CHICKEN HISTONE H1 GENES

A thesis submitted to the University of Adelaide

for the degree of Doctor of Philosophy

by

Leeanne Susan Coles, B.Sc. (Hons.)

Department of Biochemistry,

University of Adelaide,

Adelaide, South Australia.

January, 1986.

CONTENTS

SUMMARY

This thesis describes the isolation and analysis of the family of genes coding for H1 histone variants in the chicken.

The first part of this thesis describes the construction of a 5-day chicken embryo cDNA library, and the isolation, from this library, of a recombinant containing H1 coding sequences. The nature of the recombinant was determined by DNA sequencing.

cDNA recombinants containing core histone sequences were also isolated. One of these, contained sequences coding for an extremely variant H2A protein which probably represents the replication-independent variant H2A.Z.

The insert of the H1 cDNA recombinant was used to screen a chicken genomal library. H1 positive isolates containing previously uncharacterized DNA were identified and further characterized by restriction enzyme mapping and hybridization analysis. From this work and that performed by other members of this laboratory a total of six H1 genes were located. All H1 genes were found to be clustered with core histone genes.

Subsequent Southern analysis of chicken genomal DNA suggested that the six located H1 genes represented the full complement of H1 genes, homologous to the H1 cDNA probe, in the chicken genome.

Sequence analysis of the chicken H1 genes revealed that each gene coded for a different H1 protein. This is consistent with estimates of H1 variant numbers in chicken tissues. The coded proteins are quite distinct in primary sequence from the H1-related chicken H5 protein.

The chicken H1 genes, as for most histone genes analyzed, contain conserved histone gene-specific 3' sequence elements and have no introns.

Analysis of the chicken H1 mRNAs also revealed features common to other histone genes, viz., H1 mRNAs were found, by Northern analysis, to be non-polyadenylated and mRNA 3' termini were found, by S1 analysis, to map to a predicted conserved sequence. The H1 genes identified here appeared to be expressed in both embryo and adult tissue.

The chicken H1 gene sequences were analyzed in order to identify sequence elements that could potentially be involved in the various aspects of expression of these genes.

Comparison of the chicken H1 genes to H1 genes from other species resulted in the identification of conserved motifs in both 5' and 3' non-coding regions. Of particular note, a 7 base-pair sequence, 5' AAACACA 3', specific to H1 genes, was located in 5' non-coding regions. This is the first report of an H1 gene-specific sequence element. A role for this sequence in the cell cycle regulation of H1 gene expression is proposed. A G-rich promoter element, found to be required for efficient transcription, was also located. This element is also found in other histone and non-histone genes.

In addition, differences in non-coding regions between the chicken H1 genes are pointed out. Such differences could play a role in differential H1 gene expression.

Finally, H1 genes were compared to the tissue-specifically expressed chicken H5 gene. The H5 gene, in particular, was found to lack the H1 gene-specific 5' element discussed above, but, it does retain the remnants of certain conserved 3' elements. The consequences of these findings with regard to expression and evolution of the H5 gene are discussed.

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.

LEEANNE SUSAN COLES

ACKNOWLEDGEMENTS

I wish to thank Professor W.H. Elliot for permission to work in the department and Dr. J.R.E. Wells for his excellent supervision.

My thanks also to the members of Labs 20 and 20A, for their technical advice, critical discussion and friendship, during the course of my experimental work. In particular, I wish to thank Richard D'Andrea and Paul Krieg.

I give special thanks to Peter Vize for his support and assistance throughout the preparation of this thesis.

I also wish to thank Adelaide Secretarial for typing this manuscript and Lesley Crocker for photographic work.

Finally, I am indebted to my parents for their support, patience and encouragement.

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

CHAPTER 1

LITERATURE REVIEW

1.1 INTRODUCTION

This thesis describes the analysis of a family of genes that code for H1 histone proteins in the chicken.

The histones, consisting of five classes of small basic proteins, the H2A, H2B, H3, H4 and H1 histones, play a major role in packaging eukaryotic DNA into chromatin. The former four histones, termed core histones, are involved in nucleosome formation, while, the H1 histones are principally involved in the formation of higher order chromatin structures. Each class of histone consists of several non-allelic primary structure variants which are differentially expressed throughout development, the cell cycle and the differentiation of particular cell lines. Of the histone classes, the H1 histones of an organism are the most heterogeneous. It appears that histone variants, in particular H1 variants, may differentially affect chromatin structure and play a role in the formation of different functional states of chromatin within a cell.

Analysis of the histone multigene family is gradually elucidating the mechanisms involved in the expression of the histone proteins, which may themselves be involved in the processes of selective gene expression. Current knowledge of histone genes and their proteins is discussed in the following sections.

1.2 CORE HISTONE PROTEINS

1.2.1 Role in Chromatin Structure

The core histones consist of four classes of small basic proteins, H2A, H2B, H3 and H4, which are ubiquitous in eukaryotic organisms (Isenberg, 1979; Von Holt *et al.*, 1979). These proteins, particularly the H3 and H4 histones, are highly conserved throughout evolution.

Core histones are involved in the formation of the basic unit of chromatin structure, the nucleosome (reviewed in McGhee and Felsenfeld, 1980; Igo-Kemenes et al., 1982; Butler, 1983; Thomas, 1983, 1984). The nucleosome consists of a nucleosome core particle, containing two each of the core histones associated with 146 bp (base pairs) of DNA, and variable amounts of linker DNA joining adjacent core particles. Nucleosome repeat lengths vary between 166 and 241 bp depending on the source of chromatin. The three-dimensional structure of the nucleosome core particle at 7Å resolution has recently been determined (Richmond et al., 1984). The fifth class of histone proteins, the H1 or "linker" histones, bind at the exit and entry points of DNA from the nucleosome core. The presence of H1 is necessary for the condensation of nucleosome chains into higher order chromatin structures (Section 1.3.1).

1.2.2 Heterogeneity

Within a single organism there exist non-allelic primary structure variants (subtypes) of the H2A, H2B, and H3 histone classes (Isenberg, 1979; Von Holt et al., 1979; Zweidler, 1984; Wu et al.,1984). Differences distinguishing non-allelic variants of a particular histone class, within a species, are generally found within the central regions of the proteins (Zweidler, 1984). These regions are involved in the formation of histone complexes (McGhee and Felsenfeld, 1980).

Variants of a particular histone class can be differentially expressed during development, the cell cycle and the differentiation of particular cell types.

1.2.3 Expression of Variants

The most dramatic example of developmental regulation of histone

subtype expression is found in the sea urchin, where different subtypes of the histones H1, H2A, and H2B are expressed in a programmed fashion during embryogenesis. In the sea urchin Strongylocentrotus purpuratus there are three stages of histone synthesis, resulting in the production of cleavage-stage (CS) variants, the early or α-variants and the late variants (reviewed in Maxson et al., 1983a, 1983b; Von Holt et al., 1984).

The synthesis of "late" histone variants is also observed during the development of the chicken embryo (Urban and Zweidler, 1983). The chicken contains two major non-allelic variants of each of the H2A, H2B and H3 histones (Urban et al., 1979) as well as minor components such as M1 (also called H2A.Z; Zweidler, 1984). At the earliest stages of development only one variant each of H2B (H2B.1) and H3 (H3.2) can be detected. The second variant of each of these histones (H2B.2 and H3.3) becomes detectable and gradually increases during somite formation. In contrast to the situation in sea urchin, both the "early" and "late" appearing chicken variants continue to be expressed throughout life in all tissues. After hatching, the relative amounts of the "late" H2B and H3 variants remain at their low embryonic levels in those tissues that maintain a high mitotic activity, but increase in amount with different kinetics in tissues which essentially stop cell division in adults. This is most dramatic for H3.3 which eventually becomes the predominant H3 variant in liver and kidney of older animals. The two major H2A variants (H2A.1 and H2A.2) do not change in their relative proportions during development or in adult tissues.

In the mouse, in contrast to the two organisms discussed above, all non-allelic somatic core histone variants are already present one day after implantation and no significant quantitative changes occur until birth (Zweidler et al., 1978; Zweidler, 1980, 1984). After birth, however, changes in the relative proportions of histone variants occur in a similar fashion to those changes observed in the chicken after hatching. That is, the proportions of different variants are maintained at embryonic levels in

rapidly dividing tissues, but in essentially non-dividing adult cells those variants that are most prominent in embryonic tissues decrease in relative amount while, at the same time, a set of minor embryonic variants increase. The initially minor variants eventually represent as much as 80-90% of their respective histone classes.

Evidence has been provided that the observed changes in histone proportions in the mouse are due primarily to changes in the relative synthetic rates of variants during differentiation (Grove and Zweidler, 1984). The changes in synthetic rates are due mostly to the degree of linkage between the synthesis of a particular variant and DNA replication. On this basis histone variants have been classified roughly into three groups (Grove and Zweidler, 1984; Zweidler, 1984) as described below:

1) Strictly replication-dependent histones. These are induced at the beginning of the S-phase and repressed when replication stops. These variants occur in the highest amounts in rapidly dividing cells.

2) Partially replication-dependent histones. These variants are also induced at the beginning of S-phase but are not completely repressed when replication stops. These occur in high amounts in dividing cells and increase even further in non-dividing cells.

3) Replication-independent histones are constitutively expressed at a low rate. They occur in low amounts in rapidly dividing cells and generally accumulate slowly in non-dividing cells. The replication-independent variant M1 (H2A.Z), however, does not accumulate in non-dividing cells presumably as a result of a relatively high turnover rate (Grove and Zweidler, 1984). The relative amounts of some variants, therefore, may be determined by factors other than just cell growth rate. In fact, the synthesis and relative amounts of some variants appears to be responsive to tissue-specific signals (Grove and Zweidler, 1984).

Differential coupling of histone synthesis to DNA replication is also observed in mammalian tissue-culture cells (reviewed in Maxson et al.,

1983a, 1983b; Wu et al., 1984). Wu and Bonner (1981) found that different variants synthesized in cycling tissue-culture cells can be divided, as above, into variants whose synthesis is either dependent or independent of DNA replication. Also consistent with the above results, replication-dependent histone synthesis (also called S-phase regulated synthesis) usually represents the bulk of histone synthesis in dividing cells, while replication-independent synthesis or "basal" synthesis, which occurs relatively constantly throughout the cell cycle, represents only a minor fraction. The proportion of basal histone synthesis, however, may vary between cell types (Groppi and Coffino, 1980; Sittman et al., 1983a; Graves and Marzluff, 1984; Rickles et al., 1982). In yeast all histone synthesis is S-phase regulated (Nurse, 1983).

Urban and Zweidler (1983) suggest that the core histone variants of the chicken can also be divided into different classes. They suggest that the "late" chicken variants, H3.3 and H2B.2, which increase in relative amounts in non-dividing cells, represent examples of replication-independent variants. This is consistent with the finding that the H3.3 subtype in mouse and other mammals (Zweidler, 1984; Franklin and Zweidler, 1977), to which the chicken H3.3 is structurally analogous (Urban et al., 1979), is also a replication-independent variant (Grove and Zweidler, 1984; Sittman et al., 1983a; Wu and Bonner, 1981).

The isolation of a cDNA clone, pH2A.F, which contains sequences that presumably code for the chicken M1/H2A.Z variant, is discussed in Chapter 3 (Section 3.2.3 and 3.2.4; Harvey et al., 1983). Analysis of the H2A.F gene suggests that its protein product represents another example of a replication-independent chicken histone variant (Robins et al., 1985), consistent with findings in mammals that the M1/H2A.Z variants are not S-phase regulated (Grove and Zweidler, 1984; Wu and Bonner, 1981). Genes encoding the chicken H3.3 variant have also recently been isolated (Engel et al., 1982; Brush et al., 1985). The structure and expression

of genes encoding replication-independent chicken variants will be discussed in Sections 1.5.3(d) and 1.6.3(b).

In addition to non-allelic core histone variants being expressed differentially throughout development and throughout the cell cycle, some variants are only expressed in particular tissues (Isenberg, 1979; Von Holt et al., 1979). Many species of sea urchin have been found to express sperm-specific histone variants (Von Holt et al., 1984). Mammals contain testis-specific variants of the core histones H2A, H2B and H3 (Zweidler, 1984; Trostle-Weige et al., 1982, 1984; Shires et al., 1975).

1.2.4 Functional Significance of Variants

The presence of differentially expressed histone variants suggests that these variants may be functionally different, particularly in their effects on nucleosome structure. Evidence to support this idea comes from studies on sea urchin histones. During sea urchin development it was found that changes that occur in the histone composition of nucleosomes are accompanied by changes in the stability of the nucleosome core particle to thermal denaturation and nuclease digestion (Simpson, 1981). In addition, it was demonstrated that the presence of variant core histones in sea urchin sperm also leads to a change in core particle properties (Simpson and Bergman, 1980).

(a) Distribution of variants

An area of considerable interest with regard to histone variant function, is the distribution of variants into different functional states of chromatin. An organism particularly suited to this sort of analysis is the ciliated protozoan Tetrahymena thermophila. Vegetative cells of this organism contain two types of nuclei of common origin, these are the transcriptionally active macronucleus and the transcriptionally inactive micronucleus (Bannon and Gorovsky, 1984). Two of the core histone variants

found in Tetrahymena are specific to the active macronucleus (Allis et al., 1980). The two variants, hv1 and hv2, appear to represent H2A and H3 variants respectively. The timing of appearance of hv1 in developing new macronuclei during conjugation correlates closely with the onset of gene expression in these nuclei (Wenkert and Allis, 1984), hence, hv1 may be involved in the processes of gene activation.

Anti-sera specific for the Tetrahymena hv1 protein detects a determinant that is enriched in the nucleolar chromatin of several mammalian cell lines (Allis et al., 1982). The detected protein was most likely the previously discussed (Section 1.2.3) H2A variant H2A.Z (Allis et al., 1982). H2A.Z has independently been shown to be enriched in active chromatin by Gabrielli et al. (1981). The conserved nature of H2A.Z proteins through evolution has been demonstrated by Wu et al. (1982). Comparison of tryptic peptides of H2A variants from mouse and sea urchin revealed that the H2A.Z variant was more highly conserved than any other H2A variant (Wu et al., 1982). H2A.Z variants are also quite distinct in primary structure from other H2A variants. The sequence of the chicken H2A.Z-like, H2A.F protein (Harvey et al., 1983 ; derived from gene sequence), for example,was found to be 40% divergent from the major chicken H2A variant. The divergent nature of H2A.Z proteins has been demonstrated by others, as determined by peptide mapping (Wu et al., 1982, 1984) and partial determination of amino acid sequence (Ball et al., 1983), and may be relevant to their location in active chromatin. The presence of H2A.Z may affect the properties of nucleosome core particles (A. Zweidler, cited in Grove and Zweidler, 1984).

Another H2A variant, which is apparently the variant H2A.X (Wu et al., 1984; Zweidler, 1984), is also enriched in active chromatin (Bhatnagar et al., 1984). Both the above discussed H2A.Z and H2A.X proteins are replication-independent variants (Wu and Bonner, 1981).

(b) Function in active chromatin

The presence of specific core histone variants in active chromatin may be necessary to permit the conformational changes that take place in nucleosomes during transcription (reviewed in Reeves, 1984; Tsanev, 1983). Several genes have been shown to exhibit a reversible increase in susceptibility to micrococcal nuclease upon induction of transcription, which may reflect a reversible conformational change in nucleosome structure (Koropatnick et al., 1983; Lohr, 1983; Anderson et al., 1983; Smith et al., 1984b). Such reversible changes in nucleosome structure have been shown to occur in active ribosomal genes of Physarum (Prior et al., 1983). Coding sequences of transcriptionally active Physarum ribosomal genes were found to be complexed in an extended nucleosomal particle ("A particle" or "lexosome"). This particle is proposed to result from the opening of a normal nucleosome to form two smaller symmetrical particle bodies connected by a 50bp nucleoprotein bridge structure. The transition between the normal nucleosome and lexosome is hypothesized to be reversible and directly correlated with the transcriptional state of the gene.

The model presented above by Prior et al. (1983) is an extension of earlier models involving the unfolding of nucleosomes into symmetrically paired "half-nucleosomes" in active regions of DNA (Weintraub et al., 1976). It has been suggested that "half-nucleosomes" are also the structural units of active minichromosomes assembled within the germinal vesicle of Xenopus oocytes (Ryoji and Worcel, 1984, 1985; Gargiulo et al., 1984).

Studies on the nucleosome core particles of mouse myeloma cells, by Baer and Rhodes (1983), also suggest that transcribed nucleosomes are structurally altered. Their results indicated that nucleosomes containing RNA polymerase II transcribed genes may contain reduced amounts of H2A and H2B, approximately half that of a full core octamer suggesting the dissociation of one of the two H2A-H2B dimers from the nucleosome. During

very active transcription in some genes, such as the hsp 70 heat shock genes of Drosophila, the suggestion is that all histone proteins are removed from the coding region (Karpov et al., 1984).

The presence of specific variants in nucleosomes of active chromatin may also be necessary to permit the relaxation of higher order structures in such chromatin. Transcriptionally active chromatin has been extensively reviewed (Reeves, 1984; Tsanev, 1983; Weisbrod, 1982; Igo-Kemenes et al., 1982; Butler, 1983; Thomas, 1983; Cartwright et al., 1982). The amino-terminal regions of core histones may be involved in histone-DNA interactions between adjacent nucleosome cores in higher order chromatin structures and hence may play a role in stabilizing these structures (Allan et al., 1982; Harborne and Allan, 1983; McGhee et al., 1980, 1983). The presence in chromatin of a histone variant such as H2A.Z/H2A.F, which has a less basic amino-terminal domain than other H2A variants, could hence result in the destabilization of chromatin (Ball et al., 1983; Harvey et al., 1983).

As well as the presence of specific variants, certain modifications of core histones have been implicated in the control of gene activity (reviewed in Reeves, 1984). It is suggested that each of these modifications, viz., ubiquination, poly(ADP)-ribosylation and hyper-acetylation, are involved in disrupting or preventing the formation of higher order chromatin configurations (Levinger and Varshavsky, 1982; Doenecke and Gallwitz, 1982; Poirier et al., 1982; Aubin et al., 1983). It has been suggested that ubiquination may "tag" nucleosomes of active genes for proteolytic removal of histones (Varshavsky et al., 1983) and that hyper-acetylation plays a role in inducing conformational changes in nucleosome structure (Reeves, 1984). Active chromatin-specific variants could be responsible for targeting nucleosomes for these modifications (Ball et al., 1983).

Although H2A and H2B variants in yeast, which differ by only a few

amino acids, were found by genetic analysis not to be functionally different (Grunstein et al., 1984; Rykowski et al., 1981), this does not distract from the potential roles of the more extreme variants, such as H2A.Z/H2A.F, in different functional states of chromatin in other eukaryotes.

1.3 H1 HISTONE PROTEINS

1.3.1 Role in Chromatin Structure

The H1 histones represent a group of lysine-rich chromosomal proteins with an approximate protein length of 200 amino acids (reviewed in Isenberg, 1979; Von Holt et al., 1979; Cole, 1984). They consist of a short basic amino-terminal region, a central hydrophobic region of approximately 80 residues and a basic carboxyl-terminal tail. The central domain, which usually commences at amino acid 30 to 40 of the protein chain (Allan et al., 1980; Levy et al., 1982; Cole et al., 1984), can be folded into a globular structure (Hartman et al., 1977; Tiktopulo et al., 1982; Crane-Robinson and Privalov, 1983) and is the most highly conserved region of the protein between species (Section 6.2; Fig. 6.2).

H1 histones associate with the nucleosome at the point where DNA enters and exits the nucleosome core particle (Simpson, 1978; Thoma et al., 1979; Boulikas et al., 1980; Belyavsky et al., 1980) resulting in, on average, one H1 molecule per nucleosome (Bates and Thomas, 1981). It appears to be the globular domain that rests on the nucleosome core and this domain alone is capable of closing the two turns of DNA wound around the core (Allan et al., 1980).

The bulk of interphase chromatin in vivo exists as 30 nm fibres (Butler, 1983). The presence of H1 is necessary for the condensation of nucleosome chains into such fibres (Thoma and Koller, 1981, 1977; Thoma

et al., 1979; reviewed in Reeves, 1984; Butler, 1983; Thomas, 1983, 1984). The 30nm fibre is formed from the winding of the nucleosome filament into a shallow supercoiled solenoid with a pitch of approximately 11 nm, diameter of approximately 30 nm and with approximately six nucleosomes per solenoid turn (Finch and Klug, 1976; McGhee et al., 1983, 1980; Reeves, 1984; Butler, 1983, 1984).

Klug and his colleagues (Thoma et al., 1979; Finch and Klug, 1976) have proposed that H1 molecules are located at the centre of the solenoid where they interact with each other to stabilize the solenoid. McGhee et al. (1983), in contrast, have proposed that the H1 molecules alternate between the inner and outer faces of the solenoid. A recent examination of data on solenoid structure has yielded a model consistent with the former proposal, in which linker DNA between nucleosome core particles forms a reverse-loop into the central hole of the solenoid, hence also locating H1 molecules to the central hole (Butler, 1984). With this model the central solenoid hole, with a diameter of approximately 10.5 nm, provides adequate room to accommodate the longest described length of linker DNA and H1 molecules interacting with nucleosomes and linker DNA. Consistent with this model it appears, from the work of Losa et al. (1984), that physiologically bound H1 is located in the solenoid fibre axis. From these studies it was also proposed that the globular domains of the H1 molecules are involved in H1-H1 contacts within the fibre (Losa et al., 1984). However, reconstitution experiments indicate that the whole H1 molecule is involved in higher order structure formation (Thoma et al., 1983; Allan et al., 1980).

It has also been suggested that additional H1 binding sites, to those described above, may be present on the outside of the solenoid (Losa et al., 1984). Such additional binding sites might be used in the further compaction of the 30 nm fibre that is observed in metaphase chromosomes and to a lesser extent in interphase nuclei (reviewed in Reeves, 1984; Thomas, 1983; Butler, 1983; Igo-Kemenes et al., 1982).

It should be noted that yeast appear to lack H1 proteins (Certa et al., 1984), however, this is consistent with the apparent reduced level of condensation of yeast chromatin compared to that of other eukaryotes (Lohr and Hereford, 1979).

