1977).

Histone synthesis is also uncoupled from DNA replication during early *Drosophila* and *Xenopus* development. In *Drosophila*, stored maternal mRNA is utilised early in development to meet the requirements of extremely rapid cell division, which demands large quantities of histone proteins for chromatin formation (Zalokar 1976; see Maxson *et al*, 1983a and 1983c). Later in *Drosophila* development there is a rapid accumulation of histone message derived from the tandemly repeated histone genes (see section 1.6.2). Following gastrulation the rate of DNA synthesis declines and there is a drop in histone synthesis and mRNA stability (Anderson and Lengyel, 1980 and 1984.

Thus, in echinoderm and *Drosophila* embryogenesis, there are three phases of histone gene expression. Initially, stored histone mRNA is mobilised; in the second phase, tandemly repeated histone genes are transcribed, and in a third phase, as the cell cycle lengthens, there is a reduction in histone synthesis due to a decrease in synthesis and stability of histone mRNA (Maxson *et al*, 1983a and 1983c).

It is clear that different mechanisms have evolved to provide the histone proteins required during development. So far, sea urchin is the only example where a major switch in gene expression occurs. In the developing *Xenopus* embryo, the major source of histones is from stored maternal histone mRNA and protein (Woodland, 1980; Woodland *et al*, 1983; Van Dongen *et al*, 1984). It is only late in *Xenopus* development (1000-2000 cell stage) that histones are provided by the endogenous histone genes (Woodland *et al*, 1979). The *Xenopus* genome

contains approximately 40 copies of each histone gene and these are probably not sufficient to synthesise adequate histone protein during early development (Woodland, 1980).

The patterns of developmental regulation involved in histone gene expression of higher eukaryotes are still unknown, but, given the low repetition frequency of the histone genes in higher organisms, it would not be surprising if they, like *Xenopus*, utilise stored maternal mRNA in early embryonic development (when DNA replication rate is high) and change to histone protein derived from the endogenous histone genes as the rate of DNA replication slows down.

1.7.3 Introns

Direct DNA sequencing of cloned sea urchin early histone gene repeats (Schaffner *et al*, 1978; Sures *et al*, 1978) indicated that the coding sequences were colinear with the amino-acid sequences, and therefore demonstrated that no intervening sequences were present within these genes. DNA sequencing of histone genes from other species has indicated that this lack of intron sequences is a general, conserved feature of histone genes (Hentschel and Birnsteil, 1981).

Recently however, exceptional histone genes have been isolated and shown to contain intervening sequences. The first exception discovered was a chicken gene encoding a replication independent H3 gene (Engel *et al*, 1982) which contains at least two introns. Subsequently, a human H3.3 gene (Kedes and Wells, cited in Old and Woodland, 1984), the chicken $H_2A.F$ gene (A. Robins, personal communication) and the H3 and H4 genes of *Neurospora* (Woudt *et al*, 1983) have been shown to contain introns. Possibly the rapid expression of the replication dependent histone genes requires that any delay in activation due to processing of messages be minimised (Old and Woodland, 1984) so that no introns are found in these genes.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemicals & Reagents

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

Acrylamide: Sigma Chemical Co., twice recrystallised from $CHCl_3$.

Agarose (low gelling temperature): B.D.H. Australia.

Bromophenol Blue: B.D.H. Australia.

B.C.I.G.: Sigma Chemical Co.

Caesium Chloride: Harshaw.

Diethyl Pyrocarbonate: Sigma Chemical Co.

Dithiothreitol: Sigma Chemical Co.

E.D.T.A.: Sigma Chemical Co.

Ethidium Bromide: Aerosol Industries.

I.P.T.G.: Sigma Chemical Co.

2-mercaptoethanol: Sigma Chemical Co.

Nitrocellulose Filters: Sartorius, Schleicher and Schuell Inc. (BA85).

N,N-methylene bisacrylamide: Sigma Chemical Co.

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

Sephadex G-50 (Medium) : Pharmacia.

TEMED: Tokyo Kasei.

"Trizma" base: Sigma Chemical Co.

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.

Calf Intestinal Phosphatase : Sigma Chemical Co.

E. coli Deoxyribonuclease I : Sigma Chemcial Co.

E. coli DNA-polymerase I : Boehringer - Mannheim; Biotechnology Research Enterprises, South Australia, (B.R.E.S.A.).

E. coli DNA-polymerase I, Klenow fragment: Boehringer-Mannheim, B.R.E.S.A.

Lysozyme : Sigma Chemical Co.

Restriction Enzymes : New England Biolabs; Boehringer-Mannheim.

Ribonuclease A : Sigma Chemical Co., heated 80^{o} for 10 minutes in water before use.

RNA - dependent DNA - polymerase (reverse transcriptase): Molecular Genetic Resources, Inc.

T4 DNA ligase : Boehringer - Mannheim, B.R.E.S.A.

T4 Polynucleotide Kinase : Boehringer - Mannheim.

2.1.3 Nucleic Acids

Many purified nucleic acids were used in this study, the sources of which are indicated below:

Calf thymus DNA: Sigma Chemical Co.

E. coli λ DNA: prepared from a concentrated λ'phage stock obtained from Dr. J.B. Egan.

*Hin*dIII linker DNA (dCCAAGCTTGG): Collaborative Research.

M13mp83 and M13mp93 vector DNA: a gift from Dr. A. Robins.

M13 17 base sequencing primer: Biotechnology Research Enterprises, South Australia (B.R.E.S.A.).

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

pBR322 and pBR325 vector DNA: prepared from transformed *E. coli* E392, which was a gift from Dr. P.A. Krieg.

pHC79 cosmid cloning vector: prepared from transformed *E. coli* MC1061, and was a gift from Dr. R.P. Harvey.

Salmon sperm DNA: Sigma Chemical Co.

2.1.4 Bacterial Strains

(a) i) *E. coli* MC1061: *araD139Δ(ara,leu)7697, lacx74, galU⁻, galK⁻ hsr⁻, hsm⁺, strA* (Casadaban and Cohen, 1980) was a gift from Dr. R.P. Harvey.

ii) *E. coli* E392: *hsdK*($r_k^- m_k^+$) *supE, supF, trpR⁻, met⁻* was a gift from Dr. J.B. Egan.

These strains were used for plasmid transformation experiments.

(b) *E. coli* JM101: *lac, pro, supE, thi,* F' *traD36, proAB, lacI^q, ZΔM15,* was a gift from A.J. Robins. This strain was used in M13 cloning experiments.

(c) i) *E. coli* lysogen BHB2688: N205*recA⁻* (λ*imm434, cIts, b2, red3, Eam4, Sam7*)/λ

ii) *E. coli* lysogen BHB2690: N205*recA⁻* (λ*imm434, cIts, b2, red3, Dam15, Sam7*)/λ

These strains were used for preparation of packaging extracts (section 2.2.6) and were a gift from D. Kemp.

(d) *E. coli* HB101: F^-, *hsdS20*($r_B^- m_B^-$), *recA13, ara14, proA2, lacY1, galK2, rpsL20*(Sm^r), *xyl-5, mtl-1, supE44,* λ^-, was a gift from Dr. J.B. Egan and was used for propagation of cosmid recombinants.

(e) *E. coli* LE392; F^-, *hsdR514*($r_K^- m_K^-$), *supE44, supF58, lacY1* or *Δ(lacIZY)6, galK2, galT22, metB1, trpR55,* λ^-, was a gift from Dr. J.B. Egan and was used for propagation of λ derived 'phage.

2.1.5 Growth Media

(a) Luria (L) Broth for growth of *E. coli* MC1061, contains per litre: 10 g Bacto-tryptone (Difco), 5 g yeast extract (Difco), 10 g NaCl and the pH adjusted to 7.2 with NaOH. Where appropriate, the media was supplemented with ampicillin (50 ug/ml) or tetracycline (20 ug/ml).

(b) Growth media for *E. coli* JM101

(i) Minimal salts media, contains per litre: 10.5 g K_2HPO_4, 4.5 g KH_2PO_4, 1.0 g $(NH_4)_2SO_4$, 0.5 g Na_3-citrate supplemented after autoclaving with 1 ml of 20% $MgSO_4$, 0.5 ml of 1% thiamine-HCl and 10 ml of 20% glucose.

(ii) 2 X YT broth contains per litre: 16 g Bacto-tryptone (Difco), 10 g yeast Extract (Difco), and 5 g NaCl.

(c) Media NZ broth, for growth of *E. coli* lysogens BHB2688 and BHB2690, contains per litre: 10 g NZamine (Humko Sheffield Chemical Division of Kraft, Inc.), 5 g NaCl, 2 g $MgCl_2$ adjusted to pH 7.5.

(d) Agar plates were prepared by supplementing the above media with 1.5% Bacto agar (Difco). 0.7% agar was used for preparing soft agar overlays. All media was made sterile by autoclaving.

2.1.6 Antibiotics

Ampicillin was from Sigma Chemical Co., Chloramphenicol from Parke Davis & Co., Sydney Australia and Tetracycline from UPJOHN Pty. Ltd.

2.1.7 Isotopically Labelled Compounds

α-^{32}P-dCTP(1500 Ci/mmol), α-^{32}P-dATP(1500 Ci/mmol) and γ-^{32}P-ATP (2000 Ci/mmol) were from Biotechnology Research Enterprises, South Australia (B.R.E.S.A.).

2.1.8 Distilled Water

Double-distilled water was used throughout except that liquid growth media was prepared with mono-distilled water. The second distillation was from an all-glass apparatus. Solutions were sterilised by autoclaving, except where labile chemicals were involved in which case filtration through Millipore filters was used.

2.1.9 Buffer Saturated Phenol

This was prepared by vigorously mixing equal volumes of redistilled phenol and 0.1 M Tris-HCl, pH 9.5, 5 mM EDTA. The saturated phenol phase

was stored at room temperature.

2.1.10 Glassware and General Equipment

All glassware and equipment where necessary was alkali washed (in 1 N KOH, to minimise RNA'se contamination), rinsed well with double-distilled water and where possible sterilised by dry heat or autoclaving.

2.2 METHODS

2.2.1 Restriction Enzyme Digestion

Restriction endonuclease digestion of DNA was performed using the conditions of each enzyme described by Davis *et al*, (1980). A two fold excess of enzyme was generally used and the reactions were run for an hour, although this time was increased to up to eight hours for preparative digestions.

Reactions were terminated by the addition of EDTA to 5 mM, followed either by extraction with phenol and chloroform, and ethanol precipitation from a 0.2 M NaCl solution or direct ethanol precipitation after the addition of ammonium acetate to a final concentration of 2 M. Alternatively digestion was terminated by the addition of half a volume of urea load buffer (4 M urea, 50 mM EDTA, 0.5% bromophenol blue, 50% sucrose).

Redigesting fragments in LGT agarose

After initial digestion plasmid DNA was fractionated on LGT agarose (section 2.2.2a). A gel slice containing the fragment to be analysed was cut out, an equal volume of sterile water added, and after melting the agarose at 65^{o}C for 15 minutes, the DNA fragment was redigested at 37^{o}C in the appropriate enzyme buffer. After phenol/chloroform extraction and ethanol precipitation this DNA was analysed by agarose or polyacrylamide gel electrophoresis (section 2.2.2).

2.2.2 Routine Gel Electrophoresis

(a) Agarose Gel Electrophoresis

Analytical

Agarose (0.7% - 2%) was dissolved in TEA (40 mM Tris-acetate, 20 mM Na-acetate, 1 mM EDTA, pH 8.2) and cast either in 14 cm x 14 cm x 0.3 cm vertical slab-gel templates or on to 7.5 cm x 5 cm microscope slides, for horizontal gels.

Vertical gels were electrophoresed between tanks containing TEA at 65 mA for approximately three hours. Horizontal gels were run submerged in TEA at 150 mA for approximately 15 minutes.

DNA was visualised by staining with 10 ug/ml ethidium bromide for 5 minutes and examination under UV light.

Preparative

Low gelling temperature (LGT) agarose was dissolved in TEA and cast either into vertical templates or on to horizontal slides, as described above. Electrophoresis was carried out at $4^{o}C$.

DNA was detected by brief ethidium bromide staining and the desired bands excised from the gel with a scalpel. Two volumes of 200 mM NaCl, 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA were added to the slice, and the agarose melted at $65^{o}C$ for 15 minutes. An equal volume of buffer-saturated phenol at $37^{o}C$ was added, the phases rapidly mixed then immediately separated by centrifugation. The aqueous phase was re-extracted with phenol, then with ether, and the DNA ethanol precipitated.

Typically, 60% of the DNA present in any band was recovered.

(b) Polyacrylamide gel electrophoresis

Electrophoresis of DNA species of less than about 1 kb in length was carried out on vertical 14 cm x 14 cm x 0.5 mm gels containing 4% - 20% acrylamide/bis (30:1) polymerised in 90 mM Tris-borate, 2.5 mM

EDTA, pH 8.3 (TBE). Electrophoresis was performed at 250 V for approximately 90 minutes. DNA was visualised under UV light following ethidium bromide staining.

DNA fragments that had been fractionated preparatively were excised from the gel and the DNA eluted into two changes of 200 ul 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA at 37°C for between 1 and 16 hours. The eluate was adjusted to 200 mM NaCl and the DNA ethanol precipitated.

Efficiency of recovery depended on the size of the DNA fragments and ranged from 50-99%.

2.2.3 Transfer of DNA to Nitrocellulose and Hybridization with a Labelled Probe

Restricted DNA fractionated on agarose slab gels was transferred to nitrocellulose filter paper using the method of Southern (1975), as modified by Wahl *et al*, (1979).

Bidirectional transfer from agarose gels 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 one hour.

Prehybridization, hybridization and washing conditions were essentially as described by Wahl *et al*, (1979), except that both dextran sulphate and glycine were omitted from the hybridization mix.

Washed, dried, nitrocellulose filters were placed in contact with X-ray film and exposed at -80°C in the presence of one or two tungstate intensifying screens.

Dot-blot Analysis of DNA (Kafatos *et al*, 1979)

DNA (up to 5 ug/dot) was denatured in 0.5 M NaOH, neutralised with HCl, an equal volume of 20 x SSC (3 M NaCl, 0.3 M Na-citrate) added and the sample spotted on to nitrocellulose filter paper damp with 20x SSC. The filters were then processed as described above.

2.2.4 Purification by Sucrose Gradient Velocity Centrifugation

All sucrose gradients were 10-40% linear gradients constructed in a "gradient maker" using 5.5 mls each of 10% and 40% (w/v) solutions of sucrose with NET buffer (0.2 M NaCl, 10 mM Tris-HCl pH 7.6, 1 mM EDTA). Sucrose solutions were sterilised with diethyl-pyrocarbonate by incubation at 37°C for at least 2 hours.

Restriction endonuclease reaction mixes were layered directly onto the sucrose gradient after being adjusted to 5 mM EDTA and 0.2% (w/v) sodium dodecylsulphate.

Centrifugation was carried out using a Beckman SW-41 rotor at 210,000 x g for 16 hours at 4°C.

Gradients were fractionated by upward displacement with a more dense sucrose solution (50% w/v) using an ISCO Density Gradient Fractionator. DNA was recovered by ethanol precipitation.

2.2.5 Isolation of Clones from a λ-Recombinant Genomic Library

(a) Plating and Screening (Benton and Davis, 1977)

0.25 ml of a 'phage suspension in PSB (100 mM NaCl, 10 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$) were gently mixed with 0.5 ml of a mid-log phase culture of *E. coli* LE392 in L-broth and incubated at 37°C for 10 minutes. 9 ml of 0.7% L-agar, containing 10 mM $MgCl_2$, at 42°C, were added and the mixture poured on to fresh, dry 15 cm 1.5% agar plates containing 1% bacto-tryptone, 0.5% yeast extract, 0.5% NaCl, 0.2% glucose, 10 mM Tris-HCl pH 7.5, 1 mM $MgCl_2$. Plates were incubated, inverted, at 37°C overnight then stored at 4°C to harden the agar.

An unwashed, 14 cm nitrocellulose disc was lain onto the plate, orientation marks made with a needle and, when uniformly wet, peeled off and placed on to filter paper saturated with 0.5 M NaOH, 1.5 M NaCl for one minute and then sequentially on to two filter papers saturated with 0.5 M Tris-HCl pH 7.4, 1.5 M NaCl for two minutes each. A duplicate filter was lain on to the plate, the orientation marks aligned and the filter processed as described for the first filter.

Filters were air dried, baked at 80°C *in vacuo* for one hour, then pre-hybridised, probed and washed as specified above for Southern blots (section 2.2.3).

Autoradiography was carried out for one-two days.

"Dot-Benton" Procedure

1 ul of purified phage stock was spotted in duplicate onto a lawn of freshly plated LE392 (section 2.1.3) to form an ordered grid. After incubation overnight, at 37°C, to allow plaque formation, nitrocellulose filters were made from each plate using the procedure described above. These filters were hybridised with nick translated probes (see section 2.2.3) washed and exposed to X-ray film.

(b) Growth of 'phage and preparation of DNA

10^5 pfu/15 cm plate were adsorbed to *E. coli* host LE392 and plated as described above. Plates were incubated right-side up overnight at 37°C and then stored at 4°C. Plates were overlayed with 10 ml PSB and the 'phage allowed to diffuse into this solution at 4°C for eight hours. Debris was removed by centrifugation (10,000 g, 5 minutes, 4°C) and the 'phage precipitated at 4°C for two hours by adjusting the solution to 875 mM NaCl, 6% PEG. The flocculated 'phage were collected by centrifugation (10,000 *g*, 10 minutes, 4°C) and resuspended in 14 ml PSB.

This suspension was layered on to discontinuous CsCl gradients containing 2 ml blocks of CsCl in PSB, with densities of $\rho = 1.40$ and $\rho = 1.60$ and centrifuged at 210,000 *g* for 90 minutes at 15°C. 'Phage particles were collected from the 1.40/1.60 interface and stored at 4°C.

DNA was isolated from 'phage stocks by phenol/ chloroform extraction following the addition of two volumes of 10 mM Tris-HCl pH 7.4, 5 mM EDTA, and concentrated by ethanol precipitation.

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. After gradient purification, as described above, an equal volume of phenol/chloroform (1:1) was added and the aqueous phase extracted twice, then ethanol precipitated. The 'phage DNA was dissolved in water and stored at -20°C.

2.2.6 Construction of a Cosmid Recombinant Library

(a) Ligation Conditions

pHC79 DNA and 3.6 ug of partially digested *Sau*3AI chicken genomic DNA were incubated in a 2:1:2 molar ratio of arm 1:insert:arm 2, in a 10 ul reaction mix containing 1 unit T4 DNA ligase, using buffer conditions described in section 2.2.7c. This reaction mix was incubated at 4°C for 24 hours.

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

The bacterial strains used are described in section 2.1.4.

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 L-plates (section 2.1.5) and incubating them at 30°C for about 24 hours or until colonies were visible. The incubation temperature was critical, as the strains are λ lysogens and carry a thermoinducible prophage. Growth at intermediate temperatures (e.g. 34°C - 36°C) can inevitably lead to loss of the prophage, cell death and selection of unwanted mutants. To verify that the cultures still contained a thermoinducible prophage, loopfuls of cells were streaked on two L-plates and incubated, one at 30°C and the other at 42°C. Thick growth appeared on the 30°C plate after overnight incubation but not on the 42°C plate.

(c) Preparation of Packaging Extracts (B. Hohn, unpublished)

Growth and Induction

BHB2690 and BHB2688 were inoculated, to an A_{600} of 0.08, into

500 ml and 1500 ml respectively of prewarmed NZ broth (section 2.1.5) and incubated at 30°C for approximately four hours to give a final A_{600} of 0.3. At this absorbance the cultures were induced at 42°C for 15 minutes and then incubated at 37°C with vigorous aeration for 3 hours.

The cells were harvested by centrifuging at 5000 rpm for 10 minutes in the Beckman JA10 rotor.

Sonicated extract from BHB2690

All extra fluid was removed from the cell pellet and the cells were resuspended in 5 ml of cold sonication buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 3 mM $MgCl_2$, 5 mM β-mercaptoethanol), respun, and finally resuspended in 2 ml of cold sonication buffer.

Cells were sonicated while at 0-4°C, using a microtip, with short (2 second) bursts, until the solution began to clear and lose viscosity. At this stage an aliquot was removed and the rest of the solution sonicated further. Three aliquots were taken, processed separately and used in separate packaging reactions to determine packaging efficiency.

After sonication was complete the cell debris was removed from each sample by centrifugation (10 minutes, eppendorf microfuge, 4°C). An equal volume of cold sonication buffer and 1/6 volume of packaging buffer (6 mM Tris-HCl pH 8.0, 50 mM spermidine, 20 mM $MgCl_2$, 30 mM ATP, 30 mM β-mercaptoethanol) were added to the supernatant and 50 ul aliquots were dispensed into cold eppendorf tubes, immediately immersed in liquid nitrogen, and stored at -80°C.

Freeze-thaw lysate from BHB2688

After removing extra fluid from the culture pellet, cells were resuspended in 5 ml of cold sucrose solution (10% sucrose, 50 mM Tris-HCl pH 7.6), pelleted again and resuspended in 2 ml of the same buffer. The cell suspension was then transferred to a 10 ml polycarbonate Ti50 tube and 100 ul of fresh lysozyme solution (2 mg/ml lysozyme in 0.25 M Tris-HCl pH 7.6) added.

This mixture was then placed in liquid nitrogen and vortexed, alternately until frozen solid.

After thawing on ice 100 ul of packaging buffer was added and the cell debris removed by centrifugation (35,000 rpm, 35 minutes, $4^{o}C$, Ti50 rotor). The supernatant was distributed into 10 ul aliquots which were snap frozen in liquid nitrogen and stored at $-80^{o}C$.

(d) *In vitro* packaging

After thawing on ice, the freeze thaw extract and DNA (approximately 0.5 ug in 5 ul ligase buffer) were mixed. 15 ul of freshly thawed sonicate was immediately added and mixed in thoroughly with the pipette tip. After incubation for 90 minutes at $25^{o}C$ 70 ul of cold λ-dilution buffer (10 mM Tris-HCl pH 7.6, 10 mM $MgSO_4$, 0.01% gelatin) was added and the packaged DNA stored as a conventional 'phage lysate at $4^{o}C$.

(e) Infection of host bacteria

10 mls of L-broth + 0.4% maltose were inoculated with 0.2 ml of fresh HB101 overnight culture and grown to an A_{600} of 1-1.5. 200 ul of cells were incubated with 100 ul phage in 'phage storage buffer, PSB, and allowed to adsorb for 30 minutes at $37^{o}C$ before spreading onto predried L-agar plates containing 50 ug/ml ampicillin. Colonies were formed after incubation at $37^{o}C$ overnight.

(f) Amplification, storage and high density screening of cosmid recombinants (Hanahan and Meselson, 1980)

Each master filter was placed colony side up on a pad of slightly damp, sterile Whatman 3MM paper and a sterile, dampened nitrocellulose filter was carefully layered over the master. The filters were then marked with a needle, pressed firmly together, and peeled apart. The replica filter was placed on a fresh L-agar plate containing 50 ug/ml ampicillin, incubated at $37^{o}C$ to obtain colonies and then replicated, as above, to generate filters for probing. The original master

nitrocellulose filter was transferred to an L-agar plate containing ampicillin and 25% glycerol. After incubation for two hours, at 37°C, this filter was sandwiched with another sterile filter and stored in a sealed plastic bag at -80°C between six sheets of Whatman 3MM filter paper (one of which was damp to maintain humidity).

2.2.7 Construction of Recombinant DNA

(a) Blunt-ending reactions (Seeburg *et al*, 1977)

When it was required that DNA fragments be repaired to generate blunt-ends, 5'-overhangs were end-filled in a 20 ul volume containing 50 mM NaCl, 10 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 1 mM dithiothreitol and four dNTPs all at 0.5 mM. 2 units of *E. coli* DNA-polymerase I, Klenow fragment were added and the reaction mix incubated at 37°C for 30 minutes. The reaction was terminated by phenol/chloroform extraction and the DNA recovered by ethanol precipitation.

(b) Ligation of *Hin*dIII recognition sites

Oligodeoxynucleotide linker (dCCAAGCTTGG) containing the recognition sequence for *Hin*dIII restriction endonuclease was supplied free of 5' terminal phosphate groups and was phosphorylated with T4 polynucleotide kinase as described below.

Linker DNA was incubated in a 10 ul 'hot' reaction containing 60 mM Tris-HCl pH 7.5, 9 mM $MgCl_2$, 15 mM dithiothreitol, 1 mM spermine, 100 uCi of γ-^{32}P-ATP, 25 pmol of DNA, and 1 unit of T4 polynucleotide kinase (added last), for 45 minutes at 37°C. A 'cold' reaction mix contained 60 mM Tris-HCl pH 7.5, 9 mM $MgCl_2$, 15 mM dithiothreitol, 0.15 mM ATP, 500 pmoles of linker DNA and 50 units of T4 polynucleotide kinase and was incubated at 37°C for 30 minutes, then stored frozen.

Linker DNA was ligated to blunt ended DNA fragments in a 10 ul reaction mixture containing 1 ul ^{32}P-labelled linker, 1.5 ul 'cold' phosphorylated linker, from above, 60 mM Tris-HCl pH 7.5, 9 mM $MgCl_2$, 15 mM dithiothreitol, 1 mM ATP, 100-500 ng of purified insert DNA and

0.5 units of T4 DNA ligase. This reaction was incubated at 10°C for 24 hours and then made to 50 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol in a volume of 100 ul, 5 units of *Hin*dIII restriction endonuclease added and incubated at 37°C for 3 hours. After addition of EDTA to 5 mM the reaction mixture was phenol extracted and ethanol precipitated.

Linkered DNA was separated from cleaved linker DNA by electrophoresis on a 5% polyacrylamide gel and visualised by autoradiography overnight at 4°C. Sections of the gel containing the linkered DNA fragments were cut out and the DNA eluted, ethanol precipitated and ligated to *Hin*dIII cleaved pBR322 using the method described in section 2.2.7c.

(c) Ligation of insert DNA to plasmid DNA

In order to prevent self ligation, the 5' terminal phosphate groups were removed from plasmid DNA. Restricted plasmid vector (10 ug) was incubated with 1.0 units of Calf Intestinal Phosphatase, CIP, (previously dialysed against 50 mM Tris-HCl pH 9.0, 1 mM $ZnSO_4$) at 37°C in a reaction mix containing 10 mM Tris-HCl pH 9.0, 0.1% SDS for 2 hours. The reaction mix was then made 5 mM with EDTA, phenol/chloroform extracted three times, ethanol precipitated, and redissolved in an appropriate volume of water.

Vector DNA was purified from uncut vector by passaging through an LGT-agarose gel (section 2.2.2a). Restriction fragments to be subcloned were preparatively isolated from LGT-agarose or polyacrylamide gels.

Insert was ligated to vector DNA in a 10 ul volume containing 60 mM Tris-HCl pH 7.6, 9 mM $MgCl_2$, 15 mM dithiothreitol, 0.5 mM ATP, insert and vector DNA and 0.5 units T4 DNA ligase at 4-15°C for 16 hours. Sufficient vector to give a 1:1 molar ratio with insert was generally used.

Recombinant molecules were transformed into *E. coli* hosts E392

or MC1061 (section 2.2.8), selected and characterised as described (section 2.2.9b).

2.2.8 Transformation of *E. coli*

E. coli strains MC1061 or E392 were grown overnight at 37°C in Luria broth (section 2.1.5) and then diluted 1/50 into fresh L-broth and grown to an A_{600} of 0.6. The cells were chilled on ice for 30 minutes, pelleted by centrifugation (500 *g*, 5 minutes, 4°C) and washed in ½ volume of ice-cold 0.1 M $MgCl_2$. The cells were resuspended in 1/20 of the original volume of ice-cold, freshly prepared 0.1 M $CaCl_2$ and stored at 4°C for between 4 and 24 hours.

0.2 ml of these competent cells were added to 0.1 ml of the DNA (typically 5 ng - 50 ng) in 0.1 M Tris-HCl pH 7.5, and stood, with occasional mixing, on ice for 30 minutes. The cells were heated at 42°C for two minutes, kept on ice for a further 30 minutes and then allowed to warm to room temperature. 0.5 ml of L-broth was added and the transformed cells incubated at 37°C for 20-30 minutes.

The transformed cells were mixed with 3 ml of 0.7% L-agar (at 42°C) and plated on to 1.5% L-agar plates containing appropriate antibiotic. Plates were incubated overnight at 37°C.

2.2.9 Detection and Examination of Recombinant Plasmid/Cosmid Clones

(a) Colony Screening (Grunstein and Hogness, 1975).

Colonies for screening were grown on a sheet of sterile nitrocellulose that had been boiled three times in distilled water and lain onto an L-agar plate containing 50 ug/ml ampicillin. Filters were incubated overnight at 37°C until colonies reached between 0.5-1.0 mm in diameter. This filter was transferred to an L-agar plate containing chloramphenicol (200 ug/ml) and incubated overnight at 37°C before being sandwiched with another nitrocellulose filter. The colonies were lysed by transferring the filters (still sandwiched) sequentially onto Whatman 3MM paper saturated with 10% SDS for three minutes, 0.5 M NaOH,

1.5 M NaCl for three minutes and 0.5 M Tris-HCl pH 8.0, 1.5 M NaCl for 5 minutes. The nitrocellulose filters were peeled apart and baked at 80°C, under vacuum for 2 hours. Hybridisation and washing conditions were as described for Southern blot experiments (section 2.2.3).

(b) Miniscreen examination of plasmid/cosmid recombinants (Birnboim and Doly, 1979)

1.5 ml cultures of each recombinant were grown overnight in L-broth containing antibiotic. The cells were pelleted by centrifugation for 30 seconds in an Eppendorf centrifuge, resuspended in 100 ul of 15% sucrose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA, containing 4 mg/ml lysozyme, and incubated at room temperature for 5 minutes. 200 ul of freshly prepared, ice-cold 0.2 M NaOH, 1% SDS were added and the solution gently mixed and returned to ice for 10 minutes. 125 ul of ice-cold 3 M Na-acetate pH 4.6 were added and the solution incubated on ice for a further 15 minutes.

Insoluble material was removed by centrifugation (10 minutes, Eppendorf centrifuge, 4°C) and the supernatant phenol/chloroform extracted. Plasmid DNA was recovered from the aqueous phase by ethanol precipitation, resuspended in water and an aliquot analysed by restriction enzyme digestion and agarose gel electrophoresis. 1 ul of 10 mg/ml DN'ase-free pancreatic RN'ase was included in the restriction reaction.

2.2.10 Large-Scale Preparation of Recombinant Plasmid/Cosmid DNA

500 ml cultures of recombinant cells were grown in L-Broth to an A_{600} of 1.0 and then chloramphenicol added to a final concentration of 150 ug/ml. The cells were incubated for 8-16 hours to allow amplification of the plasmid DNA (Clewell, 1972). Cells were harvested by centrifugation (10,000 *g*, 5 minutes, 4°C) and plasmid DNA isolated by either Triton or alkali/SDS lysis.

(a) Triton lysis method

Cells were resuspended in 15 ml of 15% sucrose,

50 mM EDTA pH 8.0, containing 12.5 mg of lysozyme and incubated on ice for 15 minutes. 15 ml of 0.1% Triton X-100, 62.5 mM EDTA, 50 mM Tris-HCl pH 8.0 were added, with gentle mixing until the solution was homogeneous and the solution centrifuged (45,000 *g*, 30 minutes, 4°C). The supernatant was carefully removed and treated with 20 ug/ml (final concentration) DN'ase-free, pancreatic RN'ase, for 30 minutes at 37°C and 50 ug/ml (final concentration) Proteinase K for 30 minutes at 37°C. The solution was extracted with an equal volume of phenol/chloroform and the aqueous phase dialysed extensively against 10 mM Tris-HCl pH 7.4, 1 mM EDTA.

Following dialysis, the solution was adjusted to 0.2 M NaCl and the DNA recovered by ethanol precipitation. Contaminating RNA was removed by fractionating the DNA on a Sephadex G-150 column eluted with 0.2 M NaCl, 10 mM Tris-Cl pH 7.5, 1 mM EDTA. Plasmid DNA was identified by its absorbance at 254 nm and ethanol precipitated.

(b) Alkali/SDS lysis method

Plasmid DNA was liberated from the cells as described above for the miniscreen method, except that the volumes were increased 40-fold, and plasmid DNA was treated with 20 ug/ml DN'ase-free, pancreatic RN'ase prior to phenol/chloroform extraction.

The ethanol precipitate was resuspended in 1.6 ml of water, adjusted to 0.4 M NaCl, 6.5% PEG, and the DNA precipitated on ice for one hour. The precipitate was recovered by centrifugation (10 minutes, Eppendorf centrifuge, 4°C), washed with 70% ethanol and dissolved in water. Plasmid DNA was stored at either 4°C or -20°C.

(c) Isolation of supercoiled DNA

Plasmid DNA (approximately 200 ug) was resuspended in 7.00 ml H_2O/tube and 7.00 g solid CsCl added. In the dark, 0.700 ml 10 mg/ml ethidium bromide were added and the mixture centrifuged at 210,000 *g* for 40 hours at 15°C. The lower band was identified by brief exposure to weak UV light and recovered. Ethidium bromide was removed by five

extractions with isoamyl alcohol and the DNA precipitated by the addition of two volumes of water and six volumes of ethanol. DNA was recovered by centrifugation and washed three times with 70% ethanol.

2.2.11 Preparation of *In Vitro* Labelled DNA

(a) Oligo-dT-primed reverse transcription

The synthesis of cDNA was carried out in a 20 ml reaction mix containing up to 2 mg of mRNA, 1 mM each of the deoxyribonucleotides dATP, dTTP, dGTP, about 0.1 mM α-^{32}P-dCTP, 50 mM Tris-HCl pH 8.3, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol and oligo-$(dT)_{10}$ primer to a final concentration of 20 ug/ml. 10 units of reverse transcriptase were added and the solution incubated at 41°C for 60 minutes. The RNA template was removed by alkaline hydrolysis with 0.3 N NaOH for 15 minutes at 65°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 and the aqueous phase loaded on to a 0.4 cm x 10 cm Sephadex G-50 column and eluted with 10 mM Tris-HCl pH 7.6, 1 mM EDTA. 200 ul fractions were collected and the cDNA detected by Cerenkov counting.

(b) Nick-translation of double-stranded DNA

(Maniatis *et al.*, 1975)

200 ng of DNA were labelled in a 25 ul reaction mix containing 50 mM Tris-HCl pH 7.8, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 50 ug/ml bovine serum albumin, 5 uM each of α-^{32}P-dCTP and α-^{32}P-dGTP and 25 uM each of unlabelled dATP and dTTP. The DNA was nicked by the addition of 20 pg of *E. coli* DN'ase I and the reaction was started by the addition of two units of *E. coli* DNA-polymerase I. The solution was incubated at 15°C for 90 minutes, phenol/chloroform extracted and the unincorporated nucleotides removed by chromatography on Sephadex G-50 as described above. If the labelled DNA was to be used as hybridization probe, the DNA strands were separated by boiling the solution for two minutes and

then snap-cooling on ice.

2.2.12 Sequence Analysis of DNA

(a) Subcloning into M-13 'Phage Vectors; Preparation of M-13 replicative-form (Rf) DNA (Winter, 1980)

To 3 ml of 0.7% L-agar, at 45°C, were added 20 ul BCIG (8 mg/ml in dimethylformamide), 20 ul IPTG (8 mg/ml in water), 0.2 ml JM101 (A_{600}=0.6) and 0.1 ml of diluted M-13 'phage (approximately 200 pfu). This mixture was poured on to a 1.5% agar in 2 x YT plate (section 2.1.5) and incubated at 37°C for 9 hours.

A blue plaque was selected, toothpicked into 1 ml of 2 x YT broth and grown with shaking for 6 hours at 37°C. Meanwhile, a 10 ml culture of JM101 from a single colony on a minimal glucose plate was grown at 37°C to an A_{600} of 0.5, and added to 1 litre of 2 x YT. When the A_{600} of this culture reached 0.5, the 1 ml of 'phage solution was added and grown for 4 hours. Replicative form M-13 DNA was prepared by the alkali/SDS method described above (section 2.2.10b).

Ligation and transformation

M-13 (mp83 or mp93) vectors were prepared and ligations performed as described above (section 2.2.7c).

Competent cells were prepared by growing JM101 to an A_{600} of 0.6 in 2 x YT-broth, harvesting by centrifugation (500 *g*, 5 minutes, 4°C) and resuspending in freshly prepared 50 mM $CaCl_2$. Cells were used after storage at 4°C for at least 4 hours, but up to 9 days. One fifth of a ligation mix was added to 0.2 ml of competent JM101 and left on ice for 40 minutes. The cells were heat-shocked at 42°C for 2 minutes and then added to 3 ml of 0.7% agar containing 20 ul BCIG (8 mg/ml), 20 ul IPTG (8 mg/ml), and 0.2 ml of exponential JM101 (A_{600} approximately 0.5). The mixture was plated on 2 x YT-agar plates and grown for 9-12 hours at 37°C.

(b) Preparation of templates for sequencing

Recombinant plaques were toothpicked into 1 ml of 2 x YT containing 2 ml of overnight JM101 and grown with shaking for 6 hours at 37°C. Cells were pelleted by centrifugation in an Eppendorf centrifuge for 5 minutes. To each supernatant were added 0.2 ml of 2.5 M NaCl, 20% PEG 6000 and, after leaving at room temperature for 15 minutes, the 'phage pellet was collected by centrifugation. After removal of all the supernatant, the pellet was resuspended in 0.1 ml of 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA and extracted with an equal volume of neutralised phenol. The aqueous phase was re-extracted with 0.5 ml of diethyl ether and ethanol precipitated. The 'phage DNA was collected by centrifugation, resuspended in 25 ml of 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA and stored at -20°C.

(c) Complementarity testing of single-stranded M-13 recombinants

To determine which strand, of a particular sub-cloned DNA fragment, was present in a single-stranded M-13 recombinant (ssM-13 clone), hybridization analysis was carried out using an arbitrarily selected, or previously sequenced, ssM-13 clone, as a reference.

1 ul of the ssM-13 clone to be tested was added to 1 ul of reference ssM-13 clone and incubated with 1 ul of 10 x Hin buffer (100 mM Tris-HCl pH 7.4, 100 mM $MgCl_2$, 500 mM NaCl) at 65°C for 1 hour.

2 ul of loading buffer were added and the sample was electrophoresed on a horizontal, 1% agarose gel, next to 2 ul of reference clone (plus 2 ul loading buffer), until the dye had moved the desired distance. The DNA was visualised after ethidium bromide staining. Single-stranded M-13 clones with inserts identical to the reference clone co-migrate with the reference, whereas clones with the complementary strand are retarded as they have hybridized to the reference, thereby doubling their molecular weight.

(d) Di-deoxy Sequencing Procedures (Sanger *et al.*, 1977).

Hybridization

2.5 ng of universal primer (17-mer) were annealed to 1 ug of M-13 single-stranded template in a 10 ul volume containing 10 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$ by incubating at 70°C for 10 minutes, 37°C for 10 minutes and 25°C for 10 minutes.

Polymerisation

4 ul of α-^{32}P-dGTP (approximately 16 uCi) were lyophilized, the hybridization mixture was added, vortexed to resuspend the label and then 1 ul of 10 mM DTT was added. 1.5 ul each of the appropriate zero mix (T^o for ddTTP: 10 uM dTTP, 200 uM dCTP, 200 uM dATP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA; C^o for ddCTP : 200 uM dTTP, 10 uM dCTP, 200 uM dATP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA: A^o for ddATP : 200 uM dTTP, 200 uM dCTP, 10 uM dATP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA; G^o for ddGTP : 200 uM dTTP, 200 uM dCTP, 200 uM dATP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) and ddNTP solutions (0.3 mM ddTTP, 0.15 mM ddCTP, 0.5 mM ddATP, 0.35 mM ddGTP, each in water) were added together. 2 ul of the zero mix - ddNTP mixtures were added separately to four Eppendorf tubes ("reaction tubes").

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

4 ul of formamide loading buffer (formamide, deionised with mixed bed resin, 0.1% bromo-cresol purple, 0.1% xylene cyanol FF and EDTA to 20 mM) were added to stop the reactions and mixed by a short

centrifugation.

Samples were boiled for 3 minutes and then analysed on a sequencing gel.

DNA sequencing gels

1 ul of each sample was electrophoresed on a 6% polyacrylamide, 7 M urea gel (20cm x 40cm, 0.2mm) in TBE buffer (section 2.2.2b). The gels were pre-electrophoresed for 40 minutes prior to loading the samples. Electrophoresis was at 1200 volts until the tracker dyes had migrated the desired distance. After electrophoresis the gels were fixed for 10 minutes in 10% acetic acid, washed with several litres of water, to remove urea and dried. Autoradiography was generally carried out overnight at room temperature.

2.2.13 Containment Facilities

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

CHAPTER 3

CONSTRUCTION AND CLONING OF

CORE HISTONE GENE SPECIFIC PROBES

3.1 INTRODUCTION

In order to isolate and characterise recombinants containing chicken histone genes it was necessary to have homologous probes of high purity.

The histones show only limited sequence divergence throughout evolution consistent with their ubiquitous function. While there were a number of cross-species probes for individual histone genes available, results from this laboratory (Harvey *et al*, 1981) suggest that cross-species histone DNA probes can give ambiguous results. It was therefore necessary to construct specific probes which would allow unequivocal detection of chicken histone gene sequences.

RNA from 5 day old chicken embryos has been shown to be an excellent source of histone mRNA (Krieg, 1980; Crawford *et al*, 1979). A heterogenous 7-11S fraction could be resolved from 5 day old total chicken RNA by sucrose gradient centrifugation and this fraction contained mRNA for each of the 5 histone proteins, as identified by *in vitro* translation in a wheat embryo cell free system (Krieg, 1980).

Radio-labelled cDNA made to this RNA fraction was originally used as a probe to screen the chicken genomic library of Dodgson *et al* (1979). Recombinants, positive to chicken cDNA probe but negative to a combined cDNA probe made from known contaminating RNA species, were selected (Harvey and Wells, 1979).

One of these recombinants, λCHO1, was selected and characterised further (Harvey and Wells, 1979; Harvey *et al*, 1981). Figure 3.1 shows a restriction map of λCHO1 and maps of subclones, pCH3.3E and pCH3.75EH. All the coding regions within these clones have been fully sequenced.

Given the precise location of histone gene coding regions in the subclones, and the lack of intervening sequences within the genes, it was decided to isolate coding-specific fragments and subclone each

Figure 3.1

Restriction Endonuclease Map of λCHO1 and Derived Subclones

Blocked in regions indicate fragments which contain histone gene sequences (see below). The insert size of λCHO1 is 14 kb. Data is from Harvey *et al* (1981).

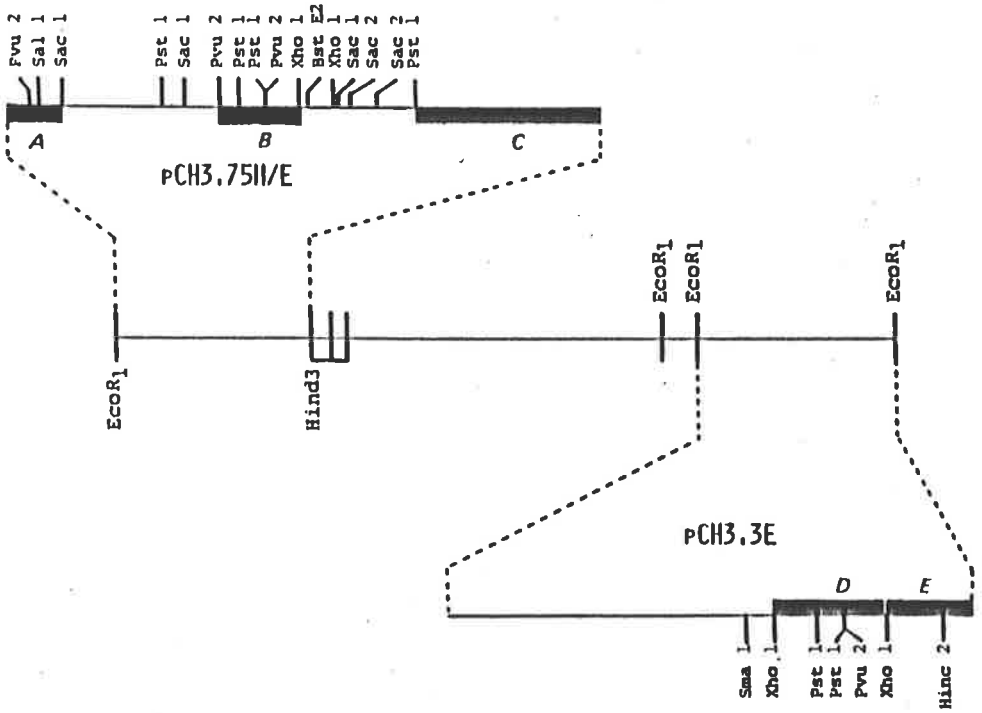
A: H3

B: H_2A

C: H4

D: H_2A

E: H_2B


Pvu 2
Sal 1
Sac 1
Pst 1
Sac 1
Pvu 2
Pst 1
Pst 1
Pvu 2
Xho 1
Bst E2
Xho 1
Sac 1
Sac 2
Sac 2
Pst 1
A
B
C
pCH3.75II/E
EcoR1
Hind3
EcoR1
EcoR1
EcoR1
pCH3.3E
D
E
Sma 1
Xho 1
Pst 1
Pst 1
Pvu 2
Xho 1
Hinc 2

of them into pBR322 for convenient use as specific probes to individual histone genes. The fragments isolated are described below.

The *Xho*I-*Xho*I fragment in pCH3.3E (Fig. 3.1) is 709 bp long and contains the entire H_2A coding region, 123 bp of 3' non-coding and 196 bp of 5' non-coding region. The sequence of this fragment has been reported (D'Andrea *et al*, 1981) and this fragment has been used previously to determine the position of other H_2A genes in chicken genomic clones (Harvey *et al*, 1981).

The 537 bp *Xho*I-*Eco*RI fragment in pCH3.3E contains the entire H_2B coding region, 146 bp of 5' non-coding and 10 bp of 3' non-coding (Harvey *et al*, 1982).

The complete H4 coding region can be excised from the subclone pCH3.75EH with *Hin*fI which generates a 338 bp fragment containing only 14 bp of 5' non-coding region and 12 bp of 3' non-coding sequence.

The complete H3 coding region is not present in λCHO1 but a 150 bp *Eco*RI-*Sal*I fragment can be isolated, from pCH3.75EH, which is entirely composed of H3-coding region (amino acid 10 to amino acid 59) and is therefore H3-gene specific.

The H_2A, H_2B and H4 fragments were treated with *E. coli* DNA Polymerase I and dNTP's to generate blunt ends which could be ligated to *Hind*III linkers and subcloned into the *Hind*III site of pBR322. The H3 specific fragment was isolated and subcloned into pBR322 digested with *Eco*RI and *Sal*I, using standard procedures.

3.2 RESULTS

3.2.1 Isolation of Coding Specific Fragments

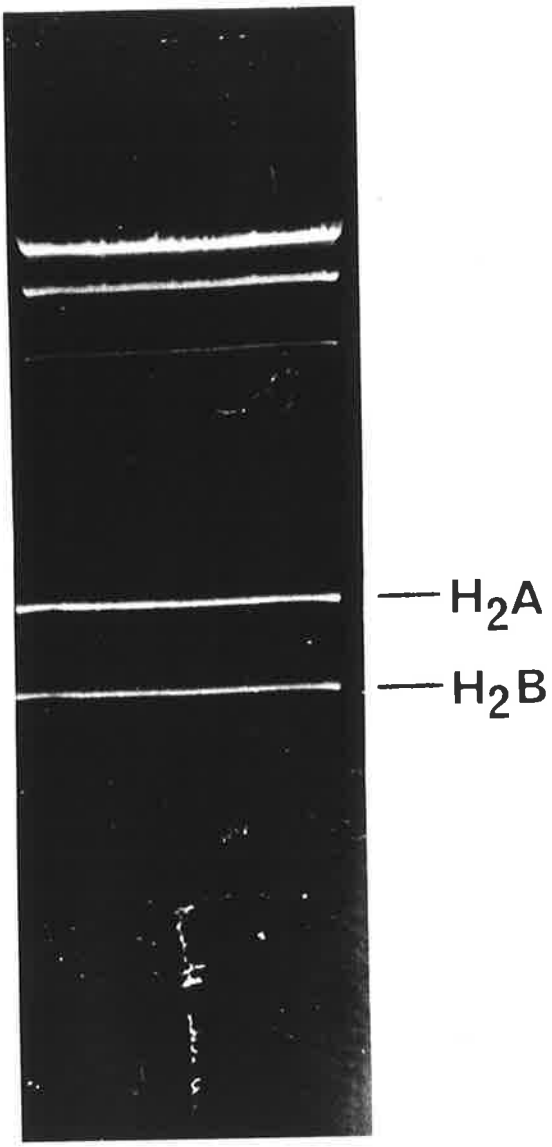
The H_2A specific *Xho*I-*Xho*I fragment and the H_2B specific *Xho*I-*Eco*RI fragment were isolated by elution from a preparative 5% polyacrylamide gel (2.2.2b) after *Xho*I-*Eco*RI double digestion of pCH3.3E (Fig. 3.2). The H4 specific *Hin*fI-*Hin*fI fragment was isolated in a similar manner. The *Eco*RI-*Sal*I fragment specific for H3 was kindly

Figure 3.2

Purification of H_2A and H_2B Gene Fragments

Isolation of H_2A and H_2B gene containing restriction fragments following *Eco*RI/*Xho*I double digestion of the chicken subclone pCH3.3E (Harvey *et al*, 1981; see Fig. 3.1).

10 ug of pCH3.3E DNA were digested with *Eco*RI and *Xho*I, electrophoresed on a 5% preparative polyacrylamide gel (section 2.2.2b) and the DNA visualised under UV light following ethidium bromide staining. The 0.7 kb *Xho*I-*Xho*I fragment containing H_2A gene sequences and the 0.5 kb *Xho*I-*Eco*RI H_2B gene fragment were excised and the DNA recovered by elution (section 2.2.2b).


H_2A
H_2B

provided in pure form by L.S. Coles from this laboratory.

After elution from polyacrylamide the H_2A, H_2B and H4 DNA fragments were treated with *E. coli* DNA Polymerase I, as described in section 2.2.7, to generate flush ended DNA suitable for ligation to linker DNA fragments.

3.2.2 Ligation of *Hin*dIII Recognition Site

To enable the H_2A, H_2B and H4 fragments to be conveniently inserted into (and removed from) the *Hin*dIII site of the plasmid vector pBR322, synthetic linker DNA encoding the *Hin*dIII recognition sequence was ligated to the termini of these fragments. *Hin*dIII does not cleave any of these inserts.

The linkers were labelled with ^{32}P using T4 polynucleotide kinase (Fig. 3.3; section 2.2.11d) and a test ligation reaction was followed by laddering of labelled linker DNA into multimers (Fig. 3.4). Ligation of H_2A, H_2B and H4 fragments with ^{32}P-labelled linkers resulted in the formation of dimers, as well as addition of linker DNA to the gene-specific fragments (Fig. 3.5).

*Hin*dIII was then used to cleave the synthetic linker DNA to monomer form leaving sticky ends suitable for ligation into the *Hin*dIII site of pBR322.

Linkered DNA fragments were separated from monomeric linker by electrophoresis on 5% preparative polyacrylamide gels (section 2.2.2b), as seen in Figure 3.5, followed by elution of DNA from the gel slices.

3.2.3 Ligation of Linkered Fragments into pBR322 Plasmid DNA

The plasmid vector DNA was treated with Calf Intestinal Phosphatase (section 2.2.7), after linearisation with *Hin*dIII, to remove 5' terminal phosphate groups. As at least one 5' terminal phosphate group is necessary for the ligation activity of T4 DNA ligase, this treatment effectively reduces the number of parental recombinants. Ligation of insert DNA to vector DNA is favoured as insert DNA has

Figure 3.3

5'-end Labelling of *Hin*dIII Oligomer

150 ng of *Hin*dIII linker DNA was 5'-end labelled by kinasing (section 2.2.7b). The reaction mix was phenol-chloroform extracted and the DNA recovered by ethanol precipitation. ^{32}P-labelled *Hin*dIII linker DNA was resuspended at a final concentration of 2.5 pmol/ul. A sample of the resultant molecules was electrophoresed on a 20% polyacrylamide gel (section 2.2.2b) and the DNA identified by autoradiography.

— 11mer
— HindIII decamer

Figure 3.4

Ligation of Kinased *Hin*dIII Linker DNA

Aliquots of a control ligation reaction containing ^{32}P-labelled linker DNA were electrophoresed on an 8% polyacrylamide gel (section 2.2.2b) before (A) and after (B) *Hin*dIII digestion, and autoradiographed for 16 hours. The ladder of bands seen in track (A) is due to multimers of synthetic linker DNA. In track (B) multimeric fragments have been cleaved with *Hin*dIII to generate monomeric linker DNA fragments.


A
B

Figure 3.5

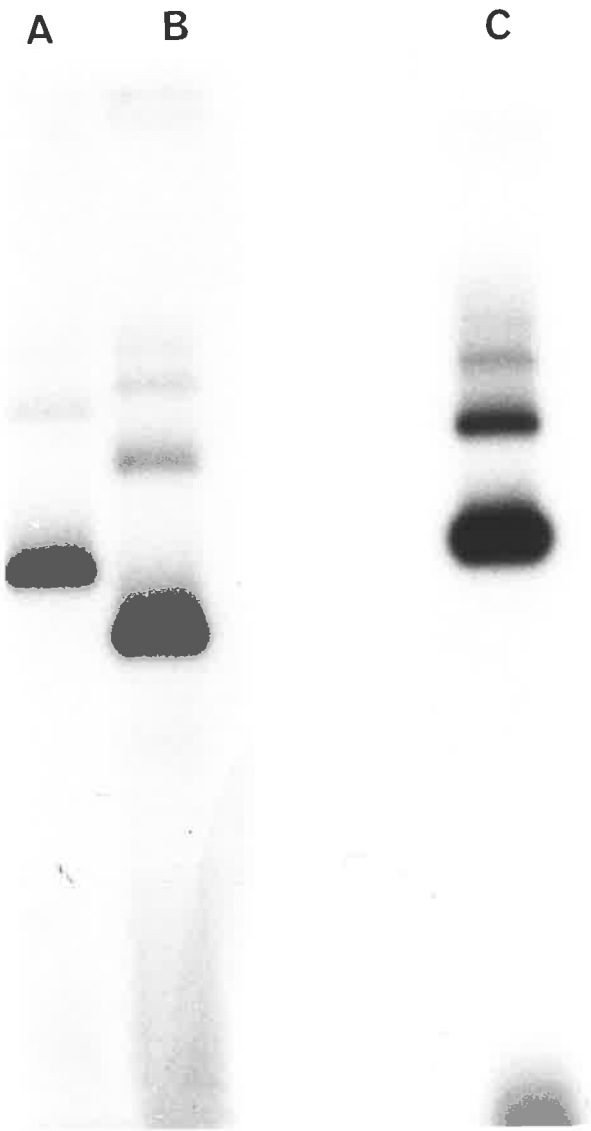
Preparative Electrophoresis of Linkered Histone Gene-Specific Fragments

Purified H_2A, H_2B and H4 gene specific fragments (see text) were ligated to ^{32}P-labelled *Hin*dIII linker DNA (section 2.2.7b). Linkered DNA was digested with *Hin*dIII and "sticky ended" gene fragments were purified from monomeric linker DNA by electrophoresis on 5% (tracks A and B) or 6% (track C) polyacrylamide gels. After autoradiography at $4^{o}C$, overnight, the linkered DNA fragments were excised from the gel and eluted (section 2.2.2b). Several concatameric forms of each gene specific fragment can be observed above the major linkered DNA fragment.

Track A: H_2A gene specific fragment

Track B: H_2B gene specific fragment

Track C: H4 gene specific fragment


A
B
C

5' terminal phosphate groups needed for T4 ligase activity.

Approximately 50-100 ng of purified H_2A, H_2B and H4 linkered inserts was ligated to 500 ng of dephosphorylated *Hin*dIII cut pBR322 DNA. Thus a 1:1 molar ratio of vector:insert was intended to give maximum ligation of linkered insert, and to ensure that no more than one DNA fragment was inserted into a single plasmid *Hin*dIII site.

3.2.4 Transformation, Selection and Screening of Recombinants

A total of approximately 50 ng of recombinant DNA (one tenth of each ligation mix) was used to transform *E. coli* strain E392 (sections 2.1.4; 2.2.8).

Amp^r*tet*^{s} colonies from each ligation were subjected to miniscreening (section 2.2.9b). Plasmid DNA from these colonies was digested with *Hin*dIII (H_2A, H_2B and H4 DNA) or *Eco*RI (H3 DNA) and run on either 6% polyacrylamide gels (section 2.2.2.b) or 1% agarose gels (section 2.2.2.a) to determine whether inserts were present. The results of this screening are shown in Figures 3.6 and 3.7.

*Hin*dIII digests of possible H_2A, H_2B and H4 hybrid plasmids revealed the presence of *Hin*dIII excisable inserts in most cases (Figs. 3.6 and 3.7b).

*Eco*RI digestion of H3 clones and comparison with *Eco*RI digested pBR322 indicated that all *tet*^{s}*amp*^{r} colonies screened were recombinants (Fig. 3.7a).

One each of the H_2A, H_2B, H3 and H4 containing clones was selected and large preparations (from 500 ml cultures) of plasmid DNA isolated from these clones (section 2.2.10). Approximately 50 ug of DNA from the H_2A, H_2B and H4 clones was digested with *Hin*dIII. 100 ug of DNA from the H3 clone was digested with *Eco*RI and *Sal*I. Each insert was purified by sucrose gradient centrifugation (section 2.2.4) and sized by electrophoresis on a 5% polyacrylamide gel (section 2.2.2b) (Fig. 3.8).

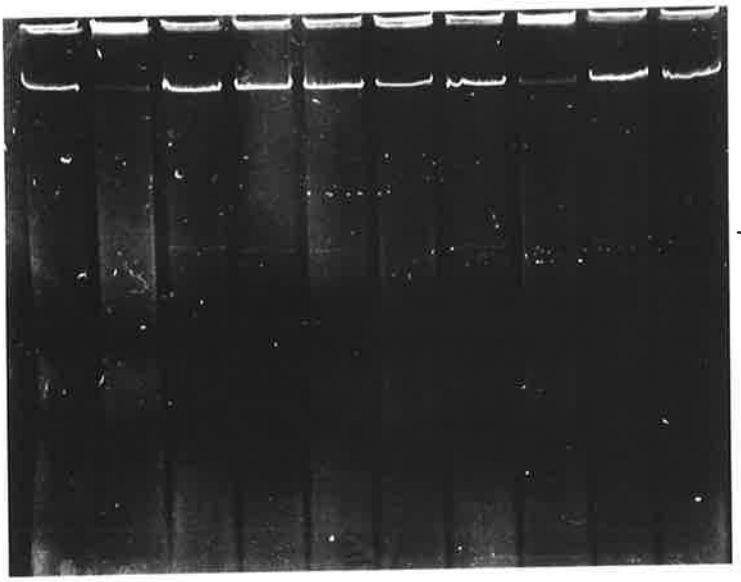
Figure 3.6

Examination of Recombinant Plasmids

Plasmid DNA prepared from *amp*^{r} colonies (section 2.2.9b) was digested with *Hin*dIII, electrophoresed on 6% polyacrylamide gels and visualised with UV light, after ethidium bromide staining.

Gel (a) shows DNA isolated from colonies generated after transformation with H_2B ligation mix and gel (b) DNA from colonies generated from the H4 ligation mix. DNA from a number of the colonies was cleaved by *Hin*dIII and generated insert fragments of the expected size. The position of the insert DNA fragments is indicated.

a)



b)

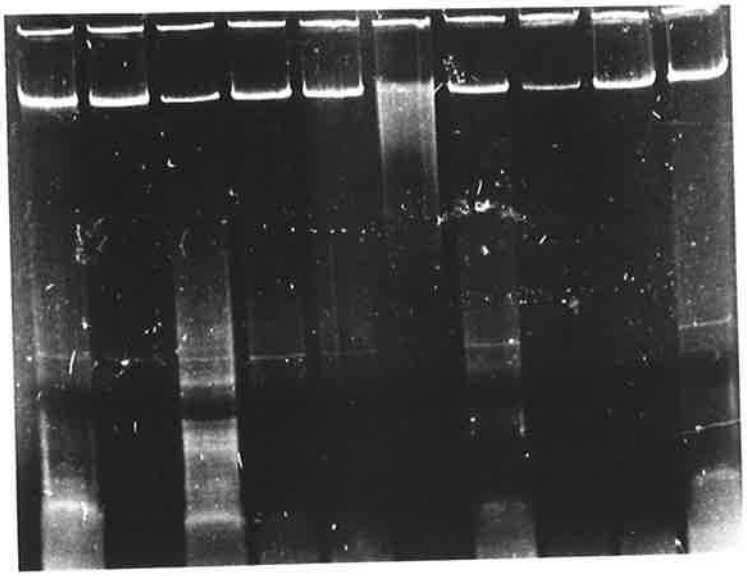


Figure 3.7

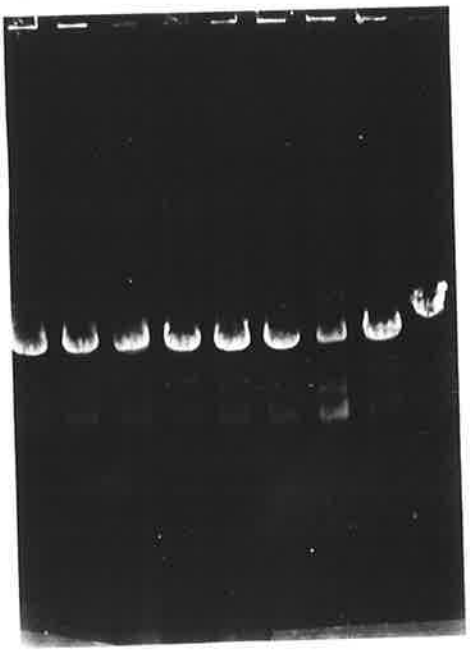
Digests of "Miniscreened" Colonies

a) Plasmid DNA prepared from putative H3 recombinants was linearised with *Eco*RI, electrophoresed on a 1% agarose gel and visualised with UV light after ethidium bromide staining.