1.3.2 Heterogeneity

Of all the histone proteins the H1 histones are the most heterogeneous (Isenberg, 1979 ; Von Holt et al., 1979) and, as is observed for the core histone proteins, the H1 histones within any one organism represent a set of non-allelic primary structure variants (Cole et al., 1984; Von Holt et al., 1984; Dupressoir and Sautiere, 1984; Risley and Eckhardt, 1981; Smith et al., 1984a; Lennox, 1984; Lennox and Cohen, 1984a; Cole, 1977, 1984; Hohmann, 1978). The expression of these variants is differentially regulated.

1.3.3 Expression of H1 Variants

(a) Developmental switches

The sea urchin species Parechinus angulosus and Strongylocentrotus purpuratus have been shown to contain , respectively, at least five and four different embryonic H1 subtypes (De Groot et al., 1983; Pehrson and Cohen, 1984). The expression of the different H1 variants, from these and other sea urchin species, is precisely regulated throughout embryogenesis, as is also observed for sea urchin core histone variants (Section 1.2.3; reviewed in Von Holt et al., 1984; Maxson et al., 1983a, 1983b).

Switches in H1 variant synthesis during embryogenesis have also been observed in another invertebrate, Urechis caupo. Germinal vesicles and cleavage-stage nuclei are enriched in a variant named H1M. During late

cleavage a second variant, H1E, appears among nuclear histones and gradually replaces H1M in chromatin. The switch from H1M to H1E synthesis occurs after fertilization (Franks and Davis, 1983).

As for the above invertebrate species, the expression of some of the H1 variants in the vertebrate species Xenopus laevis may be under developmental control (Risley and Eckhardt, 1981, 1980). This species contains at least five electrophoretically resolvable H1 variants. Three of these are found in both embryo and adult tissues, but an additional two are found only in adult tissues. Initial synthesis of "adult" variants probably occurs during or following metamorphosis. H1 proteins similar to these latter variants also exist in other amphibian species (Panyim et al., 1971; Alder and Gorovsky, 1975).

Qualitative changes in H1 synthesis, such as those discussed above, also occur during the differentiation of certain cell lines due to the expression of tissue-specific H1 variants. Such variants will be discussed in Section 1.3.3(c).

(b) Vertebrate H1 variant ratios and cell cycle regulation

In contrast to the above examples most changes that occur in vertebrate H1 variant expression are quantitative rather than qualitative. Vertebrate species generally have between four ~~to~~ and six different H1 variants and most of these variants are expressed in all tissues (Panyim et al., 1971; Kinkade, 1969; Bustin and Cole, 1968; Risley and Eckhardt, 1981; Gabrielli et al., 1985; Harris and Smith, 1983; reviewed in Cole, 1984, 1977; Lennox and Cohen, 1984a). Within any one cell type the different proteins are expressed at different relative levels. During differentiation these comparative H1 variant levels are observed to change (Risley and Eckhardt, 1981; Lennox and Cohen, 1984a; Winter, 1984; Winter et al., 1985a; Appels et al., 1972). In addition, the relative level of

any one H1 type differs from cell type to cell type, such that, a particular protein may be the predominant subtype in one tissue but not in another (refs. above).

Most emphasis has so far been placed on the changes that occur in H1 variant ratios during differentiation. This has been most extensively investigated in the mouse and the chicken.

Variant ratios in the mouse: In the mouse the changes that occur during differentiation can be mostly accounted for by differences in the linkage of synthesis of different H1 variants to the cell cycle, and to differences in stability of variants (Lennox and Cohen, 1984a, 1984b, 1983; Pehrson and Cole, 1982). The mouse contains six non-tissue-specific H1 subtypes. In most dividing cells all subtypes were found to be synthesized, with the possible exception of one variant, H1°. In non-dividing cells, where H1 synthesis occurs at a decreased rate, H1 variants were synthesized in quite different proportions, with the synthesis of some variants being almost completely dependent on cell division. The differences in properties found for the mouse H1 variants appear to be well conserved throughout mammalian evolution (Lennox, 1984; Lennox and Cohen, 1984a).

It is evident that the H1 variants of the mouse and other mammalian species can be classified, as are the core histone proteins (Section 1.2.3; Grove and Zweidler, 1984; Zweidler, 1984), according to their dependency on the cell cycle for synthesis. The variant H1° appears to be an example of a replication-independent H1 variant (Chabanas et al., 1983; D'Anna et al., 1985). This variant is quite different in primary sequence from the other non-tissue-specific H1 variants (Smith et al., 1980; Pehrson and Cole, 1981; see also Section 6.2; Fig. 6.2) and is very stable (Lennox and Cohen, 1984a). The properties of H1° are such that, during the development of tissues which have few dividing cells in the adult, it accumulates, whereas some other H1 variants may for instance decrease in relative amounts.

The differences in properties of the mouse H1 variants cannot explain, however, the differences in subtype ratios between different adult tissues with the same proliferative rate. The absence of H1° from non-dividing lymphocytes and macrophages also cannot be explained (Lennox and Cohen, 1983). It is proposed that H1° may play a role in the maintenance of the mature differentiated state of some cells (Gjerset et al., 1982), hence, its absence from the above mentioned cell types may be necessary for the cells to be able to commence rapid cell division.

From the above observations it seems, therefore, that the final H1 composition of a particular cell type is also determined by tissue-specific requirements and signals. It has been found that H1° expression is responsive to induction by hormones in rodent tissues (Gjerset et al., 1982) and by agents that induce differentiation of murine erythroleukemia cells (Osborne and Chabanas, 1984). In addition, changes in patterns of H1 synthesis can be induced by the effects of hormones on mammary tissue in vitro (Hohmann and Cole, 1969, 1971).

Variant ratios in the chicken: In the chicken there are at least five different non-tissue-specific H1 subtypes (Winter, 1984; Winter et al., 1985a; Powell, 1984; Dupressoir and Sautiere, 1984). In addition to these proteins, the chicken also has the H1-related H5 protein which is expressed only in erythroid cells (discussed in Section 1.3.3(c)). As for the mouse and other vertebrate species the ratios of the non-tissue-specific chicken H1 variants vary between tissues (Kinkade, 1969; Panyim et al., 1971; Berdikov et al., 1975) and have been shown to change during the differentiation of certain cell lines (Winter, 1984; Winter et al., 1985a; Appels et al., 1972).

Winter et al. (1984, 1985a) have shown that the different chicken H1 variants differ in both their stability and in the dependency of their synthesis on cell division. In non-dividing cells (total H1 synthesis decreased by approximately 75%) the synthesis of one protein, named H-1c,

was much less affected (reduced) by the cessation of cell division than the other H1 variants. H-1c is no doubt another example of a replication-independent H1 variant, while the other variants appear to be replication dependent. H-1c is also the most stable of the chicken H1 proteins. The properties of the H1 variants were found to account largely for the observed changes in H1 subtype ratios between dividing myoblasts and non-dividing myotubes during *in vitro* myogenesis, in particular, the increase in relative amounts of the H-1c protein. H-1c has also been found to be enriched in other cell types of the chicken with a low rate of cell division (Winter, 1984; Winter *et al.*, 1985a; Berdikov *et al.*, 1975).

Given the properties of H-1c, Winter *et al.* (Winter, 1984; Winter *et al.*, 1985a) suggested that this protein may be similar to H1°. H1°, discussed above, exists in the tissues of all mammalian species examined (Smith *et al.*, 1984a; Harris and Smith, 1983; Lennox and Cohen, 1984a) and may also be present in non-mammalian vertebrate species (Smith *et al.*, 1984a; Srebreva and Zlatanova, 1983). Although there have been conflicting results as to its existence in chicken tissues (Srebreva *et al.*, 1983; Smith *et al.*, 1981), recent work in this laboratory strongly suggests that H1° proteins do not exist in the chicken (Shannon *et al.*, 1985).

In conclusion, the properties of the chicken H1 proteins can account for changes occurring in H1 subtype ratios during differentiation, but as for the mouse these metabolic properties are probably modulated in a tissue-specific manner, resulting in unique subtype ratios in different adult chicken tissues.

Winter *et al.* (Winter, 1984; Winter *et al.*, 1985b) investigated the basis for differential coupling of H1 synthesis to DNA replication in chicken myogenic cells. H1 synthesis was found to be regulated primarily at the level of mRNA availability. The levels of mRNAs for the different H1 subtypes were, therefore, coupled to DNA replication to different extents. The level of translatable mRNA in non-dividing cells for most

H1 variants decreased significantly relative to levels in dividing cells. In contrast, the level of translatable H-1c mRNA changed little in non-dividing cells and hence was almost completely uncoupled from DNA replication. Regulation of histone protein synthesis by mRNA availability is a general mechanism in the control of expression of both core and H1 histone proteins and is discussed further in Section 1.6.3. The potential role of various sequence elements in the expression of genes for chicken H1 proteins, isolated and analyzed for this thesis, is discussed in Chapter 6.

Differential linkage of H1 subtype synthesis to DNA replication no doubt explains the observed presence of a significant level of H1 synthesis outside S-phase in some mammalian tissue-culture cells (Wu and Bonner, 1981; Tarnowka et al., 1978). However, most H1 synthesis, as for core histone synthesis (Section 1.2.3), has been found to be cell cycle regulated (Wu and Bonner, 1981; Tarnowka et al., 1978; Delegeane and Lee, 1982; Waithe et al., 1983; Plumb et al., 1984; Baumbauch et al., 1984; Stein and Stein, 1984) consistent with observations in the mouse and chicken discussed above.

(c) Tissue-specific H1 variants

As previously mentioned (Section 1.3.3(a)) some H1 variants are only synthesized during the differentiation of a particular cell line, viz., they are tissue-specific H1 variants.

The adult testis of many mammalian species contain an H1 variant, H1t, which is not found in any other tissue (Cole et al., 1984; Lennox and Cohen, 1984a, 1984b; Seyedin and Kistler, 1980, 1983; Seyedin et al., 1981; Bucci et al., 1982; Kumaroo and Irvin, 1980; Rao et al., 1983). In rat and mouse testis, H1t first appears in pachytene spermatocytes and replaces pre-existing H1 subtypes in chromatin to become the most abundant H1 subtype (Lennox and Cohen, 1984b; Seyedin and Kistler, 1980; Bucci et al.,

1982). During the development of the spermatid all histones, including H1t, are replaced by other basic proteins (Rodman et al., 1979; Bucci et al., 1982). Tissue-specific H1 variants are also found in the sperm of several sea urchin species (Von Holt et al., 1984; Strickland et al., 1980a, 1980b, 1982a, 1982b). In addition, unusual H1 variants have been isolated from the testes of sea cucumbers (Phelan et al., 1972) and a novel H1 variant has been reported to be associated with meiosis during spore formation in lilies (Sheridan and Stern, 1967). The H1 variants described above may be involved in meiotic chromosome condensation (Lennox and Cohen, 1984a, 1984b).

A tissue-specific H1-related protein, named H5, has been found in the nucleated erythrocytes of several species of bird (Neelin, 1968; Huang et al., 1977; Briand et al., 1980; Yaguchi et al., 1977, 1979; Seligy et al., 1976) including the chicken (Section 1.3.3(b)). Using monoclonal H5 antibodies Shannon et al. (1985, this laboratory) have found no evidence for H5-like proteins in any cell type other than erythroid cells. It appears that H5-like proteins may also be present in the nucleated erythrocytes of non-avian vertebrates (Rutledge et al., 1984; Smith et al., 1984a).

The appearance of H5 occurs early in avian erythropoiesis, with the protein being synthesized in dividing erythrocyte cell precursors (Ruiz-Vazquez and Ruiż-Carrillo, 1982; Appels and Wells, 1972). As cells cease division during erythrocyte maturation H5 accumulates resulting in the partial replacement of the "standard" H1 molecules in chromatin (Ruiz-Carrillo et al., 1974, 1976; Billet and Hindley, 1972; Appels et al., 1972; Weintraub, 1978). The chromatin of mature chicken erythrocytes contains approximately twice as much H5 as it does other H1 subtypes (Bates and Thomas, 1981). The accumulation of H5 correlates with the increased condensation of chromatin during maturation of cells (Huang et al., 1977).

As H5 synthesis is independent of DNA synthesis (Appels and Wells, 1972; Ruiz-Carrillo et al., 1976), H5 can be considered to be a replication-independent H1 variant. The same definition can also be applied to the tissue-specific mammalian H1t variant (Bucci et al., 1982).

The gene for the chicken H5 protein has been isolated and analyzed in this laboratory and by others (Krieg et al., 1982a, 1982b, 1983; Wigley et al., 1985; Dalton et al., 1985; Ruiz-Carrillo et al., 1983; Section 1.5.3(d)). As for the other H1 proteins of the chicken, H5 expression is regulated primarily at the level of mRNA availability (Dalton et al., 1985; Section 1.6.3(b)).

1.3.4 Functional Significance of H1 Variants

(a) Properties of H1 Variants

It has been demonstrated that the appearance and accumulation of several H1 variants can be correlated with increased compaction of whole chromatin during the differentiation of certain cell types, for example, avian H5 during erythropoiesis (Section 1.3.3(c)) and mammalian H1t during spermatogenesis (Bucci et al., 1982; Section 1.3.3(c)). Changes in H1 subtype expression during sea urchin embryogenesis are also correlated with changes in chromatin structure (Poccia et al., 1981; Arceci and Gross, 1980; Chambers et al., 1983). These observations suggest that different H1 variants can differ in the effects they have on chromatin in vivo, viz., that they may differ in their ability to condense chromatin into higher order structures (Section 1.3.1). In the case of H5, the tighter binding of this protein to chromatin than other H1s (Thomas, 1984; Thoma et al., 1983) may play a role.

In agreement with the above proposal, different H1 subtypes have been shown to differ in their abilities to condense DNA and dinucleosomes (Welch

and Cole, 1979, 1980; Liao and Cole, 1981a, 1981b, 1981c; reviewed in Cole, 1984) and H1-depleted chromatin (Hannon et al., 1984). Different H1 subtypes also differ in the nuclease protection they confer upon DNA (Gorka and Lawrence, 1979).

As mentioned previously (Section 1.3.3(b)), the set of H1 proteins in the mouse appear to be well conserved, with regard to metabolic properties, throughout mammalian evolution (Lennox, 1984; Lennox and Cohen, 1984a; Lennox and Abrams, 1982). However, the actual structures of the different H1 variants are conserved to different degrees (determined from electrophoretic mobilities). With regard to functional differences between H1 variants, Lennox and Cohen (1984a) have suggested that those subtypes, which are both relatively unstable and not well conserved in structure between species, may be less involved in interactions with chromatin than other H1 types. These variants may, therefore, be less effective in condensing chromatin than other H1 variants, consistent with the finding that such variants are the most abundant subtypes in mammalian pre-pachytene spermatocytes, the chromatin of which appears to be in a relaxed state (Lennox and Cohen, 1984a, 1984b; Bucci et al., 1982; Seyedin and Kistler, 1980, 1979).

The sorts of differences between H1 variants that could influence their functions are discussed in Chapter 6 (Section 6.2).

(b) Distribution: role in gene expression

As for core histone variants (Section 1.2.4(a)) different H1 variants appear to be non-randomly distributed between regions of chromatin with different transcriptional potentials. This is most clearly demonstrated in comparison of the H1 compositions of the transcriptionally active macronucleus and the transcriptionally inactive micronucleus of Tetrahymena (Bannon and Gorovsky, 1984; Allis et al., 1984). Macronuclei contain a

fairly typical H1 protein, while micronuclei contain three different H1-like polypeptides.

Analysis of calf chromatin also reveals a non-random distribution of H1 variants between chromatin fractions presumed to be of different transcriptional potential (Huang and Cole, 1984). As previously mentioned (Section 1.2.4(b)) active chromatin appears to be in a more relaxed state than inactive chromatin (Kimura et al., 1983; Smith et al., 1984c; reviewed in Weisbrod, 1982; Cartwright et al., 1982; Thomas, 1983; Igo-Kemenes et al., 1982; Butler, 1983; Tsanev, 1983; Reeves, 1984) and may in some cases have properties similar to a 10 nm chromatin filament (Kimura et al., 1983; Smith et al., 1984c). Inactive chromatin represents the bulk of chromatin in a cell. The structure of this chromatin was described in Section 1.3.1. Huang and Cole (1984) found that the degree of condensation of the chromatin fraction in which a particular calf H1 subtype was enriched, correlated with its ability to aggregate DNA in vitro (Liao and Cole, 1981a, 1981b, 1981c). An H1 variant, named H1c or CTL-1, found to be enriched in active chromatin, for example, was the least effective in DNA condensation. It appears, therefore, that the different H1 subtypes may play a role in the formation of the different functional chromatin states in vivo. As the relaxation of structure in active chromatin is probably necessary for transcriptional activity (Hannon et al., 1984), the different H1 variants may, therefore, also be playing a role in selective gene expression. It is apparent that the differential synthesis of H1 variants may be a means of establishing tissue-specific chromatin states.

Non-random distribution has also recently been reported in the mouse, where it appears that H1° is preferentially associated with transcriptionally repressed chromatin (Roche et al., 1985). Non-random distribution of H1 variants may explain the results of Weintraub (1984) that suggest that H1 in active chromatin is bound differently from H1 in repressed chromatin.

In addition to variation in H1 composition, differences in total H1 amounts may also play a role in determining structural differences between active and inactive chromatin. It is generally agreed that active chromatin contains reduced amounts of H1 compared to inactive (Huang and Cole, 1984; Karpov et al., 1984; Smith et al., 1984c; Reeves, 1984). Reduced amounts of H1 may result from competition between the binding of H1 proteins and transcription factors to genes during assembly of active chromatin (Schlissel and Brown, 1984). Schlissel and Brown (1984) have also suggested that factors, such as proteases which degrade H1 molecules, may be involved in gene activation. H1 variants enriched in active chromatin could be more susceptible to degradation than those of inactive chromatin.

As well as affecting higher order chromatin structure, the specific H1 composition of active chromatin may be necessary to permit the apparent alterations in nucleosome structure that occur during transcription (Section 1.2.4(b)).

How different H1 variants are distributed into appropriate regions of chromatin is not known. One possible means is via interaction with non-histone chromosomal proteins. Different H1 subtypes have been shown to have different affinities for certain high mobility group (HMG) proteins (Smerdon and Isenberg, 1976) that appear to be associated with active chromatin (Cartwright et al., 1982; Stein et al., 1983). Different subtypes may also have different affinities for sequences shown to be preferentially bound by H1 (Renz and Day, 1976; Berent and Sevall, 1984). The presence of specific histone variants and modifications in active nucleosomes (Sections 1.2.4(a) and 1.2.4(b)) could be another means of affecting H1 subtype binding.

(c) H1 modifications

H1 subtypes enriched in active chromatin may be more susceptible to modifications, such as poly(ADP)-ribosylation, suggested to facilitate relaxation of chromatin structure (Poirier et al., 1982; Aubin et al., 1983). Phosphorylation of H1 (reviewed in Hohmann, 1983; Lennox and Cohen, 1984a) also appears to be involved in destabilization of chromatin (Kaplan et al., 1984), possibly in some cases by affecting H1 protein conformation (Lennox et al., 1982; Billings et al., 1979; Glover et al., 1981). In non-dividing mammalian cells phosphorylation of H1 can be hormonally-induced (reviewed in Hohmann, 1983; Cooper and Spaulding, 1985). The absence of hormonally-induced phosphorylation sites from some H1 subtypes (Langan et al., 1971) represents a means for limiting hormone-dependent gene transcription to selected regions of chromatin via non-random distribution of H1 subtypes.

1.4 HISTONE GENES

Given the precise regulation of expression of both core and H1 histone proteins (Sections 1.2.3 and 1.3.3) and the potential involvement of these proteins, themselves, in the processes of selective gene expression (Sections 1.2.4 and 1.3.4), much emphasis has been placed on the analysis of genes coding for histone proteins. Histone genes have been analyzed in a wide variety of organisms and have been the subject of several reviews (Birnstiel, 1984; Old and Woodland, 1984; Maxson et al., 1983a, 1983b; Hentschel and Birnstiel, 1981; Kedes, 1979). Work in our laboratory and that described in this thesis involves an analysis of the histone genes from the chicken.

Current knowledge of histone gene organization and expression is discussed in the following sections.

1.5 HISTONE GENE ORGANIZATION

Studies on histone gene organization (Old and Woodland, 1984; Maxson et al., 1983a, 1983b; Hentschel and Birnstiel, 1981; Kedes, 1979; and below) have shown that the genes for the five classes of histone protein are usually clustered but that the arrangement of these genes varies even between closely related organisms. Organization ranges from randomly arranged dispersed clusters of genes to highly regular, tandemly repeated quintets. Topologies of histone genes may be the most variable of any gene family. Gene copy numbers also vary markedly between species. A summary of histone gene organization in different species is given below.

1.5.1 Lower Eukaryotes

(a) Yeast

The yeast, Saccharomyces cerevisiae, has a low histone gene copy number, containing only two copies of each of the genes coding for H2A, H2B, H3 and H4 proteins (Smith and Murray, 1983; Hereford et al., 1979). H1 genes have not been isolated from yeast, consistent with a similar lack of detection of H1 proteins (Certa et al., 1984; Section 1.3.1). The yeast genes are arranged in duplicate H2A/H2B and H3/H4 gene pairs that are not closely linked. The members of each gene pair are divergently transcribed and separated by 600 to 800 bp (Smith and Andrésson, 1983; Choe et al., 1982; Wallis et al., 1980; Smith, 1984).

(b) Other organisms

Investigation of H3 and H4 genes from Neurospora crassa (Woudt et al., 1983) reveals a single linked H3/H4 gene pair. These genes, as observed in

yeast, are divergently transcribed. Neurospora histone genes are unusual in that, unlike most histone genes (Section 1.6.2(c)), they contain introns. Other examples of intron-containing histone genes are found among chicken (Section 1.5.3(d)) and mammalian genes (Section 1.5.3(e)).

As for yeast and Neurospora, the histone genes examined in the ciliated protozoan Tetrahymena thermophila are present in low copy number (Bannon et al., 1983, 1984; Bannon and Gorovsky, 1984) and are not all closely linked. The two H4 genes of this organism, for example, are located on separate chromosomes.

1.5.2 Invertebrate Animals

(a) Sea Urchin

Histone gene organization has been analyzed in several sea urchin species (reviewed in Kedes, 1979; Hentschel and Birnstiel, 1981; Maxson et al., 1983a, 1983b). The majority of sea urchin histone genes are organized into tandemly repeated units. Each unit is approximately 6 to 7 kb (kilobase pairs) in length, depending on the species, and contains one each of the five types of histone gene, arranged in the order $\overrightarrow{H1}$-$\overrightarrow{H4}$-$\overrightarrow{H2B}$-$\overrightarrow{H3}$-$\overrightarrow{H2A}$ (arrows indicate the direction of transcription). The units are reiterated several hundred times. Several sea urchin species contain more than one type of non-allelic repeating unit, for example, in Lytechinus pictus (Cohn and Kedes, 1979a, 1979b; Holt and Childs, 1984) and Psammechinus miliaris (Gross et al., 1976; Birnstiel et al., 1978; Busslinger et al., 1980). The units within any one cluster type are essentially homogeneous.

Late histone genes: The histone genes described above are expressed during the period of rapid cleavage between fertilization and late blastula (Maxson et al., 1983a, 1983b), and code for the early sea urchin histone variants discussed in Section 1.2.3. After the late blastula stage the

activity of early genes is sharply curtailed and mRNA from a second class of histone genes becomes predominant (Maxson et al., 1983a, 1983b; Knowles and Childs, 1984; Section 1.6.4). These genes code for the previously described late histone protein variants (Section 1.2.3).

Several late histone genes have been isolated from the sea urchin species S. purpuratus (Maxson et al., 1983a, 1983b, 1983c) and L. pictus (Childs et al., 1982; Roberts et al., 1984). Unlike early histone genes, members of the late histone gene family are present in about ten copies per haploid genome. In addition, the genes are not arranged into tandemly repeated quintets but exist as isolated single genes or heterotypic pairs.

Sequence analysis of late H2B genes from S. purpuratus revealed considerable variation between the genes (Maxson et al., 1983a, 1983b). In contrast, variation between isocoding L. pictus late genes is mainly restricted to sequences outside coding regions, while coding regions themselves remain highly conserved (Roberts et al., 1984). Roberts et al. (1984) suggested that such conservation of sequences is probably the result of gene conversion mechanisms. There is also evidence for gene conversion among non-histone genes (Shen et al., 1981; Slightom et al., 1980; Baltimore, 1981; Weiss et al., 1983). A recombination mechanism such as unequal crossing-over has been suggested to be responsible for the high degree of sequence conservation between the tandemly repeated early genes (Kedes, 1979; Hentschel and Birnstiel, 1981; Coen et al., 1982; Dover et al., 1982). Unequal crossing-over also explains the striking differences in repetition frequency of histone gene repeat types observed between closely related species (Maxson et al., 1983b).

Orphon histone genes: As well as tandemly repeated early and dispersed late histone genes, sea urchins contain, what have been termed orphon histone genes (Maxson et al., 1983b; Childs et al., 1981; Liebermann et al., 1983). Orphon histone genes are nearly identical in DNA sequence to histone genes contained within tandem repeats but they exist as solitary

elements remote from the tandem cluster. There are a number of orphons (5-20) for each of the five histone coding regions (Childs et al., 1981). Drosophila melanogaster was also shown to contain orphon histone genes (Childs et al., 1981). DNA sequence analysis of a sea urchin H2B gene orphon suggests that translocation of this gene from the tandem cluster may have occurred via an RNA intermediate (Liebermann et al., 1983). Such mechanisms have been described for the formation of processed pseudogenes in other gene systems (Lewis, 1983; Sharp, 1983). The generation of other orphon histone genes, however, cannot be explained by such mechanisms (Maxson et al., 1983b). As orphon genes are removed from the homogenizing mechanisms associated with tandemly repeated genes (see above), they are relatively free to diverge in sequence and function and may come under different regulatory controls (Maxson et al., 1983b).

(b) Drosophila

Most of the histone genes of the fruit fly Drosophila melanogaster are organized into tandemly repeated quintets, as are the bulk of sea urchin histone genes (Lifton et al., 1977; Goldberg, 1979; Karp, 1979; Kedes, 1979). The quintets are reiterated approximately one hundred times and the gene order within each repeat is $\overleftarrow{H3}$-$\overrightarrow{H4}$-$\overleftarrow{H2A}$-$\overrightarrow{H2B}$-$\overleftarrow{H1}$. Two major types of repeat exist, a 4.8 kb and a 5.0 kb repeat, that differ by the presence of a 208 bp insertion in the H1-H3 gene spacer region of the longer repeat. Different strains of Drosophila differ in their relative amounts of the two repeat types (Strausbaugh and Weinberg, 1982).

Histone gene diversity in Drosophila may be generated via the formation of orphons as discussed above (Childs et al., 1981; Maxson et al., 1983b; Section 1.5.2(a)) or via the insertion of transposable elements. A transposable element was found to be inserted into the "TATA"

box sequences of two Drosophila H3 genes (Ikenaga and Saigo, 1982). Such an insertion could drastically affect the expression of these genes.

Preliminary investigation of the histone genes of a third invertebrate species, the brine shrimp Artemia, reveals, as for sea urchin early genes and most Drosophila histone genes, tandemly repeated gene clusters (Bagshaw et al., 1984).

1.5.3 Vertebrates

(a) Newt

The newt Notophthalmus viridescens has a high histone gene copy number and its genes are organized into a 9.0 kb repeating unit. This unit is reiterated six to eight hundred times and contains genes in the order $\overrightarrow{H1}$-$\overrightarrow{H3}$-$\overleftarrow{H2B}$-$\overrightarrow{H2A}$-$\overrightarrow{H4}$. The repeating units are, however, not adjacent in the genome but are separated by approximately 50 kb of spacer sequences (Stephenson et al., 1981a, 1981b).

(b) Xenopus

In Xenopus species histone genes are reiterated approximately forty to one hundred times. Gene organization varies between species. In one species, X. borealis, 70% of the genes are found in a single homogeneous repeat type with a minimum length of 16 kb, but it has not been determined whether these gene clusters are repeated in tandem (Turner and Woodland, 1983). In another species, X. laevis, the histone genes are organized into at least three different cluster types (Destrée et al., 1984). Different cluster types appear to be associated with the genes for different Xenopus H1 variants (Destrée et al., 1984; Turner et al., 1983; Zernik et al., 1980). Such organization may be relevant to the differential expression of

protein products from the various H1 genes (Risley and Eckhardt, 1981; Van Dongen et al., 1984; Section 1.3.3(a)). A non-uniform arrangement of histone genes is also observed in X. tropicalis (Ruberti et al., 1982).

(c) Trout

The histone genes of the rainbow trout, Salmo gairdnerii, are clustered in a repeating quintet unit of approximately 10 kb which is repeated approximately one hundred and forty five times. There is, however, no evidence for tandem linkage and clusters do not appear to be homogeneous (Conner et al., 1984a, 1984b).

In each of the vertebrates discussed above a certain degree of order in histone gene organization is maintained, but in each case there is deviation from the classical organization of tandemly repeated clusters observed in the invertebrate animal species. In contrast to the other vertebrates, the histone genes from both mammalian and bird species, are quite disorganized (although they do exhibit some local order), in particular, the histone genes are not organized into a quintet structure.

(d) Chicken

There are between six to ten copies of each histone gene in the chicken (D'Andrea et al., 1985; Ruiz-Carrillo et al., 1983; Section 4.2.6). The organization of these genes has been extensively analyzed in this laboratory and by others (Wang et al., 1985; D'Andrea et al., 1985; Wells et al., 1983; Harvey and Wells, 1984; Harvey et al., 1981; Sugarman et al., 1983; Engel, 1984; See Chapter 4). The majority of histone genes are organized into several dissimilar clusters of genes (D'Andrea et al., 1985; D'Andrea, 1985; Figs. 4.4, 4.19 and 4.26). Each cluster is separated by variable lengths of DNA and contains between two and eight histone genes.

Characterized histone gene-containing DNA covers approximately 175 kb, but this DNA does not appear to exist as one continuous stretch in the genome (D'Andrea et al., 1985; discussed in Section 4.2.5(c)).

No long range order is evident in the chicken histone genes but there are preferred associations of genes, such as H2A/H2B divergently transcribed gene pairs and H3/H4 gene pairs. Another feature of the chicken histone genes is the presence of three gene clusters in which genes are symmetrically ordered around central H3 genes. In one such cluster where an H2A/H4 gene pair is duplicated and inverted around a central H3 gene (Fig. 4.4), the boundaries of the duplication coincide with repeated sequences (Wang et al., 1985). The boundaries also contain highly conserved gene promoter elements. It is suggested that interaction of the transcriptional machinery with histone genes may be connected with recombination in promoter regions. Vitelli and Weinberg (1983) have also characterized a rare duplication/inversion in an unusual histone gene cluster in sea urchin DNA. The breakpoints of this event are also adjacent to promoter elements.

Sequence analysis of the clustered chicken core histone genes described above (Sugarman et al., 1983; Wang et al., 1985; Harvey et al., 1982; Grandy et al., 1982; Engel et al., 1982; D'Andrea et al., 1981; Sturm, 1985; Section 4.2.5(b)) reveals that these genes code for protein variants, described by Urban et al. (1979), whose synthesis is probably dependent on DNA replication (replication-dependent variants; Section 1.2.3; Urban and Zweidler, 1983). This is consistent with findings in this laboratory that these genes are S-phase regulated (Dalton et al., 1985; Section 1.6.3). The clustered H1 genes characterized for this thesis also appear to be S-phase regulated as discussed in Chapter 6 (Section 6.4.1; Dalton et al., 1985).

Replication-independent genes: Histone genes have been isolated, from the chicken genome, which are separated from the major clusters of genes

described above. Each of these genes codes for a replication-independent protein variant (Sections 1.2.3 and 1.3.3). One of these genes codes for the H1-related, tissue-specific H5 protein (Section 1.3.3(c)). H5 cDNA clones were isolated, in this laboratory (Krieg et al., 1982a, 1982b) and by others (Ruiz-Vazquez and Ruiz-Carrillo, 1982), and these clones were in turn used to isolate H5 gene-containing genomal clones (Krieg et al., 1983; Ruiz-Carrillo et al., 1983). The H5 gene is present as a single copy and is not closely linked to any other histone gene. The H5 gene is expressed exclusively in erythroid cells (Shannon et al., 1985). Another apparently "solitary" histone gene isolated in this laboratory is the H2A.F gene (Robins et al., 1985; Harvey et al., 1983). This gene most probably codes for the replication-independent M1/H2A.Z, H2A variant described by Urban et al. (1979; Section 1.2.3). The isolation of H2A.F cDNA sequences is discussed in Chapter 3 (Section 3.2.3 and 3.2.4).

In contrast to most sequenced histone genes (Maxson et al., 1983a, 1983b; Hentschel and Birnstiel, 1981) the H2A.F gene contains introns (Robins et al., 1985). Two other solitary chicken histone genes also contain introns (Engel et al., 1982; Brush et al., 1985). Both genes, the H3.3A and H3.3B genes, code for the replication-independent H3 variant, H3.3 (Urban et al., 1979; Urban and Zweidler, 1983; Section 1.2.3). Although the two H3.3 genes code for identical proteins they differ in several ways, including in primary DNA sequence and in location and sizes of introns. Human genes encoding the H3.3 variant probably also contain introns (Wells and Kedes, 1985; Section 1.5.3(e)). The only other histone genes found to contain introns are those of Neurospora (Section 1.5.1(b); Woudt et al., 1983).

The evolution of the replication-independent H5 gene is discussed in Chapter 6 (Section 6.6).

(e) Mammals

Histone genes in mammals are reiterated ten to twenty times. Analysis of histone gene-containing recombinants from mouse (Sittman et al., 1981, 1983b; Marzluff and Graves, 1984; Graves et al., 1985) and human (Stein et al., 1984; Zhong et al., 1983; Heintz et al., 1981; Sierra et al., 1982a, 1982b; Carozzi et al., 1984; Zwollo et al., 1984) genomal libraries reveals that the mammalian histone genes are organized, as are the chicken histone genes, into small variable clusters with no obvious repeat evident. As for the clustered chicken genes, the clustered human and mouse histone genes are S-phase regulated (Marzluff and Graves, 1984; Graves et al., 1985; Stein et al., 1984; Stein and Stein, 1984).

cDNA clones coding for the replication-independent human H3.3 variant (Wu et al., 1984) have recently been isolated (Wells and Kedes, 1985). Analysis of a cDNA clone derived from an H3.3 mRNA precursor reveals the presence of an intervening sequence. Hence the H3.3 gene probably contains introns as do the replication-independent H2A.F and H3.3 chicken genes.

Human H2A and H2B pseudogenes have also been isolated (Marashi et al., 1984).

1.5.4 Plants

Some preliminary investigations of the organization of histone genes in rice (Thomas and Padayatty, 1983) and wheat (Tabata et al., 1983, 1984; Tabata and Iwabuchi, 1984) have been made. Histone genes in wheat are reiterated approximately fifty to sixty times per hexaploid genome. As for the higher vertebrate species there was no evidence for a repeating unit of histone genes.

1.5.5 Summary

Histone gene copy number and organization vary considerably between organisms. Where histone gene copy number is high the genes are organized into a highly reiterated quintet structure. Reduced histone gene number correlates with a breakdown of such organization, which is most evident in the higher vertebrates, birds and mammals (Sections 1.5.3(d) and 1.5.3(e)), and in the lower eukaryotic organisms (Section 1.5.1). Histone genes in these organisms are however, not completely disorganized. In the chicken for example, preferred histone gene arrangements are observed (Section 1.5.3(d)). Histone gene copy numbers in animal species appear to vary according to the requirements for histone gene transcription during embryogenesis (Old and Woodland, 1984).

Different recombination mechanisms have been postulated as the major mechanisms for maintaining homogeneity between members of either tandemly repeated histone gene families or dispersed histone gene families, as discussed in Section 1.5.2(a) (Maxson *et al.*, 1983b; Roberts *et al.*, 1984). Stretches of alternating co-polymeric DNA found in the non-coding spacer DNA between histone genes of tandemly repeated sea urchin and *Drosophila* histone genes may act as foci for homogenizing recombination events (Kedes, 1979).

As well as variation between organisms, different histone gene organizations can exist within a single organism. Genes that are organized differently are also regulated differently, for example, as is observed for sea urchin early and late histone genes (Section 1.5.2(a)),and the chicken (and presumably mammalian) replication-dependent and -independent genes (Sections 1.5.3(d) and 1.5.3(e)). Different *Xenopus laevis* cluster types may also be differentially regulated (Section 1.5.3(b)).

Several processes such as orphon generation and transposon insertion, observed in sea urchin species and *Drosophila* (Sections 1.5.2(a) and

1.5.2(b)), and proposed recombination events associated with promoter elements (Wang et al., 1985; Section 1.5.3(d)), could be involved in the generation of histone gene diversity within an organism.

1.6 HISTONE GENE EXPRESSION

Unlike most genes transcribed by RNA polymerase II (Breathnach and Chambon, 1981; Nevins, 1983) most histone genes lack introns, their expression is cell cycle regulated and their mRNAs are not polyadenylated (Maxson et al., 1983a, 1983b; Old and Woodland, 1984; Hentschel and Birnstiel, 1981). Current knowledge of the regulation of histone gene expression is given below.

1.6.1 Requirements for Transcription

(a) Sequences

The 5' termini of sea urchin histone mRNAs map within a sequence having the consensus 5' $\begin{matrix}T\\C\end{matrix}$CATTC$\begin{matrix}G\\A\end{matrix}$ 3' (Hentschel et al., 1980; Sures et al., 1980). This sequence, the "cap site", is not well conserved in the histone genes of other organisms. Approximately 20 bp upstream from the cap sites of histone genes is an AT-rich sequence with a consensus 5' TATAAATA 3' (Hentschel and Birnstiel, 1981). This sequence referred to as the "TATA" box is present in nearly all RNA polymerase II transcribed genes and its main function appears to be to specify the site of initiation of transcription at a fixed position downstream (Breathnach and Chambon, 1981; Nevins, 1983; Grosschedl and Birnstiel, 1980a, 1980b, 1982; Grosschedl et al., 1981; Etkin and Di Berardino, 1983). This sequence is also required for an optimal rate of transcription. Upstream from the "TATA" box, approximately 70 to 80 bp upstream from transcription initiation sites,

many histone genes have a sequence with the consensus 5' CCAAT 3' (Hentschel and Birnstiel, 1981). Multiple copies of this sequence are found in some histone genes (Wang et al., 1985; Conner et al., 1984b and refs. therein). This sequence is also found in other RNA polymerase II transcribed genes (Breathnach and Chambon, 1981) and appears to be part of the RNA polymerase II promoter (Dierks et al., 1983; Grosveld et al., 1982). It has recently been suggested that this sequence may play a role in the developmental control of globin gene expression (Gelinas et al., 1985).

Between "TATA" and "CCAAT" boxes is a sequence 5' GATCC 3', which is mainly restricted to sea urchin histone genes (Hentschel and Birnstiel, 1981) and may be necessary for proper expression of these genes (Etkin and Di Berardino, 1983; Etkin and Maxson, 1980). More distal sequence elements also play a role in sea urchin histone gene transcription (Grosschedl et al., 1983; Grosschedl and Birnstiel, 1980b). A region of DNA, located between -165 and -111 bp from the transcription initiation site, has been shown to be required for maximal transcription of a sea urchin H2A gene when expressed in Xenopus oocyte nuclei (Grosschedl et al., 1983). This region of DNA, called the H2A gene "modulator", contains two stretches of sequence conserved between sea urchin species. One of these conserved stretches shows homology to viral enhancer sequences. Similar sequences to those conserved within the H2A gene modulator are also found upstream of some chicken H2A genes (Wang et al., 1985) and hence may not be specific to just sea urchin genes.

Ubiquitous 5' elements associated with specific histone gene types have been observed in H4 genes (Clerc et al., 1983), H2B genes (Harvey et al., 1982) and H1 genes, as discussed in Chapter 6 (Section 6.3.2; Coles and Wells, 1985). The H4 and H2B gene elements have been shown to be necessary for efficient transcription (Clerc et al., 1983; J. Mous and H. Stunnenberg, cited in Birnstiel, 1984). A role for the H1 gene-conserved

sequence is discussed in Section 6.4.1. Late sea urchin H4 genes and late sea urchin H3 genes (Section 1.5.2(a)) have also been shown to contain conserved 5' elements (Roberts et al., 1984). These sequences are not present in their early gene counterparts and hence may be involved in the differential expression of early and late sea urchin histone genes (Sections 1.5.2(a) and 1.6.4).

There are numerous examples in non-histone gene systems where specific 5' sequences are involved in gene expression (Clerc et al., 1983; Grosschedl et al., 1983; Guarente, 1984; Khoury and Gruss, 1983; Perry, 1984; North, 1984; Hogan, 1983; Groner et al., 1984; Yaniv, 1984; Sommerville, 1984; Reeder, 1984; Parker, 1983; Boss, 1983 and refs. therein).

Sequences within histone gene coding regions may also be necessary for transcriptional activity (Gargiulo et al., 1984; Birnstiel, 1984). Internal sequences have been found to be involved in the expression of human globin genes (Charnay et al., 1984), Xenopus 5S RNA genes (Brown, 1984), immunoglobulin genes (Banerji et al., 1983; Queen and Baltimore, 1983) and the Adenovirus E1A gene (Osborne et al., 1984). Histone gene coding regions may be the binding sites of factors involved in the assembly of active histone gene chromatin (Gargiulo et al., 1984).

(b) Transcription factors

An RNA polymerase II transcription factor preparation, from Drosophila Kc cells, has been shown to contain a component that binds to the promoter-initiation region of histone H3 and H4 genes and to actin 5C genes (Parker and Topol, 1984a). The binding region includes the "TATA" box, cap site and part of the leader region. It is suggested that the component is involved in promoter recognition and in the initiation of transcription by RNA polymerase II. Interestingly, this component does not bind to H2A and

H2B genes. Hence, additional factors or a different factor may be required for expression of these genes. A "TATA" box binding activity has also been isolated from HeLa cells (Davison et al., 1983).

Other sequences conserved in the 5' regions of histone genes may also be sites for the binding of regulatory factors, as discussed for H1 histone genes in Chapter 6 (Sections 6.3.2 and 6.4.1). Several sequence-specific regulatory factors have been identified in non-histone gene systems (Parker and Topol, 1984b; Gidoni et al., 1984; Plumb et al., 1985; Emerson et al., 1985; Bram and Kornberg, 1985; Séguin et al., 1984; Sassone-Corsi et al., 1985; Jones and Tjian, 1984; Giniger et al., 1985; Yaniv, 1984). Factors required for the maximal expression of the chicken H5 gene are at present being investigated (P. Wigley, this laboratory).

The specific attachment of sequences adjacent to histone genes, to the nuclear scaffold, as observed in Drosophila (Mirkovitch et al., 1984), may also be involved in the expression of these genes (Ciejek et al., 1983; Hentzen et al., 1984).

1.6.2 RNA Processing

(a) Non-polyadenylated mRNA

Histone mRNA 3' termini are generally not polyadenylated (Hentschel and Birnstiel, 1981; Old and Woodland, 1984; Birnstiel et al., 1985). Such termini are produced by processing of larger precursor RNAs (Price and Parker, 1984; Krieg and Melton, 1984; Birchmeier et al., 1984) rather than by termination of transcription at the 3' terminus site. Termination of transcription of a sea urchin H2A gene injected into Xenopus oocyte nuclei was found to occur heterogeneously within the first 100-200 bp downstream from 3' termini (Birchmeier et al., 1984).

The presence of certain sequences in sea urchin histone mRNA precursors were found to be necessary for efficient and accurate mRNA 3' end formation (Birchmeier et al., 1982, 1983, 1984; Georgiev and Birnstiel, 1985). The first of these sequences contains a region of hyphenated dyad symmetry and is highly conserved throughout evolution. The consensus sequence for this region in DNA is 5' AACGGC$\underset{C}{T}$CTTTTCAG$\underset{G}{A}$GCCACCA 3' (Birchmeier et al., 1983; Hentschel and Birnstiel, 1981). The sequence is located approximately 40 bp downstream from termination codons. Histone mRNA 3' termini generally map at or near the 3' end of this sequence (Hentschel and Birnstiel, 1981; Maxson et al., 1983b). The second sequence requirement, 5' CAAGAAAGA 3', is located 6 nucleotides downstream from the dyad symmetry element. This sequence is not well conserved in non sea urchin species (Hentschel and Birnstiel, 1981; Birnstiel et al., 1985; Section 6.5). Lastly, an additional 50 to 80 nucleotides of further 3' sequence are also required (Birchmeier et al., 1984; Georgiev and Birnstiel, 1985).