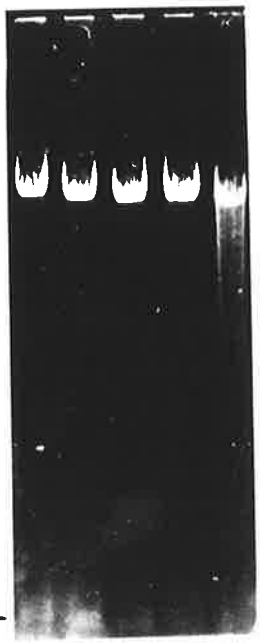
Tracks 1-8 contain DNA prepared from colonies transformed with pBR322 vector DNA ligated to *Eco*RI-*Sal*I H3 insert. Track 9 contains *Eco*RI cleaved pBR322. As the H3 insert is smaller than the *Eco*RI-*Sal*I fragment which was deleted during vector construction the linearised DNA from recombinants is smaller than the linear pBR322 DNA. All $amp^r tet^s$ colonies assayed appear to contain H3 insert.

b) *Hin*dIII digests of DNA isolated from 4 putative H_2A subclones were fractionated on a 1% agarose gel (tracks 1-4) and stained with ethidium bromide. 500 ng of *Hin*dIII cut pBR325 DNA was run as marker in track 5. At least two recombinants have a 0.7 kb insert.

a) 1 2 3 4 5 6 7 8 9



b) 1 2 3 4 5



H_2A—

Figure 3.8

Purified Histone Gene-Specific Inserts

100 ug of plasmid DNA derived from each histone gene-specific recombinant were digested with either *Hin*dIII (H_2A, H_2B and H4) or *Eco*RI and *Sal*I (H3) and fractionated on 10-40% linear sucrose gradients (section 2.2.4). The insert DNA was collected and an aliquot analysed by electrophoresis on a 6% polyacrylamide gel to check purity. The DNA was visualised with UV light after ethidium bromide staining.

Track A: H4 insert

Track B: H3 insert

Track C: H_2B insert

Track D: H_2A insert

These purified inserts were used as templates for nick-translation.


A
B
C
D

3.2.5 Discussion

DNA fragments specific for the coding regions of each core histone gene of the chicken were cloned into pBR322 in a manner generating easily excisable, highly specific fragments. These fragments could be separated easily from vector DNA, after excision, by sucrose gradient centrifugation, LGT agarose electrophoresis and elution, or elution after preparative polyacrylamide gel electrophoresis. Thus histone gene specific DNA fragments of high purity, suitable for detection of chicken histone gene sequences were available. Recombinants containing H_2A, H_2B, H4 and H3 inserts were referred to as pCH.H_2AH, pCH.H_2BH, pCH.H4H and pCH.H3ES respectively.

CHAPTER 4

ISOLATION OF HISTONE GENE RECOMBINANTS FROM BACTERIOPHAGE LIBRARIES

4.1 INTRODUCTION

The development of techniques for efficient *in vitro* packaging of DNA into 'phage particles (Hohn and Murray, 1977; Sternberg *et al*, 1977) and *in situ* hybridisation (Benton and Davis, 1977) have enabled the construction of complete 'phage libraries of eukaryotic DNA and the identification of desired recombinant clones by hybridisation.

Genomic libraries of eukaryotic DNA in bacteriophage vectors can be prepared in two ways. Genomic DNA can be digested to completion with a restriction enzyme (6 base recognition sequence) and cloned into a 'phage vector. This procedure has two disadvantages. If the desired gene contains the restriction enzyme recognition site, then this region will be cloned in two or more pieces. Certain regions of the genome will not be cloned at all if the restriction enzyme fragments generated are too small or large to be packaged. Secondly, because the average size of fragments generated using this procedure is approximately 4 kb, a large number of recombinants is generated and a laborious screening is necessary.

These problems can be avoided by generating and cloning larger DNA fragments (20 kb). Larger DNA fragments can be generated either by random shearing of eukaryotic DNA (Maniatis *et al*, 1978) or partial restriction enzyme digestion with endonucleases which recognise 4 base sequences and thus generate a series of DNA fragments which will be overlapping and essentially random.

These alternative procedures are now widely used and have two advantages, viz., sequences are unlikely to be excluded from the library due to the distribution of restriction enzyme sites and, as the fragments generated are large, 15–20 kb, fewer recombinants need be screened and the chance of isolating a single copy gene is high.

Libraries generated using these procedures are made up of a

collection of potentially random clones covering the entire genome. It is therefore possible to use a DNA fragment from one recombinant as a probe to identify overlapping clones derived from the same region of the genome. This is referred to as "chromosome walking". Obviously the aim of the cloning procedure is to generate a library of recombinants in which the total genome is represented. However there are a variety of reasons why particular sequences may be under represented (discussed in section 5.2.17).

Dodgson *et al* (1979) have constructed a library of random chicken DNA fragments (15-21 kb long) in the bacteriophage vector Charon 4a and the histone gene containing genomic clones λCHO1 and λCHO2 were isolated from this library (Harvey & Wells, 1979; Harvey *et al*, 1981) (see section 3.1).

Analysis of the gene organisation in these two clones (Harvey *et al*, 1981) indicated that although there is some clustering of histone genes, no obvious repeating unit is observed.

The aim of the work described in this chapter was to isolate clones which overlap and extend λCHO1, and thus allow an examination of the long-range organisation of histone genes in the chicken genome.

4.2 RESULTS

4.2.1 Screening of the Chicken Genomic Library

(a) Screening with Core Histone Probes

The chicken library of Dodgson *et al* (1979) consists of 5×10^5 independent recombinants containing large fragments (15-21 kb) cloned into the bacteriophage vector Charon 4a, and was constructed from a partial *Alu*I and *Hae*III digest of chicken DNA using the procedure described by Maniatis *et al* (1978).

It was initially decided to screen this library with a combination of the four core histone gene probes and generate a histone gene "clone bank", (i.e. a bank of clones probably representing the majority of

chicken histone genes) and to use this clone bank for any subsequent screenings.

The number of 'phage required for a "complete library" (i.e. a library having 99% probability of containing any sequence present in the genome) is 5.4×10^5 for chicken (Clarke and Carbon, 1976; chicken genome 2×10^9bp, see Old and Woodland, 1984).

A total of 6×10^5 'phage were screened by plating 10(15 cm diameter) plates each containing 6×10^4 'phage. Plaques from each plate were transferred to nitrocellulose filters using the procedure of Benton and Davis (1977) (section 2.2.5a). These filters were hybridised with a total of 2.4×10^7 cpm of ^{32}P-labelled DNA representing all 4 core histone gene sequences (6×10^6 cpm per histone gene sequence). Histone gene inserts to be used as probes were prepared by *Hin*dIII digestion of pCH.H_2AH, pCH.H_2BH, pCH.H4H and *Eco*RI/*Sal*I digestion of pCH.H3ES (see chapter 3) and purified by sucrose gradient centrifugation (Fig. 3.8) (section 2.2.4). It was necessary to separate these fragments from the plasmid vector to ensure a high specific activity and to avoid possible cross-reaction of vector pBR322 with *E. coli* sequences or with 'phage Charon 4a arms. These inserts were then labelled by nick-translation incorporating α-^{32}P-dATP and α-^{32}P-dCTP (section 2.2.11b).

The filters were washed several times at 65°C in IxSSC, 0.1% SDS and autoradiographed overnight at -80°C with an intensifying screen (see Fig. 4.1). Duplicate filters were not made in this case, and 30 possible positive plaques were picked and replated at lower density on small plates. Filters were again prepared and rescreened with the same probe. This enabled verification of the original signal and was necessary for eventual selection of an appropriate single plaque. Usually two extra rounds of screening were required. A final round of screening ensured purity of each isolate. Figure 4.2 shows signals from such a third round screening.

DNA was prepared from 12 clones (section 2.2.5) and subsequent restriction enzyme analysis revealed only five isolates differing from λCHO1 or λCHO2 (data not shown).

(b) Screening with probes to detect λCHO1 overlaps

In a separate experiment the chicken library was also screened with non-coding DNA fragments specific for λCHO1, and λCHO1 overlaps.

The fragments used as probes were a 2.0 kb *Eco*RI-*Sma*I fragment from the right-hand end of λCHO1 (Fig. 3.1), containing no detectable core histone gene sequences, and a 0.84 kb *Sal*I-*Pst*I fragment from the left-hand end of λCHO1. The latter fragment contains 30 bp of H3 coding region but given the overall size of the probe it was unlikely that this H3 sequence would interfere with the screening. These fragments were prepared by double digestion of either pCH3.3E (*Eco*RI and *Sma*I) or pCH3.75EH (*Pst*I and *Sal*I) and purified by LGT gel electrophoresis (section 2.2.2a). 500 ng of each purified DNA fragment was labelled by nick-translation with α-^{32}P-dATP and α-^{32}P-dCTP (section 2.2.11b).

For this screening 4×10^5 'phage were plated on five plates (15cm diameter) so that each contained 8×10^4 plaques. In this case duplicate nitrocellulose filters were made from each plate (section 2.2.5a) and hybridised with 13×10^6 cpm of ^{32}P-labelled DNA (5×10^6 cpm of 2.0 kb *Eco*RI-*Sma*I DNA; 8×10^6 cpm of 0.84 kb *Pst*I-*Sal*I DNA). Filters were washed and exposed as described in Figure 4.1. Duplicate filters were aligned and twenty eight plaques, detected in duplicate (Fig. 4.1), were picked and purified (Fig. 4.2). After three rounds of purification and screening, twenty five recombinants gave positive signals. These were picked for further characterisation (see section 4.2.3).

It should be noted that as only one complete library was screened with the presumptive λCHO1 specific probes, the number of positive

Figure 4.1

Primary Screening of a Chicken Recombinant Library

The chicken λ-library was plated at a density of $8x10^4$ pfu/plate on five 15 cm plates. Duplicate Benton and Davis filters were prepared (section 2.2.5a), and probed, in this screening, with a total of 1.3×10^7 cpm of ^{32}P-labelled probe, prepared by nick-translation (section 2.2.11b; 5.0×10^6 cpm derived from the 2.0 kb *Eco*RI-*Sma*I fragment and 8.0×10^6 cpm representing the 0.84 kb *Pst*I-*Sal*I fragment. Filters were washed in 1xSSC/0.1% SDS at 65°C and autoradiographed for two days, at -80°C with an intensifying screen.

From this screening a total of 28 positive plaques were detected in duplicate. These were picked for further purification. The signals, present in duplicate on the filters shown, are circled. The arrows indicate the positions of holes used for orientation.

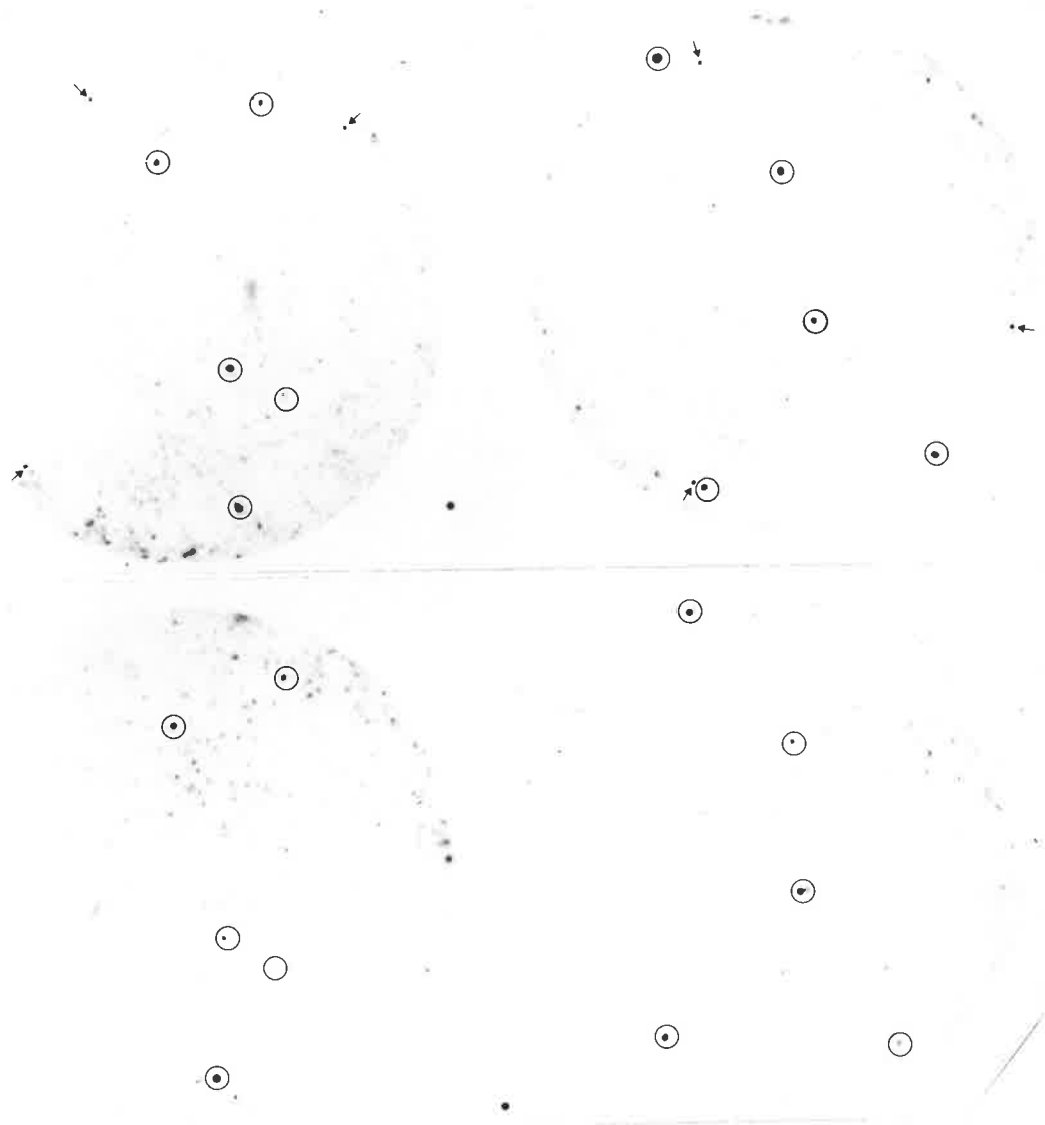


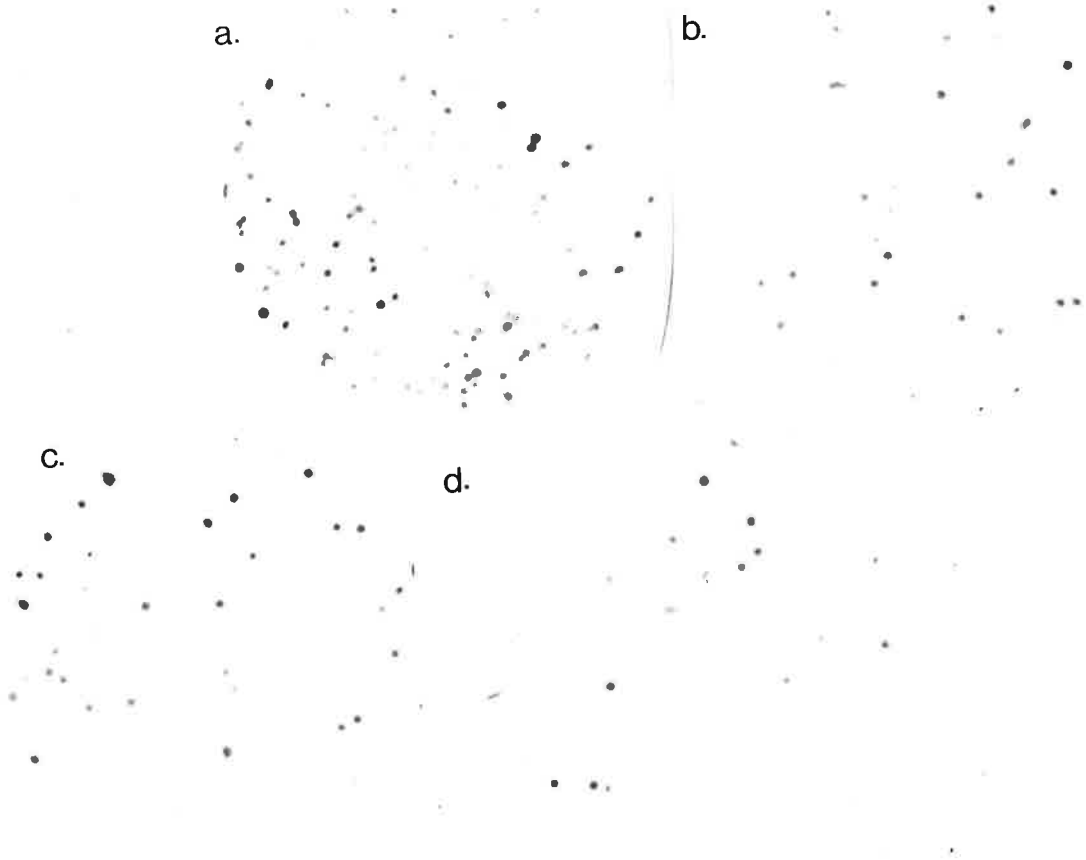
Figure 4.2

Third Round Screening of Primary Isolates

From a Chicken λ Library

Positive isolates, obtained after primary screenings of the chicken genomic λ-library (Fig. 4.1), were picked into PSB, titred and replated at a density of 1000-4000 pfu/plate. Benton and Davis filters were prepared (section 2.2.5a) and rehybridised with ^{32}P-labelled probe (details of the probe used in this screening are given in Fig. 4.1). Plaques corresponding to positive signals were picked into PSB, titred, and plated at a density of 100-500 pfu/plate. Filters were prepared and hybridised with approximately 10^6 cpm per filter of the same probe, washed in 1xSSC/0.1% SDS at 65°C and autoradiographed at -80°C, for 1-2 days with an intensifying screen.

Strong positive responses were evident after two rounds of purification on the four filters shown (a-d). Single positive plaques were picked for further analysis.


a.
b.
c.
d.

plaques expected from this screening was relatively low (i.e. less than ten). The high number of plaques detected suggested the possibility that one or both of the probes contained sequences which were not unique to λCHO1 and had detected clones containing similar sequences from elsewhere in the genome (refer to section 4.2.2).

4.2.2 The 2.0kb *Eco*RI-*Sma*I Fragment of pCH3.3E Contains an H1 Gene

A number of results (not discussed) had indicated indirectly, the possibility that an H1 gene was contained within one of the fragments used to screen the chicken library for clones overlapping λCHO1 (section 4.2.1b).

To determine whether H1 coding sequences were located in pCH3.3E or pCH3.75EH, 0.5 ug of DNA from both clones was denatured and spotted onto a nitrocellulose filter (section 2.2.3). Hybridisation with an H1 specific probe (Provided by L.S. Coles, this laboratory) indicated that pCH3.3E possibly contained H1 sequences (Fig. 4.3a).

In order to confirm this result and locate H1 sequences, 1 ug of pCH3.3E DNA was digested with *Eco*RI and *Sma*I, fractionated on a 1% agarose gel and transferred to a nitrocellulose filter. Hybridisation of this DNA with this same H1 probe showed clearly that H1 sequences were located within the 2.0 kb *Eco*RI-*Sma*I fragment from pCH3.3E (Fig. 4.3b). This fragment has now been completely sequenced (Coles and Wells, 1985) and contains a single complete H1 gene.

Thus, in the experiment described (4.2.1b), the chicken library was essentially screened for H1 genes. L.S. Coles from this laboratory simultaneously screened the same library with an H1 gene-specific fragment and purified a number of putative H1 containing genomic clones. In total, 50 positive isolates were purified from three screenings (sections 4.2.1a and 4.2.1b) and these subsequently have been shown to encode the entire chicken H1 complement (D'Andrea *et al*, 1985).

These fifty isolates have been sorted into 12 individual groups

Figure 4.3

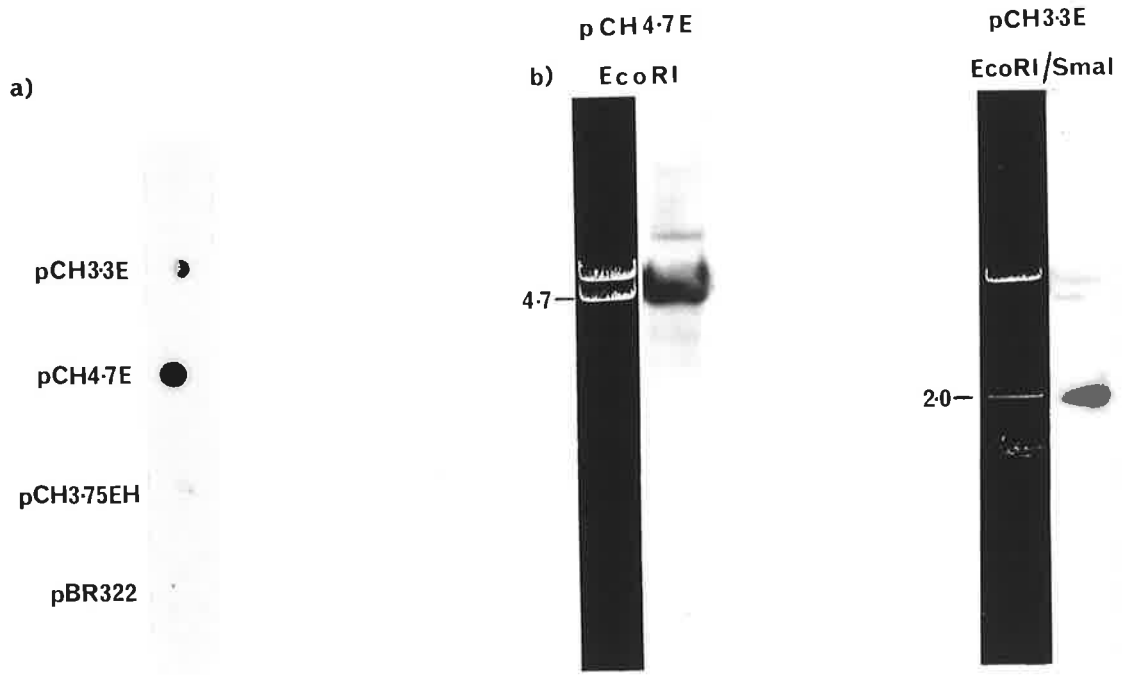
Detection of H1 Sequences in λCHO1

a) Dot-blot analysis of λCHO1 derived subclones

A dot-blot filter (section 2.2.3) was prepared using 0.5 ug samples of pCH3.3E and pCH3.75EH (see Fig. 3.1) with similar amounts of the H1 histone gene-containing chicken subclone pCH4.7E (Harvey *et al*, 1981) and pBR322 as positive and negative controls. The filter was hybridised with ^{32}P-labelled H1 probe, washed in 1xSSC/0.1% SDS at 65^{o}C and autoradiographed for 1 hour at -80^{o}C with an intensifying screen. The details of the procedures used for the probe preparation and hybridisation are given in sections 2.2.11b and 2.2.3 respectively.

b) Southern analysis of pCH3.3E DNA

1 ug of pCH3.3E DNA was digested with *Eco*RI and *Sma*I and fractionated on a 1% agarose gel. As a positive control 1 ug of pCH4.7E DNA was digested with *Eco*RI for similar analysis. Bands were detected by ethidium bromide staining and the DNA subsequently transferred to a nitrocellulose filter using the procedure described in section 2.2.3. Filter-bound DNA was hybridised with radiolabelled H1 probe, prepared by nick translation (section 2.2.11b), washed in 2xSSC/0.1% SDS at 65^{o}C and autoradiographed overnight at -80^{o}C with an intensifying screen. The sizes of the relevant fragments are given in kilobases.


a)
pCH3·3E
pCH4·7E
pCH3·75EH
pBR322
b)
pCH4·7E
EcoRI
4·7
pCH3·3E
EcoRI/SmaI
2·0

based on their gene content (as determined by dot-Bentons, this work, section 2.2.5a, data not shown) and their *Eco*RI restriction endonuclease digestion patterns (L.S. Coles, this laboratory). This data is summarised in Table 4.1.

Collectively, genomic clones have been isolated from three independent library screenings. The observation that many of these clones contain core and H1 histone genes suggests that these genes frequently occur together and suggests, further, that this collection of λ-clones should constitute a major proportion of the chicken histone genes.

4.2.3 Isolation of Clones Overlapping λCHO1

This collection of histone gene containing recombinants was then screened for clones which overlap λCHO1. Non-coding probes from each end of λCHO1 were prepared, and since the 2.0 kb *Eco*RI-*Sma*I fragment originally used was subsequently shown to contain an H1 gene, another fragment from this region was isolated. The fragment used was a 0.18 kb *Xho*I-*Sma*I fragment from between the H_2A and the H1 gene in pCH3.3E (refer Fig. 3.1). This fragment has been sequenced (D'Andrea *et al*, 1981) and is known to be non-coding. It was also easily purified by LGT gel electrophoresis after *Xho*I/*Sma*I double digestion of pCH3.3E.

The twenty five clones isolated from the screening described in section 4.2.1b were screened using the "dot Benton" procedure, described in section 2.2.5a, with either the 0.18 kb *Xho*I-*Sma*I fragment or the 0.84 kb *Pst*I-*Sal*I fragment (section 4.2.1b) as ^{32}P-labelled probes. The result of the screening is shown in Figure 4.4. The filters were also probed, separately, with nick-translated H_2B and H3 insert. Since these genes bracket λCHO1 (Fig. 3.1) they should be present in putative overlaps.

Clone 2.5 hybridised to the 0.84 kb *Sal* I-*Pst*I fragment and the H3 gene specific probe. Clearly, it did not hybridise to the 0.18 kb

Table 4.1

The table shows the core histone gene containing recombinants isolated and characterised in this laboratory.

(a) *Eco*RI digestion products of λ-recombinants. Fragment sizes are indicated in kilobases. Clones with the prefix λH1 were characterised by L.S. Coles. Data for λCHO1 and λCHO2 is from Harvey *et al* (1981).

b) Histone gene content of λ-recombinants as determined by "Dot-Bentons". (+) indicates faint hybridisation to a histone gene specific probe.

Explanatory Note

"Dot-Benton" analysis of histone gene containing λ-recombinants was carried out in association with L.S. Coles, who subsequently characterised H1 gene containing clones which did not overlap λCHO1.

Table 4.1

Core Histone Gene Containing λ-Recombinants

A.	λH1.1	λH1.2	λH1.3	λH1.4	λ4.2	λ2.1	λH1.7	λ2.5	λH1.9	λH1.10	λ7.2	λCHO1	λCHO2
						λ5.5							
	13.5	7.4	8.6	9.7	5.0	7.0	6.7	8.4	14.3	16.0	5.5	10.0	10.0
	0.2	5.0	7.2	5.7	5.0	4.0	4.4	5.0			4.9	3.3	4.7
		3.6			2.0	0.7	1.3	0.5			3.5	0.7	
		0.4			0.7	0.7					0.9		
						0.6					0.7		
						λCHO5		λCHO3			λCHO7		

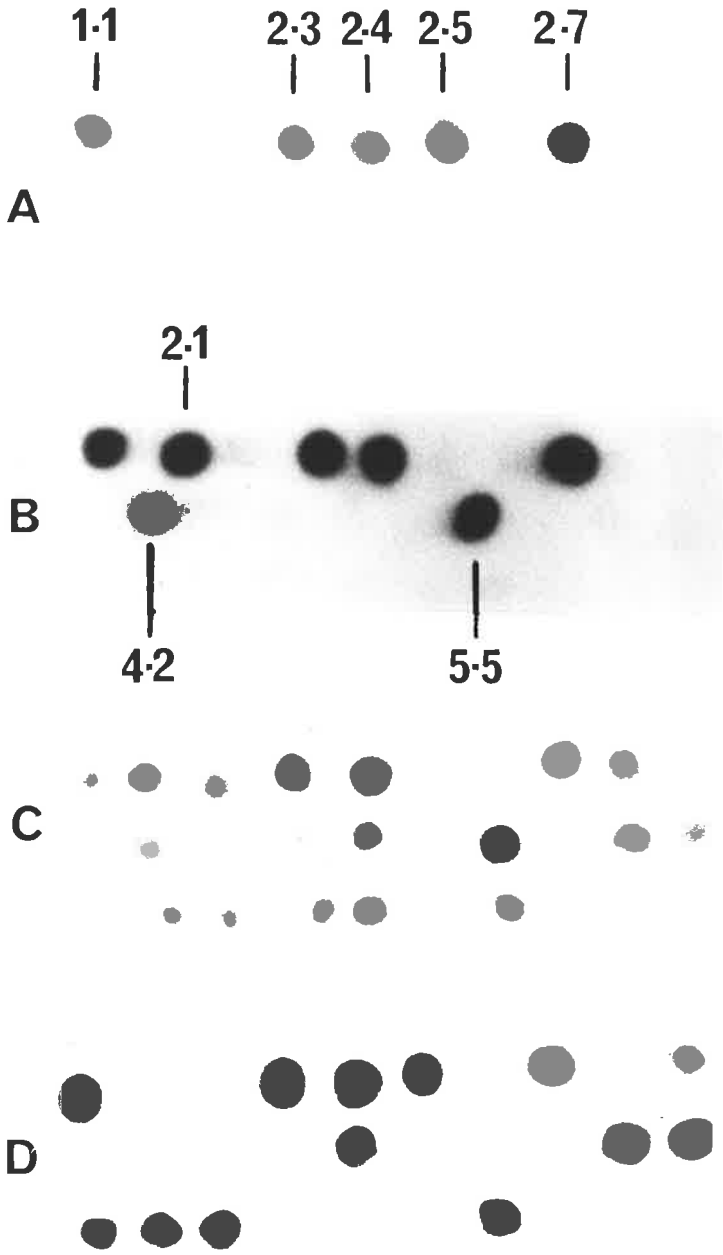
B.													
	H1 +	+	+	+	+	+	+	+	+	+		+	+
	H_2A +	(+)	+	+	+	+	(+)	+		+	+	+	(+)
	H_2B	+	+	+	+	+	+			+	+	+	+
	H3		+	+					+			+	+
	H4	+		+					+	+		+	+

Figure 4.4

Detection of λCHO1 Overlapping Recombinants

The twenty-five positive isolates derived from the library screening described in section 4.2.1b were spotted onto freshly plated lawns of LE392 and incubated, at 37°C, overnight to form ordered arrays of plaques (see section 2.2.5a). Nitrocellulose filters were prepared from these plates (section 2.2.5a) and hybridised to ^{32}P-labelled probes (see below). Filters were washed in 2xSSC/0.1% SDS at 65°C and autoradiographed overnight at -80°C with an intensifying screen. The isolates discussed in the text are indicated.

Filter	Probe
A	0.84 kb *Pst*I-*Sal*I fragment
B	0.18 kb *Xho*I-*Sma*I fragment
C	H_2B gene-specific probe
D	H3 gene-specific probe


1·1
2·3
2·4
2·5
2·7
A
2·1
B
4·2
5·5
C
D

*Xho*I-*Sma*I fragment from the right-hand end of λCHO1, or to the H_2B specific probe. These results are consistent with this clone being an overlap for the H3 end of λCHO1. Subsequently this clone is referred to as λCHO3 (section 5.2.2).

Clones 2.1, 4.2 and 5.5 are positive with the 0.18 kb *Xho*I-*Sma*I probe, but negative when hybridised to the 0.84 kb *Pst*I-*Sal*I fragment. These clones contained at least one H_2B gene consistent with them overlapping λCHO1 at the right hand end (as drawn, Fig. 3.1). None of these clones contain an H3 gene. Digestion of DNA prepared from these clones (using the high speed spin procedure described in section 2.2.5b) indicated that 2.1 and 5.5 were identical and different to 4.2 (data not shown). Clone 4.2 was partially characterised by D.L. Dodd in this laboratory but is not discussed further. Clones 2.1 and 5.5 were referred to as λCHO5 and have been fully characterised (section 5.2.6).

Four clones, viz; 2.3, 2.4 and 2.7 were positive with both H3 and H_2B gene-specific probes and both λCHO1 specific non-coding probes. These clones were all shown to be identical to λCHO1 by comparison of their restriction endonuclease digestion patterns.

4.2.4 Discussion

"Chromosome walking" involves the selection of overlapping clones from a genomic library. Non-coding probes from one clone are used to select clones generated from the same region of the genome. This procedure requires a library which is a collection of random cleavage fragments. It has been used, in this chapter, to isolate genomic clones which overlap and extend the genomic clone λCHO1 (Harvey and Wells, 1979). These clones are discussed in detail in chapter 5.

CHAPTER 5

CHARACTERISATION OF CHICKEN HISTONE GENOMIC CLONES

5.1 INTRODUCTION

The aim of the work presented in this chapter was to extend the data available on chicken histone gene organisation. The histone gene content of the available genomic clones was determined in Chapter 4. In this chapter the arrangement of the histone genes within two clones overlapping λCHO1 viz; λCHO3 and λCHO5 is presented. An overlap with λCHO5 has been isolated and is also characterised in this chapter. Two genomic clones, characterised by Sugarman *et al* (1983), overlap with λCHO3 and further extend this region.

Restriction maps for all of these clones were determined using the restriction enzymes *Eco*RI, *Hin*dIII, *Bam*HI and *Sal*I. *Eco*RI fragments containing histone gene coding regions were subcloned into pBR325 and analysed in detail to determine the location and arrangement of the genes. This method of characterisation has enabled a complete analysis of this histone gene locus.

5.2 RESULTS

5.2.1 Preparation of Bacteriophage DNA

'Phage DNA was prepared from plate stocks as described in Methods (section 2.2.5b). Initially 'phage were concentrated, after harvesting, by high speed centrifugation (20,000 rpm, 3 hours). This was time consuming and resulted in very firm 'phage pellets which were difficult to resuspend. In later experiments 'phage were concentrated by polyethylene glycol precipitation.

'Phage were further purified by centrifugation through CsCl step gradients (section 2.2.5b) and DNA purified by phenol/chloroform extraction and ethanol precipitation. This procedure resulted in good yields of 'phage DNA that would digest with most restriction enzymes.

DNA that failed to digest with restriction enzymes was purified further by centrifugation on 10-40% sucrose gradients (section 2.2.4).

This additional purification generally resulted in improved enzyme digestion.

5.2.2 Restriction Analysis of λCHO3

λCHO3 was digested in single and double digests with the enzymes *Eco*RI, *Hin*dIII, *Bam*HI and *Sal*I and electrophoresed on 1% and 1.5% agarose gels. *Hin*dIII digested λDNA and *Hin*fI digested pBR322 DNA were co-electrophoresed as molecular weight references. In addition, restriction fragments derived wholly from vector arm sequences were used as size markers.

Digestion of λCHO3 with *Eco*RI resulted in three insert fragments, 8.4 kb, 5.0 kb and 0.5 kb. Only the two large fragments were seen on the 1% agarose gel, (Fig. 5.1), while the 0.5 kb *Eco*RI-*Eco*RI fragment was observed on the 1.5% agarose gel (Fig. 5.2).

The 8.4 kb *Eco*RI-*Eco*RI fragment was cleaved by *Hin*dIII to generate an *Eco*RI-*Hin*dIII fragment of 7.5 kb (Fig. 5.1). The observation that there was no visible 0.9 kb *Eco*RI-*Hin*dIII fragment on the 1% agarose gel suggested that a number of *Hin*dIII sites were present at one end of the 8.4 kb *Eco*RI-*Eco*RI fragment and that the *Hin*dIII-*Hin*dIII and *Hin*dIII-*Eco*RI fragments generated were less than 0.5 kb and therefore not present on this gel. The presence of the small internal *Hin*dIII fragments was confirmed by *Hin*dIII digestion which produced a large internal 8.0 kb *Hin*dIII-*Hin*dIII fragment and two small internal fragments (see 1.5% agarose gel; Fig. 5.2). These two small *Hin*dIII-*Hin*dIII fragments (0.5 kb and 0.4 kb) were placed at the right of the insert to accomodate the 20.0 kb *Hin*dIII-*Hin*dIII fragment, however the relative positions of these fragments could not be determined from this data and was later determined from mapping of the subclone pCH8.4E (see section 5.2.5c). These two fragments were also present in λCHO1 (Fig. 5.2) confirming the overlap of this clone and λCHO3. Given the location of these *Hin*dIII sites in λCHO1 the extent of the overlap must be at least 4.5 kb

Figure 5.1

Restriction Analysis of λCHO3

1 ug samples of λCHO3 DNA were digested with a range of restriction enzymes either singly or in combination, and the fragments resolved on a 1% agarose gel (section 2.2.2a). Bands were observed with UV light following ethidium bromide staining. *Bam*HI and *Sma*I did not cleave the λCHO3 insert. The figure shows λCHO3 digested with the following restriction enzymes (fragment sizes are indicated in kilobases);

Track A: *Eco*RI

Track B: *Hin*dIII

Track C: *Eco*RI and *Hin*dIII

Track D: *Sal*I

Track E: *Eco*RI and *Sal*I

The sizes of DNA fragments obtained on restriction enzyme digestion of λ-recombinants and derived subclones were determined using *Hin*dIII digested λ DNA and *Hin*fI digested pBR322 DNA as molecular weight markers. In addition *Hin*dIII and *Bam*HI fragments derived wholly from vector regions also served as size standards.

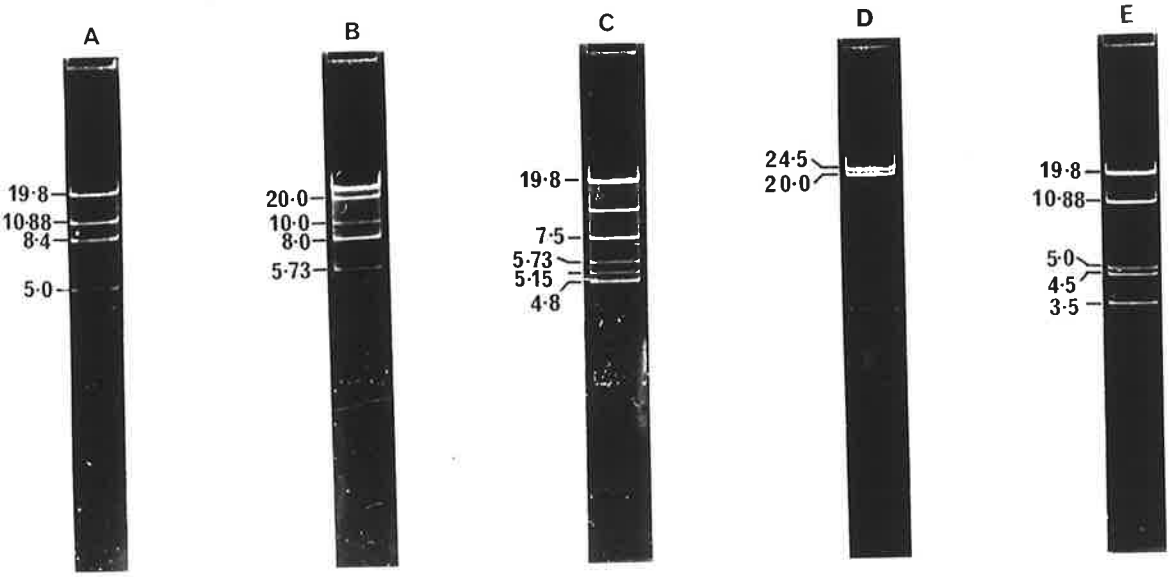

A
19·8
10·88
8·4
5·0
B
20·0
10·0
8·0
5·73
C
19·8
7·5
5·73
5·15
4·8
D
24·5
20·0
E
19·8
10·88
5·0
4·5
3·5

Figure 5.2

Restriction Analysis of λCHO3 : Comparison with λCHO1

1 ug samples of λCHO1 and λCHO3 were digested with the enzymes indicated below and fractionated on a 1.5% agarose gel (section 2.2.2a). Bands were observed under UV light after ethidium bromide staining. Fragment sizes are indicated in kilobases.

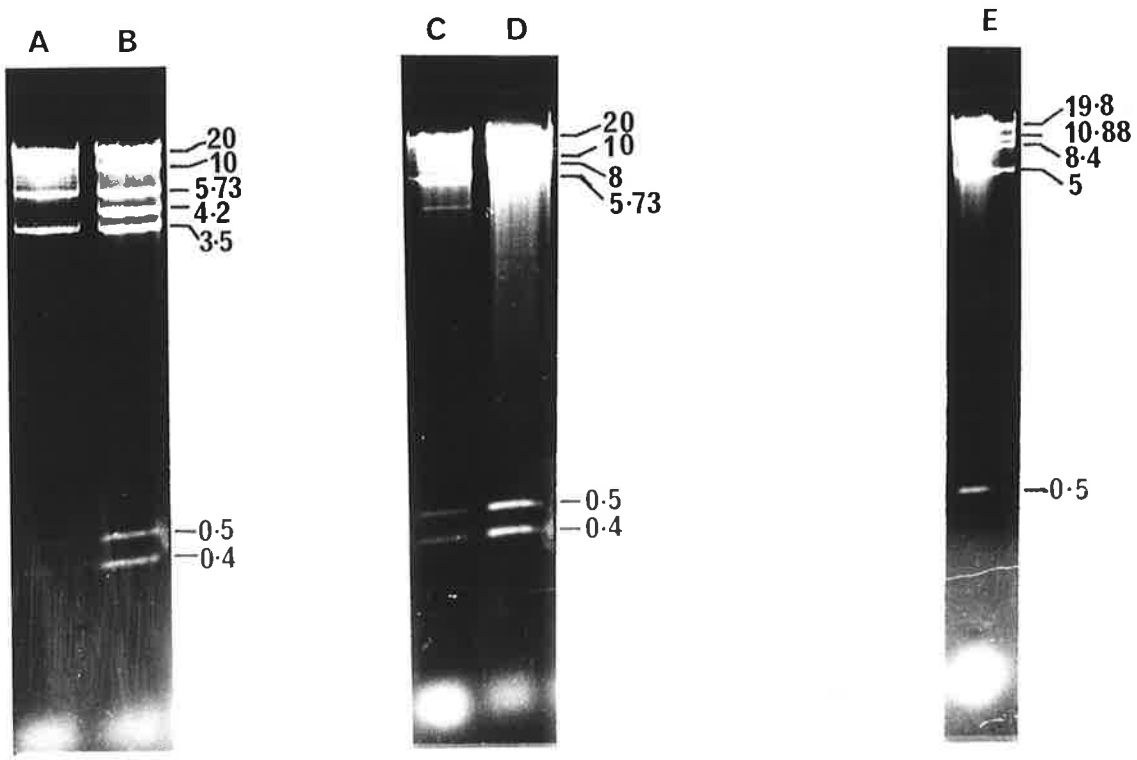
Track A: λCHO1 digested with *Sal*I and *Hin*dIII

Track B: λCHO3 digested with *Sal*I and *Hin*dIII

Track C: λCHO1 digested with *Hin*dIII

Track D: λCHO3 digested with *Hin*dIII

Track E: λCHO3 digested with *Eco*RI


A
B
20
10
5·73
4·2
3·5
0·5
0·4
C
D
20
10
8
5·73
0·5
0·4
E
19·8
10·88
8·4
5
0·5

Furthermore the 8.0 kb *Hin*dIII-*Hin*dIII fragment must link the two large *Eco*RI-*Eco*RI fragments. This *Hin*dIII-*Hin*dIII fragment is cleaved by *Eco*RI to generate the 7.5 kb *Eco*RI-*Hin*dIII fragment, however the 5.0 kb *Eco*RI-*Eco*RI on cleavage by *Hin*dIII generates a 4.8 kb *Eco*RI-*Hin*dIII fragment. The other predicted *Eco*RI-*Hin*dIII fragment is very small and was not observed in this analysis. Confirmation that this fragment existed was possible on restriction analysis of pCH5.0E (see section 5.2.5b). It is possible to account for this very small *Eco*RI-*Hin*dIII fragment by placing it adjacent to the 0.5 kb *Eco*RI-*Eco*RI fragment within the large *Hin*dIII-*Hin*dIII fragment, as shown in Figure 5.3.

*Sal*I restricted λCHO3 once within the insert and generated two very large fragments (Fig. 5.1). A *Sal*I/*Eco*RI digest indicated that the single *Sal*I restriction site was contained in the 8.4 kb *Eco*RI-*Eco*RI fragment. This fragment was cleaved to generate *Sal*I-*Eco*RI fragments sized at 3.5 kb and 4.5 kb (Fig. 5.1).

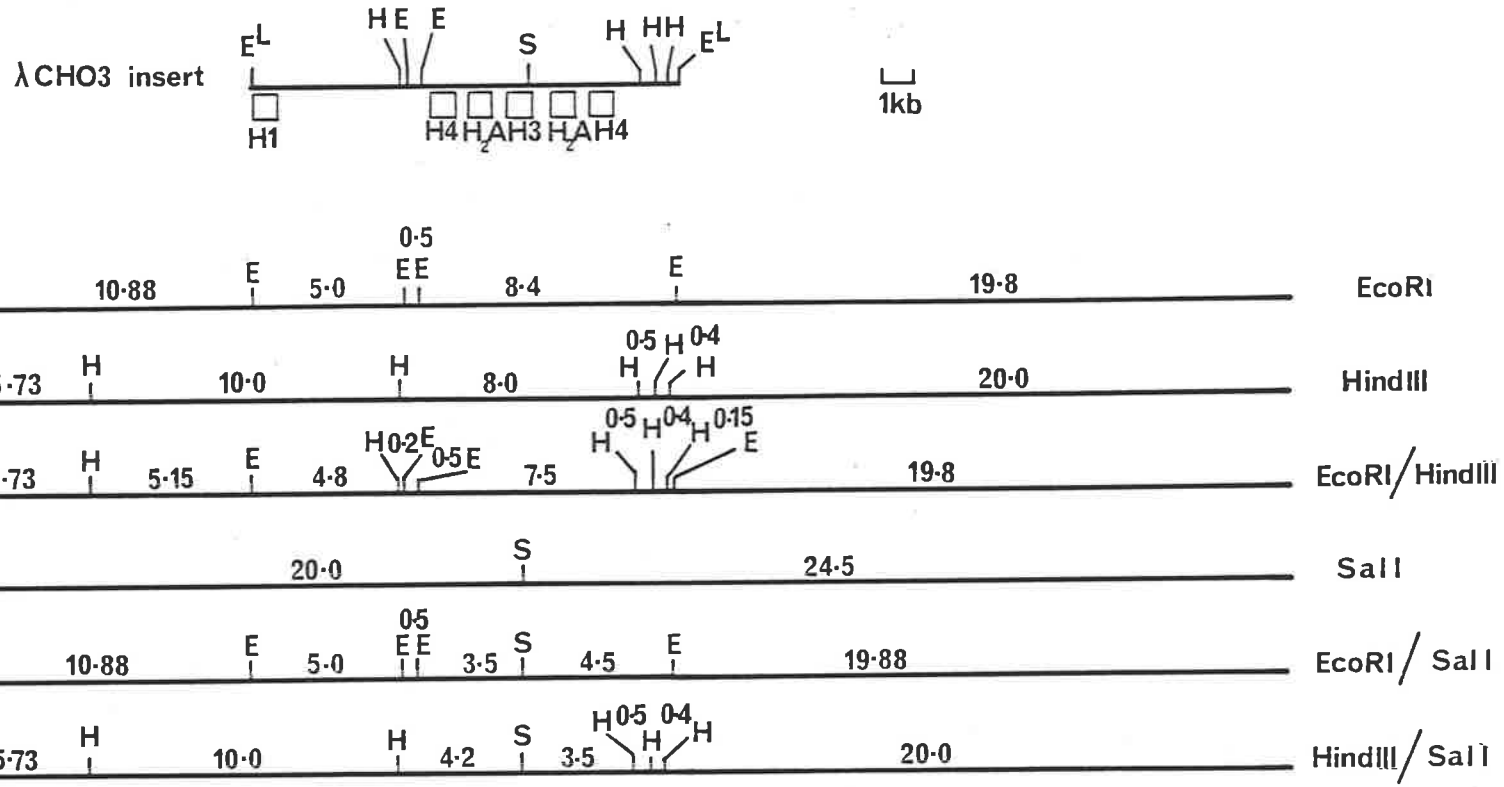
The mapping data from λCHO1 predicts a 3.5 kb *Sal*I-*Hin*dIII fragment will be common to both λCHO1 and λCHO3, assuming that both the *Sal*I and *Hin*dIII sites in λCHO1 are conserved in λCHO3. This fragment was observed on *Sal*I/*Hin*dIII digestion of the two clones (Fig. 5.2). For this *Sal*I-*Hin*dIII fragment and the two small *Hin*dIII fragments to be completely contained in the 8.4 kb *Eco*RI-*Eco*RI fragment the 4.5 kb *Sal*I-*Eco*RI fragment must span this region (Fig. 5.3). As the conserved *Sal*I site is at the end of λCHO1, the size of this *Sal*I-*Eco*RI fragment from λCHO3 effectively represents the extent of overlap of λCHO3 with λCHO1.

From this data it was clear that λCHO3 overlapped with λCHO1 but also contained approximately 10 kb of previously uncharacterised DNA.

Figure 5.3

Restriction Maps of λCHO3

The restriction fragments resulting from digestion of λCHO3 DNA with different combinations of *Eco*RI (E), *Sal*I (S) and *Hin*dIII (H) are shown, together with the overall derived restriction map of the λCHO3 insert. The boxes indicate the predicted gene positions as determined from Southern analysis. Fragment sizes are given in kilobases. *Bam*HI does not cleave the λCHO3 insert.


λCHO3 insert
E^L
H E E
S
H HH
E^L
H1
H4 H2A H3 H2A H4
1kb
10·88
E
5·0
0·5
E E
8·4
E
19·8
EcoRI
5·73
H
10·0
H
8·0
0·5 H 0·4
H H
20·0
HindIII
5·73
H
5·15
E
4·8
H 0·2 E
0·5 E
7·5
0·5 H 0·4 H 0·15
H E
19·8
EcoRI / HindIII
20·0
S
24·5
Sal I
10·88
E
5·0
0·5
E E
3·5
S
4·5
E
19·88
EcoRI / Sal I
5·73
H
10·0
H
4·2
S
3·5
H 0·5 0·4
H H
20·0
HindIII / Sal I

5.2.3 The Coding Potential of λCHO3

To determine which fragments within λCHO3 contained histone gene sequences, the digests shown in Figure 5.1 were transferred to nitrocellulose, using the bidirectional transfer technique (section 2.2.3), and filters were hybridised with individual histone gene probes (Fig. 5.4). When reprobing with a second labelled histone gene specific fragment, filters were boiled for 5-10 minutes to remove hybridised DNA, prehybridised again and reprobed.

The 8.4 kb *Eco*RI-*Eco*RI fragment is positive with the H_2A, H4 and H3 specific probes. The H3 gene probe only hybridised to the smaller 3.5 kb *Eco*RI-*Sal*I fragment. This is as expected as this fragment extends the fragment in λCHO1 from which the H3 probe was derived; that is, the 0.15 kb *Eco*RI-*Sal*I fragment in λCHO1. By analogy with λCHO1 (Harvey *et al*, 1981; Fig. 3.1) an H3 gene should span the *Sal*I site in λCHO3.

The H1 gene probe only hybridised to the 5.0 kb *Eco*RI-*Eco*RI fragment. The H_2A and H4 probes hybridised to both the 4.5 kb and 3.5 kb *Eco*RI-*Sal*I fragments, indicating that at least one copy of each of these genes exists in both of these fragments.

These results suggest that λCHO3 contains the H_2A, H4 and H3 gene cluster which is present at the left hand end of λCHO1 (Fig. 3.1). In addition, λCHO3 contains other H_2A and H4 genes within the 8.4 kb *Eco*RI-*Eco*RI fragment, and an H1 gene within the 5.0 kb *Eco*RI-*Eco*RI fragment.

5.2.4 Isolation of λCHO3 Subclones

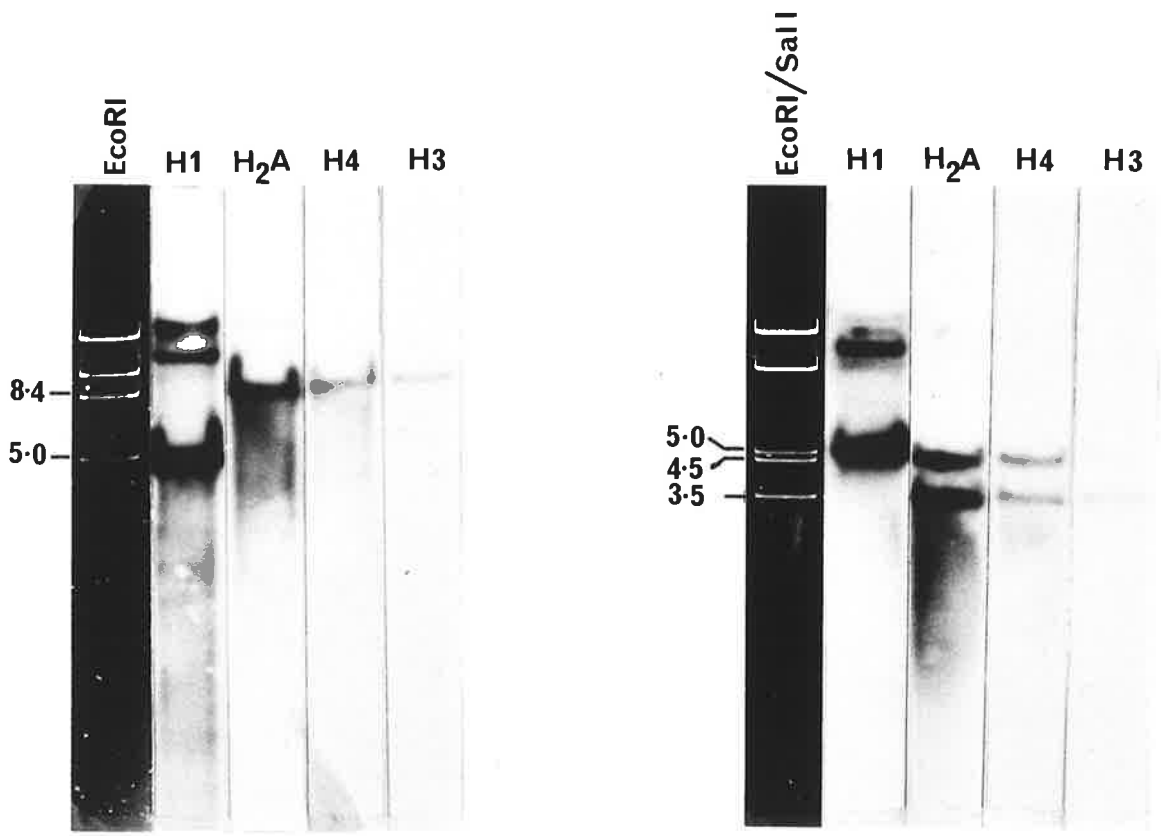
In order to define in more detail the restriction map of λCHO3 and the location, number and arrangement of the individual histone genes, the two large *Eco*RI-*Eco*RI fragments containing histone coding sequences were subcloned into the plasmid vector pBR325.

Figure 5.4

Southern Analysis of λCHO3

λCHO3 DNA which had been digested with various restriction enzymes and fractionated on a 1% agarose gel was transferred to nitrocellulose filters using the bidirectional transfer procedure (section 2.2.3). Filters were hybridised individually with nick-translated H3 and H4 histone gene specific probes, washed in 0.5xSSC/0.1% SDS at 65°C and autoradiographed at -80°C with an intensifying screen. Only the digests discussed in the text are shown in the figure. Restriction fragment sizes are in kilobases.

After exposure filters were boiled in distilled water, to remove annealed probe, prehybridised again and probed with H1 and H_2A gene specific probes.


EcoRI
H1
H2A
H4
H3
8·4
5·0
EcoRI/SalI
H1
H2A
H4
H3
5·0
4·5
3·5

After *Eco*RI digestion of λCH03 the 8.4 kb and 5.0 kb *Eco*RI-*Eco*RI fragments were separated by LGT agarose gel electrophoresis and eluted (section 2.2.2a). These fragments were ligated with plasmid DNA which had been linearised with *Eco*RI and incubated with Calf Intestinal Phosphatase to prevent self ligation (section 2.2.7). A sample of each ligation mixture was transformed into MC1061 and plated on L-agar plates containing 30 ug/ml tetracycline.

Possible recombinant transformants were identified and plasmid DNA was assayed for inserts, after digestion, by electrophoresis on a 1% agarose gel (Fig. 5.5). DNA was prepared from a single recombinant colony using the large scale procedure described in Chapter 2 (section 2.2.10). These recombinants are referred to as pCH8.4E (8.4 kb *Eco*RI-*Eco*RI insert) and pCH5.0E (5.0 kb *Eco*RI-*Eco*RI insert). Restriction enzyme mapping of these subclones is considered in detail in section 5.2.5.

5.2.5 Restriction Analysis of λCH03 Subclones

a) Introduction

Initial mapping of histone gene subclones was approached using a standard strategy. *Sac*I, *Sac*II and *Pst*I were found to have the most useful distributions of recognition sites and were of general use in subclone characterisation.

The strategy for mapping is described in the following example. For an *Eco*RI-*Eco*RI fragment with no internal *Bam*HI sites, three initial digests were carried out for each enzyme for which sites were being determined, that is, for *Sac*I sites, plasmid DNA was digested with *Sac*I, *Sac*I/*Eco*RI and *Sac*I/*Bam*HI. The following information was gained from these digestions:

1. The number of restriction sites for this enzyme (*Sac*I) within the cloned fragment.

Figure 5.5

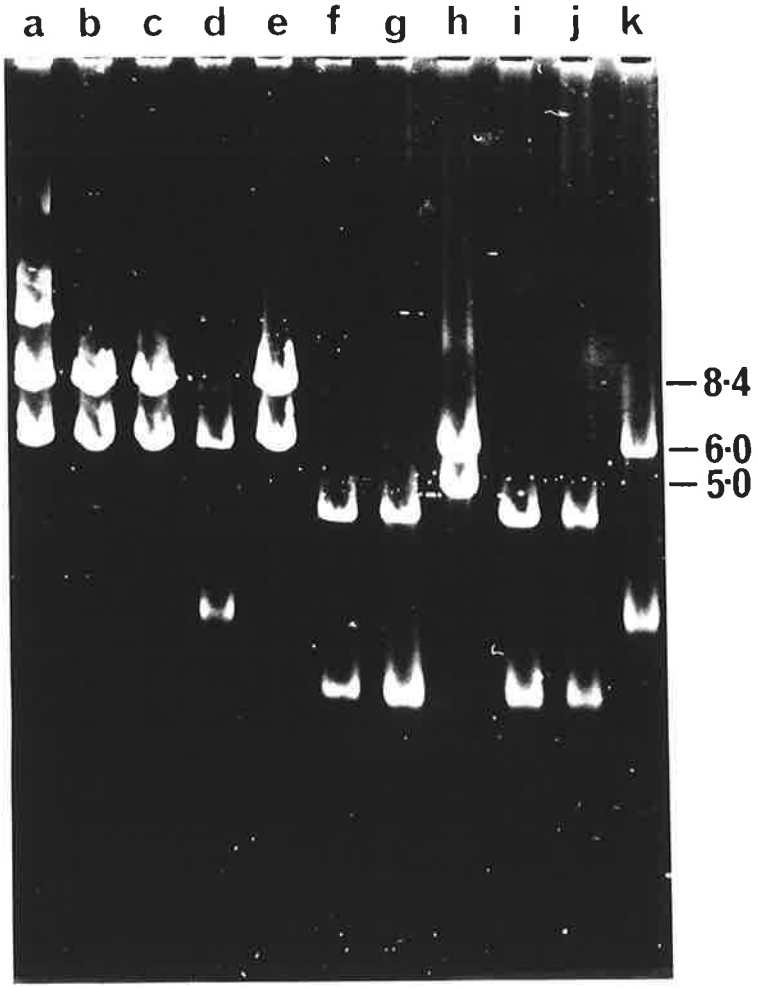
"Miniscreen" Examination of λCHO3 Subclones

A small amount of plasmid DNA was prepared from *cap*^{s} *tet*^{r} colonies using the miniscreen procedure (section 2.2.9b). An aliquot of each was digested with *Eco*RI, electrophoresed on a 1.2% agarose gel (section 2.2.2a) and visualised under UV light following ethidium bromide staining. The tracks are as follows:

Tracks (a-e): DNA from transformants derived from the ligation mix containing the 8.4 kb *Eco*RI-*Eco*RI fragment.