Processing of sea urchin H3 gene precursor mRNAs requires a small nuclear RNA of approximately 60 nucleotides which appears to act as a component of a small nuclear ribonucleoprotein (RNP) (Birchmeier et al., 1984; Stunnenberg and Birnstiel, 1982; Galli et al., 1983; Strub et al., 1984). Sequence analysis of cDNA clones made to these RNAs (named U7 RNA) reveals that sequences at their 5' ends could potentially hybridize to sequences within both the conserved dyad and 5' CAAGAAAGA 3' elements of precursor histone RNAs, leaving the 3'terminal histone mRNA sequence in a single-stranded loop out (Strub et al., 1984). A histone pre-mRNA processing activity isolated from Drosophila cells also contains RNA species (Price and Parker, 1984). Small nuclear ribonucleoproteins have been implicated in a number of cellular processes including RNA splicing (Proudfoot, 1984; Birnstiel et al., 1985; Turner, 1985).

(b) Polyadenylated mRNA

The 3' termini of histone mRNAs of yeast (Fahrner et al., 1980) and Tetrahymena (Bannon et al., 1983) are polyadenylated. This is also the case for mRNAs derived from the replication-independent chicken H2A.F (Harvey et al., 1983), H5 (Molgaard et al., 1980) and H3.3 genes (Brush et al., 1985; Section 1.5.3(d)) and for human H3.3 mRNA (Wells and Kedes, 1985; Section 1.5.3(e)).

A common feature of polyadenylated mRNAs from non-histone genes is the sequence 5' AAUAAA 3' found 10 to 30 nucleotides upstream from polyadenylation sites (Nevins, 1983; Proudfoot, 1984; Birnstiel et al., 1985). This sequence is found only in the 3' flanking regions of a few yeast histone genes (Smith, 1984; Smith and Andrésson, 1983) and the chicken H3.3B gene (Brush et al., 1985). None of the above histone genes contain the 3' sequences, viz., the dyad symmetry and 5' CAAGAAAGA 3' elements, found to be conserved between histone genes producing non-polyadenylated mRNAs (Smith, 1984; Bannon et al., 1984; Harvey et al., 1983; Robins et al., 1985; Krieg et al., 1982a, 1983; Brush et al., 1985). Alternative sequences have been postulated to be involved in the 3' end formation of yeast histone genes (Smith, 1984) and the chicken H5 gene (Krieg et al., 1982a; Section 6.6). It is not known whether the 3' termini of polyadenylated histone mRNAs are produced by processing of a precursor RNA, as observed for polyadenylated non-histone mRNAs (Nevins, 1983; Proudfoot, 1984; Birnstiel et al., 1985) and non-polyadenylated histone mRNAs.

(c) Introns

As previously discussed, some histone genes, unlike most, contain introns. Genes containing introns are the H2A.F and H3.3 genes of the

chicken (Section 1.5.3(d)), presumably the human H3.3 gene (Section 1.5.3(e)) and histone genes of Neurospora (Section 1.5.1(b)). These genes contain typical RNA splice-site consensus sequences (Woudt et al., 1983; Robins et al., 1985; Brush et al., 1985). The presence of introns in some histone genes introduces another step into the processing pathway of histone RNAs at which control mechanisms could be operating (Turner, 1985). The absence of introns in most histone genes may contribute to rapid expression of these genes during S-phase.

1.6.3 Cell Cycle Regulation

(a) Regulation of replication-dependent histone synthesis

Most core and H1 histone synthesis is found to be replication-dependent, viz., it occurs primarily during the S-phase of the cell cycle (Sections 1.2.3 and 1.3.3(b); Wu et al., 1984; Maxson et al., 1983a, 1983b). Replication-dependent core and H1 histone synthesis in mammalian tissue culture cells is achieved by the regulation of steady state histone mRNA levels (Maxson et al., 1983a, 1983b; Stein and Stein, 1984; Stein et al., 1984; Plumb et al., 1984; Marzluff and Graves, 1984). Steady state mRNA levels are, in turn, controlled at the level of transcription and mRNA turnover. Increased transcription rates and lengthening of histone mRNA half-lives results in increased levels of histone mRNA during the S-phase. Upon inhibition of DNA replication the rate of histone mRNA synthesis falls to non-S-phase levels and histone mRNAs are rapidly degraded (Heintz et al., 1983; Sittman et al., 1983a; Marzluff and Graves, 1984; Stein and Stein, 1984). Transcription rate measurements in a hamster fibroblast cell cycle mutant indicated that the triggering of replication-dependent histone mRNA synthesis occurred in late G1 (Artishevsky et al., 1984). Although H1 and core histone gene expression appears to be regulated in a similar

manner (Plumb et al., 1984; Stein and Stein, 1984; Baumbach et al., 1984; Helms et al., 1984), there is some evidence that core histone mRNA accumulation precedes that of histone H1 mRNA in S-phase (Plumb et al., 1984). Replication-dependent core and H1 histone synthesis in the chicken is probably regulated by the same mechanisms as discussed above (Dalton et al., 1985; S. Dalton and J. Coleman, unpublished; Winter, 1984; Winter et al., 1985b; Section 1.3.3(b)).

S-phase regulation of yeast histone synthesis is also regulated at both transcriptional and post-transcriptional levels (Hereford et al., 1981, 1982; Osley and Hereford, 1981; Nurse, 1983). Activation of yeast histone mRNA synthesis occurs late in G1, prior to the initiation of DNA replication. Cessation of histone mRNA synthesis was found to be dependent upon the entry of cells into S-phase (Hereford et al., 1982). Hereford et al. (1982) proposed that the changes in chromatin that precede initiation of DNA replication also bring about activation of histone mRNA synthesis. Histone gene transcription would cease once these genes had been replicated and chromatin restored to its pre-replicative state. It has recently been proposed that the replication of human H1 and core histone genes may also play a role in the periodic expression of these genes (Iqbal et al., 1984).

There is evidence that protein synthesis is necessary for cell cycle regulation of histone mRNAs in mammalian cells (Stein and Stein, 1984; Baumbach et al., 1984; Sive et al., 1984; Stimac et al., 1984; Graves and Marzluff, 1984; Stahl and Gallwitz, 1977). Inhibition of protein synthesis results in increased histone mRNA levels by increasing transcription rates and by stabilizing histone mRNAs. Such a requirement for protein synthesis is consistent with models of auto-regulation of histone gene expression by histone proteins (Stein and Stein, 1984).

Factors involved in S-phase regulation of histone gene expression are yet to be isolated. It has been observed, however, that nuclear extracts prepared from S-phase nuclei of HeLa cells had a greater capacity to

transcribe a cloned human H4 gene than extracts from non-S-phase nuclei and, hence, may contain activities necessary for cell cycle regulation *in vivo* (Heintz and Roeder, 1984). The activity(ies) in the S-phase extract appeared to be specific for the H4 gene, suggesting that there may be separate factors for different genes. H4 gene 5' non-coding sequences have recently been identified that are necessary for the specific enhancement of transcription in S-phase nuclear extracts (Hanley *et al.*, 1985). The potential role of an H1 gene-conserved 5' element in S-phase regulation is discussed in Chapter 6 (Section 6.4.1).

A sequence has also been identified in yeast that is necessary for cell cycle specific histone gene expression (Osley and Hereford, 1982). The regulatory activity in this region appears to be coincident with a sequence that supports autonomous replication in yeast. This is consistent with the proposed relationship between the events associated with the initiation of DNA replication and the activation of histone mRNA synthesis discussed above. A sequence has recently been identified in the 5' flanking region of a chicken H3 gene (Wang *et al.*, 1985) which has a high degree of homology with human DNA sequence elements that have been shown to act as autonomously replicating sequences in yeast (Montiel *et al.*, 1984). It is suggested by Wang *et al.* (1985) that this sequence may be involved in the cell cycle regulated expression of the chicken H3 gene and adjacent genes.

Nothing is known of the features of replication-dependent histone mRNA that make it susceptible to destabilization at the end of S-phase. The recent work of Lüscher *et al.* (1985) suggests that sequences in the 3' terminal portions of histone genes are involved in the post-transcriptional mechanisms of cell cycle regulation. In higher eukaryotes the conserved 3' dyad symmetry element could hence be a target for such mechanisms (Section 1.6.2(a)). The lack of polyadenylation of mRNAs could also play a role.

(b) Regulation of replication-independent histone synthesis

As discussed previously the synthesis of certain vertebrate core and H1 histone variants is independent of DNA synthesis (Section 1.2.3 and 1.3.3). Like the synthesis of replication-dependent histones, the expression of the replication-independent chicken H2A.F (Robins et al., 1985),H5 (Dalton et al., 1985) and H-1c proteins (Winter, 1984; Winter et al., 1985b; Section 1.3.3(b)), appears to be primarily regulated by steady state mRNA levels. The same appears to be true for the H3.3 variant in mammals (Sittman et al., 1983a; Marzluff and Graves, 1984). DNA synthesis inhibitors have no effect on the level of replication-independent mRNAs (Sittman et al., 1983a; Marzluff and Graves, 1984). In contrast to replication-dependent genes, investigations in this laboratory have revealed that replication-independent chicken histone genes are transcribed continuously throughout the cell cycle (Dalton et al., 1985; Robins et al., 1985).

Features of replication-independent genes, such as the lack of histone gene-specific conserved 3' elements and the production of polyadenylated mRNAs (Section 1.6.2(b)), may be important to the expression of these genes. H2A.F and H3.3 genes also differ from their replication-dependent counterparts by containing introns (Section 1.6.2(c)).

The differential coupling of histone variant synthesis to DNA replication plays a role in the establishment of tissue-specific core and H1 histone variant ratios (Sections 1.2.3 and 1.3.3(b)).

1.6.4 Developmental Regulation

Qualitative changes in histone gene expression are observed during development in some organisms (Sections 1.2.3 and 1.3.3(a)). This is most dramatic in the sea urchin (Section 1.5.2(a)). In this organism the

mechanisms involved in the regulation of the switch from early to late histone gene expression have been most extensively investigated (Maxson et al., 1983a, 1983b). During the switch it is observed that there is a dramatic decline in early histone gene mRNA synthesis and a reduction in the half-life of early gene transcripts (Weinberg et al., 1983; Maxson and Wilt, 1982, 1981; Mauron et al., 1982; Knowles and Childs, 1984). By hatching blastula stage,early variant synthesis is barely detectable and the pool of early gene transcripts has fallen to less than 10% of its peak value. Early gene transcription is controlled primarily at the level of transcription initiation (Uzman and Wilt, 1984). Coincident with the decline in the early histone mRNA pool there is a sharp increase in the concentrations of late histone gene transcripts which are detected as early as the unfertilized egg stage (Maxson et al., 1983c). These transcripts represent more than 90% of the histone mRNA pool in the one-day-old mesenchyme blastula. Late histone variants constitute a rising fraction of chromatin histone from mesenchyme blastula onward (Maxson et al., 1983a, 1983b). The increase in late histone gene transcripts is in part due to an increased rate of transcription of the late genes (Knowles and Childs, 1984).

Differences in the organization of late and early sea urchin histone genes (Section 1.5.2(a)) and the presence of late gene-specific 5' sequence elements (Section 1.6.1(a); Roberts et al., 1984) may be relevant to the expression of the two gene sets.

1.7 CHICKEN H1 GENES

At the commencement of this project there was limited sequence data of vertebrate H1 proteins and limited information about vertebrate H1 genes. Vertebrate H1 genes had only been isolated from Xenopus laevis and there was no DNA sequence data available (Zernik et al., 1980).

Work described in this thesis involves the isolation and analysis of a family of genes coding for H1 variants in the chicken. The chicken was chosen as work in this laboratory was well under way in characterizing the core histone genes of this organism. Efforts were also being directed towards the isolation of the gene for the H1-related, tissue-specifically expressed chicken H5 protein. It was of interest to determine the location of H1 genes, relative to other histone genes, and to identify by DNA sequencing potential regulatory sequences involved in the general, differential and cell cycle-specific expression of these genes. It was also of interest to compare the H1 genes to the potentially related H5 gene. Generated DNA sequences then provide a starting point for analysis of H1 gene expression, as described in later chapters.

Sequence determination of H1 genes also generates valuable protein sequence data as a step towards the elucidation of H1 protein subtype functions.

CHAPTER 2

MATERIALS AND METHODS

2.1 ABBREVIATIONS

BCIG: 5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside

bisacrylamide: N,N'-methylene-bis-acrylamide

ddNTP: dideoxynucleoside triphosphate

DN'ase: deoxyribonuclease

DTT: dithiothreitol

IPTG: isopropyl-B-D-thio-galactopyranoside

PEG: polyethylene glycol

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

SDS: sodium dodecyl sulphate

Additional abbreviations were as described elsewhere (in "Instructions to Authors" (1978) Biochem. J. 169, 1-27).

2.2 MATERIALS

2.2.1 General Reagents and Materials

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

Acrylamide: Sigma Chemical Co.

Agarose (low gelling temperature): B.R.L. Inc.

Ampicillin: Sigma Chemical Co.

BCIG: Sigma Chemical Co.

Bisacrylamide: Sigma Chemical Co.

Caesium Chloride (optical grade): Harshaw Chemical Co.

Chloramphenicol: gift from Parke-Davis.

ddNTPs: P.L. Biochemicals Inc.

dNTPs and NTPs: Sigma Chemical Co.

Glyoxal: B.D.H. Ltd.

IPTG: Sigma Chemical Co.

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

Nitrocellulose filters: Sartorius; Schleicher and Schuell Inc.

PEG(6000): B.D.H. Ltd.

Sephadex G-50 (Medium): Pharmacia.

Tetracycline: UPJOHN Pty. Ltd.

"Trizma" base: Sigma Chemical Co.

2.2.2 Enzymes

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

AMV reverse transcriptase: gift from Dr. J. Beard, Florida, U.S.A.

Calf intestinal phosphatase: Sigma Chemical Co.

E. coli DN'ase I: Sigma Chemical Co.

E. coli DNA polymerase I: Boehringer-Mannheim; BRESA, Adelaide, Australia.

E. coli DNA polymerase I, Klenow fragment: Boehringer-Mannheim; BRESA, Adelaide, Australia.

E. coli poly(A)polymerase: gift from Dr. R. H. Symons.

Exonuclease BAL-31: New England Biolabs. Inc.

Lysozyme: Sigma Chemical Co.

Pancreatic Ribonuclease: Sigma Chemical Co.

Restriction endonucleases: New England Biolabs. Inc.; Boehringer-Mannheim.

S1 nuclease: Boehringer-Mannheim.

T4 DNA ligase: Boehringer-Mannheim; BRESA, Adelaide, Australia.

Terminal deoxynucleotidyl transferase: Ratcliff Biochemicals, New Mexico.

T4 polynucleotide kinase: Boehringer-Mannheim.

2.2.3 Isotopically Labelled Compounds

α-^{32}P-dNTPs, γ-^{32}P-ATP:BRESA.

^{3}H-ATP: gift from J. Hasseloff.

^{3}H-dCTP: New England Nuclear.

2.2.4 Nucleic Acids

Bacteriophage λDNA: prepared from a concentrated 'phage stock; gift from Dr. J. B. Egan.

M13mp83, M13mp93, M13mp18, M13mp19, pUC18: gift from Dr. A. Robins.

M13 17-base sequencing primer: BRESA, Adelaide, Australia.

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

pBR322 DNA: gift from Dr. P. Krieg.

2.2.5 Bacterial Strains and Media

(a) Strains

E. coli MC1061: *ara*D139, Δ(*ara*,*leu*) 7697, Δ*lac*X74, *galU*$^{-}$, *galK*$^{-}$, *hsr*$^{-}$, *hsm*$^{+}$, *str*A; gift from Dr. R. P. Harvey.

E. coli JM101: Δ(*lac*-*pro*), F' *lacI*^{q} ZΔ M15, *traD1*; gift from Dr. A. Robins.

E. coli LE392: *F*$^{-}$ *hsdR514* ($r_k^- m_k^-$) *supE44* *SupF58* *lacY1* or Δ*lac*(*IZY*)*6* *galK2* *galT22* *metB1* *trpR55* λ-; gift from Dr. J. B. Egan.

(b) Media

L-broth (Luria broth): 1% (w/v) bacto-tryptone (Difco), 0.5% (w/v) yeast extract (Difco), 1% (w/v) NaCl, pH 7.0. Where appropriate, L-broth was supplemented with ampicillin (50 ug/ml) or tetracycline (20 ug/ml).

L-agar plates: L-broth containing 1.5% (w/v) bacto-agar (Difco).

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

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

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

CH4A-agar plates: 1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 0.2% (w/v) glucose, 10mM Tris-HCl, pH 7.5, 1mM $MgCl_2$, 1.5% (w/v) bacto-agar.

All media were sterilised by autoclaving, except heat labile reagents, which were filter sterilised. Media were prepared with mono-distilled water.

2.3 METHODS

2.3.1 Restriction Enzyme Digestion and Analysis of DNA

(a) Restriction enzyme digestion

Restriction endonuclease digestion of DNA was performed using the conditions of each enzyme described by Davis et al. (1980).

Reactions were terminated by adding EDTA to a final concentration of 5 mM. The reaction mix was then extracted with an equal volume of phenol/chloroform (v/v;1/1). The aqueous phase was adjusted to 0.2 M NaCl and the DNA precipitated by the addition of 2.5 volumes of ethanol

(-80°C, 5 min). The precipitate was collected by centrifugation (13,000 g, 5 min), washed by centrifugation in 70% ethanol and dried in vacuo. Alternatively digestion was terminated by the addition of a half-volume of 3 x urea load (4 M urea, 50mM EDTA, 0.5% (w/v) bromophenol blue, 50% (w/v) sucrose).

(b) Agarose gel electrophoresis

Analytical

Agarose (0.7%-2%) was dissolved in 1 x TEA buffer (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 onto 7.5 cm x 5 cm microscope slides, for horizontal gels.

Vertical gels were electrophoresed between tanks containing 1 x TEA at 65 mA for approximately three hours. Horizontal gels were run submerged in 1 x TEA at 150 mA for approximately 15 minutes.

DNA samples were loaded in 1 x urea load. DNA in gels was visualized by staining with 10 ug/ml ethidium bromide for 5 minutes followed by destaining with water and examination under ultraviolet (UV) light.

Preparative

Low gelling temperature (LGT) agarose was dissolved in 1 x TEA and cast either into vertical templates or on to horizontal slides, as described above. Electrophoresis was carried out at 4°C. DNA was loaded in 1 x urea load.

DNA was detected by brief ethidium bromide staining and the desired bands excised from the gel with a scalpel. Two volumes of 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA were added to the slice, and the agarose melted at 65°C for 15 minutes. An equal volume of buffer-saturated phenol at 37°C was added, the phases rapidly mixed and then immediately separated by

centrifugation. The aqueous phase was re-extracted with phenol, adjusted to 0.2 M NaCl and the DNA ethanol precipitated. Recovery of DNA from gel slices was approximately 60%.

(c) Polyacrylamide gel electrophoresis

Analytical

Electrophoresis of DNA fragments less than 1 kilobase in length was carried out in vertical 14 cm x 14 cm x 0.5 mm gels containing, usually, 6% acrylamide/bisacrylamide (30/1; w/w) polymerized in 1 x TBE buffer (90 mM Tris-borate, 2.5 mM EDTA, pH 8.3). Electrophoresis was performed at 250 volts for approximately 90 minutes. DNA samples were loaded onto gels in either 1 x urea load or 1 x acrylamide load (1 mM EDTA, 10% (w/v) sucrose, 1 mM EDTA, 0.02% (w/v) bromophenol blue, 0.02% (w/v) xylene cyanol). DNA was visualized under UV light following ethidium bromide staining.

Preparative

DNA fragments that had been fractionated preparatively were excised from the gel and the DNA eluted into two changes of 100-200 ul of 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. If ^{32}P-labelled DNA fragments (Section 2.3.7) were to be eluted from gels, such fragments were first located by exposure to X-ray film at room temperature for a few minutes. Eluted fragments were not ethanol precipitated. Recovery of DNA from gel slices was between 50% to 90% depending on the size of the DNA fragments.

(d) Transfer of DNA to nitrocellulose and hybridization with a labelled probe

Restricted DNA fractionated on vertical agarose gels was transferred

to nitrocellulose filters 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 two hours.

The prehybridization and hybridization of nitrocellulose filters was essentially as described by Wahl et al. (1979), except that both dextran sulphate and glycine were omitted from the hybridization mix. On completion of hybridization, filters were washed twice with 2 x SSC (0.3 M NaCl, 0.03 M Na_3 citrate), 0.1% (w/v) SDS for 5 minutes at room temperature and then twice with between 0.1 to 1 x SSC, 0.1% (w/v) SDS at 65°C for 30 minutes.

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

For re-use, filters were boiled twice for 5 minutes in distilled water to remove hybridized probe and then prehybridized, hybridized etc. as described above.

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 added and the sample spotted on to nitrocellulose filter paper damp with 20 x SSC. The filters were then processed as described above.

2.3.2 Construction of a cDNA Library

(a) Size fractionation of RNA

RNA was fractionated on 10-40% linear sucrose gradients. Gradients

were constructed in a "gradient maker" using 5.5 ml each of 10% and 40% (w/v) solutions of sucrose in 10 mM Tris-HCl pH 7.5, 1 mM EDTA. RNA in 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.5% SDS was heated at 65°C for 5 minutes and snap-chilled before being layered on to gradients. Gradients were centrifuged at 210,000 g for 16 hours at 4°C and fractionated by upward displacement with a 50% (w/v) sucrose solution using an ISCO Density Gradient Fractionator. RNA was identified by its absorbance at 254 nm and the desired fractions collected, adjusted to 200 mM NaCl and ethanol precipitated.

(b) Poly(A) addition to 3' termini of RNA (Gould et al., 1978; J. Hasseloff, personal communication)

10 ug of size-fractionated RNA was resuspended in 50 ul of 0.1 mM EDTA, heated at 80°C for 1 minute and snap-chilled on ice. 50 ul of 0.2 mM ATP (containing 1 uCi ^{3}H-ATP), 0.1 M Tris-HCl pH 7.9, 5 mM $MnCl_2$, 20 mM $MgCl_2$, 2 mM DTT and 13 ug of purified E. coli poly(A) polymerase (Gould et al., 1978) were added to the heat-treated RNA and incubated at 37°C for 40 minutes. The number of nucleotides added to RNA was determined by assaying the conversion of ^{3}H-ATP into a trichloroacetic acid insoluble form. RNA was phenol/chloroform extracted and ethanol precipitated from the reaction mix.

(c) Synthesis of double-stranded cDNA (Chang et al., 1978)

For first-strand cDNA synthesis approximately 4 ug of polyadenylated RNA was incubated for 30 minutes at 42°C in 20 ul of 50 mM Tris-HCl pH 8.3, 10 mM $MgCl_2$, 20 mM B-mercaptoethanol, 1 mM each of the deoxynucleotides dATP, dTTP and dCTP, 0.2 mM dGTP, approximately 10 uM α-^{32}P-dGTP (500 Ci/mmole), 400 ng of the oligonucleotide oligo-$dT_{(10)}$, and 20 units of reverse transcriptase. The first-strand mix was then boiled for 2 minutes to degrade the RNA template. 1 ul of the mix was passaged through a

Sephadex G-50 column to determine the percent incorporation of labelled nucleotide into cDNA.

For second-strand synthesis the reaction volume was increased to 50 ul and conditions adjusted to those used for synthesis of the first-strand, except that unlabelled dGTP was adjusted to 1mM, no labelled dGTP was added, and 25 units of reverse transcriptase were added. After incubation for 6 hours at 37°C, the reaction was phenol/chloroform extracted and passaged through a Sephadex G-50 column to separate cDNA from unincorporated nucleotides. Collected cDNA was then ethanol precipitated and digested with 1000 units of S1 nuclease in 30 mM Na-acetate pH 4.5, 0.03 mM $ZnSO_4$ for 15 minutes at 37°C. The reaction was then adjusted to 5 mM EDTA, phenol/chloroform extracted and loaded on to a 10-40% linear sucrose gradient (180,000 g, 16 hours, 4°C) for size-fractionation of the double-stranded (ds) cDNA. Required fractions were ethanol precipitated.

(d) dC-tailing of double-stranded cDNA (Chang et al., 1978; Roychoudhury et al., 1976)

For addition of dC residues to 3' ends in ds cDNA, approximately 30 ng of ds cDNA was incubated in 25 ul of 0.14 M Na-cacodylate/30 mM Tris-HCl (pH 7.6), 1 mM $CoCl_2$, 500 pmoles ^{3}H-dCTP (25 Ci/mmole) and 3 units of terminal deoxynucleotidyl transferase for approximately 1 hour at 37°C. The reaction was assayed by determining the conversion of ^{3}H-dCTP into a trichloroacetic acid insoluble form. The reaction was stopped by the addition of an equal volume of 20 mM EDTA.

(e) Preparation of vector and annealing to ds cDNA (Chang et al., 1978; Roychoudhury et al., 1976)

5 ug of Pst I cut pBR322 DNA was tailed with dG residues in a 50 ul reaction containing 0.14 M Na-cacodylate/30 mM Tris-HCl (pH7.6), 1 mM $CoCl_2$, 5 uCi of α-^{32}P-dGTP (500 Ci/mmole), 1 nmole unlabelled dGTP, and 8 units

of terminal deoxynucleotidyl transferase. The reaction was incubated, assayed and terminated as described above (Section 2.3.2(d)).

100 ul of annealing buffer (0.2 M NaCl, 10 mM Tris-HCl, pH 8.2), containing equimolar amounts of dG-tailed pBR322 and dC-tailed ds cDNA, was heated for 2 minutes at 65°C, followed by one hour at 45°C and then allowed to cool slowly overnight at 4°C.

Annealed DNA was used to transform E. coli (Section 2.3.2(f)) and colonies were screened for plasmids containing appropriate inserts (Section 2.3.3).

(f) Transformation of E. coli

E. coli MC1061 was grown overnight at 37°C in L-broth (Section 2.2.5(b)) 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 min, 4°C) and washed in a 1/2 volume of ice-cold 0.1 M $MgCl_2$. The cells were resuspended in 1/20 of the original volume of ice-cold 0.1 M $CaCl_2$ and stored at 4°C for between 1 and 24 hours.

0.2 ml of these competent cells were added to 0.1 ml of DNA (typically 5-50 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 overlay-agar (at 42°C) and plated on to L-agar plates containing appropriate antibiotic (Section 2.2.5(b)). Plates were incubated overnight at 37°C.

2.3.3 Detection and Examination of Recombinant Plasmid Clones

(a) Colony screening (Grunstein and Hogness, 1975)

Colonies for screening were transferred by toothpick to a master plate and to one or more sheets of nitrocellulose that had been boiled three times in distilled water and lain onto L-agar plates containing an appropriate antibiotic. The colonies were grown overnight on the nitrocellulose at 37°C, and then lysed by transferring the nitrocellulose sequentially on to Whatman 3MM paper saturated with 0.5 M NaOH for 7 minutes, 1 M Tris-HCl pH 7.4 for two minutes, 1 M Tris-HCl pH 7.4 for two minutes and 1.5 M NaCl, 0.5 M Tris-HCl pH 7.4 for 4 minutes. Nitrocellulose filters were baked at 80°C, under vacuum, for two hours. Hybridization and washing conditions were as described previously (Section 2.3.1(d)).

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

1.5 ml cultures of recombinants 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% (w/v) 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 was 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 was added and the solution incubated on ice for a further 15 minutes.

Insoluble material was removed by centrifugation (10 min, 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 ribonuclease was included in the restriction reaction.

2.3.4 Subcloning into Plasmid Vectors

(a) Blunt-ending

Restriction enzyme generated 5' or 3' overhangs were repaired to blunt-ends at 37°C for 10-30 minutes in a 10 ul reaction containing 10 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT, 75 uM of each of dATP, dTTP, dCTP and dGTP and 1.0 unit of E. coli DNA polymerase I, Klenow fragment. The reaction was terminated by phenol/chloroform extraction and the DNA recovered by ethanol precipitation.

(b) Vector preparation and ligation

To remove the 5' terminal phosphate groups from restricted vector DNA, such DNA (2-5 ug) was incubated with approximately 0.2 units of Calf Intestinal Phosphatase (CIP, previously dialyzed 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% (w/v) SDS for 2 hours. The reaction was made 5 mM with EDTA, phenol/chloroform extracted three times and ethanol precipitated. Vector DNA was purified from uncut vector by passaging through an LGT-agarose gel (Section 2.3.1(b)).

Restriction fragments to be subcloned were preparatively isolated from LGT-agarose or polyacrylamide gels.

Vector and insert DNA were ligated in a 10 ul reaction mixture

containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 0.5 mM ATP and 1 unit of T4 DNA ligase, at 4°C for 6-24 hours. Reactions usually contained about 50 ng of vector and enough insert DNA to give a 1:1 molar ratio of vector:insert.

Ligated DNA was transformed into E. coli MC1061 (Section 2.3.2(f)) and transformants analyzed (Section 2.3.3).

2.3.5 Large-Scale Preparation of Plasmid DNA

Cells containing plasmid DNA were grown up in 500 ml L-broth cultures 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. Cells were harvested by centrifugation (10,000 g, 5 min, 4°C).

Plasmid DNA was liberated from the cells as described above for the miniscreen method (Section 2.3.3(b)), except that the volumes were increased 40-fold, and plasmid DNA was treated with 20 ug/ml DN'ase-free, pancreatic ribonuclease 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 min, Eppendorf centrifuge, 4°C), washed with 70% ethanol and dissolved in water. Plasmid DNA was stored at -20°C.

2.3.6 Isolation of Clones from a λ-Recombinant Genomic Library

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

0.2 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 overlay-agar, containing 10 mM $MgCl_2$, at 42°C, were added and the mixture poured on to fresh, dry 15 cm CH4A-agar plates (Section 2.2.5(b)). Plates were incubated, inverted, at 37°C overnight then stored at 4°C to harden the agar.

An unwashed, 14 cm nitrocellulose filter was lain on to the plate, orientation marks made with a needle and, when uniformly wet, peeled off and placed on to Whatman 3MM paper saturated with 0.5 M NaOH, 1.5 M NaCl for one minute and then sequentially on to two sheets of 3MM paper saturated with 0.5 M Tris-HCl pH 7.4, 1.5 M NaCl for one minute 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 under vacuum for one hour and then probed with required sequences and autoradiographed as previously described (Section 2.3.1(d)).

Dot-Benton procedure

3 ml of overlay-agar, containing 200 ul of LE392 grown in L-broth to an A_{600} of approximately 1.5, was plated on to dry 9 cm CH4A-agar plates. Purified 'phage stocks were dotted onto the freshly plated LE392 lawn with a glass capillary after agar had set. After incubation overnight at 37°C, to allow plaque formation, nitrocellulose filters were made from plates as described above. Filters were then hybridized with probe, washed and exposed to X-ray film (Section 2.3.1(d)).

(b) Preparation of 'phage DNA

CsCl method

10^5 pfu (plaque forming units)/15 cm plate were adsorbed to 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 min, 4°C) and the 'phage precipitated at 4°C for two hours by adjusting the solution to 875 mM NaCl, 6% PEG. The 'phage were collected by centrifugation (10,000 g, 10 min, 4°C) and resuspended in 4 ml of PSB.

'Phage were then layered on to discontinuous CsCl gradients containing 2 ml blocks of CsCl in PSB, with densities of p = 1.40 and p = 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, diluted 1 in 10 with PSB and treated with 20 ug/ml DN'ase-free, pancreatic ribonuclease at 37°C for 30 minutes. The solution was made to 5 mM EDTA, 0.2 M NaCl and the DNA isolated from 'phage by phenol/chloroform extraction followed by ethanol precipitation.

Mini-preparation

'Phage were eluted into 3 ml PSB from 9 cm plates containing 10^5 pfu as described above. Debris was removed from the 'phage solution by two rounds of centrifugation (10,000 g, 5 min, 4°C). 'Phage were then pelleted by centrifugation at 45,000 rpm for 1 hour at 5°C in a Ti50 rotor. The 'phage pellet was dissolved in 400 ul PSB, treated with ribonuclease and DNA isolated by phenol/chloroform extraction and ethanol precipitation as above. The 'phage pellet dissolved in PSB was also used as a concentrated 'phage stock.

2.3.7 Preparation of In Vitro Labelled DNA

(a) Nick-translation (Maniatis et al., 1975)

100 ng of DNA was labelled in a 20 ul reaction mix containing 50 mM Tris-HCl pH 7.8, 5 mM $MgCl_2$, 10 mM B-mercaptoethanol, 50 ug/ml bovine serum albumin, 25 uCi each of α-^{32}P-dCTP and α-^{32}P-dATP (1700 Ci/mmole), 25 uM each of unlabelled dGTP and dTTP, 20 pg of E. coli DN'ase I and 5 units of E. coli DNA polymerase I. The solution was incubated at 15°C for 90 minutes, phenol/chloroform extracted and the un-incorporated nucleotides removed by chromatography on a Sephadex G-50 column. For use as a hybridization probe, the DNA strands were separated by boiling for 2 minutes followed by snap-chilling on ice.

(b) End-fill labelling

DNA fragments with 5' overhangs were incubated in a 10 ul reaction mix with E. coli DNA polymerase I, Klenow fragment, as previously described (Section 2.3.4(a)) except that certain unlabelled dNTPs were replaced with 25 uCi of labelled α-^{32}P-dNTPs (1700 Ci/mmole) depending on the nature of the overhang. The reaction was terminated by phenol/chloroform extraction and DNA recovered by ethanol precipitation.

(c) M13 probes

2.5 ng of sequencing primer (17-mer) were annealed to 500 ng of recombinant M13 single-stranded DNA in a 10 ul reaction as described elsewhere (Section 2.3.10(a)). The primer hybridizes to the 3' side of the M13 insert DNA. The hybridization mix was then added to an additional 10 ul containing 1 mM DTT, 0.15 mM each of dTTP and dGTP, 25 uCi each of

α-^{32}P-dATP and α-^{32}P-dCTP (1700 Ci/mmole) and 1 unit of E. coli DNA polymerase I, Klenow fragment. Incubation at 37°C for 15 minutes results in the synthesis of a complementary labelled DNA strand. At the end of the incubation 4 ul of chase (0.25 mM dATP, 0.25 mM dCTP) were added and the reaction incubated for a further 15 minutes at 37°C. The labelled DNA was then digested with appropriate restriction enzymes for release of double-stranded insert DNA which could then be purified from a 6% polyacrylamide gel (Section 2.3.1(c)). For use as a hybridization probe the DNA was boiled before use as previously described (Section 2.3.7(a)).

(d) 5' end-labelling with polynucleotide kinase

Prior to labelling, DNA was dephosphorylated with calf intestinal phosphatase (Section 2.3.4(b)). After dephosphorylation, the enzyme was inactivated by heating at 65°C for 5 minutes, and then phenol/chloroform extracted. DNA was recovered by ethanol precipitation.

DNA was end-labelled in a 10 ul volume containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, approximately 50 pmole γ-^{32}P-ATP (2000 Ci/mmole) and 2 units of T4 polynucleotide kinase at 37°C for 30 minutes. The reaction was then phenol/chloroform extracted and ethanol precipitated.

2.3.8 Subcloning into M13 Vectors

(a) Preparation of double-stranded M13 replicative form (RF) (Winter, 1980)

A plaque from a plate stock of M13 'phage (Section 2.3.8(c)) was toothpicked into 1 ml of 2 x YT broth (Section 2.2.5(b)) and grown with shaking for 6 hours at 37°C. Meanwhile, a 10 ml culture of the E. coli

strain JM101 (from a single colony on a minimal plus glucose plate; Section 2.2.5(b)) 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. DNA was prepared as previously described (Section 2.3.5).

(b) BAL-31 treatment of DNA

DNA for cloning into M13 vectors was frequently treated with the nuclease BAL-31 which has an exonuclease activity which simultaneously degrades both the 3' and 5' termini of duplex DNA. Usually 1 ug of restricted DNA was digested in a 25 ul reaction containing 600 mM NaCl, 20 mM Tris-HCl pH 8.0, 12 mM $CaCl_2$, 12 mM $MgCl_2$ and 1 mM EDTA. The reaction was started by the addition of 1 unit of BAL-31 and the mix incubated at 30°C. After varying lengths of time the reaction was phenol/chloroform extracted, the DNA ethanol precipitated and then incubated at 37°C for 10 minutes in a 10 ul blunt-ending reaction (Section 2.3.4(a)). BAL-31 treated DNA was fractionated on an LGT agarose gel (Section 2.3.1(b)).

(c) Ligation and transformation

M13 vectors (mp83, mp93, mp18 and mp19) were prepared and ligations performed as previously described (Section 2.3.4(b)). In ligations 20 ng of vector were used and the molar ratio of vector:insert was 1:3.

Competent cells were prepared by growing JM101 in 100 ml of 2 x YT broth to an A_{600} of 0.3, harvesting by centrifugation (500 g, 5 min, 4°C) and resuspending in 5 ml of freshly prepared 50 mM $CaCl_2$. 2 ul of the 10 ul ligation mix were 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 overlay-agar containing 20 ul BCIG (8 mg/ml), 20 ul IPTG

(8 mg/ml) and 0.2 ml of exponential JM101 (A_{600} approx. 0.5). The mixture was plated onto minimal plus glucose plates and grown for 9-12 hours at 37°C. Plaques containing either wild-type vector or recombinant molecules were respectively blue and white in colour.

Plate stocks of individual M13 clones or vectors were generated via transformation of purified double or single-stranded DNA as described above.

(d) Preparation of single-stranded M13 recombinant DNA

Recombinant plaques were toothpicked into 1 ml of 2 x YT broth containing 25 ul of an overnight culture of JM101 (grown in minimal medium; Section 2.2.5(b)) and grown with shaking for 5 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 ul of 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA and stored at -20°C.

'Phage supernatants, generated after removal of cells from 1 ml cultures, can be kept at 4°C as 'phage stocks. 20 ul of these stocks were often used to re-innoculate 1 ml cultures for single-stranded DNA preparation.

2.3.9 Screening M13 Recombinants

(a) Complementarity testing

To determine which strand of a particular sub-cloned DNA fragment was

present in a single-stranded M13 recombinant, hybridization analysis was carried out using an arbitrarily selected or previously sequenced, single-stranded M13 recombinant, as a reference.

1 ul of the clone to be tested was added to 1 ul of reference recombinant and incubated with 2 ul of 100 mM Tris-HCl pH 7.4, 100 mM $MgCl_2$, 500 mM NaCl, at 65°C for 1 hour.

2 ul of 3 x urea load buffer (Section 2.3.1(a)) were added and the sample was electrophoresed on a horizontal, 1% agarose gel (Section 2.3.1(b)), next to 2 ul of reference clone in urea load. The DNA was visualized in UV light after ethidium bromide staining. Single-stranded M13 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.

(b) Hybridization analysis

Recombinant plaques were toothpicked into 1 ml cultures, grown for 5 hours at 37°C and cells pelleted by centrifugation as previously described (Section 2.3.8(d)). 2 ul of 'phage-containing supernatants were dotted on to dry nitrocellulose filters. Nitrocellulose filters were then placed on to Whatman 3MM paper, saturated in 0.5 M NaOH, for 5 minutes and then on to 3MM paper, saturated in 0.5 M Tris-HCl pH 7.4, 1.5 M NaCl, for 5 minutes.

Alternatively 1 ul of 20 x SSC (Section 2.3.1(d)) was added to 1 ul of an M13-recombinant single-stranded DNA preparation (Section 2.3.8(d)) and dotted on to dry nitrocellulose.

Filters were air dried, baked and hybridized with appropriate probes (Section 2.3.1(d)).

2.3.10 **Di-Deoxy Sequence Analysis of M13 Clones** (Sanger et al., 1977)

(a) Hybridization

2.5 ng of universal primer (17-mer) were annealed to approximately 500 ng of M13 single-stranded template (5 ul of a 25 ul single-stranded DNA preparation; Section 2.3.8(d)), in a 10 ul volume containing 10 mM Tris-HCl pH 8.5, 10 mM $MgCl_2$, by incubating at 75°C for 3-5 minutes, 37°C for 5 minutes and then 25°C for 5-10 minutes.

(b) Polymerization

The hybridization mixture was added to 10 uCi of dried down α-^{32}P-dATP (1700 Ci/mmole), vortexed to resuspend the lable and then 1 ul of 10 mM DTT added. In addition, 1.5 ul each of the appropriate zero mix (T° for ddTTP: 10 uM dTTP, 200 uM dCTP, 200 uM dGTP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA; C°for ddCTP : 200 uM dTTP, 10 uM dCTP, 200 uM dGTP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA; A° for ddATP : 200 uM dTTP, 200 uM dCTP, 200 uM dGTP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA; G° for ddGTP : 200 uM dTTP, 200 uM dCTP, 10 uM dGTP, 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) and ddNTP solutions (usually 1.0 mM ddTTP, 0.2 mM ddCTP, 0.7 mM ddATP, 0.4 mM ddGTP) were added together. ddNTP concentrations were varied according to the nucleotide content of the DNA being sequenced. 2 ul of each of the four zero mix-ddNTP solutions was added to a separate Eppendorf tube, generating four tubes each of which contains a different zero mix-ddNTP solution (reaction tubes).

0.5 unit of the E. coli DNA polymerase I Klenow fragment was added to the hybridization mixture-lable-DTT solution, described above, and gently mixed. 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 incubation

for 10 minutes at 37°C, 1 ul of dATP chase (0.5 mM dATP, containing 1 unit of the Klenow fragment per 50 ul, added immediately before use) 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, containing 0.1% (w/v) bromo-cresol purple, 0.1% (w/v) xylene cyanol FF, 20 mM EDTA) was added to stop the reactions and mixed by a short centrifugation. Samples were boiled for 3 minutes and then analyzed on sequencing gels.

(c) Sequencing gels

1 ul of each sample was electrophoresed, usually, on a 6% polyacrylamide, 7 M urea gel (20 cm x 40 cm x 0.2 mm) in 1 x TBE buffer (Section 2.3.1(c)). The gels were pre-electrophoresed for 40 minutes prior to loading samples. Electrophoresis was at 1200 volts. In cases where secondary structure caused sequencing "compressions" samples were run on 8% acrylamide, 7 M urea gels containing between 25 and 40% formamide. After electrophoresis gels were fixed for 10 minutes in 10% acetic acid, 20% ethanol, washed with several litres of 20% ethanol, to remove urea, and dried. Autoradiography was carried out overnight at room temperature.

2.3.11 **Analysis of RNA**

(a) Electrophoresis and transfer to nitrocellulose (McMaster and Carmichael, 1977; Thomas, 1980)

Dried down RNA (up to 20 ug) was incubated at 80°C for 3 minutes in 10 ul deionized formamide and then snap-chilled on ice. 8 ul of 20 mM NaH_2PO_4, pH 6.5 and 3 ul of 6 M glyoxal (deionized with mixed bed resin) were

added to the RNA solution and heated at 50°C for 15 minutes. 4 ul of loading buffer (50% (w/v) glycerol, 0.01 M NaH_2PO_4 pH 6.5, 0.5% (w/v) bromophenol blue) were then added and the sample run on a 1.5% vertical agarose slab-gel (Section 2.3.1(b)) in 0.01 M NaH_2PO_4 pH 6.5. Tank buffer was re-circulated to maintain the pH at less than 8.0 (glyoxal dissociates from RNA at pH >8.0). Gels were run at 60 mA.

RNA was transferred to nitrocellulose filters pre-soaked in 20 x SSC as described by Thomas (1980). Transfer is complete in 15 to 20 hours. After transfer, filters were air-dried and baked under vacuum at 80°C for 2 hours and then immersed in 20 mM Tris-HCl pH 8.0 at 100°C and allowed to cool to room temperature (to remove residual glyoxal). Prehybridization, hybridization and washing of filters was as previously described (Section 2.3.1(d)). For re-use of filters, probe was removed by washing filters in 5 mM Tris-HCl pH 8.0, 0.2 mM EDTA, 0.05% (w/v) pyrophosphate, 0.1 x Denhardts solution (0.02% (w/v) each of ficoll, polyvinylpyrollidone and bovine serum albumin) at 65°C for 1 to 2 hours.