Tracks (f-k): DNA from transformants derived from the ligation mix containing the 5.0 kb *Eco*RI-*Eco*RI fragment.

Fragment sizes are indicated in kilobases.


a b c d e f g h i j k
8·4
6·0
5·0

2. The distance from the terminal *Sac*I restriction sites to the *Eco*RI sites could be determined from the sizes of the *Sac*I-*Eco*RI fragments in the *Sac*I/*Eco*RI double digest. These fragments could be identified by comparison with the *Sac*I digest.

3. As there is a known *Bam*HI site 1.6 kb from the vector cloning site, and no *Bam*HI sites within the insert, the sizes of the *Sac*I-*Bam*HI fragments can be used to position the *Sac*I sites at each end of the insert relative to the vector *Bam*HI site. The two *Sac*I-*Eco*RI fragments can then be positioned by accommodating the sizes of the *Sac*I-*Bam*HI fragments.

This procedure determines the orientation of the insert in the vector, relative to the *Bam*HI site, but the arrangement of any internal *Sac*I fragments cannot be deduced.

Once this information was determined for these enzymes (*Sac*I, *Sac*II and *Pst*I) the internal fragments could be placed by comparing single and double digests for enzymes which cleave the insert. A number of restriction patterns could be predicted from the general strategy described above, and these could be progressively confirmed or eliminated by further digestion. Restriction fragment sizes were generally determined by electrophoresis on 1%, 1.2%, 1.5% or 2% agarose gels. These fragments could be transferred to nitrocellulose filters and the location of histone gene coding regions determined by hybridisation to histone gene specific probes. Small fragments (less than 1 kb) were sized by electrophoresis alongside marker DNA on 5% or 6% polyacrylamide gels.

b) Restriction Analysis of pCH5.0E

pCH5.0E contains the 5.0 kb *Eco*RI-*Eco*RI, H1 encoding fragment of λCHO3, subcloned into the *Eco*RI site of pBR325. DNA from pCH5.0E was mapped with *Sac*I, *Sac*II and *Pst*I restriction endonucleases. Figures 5.6 and 5.7 present the relevant restriction digests which led to the restriction map presented in Figure 5.8.

Figure 5.6

Restriction Analysis of pCH5.OE

1 ug samples of pCH5.OE DNA were digested with a range of restriction enzymes, either singly or in combination, and the fragments resolved on a 1.2% agarose gel (section 2.2.2a). Bands were detected with UV light after ethidium bromide staining. Fragment sizes are indicated in kilobases. The figure shows pCH5.OE DNA digested with the following enzymes,

Track A: *Sac*I

Track B: *Sac*I and *Eco*RI

Track C: *Sac*I and *Bam*HI

Track D: *Sac*II

Track E: *Sac*II and *Eco*RI

Track F: *Sac*II and *Bam*HI

Track G: *Pst*I

Track H: *Pst*I and *Eco*RI

Track I: *Pst*I and *Sal*I

Track J: *Pst*I and *Hin*dIII

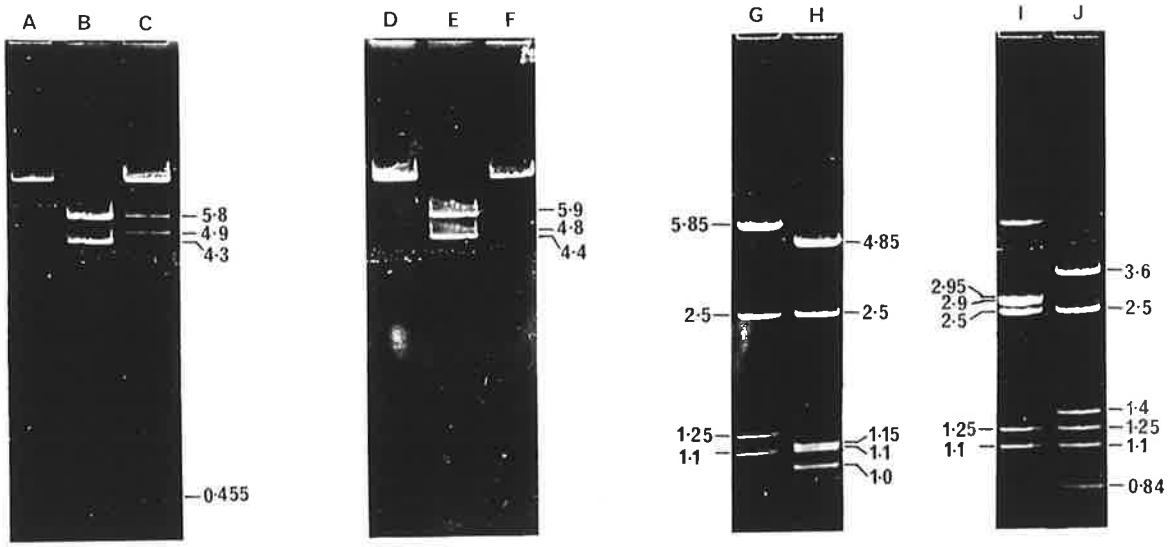

A
B
C
5·8
4·9
4·3
0·455
D
E
F
5·9
4·8
4·4
G
H
5·85
2·5
1·25
1·1
4·85
2·5
1·15
1·1
1·0
I
J
2·95
2·9
2·5
1·25
1·1
3·6
2·5
1·4
1·25
1·1
0·84

Figure 5.7

Restriction Analysis of pCH5.OE

To determine accurately the size of small restriction fragments (less than 1 kb in length) 1 ug samples of pCH5.OE DNA were digested with appropriate restriction enzymes (see below) and electrophoresed on a 5% polyacrylamide gel (section 2.2.2b). DNA was visualised with UV light after ethidium bromide staining and the bands were sized by comparison with DNA fragments generated by *Hin*fI cleavage of pBR322 DNA. The sizes determined for DNA fragments are given in kilobases alongside each track. Fragments marked with a superscript P are partial digestion products.

Samples of pCH5.OE DNA, digested with the following enzyme combinations are shown;

Track A: *Eco*RI and *Hin*dIII

Track B: *Pst*I and *Hin*dIII

Track C: *Eco*RI and *Pst*I

Track D: *Eco*RI and *Sac*I

Track E: *Eco*RI and *Sac*II

Track F: *Pst*I and *Sac*I

Track G: *Pst*I and *Sac*II

Track H: *Sac*I and *Sac*II

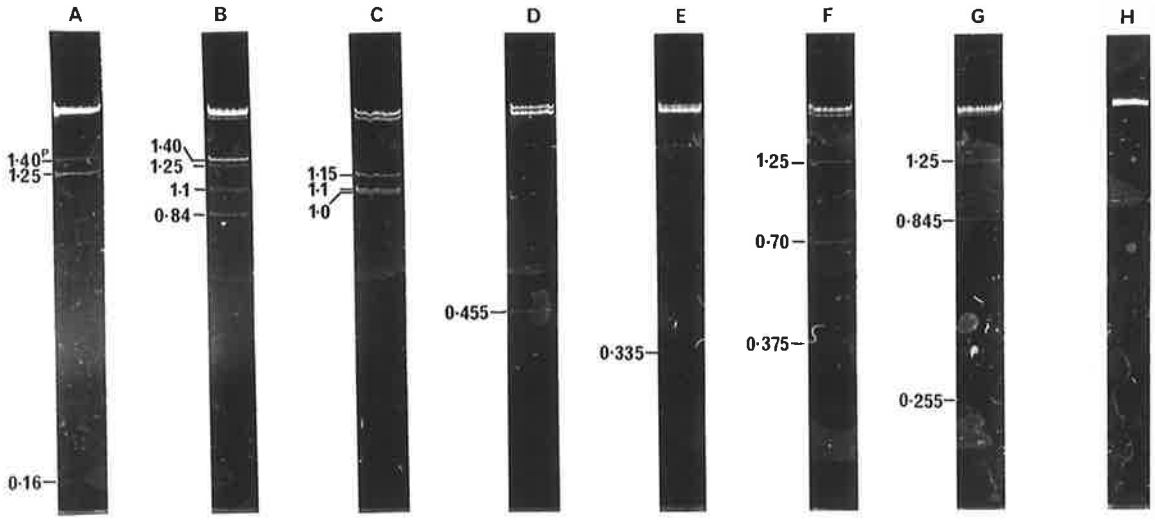

A
B
C
D
E
F
G
H
1·40P
1·25
0·16
1·40
1·25
1·1
0·84
1·15
1·1
1·0
0·455
0·335
1·25
0·70
0·375
1·25
0·845
0·255

Figure 5.8

Restriction Maps of pCH5.OE

The restriction fragments generated on digestion of pCH5.OE DNA with various combinations of restriction enzymes are indicated, together with the derived map of the pCH5.OE insert. The predicted position of the H1 histone gene, as determined from Southern analysis, is shown. Fragment sizes are in kilobases.

P: *Pst*I

E: *Eco*RI

S: *Sal*I

SI: *Sac*I

SII: *Sac*II

H: *Hin*dIII

B: *Bam*HI

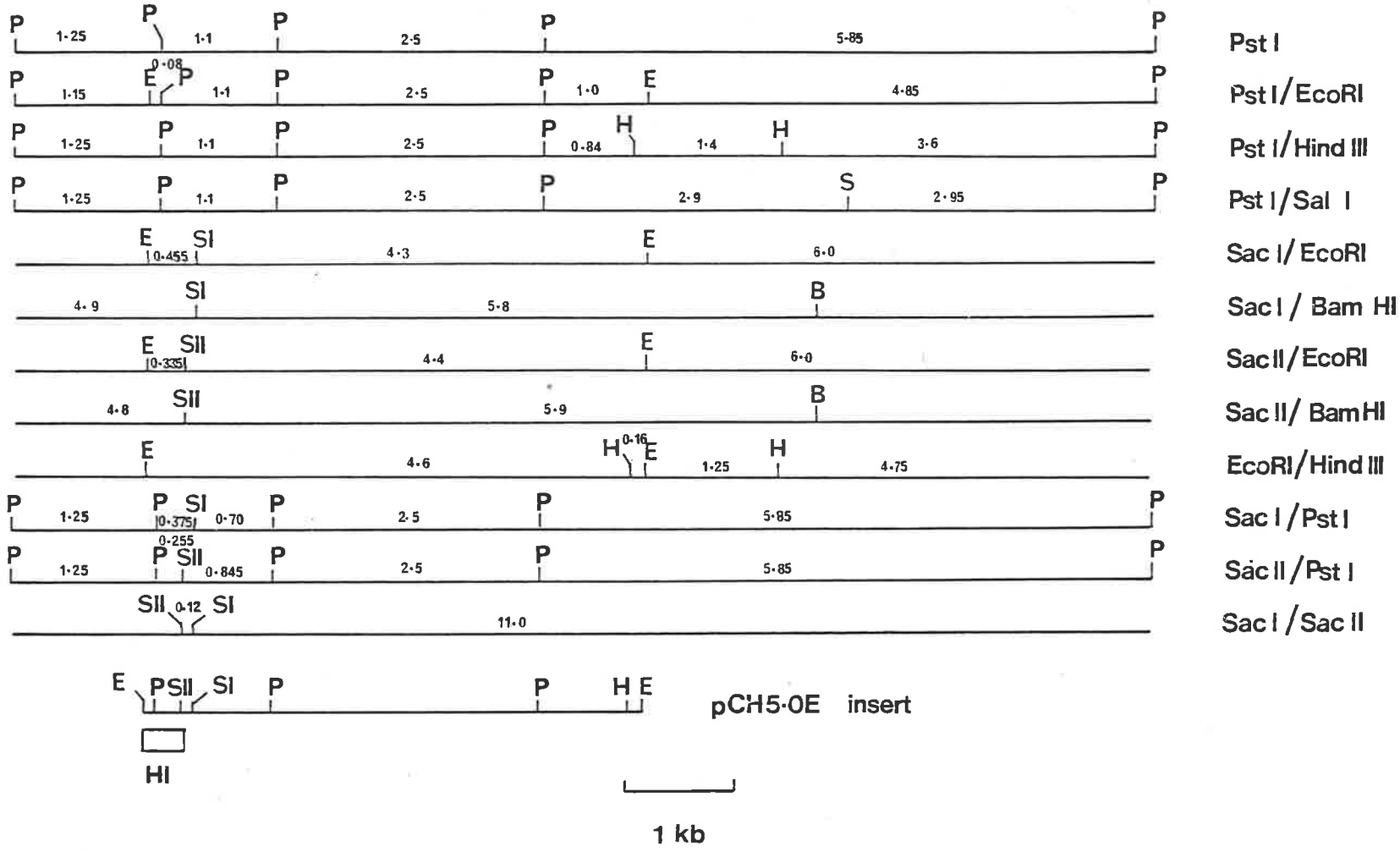

P 1·25 P 1·1 P 2·5 P 5·85 P
Pst I
P 1·15 E 0·08 P 1·1 P 2·5 P 1·0 E 4·85 P
Pst I/EcoRI
P 1·25 P 1·1 P 2·5 P 0·84 H 1·4 H 3·6 P
Pst I/Hind III
P 1·25 P 1·1 P 2·5 P 2·9 S 2·95 P
Pst I/Sal I
E 0·455 SI 4·3 E 6·0
Sac I/EcoRI
4·9 SI 5·8 B
Sac I / Bam HI
E 0·335 SII 4·4 E 6·0
Sac II/EcoRI
4·8 SII 5·9 B
Sac II/ BamHI
E 4·6 H 0·16 E 1·25 H 4·75
EcoRI/Hind III
P 1·25 P 0·375 SI 0·70 P 2·5 P 5·85 P
Sac I/Pst I
P 0·255 SII 0·845 P 2·5 P 5·85 P
Sac II/Pst I
SII 0·12 SI 11·0
Sac I/Sac II
E PSII SI P P H E
pCH5·0E insert
HI
1 kb

Figure 5.9

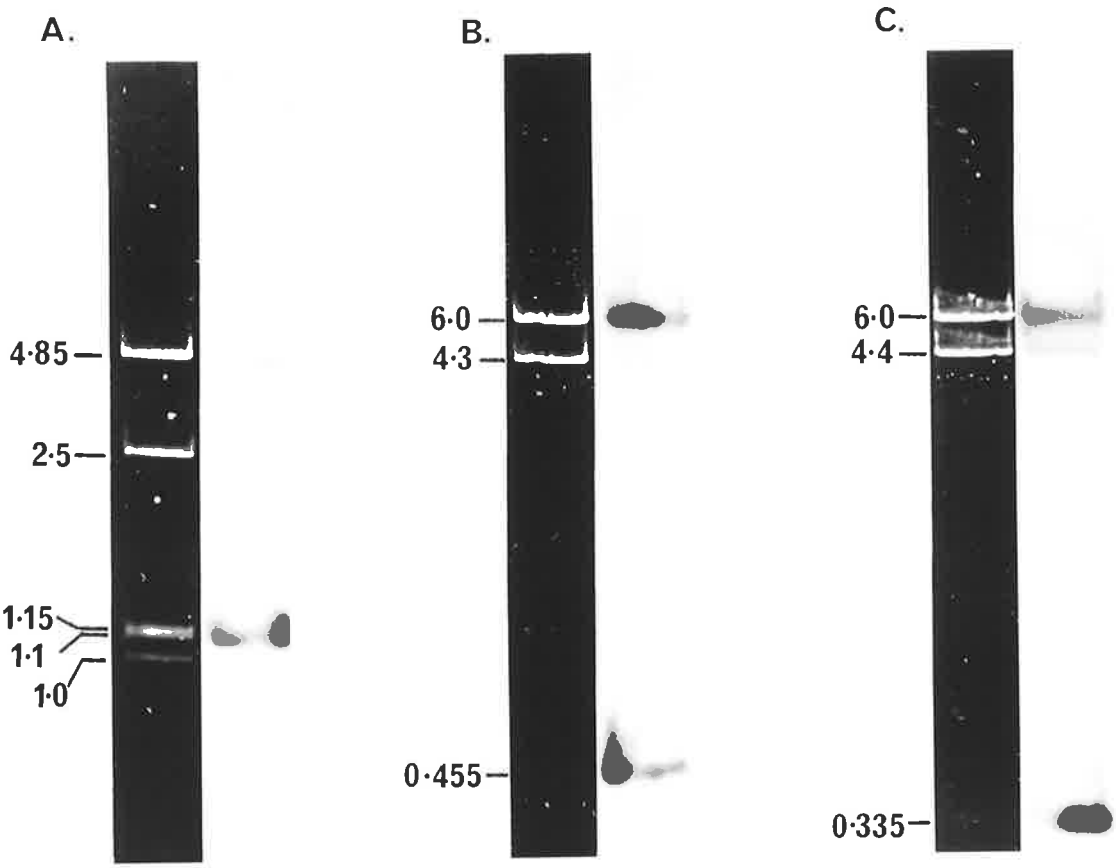
Southern Analysis of pCH5.0E

Restriction enzyme digested pCH5.0E DNA was electrophoresed on a 1.2% agarose gel and transferred to a single nitrocellulose filter (section 2.2.3). The filter was hybridised with ^{32}P-labelled H1 gene-specific probe, washed in 0.5xSSC/0.1% SDS at 65^{o}C and exposed to X-ray film overnight at -80^{o}C with an intensifying screen. Fragment sizes are given in kilobases. The following digests are shown.

Track A: *Pst*I and *Eco*RI

Track B: *Sac*I and *Eco*RI

Track C: *Sac*II and *Eco*RI


A.
4·85
2·5
1·15
1·1
1·0
B.
6·0
4·3
0·455
C.
6·0
4·4
0·335

From initial mapping, *Pst*I, *Sac*I and *Sac*II sites were predicted to occur very close together within the 5.0 kb insert. Digests generating the small restriction fragments were run on a 5% polyacrylamide gel to allow accurate sizing of these fragments. These digests are shown in Figure 5.7.

H1 Coding Region

To determine the location of the H1 gene, digests of pCH5.0E DNA were transferred to nitrocellulose filters (section 2.2.3) and hybridised with H1 probe (provided by L.S. Coles). Results of this hybridisation are shown in Figure 5.9.

The H1 probe hybridises to the small *Sac*I-*Eco*RI fragment and to both *Sac*II-*Eco*RI fragments. This probe also hybridises to the internal 1.1 kb *Pst*I-*Pst*I fragment. The 2.5 kb *Pst*I-*Pst*I fragment does not cross-react with H1 probe, indicating that the hybridising region is located at one end of the insert, beginning in the region between the *Sac*I and *Sac*II restriction sites, and given the predicted size of an H1 gene (H1 contains 121 amino acids), extending to the end of the clone.

c) Restriction Analysis of pCH8.4E

pCH8.4E contains the 8.4 kb *Eco*RI-*Eco*RI fragment from λCH03, which encodes H_2A, H4 and H3 genes, cloned into the *Eco*RI site of pBR325. This clone has been mapped with *Sac*I, *Sac*II, *Hin*dIII, *Sal*I and *Xho*I restriction enzymes. Figures 5.10 and 5.11 present the relevant digests and the derived restriction map is shown in Figure 5.13. Small restriction fragment sizes were obtained by electrophoresis on 5% polyacrylamide gels (Fig. 5.12).

H3 Coding Region

The H3 coding region is completely contained in an internal 2.0 kb *Sac*II-*Sac*II fragment (Fig. 5.14b). This fragment contains the single *Sal*I restriction site and is restricted by *Sal*I to generate 1.7 kb and 0.33 kb *Sal*I-*Sac*II fragments. Only the 0.33 kb *Sal*I-*Sac*II

Figure 5.10

Restriction Analysis of pCH8.4E

1 ug samples of pCH8.4E DNA were digested with the restriction enzymes indicated below and the fragments generated resolved on a 1% agarose gel (section 2.2.2a). DNA was detected with UV light after ethidium bromide staining. Fragment sizes are indicated in kilobases. Fragments marked with a superscript P resulted from incomplete digestion. The following restriction enzyme digests of pCH8.4E DNA are shown;

Track A: *Sac*I

Track B: *Sac*I and *Eco*RI

Track C: *Sac*I and *Sal*I

Track D: *Sac*I and *Xho*I

Track E: *Sac*I and *Sac*II

Track F: *Sac*II

Track G: *Sac*II and *Eco*RI

Track H: *Sac*II and *Sal*I

Track I: *Sac*II and *Xho*I

Track J: *Sac*II and *Hin*dIII

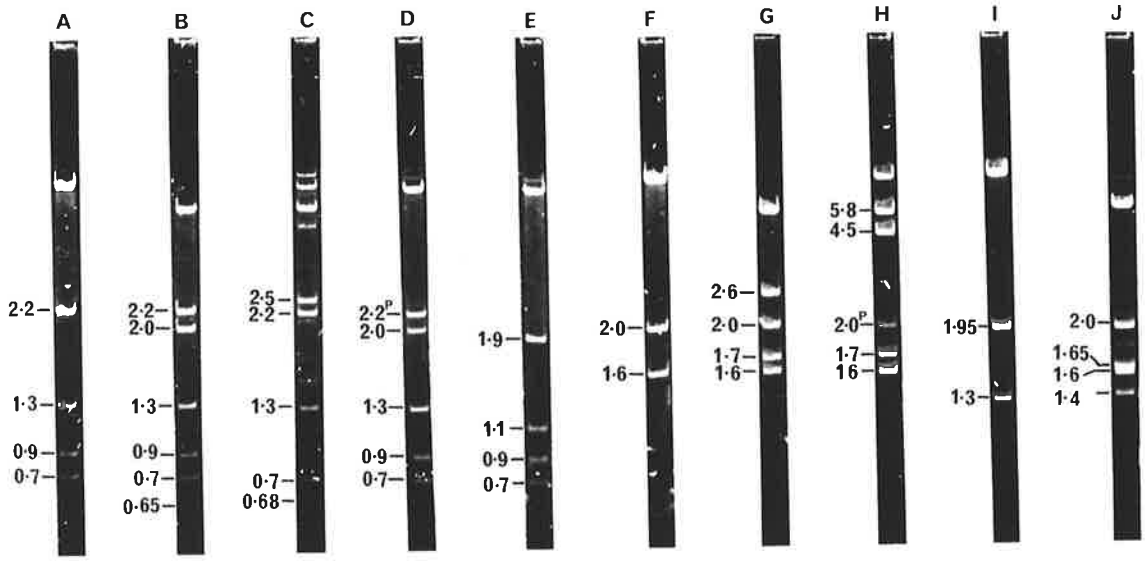

A
2·2
1·3
0·9
0·7
B
2·2
2·0
1·3
0·9
0·7
0·65
C
2·5
2·2
1·3
0·7
0·68
D
2·2P
2·0
1·3
0·9
0·7
E
1·9
1·1
0·9
0·7
F
2·0
1·6
G
2·6
2·0
1·7
1·6
H
5·8
4·5
2·0P
1·7
16
I
1·95
1·3
J
2·0
1·65
1·6
1·4

Figure 5.11

Restriction Analysis of pCH8.4E

DNA of pCH8.4E was digested with the range of restriction enzymes shown below and analysed by electrophoresis through a 1.2% agarose gel (section 2.2.2a; Fig. 5.10). Fragment sizes are given in kilobases. This figure shows pCH8.4E DNA digested with the following enzymes;

Track A: *Xho*I

Track B: *Xho*I and *Eco*RI

Track C: *Xho*I and *Sal*I

Track D: *Xho*I and *Hin*dIII

Track E: *Sac*I

Track F: *Sac*I and *Xho*I

Track G: *Sac*II

Track H: *Sac*II and *Xho*I

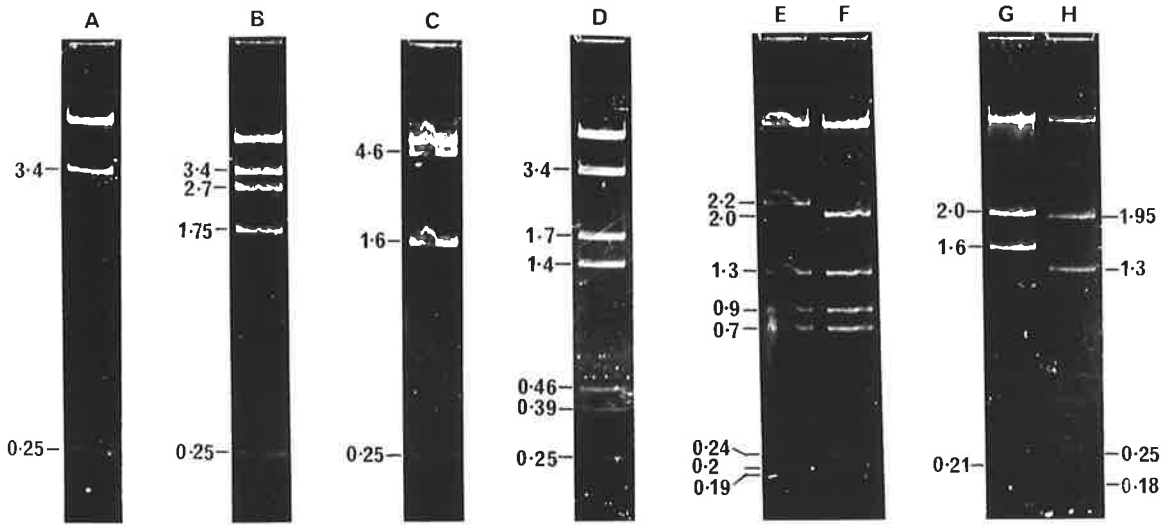

A
B
C
D
E
F
G
H
3·4
0·25
3·4
2·7
1·75
0·25
4·6
1·6
0·25
3·4
1·7
1·4
0·46
0·39
0·25
2·2
2·0
1·3
0·9
0·7
0·24
0·2
0·19
2·0
1·6
0·21
1·95
1·3
0·25
0·18

Figure 5.12

Restriction Analysis of pCH8.4E

1 ug samples of pCH8.4E DNA were digested with appropriate restriction enzymes (see below) and the resultant fragments fractionated on 5% (Tracks G-I) and 6% (tracks A-F) polyacrylamide gels (section 2.2.2b). DNA was viewed under UV light after ethidium bromide staining. Fragment sizes are indicated, in kilobases, alongside each track.

The following restriction enzyme digests are shown;

Track A: *Sac*II

Track B: *Sac*II and *Sal*I

Track C: *Sac*I

Track D: *Sac*I and *Sal*I

Track E: *Sac*I and *Eco*RI

Track F: *Sac*I and *Sac*II

Track G: *Sac*II and *Xho*I

Track H: *Hin*dIII

Track I: *Eco*RI and *Hin*dIII

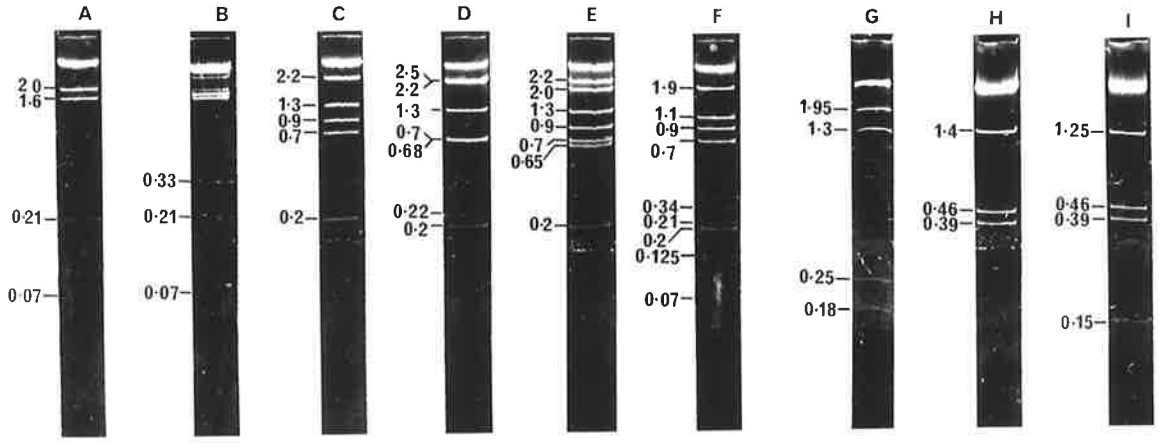

A
2·0
1·6
0·21
0·07
B
0·33
0·21
0·07
C
2·2
1·3
0·9
0·7
0·2
D
2·5
2·2
1·3
0·7
0·68
0·22
0·2
E
2·2
2·0
1·3
0·9
0·7
0·65
0·2
F
1·9
1·1
0·9
0·7
0·34
0·21
0·2
0·125
0·07
G
1·95
1·3
0·25
0·18
H
1·4
0·46
0·39
I
1·25
0·46
0·39
0·15

Figure 5.13

Restriction Maps of pCH8.4E

The sizes (in kilobases) of the fragments generated after restriction enzyme cleavage are shown with the derived map of the pCH8.4E insert. The gene positions, as determined from Southern analysis, are also shown. Restriction sites are indicated as follows.

E: *Eco*RI

H: *Hin*dIII

S: *Sal*I

SI: *Sac*I

SII: *Sac*II

X: *Xho*I

Note: The small 0.07 kb fragment indicated in Figure 5.12 has not been positioned.

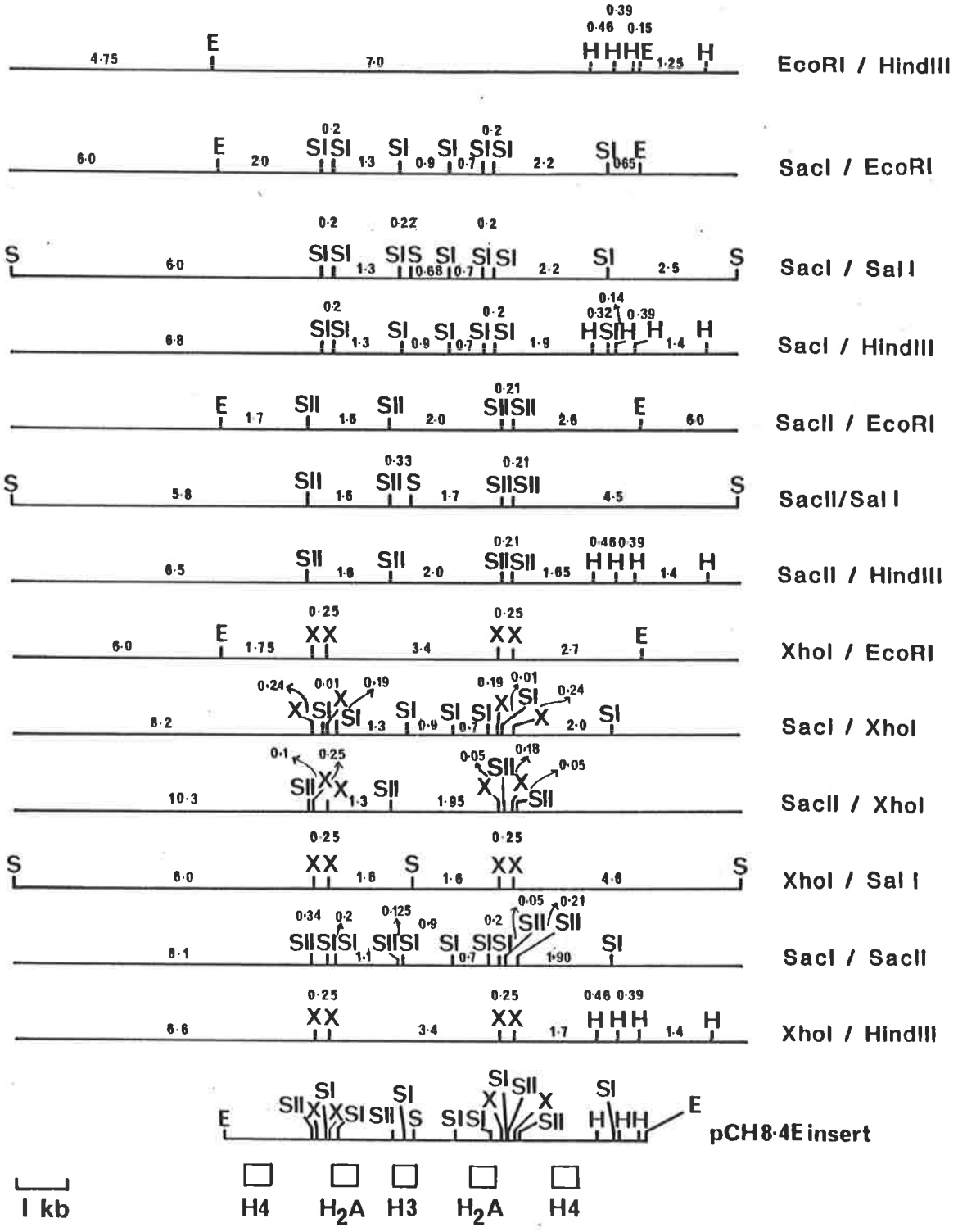

EcoRI / HindIII
SacI / EcoRI
SacI / Sal I
SacI / HindIII
SacII / EcoRI
SacII/Sal I
SacII / HindIII
XhoI / EcoRI
SacI / XhoI
SacII / XhoI
XhoI / Sal I
SacI / SacII
XhoI / HindIII
pCH8·4E insert
1 kb
H4
H2A
H3
H2A
H4

Figure 5.14

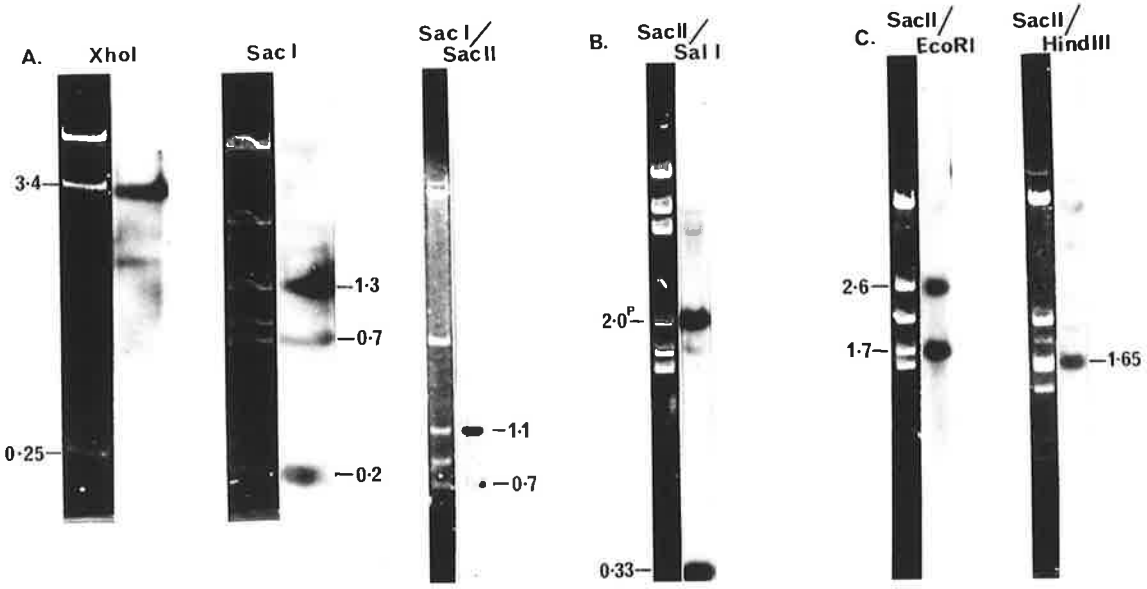
Southern Analysis of pCH8.4E

Plasmid DNA (pCH8.4E) which had been restriction enzyme digested and fractionated on agarose gels was transferred to nitrocellulose filters using the bidirectional transfer procedure detailed in section 2.2.3. Filters were hybridised to radiolabelled histone gene specific probes, washed and exposed as described previously (Fig. 5.9). Only the digests discussed in the text are shown. Fragment sizes are indicated in kilobases. Fragments marked with a superscript P are the result of incomplete restriction enzyme reactions. The digests shown were probed as follows;

A: H_2A gene specific probe

B: H3 gene specific probe

C: H4 gene specific probe


A.
XhoI
3·4
0·25
Sac I
1·3
0·7
0·2
Sac I / Sac II
1·1
0·7
B.
SacII / Sal I
2·0P
0·33
C.
SacII / EcoRI
2·6
1·7
SacII / HindIII
1·65

fragment hybridises to the H3 gene specific probe (Fig. 5). This H3 probe is specific for the coding region of H3 genes 3' to the conserved *Sal*I restriction site. It follows then that the H3 gene spans the *Sal*I site but that the coding region is almost completely contained in the 0.33 kb *Sal*I-*Sac*II fragment.

H_2A Coding Regions

Two internal *Sac*I-*Sac*I fragments (1.3 kb and 0.7 kb) are positive with H_2A probe (Fig. 5.14a). These fragments are not adjacent to each other and each lies next to a 0.2 kb *Sac*I-*Sac*I fragment. This 0.2 kb doublet is also positive with H_2A probe, implying that an H_2A gene is located within the 1.3 kb and 0.2 kb *Sac*I-*Sac*I fragments at the left hand end of the insert with another H_2A gene spanning the 0.7 kb and 0.2 kb *Sac*I-*Sac*I fragments at the right-hand end, as drawn in Figure 5.13.

H4 Coding Regions

H4 coding regions were detected in both *Sac*II-*Eco*RI fragments (1.7 kb and 2.6 kb; Fig. 5.14). Hybridisation of H4 probe to a *Sac*II/*Hin*dIII digest indicates that the right-hand H4 gene is completely contained in the 1.65 kb *Sac*II/*Hin*dIII fragment (Fig. 5.14c).

The location of the restriction sites and genes within this fragment has recently been confirmed by sequencing (Wang *et al*, 1985).

5.2.6 Restriction Analysis of λCH05

λCH05 DNA was digested in single and double digests, with *Eco*RI, *Xho*I, *Bam*HI, *Sal*I and *Hin*dIII and the resultant fragments fractionated on agarose gels, and sized by comparison with marker DNA (Fig. 5.15). The size and relative position of each fragment and a complete map of the insert is shown in Figure 5.16.

*Eco*RI/*Xho*I digestion indicated that this clone contains the 0.7 kb *Eco*RI-*Eco*RI fragment and the 0.7 kb *Xho*I-*Xho*I fragment found

Figure 5.15

Restriction Analysis of λCHO5

1 ug samples of λCHO5 DNA were digested with a range of restriction endonucleases, either singly or in combination, and fractionated on a 1% agarose gel (section 2.2.2a). Bands were observed with UV light following ethidium bromide staining. *Hin*dIII and *Sal*I did not cleave the λCHO5 insert. Fragment sizes are indicated in kilobases. The figure shows the following restriction digests;

Track A: *Eco*RI

Track B: *Bam*HI

Track C: *Eco*RI and *Bam*HI

Track D: *Xho*I

Track E: *Eco*RI and *Xho*I

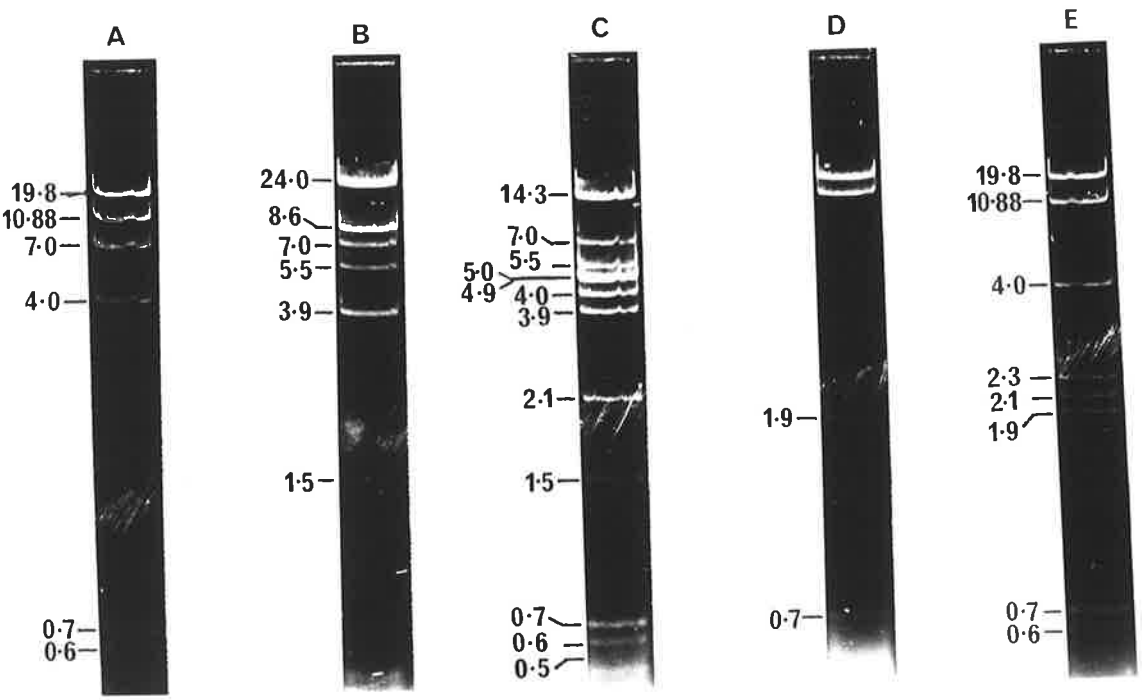
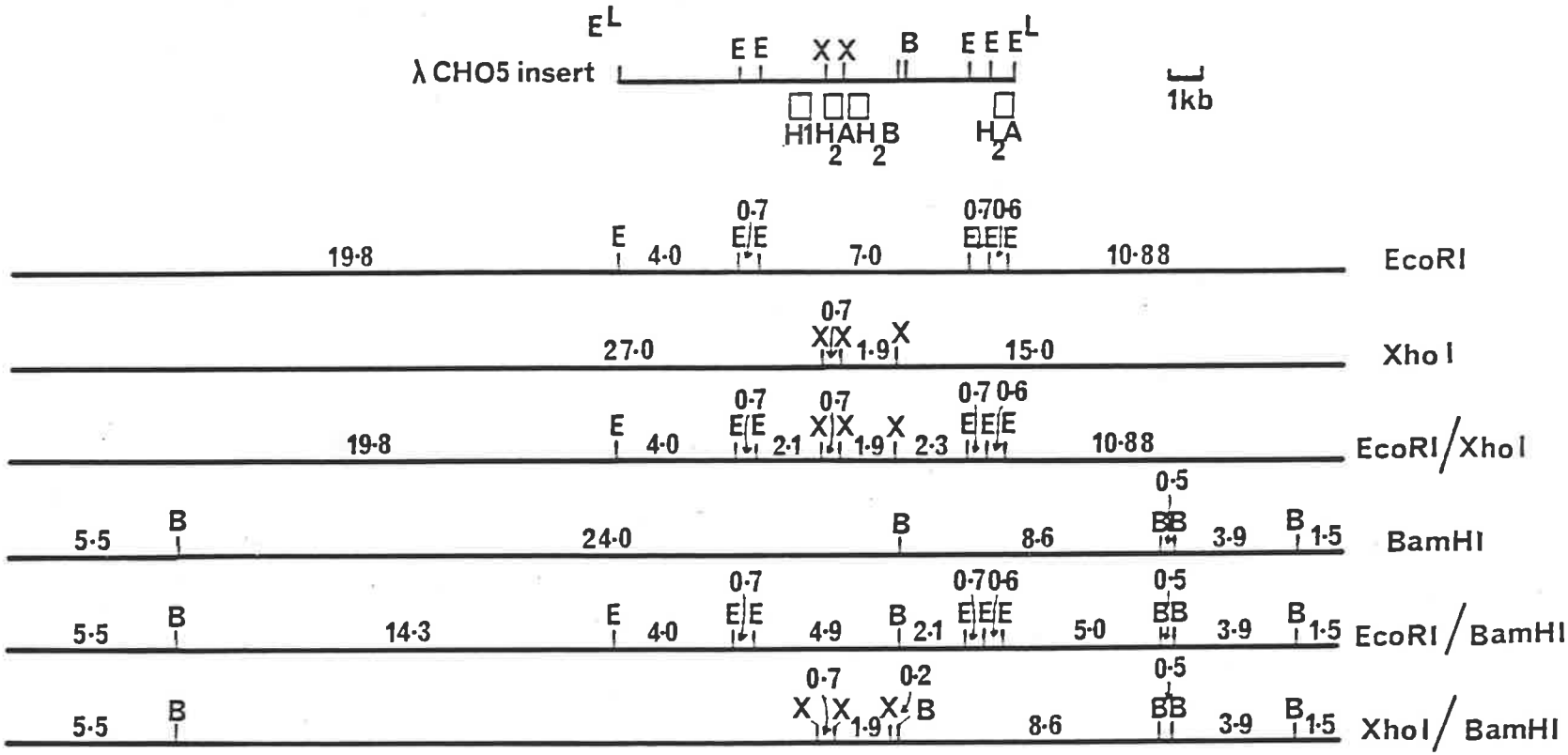

A
19·8
10·88
7·0
4·0
0·7
0·6
B
24·0
8·6
7·0
5·5
3·9
1·5
C
14·3
7·0
5·5
5·0
4·9
4·0
3·9
2·1
1·5
0·7
0·6
0·5
D
1·9
0·7
E
19·8
10·88
4·0
2·3
2·1
1·9
0·7
0·6

Figure 5.16

Restriction Maps of λCHO5

The fragments generated on restriction enzyme cleavage of λCHO5 DNA with various combinations of *Eco*RI(E), *Xho*I(X) and *Bam*HI(B) are shown. An overall derived restriction map of the λCHO5 insert showing the predicted gene positions (as determined from Southern analysis; Fig. 5.17) is also shown. *Sal*I and *Hin*dIII do not cleave the λCHO5 insert.


λ CHO5 insert
E L E E X X B E E E L
1kb
H1 H2 A H2 B
H2 A
EcoRI
Xho I
EcoRI / Xho I
BamHI
EcoRI / BamHI
Xho I / BamHI

in λCHO1 and thus confirmed that this clone overlaps with λCHO1. The 7.0 kb *Eco*RI-*Eco*RI fragment contained a single *Bam*HI site very close to a *Xho*I restriction site. The exact location of these sites, and the size of the small *Xho*I-*Bam*HI fragment was determined by detailed restriction mapping of the subclone pCH7.0E (see section 5.2.9).

5.2.7 The Coding Potential of λCHO5

Digested DNA shown in Figure 5.15 was transferred to nitrocellulose using the bidirectional transfer technique described (section 2.2.3) and the two filters obtained were probed separately with either H_2A or H_2B gene specific probe. From results of experiments previously described (section 4.2.3) it was known that λCHO5 contained only H1, H_2A and H_2B genes.

H_2A probe hybridised to the 0.7 kb *Xho*I-*Xho*I fragment and the 0.6 kb *Eco*RI-*Eco*RI fragment (Fig. 5.17). The *Xho*I-*Xho*I fragment is known to contain a single complete H_2A gene (D'Andrea *et al*, 1981) and is the fragment in λCHO1 from which the H_2A specific probe is derived. Hybridisation to the 0.6 kb *Eco*RI-*Eco*RI fragment was much weaker and suggested that this fragment may contain only part of an H_2A gene, or alternatively a variant H_2A gene sequence. This fragment was located next to a 0.7 kb *Eco*RI-*Eco*RI fragment at the right hand end of λCHO5 but the order of these two fragments could not be determined from this data. However, it was later positioned at the end of the insert with an H_2A gene only partly contained in this clone. This positioning was resolved after characterization of the clone λCHO7 which overlaps and extends this end of λCHO5 (section 5.2.11).

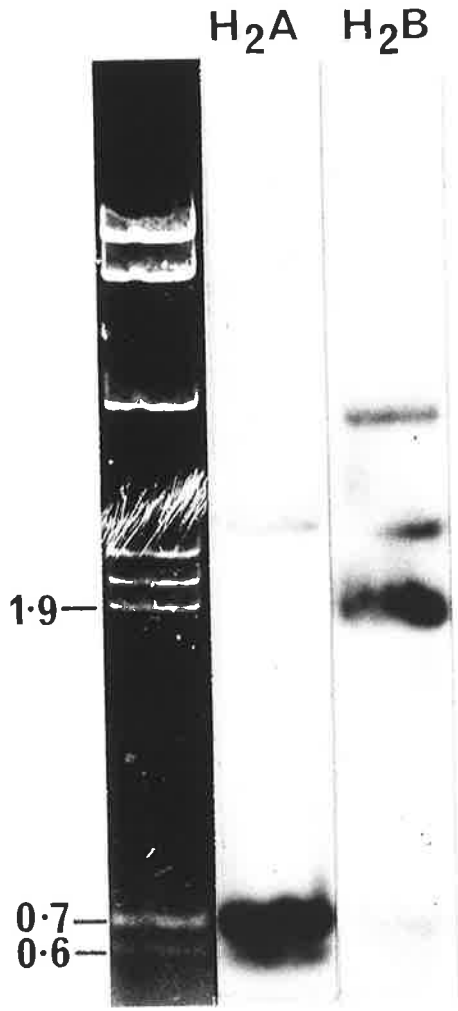
H_2B probe hybridised only to the 1.9 kb *Xho*I-*Xho*I fragment which is completely contained within the 7.0 kb *Eco*RI-*Eco*RI fragment. An H_2B gene is known to be present at one end of this *Xho*I-*Xho*I fragment (i.e. next to the H_2A gene, Harvey *et al*, 1981; Harvey *et al*, 1982).

A single complete H1 gene is known to be present in the 2.1 kb

Figure 5.17

Southern Analysis of λCHO5

λCHO5 DNA which had been digested with various restriction enzymes was fractionated on a 1% agarose gel and transferred, bidirectionally (section 2.2.3), to nitrocellulose filters. Filter-bound DNA was hybridised to either H_2A or H_2B gene specific probe, washed in 0.5xSSC/0.1% SDS at 65°C and exposed to X-ray film at -80°C with an intensifying screen. Only *Eco*RI/*Xho*I digested DNA is shown in the figure. Fragment sizes are given in kilobases.


H2A
H2B
1·9
0·7
0·6

*Eco*RI-*Xho*I fragment and Southern analysis indicated this was the only H1 gene present in λCHO5 (L.S. Coles, personal communication).

5.2.8 Isolation of a λCHO5 Subclone

The 7.0 kb *Eco*RI-*Eco*RI fragment from λCHO5 extends a 3.3 kb *Eco*RI-*Eco*RI fragment in λCHO1. To determine whether this *Eco*RI-*Eco*RI fragment from λCHO5 contains any histone coding domains in addition to those present in the 3.3 kb *Eco*RI-*Eco*RI fragment from λCHO1, it was subcloned into pBR325 and analysed further. DNA was prepared from a single recombinant colony and used for the restriction enzyme mapping described below. This clone is termed pCH7.0E.

5.2.9 Restriction Analysis of pCH7.0E

pCH7.0E DNA was digested with *Sac*I, *Sac*II, *Sma*I, *Xho*I and *Hin*dIII, in single and double digests, and the resultant fragments fractionated on agarose and polyacrylamide gels alongside marker DNA (Figs. 5.18, 5.19 and 5.20).

Maps indicating the sizes and relative positions of each fragment are shown in Figure 5.21. Digested DNA was transferred to nitrocellulose filters and gene locations determined by hybridisation.

H1 Coding Region

H1 probe hybridised only to the 0.9 kb *Sac*I-*Sac*II fragment (Fig. 5.22), indicating that pCH7.0E contains H1 coding sequences within this fragment. This region is also contained in pCH3.3E and has been completely sequenced and shown to contain a single H1 coding region (Coles and Wells, 1985).

H_2A Coding Region

The only H_2A-positive fragment in pCH7.0E was the 0.7 kb *Xho*I-*Xho*I fragment (Fig. 5.22) also present in λCHO1, and containing a complete H_2A gene (D'Andrea *et al*, 1981). This *Xho*I-*Xho*I fragment contains three *Sac*I sites. The two *Sac*I-*Sac*I fragments (340 bp and 190 bp) and the 140 bp *Sac*I-*Xho*I fragment generated after *Sac*I/*Xho*I digestion

Figure 5.18

Restriction Analysis of pCH7.OE

1 ug samples of pCH7.OE DNA were digested with various restriction enzymes, singly or in combination, and fractionated on a 1% agarose gel (section 2.2.2a) to allow size determination. DNA was detected with UV light after ethidium bromide staining. Fragment sizes are indicated in kilobases. The figure shows pCH7.OE DNA digested with the following enzymes;

Track A: *Eco*RI

Track B: *Sac*I

Track C: *Sac*I and *Eco*RI

Track D: *Sac*II

Track E: *Sac*II and *Eco*RI

Track F: *Sma*I

Track G: *Sma*I and *Eco*RI

Track H: *Xho*I

Track I: *Xho*I and *Eco*RI

Track J: *Bam*HI

Track K: *Bam*HI and *Eco*RI

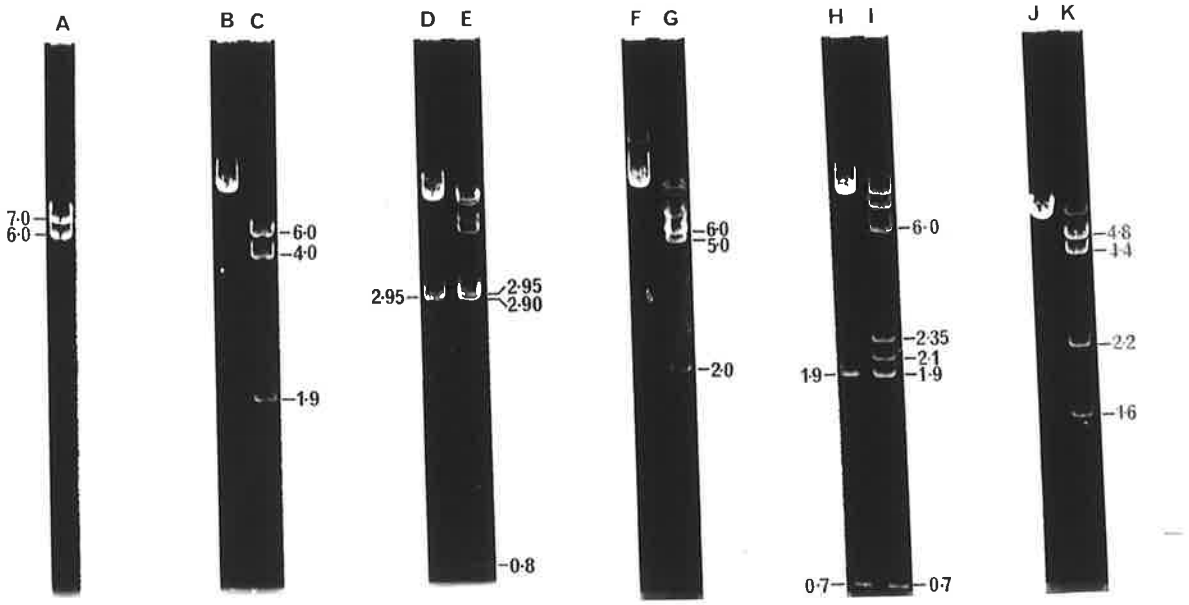

A
7·0
6·0
B C
6·0
4·0
1·9
D E
2·95
2·95
2·90
0·8
F G
6·0
5·0
2·0
H I
6·0
2·35
2·1
1·9
1·9
0·7
0·7
J K
4·8
4·4
2·2
1·6

Figure 5.19

Restriction Analysis of pCH7.0E

Analysis of pCH7.0E DNA (1 ug) digested with various restriction enzymes. Details of all procedures used are given in Chapter 2 and in Figure 5.18. Fragment sizes are indicated in kilobases. Fragments marked with a superscript P are due to incomplete digestion. The products of the following restriction enzyme reactions are shown;

Track A: *Sac*I and *Bam*HI

Track B: *Sac*I and *Xho*I

Track C: *Sac*I and *Sma*I

Track D: *Sac*I and *Sac*II

Track E: *Sac*II and *Bam*HI

Track F: *Sac*II and *Xho*I

Track G: *Sac*II and *Sma*I

Track H: *Sac*II

Track I: *Pst*I

Track J: *Pst*I and *Bam*HI

Track K: *Pst*I and *Eco*RI

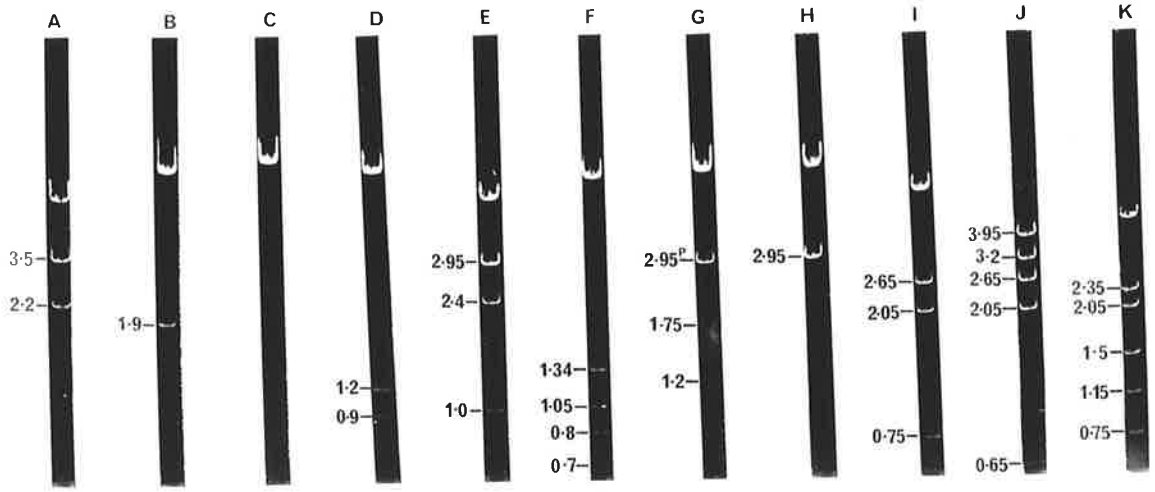

A
B
C
D
E
F
G
H
I
J
K
3·5
2·2
1·9
1·2
0·9
2·95
2·4
1·0
1·34
1·05
0·8
0·7
2·95
1·75
1·2
2·95
2·65
2·05
0·75
3·95
3·2
2·65
2·05
0·65
2·35
2·05
1·5
1·15
0·75

Figure 5.20

Restriction Analysis of pCH7.0E

1 ug samples of pCH7.0E DNA were digested with appropriate restriction enzymes (see below) and electrophoresed on a 6% polyacrylamide gel (section 2.2.2b) to allow sizing of small fragments. Staining with ethidium bromide allowed visualisation of DNA under UV light. Sizes were determined by comparison with molecular weight standards (*Hin*fI digested pBR322) and are given in kilobases. The following restriction enzyme digests are shown,

Track A: *Sac*I

Track B: *Sac*I and *Xho*I

Track C: *Sac*I and *Sac*II

Track D: *Sac*II and *Xho*I

Track E: *Sac*II and *Bam*HI

Track F: *Pst*I

Track G: *Pst*I and *Bam*HI

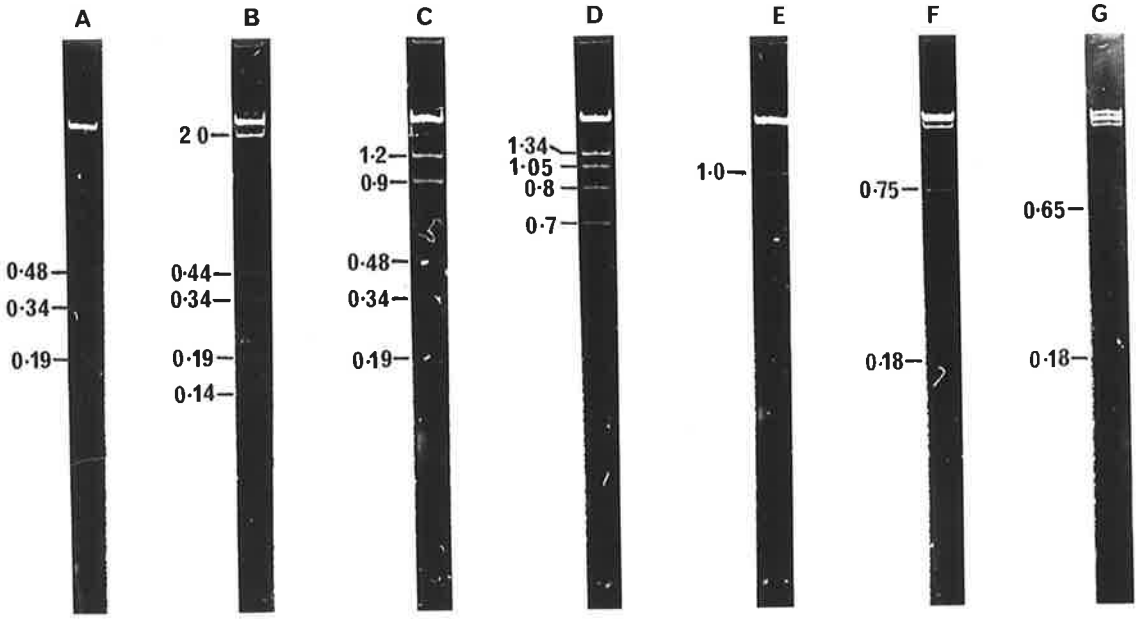

A
0·48
0·34
0·19
B
2 0
0·44
0·34
0·19
0·14
C
1·2
0·9
0·48
0·34
0·19
D
1·34
1·05
0·8
0·7
E
1·0
F
0·75
0·18
G
0·65
0·18

Figure 5.21

Restriction Maps of pCH7.OE

The sizes, in kilobases, of the fragments obtained after restriction enzyme cleavage of pCH7.OE are indicated. The overall derived restriction map, showing the predicted gene positions is also shown. Restriction enzyme sites are indicated as follows;

E: *Eco*RI

B: *Bam*HI

SI: *Sac*I

SII: *Sac*II

Sm: *Sma*I

X: *Xho*I

P: *Pst*I

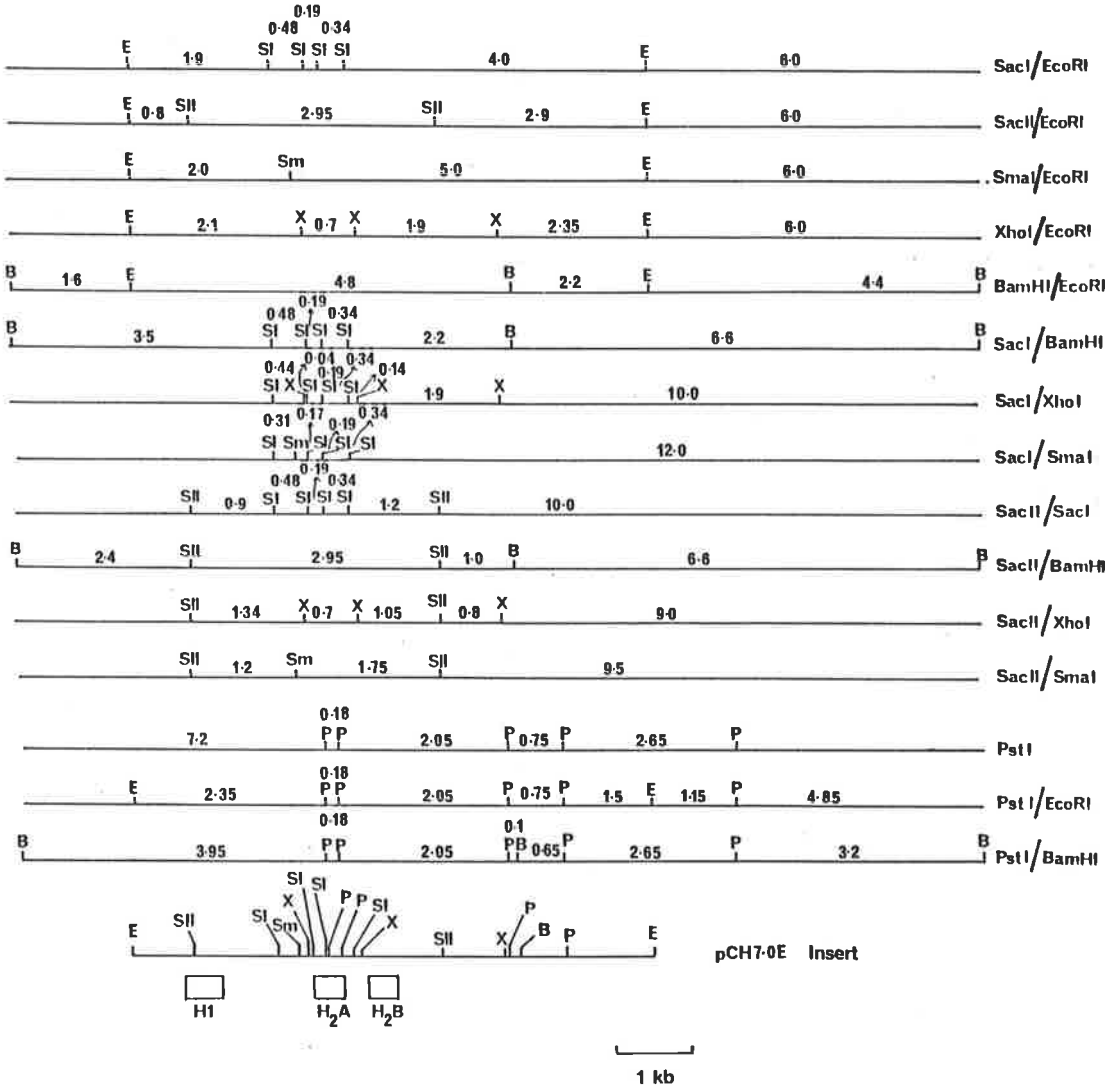
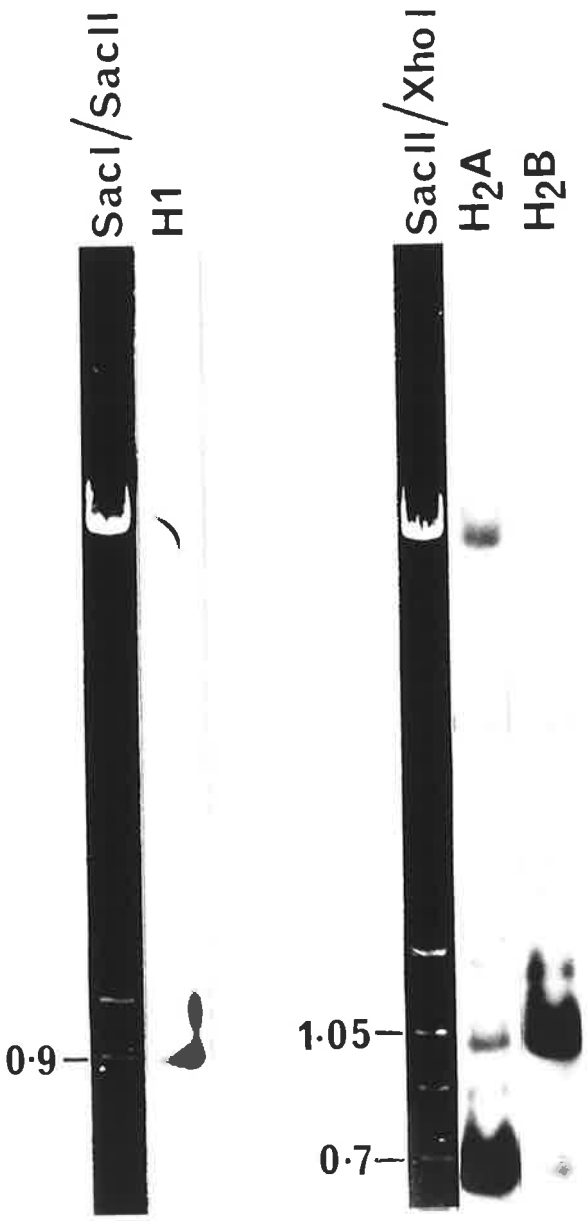

SacI/EcoRI
SacII/EcoRI
SmaI/EcoRI
XhoI/EcoRI
BamHI/EcoRI
SacI/BamHI
SacI/XhoI
SacI/SmaI
SacII/SacI
SacII/BamHI
SacII/XhoI
SacII/SmaI
PstI
PstI/EcoRI
PstI/BamHI
pCH7·0E Insert
H1
H2A
H2B
1 kb

Figure 5.22

Southern Analysis of pCH7.0E

Restriction enzyme cleaved pCH7.0E DNA was fractionated on a 1% agarose gel and transferred to nitrocellulose filters using the bidirectional transfer technique (section 2.2.3). Filter-bound DNA was analysed for gene content with H_2A and H_2B gene-specific probes. Filters were washed in 0.5xSSC/0.1% SDS at 65°C and autoradiographed overnight at -80°C with an intensifying screen. Digests discussed in the text are shown. Sizes are indicated in kilobases.


SacI/SacII
H1
0·9
SacII/XhoI
H2A
H2B
1·05
0·7

were seen on a 5% polyacrylamide gel (Fig. 5.18). The locations of the *Pst*I and *Sac*I sites within the H_2A containing *Xho*I-*Xho*I fragment were known from DNA sequencing (D'Andrea *et al*, 1981).

H_2B Coding Region

The only fragment positive with the H_2B specific probe was a 1.0 kb *Sac*II-*Xho*I fragment. A single H_2B coding region has subsequently been identified by sequencing (Harvey *et al*, 1982).

Together these results indicated that the only histone genes present in pCH7.0E were those already located and analysed in the subclone pCH3.3E from λCHO1 (Fig. 3.1).

5.2.10 Search for Clones Extending λCHO3

Fragments which did not contain histone gene sequences were isolated from both ends of λCHO3. The 2.5 kb *Pst*I-*Pst*I fragment (Fig. 5.8) and the 0.65 kb *Sac*I-*Eco*RI fragment from pCH8.4E (Fig. 5.13) were purified by LGT gel electrophoresis, eluted, labelled by nick translation and used to probe the histone clone bank described in Chapter 4, using the "dot-Benton" procedure previously described (section 2.2.5a).

One plaque, 4.1, was positive with the 2.5 kb *Pst*I-*Pst*I probe but did not hybridise to the 0.65 kb *Sac*I-*Eco*RI probe from the other end of λCHO3. This clone was characterised further (Linda Tabe, this laboratory) but was found to have resulted from rearrangements during the construction of the chicken library (Results not shown).

5.2.11 Search for Clones Extending λCHO5

The 2.1 kb *Bam*HI-*Eco*RI fragment from λCHO5 (see Fig. 5.16) was purified by electrophoresis through LGT agarose, eluted and nick-translated, and used to probe the "bank" of histone gene containing recombinants.

Two plaques, the 5.5 and 7.2 isolates, were positive with this probe (data not shown). Digestion of 7.2 DNA with *Eco*RI gave a different digestion pattern to λCHO5 and suggested this clone was a possible overlap.

Previous hybridisation results (section 4.2.3) indicated that clone 7.2 contained H_2A and H_2B genes only. This clone was subsequently referred to as λCHO7 and is discussed in more detail below.

5.2.12 Restriction Analysis of λCHO7

λCHO7 DNA was digested with *Eco*RI, *Xho*I, *Bam*HI, *Sal*I, and *Hin*dIII and resultant fragments sized on agarose and polyacrylamide gels (Figs. 5.23 and 5.24). Maps showing the size and relative position of each fragment are shown in Figure 5.25. A complete map of the insert is also presented.

Note that the 1.9 kb and the 0.7 kb *Xho*I-*Xho*I fragments present in λCHO5 are also present in this clone (Fig. 5.23). However, the 7.0 kb *Eco*RI-*Eco*RI fragment from λCHO5 which contains both of these fragments is not present (Fig. 5.23), implying that this clone begins within the 2.1 kb *Eco*RI-*Xho*I fragment in λCHO5.

5.2.13 Coding Potential of λCHO7

To determine the coding potential of this clone, restriction enzyme digestions were transferred to nitrocellulose and hybridised with H_2A and H_2B probes. Both probes hybridised to the 5.5 kb and 3.5 kb *Eco*RI-*Eco*RI fragments, indicating that there are at least two copies of each of these genes.

Furthermore, H_2A probe hybridised to the 0.7 kb *Xho*I-*Xho*I fragment (also contained in λCHO1 and λCHO5) and the 2.2 kb *Xho*I-*Eco*RI fragment unique to λCHO7.

H_2B probe was specific for the 1.9 kb *Xho*I-*Xho*I fragment (also present in λCHO5 and partly in λCHO1) and also the 2.2 kb *Xho*I-*Eco*RI fragment which was positive with H_2A probe.

These results suggested that previously uncharacterised H_2A and H_2B genes were located in the 2.2 kb *Eco*RI-*Xho*I fragment of λCHO7. In addition to these genes, λCHO7 contains H_2A and H_2B genes present in λCHO5 (also contained in λCHO1).

Figure 5.23

Restriction Analysis of λCHO7

1 ug samples of λCHO7 DNA were digested with the restriction enzymes listed below and fractionated on a 1% agarose gel (section 2.2.2a). Bands were observed with UV light after ethidium bromide staining. The *Eco*RI and *Hin*dIII reaction (track C) was incomplete leading to a number of partial digestion products (10.88 kb, 5.5 kb 4.9 kb, 3.5 kb and 3.0 kb). In track D the 7.0 kb fragment is the annealed product of the 5.5 kb and 1.5 kb terminal *Bam*HI vector fragments (see Fig. 5.25). Fragment sizes are indicated in kilobases. The following restriction digests are shown;

Track A: *Eco*RI

Track B: *Hin*dIII

Track C: *Eco*RI and *Hin*dIII

Track D: *Bam*HI

Track E: *Hin*dIII and *Bam*HI

Track F: *Bam*HI

Track G: *Eco*RI and *Bam*HI

Track H: *Xho*I

Track I: *Eco*RI and *Xho*I

Track J: *Bam*HI and *Xho*I

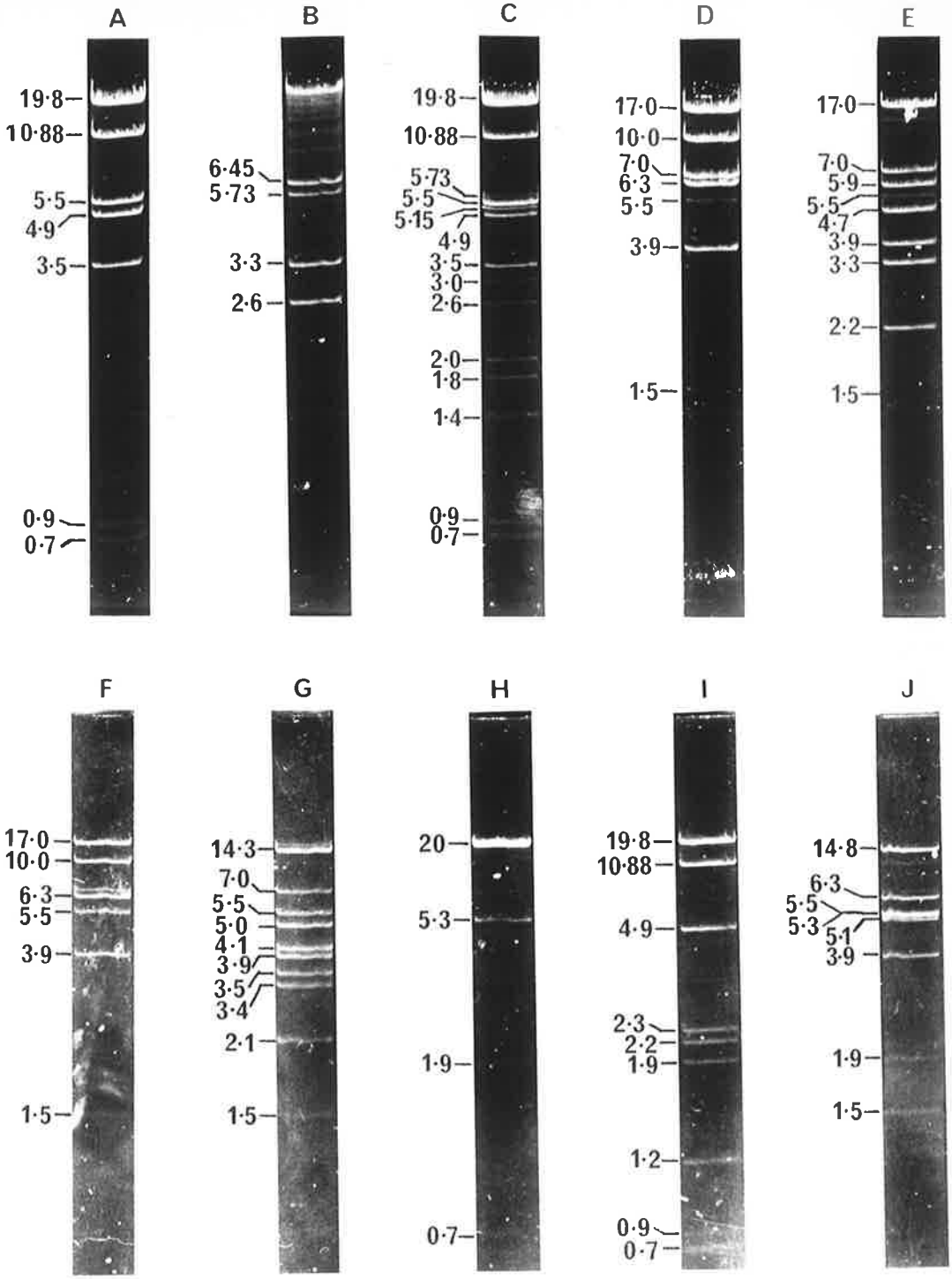

A
19·8
10·88
5·5
4·9
3·5
0·9
0·7
B
6·45
5·73
3·3
2·6
C
19·8
10·88
5·73
5·5
5·15
4·9
3·5
3·0
2·6
2·0
1·8
1·4
0·9
0·7
D
17·0
10·0
7·0
6·3
5·5
3·9
1·5
E
17·0
7·0
5·9
5·5
4·7
3·9
3·3
2·2
1·5
F
17·0
10·0
6·3
5·5
3·9
1·5
G
14·3
7·0
5·5
5·0
4·1
3·9
3·5
3·4
2·1
1·5
H
20
5·3
1·9
0·7
I
19·8
10·88
4·9
2·3
2·2
1·9
1·2
0·9
0·7
J
14·8
6·3
5·5
5·3
5·1
3·9
1·9
1·5

Figure 5.24

Restriction Analysis of λCHO7

Small restriction fragments generated on digestion of λCHO7 DNA (1 ug) were sized by electrophoresis on a 5% polyacrylamide gel (section 2.2.2b). DNA was visualised under UV light after ethidium bromide staining. *Hin*fI cut pBR322 was co-electrophoresed as marker DNA. The following restriction enzyme digests of λCHO7 are shown:

Track A: *Eco*RI

Track B: *Eco*RI and *Hin*dIII

Track C: *Bam*HI

Track D: *Eco*RI and *Bam*HI

Track E: *Bam*HI and *Hin*dIII

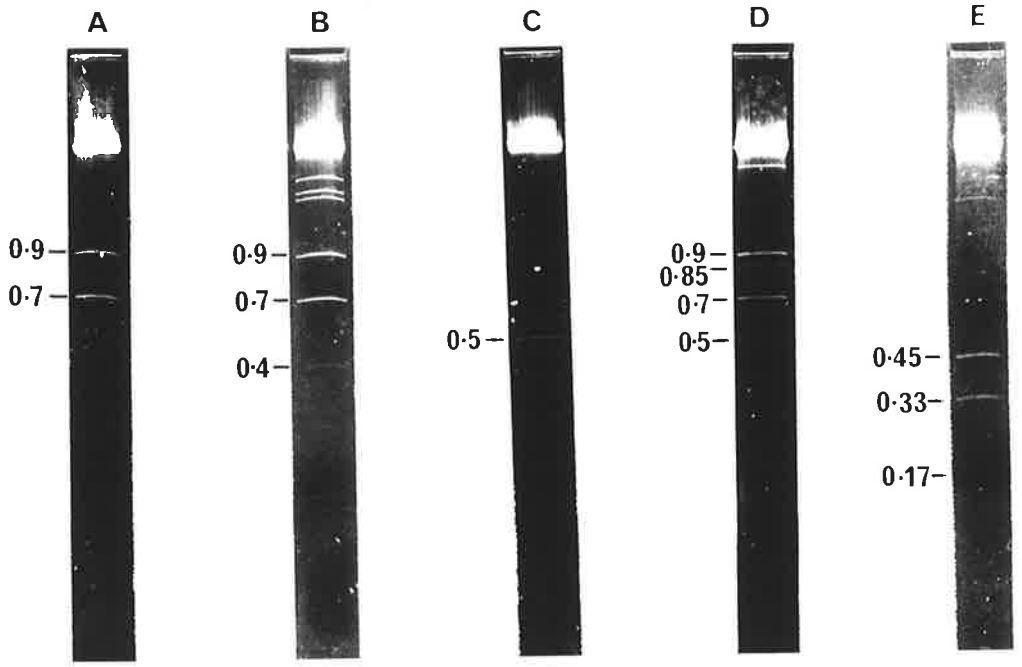

A
B
C
D
E
0·9
0·7
0·9
0·7
0·4
0·5
0·9
0·85
0·7
0·5
0·45
0·33
0·17

Figure 5.25

Restriction Maps of λCHO7

The fragments obtained after restriction enzyme cleavage of λCHO7 with various enzymes are indicated (sizes are given in kilobases). The overall derived restriction map of the λCHO7 insert is also shown. Gene positions were determined from Southern analysis (see Figs 5.26 & 5.30). *Sal*I does not cleave the λCHO7 insert.

E: *Eco*RI

H: *Hin*dIII

B: *Bam*HI

X: *Xho*I

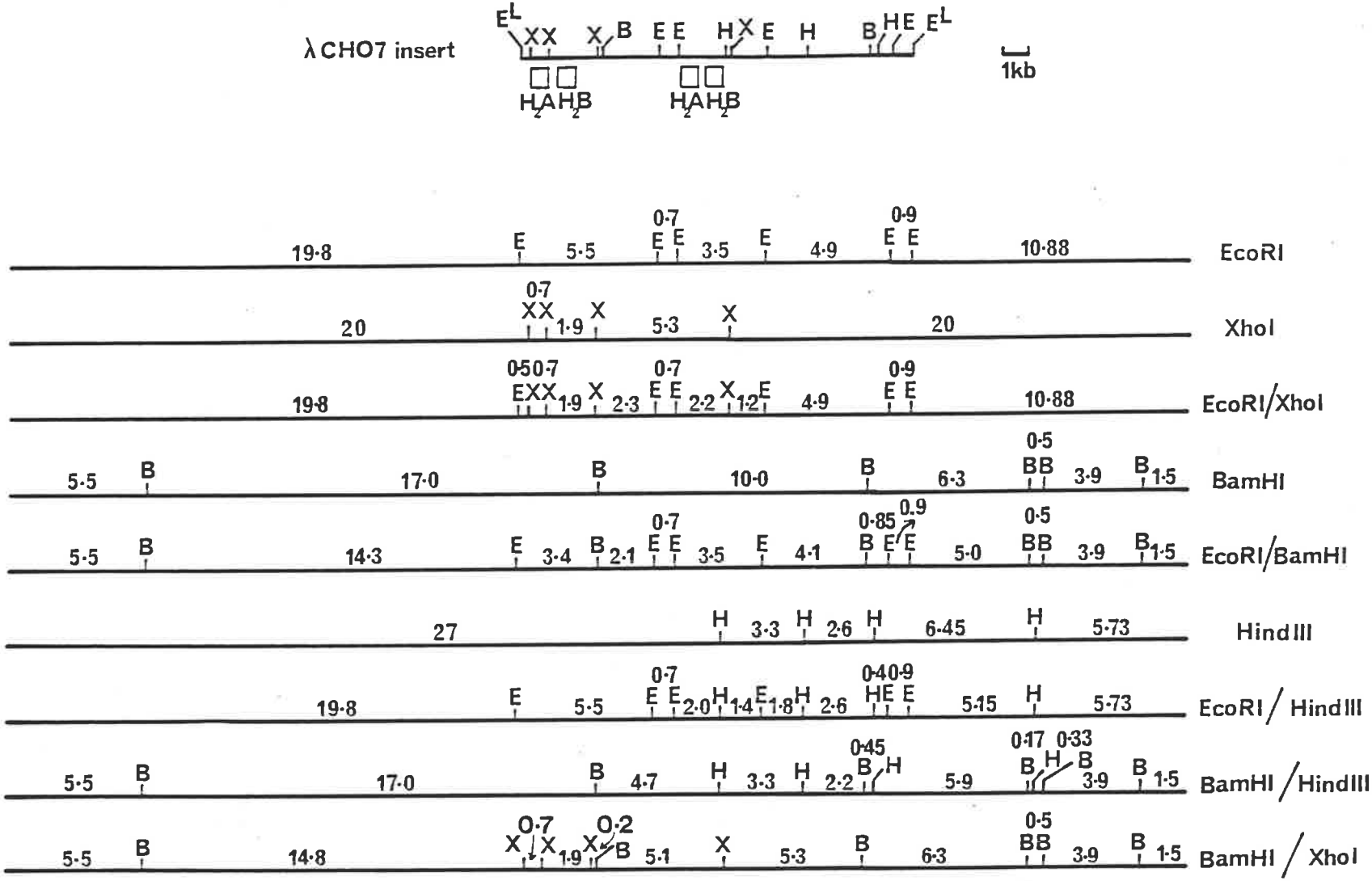
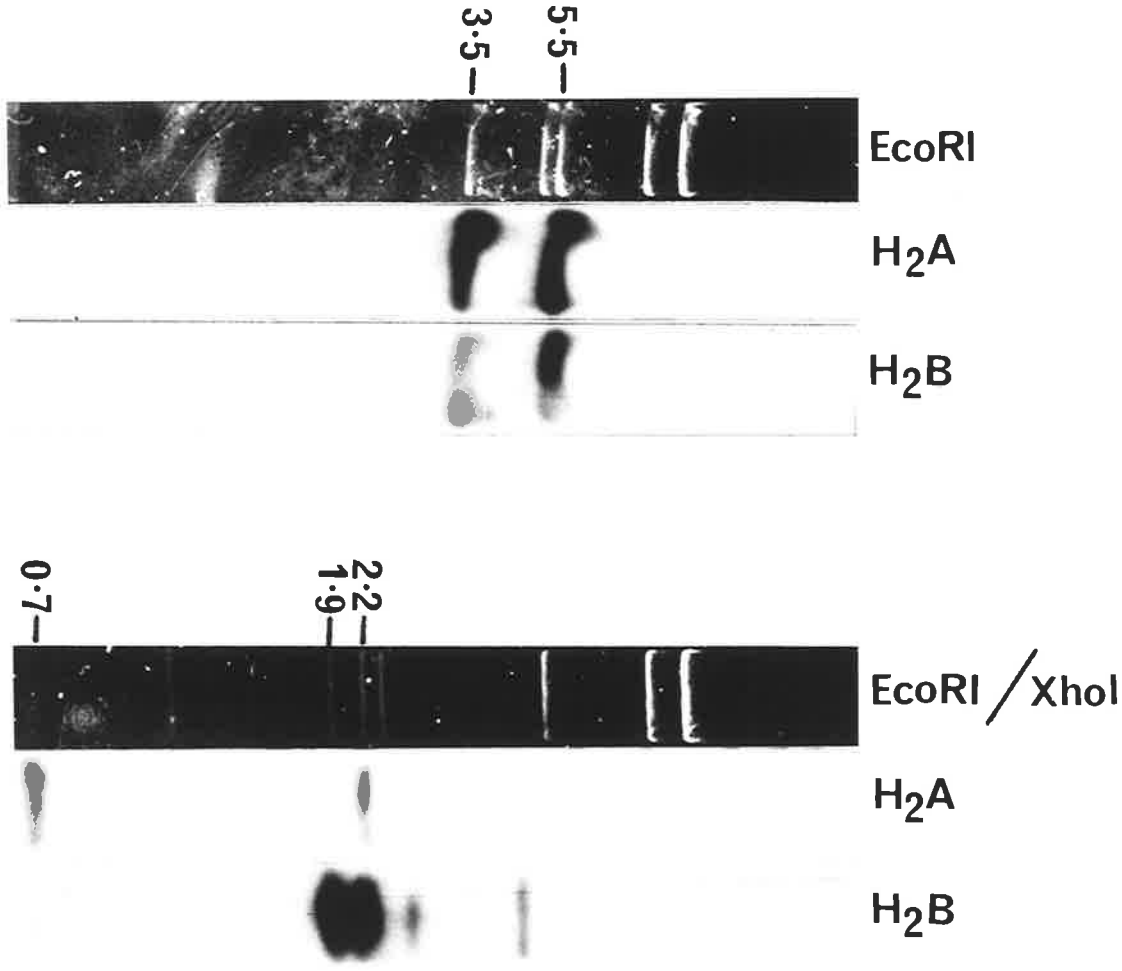

λCHO7 insert
1kb
EcoRI
XhoI
EcoRI/XhoI
BamHI
EcoRI/BamHI
HindIII
EcoRI / HindIII
BamHI / HindIII
BamHI / XhoI

Figure 5.26

Southern Analysis of λCHO7

Cleaved λCHO7 DNA was transferred to nitrocellulose filters using the procedure described (section 2.2.3), and hybridised to either H_2A or H_2B gene specific probe. Filters were washed in 0.5xSSC/0.1% SDS at 65°C and autoradiographed overnight. Only the digests discussed in the text are shown. Fragment sizes are indicated in kilobases.


3·5
5·5
EcoRI
H2A
H2B
0·7
1·9
2·2
EcoRI / XhoI
H2A
H2B

5.2.14 Isolation of λCHO7 Subclones

The three *Eco*RI-*Eco*RI fragments from λCHO7 (5.5 kb, 4.9 kb and 3.5 kb) were subcloned into the plasmid pBR325 for further characterisation. Insert containing colonies were identified by antibiotic selection and confirmed by the miniscreen procedure (section 2.2.9b). Large-scale DNA preparations were carried out for recombinants with inserts of correct size. These clones were subsequently referred to as pCH5.5E, pCH4.9E and pCH3.5E. Clones pCH5.5E and pCH4.9E do not contain histone gene sequences and are not discussed further.

5.2.15 Restriction Analysis of pCH3.5E

pCH3.5E contains the 3.5 kb, H_2A and H_2B positive, *Eco*RI-*Eco*RI fragment of λCHO7 subcloned into the *Eco*RI site of pBR325. This DNA has been restriction mapped with *Sac*I, *Sac*II, *Xho*I, *Sma*I and *Hin*dIII, and the location of the H_2A and H_2B coding regions accurately determined by hybridisation analysis.

The size and relative positions of restriction fragments is indicated in Figure 5.29.

H_2A Coding Region

H_2A probe hybridised to the 0.55 kb *Sac*I-*Sac*II fragment and the adjacent 0.64 kb *Sac*I-*Sac*I fragment (Fig. 5.30). The relative intensity of hybridisation suggested that most of the H_2A coding region was contained within the 0.64 kb *Sac*I-*Sac*I fragment but spanned the *Sac*I site as shown (Fig. 5.29).

H_2B Coding Region

The H_2B specific probe hybridised to the 0.8 kb *Sac*I-*Hin*dIII fragment (Fig. 5.30) and on longer exposure it became clear that the adjacent 0.64 kb *Sac*I-*Sac*I fragment was also hybridising to this probe. This suggested that the hybridising region spanned the *Sac*I site but was predominantly located in the 0.8 kb *Sac*I-*Hin*dIII fragment as shown in Figure 5.29.

Figure 5.27

Restriction Analysis of pCH3.5E

1 ug samples of pCH3.5E DNA were digested with a range of restriction enzymes (see below) and electrophoresed on a 1.2% agarose gel (section 2.2.2a). Bands were detected with UV light after staining with ethidium bromide. Fragment sizes are indicated in kilobases. The figure shows pCH3.5E cleaved with the following restriction enzymes;

Track A: *Sac*I

Track B: *Sac*I and *Eco*RI

Track C: *Sac*I and *Hin*dIII

Track D: *Sac*II

Track E: *Sac*II and *Eco*RI

Track F: *Sac*II and *Hin*dIII

Track G: *Sma*I and *Eco*RI

Track H: *Sma*I and *Bam*HI

Track I: *Sma*I and *Xho*I

Track J: *Xho*I and *Bam*HI

Track K: *Xho*I and *Hin*dIII

Track L: *Eco*RI and *Hin*dIII

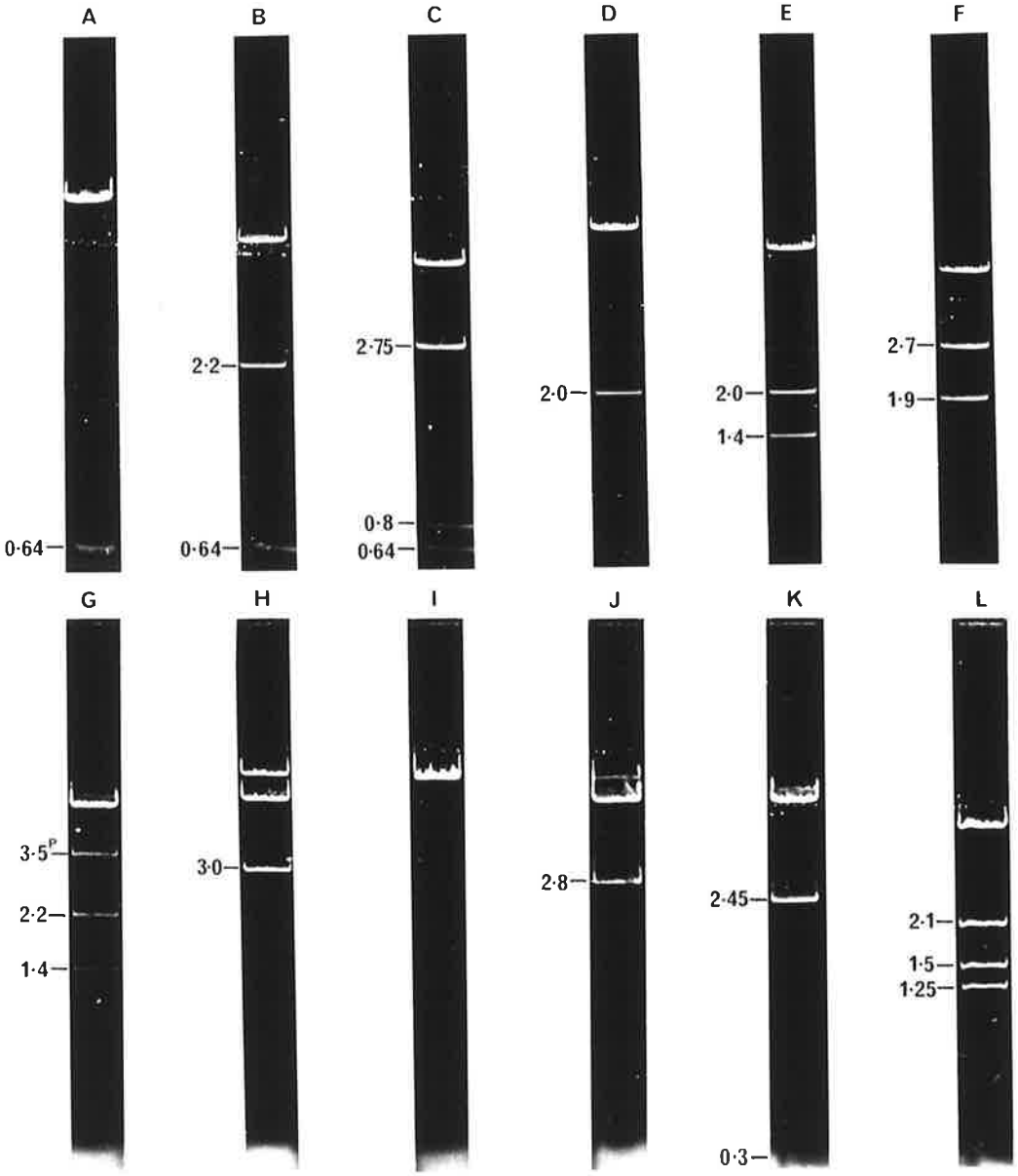

A
B
C
D
E
F
0·64
2·2
0·64
2·75
0·8
0·64
2·0
2·0
1·4
2·7
1·9
G
H
I
J
K
L
3·5P
2·2
1·4
3·0
2·8
2·45
0·3
2·1
1·5
1·25

Figure 5.28

Restriction Analysis of pCH3.5E

1 ug samples of pCH3.5E DNA were digested with various restriction enzymes, as indicated below, and fractionated on a 6% polyacrylamide gel. *Hin*fI digested pBR322 DNA (1 ug) was co-electrophoresed as marker DNA and fragments were visualised with UV light after staining with ethidium bromide. The following restriction enzyme digests of pCH3.5E are shown;

Track A: *Sma*I and *Xho*I

Track B: *Sac*II and *Xho*I

Track C: *Sac*II and *Hin*dIII

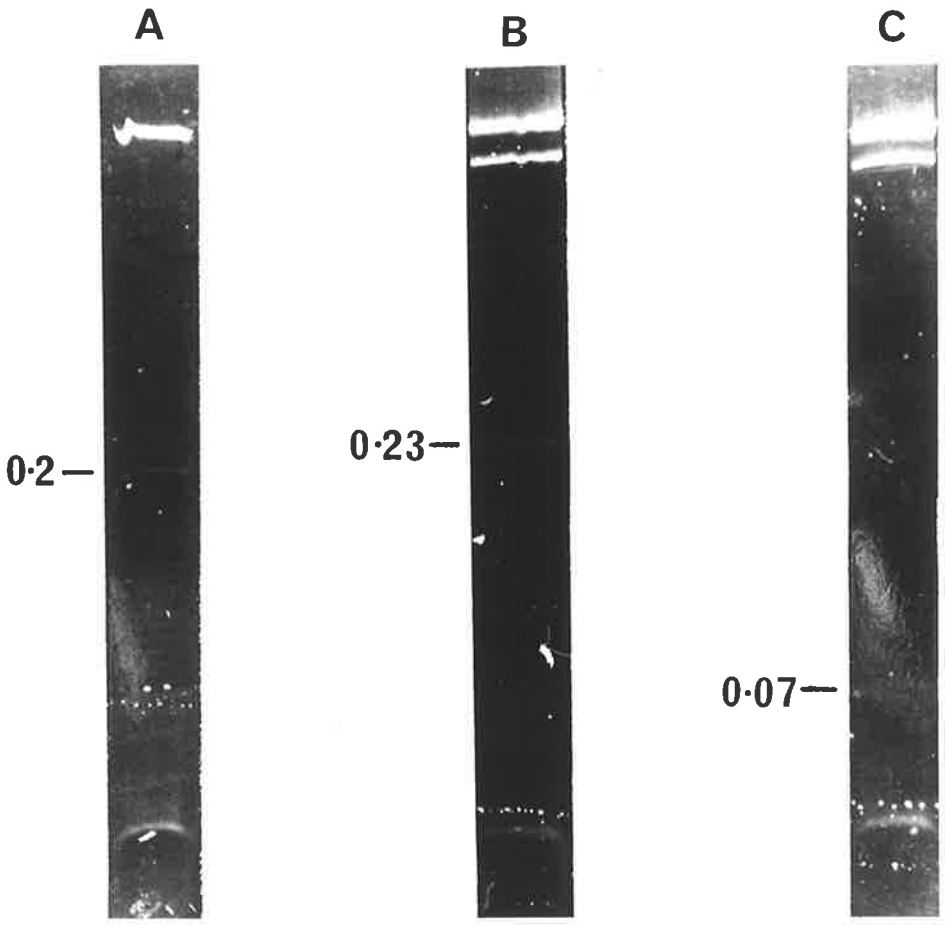

A
B
C
0·2
0·23
0·07

Figure 5.29

Restriction Maps of pCH3.5E

The restriction fragments generated on digestion of pCH3.5E DNA with various combinations of restriction enzymes are shown together with the derived map of the pCH3.5E insert. The predicted positions of the H_2A and H_2B genes, as determined from Southern analysis, are indicated. Fragment sizes are in kilobases.

E: *Eco*RI

SI: *Sac*I

SII: *Sac*II

X: *Xho*I

Sm: *Sma*I

H: *Hin*dIII

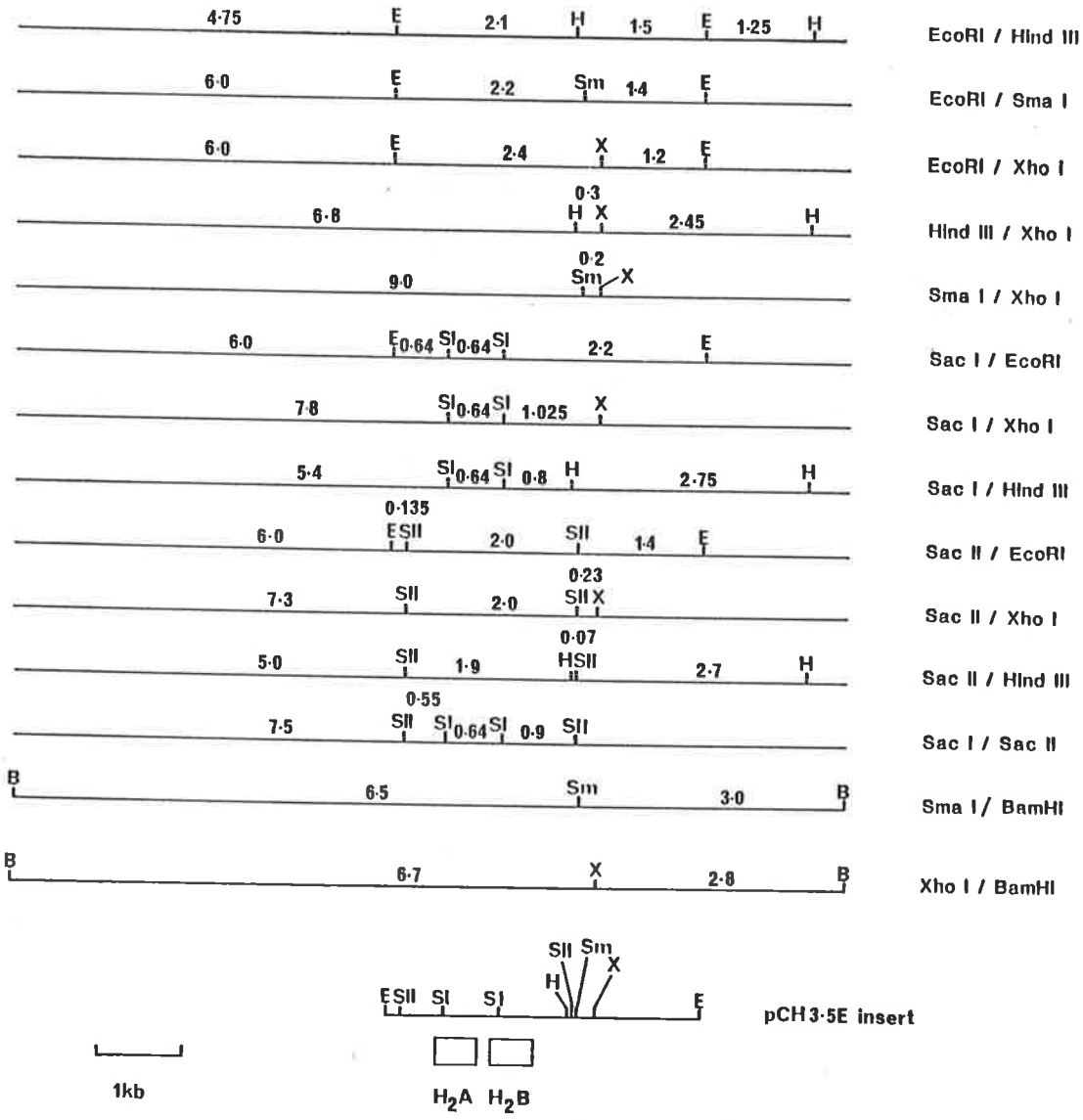
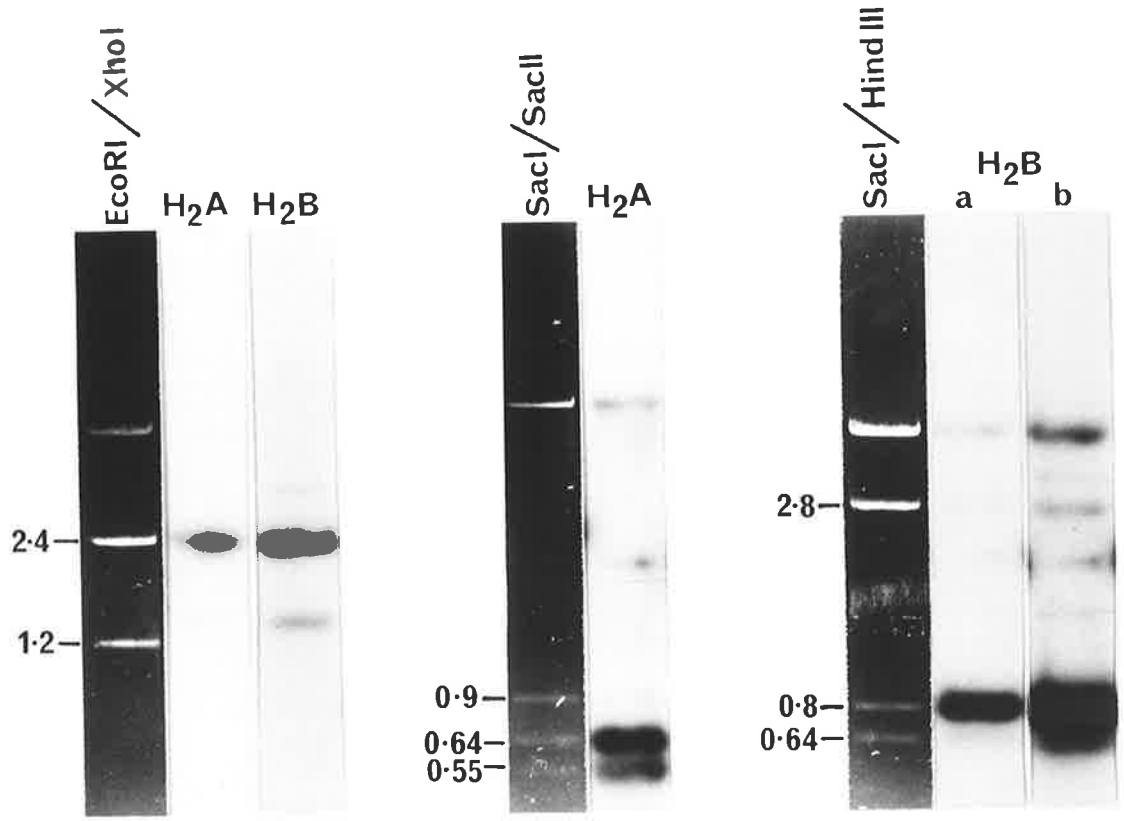

4·75 E 2·1 H 1·5 E 1·25 H
EcoRI / Hind III
6·0 E 2·2 Sm 1·4 E
EcoRI / Sma I
6·0 E 2·4 X 1·2 E
EcoRI / Xho I
0·3
6·8 H X 2·45 H
Hind III / Xho I
0·2
9·0 Sm X
Sma I / Xho I
6·0 E 0·64 SI 0·64 SI 2·2 E
Sac I / EcoRI
7·8 SI 0·64 SI 1·025 X
Sac I / Xho I
5·4 SI 0·64 SI 0·8 H 2·75 H
Sac I / Hind III
0·135
6·0 E SII 2·0 SII 1·4 E
Sac II / EcoRI
0·23
7·3 SII 2·0 SII X
Sac II / Xho I
0·07
5·0 SII 1·9 H SII 2·7 H
Sac II / Hind III
0·55
7·5 SII SI 0·64 SI 0·9 SII
Sac I / Sac II
B 6·5 Sm 3·0 B
Sma I / BamHI
B 6·7 X 2·8 B
Xho I / BamHI
SII Sm
H X
E SII SI SI E
pCH3·5E insert
1kb
H2A H2B

Figure 5.30

Southern Analysis of pCH3.5E

Restriction enzyme digested pCH3.5E DNA which had been fractionated on a 1.2% agarose gel was transferred to two nitrocellulose filters (section 2.2.3). Filter-bound DNA was hybridised with either H_2A or H_2B gene specific probe, washed in 0.5xSSC/0.1% SDS at 65°C and autoradiographed for 4 hours at -80°C with an intensifying screen. Fragment sizes are given in kilobases. A 4 hour exposure (a) and an overnight exposure (b) are shown for the filter probed with H_2B to indicate the hybridisation of the 0.64 kb fragment.


EcoRI / XhoI
H2A
H2B
2·4
1·2
SacI/SacII
H2A
0·9
0·64
0·55
SacI/HindIII
H2B
a
b
2·8
0·8
0·64

5.2.16 Search for Clones Extending λCHO7

The 0.85 kb *Bam*HI-*Eco*RI fragment from the right-hand end of λCHO7 (Fig. 5.25) was isolated after digestion of pCH4.9E with *Bam*HI and *Eco*RI. This fragment was eluted from LGT agarose, labelled with ^{32}P by nick-translation and used to score the histone clone bank.

The only clone which hybridised with this probe was λCHO7 (isolate 7.2; result not shown) implying that none of the other genomic clones contain sequences derived from this region. Consequently no overlapping clones were isolated.

5.2.17 Other Histone-Gene Containing Genomic Clones

Sugarman *et al* (1983) have characterised a series of clones which were also isolated from the chicken genomic library constructed by Dodgson *et al* (1979). One of these clones, λCH10a, was recognised as a possible extension of λCHO3. This clone was kindly provided by Dr. D. Engel, and the possibility that this clone overlapped with λCHO3 was further analysed.

An *Eco*RI/*Hin*dIII digest of λCH10a DNA was transferred to nitrocellulose and probed with ^{32}P-labelled 2.5 kb *Pst*I-*Pst*I fragment from λCHO3. This non-histone coding probe hybridised to a 4.5 kb *Eco*RI-*Hin*dIII fragment from λCH10a (Fig. 5.31) and confirmed that this genomic clone, isolated and characterised independently from the same genomic library, extends the region spanned by the four clones, λCHO3, λCHO1, λCHO5 and λCHO7. λCH10a contains the H1 gene present in λCHO3 and additional H3 and H4 histone gene sequences (Sugarman *et al*, 1983).

5.2.18 Discussion

An overall representation of the region characterised and the clones described is presented in Figure 5.32. This region spans 50 kb and contains 13 histone genes. Further extension of this region was

Figure 5.31

Southern Analysis of λCH10a

1 ug of λCH10a DNA was digested with *Eco*RI and *Hin*dIII, fractionated on a 1% agarose gel, transferred to a nitrocellulose filter and hybridised with ^{32}P-labelled probe (2.5 kb *Pst*I-*Pst*I fragment from pCH5.0E, see Fig. 5.8). The filter was washed in 0.5xSSC/0.1% SDS at 65^{o}C and bands were detected by autoradiography. Details of all procedures used are given in Chapter 2.

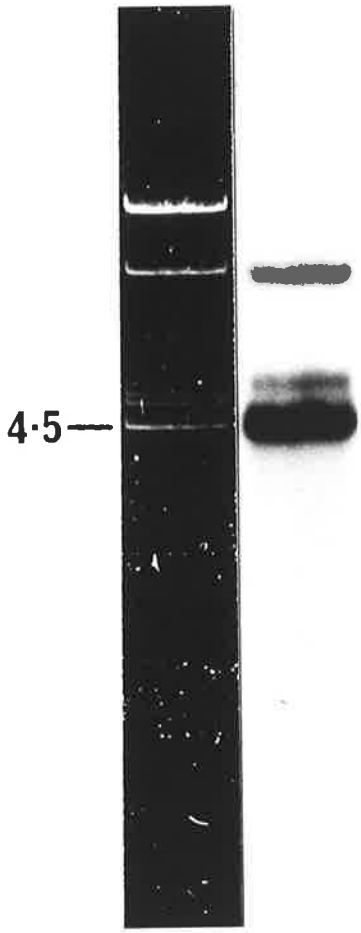

4·5

Figure 5.32

Schematic Representation of λ-recombinants

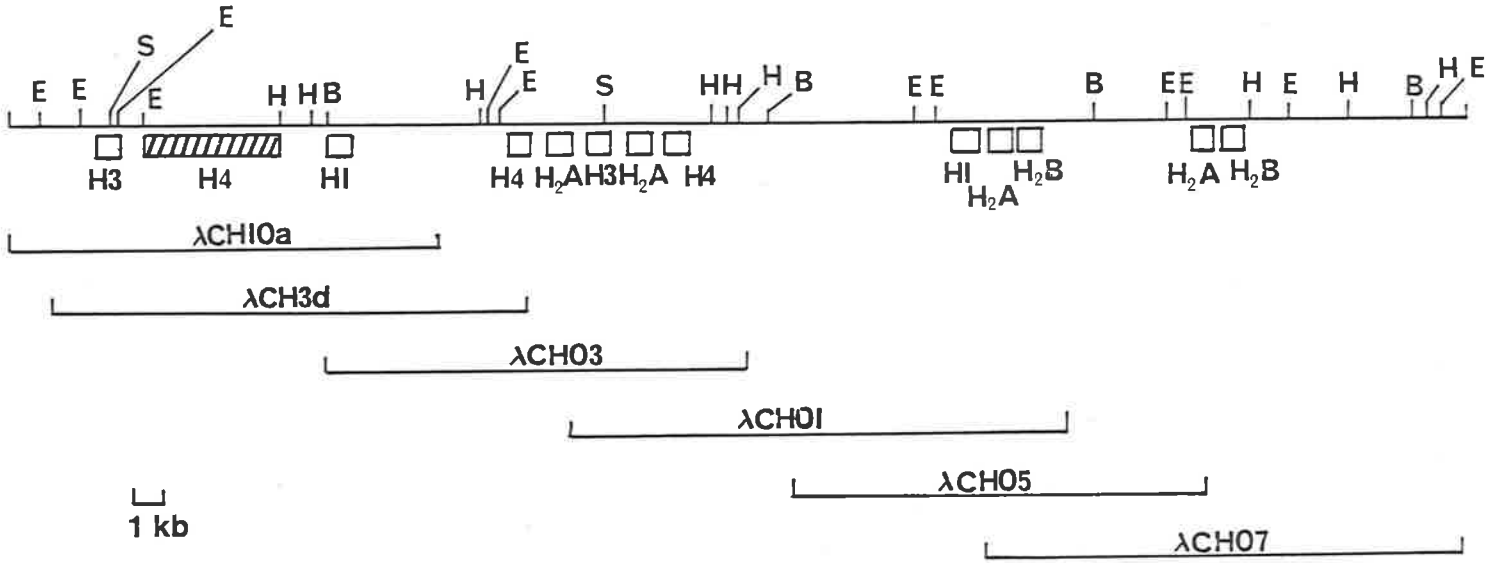
Mapping data for λCHO1 is from Harvey *et al* (1981). Data for λCH1Oa and λCH3d is from Sugarman *et al* (1983). Restriction sites are indicated (see below). Histone genes are indicated by the boxes. A 4.0 kb *Eco*RI-*Hin*dIII fragment in λCH1Oa and λCH3d hybridises to H4 probe (Sugarman *et al*, 1983) and is indicated; the number of H4 genes in this region is unknown.

E: *Eco*RI

S: *Sal*I

H: *Hin*dIII

B: *Bam*HI


E
S
E
E E E H H B H E E S H H H B E E B E E H E H B H E
H3 H4 HI H4 H2A H3 H2A H4 HI H2A H2B H2A H2B H2A
λCH10a
λCH3d
λCH03
λCH01
λCH05
λCH07
1 kb

attempted by screening the clone bank described in Chapter 4 for overlapping clones, but this proved unsuccessful. There are two possible reasons why this screening did not reveal overlaps.

Firstly, the histone genes may not all be present at one "locus". There may be separate, unlinked histone gene loci or a few linked but non-continuous loci. Alternatively, the histone genes could be restricted to a single locus, but with relatively large distances between the genes or clusters. In these cases, the histone gene region could not be completely covered by analysing histone gene containing clones.

Secondly, a non-random distribution of *Hae*III or *Alu*I recognition sequences in a particular region could prevent this region being represented in a genomic library made from *Hae*III and *Alu*I partially digested DNA. Such factors as local G + C content and the presence of repeated DNA sequences can affect the distribution of restriction sites and lead to the generation of fragments too small or large to be cloned.

There were two alternative approaches which would extend the data already obtained.

Non-coding probes, representing the ends of the region so far characterised, could have been used to screen the chicken λ-library for overlapping recombinants. However, extending a genomic region in this manner requires many successive library screenings, as the clones obtained will represent only between 11-20 kb of DNA. The development of techniques for generating libraries of large genomic DNA fragments (between 35-45 kb of DNA) in small "cosmid" vectors (see Chapter 6) overcomes this disadvantage and therefore it was decided to attempt the construction of a complete chicken cosmid library to facilitate future chromosome walking experiments.

CHAPTER 6

CONSTRUCTION, ISOLATION AND CHARACTERISATION OF CHICKEN HISTONE COSMID RECOMBINANTS

6.1 INTRODUCTION

Plasmid vectors that can be packaged into 'phage heads ("cosmids") were originally constructed by Collins and Hohn (1978). These vectors accept large DNA inserts (about 40 kb) and are introduced into host bacteria via *in vitro* packaging and infection.

The essential component of a cosmid vector is the λ-derived *cos* sequence, which is recognised and cleaved by the λ A-gene product during the packaging of 'phage DNA. Cleavage generates 12 base single strand cohesive termini which anneal and are sealed by *E. coli* DNA ligase inside the host cell. It has been shown (Syvanen, 1974) that DNA 39-45 kb long containing *cos* sequences can be packaged *in vitro*, irrespective of origin.

A number of small (6-10 kb) cosmid vectors have been constructed containing this *cos* sequence, a plasmid origin of replication enabling propagation, drug resistance markers and unique restriction sites for cloning. The small size of these vectors allows very large fragments to be cloned.

Vector DNA is ligated to genomic insert DNA with the aim of forming chimeric molecules in which the foreign DNA is linked to a pair of cosmid molecules, such that the *cos* sequences are in the same orientation and approximately 45 kb apart, with all the plasmid information represented between the *cos* sites. These molecules will be packaged and, after adsorption to *E. coli*, recombinant DNA is injected into the host cell, circularised and propagated as a large plasmid.

Protocols for cloning into cosmid vectors must overcome two major technical difficulties viz;

1. Formation of vector concatamers leading to recombinant molecules with small genomic inserts which can be packaged. These molecules are subject to intramolecular recombination between vector sequences and are capable of generating faster replicating cosmids, resulting in

the loss of recombinants by segregation.

2. Scrambling of genomic sequences. This occurs if two non-contiguous genomic fragments are inserted into the same vector molecule.

Protocols have been developed which overcome both of these problems. There are two approaches, differing in the method used to prevent self-ligation of vector DNA. In one case (Meyerowitz *et al*, 1980; Grosveld *et al*, 1981) the linear vector is dephosphorylated. The alternative procedure, developed by Ish-Horowitz and Burke (1981) was used here and is summarized in Figure 6.2.

The success of these two protocols, and the development of the high-density colony screening technique by Hanahan and Meselson (1980) has made cosmid cloning the method of choice for isolation of large genomic fragments, which facilitate chromosome walking experiments, and isolation of large genes.

Cosmid cloning has been advantageous in the study of a number of gene clusters (Grosveld *et al*, 1981; Cattaneo *et al*, 1981; Steinmetz *et al*, 1982) and in the transformation of yeast and mammalian cells with large DNA segments (Morris *et al*, 1981; Grosveld *et al*, 1982).

6.2 RESULTS

6.2.1 Preparation of High Molecular Weight Chicken DNA

In constructing a cosmid library it is essential that the genomic DNA used is of high molecular weight, to ensure that fragments obtained after partial digestion have two cohesive ends available for ligation into vector molecules.

Chicken DNA was prepared by Richard Harvey (this laboratory), from chicken erythrocytes using a modification of the method described by Gross-Bellard *et al* (1973). The DNA was greater than 100 kb in length as determined by 0.3% agarose gel electrophoresis.

6.2.2 Partial Digestion of Chicken DNA (Linda Tabe, this laboratory)

High molecular weight chicken DNA was partially digested with

*Sau*3AI. The resultant DNA fragments with *Sau*3AI cohesive ends can be ligated into the *Bam*HI site of vector DNA. Conditions of digestion were selected to produce DNA fragments of average size 35-45 kb. (Conditions used were 6 units of enzyme for 5 minutes at 37°C with 100 ug of chicken DNA). Chicken DNA was digested in six 500 ul digestions, each containing 100 ug of DNA. Two digestions were carried out for the optimal time determined, two for one minute longer and two for one minute shorter. Reaction mixes were prewarmed to 37°C and digestions were stopped by the addition of SDS to a final concentration of 0.1%.

Insert DNA was dephosphorylated in the same reaction mix using Calf Intestinal Phosphatase (section 2.2.7), phenol/chloroform extracted three times to ensure no trace of phosphatase activity remained, ethanol precipitated, resuspended in water and pooled. The DNA was sized by electrophoresis alongside marker DNA on 0.3% agarose gels.

6.2.3 Size Fractionation of *Sau*3AI Partially Digested Chicken DNA

(Linda Tabe, this laboratory)

In their original procedure, Ish-Horowitz and Burke (1981) did not select DNA fragments of a particular size class for ligation to the cosmid vector. We have modified this protocol to include a sizing step and insert dephosphorylation. There are two reasons for including these procedures.