(b) S1 nuclease analysis of RNA

An appropriately end-labelled DNA restriction fragment was hybridized to an excess of a single-stranded M13 recombinant (containing sequences complementary to the mRNA being analyzed) in 50 ul of 100 mM NaCl, 20 mM Tris-HCl pH 7.4, 0.1% (w/v) SDS. The solution was heated at 100°C for 3 minutes and then at 65°C for 30-60 minutes. The required unhybridized labelled DNA strand was isolated from a 6% polyacrylamide gel (Section 2.3.1(c)).

The isolated single-stranded DNA was then hybridized to 5-10 ug of fractionated RNA in a 30 ul reaction containing 80% deionized formamide, 40 mM PIPES, pH 6.4, 1 mM EDTA and 400 mM NaCl, heated at 90°C for 5 minutes and then 12-16 hours at 30°C. Annealing reactions were diluted with

300 ul of ice-cold S1 buffer (200 mM NaCl, 2 mM $ZnSO_4$, 50 mM Na-acetate, pH 4.6) and 2,500 units of S1 nuclease added. Digestions were at 37°C for 1 hour. The mixture was then phenol/chloroform extracted, nucleic acids precipitated with ethanol, and dissolved in formamide loading buffer and electrophoresed on 6% sequencing gels (Section 2.3.10(c)).

2.3.12 **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

ISOLATION AND CHARACTERIZATION OF HISTONE cDNA SEQUENCES

3.1 INTRODUCTION

In order to investigate chicken H1 genes a suitable probe was required. H1 genes had been isolated from a few species at the commencement of this project, viz., from the sea urchin (Birnstiel et al., 1978; Sures et al., 1978), Drosophila (Goldberg, 1979) and Xenopus laevis (Zernik et al., 1980). Such genes could have potentially been used as probes but due to

(i) the divergence of H1 protein sequences between species (Isenberg, 1979; Von Holt et al., 1979); and

(ii) the previous observation, in this laboratory, that the use of cross-species probes to investigate histone genes can give ambiguous results (Harvey et al., 1981),

it was decided that an homologous probe would be prepared.

RNA from 5 day old chicken embryos had been shown to be a good source of histone mRNA (Crawford et al., 1979). It was therefore decided to construct a library of cDNA recombinants from this RNA, from which H1 sequences could be isolated for use as a probe to investigate genomal chicken H1 genes.

The construction of a chicken embryo cDNA library and the isolation from this library of recombinants containing both H1 and core histone sequences is discussed below.

3.2 RESULTS

3.2.1 Preparation, Fractionation and Polyadenylation of Embryo RNA

Total RNA from 5 day old chicken embryos was prepared using the guanidinium thiocyanate method of Chirgwin et al. (1979) and the 5-13S fraction collected after centrifugation on 10-40% linear sucrose gradients

(Section 2.3.2(a); RNA was prepared and fractionated by S. Bruschi). This fraction of RNA had previously been shown to be enriched for mRNAs coding for the five histone proteins (H1, H2A, H2B, H3 and H4) by translation in a wheat germ cell-free system (Crawford et al., 1979). Histone mRNAs are generally about 9S in size (Kedes, 1979, 1976). The 5-13S RNA was further purified by a second round of sucrose gradient centrifugation (Section 2.3.2(a)).

The method of cDNA synthesis used in the construction of cDNA recombinants (Section 2.3.2(c); and below) required that RNA templates be polyadenylated at their 3' termini. Crawford et al. (1979) found that the bulk of translatable histone mRNAs, including H1 mRNAs, in chicken embryo RNA were non-polyadenylated. Lack of polyadenylation is a common feature of histone mRNAs from many species (Old and Woodland, 1984; Section 1.6.2(a)). 20 ug of purified 5-13S chicken embryo RNA was polyadenylated with E. coli poly(A) polymerase to approximately 11 dA residues per 3' end in two 100 ul reactions as described in Section 2.3.2(b).

3.2.2 Construction of cDNA Recombinants

A library of cDNA recombinants was prepared from the polyadenylated chicken embryo RNA using basically the same protocol as described by Chang et al. (1978), except that reverse transcriptase was used instead of E. coli DNA polymerase I for second-strand cDNA synthesis.

(a) Synthesis and dC-tailing of double-stranded (ds) cDNA

12 ug of polyadenylated 5-13S chicken embryo RNA was used as template for first-strand cDNA synthesis by reverse transcriptase, with oligo-$dT_{(10)}$ as a primer (Section 2.3.2(c)). A total of approximately 430 ng of cDNA was synthesized (3.5% copy efficiency). First-strand reaction mixes were

boiled to degrade the RNA template and then reaction volumes were increased for second-strand cDNA synthesis by reverse transcriptase (Section 2.3.2(c)). The 3' ends of the single-stranded cDNAs form short hairpin loops which act as a self-primer for second-strand cDNA synthesis. After incubation of second-strand reaction mixes and recovery of ds cDNA (Section 2.3.2(c)), hairpin loops and any remaining single-stranded cDNAs were digested with S1 nuclease (Section 2.3.2(c)). Efficiency of second-strand synthesis was approximately 90%. The ds cDNA was fractionated by sucrose gradient centrifugation and the highest molecular weight fractions, approximately 90 ng of ds cDNA, collected (Section 2.3.2(c)).

Approximately 30 ng of fractionated ds cDNA was tailed with dC residues as described in Section 2.3.2(d). Approximately 20 dC residues were added per ds cDNA 3' end.

The above described synthesis and dC-tailing of ds cDNA was carried out in association with Paul Krieg (this laboratory).

(b) Preparation of vector, annealing to ds cDNA and transformation

5 ug of the 4.3 kb plasmid vector pBR322 (Bolivar et al., 1977; Sutcliffe, 1978), which contains genes for resistance to both ampicillin and tetracycline, was digested with Pst I (pBR322 contains a unique Pst I site within the ampicillin resistance gene) and dG residues added to 3' ends (approx. 10 per 3' end) by terminal deoxynucleotidyl transferase as described in Section 2.3.2(e).

10 ng of dC-tailed ds cDNA was annealed to 100 ng of dG-tailed vector in 100 ul of annealing buffer (Section 2.3.2(e)). Annealed DNA (5 ng at a time) was used to transform competent E.coli MC1061 (Section 2.3.2(f)), which were then plated onto L-agar plates containing tetracycline. A total of 5,215 tetracycline resistant (tet^R) colonies were obtained. 1,176 of the tet^R colonies were also found to be ampicillin sensitive (amp^S), and

hence represented colonies containing recombinant molecules.

3.2.3 Detection of cDNA Recombinants Containing Histone Sequences

To identify cDNA recombinants containing H1 coding sequences it was originally decided to screen the cDNA library with a probe that would detect all histone sequences and then to screen out core histone sequences with available homologous core histone gene probes. The highly conserved 23 bp histone gene 3' dyad-symmetry element (Section 1.6.2(a)), was chosen as a positive probe for histone sequences. This element is present in most sequenced histone genes, including most chicken histone genes, and is located approximately 40 bp downstream from termination codons (Section 1.6.2(a)). Histone mRNA 3' termini generally map at or near the last base of the conserved sequence (Hentschel and Birnstiel, 1981; Maxson *et al.*, 1983b).

A 35 bp *Alu* I restriction fragment containing most of the conserved 3' element was isolated from a chicken H2A gene sequenced in this laboratory (D'Andrea *et al.*, 1981). This fragment, when ^{32}P-labelled with polynucleotide kinase (Section 2.3.7(d)) and used to probe restriction enzyme digested chicken genomal clones (Section 2.3.1(d)), could detect regions of DNA containing all four core histone coding sequences. The *Alu* I fragment was therefore a suitable probe to detect chicken core histone and presumably H1 sequences.

During the final stages of the preparation of the chicken embryo cDNA library and the preparation of the probe described above, a chicken H1 gene was identified in the chicken genomal clone λCHO2 (A. Robins and S. Clark, unpublished). λCHO2 was one of two genomal clones (λCHO1 and λCHO2) that had been characterized in this laboratory at the commencement of this project (Harvey and Wells, 1979; Harvey *et al.*, 1981). The H1 gene was identified by random sequencing of restriction fragments derived from

λCH02. The complete sequence of this gene has since been determined by Sugarman et al. (1983; discussed in Section 5.2.5). Even though the location of this gene removed the need to isolate an H1 cDNA recombinant for use as a probe, screening of the cDNA library still went ahead as it was of interest to examine expressed H1 sequences and to compare these to genomal H1 sequences. As a continuation of work in this laboratory, directed towards the analysis of chicken core histone genes, the library was also screened for core histone sequences.

(a) Probe preparation

DNA fragments containing each of the four core histone genes had previously been isolated from the chicken genomal clone λCH01 and sub cloned into pBR322 by R. D'Andrea (1985) for use as sequence-specific probes. The constructed subclones have been named pCH-H2AH, pCH-H2BH, pCH-H4H and pCH-H3ES. Inserts of the former two clones were derived from the λCH01 subclone pCH3.3E, while inserts for the latter two were derived from the subclone pCH3.75EH (Harvey et al., 1981). All cloned fragments have been sequenced (D'Andrea et al., 1981; Harvey et al., 1982; Wang et al., 1985).

The insert of pCH-H2AH is 709 bp long and contains an entire H2A coding region, 123 bp of 3' non-coding and 196 bp of 5' non-coding DNA. The pCH-H2BH insert is 537 bp long and contains an entire H2B coding region, 146 bp of 5' non-coding and 10 bp of 3' non-coding DNA. A complete H4 coding region plus 14 bp of 5' non-coding and 12 bp of 3' non-coding DNA is contained within the 338 bp insert of pCH-H4H, and the 150 bp insert of pCH-H3ES contains a portion of an H3 coding region which codes for amino acid 10 to amino acid 59.

Inserts of pCH-H2AH, pCH-H2BH and pCH-H4H are excisable with HindIII while the pCH-H3ES insert is excisable with a combination of EcoRI and

SalI.

In order to avoid the problems of insert contamination by pBR322 sequences, which would result in cross-reaction of probes with the vector sequences of cDNA recombinants, the inserts of the four histone recombinants (pCH-H2AH etc.) were re-cloned into M13 vectors (Section 2.3.8) as described below.

The four histone subclones were digested with the appropriate restriction enzymes (Section 2.3.1(a)) for release of inserts which were subsequently isolated by elution from a preparative 6% polyacrylamide gel (Section 2.3.1(c)). H2A, H2B and H4 inserts were ligated to HindIII digested M13 mp83 vector and the H3 insert was ligated to EcoRI/SalI double-digested M13 mp83 (Section 2.3.8(c)). Ligated DNA was used to transform the E. coli host JM101 (Section 2.3.8(c)) and double-stranded DNA was prepared from recombinant 'phage (Section 2.3.8(a)). Digestion of DNA with appropriate restriction enzymes followed by analysis on a 6% polyacrylamide gel (Section 2.3.1(c)) revealed that each of the four putative recombinant clones contained the expected size insert (Fig. 3.1). The recombinants were subsequently referred to as MCH-H2A, MCH-H2B, MCH-H3 and MCH-H4, containing respectively H2A, H2B, H3 and H4 inserts.

An M13 recombinant (CHE-7R) containing the λCHO2 H1 gene was kindly provided by A. Robins. CHE-7R was constructed by blunt-end ligation of a 1.3 kb DNA fragment into the HincII site of the M13 vector mp7. The 1.3 kb insert contains the entire H1 coding region, approximately 40 bp of 3' non-coding DNA and approximately 600 bp of 5' non-coding DNA. Double-stranded DNA was prepared from the recombinant (Section 2.3.8(a)). The insert was excised by digestion with BamHI (BamHI sites immediately flank HincII sites in mp7).

For screening of cDNA recombinants, DNA from the five M13 clones (MCH-H2A, MCH-H2B, MCH-H3, MCH-H4 and CHE-7R) was digested with restriction enzymes to release inserts. Inserts were purified on 10-40% linear sucrose

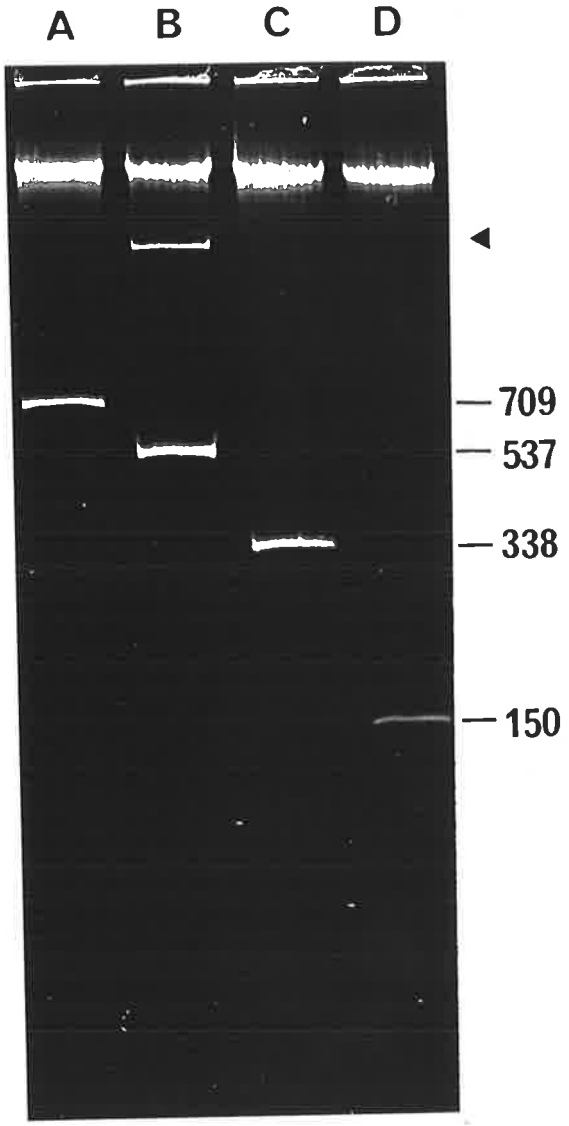
FIGURE 3.1

Digestion of histone gene-containing M13 recombinants

1 ug double-stranded DNA of each of the M13 recombinants MCH-H2A, MCH-H2B, MCH-H4 and MCH-H3 (Section 3.2.3(a)) was digested with restriction enzymes for release of histone gene-containing inserts. DNA was electrophoresed on a 6% polyacrylamide gel (Section 2.3.1(c)) and visualized under UV light following ethidium bromide staining.

Track: A. MCH-H2A digested with HindIII
B. MCH-H2B digested with HindIII
C. MCH-H4 digested with HindIII
D. MCH-H3 digested with EcoRI and SalI

The sizes of released inserts are indicated in base pairs. MCH-H2B contains a "double" H2B insert, viz., two H2B gene-containing HindIII fragments ligated together. The arrow head indicates the insert dimer, which resulted from incomplete digestion of MCH-H2B.


A
B
C
D
709
537
338
150

gradients (180,000 g, 16 hrs, 4°C), labelled by nick-translation incorporation of α-^{32}P-dATP and α-^{32}P-dCTP (Section 2.3.7(a)), and further purified by elution from a 6% polyacrylamide gel (Section 2.3.1(c)).

(b) Screening cDNA recombinants

The five probes prepared above were used to screen the 1,176 bacterial colonies containing recombinant molecules (*amp*^{S} *tet*^{R} colonies; Section 3.2.2(b)) as described in Section 2.3.3(a). Colonies were dotted in duplicate on to nitrocellulose filters. Approximately 1 x 10^6 dpm of ^{32}P-labelled probe were used per nitrocellulose filter (each filter contained 100 duplicate colonies) and filters were washed in 0.5 x SSC, 0.1% (w/v) SDS as described in Section 2.3.1(d). Positive and negative controls were included in each hybridization.

One H1, three H2A, five H2B, one H3 and five H4 positive colonies were detected. The single H1 positive response is shown in Figure 3.2. The cDNA clones were named cH1.1, cH2A.1, cH2A.2, cH2A.3. etc. (See Fig. 3.3).

3.2.4 Analysis of Core Histone cDNA Clones

(a) Miniscreen and hybridization analysis

DNA was prepared from each of the recombinants positive for core histone probes using the miniscreen procedure (Section 2.3.3(b)). This DNA was digested with EcoRI, to linearize recombinants, and analyzed on 0.8% agarose gels (Section 2.3.1(b); Fig. 3.3). A small amount of linearized pBR322 DNA was also included in each gel track so that it could be determined whether or not potential recombinants contained inserts - linearized molecules containing inserts will migrate more slowly than the linearized vector. Alternatively inserts of recombinant molecules could

FIGURE 3.2

Detection of an H1 positive cDNA recombinant

1,176 bacterial colonies containing recombinant molecules (*amp*^{S}, *tet*^{R}; Section 3.2.2(b)) were screened with the ^{32}P-labelled insert of the H1 gene-containing M13 recombinant CHE-7R (Section 3.2.3(a)) as described in Section 3.2.3(b). Colonies were dotted on to nitrocellulose filters in duplicate. 100 duplicate colonies were placed on each nitrocellulose filter in a 10 x 10 grid pattern. Each filter was probed with approximately 1 x 10^6 dpm of H1 probe, washed (final wash in 0.5 x SSC, 0.1% (w/v) SDS at 65°C) and autoradiographed as described in Section 2.3.1(d). Resultant autoradiograms revealed only a single H1 positive response. The autoradiogram of the filter containing the H1 positive recombinant, cH1.1, is shown. The positive signal is indicated with an arrow. The grid pattern on the nitrocellulose filter is also shown.

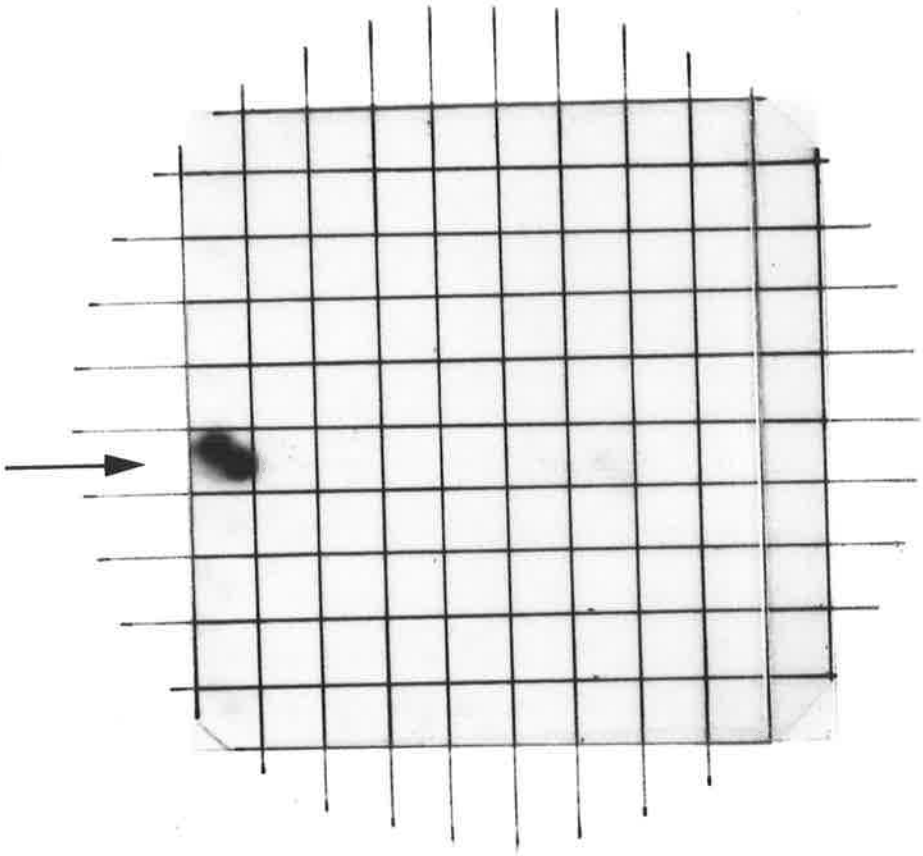


FIGURE 3.3

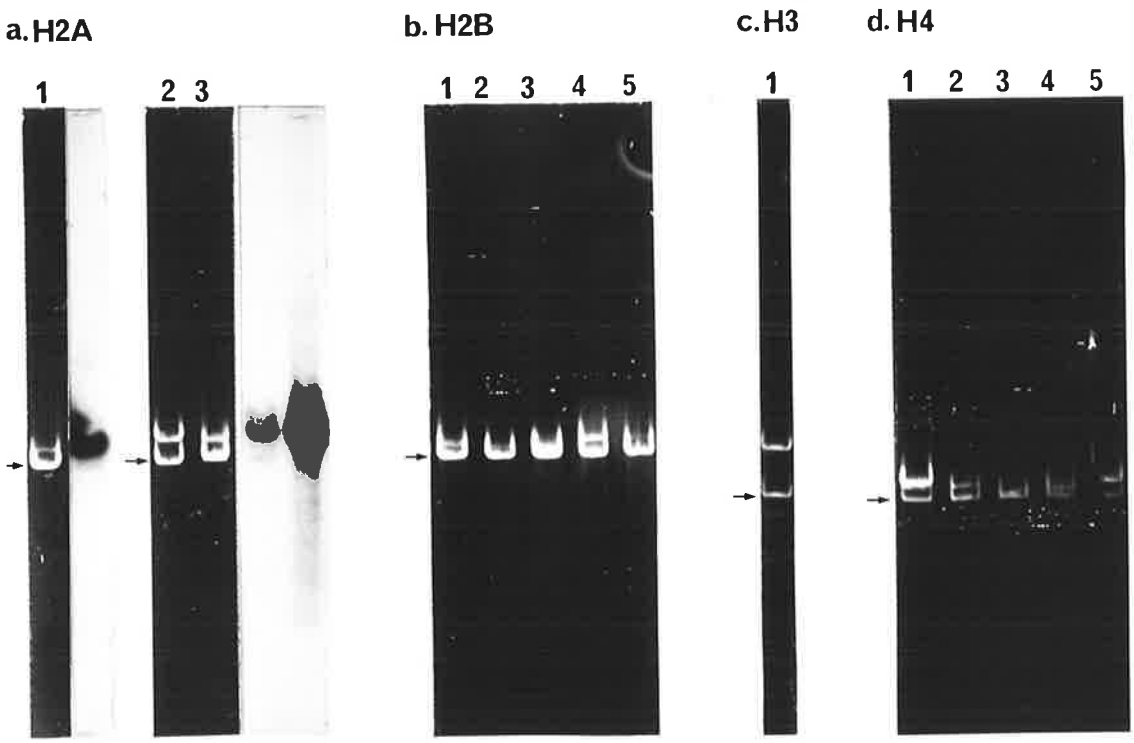
Miniscreen and hybridization analysis of core histone cDNA clones

DNA was prepared from cDNA recombinants positive to core histone gene probes by the miniscreen procedure (Section 2.3.3(b)), digested with EcoRI and analyzed on 0.8% vertical agarose gels (Section 2.3.1(b)). 50-100 ng of EcoRI digested pBR322 DNA was also included in each gel track as a marker. The digestions of recombinants are shown:

a) H2A positive recombinants
   Track: 1. cH2A.1
          2. cH2A.2
          3. cH2A.3
b) H2B positive recombinants
   Track: 1. cH2B.1
          2. cH2B.2
          3. cH2B.3
          4. cH2B.4
          5. cH2B.5
c) H3 positive recombinant
   Track: 1. cH3.1
d) H4 positive recombinants
   Track: 1. cH4.1
          2. cH4.2
          3. cH4.3
          4. cH4.4
          5. cH4.5

Arrows indicate EcoRI digested pBR322. Linearized recombinant molecules migrate more slowly than the vector DNA.

DNA from gels was transferred to nitrocellulose and hybridized with appropriate ^{32}P-labelled core histone gene probes (Section 3.2.3(a)) as described in Section 2.3.1(d). Final filter washes were in 0.5 x SSC, 0.1% (w/v) SDS. Resultant autoradiograms, after hybridization of H2A positive recombinants with the H2A gene probe, are shown adjacent to respective gel tracks. From left to right, autoradiogram tracks represent hybridization to cH2A.1, cH2A.2 and cH2A.3.


a. H2A
1
2 3
b. H2B
1 2 3 4 5
c. H3
1
d. H4
1 2 3 4 5

have been released by digestion with PstI as the cloning procedure results in the regeneration of PstI sites flanking ds cDNA inserts. Contaminants in miniscreen DNA preparations however, usually inhibit PstI activity. The regeneration of PstI sites is also not 100% efficient (Villa-Komaroff et al., 1978).

Most recombinant molecules appeared to have inserts (Fig. 3.3) and most of these varied in size from about 100 to 600 bp. Surprisingly the single H3 recombinant (cH3.1) contained an insert of approximately 1.4 kb. S-phase regulated chicken H3 genes, which represent most of the H3 genes in the chicken genome (Section 1.5.3(d)) have transcripts of approximately 500 bases in length (S. Dalton, unpublished; Engel et al., 1982; Sugarman et al., 1983). The 1.4 kb insert may therefore represent the transcript of a different type of H3 gene. It has recently been reported that the genes coding for the replication-independent chicken H3.3 protein variant (Section 1.5.3(d)) produce large transcripts (Brush et al., 1985), hence the cDNA clone could be derived from such a gene. This cDNA clone was not investigated further.

To confirm the nature of recombinants, EcoRI digested DNA was transferred to nitrocellulose and hybridized with appropriate ^{32}P-labelled core histone gene probes (Section 3.2.3(a)) as described in Section 2.3.1(d). All clones were positive with their respective probes. Probes did not cross-react with pBR322 sequences, therefore, positive results were due to cross-reaction of probes with insert sequences.

The results of hybridization analysis of the three H2A clones is shown in Figure 3.3. Interestingly the second H2A clone (cH2A.2) hybridized weakly to the H2A probe, in comparison to the adjacent H2A clone (cH2A.3) which had a similar size insert and of which there was a similar amount of DNA. This suggested that cH2A.2 contained sequences that may diverge considerably from the H2A gene used as probe (Section 3.2.3(a)). Variations in hybridization efficiency were also observed between the five

H2B clones.

cH2A.2 (renamed pH2A.F) and the five H2B cDNA clones (cH2B.1-cH2B.5) were further analyzed by DNA sequencing using the M13 subcloning and dideoxy sequence analysis procedures (Sections 2.3.8, 2.3.9 and 2.3.10). This work was done by other members of the laboratory and is discussed below.

(b) Sequence analysis

pH2A.F - an extremely variant H2A sequence

The 650 bp insert of pH2A.F contains a complete coding region for an H2A-like protein, 70 bp 5' to the ATG codon and 193 bp 3' to the termination codon (Harvey et al., 1983). The H2A variant coded for by pH2A.F is 40% divergent from the most abundant chicken H2A protein, H2A.1. The H2A gene probe used to isolate pH2A.F (Section 3.2.3(a)) represents one of the genes that code for the H2A.1 subtype (D'Andrea et al., 1981). This degree of divergence is striking given that the H2A.1 sequences of calf and chicken show only 5% divergence. The H2A.F protein most probably represents the replication-independent H2A variant H2A.Z (also called M1; Section 1.2.3), as determined by comparison to partial amino acid sequence data of calf thymus H2A.Z (Ball et al., 1983) and by comparison to amino acid composition data of chicken H2A.Z (Urban et al., 1979) - J. Whiting, personal communication. The partial amino acid sequence of calf thymus H2A.Z, which covers the first 30 residues of the protein, differs by only one amino acid from the predicted H2A.F amino acid sequence over this region.

Consistent with the postulate that H2A.F is H2A.Z is the finding that the expression of mRNA from the H2A.F gene is not S-phase regulated (Dalton et al., 1985).

pH2A.F has been used to isolate the genomal H2A.F gene from a chicken

genomal library (Robins et al., 1985). The H2A.F gene represents one of a set of replication-independent chicken genes that are not closely linked to each other or to the major cluster of replication-dependent histone genes (Section 1.5.3(d)). The properties of the replication-independent genes that distinguish them from the majority of chicken histone genes are discussed in Sections 1.5.3(d) and 1.6.3(b).

Antibodies to protein derived from pH2A.F are at present being used to determine the distribution of H2A.F in chromatin (J. Whiting, unpublished) as it has been found that H2A.Z-like proteins appear to be enriched in active chromatin (Section 1.2.4(a)). H2A.Z-like proteins may play a part in determining the structure of active chromatin (Section 1.2.4(b); Ball et al., 1983).

H2B cDNA clones

The five H2B cDNA clones were sequenced by P. Wigley (1982). The insert of cH2B.1 was 325 bp long and contained sequences coding for amino acid 12 to 119 of an H2B protein. The insert of cH2B.2 was 228 bp long, commenced at sequence coding for amino acid 63 and ended 35 bp 3' to the termination codon. The 416 bp insert of cH2B.4 contained sequence commencing at that coding for amino acid 5 and ending 51 bp 3' to the termination codon. The 422 bp cH2B.5 insert contained an entire H2B coding region, 18 bp 5' to the ATG codon and 23 bp 3' to the termination codon. The cH2B.3 insert contained a small stretch of H2B coding sequence linked to non-histone coding sequences and will not be discussed further.

The cH2B.1, .2, .4 and .5 sequences were compared to the H2B genes from the genomal clones λCH01 (clone from which the H2B probe was derived; Section 3.2.3(a)) and λCH02 (Harvey et al., 1982). The insert sequence of cH2B.1 is identical to sequences within the λCH02 H2B coding region and hence may be derived from a transcript of the λCH02 gene. Inserts of

cH2B.2, .4 and .5 did not appear to be derived from either genomal H2B sequence, and the amino acid sequences derived from coding regions of cH2B.4 and .5 differed from the predicted λCHO1 and λCHO2 H2B protein sequences (both genomal genes encode identical proteins) by one and two amino acids respectively. None of the H2B cDNA clones, therefore, appear to code for extreme H2B variants.

Urban et al. (1979) have identified two major chicken H2B variants (H2B.1 and H2B.2; Section 1.2.3) and have determined by peptide mapping that there are at least 2 amino acid differences between the two proteins. H2B.1 has serine at position 32 and glycine at position 60, whereas, H2B.2 has alanine and serine respectively at position 32 and 60. The proteins encoded by the genomal H2B genes and cH2B.4 contain the two amino acids that distinguish H2B.1 from H2B.2. In cH2B.5, however, the serine at position 32 of H2B.1 proteins has been replaced by a threonine residue. None of the H2B sequences code for the H2B.2 subtype. It has been proposed that this variant is a replication-independent variant (Urban and Zweidler, 1983; Section 1.2.3).

3.2.5 Analysis of an H1 cDNA Clone

A large-scale DNA preparation (Section 2.3.5) was performed on the single H1 positive cDNA recombinant cH1.1 (Section 3.2.3(b)). cH1.1 DNA was digested with PstI and analyzed on an agarose gel to determine the cH1.1 insert size. Digestion released an insert of approximately 600 bp.

The DNA sequence of the cH1.1 insert was determined from both BAL-31 nuclease (Section 2.3.8(b)) and restriction enzyme generated DNA fragments, using the M13-cloning and di-deoxy sequencing procedures (Section 2.3.8, 2.3.9 and 2.3.10).

(a) Preparation of M13 recombinants

Cloning BAL-31 generated fragments

cH1.1 DNA was digested with PstI, and the cDNA insert isolated from a 1% preparative LGT agarose gel (Section 2.3.1(b)). Insert DNA (approx. 2 ug) was then digested with BAL-31 in a 25 ul reaction mix as described in Section 2.3.8(b). After varying lengths of incubation (viz., after 15 s, 1 min, 2 min and 3 min) 6 ul aliquots of the reaction mix were removed and added to phenol to stop digestion. Digested DNA was then phenol/chloroform extracted, ethanol precipitated and treated to generate blunt ends (Section 2.3.8(b)). Digestion of duplex DNA with BAL-31 does not always produce blunt-ends, as removal of 5' and 3' termini by the nuclease is not synchronous. BAL-31 treated, blunt-ended DNA was then fractionated on a 6% polyacrylamide gel and four size fractions of digested DNA were isolated (Fig. 3.4.(a)). The four fractions contained respectively, DNA fragments between, approximately 500-400 bp, 400-300 bp, 300-200 bp and 200-150 bp.

DNA from each size fraction was cloned into the SmaI site of the M13 vector mp93 as described in Section 2.3.8(c). Single-stranded DNA was then prepared from recombinant 'phage (Section 2.3.8(d)) and analyzed by the complementarity testing (c-testing) procedure (Section 2.3.9(a)). Recombinants containing DNA from each BAL-31 size fraction, in both orientations (as determined by c-testing), were then used as templates for DNA sequence analysis (Section 2.3.10; Fig 3.5).

Cloning FnuDII restriction fragments

Approximately 3 ug of cH1.1 DNA was digested with FnuDII and fractionated, next to 3ug of FnuDII digested pBR322 vector DNA, on a 6% polyacrylamide gel (Fig. 3.4(b)). The three DNA fragments (indicated in Fig. 3.4(b)) unique to digested cH1.1 DNA were

FIGURE 3.4

Preparation of DNA fragments from cH1.1 for sequence determination

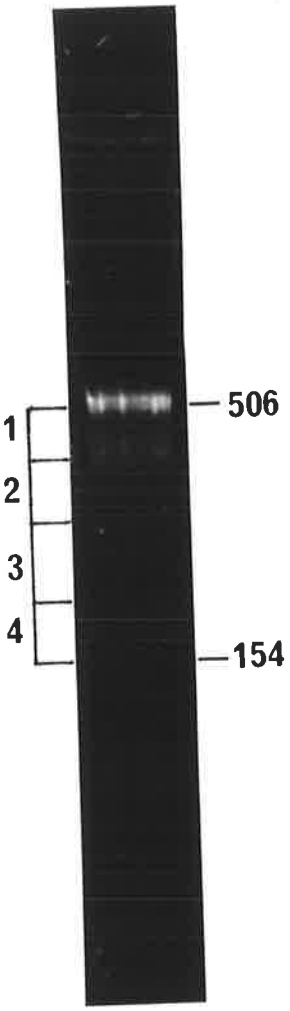
a) BAL-31 generated DNA fragments

Insert DNA from cH1.1 was prepared, treated with BAL-31 and blunt-ended as described in Section 3.2.5(a), and then fractionated on a 6% polyacrylamide gel. DNA was visualized after staining with ethidium bromide. Four size fractions of DNA, numbered 1, 2, 3 and 4, were recovered from the gel, as described in Section 2.3.1(c), for cloning into an M13 vector. DNA marker sizes are indicated in base pairs.

b) FnuDII restriction fragments

Approximately 3 ug of cH1.1 DNA was digested with FnuDII and fractionated next to 3 ug of FnuDII digested pBR322 DNA, on a 6% polyacrylamide gel. DNA was visualized after staining with ethidium bromide. Gel tracks containing cH1.1 and pBR322 DNA are labelled respectively cH1.1 and 322. Sizes of pBR322 DNA fragments are given in base pairs. Three DNA fragments unique to cH1.1 are indicated with arrows. These fragments were recovered from the gel for cloning.

a)



b)

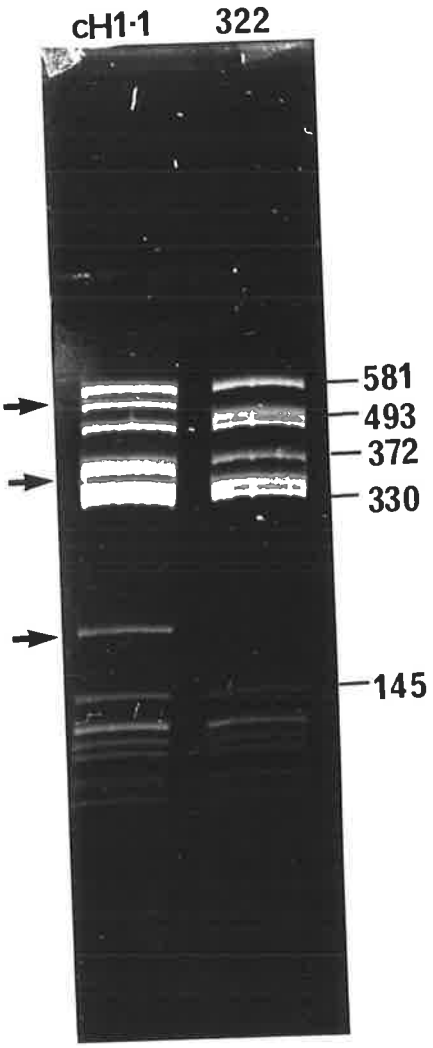
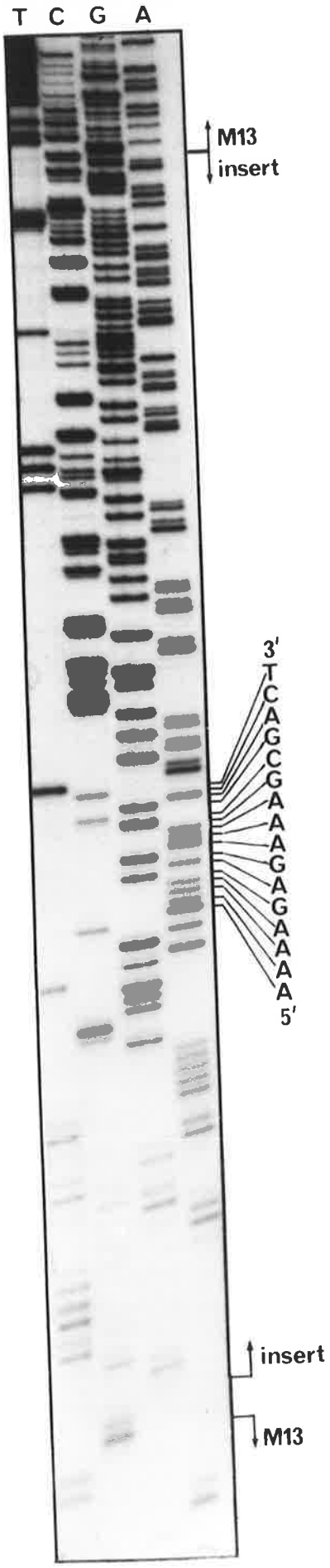


FIGURE 3.5

DNA sequence determination of BAL-31, cH1.1 insert fragments

M13 recombinants were constructed from BAL-31 generated fragments, derived from the insert of cH1.1, as described in Section 3.2.5(a). Single-stranded recombinant DNA was prepared (Section 2.3.8(d)), analyzed by complementarity testing (Section 2.3.9(a)) and recombinants containing DNA from different BAL-31 size fractions, in both orientations, were used as templates for DNA sequence analysis (Section 2.3.10).

The sequencing ladder of a BAL-31 fragment, cloned into the SmaI site of M13 mp93, is shown here. Samples were electrophoresed on a 6% polyacrylamide sequencing gel and bands were detected by autoradiography (Section 2.3.10(c)). The cloned BAL-31 fragment is 204 bp long and the DNA sequence reads, in a 5' to 3' direction, from residue number 163 to residue number 366 of the final cH1.1 insert sequence shown in Figure 3.7. A small portion of the sequence is shown next to the sequencing ladder.


T C G A
M13
insert
3′
TCAGCGAAAGAGAAAA
5′
insert
M13

recovered as described in Section 2.3.1(c). DNA was cloned into the SmaI site of M13 mp93 and recombinants were c-tested and sequenced as described above. The largest two FnuDII restriction fragments spanned 5' and 3' vector-insert junctions, while the third fragment (189 bp) was derived from within the insert.

(b) Sequence of the cH1.1 insert

The regions of the cH1.1 insert sequenced from the various cloned fragments described above are shown in Figure 3.6. 96% of the insert was sequenced in both directions.

The insert sequence (Fig. 3.7) is 536 bp long, not including GC-tails, and contains 510 bp of coding DNA (including termination codon) and 26 bp of 3' non-coding DNA. The sequence is identical to that derived from the same region of the λCHO2 H1 gene (the gene used as probe to isolate cH1.1; Section 3.2.3(a)) except for a single silent third base change (see Fig. 3.7) within the coding region. The cH1.1 sequence may therefore be derived from a transcript of the λCHO2 H1 gene. The region of the λCHO2 H1 coding region present in cH1.1 codes for amino acids 49 to 217 (terminal amino acid) of an H1 protein.

No poly(A) tract was present at the 3' end of the cDNA insert sequence, therefore the cloning procedure (Section 2.3.2) has truncated the 3' end of the mRNA sequence as well as the 5' end (the insert sequence commences within the coding region).

H1 gene coding and flanking sequences are discussed further in Chapter 6.

3.2.6 Discussion

Recombinants containing core and H1 histone coding sequences were

FIGURE 3.6

cH1.1 insert sequencing strategy

The regions of the cH1.1 cDNA insert sequenced, from cloned FnuDII restriction fragments and BAL-31 generated DNA fragments, are indicated by horizontal arrows, the direction of arrows represents the 5' to 3' direction of generated DNA sequence. The location of FnuDII sites (F) in the insert are shown. The large arrow indicates the H1 protein coding regions of the cDNA insert and the direction of gene transcription.

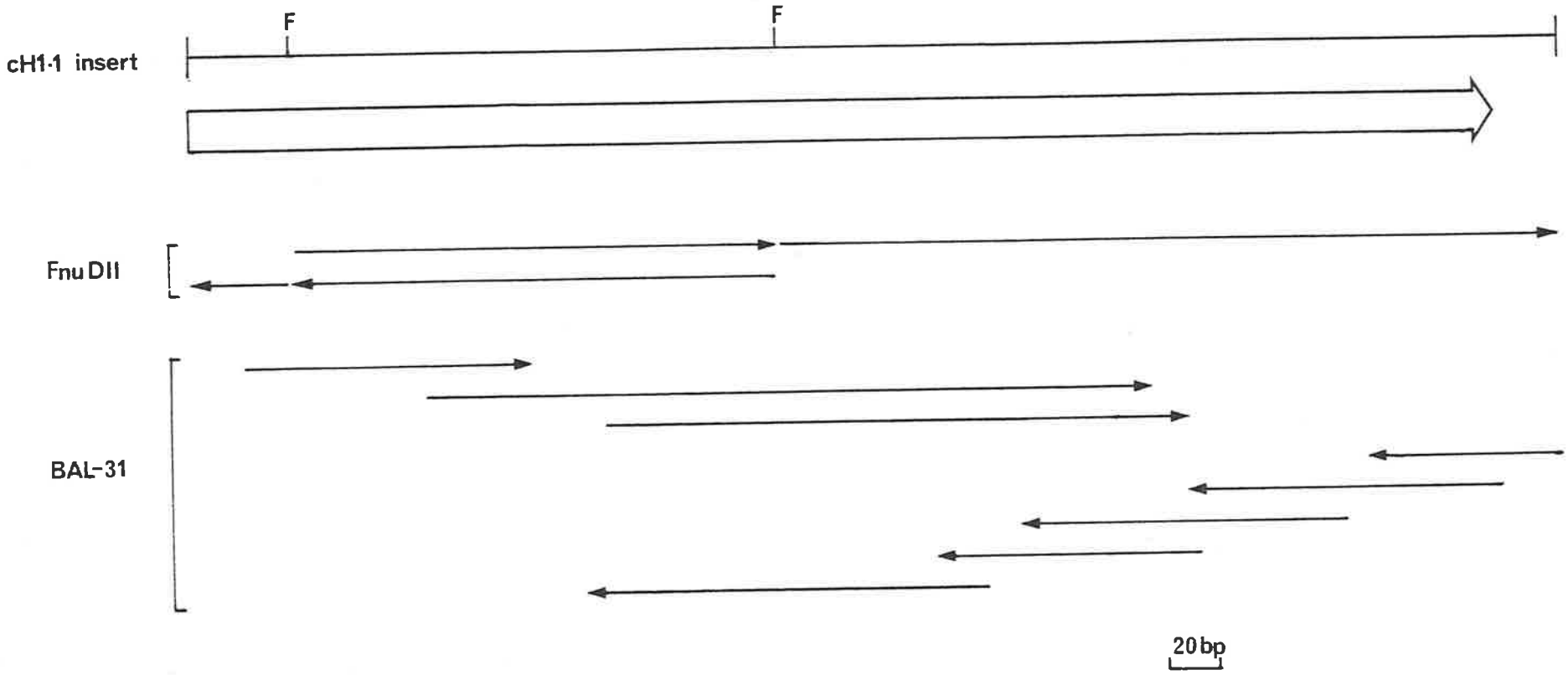

F
F
cH1-1 insert
Fnu DII
BAL-31
20 bp

FIGURE 3.7

Nucleotide sequence of the cH1.1 insert

The nucleotide sequence of the 536 bp cH1.1 insert (not including GC-tails) is presented. Only the coding strand is given and is presented in a 5' to 3' direction. FnuDII restriction sites used in sequence analysis are indicated. The sequence is identical to that derived from the corresponding region of the λCHO2 H1 gene, except that the λCHO2 H1 gene has a T residue (marked by an *) instead of the G residue found at nucleotide number 54 of the cH1.1 sequence.