Firstly, DNA fragments greater than 45 kb or less than 30 kb cannot be incorporated into viable recombinant 'phage. By removing these insert molecules the productive ligation efficiency is increased. Also, by selecting DNA with an average size of 40 kb, large inserts are ensured and the possibility of introducing more than one non-continuous fragment of chicken DNA into a single recombinant 'phage is reduced. (This event should occur rarely with dephosphorylated insert).

Chicken DNA, partially digested with *Sau*3AI, was fractionated after centrifugation on 5-25% sucrose gradients (25,000 rpm, 16 h,

4°C, SW41 rotor; section 2.2.4). 0.5 ml fractions were collected, analysed on 0.3% agarose gels, and those containing DNA in the range 35-45 kb were pooled, ethanol precipitated and resuspended at a concentration of 1 ug/ul.

6.2.4 Preparation of Cosmid Vector

The vector used for construction of the cosmid library, pHC79 (Hohn and Collins, 1980), is shown in Fig. 6.1.

pHC79 DNA was prepared by the alkali lysis method (2.2.10b) and supercoiled DNA purified by centrifugation through CsCl density gradients (2.2.10c).

This DNA was digested in two separate aliquots, with either *Sal*I or *Eco*RI which cleave on opposite sides of the *Bam*HI cloning site. (see Fig. 6.2). After digestion these cohesive ends were made incapable of ligation by treatment, in the same reaction mix, with Calf Intestinal Phosphatase (section 2.2.7c). The efficiency of the dephosphorylation reaction was checked by assaying ligation efficiency.

Linear, dephosphorylated pHC79 molecules were then digested with *Bam*HI. 100 mM ATP was included in this reaction to prevent dephosphorylation of *Bam*HI termini by contaminating phosphatase. Vector fragments containing the *cos* sequence were purified either by LGT agarose gel electrophoresis (section 2.2.2a; Fig. 6.3) or by sucrose gradient centrifugation (section 2.2.4). Best yields were obtained by centrifugation on 10-40% sucrose gradients but this resulted in minor contamination with small vector fragments. After purification, vector fragments were precipitated with ethanol and resuspended at a concentration of 450 ng/ul.

Purified vector fragments were incubated with DNA ligase and genomic DNA and ligation efficiency was assayed on 1% agarose minigels (section 2.2.2a). The result of this assay is shown in Figure 6.4. Greater than 80% of each arm was incorporated into concatamer suggesting

Figure 6.1

Cosmid Vector pHC79

A restriction map of the plasmid vector pHC79. Data from Hohn and Collins (1980).

Restriction sites are indicated as follows:

E: *Eco*RI

B: *Bam*HI

Sal: *Sal*I

Ava: *Ava*I

Bgl: *Bgl*II

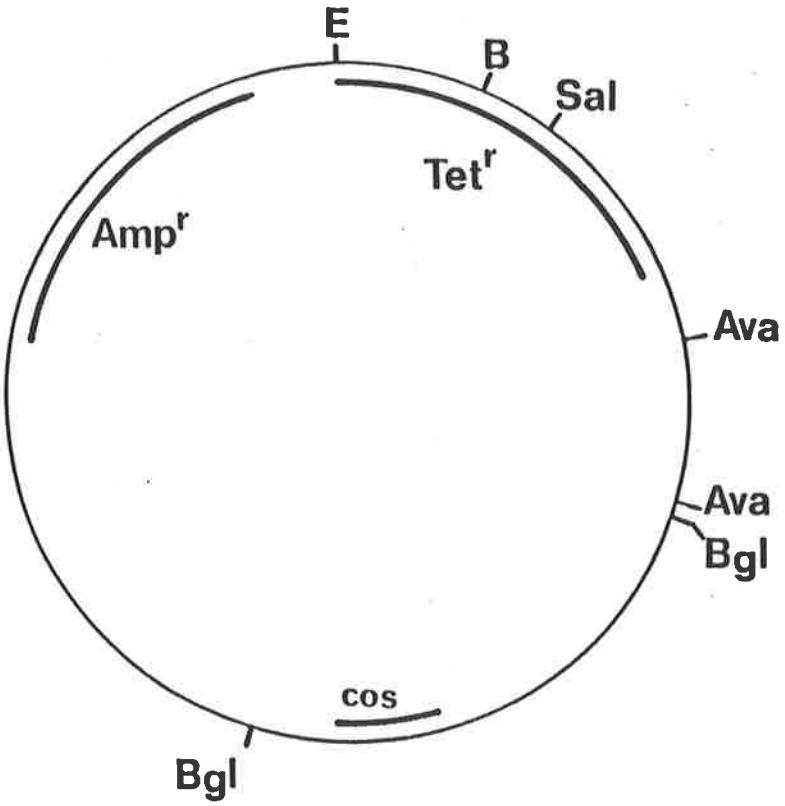

E
B
Sal
Tet^r
Amp^r
Ava
Ava
Bgl
cos
Bgl

Figure 6.2

Protocol for Cosmid Library Construction

This procedure is a modification of the protocol described by Ish-Horowicz and Burke (1981). Following digestion of pHC79 DNA with *Eco*RI or *Sal*I and inactivation of the protruding termini by Calf Intestinal Phosphatase, CIP, (section 2.2.7), the linear vector DNA is cleaved by *Bam*HI. The two vector fragments containing the *cos* sequences are isolated and ligated to genomic fragments (size fractionated; 35-45 kb) of eukaryotic DNA generated by partial digestion with *Sau*3AI. An entire complement of plasmid DNA is contained between the two *cos* sequences. The concatamers formed are used as substrates for *in vitro* packaging (section 2.2.6d). On introduction into *E. coli* host HB101 (section 2.1.4) the DNA is circularised and replicates in the form of a large plasmid, conferring resistance to ampicillin. Details of all procedures used in cosmid library construction are given in section 2.2.6.

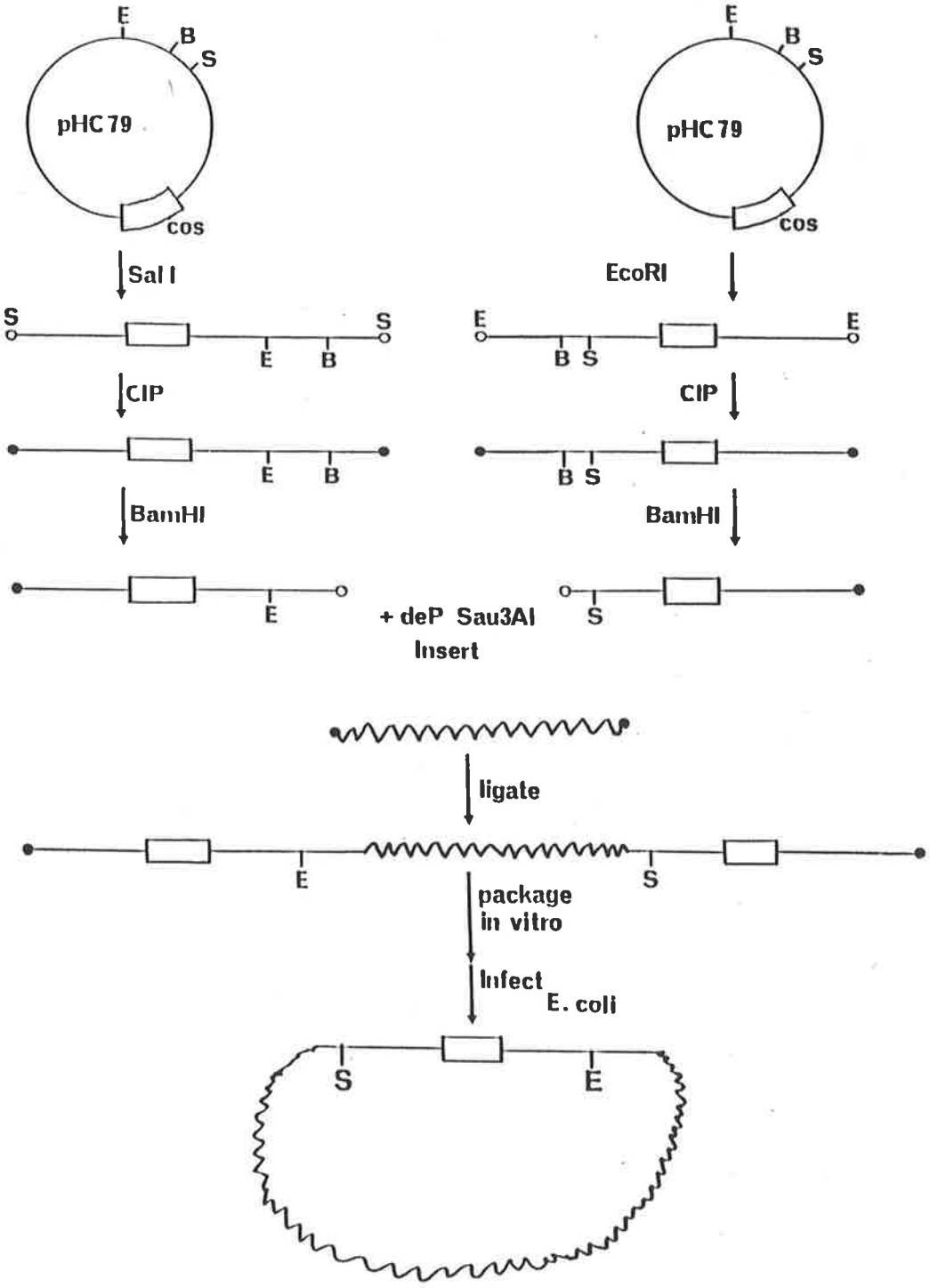

E
B
S
pHC 79
cos
Sal I
S
S
E
B
CIP
E
B
BamHI
E
E
B
S
EcoRI
E
E
B S
CIP
B S
BamHI
S
+ deP Sau3AI
Insert
ligate
E
S
package
in vitro
Infect
E. coli
S
E

Figure 6.3

Purification of Vector Arm Fragment

50 ug of pHC79 DNA which had been digested with *Eco*RI, dephosphorylated and redigested with *Bam*HI were electrophoresed on a 1.2% LGT agarose gel (section 2.2.2a) and visualised under UV light after brief ethidium bromide staining. The 5.8 kb vector arm fragment was excised from the gel, the DNA recovered by elution (section 2.2.2a) and concentrated by precipitation with ethanol.


5·8kb
vector arm
0·35kb

Figure 6.4

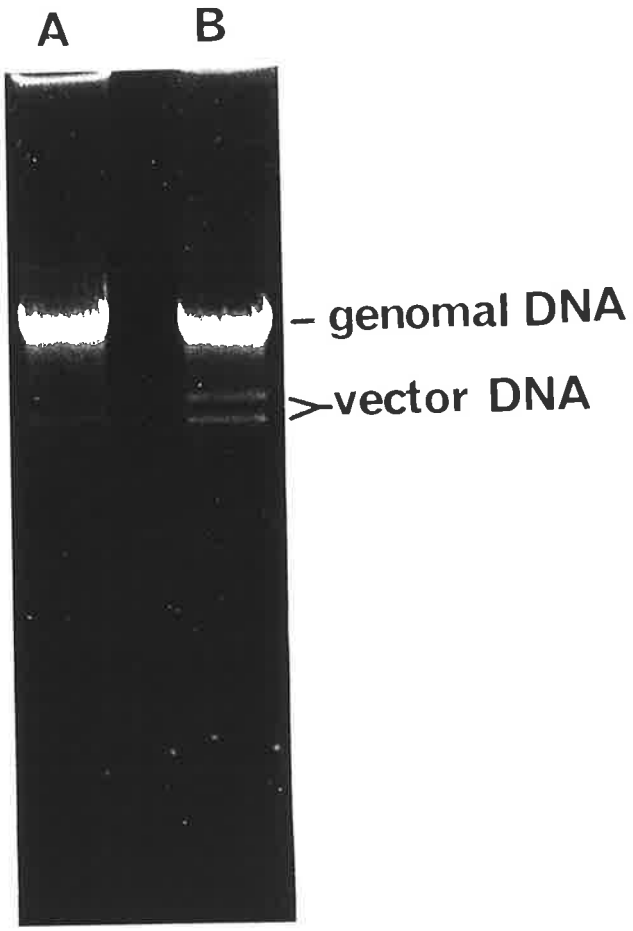
Trial Ligation of Vector and Genomic DNA

Vector arm fragments (100 ug of each) were incubated with 500 ug of partial *Sau*3AI chicken DNA (35–45 kb fraction) in a 10 ul ligation reaction at $4^{o}C$ overnight. After incubation DNA was fractionated on a 1% agarose minigel (section 2.2.2a) and viewed under UV light following ethidium bromide staining.

track A : + DNA ligase

track B : – DNA ligase

In the presence of DNA ligase greater than 80% of the vector fragments were incorporated into concatamers.


A
B
genomal DNA
vector DNA

these vector fragments were suitable for library construction.

6.2.5 Ligation of Chicken DNA to pHC79 Vector DNA

Vector arms were ligated to partial *Sau*3AI 35-45 kb chicken DNA using approximately a 2 fold molar excess of each arm. The desired concatamer is : vector 1 (*Sal*I-*Bam*HI) - insert - vector 2(*Bam*HI-*Eco*RI) (see Fig. 6.2). If it is assumed that 50-80% of vector molecules have *Bam*HI ends available for ligation (i.e. phosphorylated) then this ratio will ensure that there is not an excess of insert molecules over effective vector molecules. The total concentration of DNA in the preparative ligation was deliberately very high (540 ug/ul) to ensure a high ligation efficiency.

The cosmid library was prepared from two preparative ligation reactions, each containing 3.6 ug of size fractionated chicken DNA and 0.9 ug of each vector arm. These reactions were incubated overnight at 4^{o}C with 1.0 Weiss unit of T4 DNA ligase (section 2.2.6a).

6.2.6 *In Vitro* Packaging & Cosmid Infection

The principle of *in vitro* packaging is as follows: 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, DNA is packaged into 'phage heads. The full heads are then matured *in vitro* into viable infectious 'phage in the presence of proteins from genes *W* and *F*11 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*, strain BHB2690) and therefore accumulates these precursors, while the other (BHB2688, *Eam*) is inhibited in forming capsid protein. Lysates from these two strains provide all the necessary components to convert λ DNA into infectious 'phage particles.

Using the packaging procedure described in section 2.2.6d,

packaging extracts were prepared which gave between 1.1 and 2.5 x 10^8 pfu/ug of λ DNA.

Ligated cosmid/chicken DNA was packaged into bacteriophage particles using the procedure detailed (section 2.2.6d). The resultant 'phage stocks were used to infect *E. coli* host HB101 and gave between 0.5 and 5 x 10^4 cfu/ug insert DNA when spread on ampicillin plates.

The two large scale ligation reactions (section 6.2.5) were each divided into 8 aliquots (0.45 ug insert DNA in each) and packaged separately (16 packaging reactions). The efficiency of this large scale packaging was 1.7 x 10^4 cfu/ug insert DNA or approximately 120,000 cfu total. (7.2 ug total ligated insert).

These packaged stocks were pooled and adsorbed to host bacteria. Cells were plated on eight large detergent free nitrocellulose filters on ampicillin plates. This procedure was expected to generate a library of chicken genomic DNA consisting of 120,000 colonies (15,000 colonies per plate), however, an average of 6,600 colonies per plate, and a total population of approximately 52,800 were obtained.

The high concentration of free 'phage tails present during the large scale adsorption may be a possible cause of the low plating efficiency. These 'phage tails may adsorb to λ receptors on the host cell surface and compete with intact 'phage for these sites.

To test this, 200 cfu were adsorbed to *E. coli* strain HB101 (0.3 ml, A_{600} = 1.0) in the presence and absence of packaging extracts. The number of ampicillin resistant colonies was reduced approximately five-fold after adsorption in the presence of the packaging lysates (data not shown). It was shown, subsequently, that this inhibitory effect could be overcome by adsorbing the packaged stock (10^4-10^5 cfu) to a larger number of cells (2 mls). After adsorption cells were pelleted by centrifugation and resuspended in a smaller volume (0.5 ml) before spreading on ampicillin plates. This procedure gave an equivalent number

of colonies when recombinant 'phage particles were adsorbed and plated in the presence and absence of packaging extracts.

As the chicken genome is 2×10^6 kb in length and the average size of the cloned segments should be 40 kb (35-45 kb fragments used for ligation) 50,000 colonies should equal one genome equivalent.

Whilst a higher titre of cosmid recombinants would have been more acceptable, it was considered that this bank of recombinants provided a sufficient representation of the chicken genome to permit screening for histone gene containing clones.

6.2.7 Replication & Screening of Cosmid Recombinants

Replica filters were prepared from the set of eight master filters (section 6.2.6) using the filter to filter procedure described by Hanahan and Meselson (1980) and detailed in section 2.2.6f.

Colonies on the master filters were regrown on ampicillin plates, then transferred to ampicillin plates containing 25% glycerol for incubation at $37^{o}C$, for 1 hour, before being sandwiched with a fresh sterile filter and stored in a sealed plastic bag at $-80^{o}C$.

The new filters generated were incubated at $37^{o}C$ on ampicillin plates overnight, further replicated, and then, after reincubation at $37^{o}C$, stored, inverted and sealed, at $4^{o}C$ until hybridisation results were available.

After colonies had grown to 0.5-1 mM in diameter on the filters to be screened, these filters were sandwiched with a fresh filter and DNA liberated from the sandwiched colonies using the procedure of Grunstein and Hogness (1975) with modifications as described in section 2.2.9a. This procedure allowed colonies to be screened in duplicate.

The hybridisation probe used for screening was a mixture of the histone gene probes (16×10^6 cpm; representing H1, H_2A, H_2B, H3 and H4). All histone gene-specific fragments were purified by polyacrylamide gel electrophoresis, eluted and labelled as a mixture

by nick-translation (section 2.2.11b).

Figure 6.5 shows an autoradiograph of a selection of the nitrocellulose filters and gives the details of the hybridisation and washing conditions used. Colonies showing hybridisation to the probe with varying degrees of intensity can be seen in the autoradiogram.

Figure 6.6 shows the result of hybridising the mixed histone gene probe with a filter onto which DNA representing the five histone genes and pBR322 had been spotted. This result indicated clearly that the probe used for screening was detecting each histone gene sequence and did not contain labelled vector DNA.

From the filters screened (50,000 colonies) five strong duplicate positives were obtained. These were picked into L-broth and ampicillin and replated on small ampicillin plates at a density low enough to allow single colony isolation after rescreening.

6.2.8 Preliminary Characterisation of Cosmid Recombinants

(a) Digestion Pattern

DNA was prepared from 5 ml overnight cultures of each of the purified recombinants using the small scale plasmid DNA isolation method described in section 2.2.9b. DNA from each clone was digested with *Eco*RI and *Xho*I and analysed by electrophoresis on a 1% agarose gel (Fig. 6.7).

Addition of approximate fragment sizes indicated that all recombinants contained large inserts (greater than 40 kb). The digestion pattern of two of the recombinants, 2.1C and 5.1C, was very similar (Fig. 6.7) indicating that the inserts for these two clones were probably generated from the same region of the genome. Further analysis of these clones is reported in section 6.2.15. It was not possible to say from this analysis whether any of the other recombinants overlapped.

(b) Gene Content

To determine the histone gene content of each clone, 500 ng

Figure 6.5

High Density Screening of a Cosmid Library

A cosmid library (approximately 50,000 colonies) was prepared in *E. coli* HB101 using the plasmid vector pHC79 and size fractionated *Sau*3AI treated chicken DNA (section 2.2.6). Colony filters were prepared and screened in duplicate (section 2.2.9a) by hybridisation with a ^{32}P-labelled mixed histone gene probe (see text). After washing in 1xSSC/0.1% SDS at 65^{o}C filters were autoradiographed at -80^{o}C overnight with an intensifying screen.

The figure shows autoradiographs of two colony filters screened in duplicate. Positive signals, present in duplicate, are circled. Orientation marks are indicated by the arrows.

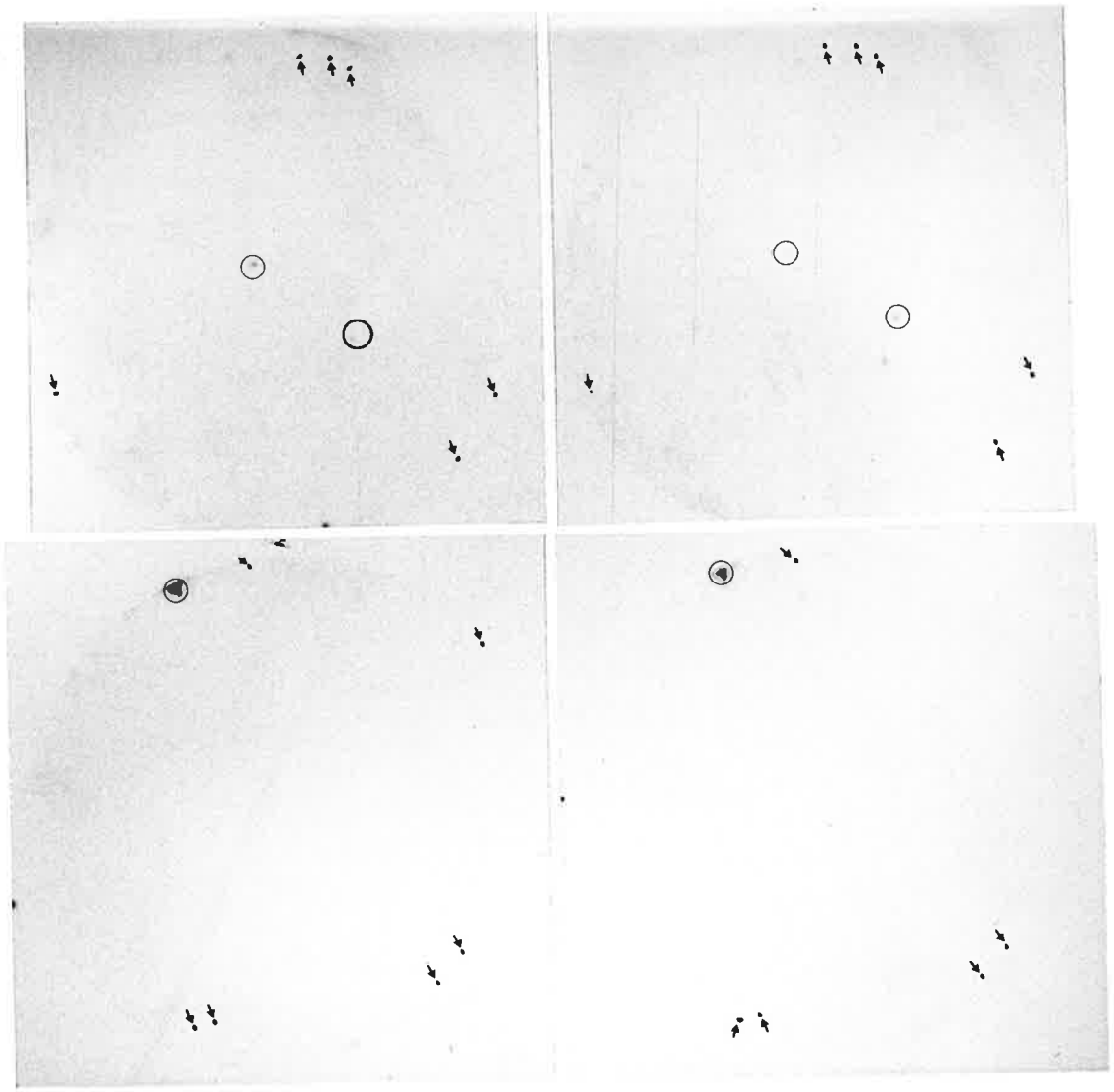


Figure 6.6

Dot-Blot Analysis Using Mixed Histone Probe

A dot-blot filter was prepared with duplicate 1 ug samples of each histone gene-specific plasmid (Chapter 3) and pBR322 DNA as a control. Filter bound DNA was hybridised with the ^{32}P-labelled mixed histone gene probe used to screen the chicken cosmid genomic library (see text), washed and autoradiographed overnight (see Fig. 6.5).

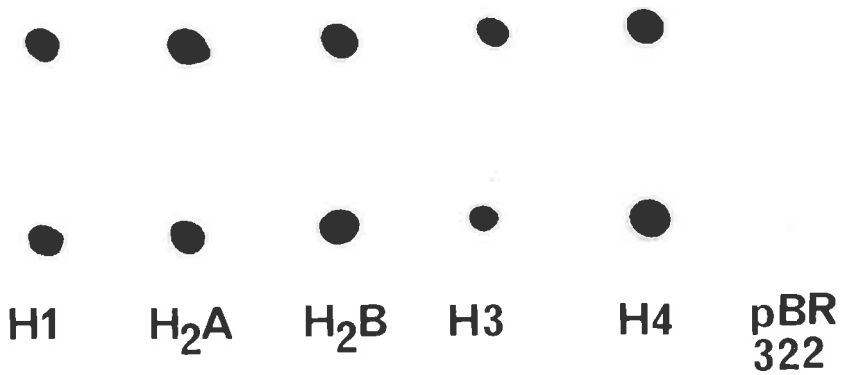

H1
H2A
H2B
H3
H4
pBR
322

Figure 6.7

Miniscreen Analysis of Cosmid Recombinants

Colonies giving rise to positive responses with histone gene specific probe on colony screening (Fig. 6.5) were picked from master plates and purified. Plasmid DNA was isolated from 5 ml cultures using the "mini-prep" method (section 2.2.9b). An aliquot of each was digested with *Eco*RI (A) or *Xho*I (B) and electrophoresed on a 1% agarose gel (section 2.2.2a). DNA was viewed under UV light after ethidium bromide staining. Reference sizes are indicated at left, in kilobases, and correspond to marker DNA fragments.

A.
2·1C
5·1C
4·2C
6·1C
6·3C
9·0
4·2
2·2
B.
2·1C
5·1C
4·2C
6·1C
6·3C

of recombinant cosmid DNA was spotted onto nitrocellulose filters (section 2.2.3), in duplicate, and probed separately with each individual histone gene probe. 100 ng of the λCHO1 derived histone containing subclones, pCH3.3E and pCH3.75EH (Harvey *et al*, 1981; Fig. 3.1), and pBR322 DNA were also spotted onto these filters as positive and negative controls respectively. The results of this analysis are shown in Figure 6.8 (A-E).

6.2.9 Probing for Cosmid Recombinants Overlapping λCHO3 & λCHO7

To determine whether any of these clones extended the region bracketed by λCHO3 and λCHO7 (see Fig. 5.32) non-coding fragments were isolated from the λCHO3 subclone pCH5.0E (2.5 kb *Pst*I-*Pst*I fragment; Fig 5.8) and from λCHO7 (0.85 kb *Bam*HI-*Eco*RI fragment; Fig. 5.25), labelled with ^{32}P by nick-translation (section 2.2.11b) and used separately to probe filters carrying DNA from each of the five cosmid recombinants (see Fig. 6.9).

Clone 6.3C hybridised to the 0.85 kb *Bam*HI-*Eco*RI probe from λCHO7 and possibly represents an overlapping region of the genome. None of the cosmid recombinants were positive with the non-coding probe specific for λCHO3. Cosmid 6.3C has been fully characterised and is discussed in the next section.

6.2.10 Restriction Analysis of Cosmid 6.3C

DNA was prepared from recombinant 6.3C using the large scale plasmid preparation procedure described in section 2.2.10b, and digested with *Eco*RI, *Xho*I, *Bam*HI, *Sal*I and *Hin*dIII, in single and double digests. Resultant fragments were fractionated on agarose gels and sized by comparison with marker DNA (Figs. 6.10, 6.11). Fragment sizes and a complete map indicating their relative positions are shown in Figure 6.12.

The presence of the 0.7 kb *Eco*RI-*Eco*RI fragments and the 0.7 kb *Xho*I-*Xho*I fragment is indicative of the region spanned by λCHO1, λCHO5 and λCHO7 (Chapter 5) and confirms that the insert in cosmid

Figure 6.8

Dot-Blot Analysis of Cosmid Recombinants With Histone Gene Probes

Dot-blot filters were prepared with duplicate 500 ng samples of DNA from each cosmid recombinant (section 2.2.3). 100 ng samples of pCH3.3E, pCH3.75EH (see Fig. 3.1) and pBR322 were also included as positive and negative controls. To determine the histone gene content of each recombinant, individual filters (A-E) were hybridised with different ^{32}P-labelled histone gene specific probes. At a later stage a filter (F) was hybridised with a radio-labelled non-histone encoding DNA fragment isolated from cosmid 6.3C (see section 7.2.1). Filters were washed in 0.5xSSC/0.1% SDS at 65^{o}C and autoradiographed overnight at -80^{o}C with an intensifying screen. Overall results are summarised in the table below.

Probes

A	H_2A
B	H_2B
C	H1
D	H3
E	H4
F	2.0 kb *Bam*HI-*Eco*RI fragment from 6.3C.

	Recombinant				
	2.1C	4.2C	5.1C	6.1C	6.3C
H_2A	-	+	-	+	+
H_2B	-	-	-	+	+
H1	-	-	-	+	+
H3	+	+	+	+	+
H4	-	-	-	+	(+)

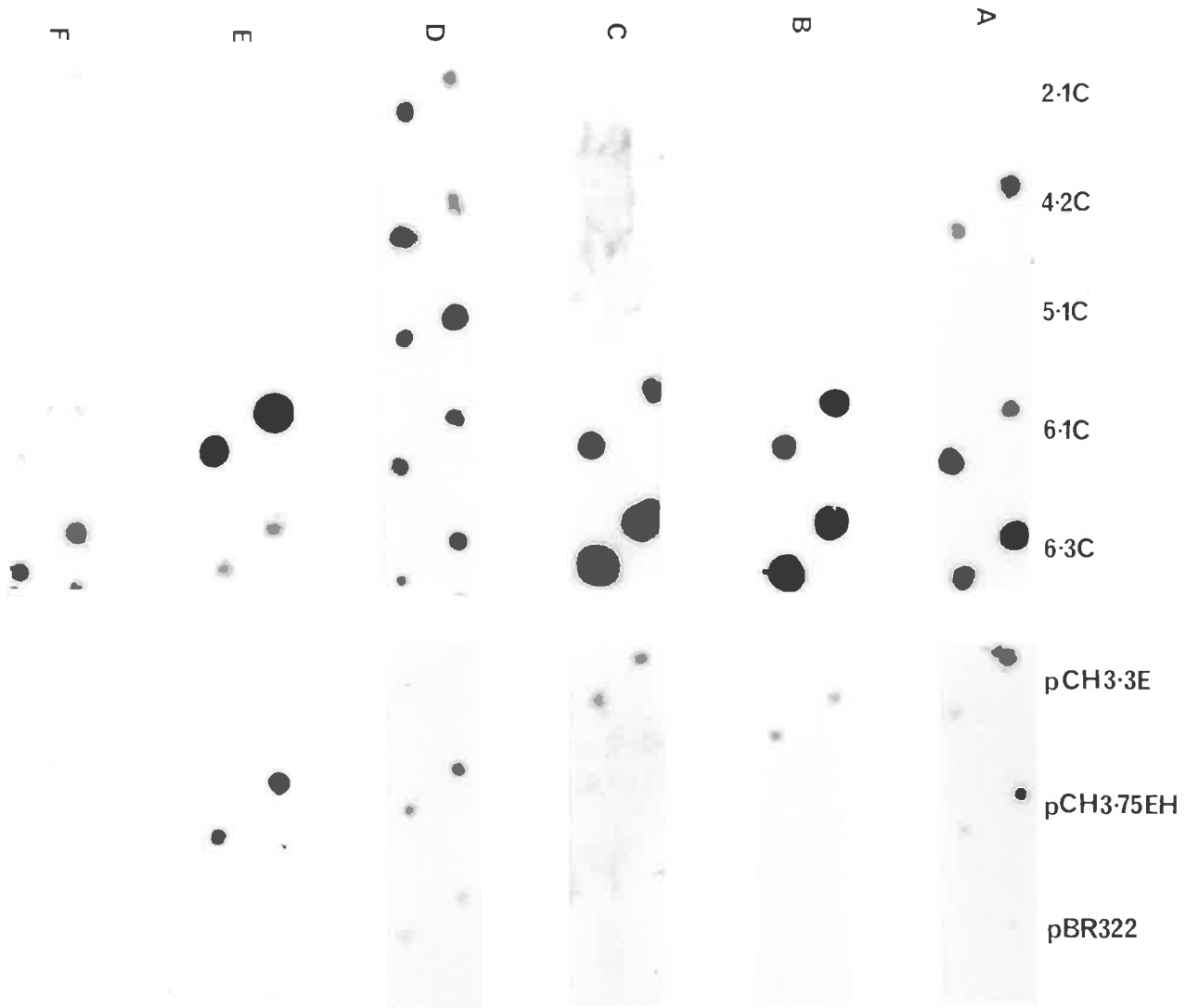

F
E
D
C
B
A
2·1C
4·2C
5·1C
6·1C
6·3C
pCH3·3E
pCH3·75EH
pBR322

Figure 6.9

Searching for Overlapping Cosmid Recombinants

Dot-blot filters were prepared (section 2.2.3) with duplicate 500 ng samples of DNA from each cosmid recombinant. 100 ng samples of pCH5.0E (derived from λCHO3), pCH4.9E (derived from λCHO7) and pBR322 were included as positive and negative controls. To determine whether any of the histone gene containing cosmid recombinants overlapped with the region bordered by λCHO3 and λCHO7 (see Fig. 5.32), individual dot-blot filters were hybridised with separate non-histone encoding ^{32}P-labelled DNA fragments from λCHO3 (2.5 kb *Pst*I - *Pst*I fragment in pCH5.0E, see Fig. 5.8) and λCHO7 (0.85 kb *Bam*HI-*Eco*RI fragment, see Fig. 5.25). Filters were washed at 65^{o}C in 0.5xSSC/0.1% SDS and exposed to X-ray film overnight at -80^{o}C with an intensifying screen.

A : Filter probed with ^{32}P-labelled probe from λCHO3

B : Filter probed with ^{32}P-labelled probe from λCHO7

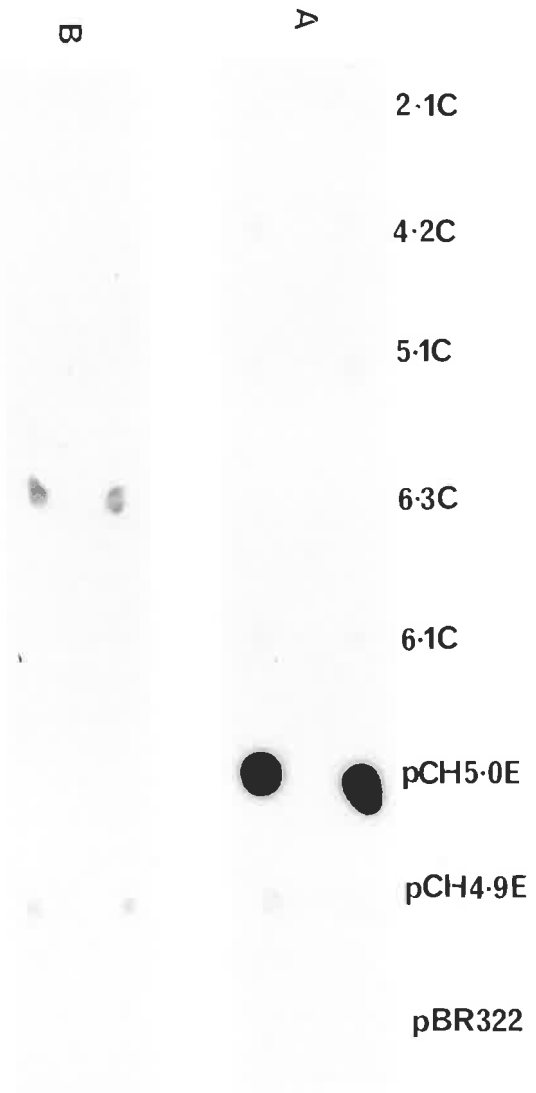

B
A
2·1C
4·2C
5·1C
6·3C
6·1C
pCH5·0E
pCH4·9E
pBR322

Figure 6.10

Restriction Analysis of Cosmid Recombinant 6.3C

1 ug samples of 6.3C DNA were digested with various combinations of restriction enzymes (see below) and electrophoresed on a 1% agarose gel (section 2.2.2a). DNA was detected with UV light after ethidium bromide staining. Fragment sizes are indicated in kilobases. The sizes of fragments obtained on restriction enzyme cleavage of DNA from cosmid recombinants, and their derived subclones, were determined using *Hin*dIII digested λDNA and *Hin*fI digested pBR322 DNA as molecular weight markers. The figure shows 6.3C digested with the following enzymes.

Track A : *Eco*RI

B : *Eco*RI and *Xho*I

C : *Xho*I

D : *Hin*dIII

E : *Hin*dIII and *Xho*I

F : *Sal*I

G : *Sal*I and *Xho*I

H : *Eco*RI

I : *Eco*RI and *Sal*I

J : *Sal*I

K : *Bam*HI

L : *Bam*HI and *Sal*I

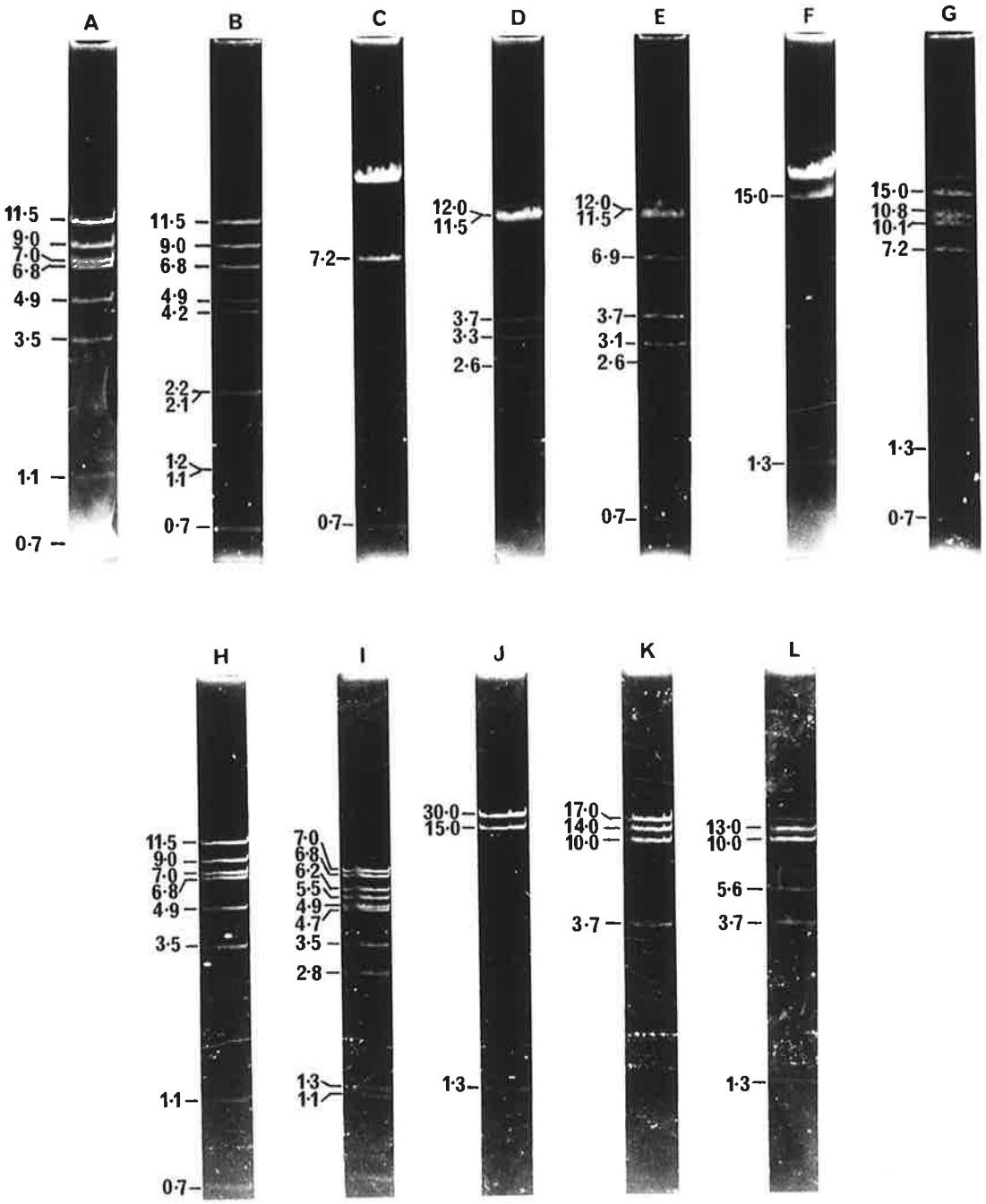

A
11·5
9·0
7·0
6·8
4·9
3·5
1·1
0·7
B
11·5
9·0
6·8
4·9
4·2
2·2
2·1
1·2
1·1
0·7
C
7·2
0·7
D
12·0
11·5
3·7
3·3
2·6
E
12·0
11·5
6·9
3·7
3·1
2·6
0·7
F
15·0
1·3
G
15·0
10·8
10·1
7·2
1·3
0·7
H
11·5
9·0
7·0
6·8
4·9
3·5
1·1
0·7
I
7·0
6·8
6·2
5·5
4·9
4·7
3·5
2·8
1·3
1·1
J
30·0
15·0
1·3
K
17·0
14·0
10·0
3·7
L
13·0
10·0
5·6
3·7
1·3

Figure 6.11

Restriction Analysis of Cosmid Recombinant 6.3C

1 ug samples of 6.3C DNA were digested with the enzymes indicated below and analysed by electrophoresis on 1% (tracks A–D) or 1.2% (tracks E and F) agarose gels (see Fig. 6.10).

The figure shows the following restriction enzyme digests of 6.3C DNA;

Track A : *Eco*RI

B : *Hin*dIII

C : *Eco*RI and *Hin*dIII

D : *Hin*dIII and *Bam*HI

E : *Eco*RI

F : *Eco*RI and *Bam*HI

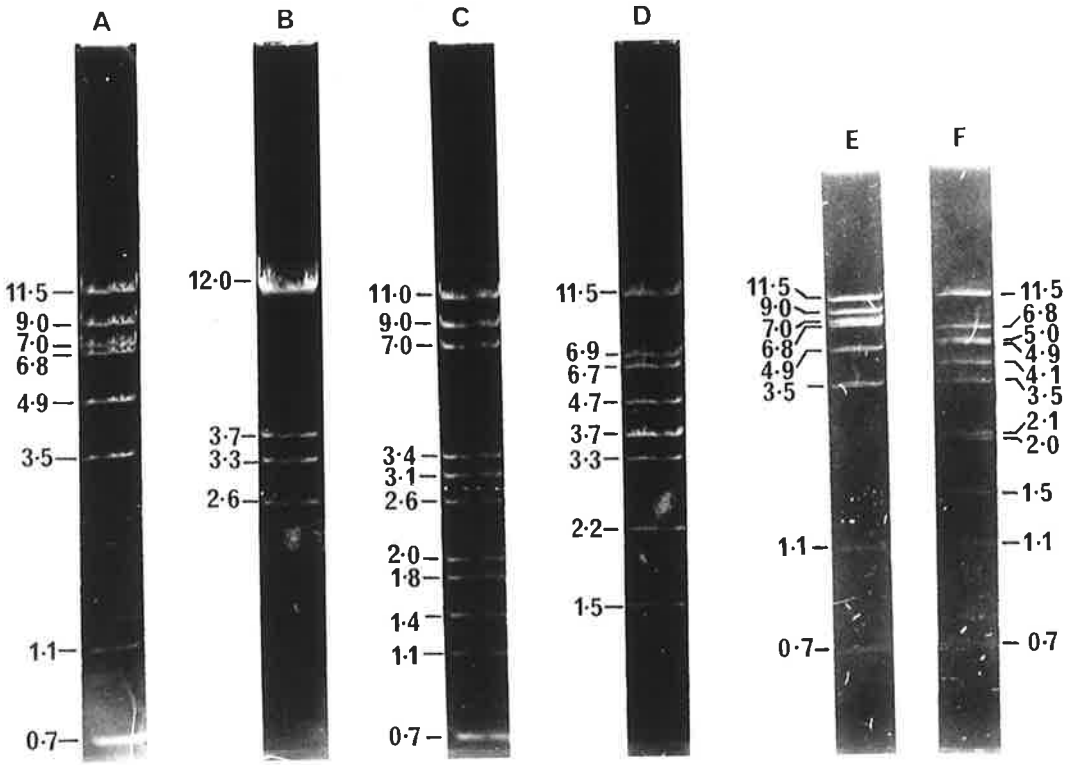

A
B
C
D
E
F
11·5
9·0
7·0
6·8
4·9
3·5
1·1
0·7
12·0
3·7
3·3
2·6
11·0
9·0
7·0
3·4
3·1
2·6
2·0
1·8
1·4
1·1
0·7
11·5
6·9
6·7
4·7
3·7
3·3
2·2
1·5
11·5
9·0
7·0
6·8
4·9
3·5
1·1
0·7
11·5
6·8
5·0
4·9
4·1
3·5
2·1
2·0
1·5
1·1
0·7

Figure 6.12

Restriction Maps of Cosmid Recombinant 6.3C

The sizes, in kilobases, of the restriction fragments generated on cleavage of 6.3C DNA with various restriction enzymes are shown. An overall derived restriction map of 6.3C, showing predicted gene positions (as determined from Southern analysis) is also presented. Vector DNA is represented by a bold line. Restriction sites are indicated as follows;

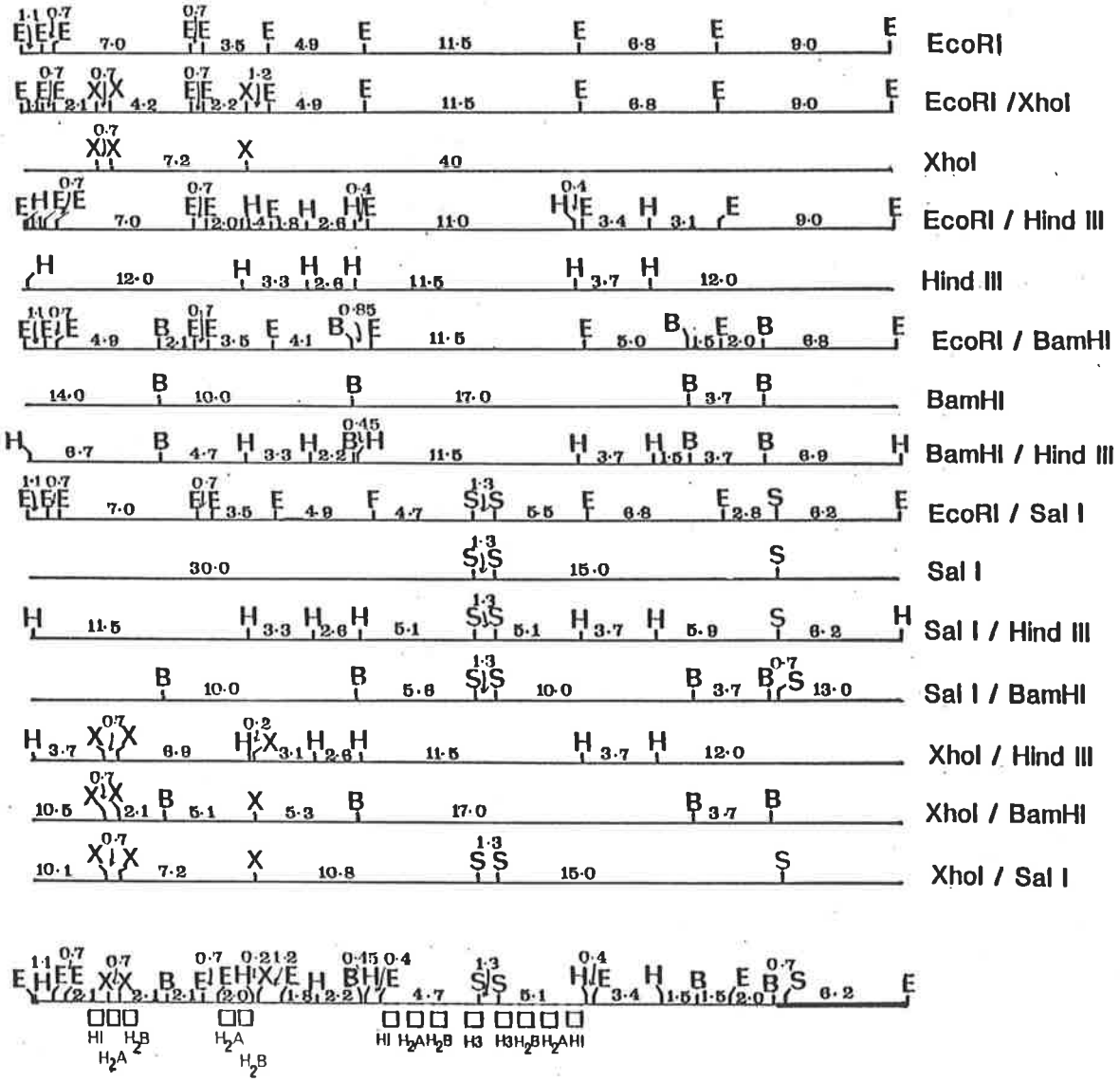
E : *Eco*RI

H : *Hin*dIII

S : *Sal*I

B : *Bam*HI

X : *Xho*I


EcoRI
EcoRI /Xhol
Xhol
EcoRI / Hind III
Hind III
EcoRI / BamHI
BamHI
BamHI / Hind III
EcoRI / Sal I
Sal I
Sal I / Hind III
Sal I / BamHI
Xhol / Hind III
Xhol / BamHI
Xhol / Sal I

6.3C was generated from this region.

The 1.9 kb, H_2B containing, *Xho*I-*Xho*I fragment observed in λCHO5 and λCHO7 is not present in cosmid 6.3C (Fig. 6.10c). In this clone this H_2B gene is present in a 4.2 kb *Eco*RI-*Xho*I fragment derived from the 7.0 kb *Eco*RI - *Eco*RI fragment (Fig. 6.13b). The *Xho*I site located next to the *Bam*HI site in λCHO5 and λCHO7 (Figs. 5.16 and 5.25) is not present in this clone. Except for this difference, which presumably represents a polymorphic restriction site difference between individual chickens, the restriction maps for λCHO1, λCHO5, λCHO7 and 6.3C agree in overlapping regions.

6.2.11 The Coding Potential of 6.3C

DNA from agarose gels was transferred to nitrocellulose filters using the bidirectional transfer procedure described (section 2.2.3) and hybridised with individual histone gene-specific fragments to determine gene locations.

Hybridisation of digested 6.3C DNA to H4 specific probe gave a negative result for all digestions and indicated that the faint cross-reaction of 6.3C with H4 probe, observed in Figure 6.8 was artefactual.

H_2A probe is specific for three regions of this clone viz; The 0.7 kb *Xho*I - *Xho*I fragment (Fig. 6.12b) which is contained within the 7.0 kb *Eco*RI-*Eco*RI fragment and has been discussed in Chapter 5. The 2.2 kb *Eco*RI-*Xho*I fragment (Fig. 6.13b) and the 2.0 kb *Eco*RI-*Hin*dIII fragment (Fig. 6.13a), both located within the 3.5 kb *Eco*RI-*Eco*RI fragment and also found in λCHO7 (section 5.2.11). The 11.5 kb *Eco*RI-*Eco*RI fragment (Fig. 6.13b) which on digestion with *Hin*dIII generates an 11.0 kb *Eco*RI-*Hin*dIII fragment (Fig. 6.13a).

H_2B sequences are evident in the 4.2 kb *Eco*RI-*Xho*I fragment (discussed above), the 11.5 kb *Eco*RI-*Eco*RI fragment and the 2.2 kb

Figure 6.13

Southern Analysis of 6.3C

6.3C DNA which had been restriction enzyme digested and fractionated on agarose gels was transferred to nitrocellulose filters using the bidirectional transfer procedure described in section 2.2.3. Filter bound DNA was hybridised to ^{32}P-labelled histone gene-specific probes, washed in 0.5xSSC/0.1% SDS at 65^{o} and autoradiographed, overnight, at -80^{o}C with an intensifying screen.

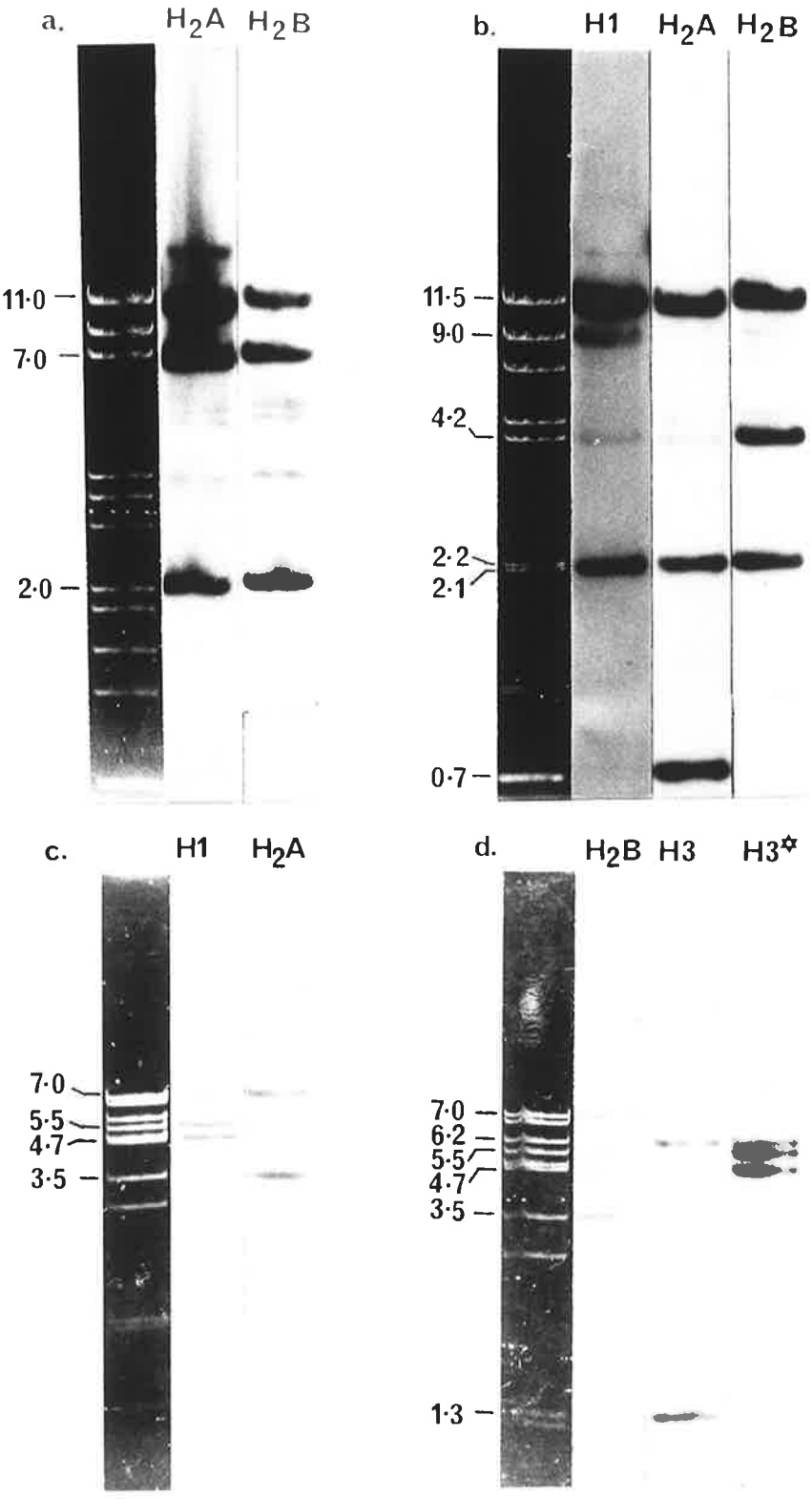
The following restriction enzyme digests of 6.3C are shown (restriction fragment sizes are indicated in kilobases);

a : *Eco*RI and *Hin*dIII

b : *Eco*RI and *Xho*I

c & d : *Eco*RI and *Sal*I

Note: An *Eco*RI and *Sal*I digest (d) was also hybridised to ^{32}P-labelled 0.85 kb *Sal*I-*Pst*I fragment, (H3*), derived from pCH3.75EH (see Fig. 3.1). This probe is specific for the 5' region of the H3 gene (see text).


a.
H2A
H2B
11·0
7·0
2·0
b.
H1
H2A
H2B
11·5
9·0
4·2
2·2
2·1
0·7
c.
H1
H2A
7·0
5·5
4·7
3·5
d.
H2B
H3
H3*
7·0
6·2
5·5
4·7
3·5
1·3

*Eco*RI-*Xho*I fragment (Fig. 6.13b).

After codigestion with *Eco*RI and *Hin*dIII the 11.0 kb and 2.0 kb *Eco*RI-*Hin*dIII fragments and the 7.0 kb *Eco*RI-*Eco*RI fragment hybridise with H_2A specific probe (Fig. 6.13a).

H1 probe cross-reacts with the 2.1 kb *Eco*RI-*Xho*I fragment (Fig. 6.13b) found in λCHO5 (Fig. 5.16) and λCHO1 and the 11.5 kb *Eco*RI-*Eco*RI fragment (Fig. 6.13b).

H3 sequences are contained only within the 11.5 kb *Eco*RI-*Eco*RI fragment. Digestion of 6.3C with *Sal*I results in cleavage at two sites within the insert (another *Sal*I restriction site is located in pHC79), both located in the large *Eco*RI-*Eco*RI fragment. *Sal*I cleaves this fragment and generates 5.5 kb *Eco*RI-*Sal*I, 1.3 kb *Sal*I-*Sal*I and 4.7 kb *Eco*RI-*Sal*I fragments. The 5.5 kb *Eco*RI-*Sal*I and 4.7 kb *Eco*RI-*Sal*I fragments are both positive with H_2A, H_2B and H1 gene probes (Figs. 6.13c and 6.13d) and the 1.3 kb *Sal*I-*Sal*I internal fragment is positive with the 0.15 kb *Eco*RI-*Sal*I H3 probe (Fig. 6.13d). This probe is specific for the H3 coding region 3' to the conserved *Sal*I site. The observation that there are two *Sal*I sites suggests the possibility that there are two H3 genes each spanning a conserved *Sal*I restriction site (among chicken histone gene recombinants *Sal*I restriction sites so far have only been observed in H3 coding regions). To investigate this, a probe from the other side of the conserved *Sal*I site representing sequences 5' to the H3 gene was isolated (0.85 kb *Pst*I-*Sal*I fragment from pCH3.75EH; Harvey *et al*, 1981; Fig. 3.1) and used to probe an *Eco*RI/*Sal*I digest of 6.3C (Fig. 6.13d). This probe hybridised to both the 5.5 kb *Eco*RI-*Sal*I fragment and the 4.5 kb *Eco*RI-*Sal*I fragments of 6.3C and indicated that this 11.5 kb *Eco*RI-*Eco*RI fragment contains two central, convergently transcribed, H3 genes with at least one copy of the H_2A, H_2B and H1 genes each side.

6.2.12 Construction of 6.3C Subclones

Three *Eco*RI-*Eco*RI fragments (11.5 kb, 9.0 kb and 6.8 kb) from cosmid 6.3C were subcloned into the plasmid vector pBR325 (section 2.2.7) and recombinants were identified by miniscreening colonies which were $cap^{s}tet^{r}$ (Fig. 6.14). Large scale plasmid DNA preparations (section 2.2.10) were performed on colonies containing appropriate inserts (referred to as pCH11.5E, pCH9.0E and pCH6.8E).

6.2.13 Restriction Analysis of pCH11.5E

pCH11.5E contains the 11.5 kb *Eco*RI-*Eco*RI fragment of 6.3C subcloned into the *Eco*RI site of pBR325. From previous mapping data (section 6.2.11) this fragment is known to contain at least two copies each of the genes for H1, H_2A, H_2B and H3.

This subclone was digested in single or double digestions with restriction enzymes *Sac*I, *Sac*II, *Hin*dIII, *Sma*I and *Sal*I (Figs. 6.15 and 6.16) and after Southern transfer, hybridised with H_2A, H_2B, H3 and H1 specific probes. The results of this allowed the construction of a restriction map and the location of the above genes (see Fig. 6.17).

From the available mapping data for this clone it was not possible to locate the internal 0.6 kb *Sac*I-*Sac*I fragment (left or right hand end of the insert, as drawn in Figure 6.17). However, when the 5.5 kb *Sal*I-*Eco*RI fragment is redigested with *Sac*I using the procedure described in section 2.2.1 the 0.6 kb *Sac*I-*Sac*I fragment is observed (Fig. 6.16b) and therefore is located within this fragment.

H1 Coding Domain

H1 coding sequences are located within the internal 1.2 kb *Sac*I-*Sac*II fragment and one of the 1.3 kb *Sac*I-*Sac*II fragments (Fig. 6.18a). The 0.80 kb *Sac*II-*Eco*RI fragment is also positive with H1 probe (Fig. 6.18d) suggesting an H1 gene spans this restriction site as shown in Figure 6.17. There are no detectable H1 specific sequences in the

Figure 6.14

"Miniscreen" Examination of 6.3C Subclones

The 11.5 kb, 9.0 kb and 6.5 kb *Eco*RI-*Eco*RI fragments from 6.3C were subcloned into *Eco*RI cleaved pBR325 using the procedures described in Chapter 2. DNA was isolated from cap^s tet^r colonies using the "miniprep" method (section 2.2.9b). An aliquot of DNA from each was digested with *Eco*RI, electrophoresed on a 1% agarose gel and viewed under UV light after ethidium bromide staining.

Track a: *Hin*dIII digested λ DNA

Tracks b-f: *Eco*RI digested DNA from transformants derived from the ligation mix containing the 6.5 kb *Eco*RI-*Eco*RI fragment.

Tracks g & h: *Eco*RI digested DNA from transformants derived from the ligation mix containing the 9.0 kb *Eco*RI-*Eco*RI fragment.

Tracks i-l: *Eco*RI digested DNA from transformants derived from the ligation mix containing the 11.5 kb *Eco*RI-*Eco*RI fragment.

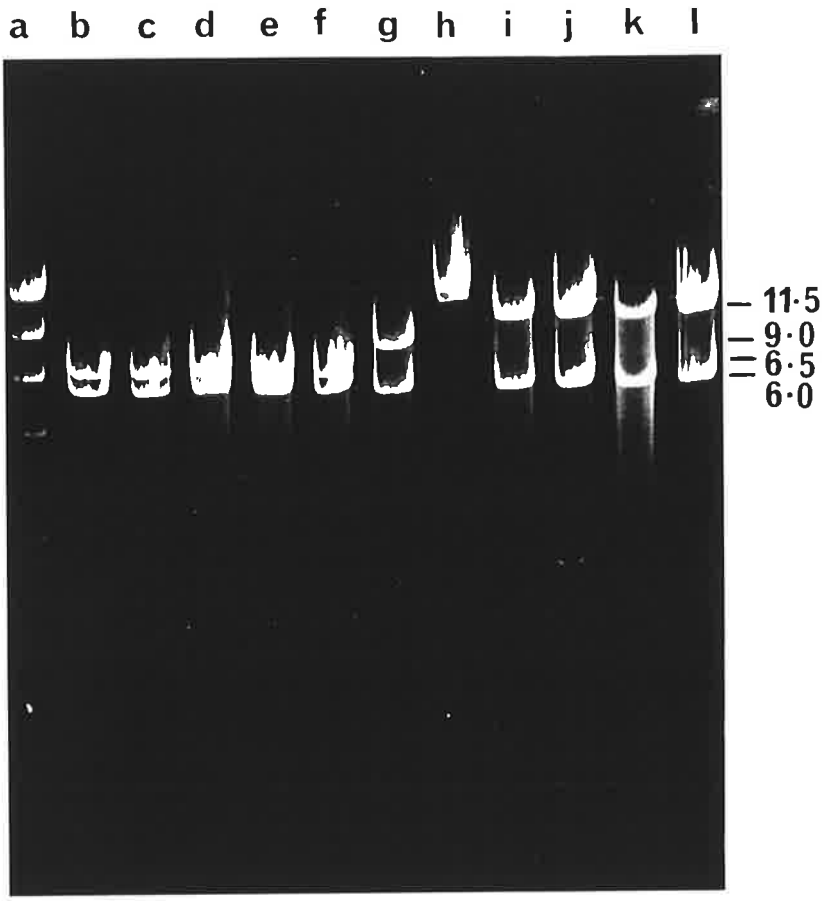

a b c d e f g h i j k l
11·5
9·0
6·5
6·0

Figure 6.15

Restriction Analysis of pCH11.5E

1 ug samples of pCH11.5E DNA were digested with the restriction enzymes indicated below and the resultant fragments resolved on a 1.2% agarose gel (section 2.2.2a). DNA was detected with UV light after ethidium bromide staining. Fragment sizes are given in kilobases. The following restriction enzyme digests are shown.

Track A: *Sac*II

Track B: *Sac*II and *Hin*dIII

Track C: *Sac*II and *Eco*RI

Track D: *Sac*II and *Sac*I

Track E: *Sac*I and *Sma*I

Track F: *Sac*I and *Hin*dIII

Track G: *Sac*I and *Eco*RI

Track H: *Sac*I

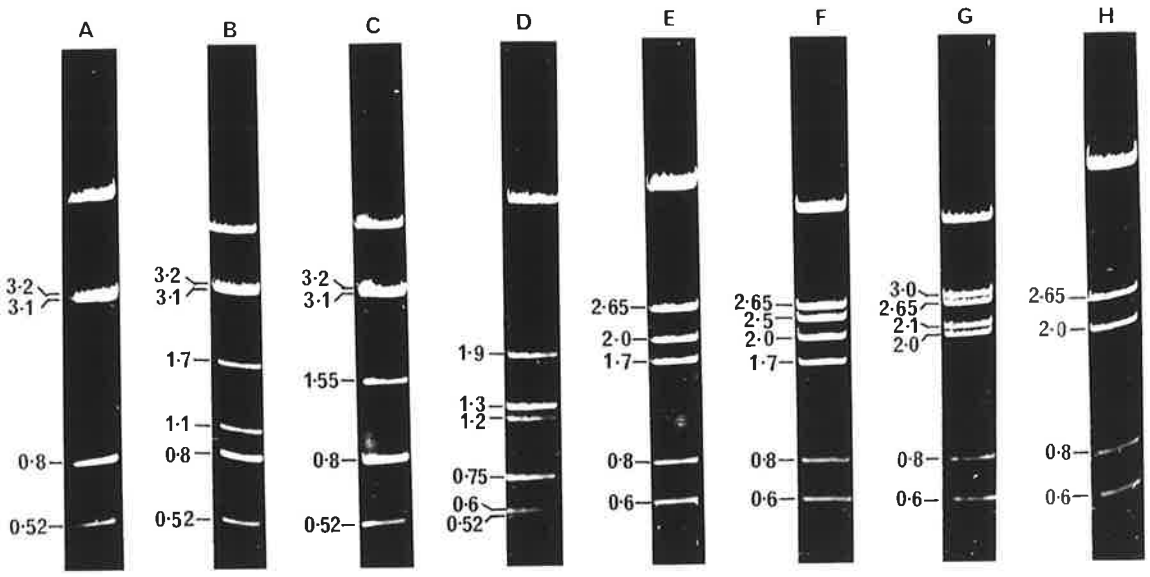

A
3·2
3·1
0·8
0·52
B
3·2
3·1
1·7
1·1
0·8
0·52
C
3·2
3·1
1·55
0·8
0·52
D
1·9
1·3
1·2
0·75
0·6
0·52
E
2·65
2·0
1·7
0·8
0·6
F
2·65
2·5
2·0
1·7
0·8
0·6
G
3·0
2·65
2·1
2·0
0·8
0·6
H
2·65
2·0
0·8
0·6

Figure 6.16

Restriction Analysis of pCH11.5E

(a) 1 ug samples of pCH11.5E DNA were digested with appropriate restriction enzymes (see below) and fractionated on a 5% polyacrylamide gel (section 2.2.2b). DNA was visualised under UV light after ethidium bromide staining. *Hin*fI cleaved pBR322 DNA was co-electrophoresed to provide molecular weight standards. Fragment sizes are indicated in kilobases alongside each track.

The following restriction enzyme digests are shown;

Track 1: *Eco*RI and *Hin*dIII

Track 2: *Sma*I and *Hin*dIII

Track 3: *Sac*II

Track 4: *Sac*II and *Sal*I

Track 5: *Sac*I

Track 6: *Sac*I and *Sal*I

Track 7: *Sac*I and *Sac*II

(b) 10 ug of pCH11.5E DNA were digested with *Eco*RI and *Sal*I and fractionated on a 1.2% LGT agarose gel (section 2.2.2a). A slice containing the 5.5 kb *Eco*RI-*Sal*I fragment was excised from the gel and melted. The DNA was redigested with *Sac*I using the procedure detailed in section 2.2.1, electrophoresed on a 1.5% agarose gel and visualised with UV light after ethidium bromide staining.

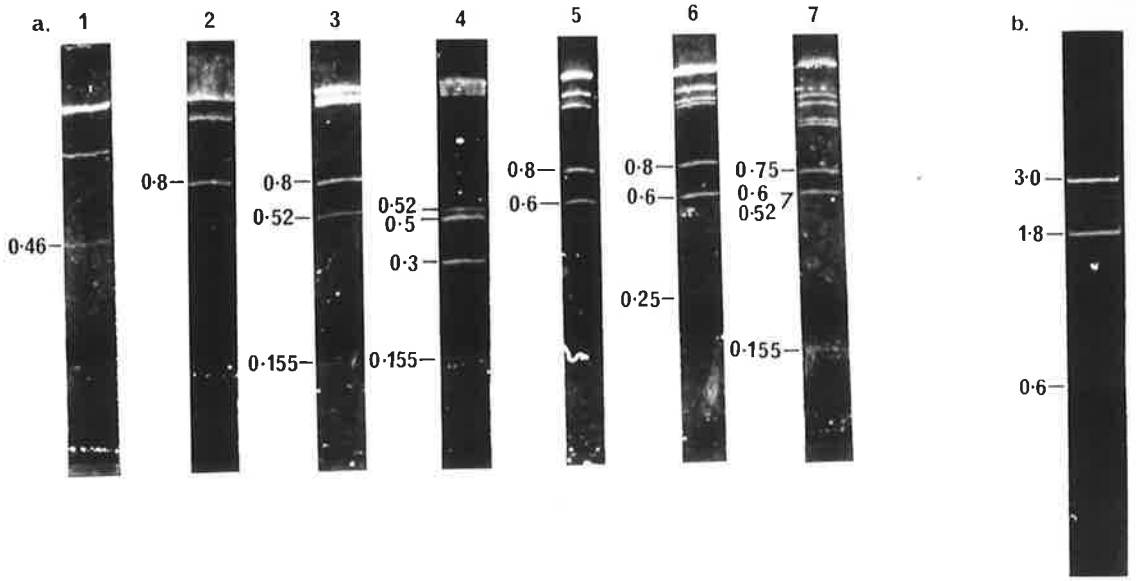

a.
1
2
3
4
5
6
7
0·46
0·8
0·8
0·52
0·155
0·52
0·5
0·3
0·155
0·8
0·6
0·8
0·6
0·25
0·75
0·6
0·52
7
0·155
b.
3·0
1·8
0·6

Figure 6.17

Restriction Maps of pCH11.5E

The figure indicates the sizes, in kilobases, of the fragments obtained after restriction enzyme cleavage of pCH11.5E. The overall derived restriction map is also shown indicating predicted gene positions as determined from Southern analysis. Restriction enzyme sites are indicated as follows;

E : *Eco*RI

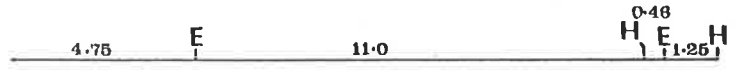
H : *Hin*dIII

S : *Sal*I

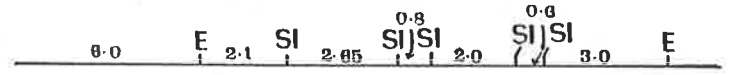
SI: *Sac*I

SII: *Sac*II

Sm: *Sma*I

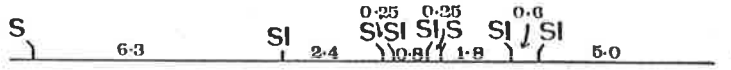


EcoRI/ Hind III

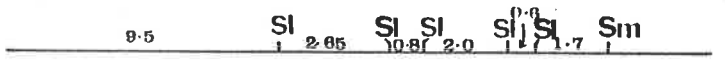


Sac I/ EcoRI

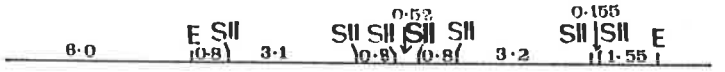
Sac I/Hind III



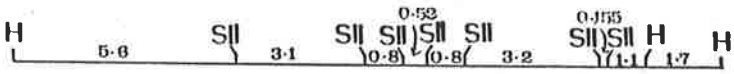
Sac I/Sal I



Sac I/Sma I

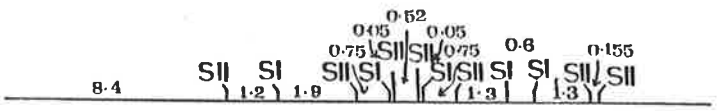


Sac II/EcoRI

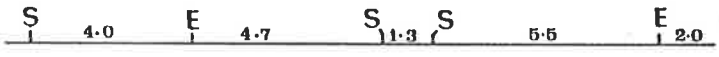


Sac II/Hind III

Sac II/Sal I

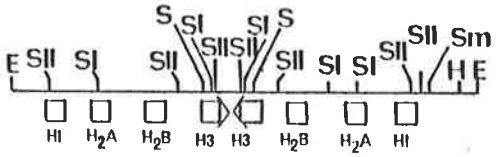


Sac I/Sac II



Sal I/EcoRI

Sma I/Hind III



pCH 11·5E

insert

Figure 6.18

Southern Analysis of pCH11.5E

Restriction enzyme cleaved pCH11.5E DNA which had been fractionated on 1% agarose gels was transferred, bidirectionally, to nitrocellulose filters (section 2.2.3) and analysed for gene content by hybridisation to ^{32}P-labelled histone gene specific probes. Filters were washed in 0.5xSSC/0.1% SDS at 65^{o}C and autoradiographed for 2-6 hours at -80^{o}C with an intensifying screen. Digests discussed in the text are shown. Between hybridisation experiments the filters were boiled in distilled water to remove annealed label. Fragment sizes are indicated in kilobases.

a: pCH11.5E digested with *Sac*I and *Sac*II

b: pCH11.5E digested with *Sac*I and *Eco*RI

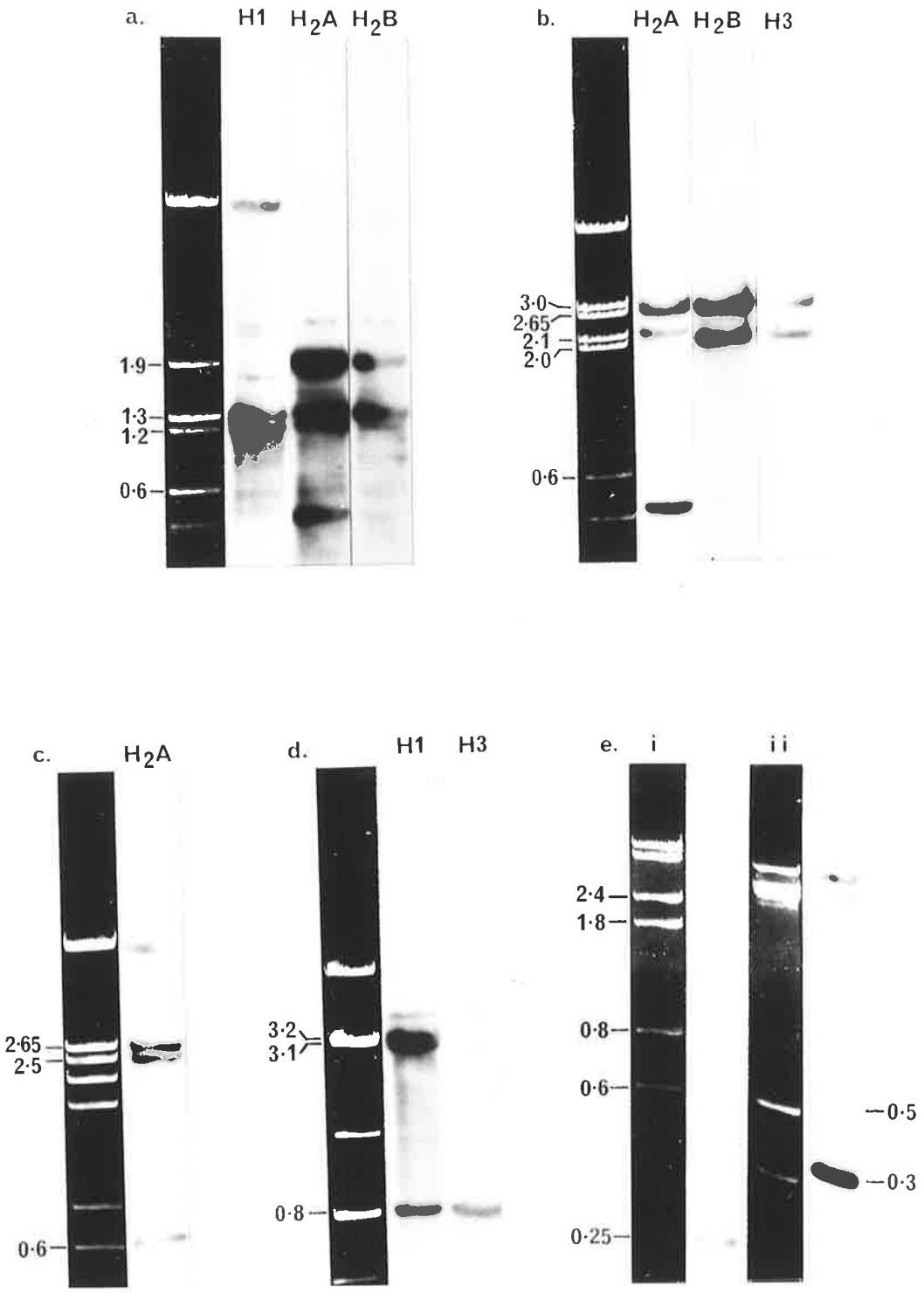
c: pCH11.5E digested with *Sac*I and *Hin*dIII

d: pCH11.5E digested with *Sac*II and *Eco*RI

e: pCH11.5E digested with

i) *Sac*I and *Sal*I and

ii) *Sac*II and *Sal*I and probed with H3 gene specific probe.


a.
H1
H2A
H2B
1·9
1·3
1·2
0·6
b.
H2A
H2B
H3
3·0
2·65
2·1
2·0
0·6
c.
H2A
2·65
2·5
0·6
d.
H1
H3
3·2
3·1
0·8
e.
i
ii
2·4
1·8
0·8
0·6
0·25
0·5
0·3

1.55 kb *Sac*II-*Eco*RI fragment located at the other end of the insert but sequencing has shown that the coding region of the H1 gene begins within the 0.155 kb *Sac*II-*Sac*II fragment (L.S. Coles, this laboratory).

H_2A Coding Domain

An H_2A gene spans the *Sac*I site at the far left of the insert such that H_2A sequences are located within the 2.1 kb *Sac*I-*Eco*RI fragment and the internal 2.65 kb *Sac*I-*Sac*I fragment (Fig. 6.18b). Another H_2A gene begins within the 0.6 kb *Sac*I-*Sac*I fragment and spans a *Sac*I restriction site such that both the 3.0 kb *Sac*I-*Eco*RI fragment and the internal 0.6 kb *Sac*I-*Sac*I fragment are positive with H_2A probe. It is difficult to observe the cross-reaction to the 3.0 kb *Sac*I-*Eco*RI fragment in Figure 6.18b but it is confirmed by the fact that the 2.5 kb *Sac*I-*Hin*dIII fragment from the same region is clearly positive (Fig. 6.18c).

H_2B Coding Domain

H_2B genes are located within the 1.9 kb internal *Sac*I-*Sac*II fragment which also contains H_2A sequences and the 1.3 kb internal *Sac*I-*Sac*II fragment (Fig. 6.18a) generated from the 2.0 kb *Sac*I-*Sac*I fragment (Fig. 6.18b).

H3 Coding Domain

H3 probe was specific for the two central 0.8 kb *Sac*II-*Sac*II fragments (Fig. 6.18d) and also the 2.65 kb and 2.0 kb internal *Sac*I-*Sac*I fragments (Fig. 6.18b). After *Sal*I/*Sac*I and *Sal*I/*Sac*II digestion H3 probe hybridised only to the 0.3 kb *Sal*I-*Sac*II and 0.25 kb *Sal*I-*Sac*I fragments (Fig. 6.18e). As discussed previously the H3 probe is specific for H3 coding region 3' to the conserved *Sal*I site. H3 genes, therefore, begin just outside the internal *Sal*I-*Sal*I fragment and are arranged covergently as shown in Fig. 6.17.

6.2.14 Partial Sequencing of Isolated Restriction Fragments

Using the M13 dideoxy sequencing technique detailed in section

Figure 6.19

DNA Sequence Determination

The 1.8 kb and 2.4 kb *Sal*I-*Sac*I fragments from pCH11.5E were subcloned into the M-13 vector, mp83, and single stranded 'phage isolated (section 2.2.12). Single stranded DNA was used as a template for chain-termination sequencing (section 2.2.12d) and the products were analysed on a 6% polyacrylamide DNA sequencing gel. The sequences shown here are from the central region of pCH11.5E running outwards from the central *Sal*I sites. The sequences are presented in Figure 6.20. Bands were detected by autoradiography.

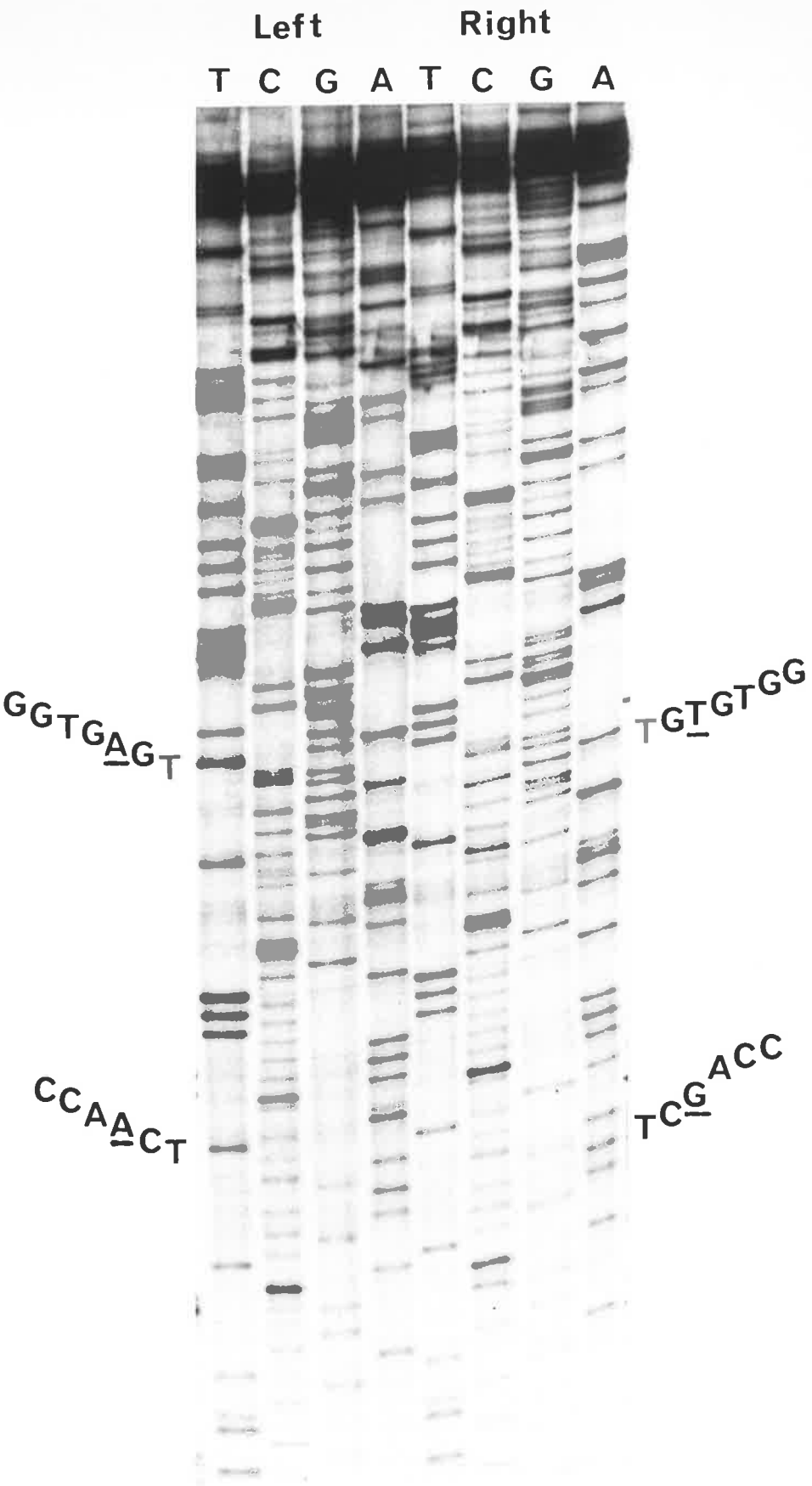
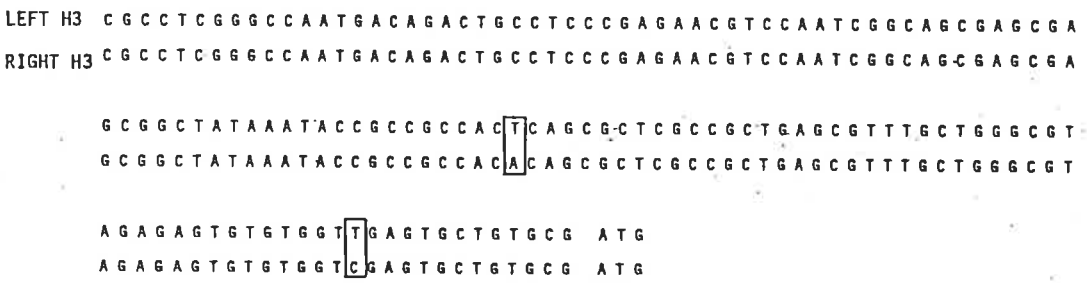

Left
Right
T C G A T C G A
GGTGAGT
TGTGTGG
CCAACT
TCGACC

Figure 6.20

DNA Sequences Obtained from pCH11.5E

The DNA sequence was determined for the 5' region of each central H3 gene in pCH11.5E (see Figure 6.19), and the sequences are shown 5'-3' reading towards the *Sal*I site. Differences are indicated by the boxes.


LEFT H3 CGCCTCGGGCCAATGACAGACTGCCTCCCGAGAACGTCCAATCGGCAGCGAGCGA
RIGHT H3 CGCCTCGGGCCAATGACAGACTGCCTCCCGAGAACGTCCAATCGGCAGCGAGCGA
GCGGCTATAAATACCGCCGCCACTCAGCGCTCGCCGCTGAGCGTTTGCTGGGCGT
GCGGCTATAAATACCGCCGCCACACAGCGCTCGCCGCTGAGCGTTTGCTGGGCGT
AGAGAGTGTGTGGTTGAGTGCTGTGCG ATG
AGAGAGTGTGTGGTCGAGTGCTGTGCG ATG

2.2.12, the 1.8 kb and 2.4 kb *Sal*I-*Sac*I fragments were sequenced from the *Sal*I site.

pCH11.5E was digested with *Sac*I and incubated with dNTP's and *E. coli* DNA-polymerase I, Klenow fragment (section 2.2.7) to create flush-ended termini. This DNA was then redigested with *Sal*I, the 2.4 kb and 1.8 kb *Sac*I-*Sal*I fragments purified by LGT agarose gel electrophoresis and cloned into the vector M13mp83 to allow sequencing through the *Sal*I restriction site.

DNA sequences obtained were analysed for histone coding sequences. The sequences from the *Sal*I restriction sites confirmed the presence of two convergent H3 genes. The autoradiograph is shown in Figure 6.19 and the sequence is presented in Figure 6.20. The two sequences, Left and Right, were identical, except for two base-pairs (marked in Figs. 6.19 and 6.20), and suggest a very recent duplication/inversion event has occurred (discussed further in Chapter 8).

6.2.15 Restriction Analysis of Cosmid Clones 2.1C and 5.1C

Cosmids 2.1C and 5.1C are two extensively overlapping cosmid recombinants containing an H3 gene in a region of about 45 kb. The single H3 gene is contained within a 2.2 kb *Eco*RI-*Bam*HI fragment (Fig. 6.21) and has been identified from sequence data (Alan Robins, this laboratory; data not shown) as the split H3 gene previously characterised by Engel *et al* (1982).

This region of DNA had not been restriction mapped previously (no other genomic clones covered this segment of DNA) and digestion with *Eco*RI and *Bam*HI generated a large number of DNA fragments (Fig. 6.21).

To enable complete mapping of clone 2.1C it was decided to approach the analysis differently.

A number of *Eco*RI fragments (12.0 kb, 2.2 kb and 1.9 kb) were subcloned and further analysed with *Bam*HI. A large number of other

Figure 6.21

Restriction and Southern Analysis of Cosmid Recombinants 2.1C and 5.1C

1 ug samples of 2.1C and 5.1C DNA were digested with *Eco*RI and *Bam*HI, singly and in combination (see below). Enzyme digests were fractionated on 1% agarose gels (section 2.2.2a) and transferred to nitrocellulose filters (section 2.2.3). Filter bound DNA was hybridised to ^{32}P-labelled H3 gene-specific probe, prepared by nick-translation (2.2.11b), and then washed in 1xSSC/0.1% SDS at 65^{o}C. Bands were detected by autoradiography, overnight at -80^{o}C with an intensifying screen. DNA fragments were observed, before transfer, by staining the gel with ethidium bromide and viewing under UV light. Fragment sizes are indicated in kilobases. The figure shows the following;

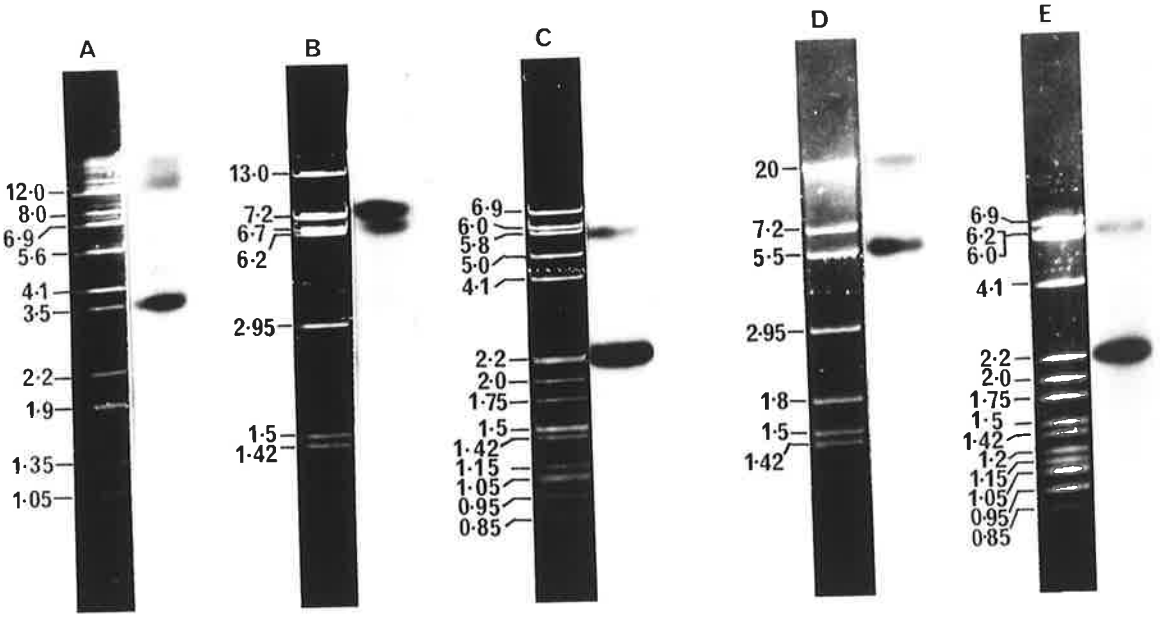
Track A: 2.1C digested with *Eco*RI

Track B: 2.1C digested with *Bam*HI

Track C: 2.1C digested with *Eco*RI and *Bam*HI

Track D: 5.1C digested with *Bam*HI

Track E: 5.1C digested with *Eco*RI and *Bam*HI


A
12·0
8·0
6·9
5·6
4·1
3·5
2·2
1·9
1·35
1·05
B
13·0
7·2
6·7
6·2
2·95
1·5
1·42
C
6·9
6·0
5·8
5·0
4·1
2·2
2·0
1·75
1·5
1·42
1·15
1·05
0·95
0·85
D
20
7·2
5·5
2·95
1·8
1·5
1·42
E
6·9
6·2
6·0
4·1
2·2
2·0
1·75
1·5
1·42
1·2
1·15
1·05
0·95
0·85

restriction enzyme fragments were cut from LGT agarose gels and redigested further, before extraction (section 2.2.1). The DNA fragments generated using this procedure were analysed by electrophoresis on agarose gels, alongside digests of the complete recombinant.

This analysis allowed specific restriction fragments to be located within a particular region (DNA fragment) of the clone and enabled restriction maps of separate fragments to be determined and then put together. The fragments analysed in this way are indicated in Figure 6.24 and restriction enzyme digestions of each fragment can be seen in Figures 6.22 and 6.23.

The size of the *Bam*HI fragment containing the single H3 gene is different in 2.1C and 5.1C (Fig. 6.21). This is due to the presence of an allelic *Bam*HI site, present in 5.1C but absent in 2.1C. (The 7.2 kb *Bam*HI-*Bam*HI fragment in 2.1C is replaced by 5.5 kb and 1.8 kb *Bam*HI-*Bam*HI fragments in 5.1C; Fig. 6.21).

All other internal fragments from these clones are identical except those containing the vector-insert junctions. The relationship between these two overlapping recombinants is shown in Figure 6.24.

6.2.16 Discussion

Using a modification of the protocol described by Ish-Horowicz and Burke (1981) a chicken genomic library (1 genome equivalent, approximately 50,000 clones) was constructed in the cosmid vector pHC79 (Hohn and Collins, 1980). This library was screened, using the procedure described by Hanahan and Meselson, with a mixture of core and H1 chicken histone gene probes. Five cosmid recombinants were isolated, of which three (6.3C, 2.1C and 5.1C) have been fully characterised. Another recombinant 6.1C was characterised by C.L. Lesnikowski (1983) and is discussed in chapter 7.

Using a non-coding probe specific for the end of λCHO7, cosmid 6.3C was shown to overlap with this recombinant. This clone extends

Figure 6.22

Restriction Analysis of 2.1C and Internal Fragments

5 ug samples of 2.1C DNA were digested with *Eco*RI, *Bam*HI or *Hin*dIII and fractionated on a 1% LGT agarose gel. A number of individual fragments were excised from the gel, melted and redigested using the procedure described in section 2.2.1. Resultant fragments were resolved on a 1% agarose gel (section 2.2.2a) and visualised with UV light after ethidium bromide staining. 2.1C DNA (1 ug) digested with different combinations of *Eco*RI and *Bam*HI was co-electrophoresed to give reference fragments.

1 ug samples of plasmids derived from 2.1C (12.0 kb and 1.9 kb *Eco*RI-*Eco*RI fragments) were also restriction enzyme digested and analysed by electrophoresis on the 1% agarose gels described above.

Fragment sizes are indicated, in kilobases, alongside each track. The figure shows the following;

Track A: 2.1C DNA digested with *Eco*RI

Track B: 2.1C DNA digested with *Bam*HI

Track C: p12.0E DNA digested with *Eco*RI

Track D: p12.0E DNA digested with *Eco*RI and *Bam*HI

Track E: 2.1C DNA digested with *Bam*HI

Track F: 2.1C DNA digested with *Eco*RI and *Bam*HI

Track G: 5.6 kb *Eco*RI-*Eco*RI fragment digested with *Bam*HI

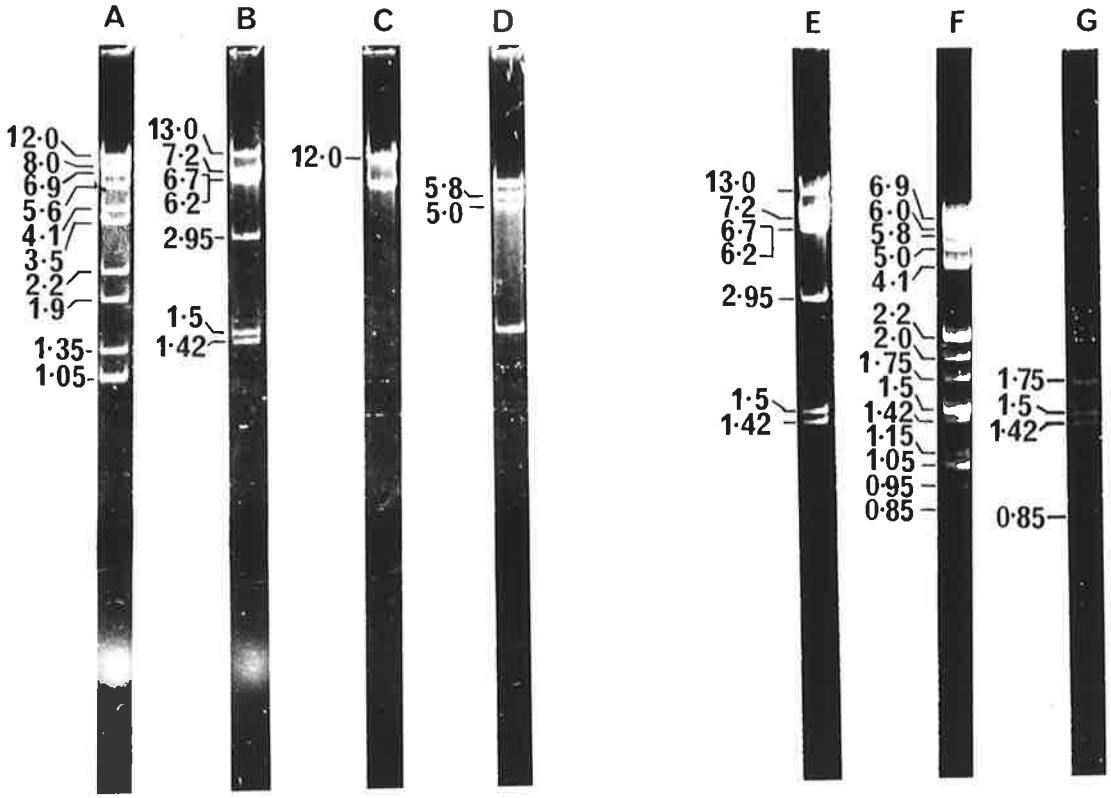

A
B
C
D
E
F
G
12·0
8·0
6·9
5·6
4·1
3·5
2·2
1·9
1·35
1·05
13·0
7·2
6·7
6·2
2·95
1·5
1·42
12·0
5·8
5·0
13·0
7·2
6·7
6·2
2·95
1·5
1·42
6·9
6·0
5·8
5·0
4·1
2·2
2·0
1·75
1·5
1·42
1·15
1·05
0·95
0·85
1·75
1·5
1·42
0·85

Figure 6.23

Restriction Analysis of 2.1C and Internal Fragments

Restriction fragments and subclones derived from 2.1C were digested with *Eco*RI and *Bam*HI and analysed by electrophoresis on 1% agarose gels. Details of the procedure used are given in Figure 6.22 and section 2.2.1.

The figure shows the following;

Track A: 2.1C DNA digested with *Eco*RI

Track B: 2.1C DNA digested with *Bam*HI

Track C: 2.1C DNA digested with *Eco*RI and *Bam*HI

Track D: p1.9E DNA digested with *Eco*RI and *Bam*HI

Track E: 14.0 kb *Bam*HI-*Bam*HI fragment digested with *Eco*RI

Track F: 12.0 kb *Hin*dIII-*Hin*dIII fragment digested with *Eco*RI

Track G: 12.0 kb *Hin*dIII-*Hin*dIII fragment digested with *Bam*HI

Track H: 6.0 kb *Hin*dIII-*Hin*dIII fragment digested with *Eco*RI

Track I: 6.0 kb *Hin*dIII-*Hin*dIII fragment digested with *Bam*HI

Track J: 2.1C DNA digested with *Eco*RI

Track K: 2.1C DNA digested with *Bam*HI

Track L: 2.1C DNA digested with *Eco*RI and *Bam*HI

Track M: 1.4 kb *Bam*HI-*Bam*HI fragment digested with *Eco*RI

Track N: 1.5 kb *Bam*HI-*Bam*HI fragment digested with *Eco*RI

Track O: 1.35 kb *Eco*RI-*Eco*RI fragment digested with *Bam*HI

Track P: 2.95 kb *Bam*HI-*Bam*HI fragment digested with *Eco*RI

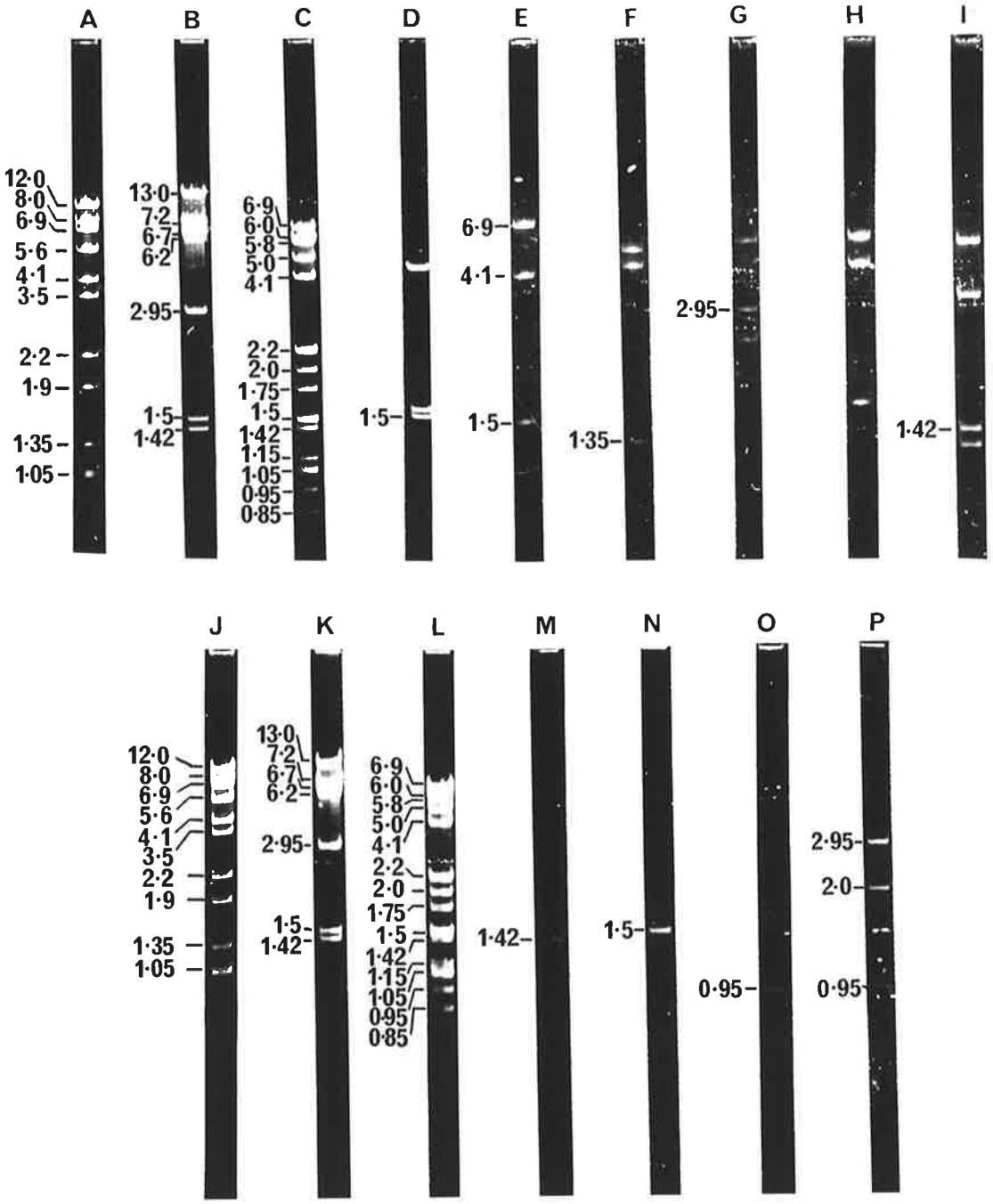

A
12·0
8·0
6·9
5·6
4·1
3·5
2·2
1·9
1·35
1·05
B
13·0
7·2
6·7
6·2
2·95
1·5
1·42
C
6·9
6·0
5·8
5·0
4·1
2·2
2·0
1·75
1·5
1·42
1·15
1·05
0·95
0·85
D
1·5
E
6·9
4·1
1·5
F
1·35
G
2·95
H
I
1·42
J
12·0
8·0
6·9
5·6
4·1
3·5
2·2
1·9
1·35
1·05
K
13·0
7·2
6·7
6·2
2·95
1·5
1·42
L
6·9
6·0
5·8
5·0
4·1
2·2
2·0
1·75
1·5
1·42
1·15
1·05
0·95
0·85
M
1·42
N
1·5
O
0·95
P
2·95
2·0
0·95

Figure 6.24

Restriction Maps of 2.1C and 5.1C

The sizes, in kilobases, of the restriction fragments obtained on cleavage of 2.1C (A) and 5.1C (B) with *Eco*RI and *Bam*HI are shown. The region of the clones derived from vector DNA is indicated by the bold line. The shaded regions show the position of H3 coding sequences as determined by Southern analysis (Fig. 6.21). An allelic *Bam*HI site is present in 5.1C DNA but absent in 2.1C. The DNA fragments analysed during the mapping procedure are indicated below the overall *Bam*HI/*Eco*RI map of 2.1C and are as follows;

a: p12.0E insert

b: p1.9E insert

c: 13.0 kb *Bam*HI-*Bam*HI fragment

d: 1.35 kb *Eco*RI-*Eco*RI fragment

e: 2.95 kb *Bam*HI-*Bam*HI fragment

f: 12.0 kb *Hin*dIII-*Hin*dIII fragment

g: 5.6 kb *Eco*RI-*Eco*RI fragment

h: 6.0 kb *Hin*dIII-*Hin*dIII fragment

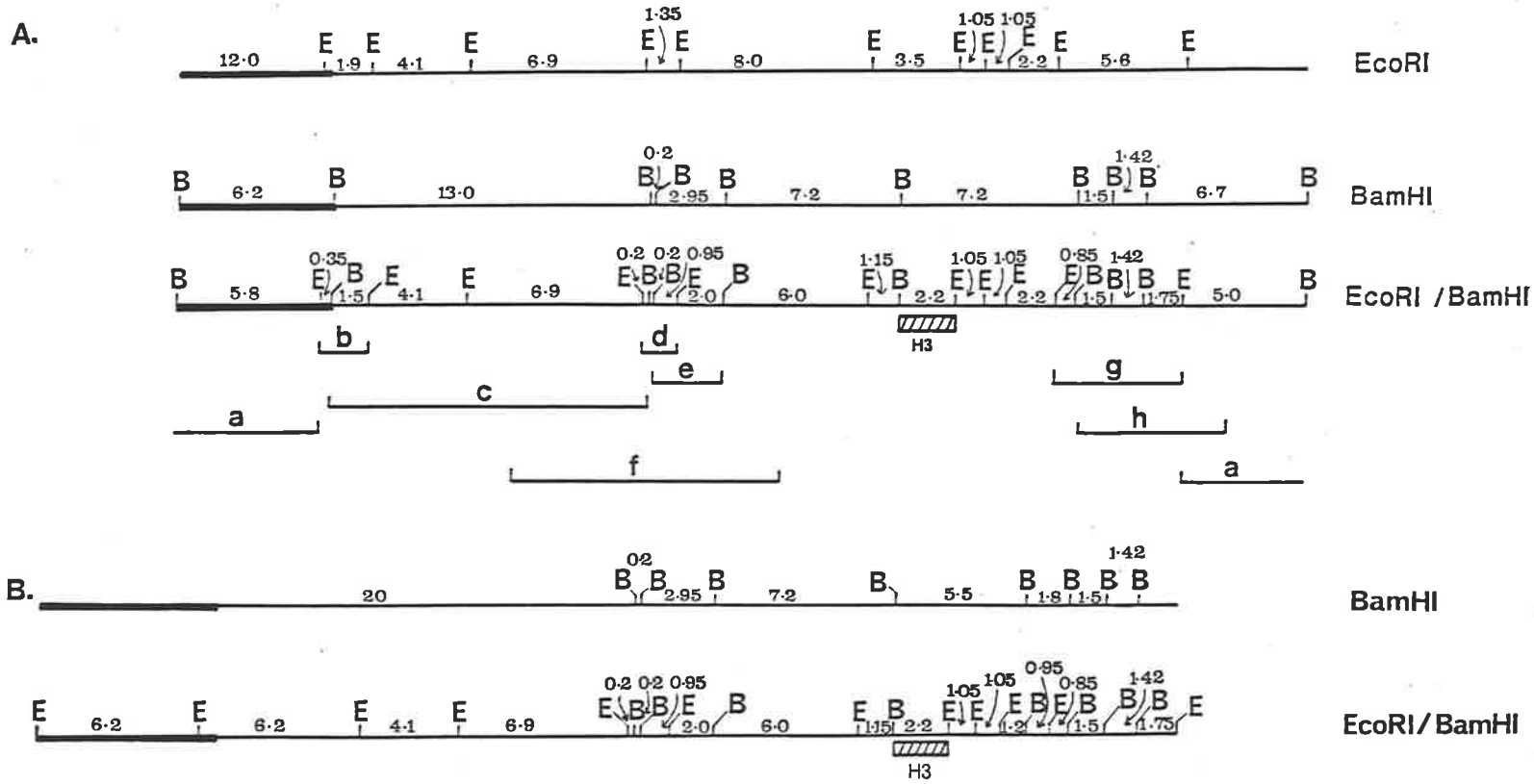

A.
EcoRI
BamHI
EcoRI /BamHI
H3
B.
BamHI
EcoRI/BamHI
H3

Figure 6.25

Structure of Recombinant Clones

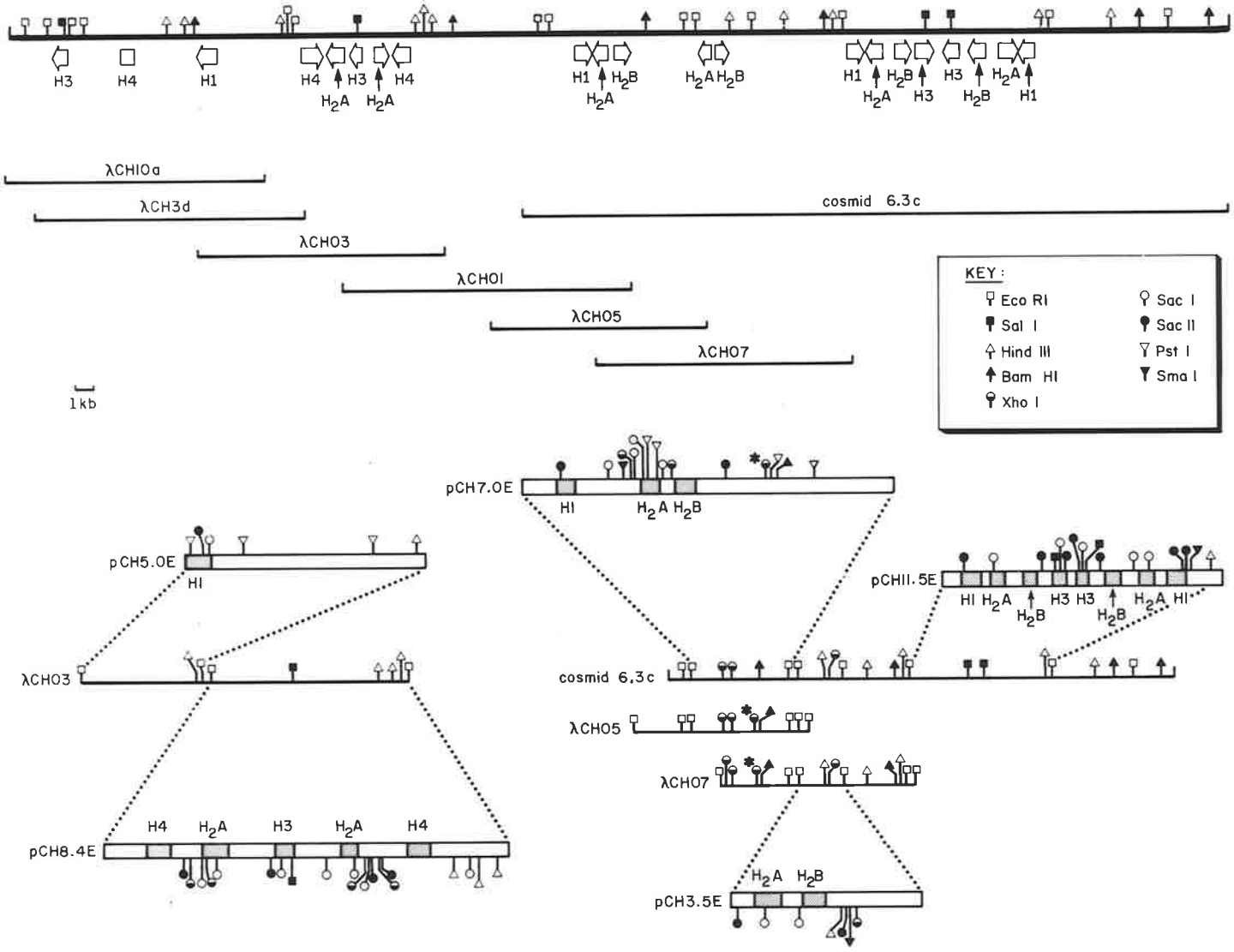
Surrounding λCHO1

The overall gene organisation in the region surrounding λCHO1 is shown. Histone genes are indicated in the top diagram by the arrows and the clones derived from this region are shown. The direction of transcription of the histone genes (5'-3'), indicated by the horizontal arrows, was determined by sequencing (R. Sturm, L.S. Coles, J. Powell, personal communication; Wang *et al*, 1985).

Clones λCH10a and λCH3d have been described previously (Sugarman *et al*, 1983). Detailed maps of histone gene containing subclones, derived from the genomic clones, are also shown with the precise locations of genes indicated.

λCH10a and λCH3d contain a 4.0 kb *Eco*RI-*Hin*dIII fragment which hybridises to H4 probe (Sugarman *et al*, 1983). The number of H4 genes within this fragment is undetermined.

A *Xho*I restriction site present in λCHO5, λCHO7 and pCH7.0E (derived from λCHO5) is not present in the cosmid recombinant 6.3C. Terminal *Eco*RI restriction sites in λCHO5 and λCHO7 were generated by the addition of *Eco*RI linker DNA during library construction.


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

the region already described (Chapter 5) approximately 20 kb and contains another histone gene cluster, giving a fully characterised region spanning about 75 kb and including 21 accurately located histone coding regions (Fig. 6.25).

These genes are arranged into clusters of 2-8 separated by non-histone encoding regions of up to 8 kb (these regions are further discussed in chapter 7). It is clear from this data that no long range histone gene repeat exists at this locus. Except for the frequent occurrence of divergently transcribed, H_2A-H_2B gene pairs the arrangements of the histone genes within clusters are unrelated (discussed further in chapter 8).

Clones 2.1C and 5.1C have overlapping restriction maps and contain a single H3 gene subsequently identified as the split H3 gene previously reported by Engel *et al* (1982).

CHAPTER 7

FINAL ANALYSIS OF HISTONE

GENE RECOMBINANTS

7.1 INTRODUCTION

From studies described in the previous chapters it has been possible to locate, in recombinant DNA clones, core and H1 histone genes related to those in λCHO1 (Harvey *et al*, 1981). The analysis described covers a region of approximately 70 kb (Fig. 6.25) and allows conclusions to be drawn concerning the overall organisation of this group of histone genes (discussed further in chapter 8).

With other data from this laboratory (Figs. 7.1 and 7.2) at least 40 core and H1 histone genes can be accounted for. Rather than extend these regions by further chromosome crawling, it was decided to complete the study and investigate some other aspects relevant to the organisation of the histone genes.

7.2 RESULTS

7.2.1 Relationship of Core and H1 Histone Gene Containing Regions to Each Other

In addition to the region shown in Figure 6.25 two other regions of DNA, containing core and H1 histone gene sequences, have been analysed in detail.

C. Lesnikowski (1983) and L.S. Coles (this laboratory) have identified and characterised recombinants which overlap with λCHO2 (Harvey *et al*, 1981). This data is presented in Figure 7.1. Two other overlapping λ recombinants (λH1.9 and λH1.10) have also been characterised (L.S. Coles, this laboratory) and are detailed in Figure 7.2.

It was relevant, for the completion of this study, to determine whether any of these histone gene regions were linked to the stretch of DNA spanning λCHO1, or to each other.

Non-histone encoding DNA fragments from the ends of λCH10a (0.7 kb *Eco*RI-*Eco*RI fragment, Fig. 6.25), 6.3C (2.0 kb *Bam*HI-*Eco*RI fragment, Fig. 6.12), λH1.1 (0.7 kb *Eco*RI-*Hin*dIII fragment, Fig. 7.1)

Figure 7.1

Structure of Recombinants Surrounding λCHO2

The genes present in this region are shown below the overall restriction map. The direction of transcription of some genes has been determined by sequencing (R. Sturm, L.S. Coles, A. Robins, personal communication; Harvey *et al*, 1981) and is indicated.

The order of the *Hin*dIII-*Hin*dIII fragments (marked x) in λH1.1 has not been determined. The *Hin*dIII sites in λH1.4 marked with an asterisk are polymorphic (i.e. not present in cosmid 6.1C which is derived from DNA of a different chicken; see section 8.1.4). Terminal *Eco*RI sites in λH1.1 and λH1.4 are derived from linker DNA. Subclones pCH22.0B and pCH2.7B were constructed using *Bam*HI-*Bam*HI fragments from cosmid 6.1C. Detailed maps of these clones are also shown, indicating the precise gene locations. Mapping data is from Lesnikowski (1983) and L.S. Coles, personal communication.

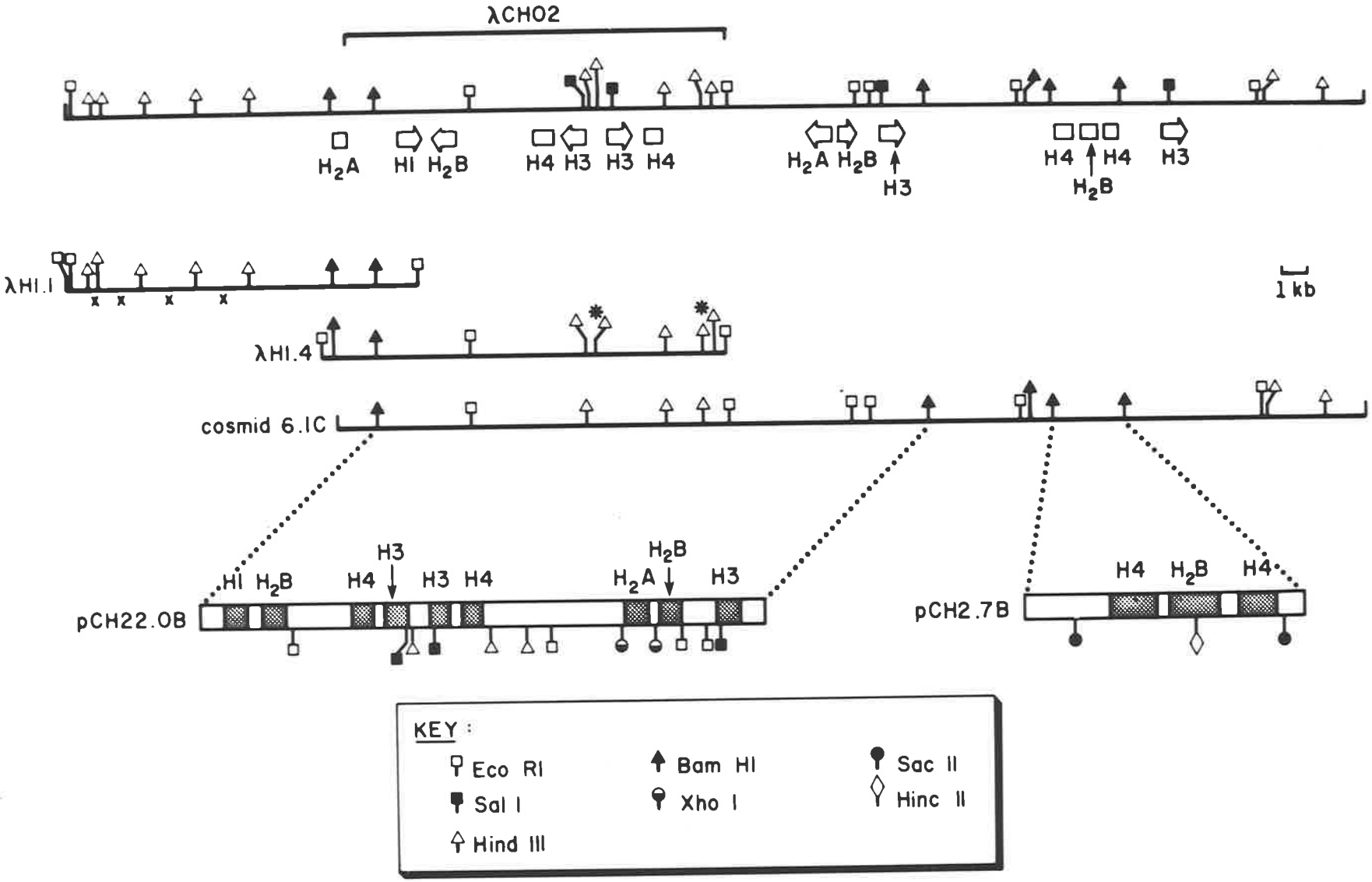
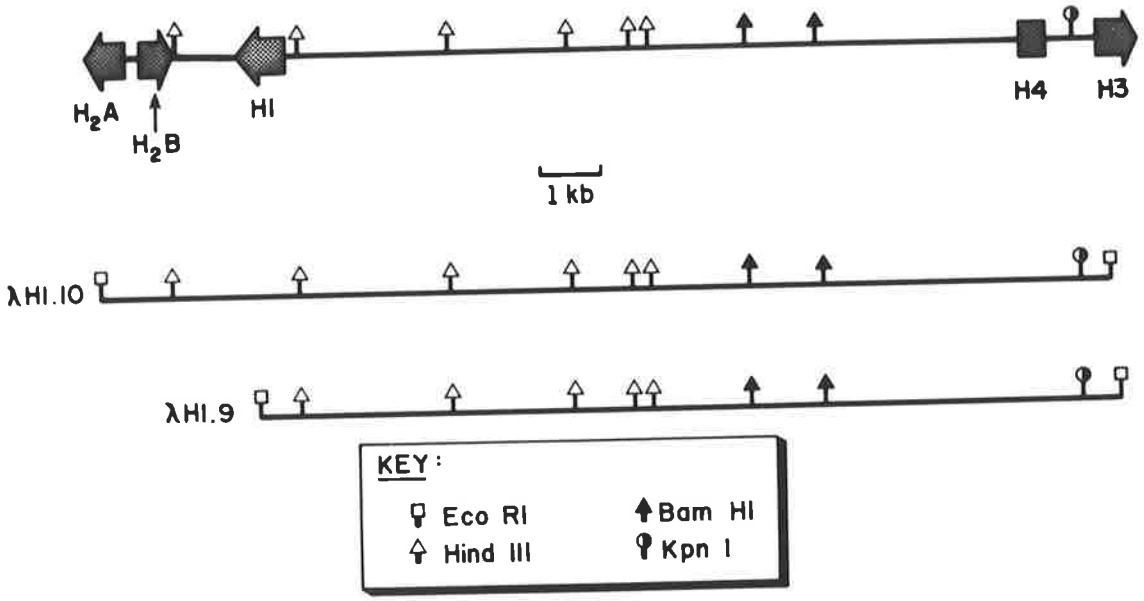

λCH02
H2A
H1 H2B
H4 H3 H3 H4
H2A H2B
H3
H4 H4
H2B
H3
λH1.1
1 kb
λH1.4
cosmid 6.1C
H1 H2B
H3
H4
H3 H4
H2B
H2A
H3
pCH22.0B
H4 H2B H4
pCH2.7B
KEY :
Eco RI
Sal I
Hind III
Bam HI
Xho I
Sac II
Hinc II

Figure 7.2

Structure of Genomic Clones λH1.9 and λH1.10

An overall restriction map of this histone gene-containing region is presented. Data was provided by L.S. Coles. The orientation of all the genes except H4 has been determined (R. Sturm, L.S. Coles, personal communication) and is indicated (5' - 3') by the arrows. The order of the *Hin*dIII-*Hin*dIII fragments has not been determined. Terminal *Eco*RI cleavage sites are derived from linker DNA.


H2A
H2B
H1
H4
H3
1 kb
λH1.10
λH1.9
KEY:
Eco RI
Hind III
Bam HI
Kpn I

and 6.1C (2.3 kb *Hin*dIII-*Hin*dIII fragment, Fig. 7.1) were purified by electrophoresis through LGT agarose, (section 2.2.2), eluted, labelled with ^{32}P and used to probe recombinants representing the ends of all the regions discussed. These probes were all specific for the recombinants from which they were derived and did not hybridise to DNA fragments from other clones when washed at high stringency (0.1xSSC/0.1% SDS) (data not shown). These results indicate that the region characterised in this work (Fig. 6.25) does not overlap with other core and H1 containing regions characterised in this laboratory. Furthermore, hybridisation with probes specific for cosmid recombinant 6.1C and for λH1.1 confirmed that this region (Fig. 7.1) does not overlap with any other characterised region. Taken together these results clearly demonstrate that the three characterised regions containing core and H1 genes (Figs. 6.25, 7.1 and 7.2) are non-overlapping and represent separate stretches of histone containing DNA.

7.2.2 Relationship of Core and H1 Histone Gene Containing Regions to 2.1C and 5.1C

All of the probes described in section 7.2.1 were used to probe DNA from the cosmid recombinants 2.1C and 5.1C (section 6.2.15). None of these probes hybridised to DNA from 2.1C or 5.1C clearly indicating that these two recombinants are not derived from the regions of the genome which generated λCHO1 or λCHO2.

The genomic region spanned by λH1.9 and λH1.10 (Fig. 7.2) has an H_2A-H_2B-H1 gene cluster at one end and an H3-H4 gene pair at the other. By comparison with the 2.1C/5.1C data, presented in Figure 6.24, it is not possible for these regions to overlap.

Therefore the region spanned by 2.1C and 5.1C (containing the single split H3 gene) is unlinked to the regions containing core and H1 histone genes.

7.2.3 Relationship of the Histone H5 Gene to Core and H1 Histone Genes

H5 is an H1 related, tissue-specific histone variant expressed in erythroid cells. The isolation of cDNA clones containing histone H5 sequences (Krieg *et al*, 1982b) has enabled the selection of recombinants containing the chromosomal H5 gene from the chicken genomic library (Krieg *et al*, 1983; Fig 7.3). Results from this laboratory indicate that the H5 gene is unique and contains no introns (Krieg *et al*, 1983) and also show that the polyadenylated H5 mRNA lacks the termination element typical of other histone messages (Krieg *et al*, 1982a).

The single chicken H5 gene is completely contained within a λ-recombinant referred to as λCH501. It was of interest to know whether this H5 gene is interspersed with the core and H1 histone genes in the chicken genome. The 50 λ-recombinants isolated from the chicken genomic library (Chapter 4) and known to contain H1 and/or core histone sequences all failed to hybridise to H5 specific probes (data not shown). The recombinant λCH501, containing the H5 gene, similarly failed to hybridise to gene-specific core and H1 histone probes, setting a limit of approximately 7 kb each side of the H5 gene in which no other histone sequences were present. As intergene regions of approximately 8 kb exist between chicken histone gene clusters (see Figure 6.25) it was decided to extend this analysis to recombinants which overlap λCH501. These clones were obtained after rescreening the chicken library with DNA fragments from each end of the λCH501 insert (Paul Krieg, this laboratory). Together with the overlapping clones a continuous genomic region of 41 kb is represented (Fig. 7.3). Using the procedure described in section 2.2.5, plaques of these three overlapping clones were spotted, in duplicate, onto a fresh lawn of LE392. λCH01, which contains H1, H_2A, H_2B, H3 and H4 gene sequences was included as a positive control,

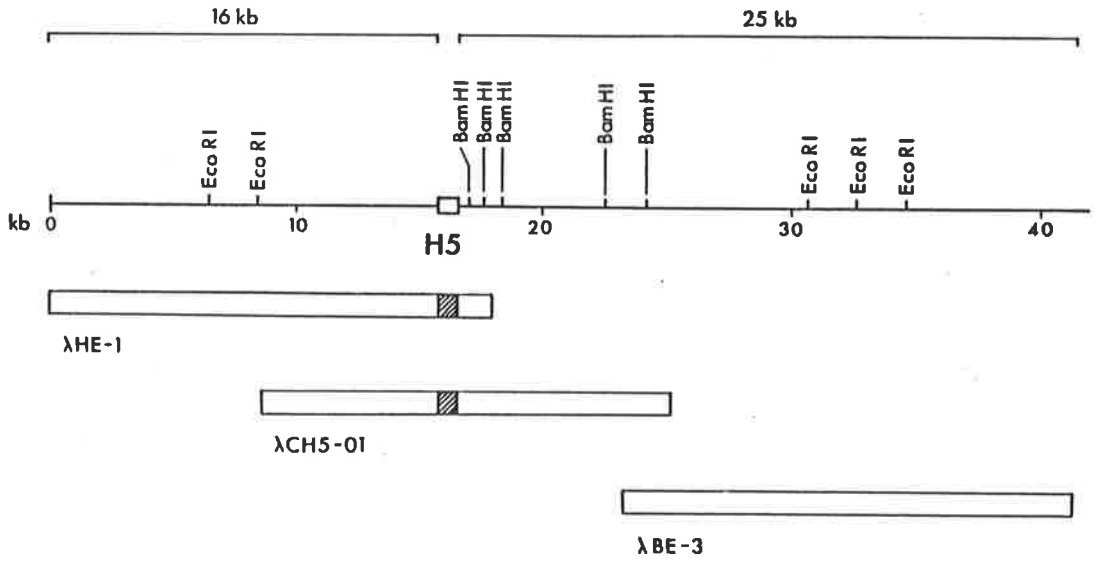
Figure 7.3

Restriction Map of Chicken Genomic DNA Surrounding the Histone H5 Gene

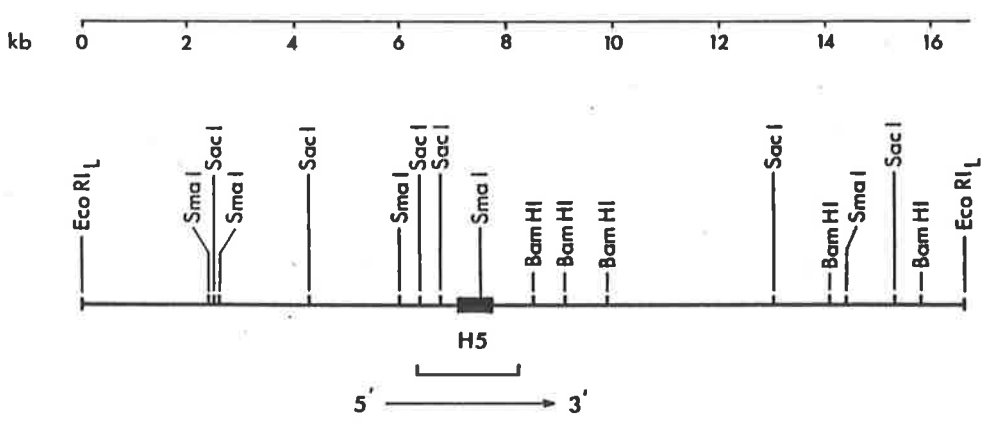
A) Schematic representation of the overlapping genomic clones in the chicken H5 gene region. Data from Krieg *et al* (1983).

B) Restriction map of the λCH501 insert. The blocked in region represents the protein coding sequences of the H5 gene and the direction of transcription is from left to right, as indicated by the arrow. The λCH501 insert contains no *Eco*RI sites. The *Eco*RI sites at the end of the insert are artificial, having been created during construction of the genomic library.

A.


16 kb
25 kb
Eco RI
Eco RI
Bam HI
Bam HI
Bam HI
Bam HI
Bam HI
Eco RI
Eco RI
Eco RI
kb 0
10
20
30
40
H5
λHE-1
λCH5-01
λBE-3

B.


kb 0
2
4
6
8
10
12
14
16
Eco RI L
Sma I
Sac I
Sma I
Sac I
Sma I
Sac I
Sac I
Sma I
Bam HI
Bam HI
Bam HI
Sac I
Bam HI
Sma I
Sac I
Bam HI
Eco RI L
H5
5'
3'

and λ(charon 4a) as a negative control. These plaques were transferred to nitrocellulose (section 2.2.5a) and hybridised to histone gene specific DNA probes (Fig. 7.4).

This result shows that none of the histone genes present in λCHO1 exist within the 41 kb region surrounding the unique chicken H5 gene. It is possible that other histone gene sequences, which would not hybridise with these core and H1 sequences, may exist in this 41 kb region, but, taken together, the results described here suggest that the histone genes examined in this thesis are not closely linked to the variant H5 histone gene.

7.2.4 The Region Between the Histone Genes

(a) Detection of Non-Histone Gene Sequences

Given the location of the histone gene coding regions, it was of interest to determine whether other coding regions exist, interspersed with these genes. DNA sequencing of specific regions can be useful in identifying possible gene sequences in localised regions. For example, a possible non-histone, processed pseudogene and a small reading frame, with promotor elements resembling those found in polymerase II genes, have been identified by sequencing histone gene containing clones (S.W. Wang and L.S. Coles, personal communication). However sequencing is not a practical way of analysing extensive chromosomal regions.

A preliminary attempt was made to identify any non-histone encoding genes within the histone gene containing recombinants by probing with cDNA.

DNA from all characterised recombinants (see table 7.1) was digested with appropriate restriction endonucleases, electrophoresed on 1.2% agarose gels and transferred to nitrocellulose (sections 2.2.2a and 2.2.3).

The probe used for hybridisation was oligo-dT primed cDNA prepared as described in section 2.2.11a using, as template, poly (A) plus RNA

Figure 7.4

Probing of the H5 Gene Region With Histone Gene Sequences

Duplicate plaques of λCH01, λCH501, λHE-1 and λBE-3 (see Fig. 7.3) and a control of λCharon 4A vector were grown on a lawn of *E. coli* LE392 (section 2.2.5a), transferred to nitrocellulose and hybridised with the radiolabelled gene specific probes indicated. Filters were washed in 2xSSC/0.1% SDS at 65°C and autoradiographed overnight at -80°C with an intensifying screen.

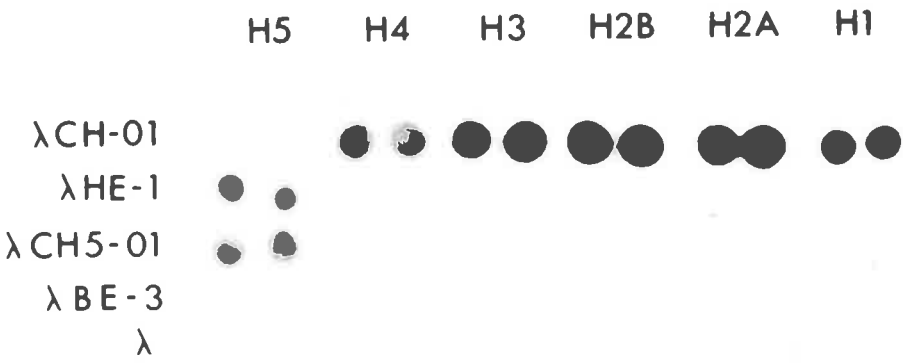

H5
H4
H3
H2B
H2A
H1
λCH-01
λHE-1
λCH5-01
λBE-3
λ

Table 7.1

Southern Analysis of Chicken Recombinants

The table shows the recombinants analysed in this chapter and the restriction enzyme digests carried out.

Restriction enzyme cleaved DNA was fractionated on 1% agarose gels and transferred to nitrocellulose filters using the bidirectional transfer procedure (section 2.2.3). Filter bound DNA was hybridised to,

a) ^{32}P-labelled cDNA made to poly (A) plus RNA from 5-day old chicken embryos.

b) ^{32}P-labelled total genomic DNA (see text for details).

Table 7.1

Recombinant	Figure Reference	Restriction Enzyme Digest
pH5(cDNA)	-	*Pst*I
pCH9.0E(6.3C)	Fig. 6.25	*Eco*RI/*Bam*HI
pCH7.0E	Figs. 6.25; 5.21	*Sac*II/*Xho*I
pβ-globin	-	*Hin*dIII
$\lambda H_2A.F$	-	*Eco*RI/*Bam*HI
λHE-1	Fig. 7.3	*Eco*RI
λCH501	Fig. 7.3	*Eco*RI/*Bam*HI
λBE-3	Fig. 7.3	*Eco*RI
λCH10a	Fig. 6.25	*Eco*RI/*Hin*dIII
λCH03	Figs. 6.25; 5.3	*Eco*RI/*Hin*dIII
λCH01	Figs. 6.25; 3.1	*Eco*RI/*Hin*dIII
λH1.1	Fig. 7.1	*Hin*dIII
λH1.2	Table 4.1	*Eco*RI/*Hin*dIII
λH1.3	Table 4.1	*Eco*RI/*Hin*dIII
λH1.4	Fig 7.1	*Eco*RI/*Hin*dIII
λH1.7	Table 4.1	*Eco*RI/*Hin*dIII
λH1.9	Fig. 7.2	*Eco*RI/*Hin*dIII
λH1.10	Fig. 7.2	*Eco*RI/*Hin*dIII
6.3C	Figs. 6.25; 6.12	*Eco*RI/*Xho*I
6.1C	Fig. 7.1	*Eco*RI/*Hin*dIII
4.2C	-	*Eco*RI
5.1C	Fig. 6.24	*Eco*RI
2.1C	Fig. 6.24	*Eco*RI

prepared from 5-day chicken embryos (Alan Colman, this laboratory). Figure 7.5 shows the result of the hybridisation and gives details of the washing conditions used.

This cDNA preparation did not detect DNA fragments containing core and H1 histone genes and was therefore free of labelled histone cDNA sequences (with the exception of histone H5). Furthermore, it hybridised strongly to a control DNA fragment containing β-globin sequences and also easily detected DNA encoding histone H5 (Fig. 7.5) which is present at a low level in the chicken embryo. However, the probe did not detect the 3.5 kb *Eco*RI-*Eco*RI fragment of 2.1C containing a split H3 gene which is reported to give rise to a polyadenylated product (Engel, 1984). This can be explained if this H3 gene encodes a product which replaces the "embryonic" histone gene products later in chicken development, or alternatively if it encodes a low abundance, constitutively expressed (i.e. cell-cycle independent) H3 gene (Engel, 1984).

A 2.2 kb *Eco*RI-*Eco*RI fragment from the cosmid recombinants 2.1C and 5.1C was the only restriction fragment from all the tested recombinants to hybridise to this cDNA probe (Fig. 7.5). The specificity of the cDNA probe for this restriction fragment suggests that a gene exists within this fragment, and that polyadenylated message from this gene is present at a low level in the 5-day chicken embryo.

This 2.2 kb *Eco*RI-*Eco*RI fragment was purified by electrophoresis through LGT agarose (section 2.2.2a), ^{32}P-labelled by nick-translation and used to probe *Eco*RI digested chicken genomic DNA (see Fig. 7.6). This fragment is specific for the region from which it is derived, indicating that the sequences represented within the 2.2 kb *Eco*RI-*Eco*RI fragment are unique in the chicken genome. The identity of the transcript and the extent to which it is expressed have not been investigated further but it is of interest that it is expressed early in development, as are the histone genes.

Figure 7.5

Probing Recombinants With Embryonic cDNA Probe

Samples of DNA (1-2 ug) from all characterised recombinants were digested with restriction enzymes (see Table 7.1), fractionated on 1% agarose gels (section 2.2.2a), visualised by ethidium bromide staining and transferred to nitrocellulose using the bidirectional transfer procedure (section 2.2.3). One set of filters was hybridised with ^{32}P-labelled cDNA which had been prepared by extending oligo-dT primer (section 2.2.11a) while using poly (A) plus RNA isolated from 5 day old chicken embryos as template. Filters analysed in this way were washed in 1xSSC/0.1% SDS at 65^{o}C and exposed to X-ray film for 42 hours at -80^{o}C with an intensifying screen. Only the digests discussed in the text are shown.

Track A: *Hin*dIII digest of a ß-globin cDNA clone (generates vector + insert DNA fragments)

Track B: *Pst*I digest of an H5 cDNA clone (generates vector + insert DNA fragments)

Track C: *Eco*RI digest of 2.1C DNA

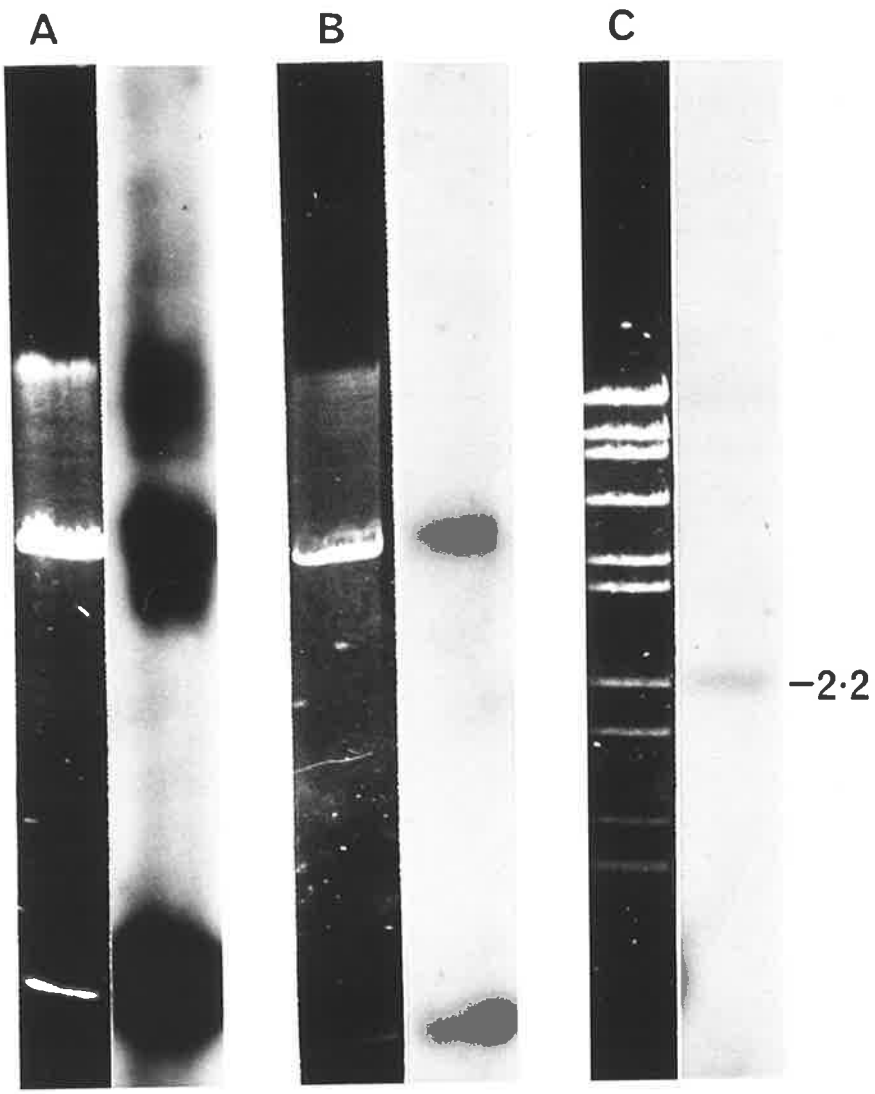
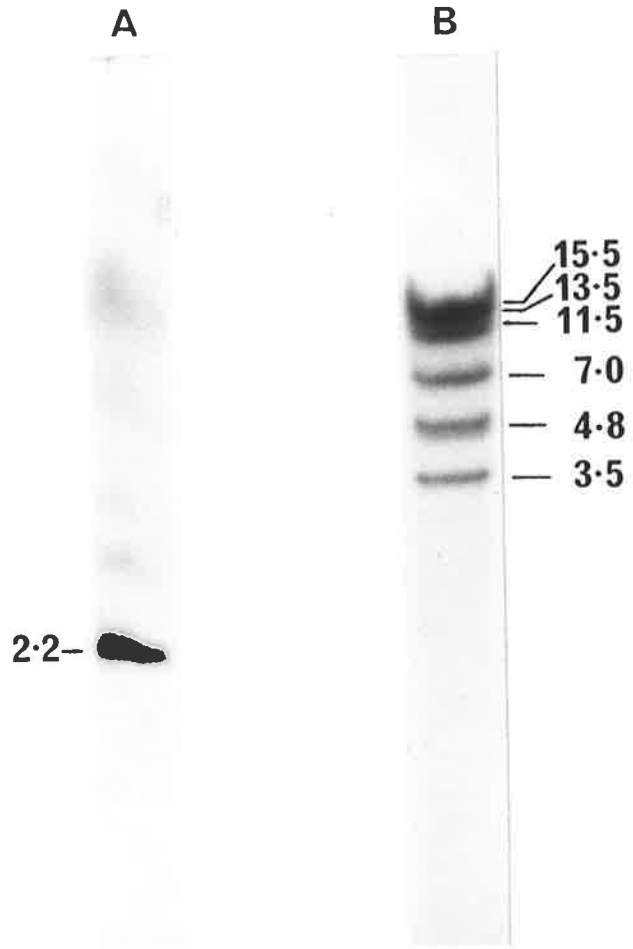

A
B
C
–2·2

Figure 7.6

Southern Analysis of Chicken Genomic DNA

High molecular weight DNA isolated from chicken reticulocytes (20 ug, courtesy of Dr. R.P. Harvey) was digested with *Eco*RI. The reaction was split into two and fractionated in separate tracks on a 1% agarose gel. After transfer to nitrocellulose (section 2.2.3) the filter was bisected and hybridised with either H_2A gene specific probe (track B) or ^{32}P-labelled 2.2 kb *Eco*RI-*Eco*RI fragment from 2.1C (track A). Filters were washed in 1xSSC/0.1% SDS at 65°C and autoradiographed overnight at -80°C with an intensifying screen. The sizes (in kilobases) of the H_2A gene containing fragments, as predicted from the mapping data, are indicated.


A
B
15·5
13·5
11·5
7·0
4·8
3·5
2·2

The fact that no DNA fragments from any other clones were positive with the cDNA probe may suggest there is no non-histone mRNA produced from these regions. However, an obvious limitation to this approach is that the cDNA probe used will not be representative of all chicken polymerase II genes. Therefore it cannot be concluded that clones containing the core and H1 histone genes contain no other protein coding regions.

(b) Detection of Repeat-Sequence DNA

Repeat-sequence DNA can be located by using total ^{32}P-labelled genomic DNA as a hybridisation probe to filter bound DNA fragments (Shen and Maniatis, 1980). This method provides an indication of the relative repetition frequency of a DNA sequence and is reported to detect sequences in the genome which are repeated fifty fold or greater.

DNA from recombinants representing all characterised regions (table 7.1) was digested and bound to nitrocellulose. 800 ng of chicken genomic DNA (prepared by R. Harvey, this laboratory) was ^{32}P-labelled by nick translation (section 2.2.11b) and used as a hybridisation probe. In order to prevent background hybridisation carrier DNA was not included in the prehybridisation or hybridisation mixes for this experiment and filters were prehybridised overnight. Filters were washed at medium stringency (1xSSC/0.1% SDS, 65^{o}C) and exposed. A 24 hour exposure revealed a number of strong signals and after a 5 day exposure a number of specific, less intense signals were also detected.

If the region including λCHO1 is considered, the only recombinant to hybridise to this probe was cosmid 6.3C. The fragments hybridising from this clone were the 9.0 kb *Eco*RI-*Eco*RI fragment, the 6.8 kb *Eco*RI-*Eco*RI fragment and the 4.2 kb *Eco*RI-*Xho*I fragment (Fig. 7.7a). The region within the 9.0 kb *Eco*RI-*Eco*RI fragment was localised further by probing an *Eco*RI/*Bam*HI digest of pCH9.0E with the same probe (Fig. 7.8b). The only insert fragment to hybridise to this probe was the

Figure 7.7

Probing Recombinants with ^{32}P-labelled Genomic DNA

Filter-bound, restriction enzyme cleaved DNA from all characterised recombinants (see Fig. 7.5; Table 7.1) was hybridised to total chicken genomic DNA which had been ^{32}P-labelled by nick-translation (section 2.2.11b). Filters were washed in 2xSSC/0.1% SDS at 65^{o}C and bands were detected by autoradiography, for 24 hours and 1 week, at -80^{o}C with an intensifying screen. After exposure, filters were boiled for 15 minutes in distilled water, prehybridised and hybridised with either the 0.8 kb *Sac*II-*Xho*I fragment (derived from pCH7.0E) or the 2.0 kb *Bam*HI-*Eco*RI fragment from pCH9.0E. Filters were washed (as above) and exposed overnight. Sizes are indicated in kilobases.

a: 6.3C DNA digested with *Eco*RI and *Xho*I; probed with

1. total genomic DNA
2. 0.8 kb *Sac*II-*Xho*I fragment
3. 2.0 kb *Bam*HI-*Eco*RI fragment

b: 2.1C DNA digested with *Eco*RI and probed with total genomic DNA.

c: 5.1C DNA digested with *Eco*RI and probed with total genomic DNA.

d: 2.1C DNA digested with *Eco*RI and probed with 0.8 kb *Sac*II-*Xho*I fragment.

e: 5.1C DNA digested with *Eco*RI and probed with 0.8 kb *Sac*II-*Xho*I fragment.

f: 6.1C DNA digested with *Eco*RI and *Hin*dIII; probed with

1. total genomic DNA
2. 0.8 kb *Sac*II-*Xho*I fragment

g: λH1.4 DNA digested with *Eco*RI and *Hin*dIII; probed with

1. total genomic DNA
2. 0.8 kb *Sac*II-*Xho*I fragment

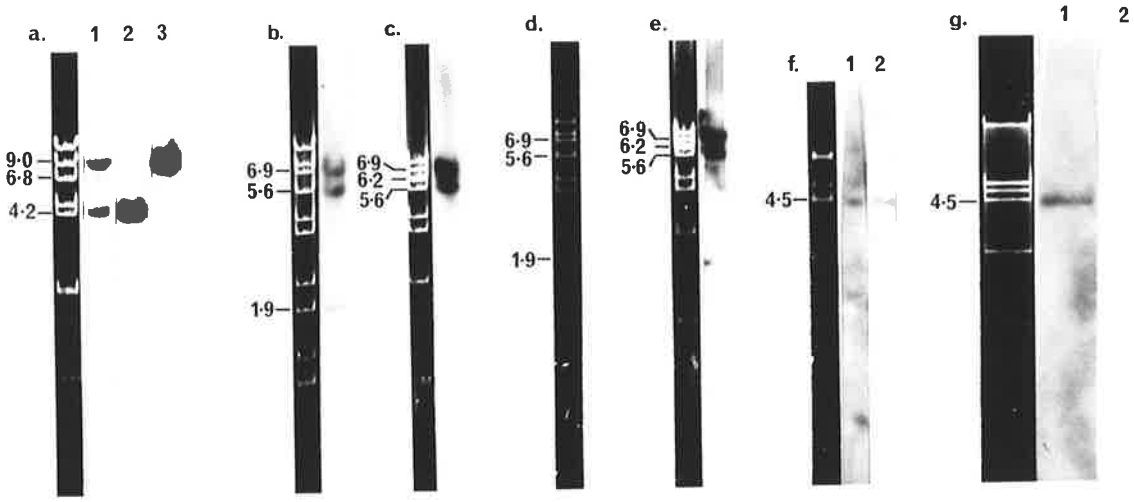

a.
1
2
3
9·0
6·8
4·2
b.
6·9
5·6
1·9
c.
6·9
6·2
5·6
d.
6·9
5·6
1·9
e.
6·9
6·2
5·6
f.
1
2
4·5
g.
1
2
4·5

Figure 7.8

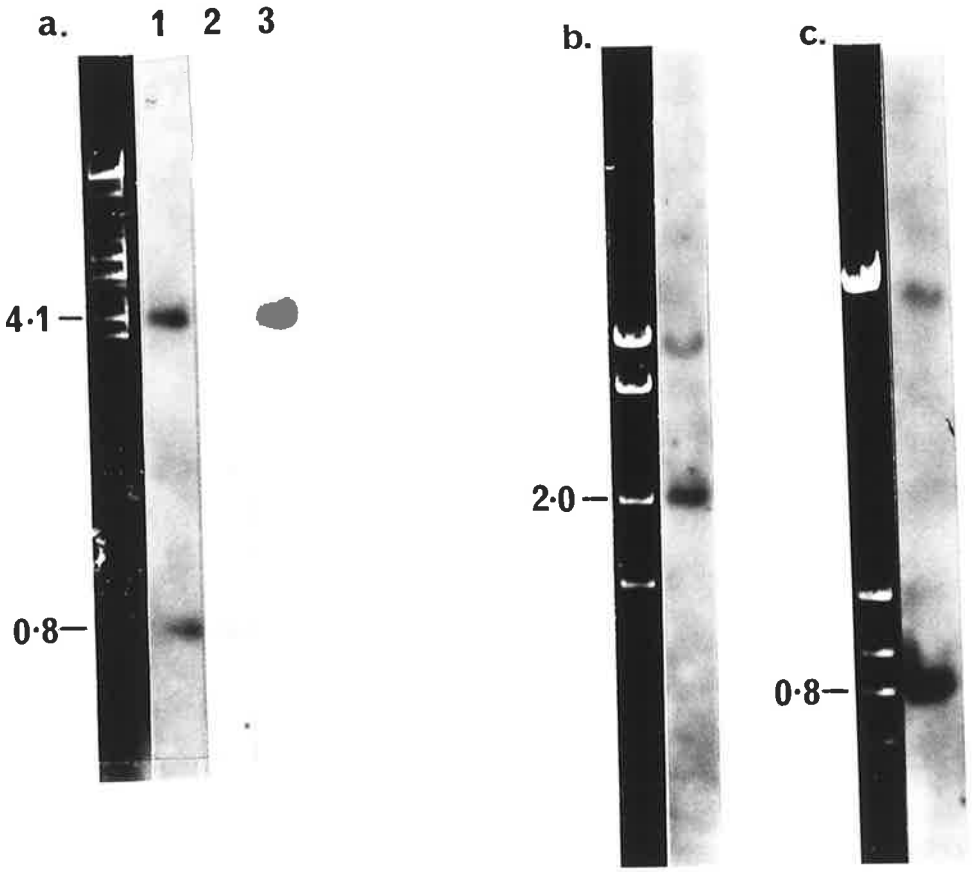
Probing Recombinants with ^{32}P-labelled Genomic DNA (see Figure 7.7)

a: λCH501 DNA digested with *Eco*RI and *Bam*HI; probed with,

1. total genomic DNA
2. 0.8 kb *Sac*II–*Xho*I fragment
3. 2.0 kb *Bam*HI–*Eco*RI fragment

b: pCH9.0E DNA digested with *Eco*RI and *Bam*HI and probed with total genomic DNA.

c: pCH7.0E DNA digested with *Sac*II and *Xho*I and probed with total genomic DNA.


a.
1
2
3
4·1
0·8
b.
2·0
c.
0·8

2.0 kb *Bam*HI-*Eco*RI fragment indicating that the repeat-sequence DNA is contained within this region. Similarly hybridisation to a *Sac*II/*Xho*I digest of pCH7.0E locates repeat-sequence DNA within the 0.8 kb *Sac*II-*Xho*I fragment (Fig. 7.8c).

These two fragments were purified and labelled by nick-translation for use as specific repeat-sequence probes. This allowed the relationship between repeats to be determined. These two fragments, when hybridised to the recombinants covering the λCH01 region, only detect the fragments from which they were derived (Fig. 7.7a). Neither fragment hybridises to the 6.8 kb *Eco*RI-*Eco*RI fragment which was also detected by ^{32}P-labelled genomic DNA. These results demonstrate that the three repeat-sequences detected in the λCH01 region are unrelated.

One fragment (4.5 kb *Eco*RI-*Hin*dIII) from the 6.1C region was detected faintly by labelled genomic DNA (Fig. 7.7f). The 0.8 kb *Sac*II/*Xho*I probe was specific for this fragment. Consistent with this result the 4.5 kb *Eco*RI/*Hin*dIII fragment in λH1.4, which represents the same chromosomal region, was also detected (Fig. 7.7g). The repeat-sequence DNA element in this region is presumably related to sequences found in the 0.8 kb *Sac*11-*Xho*1 fragment of recombinant 6.3C.

In the region covered by cosmid recombinants 2.1C and 5.1C there are three repeat-DNA sequences detected. The 6.9 kb and 5.6 kb *Eco*RI-*Eco*RI fragments, present in both recombinants, hybridise to nick-translated genomic DNA (Figs. 7.7b and 7.7c). Both of these fragments are detected by the 0.8 kb *Sac*II-*Xho*I fragment derived from cosmid 6.3C (Figs. 7.7d and 7.7e). The third sequence detected is contained in the 1.9 kb *Eco*RI-*Eco*RI fragment from 2.1C (containing the end of the insert) and the 6.2 kb *Eco*RI-*Eco*RI fragment from the same region of 5.1C. Both of these fragments are detected by ^{32}P-labelled 0.8 kb *Sac*II-*Xho*I fragment (Figs. 7.7d and 7.7e).

When total genomic DNA is used to probe λCH501, representing

the H5 genomic region, two distinct DNA fragments hybridise; notably the 4.1 kb *Bam*HI-*Bam*HI fragment located near the H5 gene (Fig. 7.8a). Hybridisation shows that this fragment contains sequences related to those present in the 2.0 kb *Bam*HI-*Eco*RI fragment. The smaller fragment (0.8 kb) from λCH501 which is also positive with labelled chicken DNA hybridises with the 0.8 kb *Sac*II-*Xho*I probe.

From this data it is clear that sequences repeated many times in the chicken genome are found interspersed with the histone genes. Two repeat sequences, A and B, are represented by the 0.8 kb *Sac*II-*Xho*I fragment, from pCH7.0E, and the 2.0 kb *Bam*HI-*Eco*RI fragment, isolated from pCH9.0E, respectively. Sequences related to the repeat-sequence DNA in these fragments are found in other histone gene containing regions characterised in this laboratory. A third, distinct, repeat-DNA sequence is located within the 6.8 kb *Eco*RI-*Eco*RI fragment of recombinant 6.3C. The results of these experiments are summarised in Figure 7.9.

7.2.5 Discussion

The aim of the work described in this chapter was to extend the analysis of the histone encoding regions to include their relationship to each other, and the nature of the regions between histone genes.

In addition to the region shown in Figure 6.25, containing λCHO1, a number of other histone gene containing regions have been characterised. These include the region described in section 6.2.15 containing the split H3 gene (cosmid recombinants 2.1C and 5.1C, Figure 6.24) and two regions containing core and H1 histone genes related to those in λCHO1 (cosmid recombinant 6.1C, λH1.1 and λH1.4; λH1.9 and λH1.10; see Figures 7.1 and 7.2).

To complete this analysis on the organisation of the histone genes it was necessary to determine whether all of these regions were independent. Using non-histone encoding hybridisation probes specific

Figure 7.9

Summary of Hybridisation Data

The figure shows a schematic representation of the DNA hybridisation data presented in this chapter. Fragments which hybridised to ^{32}P-labelled total genomic DNA are indicated with respect to the histone genes and are subdivided into those which hybridised to the 0.8 kb *Sac*II-*Xho*I fragment from pCH7.0E (marked A), the 2.0 kb *Bam*HI-*Eco*RI fragment from pCH9.0E (marked B) and fragments which did not hybridise to either of these fragments (marked C).

A 0.8 kb fragment from λCH501 hybridised to the 0.8 kb *Sac*II-*Xho*I probe derived from pCH7.0E. However, there are two fragments of this size in λCH501 (marked x in d) and it has not been determined which of these cross-reacts with this probe.

The 2.2 kb *Eco*RI-*Eco*RI fragment from 2.1C and 5.1C was positive with cDNA made to poly (A) plus, 5 day chicken embryo RNA (indicated in b).

Restriction sites are indicated by the letters shown below. Sizes of hybridising DNA fragments are given in kilobases. The *Xho*I restriction site marked + in pCH7.0E (a) is not present in 6.3C DNA. Mapping data in (c) is from Lesnikowski (1983) and L.S. Coles (personal communication). Restriction mapping data for λCH501 is from Krieg *et al*, (1983).

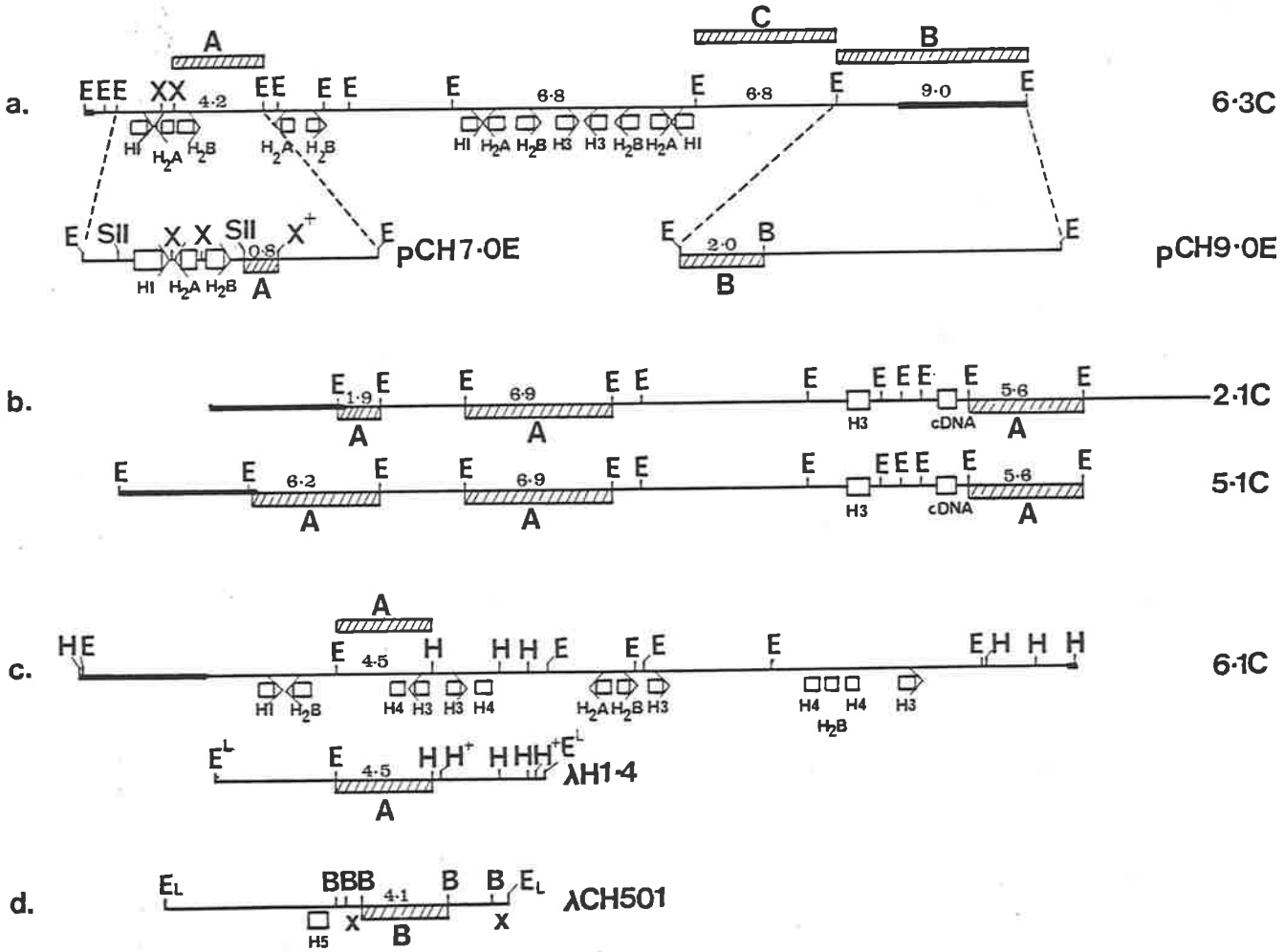
E: *Eco*RI

H: *Hin*dIII

B: *Bam*HI

X: *Xho*I

SII: *Sac*II


a.
A
C
B
EEE XX 4·2 EE E E E 6·8 E 6·8 E 9·0 E
HI H2A H2B H2A H2B HI H2A H2B H3 H3 H2B H2A HI
E SII X X SII X+ E
0·8
HI H2A H2B A
pCH7·0E
E 2·0 B E
B
pCH9·0E
6·3C
b.
E 1·9 E E 6·9 E E E E E E E 5·6 E
A A H3 cDNA A
2·1C
E E 6·2 E E 6·9 E E E E E E E 5·6 E
A A H3 cDNA A
5·1C
c.
A
HE E 4·5 H H H E E E E EH H H
HI H2B H4 H3 H3 H4 H2A H2B H3 H4 H4 H3 H2B
EL E 4·5 H H+ H HH+ EL
A
λH1·4
6·1C
d.
EL BBB 4·1 B B EL
H5 X B X
λCH501

for each region it was shown that these regions are non-overlapping.

Another significant question, concerning the structure of a multigene locus, involved the nature of intergenic DNA sequences. These may function as non-expressed spacer DNA or contain repeated DNA sequences or sequences recognised by DNA binding proteins. It was also possible that some regions encoded non-histone proteins, possibly expressed co-ordinately with the histone genes.

The universal occurrence of repeat-sequence DNA suggested the possibility that repeat-sequence DNA elements could be located in regions between the chicken histone genes. A number of repeated DNA sequences were detected when chicken histone gene containing recombinants were hybridised to *in vitro* labelled total chicken DNA. This method is reported to detect sequences repeated fifty fold or more in the genome (Shen and Maniatis, 1980). Consistent with this is the finding that histone gene sequences (10 fold reiteration frequency) were not detectable. Further hybridisation analysis with specific probes identified two major, unrelated repeats interspersed with the histone genes. A third, unrelated repeat sequence is represented only once in the regions analysed in a single *Eco*RI-*Eco*RI fragment contained within the cosmid recombinant 6.3C.

Until these individual repeat sequence elements are isolated and characterised it is not possible to comment on the relationship of these sequences to other interspersed repetitive sequence families studied in eukaryotes (Jelinek and Schmid, 1981).

In a preliminary attempt to determine whether any intergenic regions are transcribed by polymerase II *in vivo*, i.e. contain potential genes, DNA from the representative recombinants was hybridised to cDNA prepared from total poly (A) plus 5 day chicken embryo RNA. A single DNA fragment, from the region containing the split H3 gene, appears to give rise to a non-histone transcript. Further investigation is

necessary to identify the nature of this 'gene'.

The overall relationships of core and H1 genes in the chicken genome is discussed in Chapter 8.

CHAPTER 8

FINAL DISCUSSION

8.1 ORGANISATION OF CHICKEN HISTONE GENES

8.1.1 Gene Number

In this work, twenty-one histone genes were located in a region of DNA spanning the recombinant λCHO1, and covering approximately 70 kb of DNA. Together with two other characterised regions (Figures 7.1 and 7.2 see D'Andrea *et al*, 1984) 40 core and H1 genes have been positioned.

In addition, two overlapping cosmids containing only the split H3 gene (Engel *et al*, 1982) were isolated and characterised (see Chapter 6). I have not attempted exhaustively to establish whether there is linkage between any of these regions, however using probes specific for the ends of each characterised region it was possible to show that these stretches of histone gene containing DNA did not overlap (see Chapter 7). Preliminary analysis of another cosmid recombinant, 4.2C, containing an H_2A and an H3 gene shows this clone to be independent of these regions (data not shown). This recombinant has not been mapped in detail. The regions characterised also appear to be unlinked to the erythroid specific H5 gene (see Chapter 7; Krieg *et al*, 1983, Ruiz-Carrillo *et al*, 1983).

It was originally estimated that the histone genes were reiterated approximately 10 times each per chicken haploid genome (Crawford *et al*, 1979). Estimates of histone gene copy number have recently been made, from genomic Southerns, by Ruiz-Carrillo *et al*, (1983). They predicted the maximum number of copies of each histone gene per haploid genome to be H1(6), H_2A(8), H_2B(9), H3(11) and H4(7). In the recombinants detailed (including 4.2C and excluding variants) the histone genes are represented as follows H1(6), H_2A(10), H_2B(8), H3(10), H4(8), totalling 42 core and H1 histone genes. The gene numbers observed are thus consistent with previous estimates of gene copy number, suggesting that essentially

all the non-variant core and H1 histone genes are present in the recombinants described (Figs. 6.25, 7.1 and 7.2).

Comparison of the mapping data presented here and the patterns observed on extensive genomic southern analysis using an H1 gene probe (L.S. Coles, unpublished; D'Andrea *et al*, 1984) shows that all the H1 genes in the chicken genome are present in these clones. The restriction enzyme data presented in Figures 6.25, 7.1 and 7.2 is consistent with the restriction enzyme patterns observed in the above genomic Southern analysis (see D'Andrea *et al*, 1985) and with restriction enzyme patterns observed after Southern analysis with core histone gene probes (Ruiz-Carrillo *et al*, 1983; Harvey, 1982; L.S. Coles, unpublished; see Fig. 7.6).

8.1.2 Gene Organisation

Two extreme patterns of organisation have been established in eukaryotic multigene families viz; genes can either be clustered or dispersed, i.e., scattered at different chromosomal locations. It is clear that such patterns are not static and that clustered gene families can give rise to individual dispersed members, that is, orphons (Childs *et al*, 1981; Leder *et al*, 1981). The histone gene system like many other gene families, is composed of distinct subfamilies (i.e.; H1, H_2A, H_2B, H3 and H4 genes). In cases where subfamilies can be distinguished two gene arrangements have been found.

a) $\alpha_1\alpha_2/\beta_1\beta_2$; that is, genes of a particular subtype clustered together and separate from genes in another subfamily. The globin genes are arranged in this manner, with α-globin and β-globin genes existing in seperate clusters. (Efstradiatis *et al*, 1980; Jeffreys, 1982).

b) $\alpha_1\beta_1/\alpha_2\beta_2$; that is, the genes for the different subtypes are intermingled and arranged in clusters. The A and B chorion proteins (Jones and Kafatos, 1980) and the histone genes of most organisms (Maxson *et al*, 1983a; Hentschel and Birnstiel, 1981; see section 1.6) are examples

of gene families in this category.

The work described in this thesis involved a detailed analysis of the chicken histone gene system and showed that the histone genes for the separate subtypes are intermingled, and arranged in clusters which display a large degree of disorder. There is no evidence of a repeating unit, although some preferred associations are obvious (section 8.1.3). These conclusions are consistent with other data presented previously (Engel and Dodgson, 1981; Harvey *et al*, 1981; Sugarman *et al*, 1983; Ruiz-Carrillo *et al*, 1983).

The organisation observed in the chicken bears little resemblance to the repeated quintet structures observed in sea urchin, *Drosophila*, newt or *Xenopus*, but is typical of the vertebrate histone gene organisation (see section 1.6). Similar organisation has been reported for mouse (Marzluff and Graves, 1984) and human (Stein *et al*, 1984) histone gene systems.

It is clear that the organisation of histone genes is closely related to the reiteration frequency of the genes. It appears that clustering of the histone genes has been maintained throughout evolution, while the tandem quintet arrangement is only correlated with high histone gene copy number (see section 1.6.9), and may be a consequence of recombinational events which operate amongst highly repeated tandem genes (Maxson *et al*, 1983c; Old and Woodland, 1984; see section 1.6.9).

It is not clear why the histone genes have remained clustered in organisms with relatively few histone genes. Possibly there is some unknown selective pressure, relating to the co-ordinate expression or regulation of the histone genes, which maintains the genes in clusters.

8.1.3 Preferred Associations of Histone Genes

Although no overall repeat structure is discernable in the chicken histone gene system, the organisation is not totally random. Six of the eight H_2A genes, shown in Figures 6.25, 7.1 and 7.2 are closely

associated with an H_2B gene and in these cases each gene pair is divergently transcribed (R. Sturm, personal communication). This arrangement may be related to balanced expression of the two genes, which in these cases possibly share regulatory elements. This association is also observed in mammals, yeast, some sea urchin late gene clusters and amphibians (see Old and Woodland, 1984; Maxson *et al*, 1983a). The significance of this association is unclear as it is not strictly maintained. It is interesting that in yeast the H_2A gene has an essential regulatory enhancer sequence on the 3' side of the H_2B gene (Osley and Hereford, 1982).

In addition to this association there is a non-random association of H1 genes with H_2A and H_2B genes, in fact five of the six chicken H1 genes show such a linkage. The relative order of these three genes varies however, with H_2A being central in three cases and H_2B and H1 being central in the other two clusters.

In total, ten H3 genes have been mapped in this laboratory (see Figures 6.25, 7.1 and 7.2) and in six cases these are adjacent to H4 genes.

Thus, although there is not an overall repeating structure obvious for the histone genes of the chicken, it is apparent that these histone genes broadly reflect the situation observed in other organisms, that is H_2A occurring with H_2B genes and H3 occurring with H4 genes (Maxson *et al*, 1983a). The fact that there are exceptions to all of these preferred associations within the chicken histone gene system suggests that the expression of individual genes is not entirely dependent on any precise linkage. Consistent with this, the sequencing of the histone gene cluster contained in pCH8.4E, which has genes in the atypical order H4, H_2A, H3, H_2A, H4, indicates that all of these genes have the highly conserved 5' and 3' elements presumed necessary for function (Wang *et al*, 1985). There is also evidence that at least one of these

H4 genes is expressed (Sugarman *et al*, 1983). Possibly the observed preferred arrangements are remnant features of a common ancestral cluster.

8.1.4 Restriction Enzyme Site Polymorphisms

Allelic base differences, present in a population at polymorphic frequencies (greater than 0.1%) can lead to different restriction enzyme digestion patterns on Southern analysis of DNA from various individuals. A number of polymorphic restriction sites are associated with the human globin genes (Antonarakis *et al*, 1982). Using γ- and β-globin probes and various restriction enzymes Jeffreys (1979) has estimated the incidence of polymorphism at the DNA level to be 1 per 100 nucleotides, outside of the coding regions, in man.

Restriction enzyme site polymorphisms in chicken DNA have been described by Sugarman *et al*, (1983). DNA was isolated from seven individual chickens and the *Hin*dIII restriction patterns analysed using an H3 histone gene-specific probe. This analysis showed that the H3 genes were in identical chromosomal positions in different individuals and indicated the existence of polymorphic restriction enzyme sites.

The work described in this thesis has allowed the comparison of DNA cloned from different individuals, but representing the same region of the genome. Clones representing the same region of the genome are identical in structure in the overlapping regions except for some restriction enzyme sites which are present in one recombinant but absent in the other (shown in Figs. 6.25, 7.1 and 7.2). Two *Hin*dIII sites present in λH1.4 are not present in the cosmid recombinant 6.1C (Fig 7.1), and a *Xho*I restriction site present in λCHO5 and λCHO7 is not present in 6.3C (Fig 6.25). Furthermore, polymorphic restriction sites have been identified in overlapping clones which were isolated from a library of DNA fragments generated from a single individual (data not shown). In these cases the diploid individual is heterozygous for these polymorphic restriction sites.

A most important conclusion from the genomic blot comparison carried out by Sugarman *et al*, (1983) is that histone gene structure is stable and probably identical in individual chickens. The differences which were observed between chickens on digestion and Southern analysis with histone probes can be accounted for by restriction enzyme site polymorphisms.

8.1.5 Spacer DNA

Preliminary experiments have been carried out to analyse the nature of the spacer DNA between chicken histone gene clusters. Some of these regions hybridised to total ^{32}P-labelled chicken DNA under conditions which detect repeat-sequence DNA (Shen and Maniatis, 1980). Further analysis of these regions suggested three distinct repeat sequences are present, although the nature of these is unknown (see Chapter 7). Stretches of alternating co-polymeric DNA have been found in the spacer DNA of the early histone genes in sea urchin (Schaffner *et al*, 1978; Sures *et al*, 1978) and *Drosophila* (Goldberg, 1979). These sequences do not appear to exert any effect on the expression of the genes but may have some role in the evolution of the gene family (Kedes, 1979). Possibly they serve as sites for recombination events between repeats on the same or different chromosomes. Local melting of these simple sequences could initiate the formation of DNA hybrids, involving homologous sequences from elsewhere in the genome, and thus lead to recombination events (Hentschel, 1982; see Maxson *et al*, 1983a).

Sierra *et al*, (1982) found that some human histone genes are interspersed with multiple copies of the *Alu* DNA family (reviewed in Jelinek and Schmid, 1982). A transposon-like sequence which is highly repeated in *Xenopus* genomic DNA has been found upstream of an H3 gene in *Xenopus* (Van Dongen *et al*, 1984). Related sequences have since been found in other cloned *Xenopus* histone gene clusters (Van Dongen *et al*, 1984). In the newt, *Notophthalmus*, most of the histone genes are

in the form of tandem repeats which are separated by up to 50 kb of highly repetitive spacer DNA (Stephenson *et al*, 1981). Thus there are precedents for repeat-sequence DNA occurring in association with histone genes in both primitive and higher eukaryotes but the significance of this is unknown. Further analysis is required to identify the repeats associated with the chicken histone genes.

In other experiments (see Chapter 7) it was found that a rare protein-coding gene may be associated with the split H3 gene. Probing all recombinants with ^{32}P-labelled cDNA, made from poly (A) plus 5-day chicken embryo RNA, failed to detect any hybridising sequences linked to the *"embryonic"* genes, shown in Figures 6.25, 7.1 and 7.2. Embryonic RNA will not be representative of all chicken mRNA and so it is still not known whether the regions between these histone genes encode any other genes.

A single fragment in close proximity to the variant H3 gene (cosmids 2.1C and 5.1C; Chapter 6) was positive with this ^{32}P-labelled cDNA probe. Further work is required to detect any transcript from this region and to identify its function. The cDNA probe did not detect the split H3 gene but did detect the H5 gene present in another recombinant (fragments encoding non-polyadenylated histone transcripts were not detected with this probe).

Little is known of non-histone encoding genes linked to the histone genes in any species. Sierra *et al*, (1982) found a histone gene containing clone which also specified an RNA of approximately 330 nucleotides. This RNA was synthesised only during the G1 phase of the cell-cycle. The coding capacity of this RNA is still undetermined.

8.2 EXPRESSION OF THE CHICKEN HISTONE GENES

The aim of this project was purely to determine the spatial arrangement of the histone genes in the chicken and no attempt was made to examine the expression of these genes. The expression of a number

of these genes has been analysed elsewhere, but the results obtained are difficult to interpret.

Many of the core and H1 genes represented in this work (see Figs 6.25, 7.1 and 7.2) have been partly or completely sequenced (Grandy *et al*, 1982; Wang *et al*, 1985; Coles and Wells, 1985; J. Powell, R. Sturm, unpublished). From these data it is clear that the genes for a particular subtype within this group of clones are generally isocoding. Probes representative of these genes have been hybridised to RNA isolated from embryonic and adult chicken (Engel, 1984; Sugarman *et al*, 1983). Each of the probes gave an identical pattern, that is, each was specific for a single electrophoretic form of RNA from early embryonic tissues and did not hybridise to adult derived RNA. Given the similar amino-acid sequences of genes of a particular subtype in this set of clones, and the parallel expression patterns observed, it is likely that all the genes in this set of recombinants are expressed in the same fashion and thus represent a particular class of co-ordinately regulated histone genes.

The patterns observed on Northern analysis can be explained in two ways. Firstly these genes may represent an "*embryonic*" class of histone genes which is expressed primarily in embryonic tissues. Alternatively, these genes may be expressed throughout development and coupled to S-phase. In the latter case, the higher concentrations of histone message in embryonic RNA would represent the presence of many more dividing cells in the embryonic tissue samples. Evidence from this laboratory (S. Dalton, unpublished) indicates that expression of the H_2A gene in pCH7.0E is tightly coupled to S-phase in synchronised chicken ts34 cells (H5 shows no coupling).

In contrast to the results obtained for this major class of histone genes it was found that the split H3 gene (contained in 2.1C and 5.1C) hybridised to both embryonic and adult chicken RNA (Engel,

1984). This gene may be an example of an *"adult"* histone gene, which replaces the *"embryonic"* genes in later development. Alternatively, it may represent a low abundance, cell-cycle independent histone gene expressed throughout development. In fact, this H3 gene encodes a protein which is very similar to a minor, replication independent histone variant found in somatic adult chicken tissue (Engel *et al*, 1982).

Most of the genes contained in the region spanning λ CHO1 have not been analysed to determine whether they are expressed. However, extensive sequence analysis, from this laboratory, of flanking and coding regions of twenty seven chicken histone genes suggest these are capable of expression. The possibility that some of these genes are pseudogenes cannot be excluded. Histone pseudogenes have been reported for human (Marishi *et al*, 1984) and *Xenopus* (Turner *et al*, 1983).

8.3 EVOLUTION OF THE CHICKEN HISTONE GENES

8.3.1 Homogeneity

As discussed in section 1.2.3, gene conversion is expected to be the most frequent type of homogenising recombination involving members of a dispersed family of genes, such as the chicken histone genes. However the rate of gene conversion in a dispersed gene family must be quite low compared with the frequency of recombination events between highly repeated quintets.

A number of histone gene regions which were isolated and mapped in this work have now been sequenced (Wang *et al*, 1985; L.S. Coles, J. Powell, R. Sturm, unpublished). In addition to these genes, DNA sequence has been determined for genes previously identified in λCHO1 or λCHO2 (Harvey *et al*, 1982; Coles and Wells, 1985; D'Andrea *et al*, 1981) and for genes from chicken histone clones isolated in other laboratories (Sugarman *et al*, 1983; Grandy *et al*, 1982). Comparisons of coding and flanking regions of these genes has revealed considerably heterogeneity. The untranslated regions and promoter regions of isocoding genes show

little homology apart from conserved promoter elements required for expression (see Maxson *et al*, 1983a; Hentschel and Birnstiel, 1981) and a number of conserved subtype-specific sequence elements (Harvey *et al*, 1982; Coles and Wells, 1985; Sugarman *et al*, 1983). Comparison of the coding regions of iso-coding histone genes indicates that numerous base changes have occurred, generating sequence heterogeneity but, in general, maintaining the amino-acid sequence of particular subtypes.

So far, there is no evidence for conversion events between isocoding, non-allelic chicken histone genes, but this is not surprising given the organisation and reiteration frequency observed. The sea urchin "late" histone genes are organised in a similar fashion (Childs *et al*, 1982; Maxson *et al*, 1983b), however in this case, while the sequences surrounding the coding regions are not conserved, there is strong DNA sequence homology between the coding regions of some iso-coding genes (Roberts *et al*, 1984). It is suggested that conversion has acted separately on individual coding regions, maintaining homogeneity in these regions while allowing untranslated regions and spacer regions to diverge (Roberts *et al*, 1984).

Sequence analysis of pCH8.4E (Wang *et al*, 1985) has revealed striking homology between the two H_2A-H4 gene pairs present in this clone (see Figure 8.1). DNA sequence analysis of pCH11.5E (this work) shows there is also striking homology between the 5' regions of the two central H3 genes within this fragment (Chapter 6). It is clear from these results that recent recombination events (see below) have generated symmetrical structures (section 8.3.2) and have led to sequence homogeneity within small histone gene clusters.

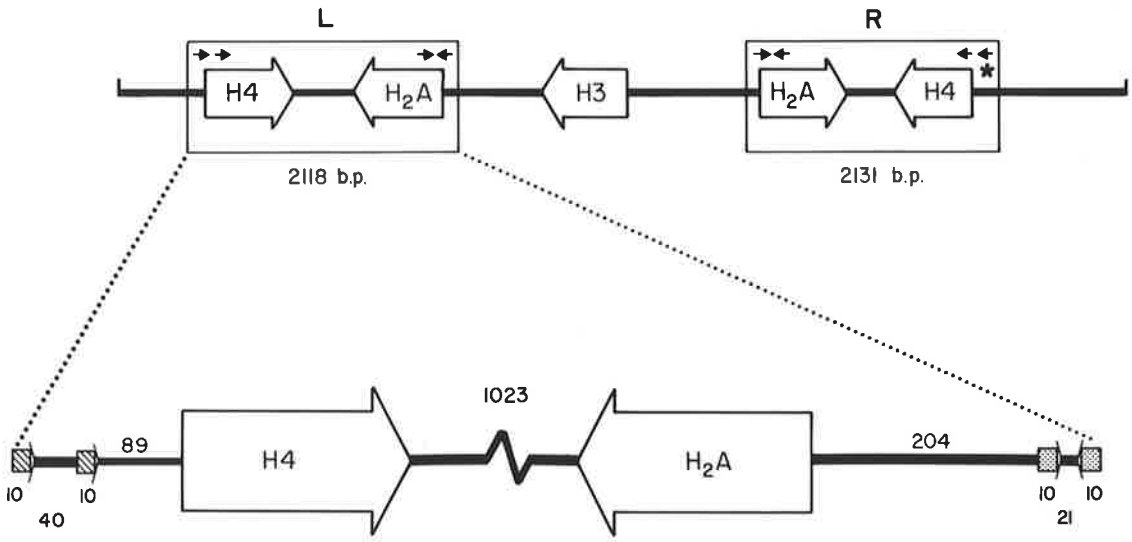
8.3.2 Symmetry Within Histone Gene Clusters

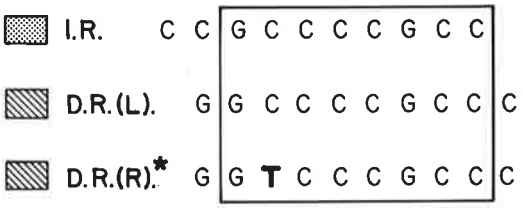
The detailed restriction mapping of subclones containing individual histone gene clusters revealed two regions containing symmetrically arranged enzyme recognition sites. In these cases clustered

Figure 8.1

Structure of pCH8.4E

The area covered by the inverted duplication is boxed. The position and orientation of the direct and inverted repeats is shown relative to the histone genes and the boundaries of the duplicated regions. The octamer sequence shared by the direct repeats (DR) and the inverted repeats (IR) is shown. Distances are indicated in base-pairs.


L
R
H4
H2A
H3
H2A
H4
*
2118 b.p.
2131 b.p.
1023
89
H4
H2A
204
10
10
40
10
10
21


I.R. C C G C C C C G C C
D.R.(L). G G C C C C G C C C
D.R.(R).* G G T C C C G C C C

histone genes were found as an inverted duplication centred around H3 genes. Two of these arrangements were found during this work and are shown in Figure 6.25 viz., pCH8.4E (H4. H_2A. H3. H_2A. H4) and pCH11.5E (H1. H_2A. H_2B. H3. H3. H_2B. H_2A. H1). Another region containing H3 genes and representing a possible inverted duplication is shown in Figure 7.1 as part of pCH22.0B (H4. H3. H3. H4). The fact that this inverted histone gene symmetry appears three times in the chicken genome suggests there may be either some form of positive selection or a predisposition for this type of arrangement.

The only symmetrical region which is fully characterised is pCH8.4E. In this case DNA sequence analysis (Wang *et al*, 1985) shows that a 2.1 kb region is almost exactly duplicated in the reverse orientation (97% homology - indicated by the boxes in Figure 8.1). Outside this duplicated region the sequences diverge considerably, that is, the boundaries of the duplication are well defined.

In pCH11.5E it appears that approximately 5 kb of DNA have been duplicated and inverted. Limited DNA sequencing from this clone (Chapter 6) suggests the duplicated regions are highly homologous.

As the complete sequence of the pCH8.4E insert had been determined (Wang *et al*, 1985), it was possible to look in this cluster for sequences which may have been involved in the generation of the symmetrical structure.

A significant feature is that the precise boundaries of the duplicated region are characterised by a ten base pair direct repeat at the H4 end and a ten base pair inverted repeat at the H_2A end of each 2.1 kb region (Fig. 8.1). These repeats are closely related to each other and share the eight base pair sequence 5' GCCCCGCC 3'. In the case of the inverted repeat, the sequence is, itself, a short direct repeat 5' CCGCC CCGCC3'. The direct and inverted repeats are separated by forty and twenty-one base pairs of DNA respectively.

A Possible Recombination Event

There is little information concerning the generation of naturally occurring inversions in eukaryotes. Vitelli and Weinberg (1983) characterised a duplication/inversion, in an unusual sea urchin histone gene cluster, which has similar features to those found in pCH8.4E. They suggest that a recombination event between two duplexes, arranged alongside each other with histone gene coding regions oriented in opposite directions ("reverse duplex pairing"), may generate an inverted gene arrangement. After comparing the sequence of parent molecules they concluded that very little homology is necessary for this type of event. A similar event has been suggested to explain the arrangements seen in integrated viral sequences in a polyoma transformed cell line. (Ruley and Fried, 1983). In this case it appears that multiple illegitimate recombination events involving very short homologous stretches (3-4 base pairs) have occurred within small regions. The inverted sequence arrangements observed were probably generated by recombination events between oppositely oriented viral molecules before integration into host DNA.

I speculate that the symmetrical arrangements observed in some chicken histone gene clusters were generated by recombination between paired short repeat sequences during "reverse duplex pairing". Figure 8.2 illustrates the possible role of the short repeat DNA sequences and the type of recombination event proposed to generate the inverted duplications.

The degree of sequence conservation within the duplicated regions of pCH8.4E (Wang *et al*, 1985) and pCH11.5E (see Chapter 6) suggest that the recombination events involved occurred relatively recently.

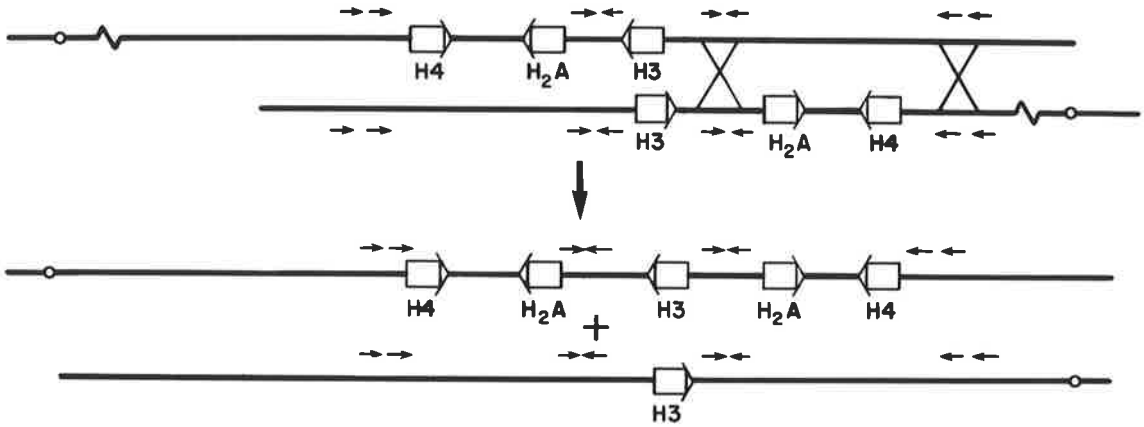
In the sea urchin duplicated inversion (Vitelli and Weinberg, 1983) the recombination event has occurred in the region between the

Figure 8.2

Generation of Inverted Duplications by Recombination at Paired Repeat Sequences During "Reverse Duplex Pairing"

The diagram illustrates the event required to generate the type of structure seen in pCH8.4E.

Small repeat sequences are indicated by arrows above the line representing each duplex. Gene regions are indicated by arrows.


H4
H2A
H3
H3
H2A
H4
H4
H2A
H3
H2A
H4
+
H3

TATA box and the transcription initiation site of H1 and H4 genes. The end points of the duplicated region in pCH8.4E are 90 base pairs from the H4 TATA box and 158 base pairs from the TATA box of the H_2A gene.

The conserved sequence element found as a direct repeat or inverted repeat at the end of the duplicated region in pCH8.4E may, itself, form part of the promoter regions of the H_2A and H4 genes. Repeat-sequence elements have been postulated as binding sites for trans-activating regulatory factors. In some cases direct evidence exists for such interactions (Parker and Topol, 1984). Furthermore, sequence elements with strong homology to known modulators of transcription are found between the short repeat sequences of pCH8.4E (Wang *et al*, 1985).

The above observations suggest that the recombination event which produced the inverted symmetry has occurred in, or close to, the promoter regions of these histone genes. It is possible that promoter elements provide a DNA structure susceptible to recombination events, possibly because the duplexes are unwound in these regions thus allowing interduplex base pairing (Vitelli and Weinberg, 1983).

8.3.3 Secondary Structure-Induced Recombination

The sequence found at the H_2A boundary of the duplication in pCH8.4E is an inverted repeat and thus potentially can form a stem-loop structure. Models have been postulated which implicate secondary structure in the initiation of recombination events (Sobell, 1975; Wagner and Radman, 1975; Ripley and Glickman, 1982). Secondary structures could lead to preferential nicking of DNA strands and thus promote recombination events. It is also possible that secondary structures are involved in the termination of recombination (Todd and Glickman, 1982).

Ripley and Glickman (1982) present a model involving strand-switching, mediated by palindromic sequences, during DNA synthesis. This type of event would give rise to a large inverted repeat of the type seen in the chicken symmetrical histone gene clusters, but a model

of this type would not be expected to generate the same sequence at both end points as seen in the pCH8.4E duplicated/inversion.

8.3.4 Transposition Associated Recombination

The overall structure of the H_2A-H4 duplication in pCH8.4E shows some similarities to an inversion associated with transposition (see Cohen and Shapiro, 1980; Shapiro, 1979; section 1.2.4).

In this case the duplicated region would represent a transposable element. There are three reasons why I do not consider the three H3 associated symmetrical duplications to be associated with transposition events. Firstly the duplicated regions have not been found elsewhere in the genome i.e. H_2A-H4 pairs only occur in pCH8.4E. Secondly there is no short duplication outside the boundary of the H_2A-H4 regions as would be predicted if this arrangement was associated with transposition (see Shapiro, 1979; Cohen and Shapiro, 1980). Furthermore, the duplicated region does not have an inverted repeat at both ends and therefore does not resemble a typical transposon in structure.

It is also unlikely that any of the symmetrical arrangements arose in *E. coli*, before or after the construction of the λ or cosmid genomic libraries. A number of sequence differences are observed between the left and right duplicated regions in pCH8.4E (Wang *et al*, 1985) and this would not be expected given recombination of amplified copies in *E. coli*. Furthermore, the divergent H3 pair found in pCH22.0B, which was generated from cosmid clone 6.1C, has also been found in the overlapping recombinants λCH02 and λH1.4 (see Fig. 7.1) which were isolated from the λ-genomic library, constructed from a different chicken.

Genomic Southern bolt analysis with histone gene probes (D'Andrea *et al*, 1985; Ruiz-Carrillo *et al*, 1983; Harvey, 1982; L.S. Coles, unpublished) reveals bands of molecular weights predicted from mapping of the various clusters (Figs. 6.25, 7.1 and 7.2).

8.3.5 Cellular Processes & The Rearrangement of Histone Genes

In this section processes involved in transcription and DNA replication are considered and lead to speculation that there is a connection between the cell-cycle regulation of histone gene transcription and the recombination events involved in the generation of prevalent symmetrical histone gene arrangements.

It was noted in section 8.3.2 that the cross-over sites for the postulated recombination event leading to the structure observed in pCH8.4E occur at specific sequences within the promoter regions of the H_2A and H4 genes in this clone. This association of the predicted cross-over sites with histone gene promoter elements suggests the possibility that proteins (possibly sequence-specific) or processes affecting transcription are involved in the recombination event.

Because transcription of the histone genes is tightly coupled to DNA replication (Maxson *et al*, 1983a and 1983c), it is probable that regulatory factors are associated with these genes during S-phase. Possibly interactions between regulatory factors, either directly or mediated through other molecules, such as RNA polymerase, bring promoter sequences into juxtaposition so that recombination can occur. It has been postulated (Ohtsubo and Ohtsubo, 1978) that RNA polymerase may play an important role in site-specific recombination. Vitelli and Weinberg (1982) have speculated that the basis of many eukaryotic rearrangements may be the fortuitous apposition of small regions of homology which have particular secondary structure due to interaction with protein.

Hereford *et al*, (1982) have shown that activation of histone gene transcription in yeast occurs in late G1 phase just prior to initiation of DNA replication. To account for this control of histone gene transcription they propose a model in which changes in chromatin structure precede the initiation of replication and also activate histone gene

transcription.

In yeast, the histone genes are in close proximity to autonomously replicating DNA sequences (*ars* sequences) (Osley and Hereford, 1982) possibly lending support to this model. If this association is common to other species, including chicken, it is conceivable that there are events involved in the initiation of DNA replication, which involve changes in DNA topology in the vicinity of the replication origin and the histone genes, and which may predispose these regions to genomic rearrangements.

Changes in DNA topology such as DNA supercoiling have a well documented association with the initiation of DNA replication in prokaryotes (reviewed in Gellert, 1981) and DNA supercoiling has been implicated in eukaryotic replication initiation (Mattern and Painter, 1979). Supercoiling of the region surrounding the origin of replication presumably assists the unwinding of DNA strands and favours binding of factors involved in replication initiation. Liu *et al*, (1980) propose that the initiation of DNA replication in bacteriophage T4, and possibly in eukaryotes, involves a topo-isomerase, and perhaps additional factors, functioning as a site-specific DNA Gyrase. They further suggest that this process may involve topo-isomerase recognition of specific sequences either side of the origin of replication. In this way, supercoiling would be induced in a local domain at the replication origin facilitating formation of a replication bubble.

The degree of supercoiling is also known to differentially activate promoters in prokaryotes (Cozzarrelli, 1980; Smith, 1981; Gellert, 1981; Fisher, 1984) and has a suggested role in the regulation of eukaryotic genes (Luchnik *et al*, 1981; Weisbrod, 1982; Smith, 1981; see section 1.3.4).

It is conceivable then, that supercoiling may be involved in the co-ordinated initiation of DNA replication and activation of histone gene transcription, particularly if there is a direct association between

the histone genes and replication origins.

The proteins that are known to be involved in altering DNA topology include gyrases and topo-isomerases (in prokaryotes) and topo-isomerases in eukaryotes (Gellert, 1981; see section 1.3.4). These proteins have also been implicated in recombination events. It is proposed that DNA Gyrase is involved in illegitimate recombination in prokaryotes (Ikeda *et al*, 1981, 1982; Marvo *et al*, 1983) and eukaryotic type I topo-isomerases have been shown to catalyse strand exchange reactions (Been and Champoux, 1981; Halligan *et al*, 1982; Dean *et al*, 1982). Because topo-isomerases are strand nicking enzymes (Gellert, 1981) they could play an important role in illegitimate recombination events.

In summary it is speculated that a topoisomerase - like enzyme (possibly sequence-specific) is involved in pre-replicative chromatin alterations and that this may,

1) provide a means of co-ordinating rapid and efficient expression of histone gene clusters with the initiation of DNA replication (late G1 phase) possibly via an association between histone gene clusters and origins of replication.

2) explain recombinational events leading to duplicated inversions and the scrambled nature of the histone gene organisation in the chicken.

8.4 FINAL SUMMARY

The chicken core and H1 genes represent an attractive system for studying the organisation and transcriptional control of a multigene family. The chicken histone genes have evolved sufficiently to allow useful DNA sequence comparisons to be made between isocoding genes and the total number of genes to be analysed is within practical bounds. The data presented here represents the first detailed picture of a vertebrate dispersed histone gene family. At the organisational level, the most obvious feature is the lack of a tandem repeat structure and one inference is that individual genes are regulated independently of

their chromosomal location. However, I have noted preferred associations and symmetrical arrangements within clusters which possibly have functional significance for co-ordinate expression. If transcription of each cluster in the chicken genome is regulated in a similar fashion to the yeast histone genes (Hereford *et al*, 1982) the timing of transcription of genes within a cluster may be related to the activation of an adjacent *ars* sequence. As many of the iso-coding chicken histone genes have characteristic 5' untranslated regions (Maxson *et al*, 1983a) it is possible to study the expression of individual histone genes in relation to cellular events such as the cell-cycle. Experiments are underway in this laboratory to detect chicken *ars* sequences, if present, in the described histone gene regions. In addition, investigations are being carried out to determine the expression profile of a number of these histone genes in relation to the cell cycle.

Whether expression of individual chicken genes is modulated by adjacent genes or is related to the composition of a cluster is not known at present. Clearly the organisation of the chicken histone genes, although disordered, is not entirely random. It is clear also, that an ordered array of genes is not required for physiologically acceptable, balanced expression of a multigene family.

BIBLIOGRAPHY

Adams, J.M. and S. Cory (1983) **in** "Eukaryotic Genes: Their Structure Activity and Regulation", Maclean, N., Gregory, S.P., and R.A. Flavell, eds., Butterworths and Co., Cambridge U.K., pp. 343-364.

Adesnick, M. and J.E. Darnell (1972) J. Mol. Biol., **67,** 397-406.

Anderson, K.V. and J.A. Lengyel (1980) Cell, **21,** 717-727

Anderson, K.V. and J.A. Lengyel (1984) **in** "Histone Genes: Structure, Organisation and Regulation", Stein, G.S., Stein, J.L. and W.F. Marzluff, eds., J. Wiley and Sons, New York, pp. 135-161.

Antonarakis, S.E., Boehm, C.D., Giardina, P.J.V. and H.H. Kazazian (1982) Proc. Natl. Acad. Sci. U.S.A., **79,** 137-141.

Appels, R., Wells, J.R.E. and A.F. Williams (1972) J. Cell. Sci, **10,** 47-59.

Arceci, R.J. and P.R. Gross (1977) Proc. Natl. Acad. Sci. U.S.A., **74,** 5016-5020.

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

Baer, B.W. and D. Rhodes (1983) Nature, **301,** 482-488.

Baker, E.J. and A.A. Infante (1982) Proc. Natl. Acad. Sci. U.S.A., **79,** 2455-2459.

Baltimore, D. (1981) Cell, **24,** 592-594.

Banerjee, A.K. (1980) Microbiol. Rev., **44,** 175-205.

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

Bauer, W.R. (1978) Ann. Rev. Biophys. Bioeng., **7,** 287.

Been, M.D. and J.J. Champoux (1981) Proc. Natl. Acad. Sci. U.S.A., **78,** 2883-2887.

Bellard, M., Dretzen, G., Bellard, F., Oudet, P. and P. Chambon (1982) EMBO J., **1,** 223-230.

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

Bienz, M. and H.R.B. Pelham (1982) EMBO J., **1**, 1583-1588.

Billet, M.A. and J. Hindley (1972) Eur. J. Biochem., **28**, 451-462.

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

Bird, A.P. (1980) Nuc. Acids Res., **8**, 1499-1504.

Bird, A.P. (1983) **in** "Eukaryotic Genes : Their Structure, Activity and Regulation", Maclean, N., Gregory, S.P. and R.A. Flavell, eds., Butterworths and Co., Cambridge, U.K., pp. 53-67.

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

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

Bishop, J.M. (1981) Cell, **23**, 5-6.

Bishop, J.M. (1983) Ann. Rev. Biochem., **52**, 301-354.

Bogenhagen, D.F., Sakonju, S. and D.D. Brown (1980) Cell, **19**, 27-35.

Bolivar, F. (1978) Gene, **4**, 121.

Bolivar, F. and K. Backman (1979) Methods in Enzymology, **68**, 245.

Boss, M.A. (1983) Nature, **303**, 281-282.

Bradbury, E.M., Inglis, R.J., and H.R., Mathews (1974) Nature, **247**, 257-261.

Brandhorst, B.P. (1980) Develop. Biol., **79**, 139-148.

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

Britten, R.J. and E.H. Davidson (1969) Science, **165**, 349-357.

Brown, D.D. (1982) The Harvey Lectures, Academic Press, New York, Series **76**, 27-44.

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

Bukhari, A.I., Shapiro, J.A. and S.L. Adhya (eds): "DNA Insertion Elements, Plasmids and Episomes", Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1977).

Burch, J.B.E. and H. Weintraub (1983) Cell, **33,** 65-76.

Busslinger, M., Portmann, R., Irminger, J. and M. Birnstiel (1980) Nuc. Acids Res., **8**, 957-977.

Busslinger, M., Rusconi, S. and M.L. Birnstiel (1982) EMBO J., **1,** 27-33.

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

Butler, W.B. and G.C. Mueller (1973) Biochim. Biophys. Acta, **294,** 481-496.

Casadaban, M.J. and S.N. Cohen (1980) J. Mol. Biol., **138**, 179-207.

Catcheside, D.G. (1974) Ann. Rev. Genetics, **8,** 279-300.

Catcheside, D.G. (1977) "The Genetics of Recombination; Genetics: Principles and Perspectives, vol 2" Edward Arnold Publishers, London.

Cattaneo, R., Gorski, J. and G. Mach (1981) Nuc. Acids. Res., **9,** 2777-2790.

Chandler, D.W. and J. Gralla (1981) Nuc. Acids Res., **9,** 6031-6046.

Chandler, M.E., Kedes, L.H., Cohn, R.H. and J.J. Yunis (1979) Science, **205**, 908-910.

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

Chao, M.V. Martinson, H.G. and J.D. Gralla (1980) Biochemistry, **19,** 3260-3269.

Chao, M.V., Mellon, P., Charnay, P., Maniatis, T. and R. Axel (1983) Cell, **32**, 483-493.

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

Childs, G., Maxson, R., Cohn, R. and L.H. Kedes (1981) Cell, **23,** 651-663.

Childs, G., Maxson, R. and L.H. Kedes (1979) Develop. Biol, **73**, 153-173.

Childs, G., Nocente-McGrath, C., Lieber, T., Holt, C. and J. Knowles (1982) Cell, **31**, 383-393.

Clark, A.J. (1973) Ann. Rev. Genet. **7**, 67-86.

Clarke, L. and J. Carbon (1976) Cell, **9**, 91-99.

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

Coen, E.S. and G.A. Dover (1983) Cell, **33**, 849-855.

Coen, E.S., Strachan, T. and G. Dover (1982) J. Mol. Biol, **158**, 17-35.

Coffino, P., Stimic, E., Groppi, V.E. and D. Bieber (1984) **in** "Histone Genes : Structure, Organisation and Regulation", Stein, G.S., Stein, J.L., and W.F. Marzluff, eds., J. Wiley and sons, New York, pp. 317-338.

Cohen, L.H., Newrock, K.M. and A. Zweidler (1975) Science, **190**, 994-997.

Cohen, S.N. and J.A. Shapiro (1980) Sci. Am., **242**, 36-45.

Cohn, R.H. Lowry, J.C. and L.H. Kedes (1976) Cell, **9**, 147-161.

Cohn, R.H. and L.H. Kedes (1979a) Cell, **18**, 843-854.

Cohn, R.H. and L.H. kedes (1979b) Cell, **18**, 855-864.

Cold Spring Harbor Laboratory (1981) "Movable Genetic Elements", Cold Spring Harbor Symp. Quant. Biol. **45**, Cold Spring Harbor, New York, pp. 1-1025.

Coles, L.S. and J.R.E. Wells (1985) Nuc. Acids Res., in press.

Collins, J. and B. Hohn (1978) Proc. Natl. Acad. Sci. U.S.A., **75**, 4242-4246.

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

Cozzarelli, N.R. (1980) Science, **207**, 953-960.

Crawford, R.J., Krieg, P.A., Harvey, R.P., Hewish, D.A. and J.R.E. Wells (1979) Nature, **279**, 132-136.

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

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

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

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

Davis, R.W., Botstein, D. and J.R. Roth (1980) **in** "Advanced Bacterial Genetics" Cold Spring Harbor Laboratory, New York, pp. 227-230.

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

Dean, F., Kransnow, M.A., Otter, R., Matzuk, M.M., Spengler, S.J. and N.R. Cozzarelli (1982) Cold Spring Harbor Symp. Quant. Biol., **47**, 769-777.

Deisseroth, A., Bode, Y., Fontana, J. and D. Hendrick (1980) Nature, **285**, 36-38.

Deisseroth, A. and D. Hendrick (1978) Cell, **15**, 55-63.

Destrée, O.H.J., Bendig, M.M., De Laaf, R.T.M. and J.G. Koster (1984) Biochim. Biophys. Acta, **782**, 132-141.

De Villiers, J., Olson, L., Tyndall, C. and W. Schaffner (1982) Nuc. Acids Res., **10**, 7965-7976.

Dodgson, J.B., Strommer, J. and J.D. Engel (1979) Cell, **17**, 659-668.

Dover, G. (1982) Nature, **299**, 111-116.

Dressler, D. and H. Potter (1982) Ann. Rev. Biochem. **51,** 727-762.

Dudler, R. and A.A. Travers (1984) Cell, **38,** 391-398.

Dunn, A.R. and N. Gough (1984) Trends Biochem. Sci., **9,** 81-82.

Dynan, W.S. and R. Tjian (1983) Cell, **35,** 79-87.

Efstradiatis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., Deriel, J.K. Forget, B.G., Weissman, S.M., Slightom, J.L. Blechl, A.E., Smithies, O., Baralle, F.E. Shoulders, C.C. and N.J. Proudfoot (1980) Cell, **21,** 653-688.

Elgin, S.C.R. (1981) Cell, **27,** 413-415.

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

Engel, D. (1984) **in** "Histone Genes, Structure, Organisation and Regulation", Stein, G.S., Stein, J.L. and W. Marzluff, eds., J. Wiley and Sons, New York, pp. 135-161.

Engel, J. and J. Dodgson (1981) Proc. Natl. Acad. Sci. U.S.A., **78,** 2856-2860.

Engel, J.D., Sugarman, B.D. and J.B. Dodgson (1982) Nature, **297,** 434-436.

Engelke, D.R. Ng, S-Y., Shastry, B.S. and R.G. Roeder (1980) Cell, **19,** 717-728.

Felsenfeld, G. and J. McGhee (1982) Nature, **296,** 602-603.

Fincham, J.R.S., Day, P.R. and A. Radford (1979) "Fungal Genetics", 4th Ed., Blackwell, Oxford.

Fink, G.R. and T.D. Petes (1984) Nature, **310,** 728-729.

Fisher, L.M. (1984) Nature, **307,** 686-687.

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

Flavell, A.J. (1982) Nature, **295**, 370.

Flavell, R.A. and F.G. Grosveld (1983) **in** "Eukaryotic Genes: Their Structure, Activity and Regulation", Maclean, N., Gregory, S.P., and R.A. Flavell, eds., Butterworths and Co., Cambridge, U.K., pp. 299-314.

Flynn, J.M. and H.R. Woodland (1980) Develop. Biol., **75**, 222-230.

Folger, K., Anderson, J.N., Hayward, M.A. and D.J. Shapiro (1983) J. Biol, Chem., **258**, 8908-8914.

Ford, J.P. and M.-T., Hsu (1978) J. Virol., **28**, 795-801.

Fritton, H.P., Igo-Kemenes, T., Nowock, J., Strech-Jurk, U., Theisen, M. and A.E. Sippel (1984) Nature, **311**, 163-165.

Gargiulo, G., Razui, F. and A. Worcel (1984) Cell, **38**, 511-521.

Gellert, M. (1981) Ann. Rev. Biochem., **50**, 879-910.

Gerber-Huber, S., May, F.E.B., Westley, B.R., Felber, B.K., Hosbach, H. A., Andres, A.-C., and G.U. Ryffel (1983) Cell, **33**, 43-51.

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

Goldberg, M. (1979) Ph.D. Thesis, Stanford University.

Goldknopf, I.L. and H. Busch (1977) Proc. Natl. Acad. Sci. U.S.A., **74**, 864-568.

Goldstein, G., Scheid, H., Hammerling, U., Boyse, E.A., Schlesinger, D.H. and H.D. Niall (1975) Proc. Natl. Acad. Sci. U.S.A., **72**, 11-15.

Gough, N.M. (1981) Trends Biochem. Sci. **6**, 203-205.

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

Graves, R.A. and W.F. Marzluff (1984) Mol. Cell. Biol, **4**, 351-357.

Groppi, V.E. Jr. and P. Coffino (1980) Cell, **21**, 195-204.

Gross-Bellard, M., Oudet, P. and P. Chambon (1973) Eur. J. Biochem., **36**, 32-38.

Grosveld, F.G., Dahl, H.-H.M., deBoer, E. and R.A. Flavell (1981) Gene, **13**, 227-237.

Grosveld, F., Lund, T., Murray, E., Mellor, A., Dahl, H. and R.A. Flavell (1982) Nuc. Acids Res., **10**, 6715-6732.

Groudine, M. and H. Weintraub (1981) Cell, **24**, 393-401.

Grunstein, M. (1978) Proc. Natl. Acad. Sci. U.S.A., **75**, 4135-4139.

Grunstein, M. and D. Hogness (1975) Proc. Natl. Acad. Sci. U.S.A, **72**, 3961-3965.

Gruss, P. and G. Khoury (1980) Nature, **286**, 634-637.

Hagenbuchle, O. and U. Schibler (1981) Proc. Natl. Acad. Sci. U.S.A., **78**, 2283-2286.

Halligan, B.D., Davis, J.L. Edwards, K.A. and L.F. Liu (1982) J. Biol. Chem., **257**, 3995-4000.

Hanahan, D. and M. Meselson (1980) Gene, **10**, 63-67.

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

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

Harvey, R.P. and J.R.E. Wells (1979) Nuc. Acids Res., **7**, 1787-1798.

Harvey, R.P. Whiting, J.A., Coles, L.S., Krieg, P.A. and J.R.E. Wells (1983) Proc. Natl. Acad. Sci. U.S.A., **80**, 2819-2823.

Hedges, R.W. and A.E. Jacob (1974) Mol. Gen. Genet., **132**, 31-40.

Heintz, N., Sive, H.L. and R.G. Roeder (1983) Mol. Cell Biol., **3**, 539-550.

Heintz, N., Zernik, M. and R.G. Roeder (1981) Cell, **24**, 661-668.

Hentschel, C. (1982) Nature, **295**, 714-716.

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

Hereford, L., Bromley, S. and M.A. Osley (1982) Cell, **30**, 305-310.

Hereford, L.M. Fahrner, K., Woolford, J. Jr., Rosbash, M. and D.B. Kaback (1979) Cell, **18**, 1261-1271.

Hershko, A. (1983) Cell, **34**, 11-12.

Hofer, E. and J.E. Darnell (1981) Cell, **23**, 585-593.

Hofer, E., Hofer-Warbinek, R. and J.E. Darnell Jr. (1982) Cell, **29**, 887-893.

Hohn, B. and J. Collins (1980) Gene, **11**, 291-298.

Hohn, B. and K. Murray (1977) Proc. Natl. Acad. Sci. U.S.A., **74**, 3259-3263.

Holliday, R. (1964) Genet. Res., **5**, 282-304.

Holt, C.A. and G. Childs (1984) Nuc. Acids Res., **12**, 6455-6471.

Hossenlopp, P., Oudet, P. and P. Chambon (1974) Eur. J. Biochem., **41**, 397-411.

Howard-Flanders, P., West, S.C. and A. Stasiak (1984) Nature, **309**, 215-220.

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

Ikeda, H., Kazuko, A. and A. Naito (1982) Proc. Natl. Acad. Sci. U.S.A., **79**, 3724-3728.

Ikeda, H., Moriya, K. and T. Matsumoto (1981) Cold Spring Harbor Symp. Quant. Biol., **45**, 399-408.

Ikenaga, H. and K. Saigo (1982) Proc. Natl. Acad. Sci. U.S.A., **79**, 4143-4147.

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

Ish-Horowicz, D. and J.F. Burke (1981) Nuc. Acids Res., **9**, 2989-2998.

Jackson, J.A. and G. Fink (1981) Nature, **292**, 306-307.

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

Jeffreys, A.J. (1982) **in** "Genome Evolution", Dover, G.A. and R.B. Flavell, eds., Academic Press, London, pp. 157-176.

Jeffreys, A.J. and S. Harris (1982) Nature, **296**, 9-10.

Jelenik, W.R. and C.W. Schmid (1982) Ann. Rev. Biochem, **51**, 813-844.

Johns, E.W. (1982) "The HMG Chromosomal Proteins", Academic Press, New York.

Jones, C. and F.C. Kafatos (1980) Cell, **22**, 855-867.

Jones, R.W., Old, J.M. Trent, R.J. Clegg, J.B. and D.J. Weatherall (1981) Nature, **291**, 39-44.

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

Kofatos, F.C., Jones, C.W., and A. Efstratiadis (1979). Nuc. Acids Res., **7**, 1541-1552.

Karpov, V.L., Preobrazhenskaya, O.V. and A.D. Mirzabekov (1984) Cell, **36**, 423-431.

Kedes, L.H. (1979) Ann. Rev. Biochem., **48**, 837-870.

Keene, M.A. and S.C.R. Elgin (1981) Cell, **27**, 57-64.

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

Kikuchi, A. and K. Asai (1984) Nature, **309**, 677-681.

Klar, A.J.S. and J.N. Strathern (1984) Nature, **310**, 744-748.

Klein, H.L. and T.D. Petes (1981) Nature, **289**, 144-148.

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

Kornberg, R.D. (1974) Science, **184**, 868-871.

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

Krieg, P.A. (1980) Ph.D. Thesis, Adelaide University.

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

Krieg, P.A., Robins, A.J., Colman, A. and J.R.E. Wells (1982a) Nuc. Acids Res., **10**, 6777-6785.

Krieg, P.A., Robins, A.J., D'Andrea, R.J. and J.R.E. Wells (1983) Nuc. Acids Res., **11**, 619-627.

Krieg, P.A. Robins, A.J. Gait, M.J., Titmas, R.C. and J.R.E. Wells (1982b) Nuc. Acids Res., **10**, 1495-1502.

Kriegler, M. and M. Botchan (1983) Mol. Cell. Biol. **3**, 325-339.

Kutsukake, K. and T. Iino (1980) nature, **284**, 479-481.

Labhart, P. and T.L. Koller (1982) Cell, **28**, 279-292.

Laimins, L.A., Khoury, G., Gorman, C., Howard, B. and P. Gruss (1982) Proc. Natl. Acad. Sci. U.S.A., **79**, 6453-6457.

Lamb, M.M. and B. Daneholt (1979) Cell, **17**, 835-848.

Leder, A., Swan, D., Ruddle, F., d'Eustachio, P. and P. Leder (1981) Nature, **293**, 196-200.

Lescure, B., Chestier, A. and M. Yaniv (1978) J. Mol. Biol. **124**, 73-85.

Lesnikowski, C.M. (1983) Honours Thesis, Adelaide University.

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

Levy, A. and M. Noll (1981) Nature, **289**, 198-203.

Lewin, R. (1983) Science, **219**, 1052-1054.

Lichtler, A.C. Detke, S., Phillips, I.K., Stein, G.S. and J.L. Stein (1980) Proc. Natl. Acad. Sci. U.S.A., **77**, 1942-1946.

Lichtler, A.C. Sierra, F., Clark, S., Wells, J.R.E., Stein, J. and G. Stein (1982) Nature, **298**, 195-198.

Liebermann, D., Hoffman-Liebermann, B., Weinthal, J., Childs, G., Maxson, R., Mauron, A., Cohen, S.N. and L. Kedes (1983) Nature **306**, 342-347.

Lifton, R.P., Goldberg, M.L. Karp, R.W. and D.S. Hogness (1977) Cold Spring Harbor Symp. Quant. Biol., **42**, 1047-1051.

Lilley, D.M.J. and M. Houghton (1979) Nuc. Acids Res., **6**, 507-523.

Liu, L.F., Liu, C.-C. and B.M. Alberts. (1980) Cell, **19**, 697-707.

Luchnik, A.N., Bakayev, V.V. and V.M. Glaser (1982) Cold Spring Harbor Symp. Quant. Biol., **47**, 793-801.

Lueders, K., Leder, A., Leder, P. and E. Kuff (1982) Nature, **295**, 426-428.

Mandel, J.-L. and P. Chambon (1974) Eur. J. Biochem., **41**, 367-378.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, D.K. and A. Efstratiadis (1978) Cell, **15**, 687-701.

Maniatis, T., Jeffrey, A. and D.G. Kleid (1975) Proc. Natl. Acad. Sci. U.S.A., **72**, 1184-1188.

Marashi, F., Prokopp, J., Stein, J. and G. Stein (1984) Proc. Natl. Acad. Sci. U.S.A., **81**, 1936-1940.

Markham, P. and H.L.K. Whitehouse (1982) Nature, **295**, 421-423.

Marvo, S.L., King, S.R. and S.R. Jaskunas (1983) Proc. Natl. Acad. Sci. U.S.A., **80**, 2452-2456.

Marzluff, W.F. and R.A. Graves (1984) **in** "Histone Genes, Structure Organisation and Regulation". Stein, G.S. Stein, J.L. and W.F. Marzluff, eds., J. Wiley and Sons, New York, pp. 281-315.

Mathis, D. Oudet, P. and P. Chambon (1980) Progress in Nuc. Acids Res. and Mol. Biol., **24**, 1-55.

Matsui, T., Segall, J., Weil, P.A. and R.G. Roeder (1981) J. Biol. Chem. **255**, 11992-11996.

Mattern, M.R. and R.B. Painter (1979) Biochim. Biophys. Acta, **563**, 306-312.

Max, E.E. (1984) Nature, **310**, 100.

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

Maxson, R., Mohun, T., Gormezano, G., Childs, G. and L. Kedes (1983b) Nature, **301**, 120-125.

Maxson, R., Mohun, T., and L. Kedes (1983c) **in** "Eukaryotic Genes : Their Structure, Activity and Regulation", Maclean, N., Gregory, S.P. and R.A. Flavell, eds., Butterworths and Co., Cambridge U.K., pp. 277-298.

Maxson, R. and F. Wilt (1982) Develop. Biol., **94**, 435-440.

McClintock, B. (1950) Proc. Natl. Acad. Sci. U.S.A., **36**, 344-355.

McClintock, B. (1951) Cold Spring Harbor Symp. Quant. Biol., **16**, 13-47.

McKnight, S.L. (1982) Cell, **31**, 355-365.

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

Meijlink, F.C.P.W., Philipsen, J.N.J., Gruber, M. and G. AB (1983) Nuc. Acids Res., **11**, 1361-1373.

Meselson, M. and C. Radding (1975) Proc. Natl. Acad. Sci. U.S.A., **72**, 358-361.

Messing, J., Crea, R. and P.H. Seeburg (1981) Nuc. Acids Res., **9**, 309-321.

Meyerowitz, E.M., Guild, G.M., Prestidge, L.S. and D.S. Hogness (1980) Gene, **11**, 271-282.

Miller, J.R. (1983) **in** "Eukaryotic Genes, Their Structure, Activity and Regulation", Maclean, N., Gregory, S.P. and R.A. Flavell, eds., Butterworths and Co., Cambridge U.K., pp 225-237.

Mirault, M.E., Southgate, R. and E. Delward (1982) EMBO J., **1**, 1279-1285.

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

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and P. Chambon (1981) Nuc. Acids Res., **9**, 6047-6067.

Morris, D.W. Noti, J.D., Osborne, F.A. and A.A. Stalay (1981) DNA, **1**, 27-36.

Moss, B.A. Joyce, W.G. and V. Ingraham (1973) J. Biol. Chem., **248**, 1025-1031.

Nasmyth, K.A. (1982) Ann. Rev. Genet., **16**, 439-500.

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

Nevins, J.R. and J.E. Darnell (1978) Cell, **15**, 1477-1493.

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

Noll, M. and R.D. Kornberg (1977) J. Mol. Biol., **109**, 393-404.

Nordheim A. and A. Rich (1983) Nature, **303**, 647-679.

Ohtsubo, H. and E. Ohtsubo (1978) Proc. Natl. Acad. Sci. U.S.A., **75**, 615-619.

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

Osley, M.A. and L. Hereford (1982) Proc. Natl. Acad. Sci. U.S.A., **79**, 7689-7693.

Pardue, M.L., Kedes, L.H. Weinberg, E.S. and M.L. Birnstiel (1977) Chromosoma, **63**, 135-151.

Parker, C.S. and J. Topol (1984a) Cell, **36**, 357-369.

Parker, C.S. and J. Topol (1984b) Cell, **37**, 273-283.

Pehrson, J.R. and R.D. Cole (1981) Biochemistry, **20**, 2298-2301.

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

Pelham, H.R.B. and M. Bienz (1982) EMBO J., **1**, 1473-1477.

Petes, T.D. (1980) Cell, **19**, 765-774.

Peterson, N.S. and H.K. Mitchell (1984) "Heat Shock Proteins" **in** "Comprehensive Insect Physiology, Biochemistry and Pharmacology", Kercut, G.A. and L. Gilbert, eds., Pergamon Press, New York.

Price, D. and C.S. Parker (1984) Cell, **38**, 423-429.

Proudfoot, N. (1984) Nature, **307**, 412-413.

Proudfoot, N.J. and G.G. Brownlee (1976) Nature, **263**, 211-214.

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

Radding, C.M. (1978) Ann. Rev. Biochem., **47**, 847-880.

Reudelhuber, T. (1984) Nature, **311**, 301.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. and A. Klug (1984) Nature, **311**, 532-537.

Ripley, L.S. and B.W. Glickman (1982) Cold Spring Harbor Symp. Quant. Biol., **47**, 851-860.

Risley, M.S. and R.A. Eckhardt (1981) Develop. Biol., **84**, 79-87.

Roberts, S.B., Weisser, K.E. and G. Childs (1984) J. Mol. Biol., **174**, 647-662.

Robertson, M. (1981) Nature, **290**, 625-627.

Roeder, R.G. (1976) **in** "RNA Polymerase". Losick, R. and M. Chamberlain, eds., Cold Spring Harbor Press, Cold Spring Harbor, New York, pp. 285-329.

Roeder, G.S., Smith, M. and E.J. Lambie (1984) Mol. Cell. Biol., **4**, 703-711.

Rubnitzt, J. and S. Subramani (1984) Mol. Cell. Biol., **4**, 2253-2258.

Ruderman, J.V. and M.R. Schmidt (1981) Develop. Biol., **81**, 220-228.

Ruiz-Carrillo, A., Affolter, M. and J. Renaud (1983) J. Mol. Biol., **170**, 843-859.

Ruiz-Carrillo, A., Waugh, L.J., Littau, V.C. and V.G. Allfrey (1974) J. Biol. Chem., **249**, 7358-7368.

Ruley, H.E. and M. Fried. (1983) Nature, **304**, 181-184.

Rykowski, M., Wallis, J., Choe, J. and M. Grunstein (1981) Cell, **25**, 477-487.

Saigo, K., Millstein, L. and L.A. Thomas (1981) Cold Spring Harbor Symp. Quant. Biol., **45**, 815-827.

Sakonju, S., Bogenhagen, D.F. and D.D. Brown (1980) Cell, **19**, 13-25.

Salditt-Georgieff, M. and J.E. Darnell Jr. (1983) Proc. Natl. Acad. Sci. U.S.A., **80**, 4694-4698.

Sanger, F., Nicklen, S. and A.R. Coulson (1977) Proc. Natl. Acad. Sci. U.S.A., **74**, 5463-5467.

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

Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith H.O. and M.L. Birnsteil (1978) Cell, **14**, 655-671.

Schlissel, M.S. and D.D. Brown (1984) Cell, **37**, 903-913.

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

Seeburg, P.H., Shine, J., Martial, J.A. Baxter, J.D. and H.M. Goodman (1977) Nature, **270**, 486-494.

Segall, J., Matsui, T. and R.G. Roeder (1980) J. Biol. Chem., **255**, 11986-11991.

Shapiro, J.A. (1969) J. Mol. Biol., **40**, 93-105.

Shapiro, J.A. (1979) Proc. Natl. Acad. Sci. U.S.A., **76**, 1933-1937.

Shapiro, J.A. (1982) Natl. Cancer Inst. Monograph, **60**, 87-110.

Shatkin, A.J. (1976) Cell, **9**, 645-653.

Shen, C.-K.J. and T. Maniatis (1980) Cell, **19**, 379-391.

Shen, S.-L., Slightom, J.L. and O. Smithies (1981) Cell, **26**, 191-203.

Sierra, F., Lichtler, A., Marashi, F., Rickless, R., Van Dyke, T., Clark, S., Wells, J., Stein, G. and J. Stein (1982) Proc. Natl. Acad. Sci. U.S.A., **79**, 1795-1799.

Simon, M., Zieg, J. and M. Silverman (1980) Science, **209**, 1370-1374.

Sittman, D., Chiu, I., Pan, C., Cohn, R., Kedes, L. and W. Marzluff (1981) Proc. Natl. Acad. Sci. U.S.A., **78**, 4078-4082.

Sittman D.B., Graves, R.A. and W.F. Marzluff (1983). Proc. Natl. Acad. Sci. U.S.A., **80**, 1849-1853.

Slightom, J.L., Blechi, A.E. and O. Smithies (1980) Cell, **21**, 627-638.

Smith, B.J., Harris, M.R. Sigournay, C.M., Mayes, E.L.V. and M. Bustin (1984). Eur. J. Biochem., **138**, 309-317.

Smith, G.E. and M.D. Summers (1980) Analytical Biochem., **109**, 123-129.

Smith, G.P. (1976) Science, **191**, 528-535.

Smith, G.R. (1981) Cell, **24**, 599-600.

Smith, G.R. (1983) Cell, **34**, 709-710.

Smith, M.M. and K. Murray (1983) J. Mol. Biol. **169**, 641-661.

Sobell, H.M. (1975) Proc. Natl. Acad. Sci. U.S.A., **72**, 279-283.

Southern, E.M. (1975) J. Mol. Biol., **98**, 503-517.

Spofford, J.B. (1976) **in** "The Genetics and Biology of *Drosophila*". Ashburner, M. and E. Novitski, eds., Academic Press, New York, Vol. 1C., pp. 955-1018.

Stahl, F.W. (1979) Ann. Rev. Genet., **13**, 7-24.

Stahl, H. and D. Gallwitz (1977) Eur. J. Biochem., **72**, 385-392.

Stalder, J., Larsen, A., Engel, J.D., Dolan, M., Groudine, M. and H. Weintraub (1980) Cell, **20**, 451-460.

Stein, G.S., Sierra, F., Stein, J.L., Plumb, M., Marashi, F., Corozzi, N., Prokopp, K. and L. Baumbach (1984) **in** "Histone Genes, Structure, Organisation and Regulation". Stein, G.S., Stein, J.L. and W. Marzluff, eds., J. Wiley and sons, New York, pp. 397-455.

Stein, G., Stein, J. and L.J. Kleinsmith (1983) **in** "Eukaryotic Genes: Their Structure, Activity and Regulation", Maclean, N., Gregory, S.P., and R.A. Flavell, eds., Butterworths and Co., Cambridge U.K., pp. 31-51.

Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Srelinger, J., Wake, C., Long, E., Mach, B. and L. Hood (1982) Nature, **300**, 35-42.

Stephenson, E., Erba, H. and J. Gall (1981) Nuc. Acids Res., **9**, 2281-2295.

Sternberg, N., Tiemeier, D. and L. Enquist (1977) Gene, **1**, 255-280.

Stunnenberg, H.G. and M.L. Birnstiel (1982) Proc. Natl. Acad. Sci, U.S.A., **79**, 6201-6204.

Sugarman, B.J. Dodgson, J.B. and J.D. Engel (1983) J. Biol. Chem., **258**, 9005-9016.

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

Syvanen, M. (1974) Proc. Natl. Acad. Sci. U.S.A., **71**. 2496-2499.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. and F.W. Stahl (1983) Cell, **33**, 25-35.

Szostak, J.W. and R.Wu (1980) Nature, **284**, 426-430.

Taylor, J.M., Illmensee, R. and J. Summers (1976) Biochim. Biophys. Acta, **442**, 324-330.

Thomas, J.O. (1983) **in** "Eukaryotic Genes : Their Structure, Activity and Regulation", Maclean, N., Gregory, S.P., and R.A. Flavell, eds., Butterworths and Co., Cambridge U.K., pp. 9-30.

Todd, P.A. and B.W. Glickman (1982) Proc. Natl. Acad. Sci. U.S.A., **79**, 4123-4127.

Turner, P.C., Aldridge, T.C., Woodland, H.R. and R.W. Old (1983) Nuc. Acids Res., **11**, 4093-4107.

Turner, P.C. and H.R. Woodland (1983) Nuc. Acids Res., **11**, 971-986.

Van de Putte, P., Cramer, S. and M. Giphant-Gassler (1980) Nature, **286**, 218-222.

Van Dongen, W.M.A.M., Moorman, A.F.M. and O.H.J. Destrée (1984) **in** "Histone Genes, Structure, Organisation and Regulation", Stein, G.S., Stein, J.L. and W.F. marzluff, eds., J. Wiley and sons, New York, pp. 199-224.

Vidali, G., Boffa, L.C. Bradbury, E.M. and V.G. Allfrey (1978) Proc. Natl. Acad. Sci. U.S.A., **75**, 2239-2243.

Vitelli, L. and E.S. Weinberg (1983) Nuc. Acids Res., **11**, 2135-2153.

Wagner, R.E. and M. Radman (1975) Proc. Natl. Acad. Sci. U.S.A., **72**, 3619-3622.

Wahl, G.M., Stern, M. and G.R. Stark (1979) Proc. Natl. Acad. Sci. U.S.A., **76**, 3683-3687.

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

Wang, A.H.-J., Quigley, G.J., Kolpack, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and A. Rich (1979) Nature, **282**, 680-686.

Wang, J.C. (1984) Nature, **309**, 669-670.

Wang, S.W., Robins, A.J., D'Andrea, R.J. and J.R.E. Wells (1985) Nuc. Acids Res., in press.

Wasylyk, B., Wasylyk, C., Augereau, P. and P. Chambon (1983) Cell, **32**, 503-514.

Weatherall, D.J. and J.B. Clegg (1979) Cell, **16**, 467-479.

Weatherall, D.J. and J.B. Clegg (1981) "The Thalassaemia Syndromes", Blackwell Scientific Publications, Oxford, U.K.

Weiher, H., Konig, M. and P. Gruss (1983) Science, **219**, 626-631.

Weinberg, E.S., Hendricks, M.B., Hemminki, K., Kuwabara, P.E. and L.A. Farrelly (1983) Develop. Biol., **98**, 117-129.

Weinberg, E.S., Overton, G.C., Hendricks, M.B., Newrock, K.M. and L.H. Ohen (1978) Cold Spring Harbor Symp. Quant. Biol., **42**, 1093-1100.

Weiner, A.M. and R.A. Denison (1982) Cold Spring Harbor Symp. Quant. Biol., **47**, 1141-1149.

Weintraub, H. (1984) Cell, **38**, 17-27.

Weisberg, R.A. and S. Adhya (1977) Ann. Rev. Genet., **11**, 451-473.

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

Weiss, E.H., Golden, L., Fahrner, K., Mellor, A.L., Devlin, J.J., Bullman, H., Tiddens, H., Bud, H. and R.A. Flavell (1984) Nature, **310**, 650-655.

Weiss, E.H., Mellor, A., Golden, L., Fahrner, K., Simpson, E., Hurst, J. and R.A. Flavell (1983) Nature, **301**, 671-679.

Whitehouse, H.L.K. (1983) Nature, **306**, 645-646.

Whitehouse, H.L.K. and P.J. Hastings (1965) Genet. Res., **6**, 27-92.

Widom, J. (1984) Nature, **309**, 312-313.

Wilks, A.F., Cozens, P.J., Mattaj, I.W. and J.-P. Jost (1982) Proc. Natl. Acad. Sci. U.S.A., **79**, 4252-4255.

Willing, M.C. Nienhuis, A.W. and W.F. Anderson (1979) Nature, **277**, 534-538.

Wilt, F.H. (1970) Develop. Biol., **23**, 444-455.

Winkler, H. (1930) Die Konversion der Gene, G. Fischer, Jena.

Winter, G. (1980) E.M.B.O. Summer Course Notes.

Woodland, H.R. (1980) FEBS Lett., **121**, 1-7.

Woodland, H.R., Flynn, J.M. and A.J. Wyllie (1979) Cell, **18**, 165-171.

Woodland, H.R., Old, R.W., Sturgess, E.A., Ballentine, J.E.M., Aldridge, T.C. and P.C. Turner (1983), **cited in** Maxson, R. Cohn, R. and L. Kedes (1983) Ann. Rev. Genet., **17**, 239-277.

Wormington, W.M., Bogenhagen, D.F., Jordan, E. and D.D. Brown (1981) Cell, **24**, 809-817.

Woudt, L.P., Pastink, A., Kempers-Veenstra, A.E., Jansen, A.E.M., Mager, W.H. and R.J. Planta (1983) Nuc. Acids Res., **11**, 5347-5359.

Wright, S., de Boer, E., Grosveld, F.G. and R.A. Flavell (1983) Nature, **305**, 333-336.

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

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

Wu, C. and W. Gilbert (1981) Proc. Natl. Acad. Sci. U.S.A., **78**, 1577-1580.

Wu, C., Wong, Y.C. and S.C.R. Elgin (1979) Cell, **16**, 807-814.

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

Wu, R.S. West M.H.P. and W.M. Bonner (1984) **in** "Histone Genes, Structure, Organisation and Regulation", Stein, G.S., Stein, J.L. and W.F. Marzluff, eds., J. Wiley and sons, New York, pp. 457-484.

Yamamoto, K.R. (1983) **in** "Steroid Hormone Receptors; Structure and Function", Nobel Symposium, **57**, Erikkson, H., Gustafsson, J.A. and B. Hogberg, eds., Elsevier/North Holland Biomedical Press, Amsterdam, pp. 285-306.

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

Zalokar, M. (1976) Develop. Biol., **49**, 425-437.

Zaret, K.S. and K.R. Yamamoto (1984) Cell, **38**, 29-38.

Zernik, M., Heintz, N., Boime, I. and R. Roeder (1980) Cell, **22**, 807-815.

Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W. and A.C. Wilson (1980) Proc. Natl. Acad. Sci. U.S.A., **77**, 2158-2162.

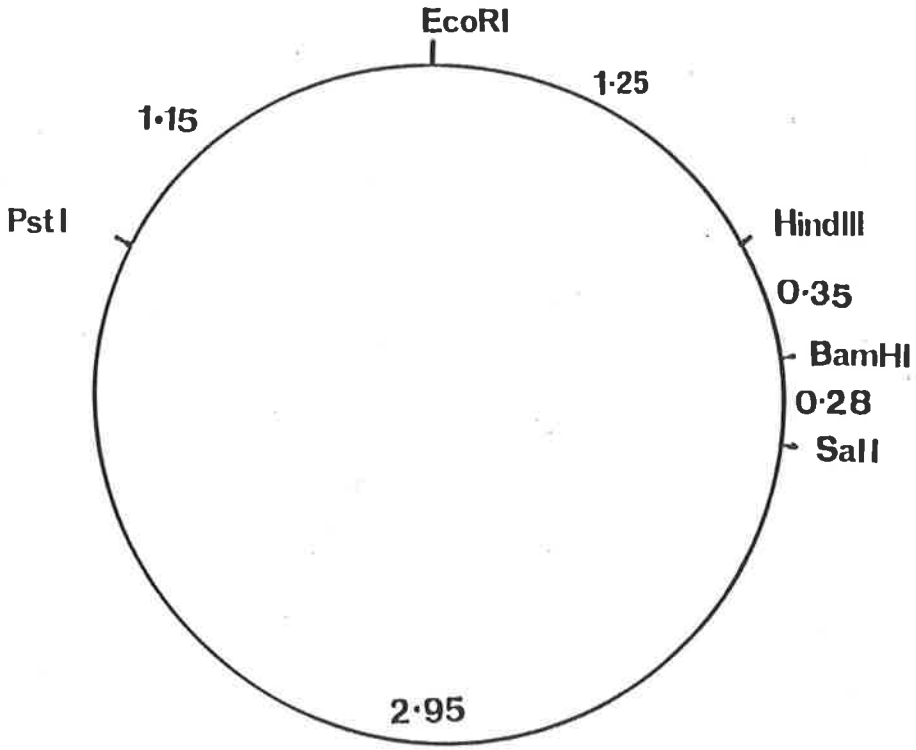
Zlatanova, J.S. (1980) FEBS Lett., **112**, 199-202.

Zweidler, A. (1980) **in** "Gene Families of Collagen and other Proteins", Prockop, D. and P.C. Champe, eds., Elsevier/North-Holland, New York, pp. 47-56.

Zweidler, A. (1984) **in** "Histone Genes, Structure, Organisation and Regulation", Stein, G.S., Stein, J.L. and W.F. Marzluff, eds., J. Wiley and Sons, New York, pp. 339-371.

APPENDIX : A

A restriction map of the plasmid vector pBR325.


EcoRI
1·25
HindIII
0·35
BamHI
0·28
SalI
2·95
PstI
1·15

APPENDIX: B - PUBLICATIONS

Work described in this thesis has been presented in the following papers:

A. Papers Published or Submitted

D'Andrea, R., Harvey, R. and J.R.E. Wells (1981). Vertebrate histone genes: Nucleotide sequence of a chicken H_2A gene and regulatory flanking sequences. Nucleic Acids Research, **9**, 3119-3128.

D'Andrea, R.J., Coles, L.S., Lesnikowski, C., Tabe, L. and J.R.E. Wells (1985). Chromosomal organisation of chicken histone genes: Preferred associations and inverted duplications. Submitted to Molecular and Cellular Biology.

Krieg, P.A., Robins, A.J., D'Andrea, R.J. and J.R.E. Wells (1983). The chicken H5 gene is unlinked to core and H1 histone genes. Nucleic Acids Research, **11**, 619-627.

Wang, S.-W., Robins, A.J., D'Andrea, R. and J.R.E. Wells (1985) Inverted duplications of histone genes in chicken and disposition of regulatory sequences. Submitted to Nucleic Acids Research.

Wells, J.R.E., Coles, L.S., D'Andrea, R., Harvey, R.P., Krieg, P.A., Robins, A.J., and J. Whiting (1983). Chicken histone genes; Analysis of core, H1, H5 and variant sequences. Manipulation and Expression of Genes in Eukaryotes, Academic Press, Australia.

B. Work Presented at Meetings

D'Andrea, R.J., Harvey, R.P. and J.R.E. Wells. DNA sequence of chicken H_2A and H_2B genes and their flanking regions: Talk presented at Sixth Annual Conference on Protein Structure and Function, Lorne, Victoria, February, 1981.

D'Andrea, R. and L. Tabe. Cosmid library construction: Talk presented at Fifth Annual Conference on the Organisation and Expression of the Eukaryotic Genome, Lorne, Victoria, February 1983.

D'Andrea, R., Coles, L.S., Lesnikowski, C. and J.R.E. Wells. Disorganised chicken histone genes: Poster presented at the Sixth Annual Conference on the Organisation and Expression of the Eukaryotic Genome, Lorne, Victoria, February 1984.