The complete λCHO2 H1 gene sequence was determined by Sugarman et al. (1983) and is discussed further in Section 5.2.5.

cH1·1

ser ala ser lys glu arg lys gly leu ser leu ala ala leu lys lys ala leu ala ala
T C C G C C T C C A A G G A G C G C A A G G G G C T C T C C C T C G C C G C G C T C A A G A A G G C G C T G G C C G C C 60
FnuDII
T*

gly gly tyr asp val glu lys asn asn ser arg ile lys leu gly leu lys ser leu val
G G C G G C T A C G A C G T G G A G A A G A A C A A C A G C C G C A T C A A G C T G G G G C T C A A G A G C C T C G T C 120

ser lys gly thr leu val gln thr lys gly thr gly ala ser gly ser phe lys leu asn
A G C A A G G G C A C C C T G G T G C A G A C C A A G G G C A C C G G C G C C T C G G G C T C T T T C A A G C T G A A T 180

lys lys pro gly glu thr lys glu lys ala thr lys lys lys pro ala ala lys pro lys
A A A A A G C C G G G T G A G A C A A A A G A G A A A G C G A C T A A G A A G A A G C C C G C G G C C A A G C C C A A G 240
FnuDII

lys pro ala ala lys lys pro ala ala ala ala lys lys pro lys lys ala ala ala val
A A G C C G G C G G C C A A G A A G C C T G C G G C T G C T G C C A A G A A G C C C A A G A A G G C A G C G G C G G T G 300

lys lys ser pro lys lys ala lys lys pro ala ala ala ala thr lys lys ala ala lys
A A G A A G A G C C C C A A G A A A G C C A A G A A G C C G G C A G C T G C T G C C A C C A A G A A G G C G G C C A A G 360

ser pro lys lys ala thr lys ala gly arg pro lys lys thr ala lys ser pro ala lys
A G C C C C A A G A A G G C T A C C A A G G C T G G C C G C C C C A A G A A G A C T G C C A A G A G C C C G G C C A A G 420

ala lys ala val lys pro lys ala ala lys ser lys ala ala lys pro lys ala ala lys
G C A A A G G C G G T G A A G C C C A A A G C T G C C A A G T C A A A G G C G G C C A A A C C C A A G G C G G C C A A G 480

ala lys lys ala ala thr lys lys lys ***
G C A A A G A A G G C A G C G A C C A A A A A G A A G T A A G A T G A C A G A A G A A A T T C G A G T C T G C T 536

isolated from a cDNA library that was constructed from the 5-13S fraction of 5 day old chicken embryo RNA.

Sequence analysis of an H2A recombinant (pH2A.F), that hybridized weakly to the probe with which it was isolated, resulted in the indentification of a sequence coding for an extremely variant H2A protein (Harvey et al., 1983). The protein is most probably the previously described replication-independent H2A variant H2A.Z.

The single H3 recombinant (cH3.1) was also unusual in that it had an insert much larger than the transcripts derived from the replication-dependent chicken H3 genes and, as for pH2A.F, may be derived from a replication-independent gene, such as the chicken H3.3 genes (Brush et al., 1985).

Sequence analysis of H2B recombinants (Wigley, 1982) did not reveal any extremely variant sequences. None of the H2B sequences coded for the replication-independent variant H2B.2.

The 536 bp insert of the single H1 recombinant, cH1.1, was most probably derived from a transcript of the H1 gene used to isolate cH1.1. The clone insert consisted mostly of coding DNA and as such was a convenient probe for further investigation of genomal H1 sequences, as discussed in the following chapters.

CHAPTER 4

ISOLATION AND CHARACTERIZATION OF H1 GENE-CONTAINING GENOMAL CLONES

4.1 INTRODUCTION

The two chicken genomal clones, λCHO1 and λCHO2 (Harvey and Wells, 1979; Harvey et al., 1981), characterized at the commencement of this project were isolated from a library of clones consisting of randomly generated chicken DNA fragments (15-21 kb) in the bacteriophage vector Charon 4A. The library was constructed by Dodgson et al.(1979), using the procedure described by Maniatis et al. (1978), and contains approximately 5 x 10^5 independent recombinants.

As previously discussed (Section 3.2.3) an H1 gene was indentified in the λCHO2 clone by random sequencing (A. Robins and S. Clark, unpublished). Subsequently λCHO1 was also found by hybridization analysis to contain H1 coding sequences (D'Andrea, 1985).

This chapter describes the isolation and characterization of additional H1 gene-containing genomal clones, as a first step in the analysis of the chicken H1 gene complement.

4.2 RESULTS

4.2.1 Screening the Chicken Genomal Library for H1 Sequences

The chicken library constructed by Dodgson et al. (1979) was screened for H1 sequences using the insert of the H1 cDNA clone, cH1.1 (Section 3.2.5), as probe.

Approximately 5 x 10^5 'phage from the chicken library were plated on to five plates (15 cm in diameter) and transferred to duplicate nitrocellulose filters (Section 2.3.6(a)). The number of 'phage required for a "complete library" (viz., a library having 99% probability of containing any sequence present in the genome) is 5.4 x 10^5 for chicken (Clarke and Carbon, 1976; the chicken genome is 2 x 10^9 bp, see Old and

Woodland, 1984). The insert of cH1.1 was isolated from an LGT agarose gel (Section 2.3.1(b)) after digestion with PstI (see Section 3.2.5) and ^{32}P-labelled by nick-translation (Section 2.3.7(a)). Nitrocellulose filters were hybridized with a total of 2.4 x 10^7 dpm of labelled insert (2.4 x 10^6 dpm per filter) washed (final wash, 0.5 x SSC, 0.1% (w/v) SDS at 65°C) and autoradiographed as previously described (Section 2.3.1(d)). A total of nineteen plaques were positive on duplicate filters (Fig. 4.1). 'Phage from positive plaques were subjected to two more rounds of screening in order to obtain pure recombinants (see Fig. 4.2). Eighteen pure recombinants, positive to the cH1.1 insert, were obtained after the three rounds of screening.

R. D'Andrea (this laboratory) independently isolated twenty seven H1 positive recombinants after screening a total of 10 x 10^5 'phage from the chicken library for histone sequences. From these, clones were isolated (λCHO3 and λCHO5) which contained genomal DNA overlapping that within the λCHO1 clone (isolated by R. D'Andrea; D'Andrea, 1985; D'Andrea et al., 1985). The remaining recombinants, in addition to the eighteen isolates discussed above, were further analyzed.

4.2.2 Analysis of Genomal Clones Containing H1 Sequences

(a) Digestion and hybridization analysis

DNA was prepared from H1 positive recombinants by the "mini"-preparation procedure described in Section 2.3.6(b), digested with EcoRI and analyzed on 1% agarose gels. Recombinants were also screened for their histone coding potential using the "Dot-Benton" procedure (Section 2.3.6(a); done in conjunction with R. D'Andrea). Prepared nitrocellulose filters were hybridized with inserts from cH1.1 (Section 3.2.5), pCH-H2AH, pCH-H2BH, pCH-H4H and pCH-H3ES (Section 3.2.3(a)) which had been

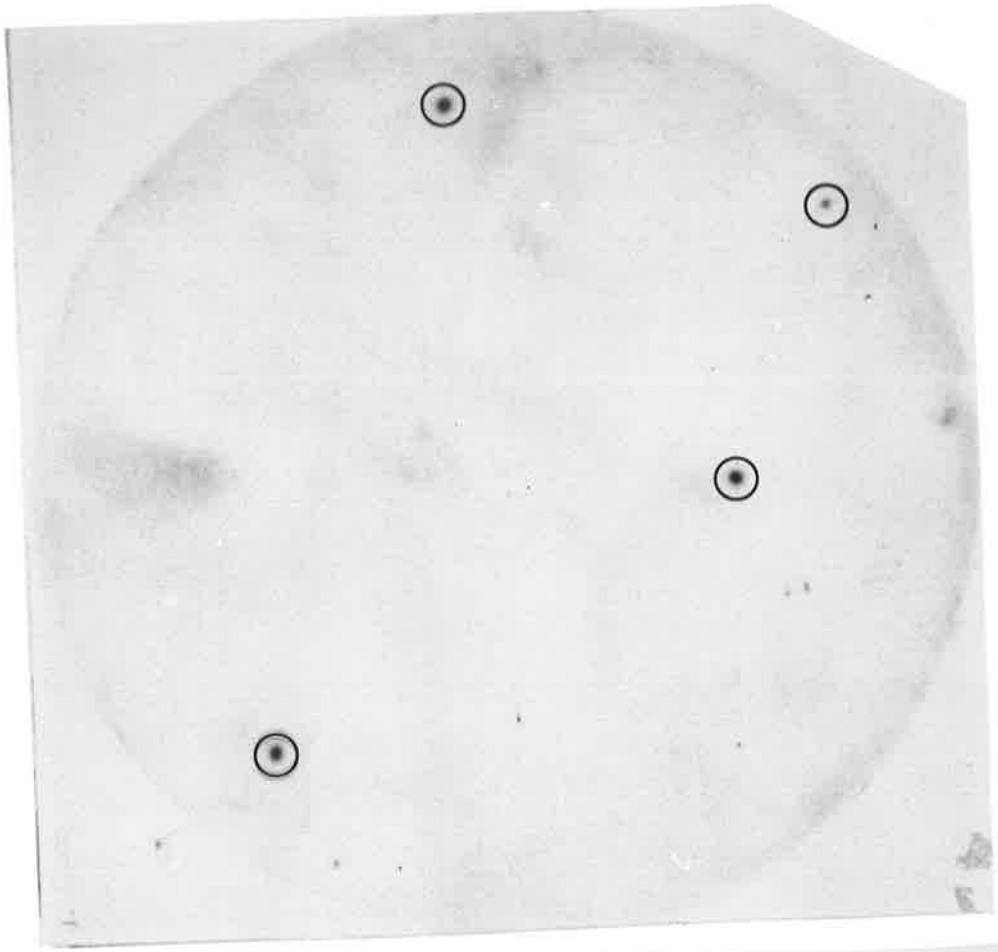
FIGURE 4.1

Primary screening of the chicken genomal library for H1 sequences

Approximately 5 x 10^5 'phage from a chicken genomal library (Dodgson et al., 1979) were plated on to five plates (15 cm in diameter) and transferred to duplicate nitrocellulose filters as described in Section 2.3.6(a). Filters were hybridized with the ^{32}P-labelled insert of the H1 cDNA clone cH1.1 (2.4 x 10^6 dpm per filter; probe prepared as described in Section 4.2.1), washed (final wash in 0.5 x SSC, 0.1% (w/v) SDS at 65°C) and autoradiographed (Section 2.3.1(d)).

A total of nineteen plaques were positive in duplicate. Four positive responses, on the autoradiograms of a single filter(a) and its duplicate(b), are shown circled.

a.



b.

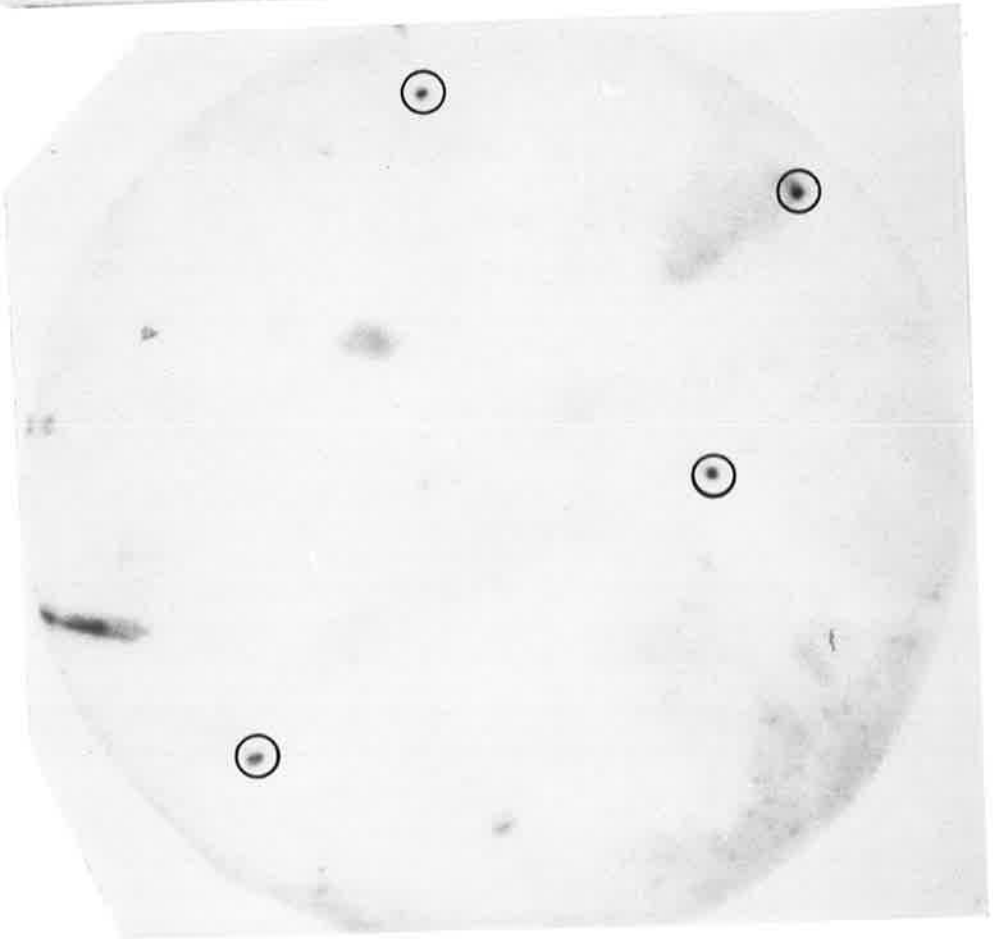


FIGURE 4.2

Third round screening of primary H1 positive recombinants

H1 positive recombinants obtained after primary screening of the chicken genomic library (Fig. 4.1) were picked into PSB, replated (approx. 500 pfu/9 cm plate) and nitrocellulose filters prepared as described in Section 2.3.6(a). Filters were hybridized with cH1.1 probe, as for primary screenings (Fig. 4.1), and plaques containing cross-reacting 'phage were picked into PSB. Such 'phage were plated (approx. 100 pfu/plate) for third round screening, filters prepared and probed as above.

Positive responses were evident on autoradiograms of filters after three rounds of screening. Autoradiograms of two filters (a. and b.) are shown.

a.

b.

^{32}P-labelled by nick-translation (Section 2.3.7(a)). Inserts contained respectively H1, H2A, H2B, H4 and H3 coding sequences and were purified from LGT agarose gels (Section 2.3.1(a)), after digestion of insert-containing recombinants (cH1.1 etc.) with appropriate restriction enzymes (Section 3.2.5 and 3.2.3(a)). As expected all clones were positive with the H1 gene probe, hence confirming the nature of these recombinants.

By virtue of their EcoRI digestion patterns and histone gene content it was found that the "bank" of H1 positive recombinants contained, in addition to genomal clones that had been previously isolated in this laboratory (viz., λCHO1, λCHO2, λCHO3 and λCHO5), seven "new" recombinants. These recombinants were named λH1.1, λH1.2, λH1.3, λH1.4, λH1.7, λH1.9 and λH1.10 (collectively called λH1 clones). The EcoRI digestion patterns and histone gene coding potentials (as determined by dot-blot analysis, Section 2.3.1(d)) of each of the seven recombinants is shown in Figure 4.3. Results of dot-blot analysis agree with those of "Dot-Benton" analysis.

The 19.94 and 11.04 kb fragments generated by EcoRI digestion of recombinants (Fig. 4.3) represent vector "arms" (chicken DNA fragments were inserted into the Charon 4A vector via addition of synthetic EcoRI linkers; Dodgson et al., 1979) while the remaining fragments were derived from insert sequences. The recombinants contained insert DNA of between approximately 12 kb (λH1.7) and 16 kb (λH1.10). Each recombinant contained, in addition to H1 gene sequences, core histone gene sequences (Fig. 4.3). Hence, as for the H1 genes from λCHO1 and λCHO2, H1 genes in the recombinants described above appeared to be associated with core histone genes.

(b) Relationship of λH1 clones to other genomal clones

From extensive mapping of chicken histone gene-containing DNA which overlapped DNA within the genomal clones λCHO1 and λCHO2 (by R. D'Andrea

FIGURE 4.3

Digestion and hybridization analysis of H1 positive genomal clones

EcoRI digestion patterns and results of hybridization analysis of seven previously unidentified H1 positive recombinants (Section 4.2.2(a)) are shown.

a) DNA (approx. 1 ug) prepared from H1 positive recombinants, by the "mini"-preparation procedure (Section 2.3.6(b)), was digested with EcoRI and analyzed on 1% vertical agarose gels (Section 2.3.1(b)). Digestions of λH1.1, λH1.2, λH1.3, λH1.4, λH1.7, λH1.9 and λH1.10 are labelled 1, 2, 3, 4, 7, 9 and 10 respectively. Fragment sizes are given in kilobases. The 19.94 and 11.04 kb vector arms are indicated with arrows.

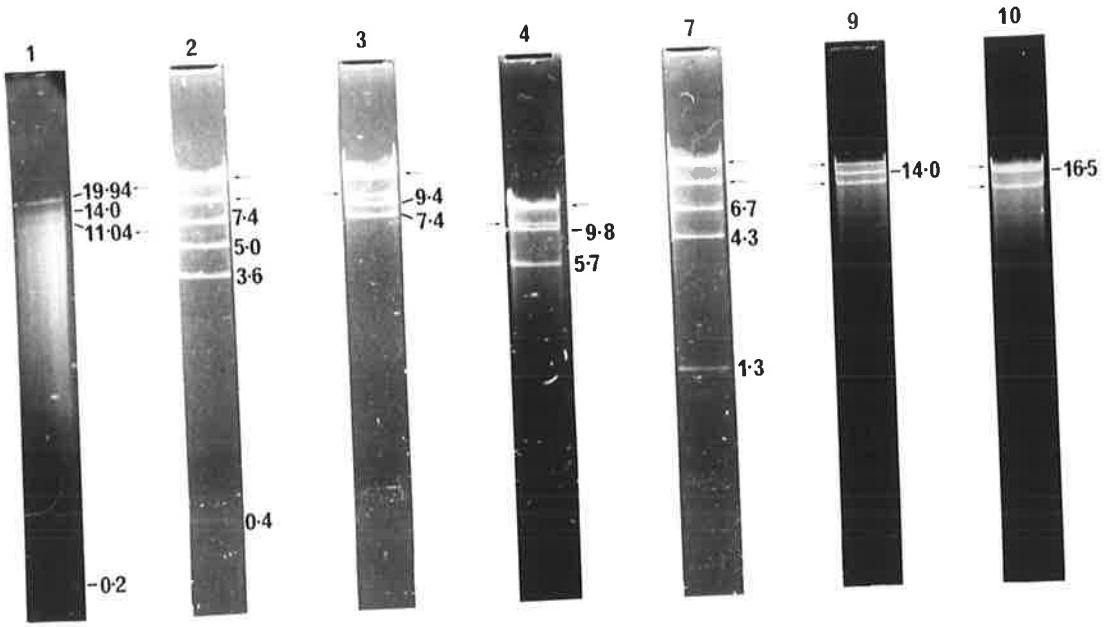
b) DNA (approx. 100 ng) of each recombinant was denatured and dotted in duplicate on to nitrocellulose filters, using the dot-blot procedure (Section 2.3.1(d)), and probed with histone gene probes (prepared as described in Section 4.2.2(a)). Final filter washes were in 0.5 x SSC, 0.1% (w/v) SDS at 65°C. Resultant autoradiograms are shown. Recombinants are numbered 1, 2, 3, 4, 7, 9 and 10 as above. Results of hybridization analysis are given below:

		Recombinant:						
		λH1.1	λH1.2	λH1.3	λH1.4	λH1.7	λH1.9	λH1.10
Gene probe:	H3	-	-	+	+	-	+	-
	H4	-	+	-	+	-	+	+
	H2A	+	*faint +	+	+	*faint +	-	+
	H2B	-	+	+	+	+	-	+
	H1	+	+	+	+	+	+	+

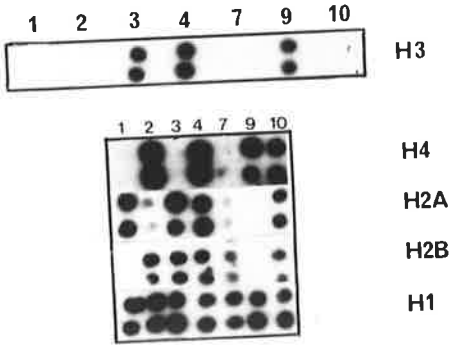
+ and - represent positive and negative responses respectively. No probes cross-reacted with λ vector sequences.

*λH1.2 and λH1.7 were found to cross-react faintly with the H2A probe. This was confirmed by Southern analysis (Section 2.3.1(d)), but was not investigated further.

a)


1
2
3
4
7
9
10
19·94
14·0
11·04
0·2
7·4
5·0
3·6
0·4
9·4
7·4
9·8
5·7
6·7
4·3
1·3
14·0
16·5

b)


1 2 3 4 7 9 10
H3
1 2 3 4 7 9 10
H4
H2A
H2B
H1

(1985) and C. Lesnikowski (1983); Figs. 4.4 and 4.5) a total of five H1 gene-containing regions were located (each representing a single H1 gene; see Section 4.2.5(c) and Chapter 5). Hybidization of restriction enzyme digested chicken genomal DNA with an H1 gene probe suggested that this did not represent the full complement of chicken H1 genes (see Section 4.2.6). Some of the seven "new" H1 positive genomal clones (λH1.1 etc.) may have therefore contained previously unidentified H1 genes.

In order to eliminate λH1 clones containing any of the five already located H1 genes various non-coding DNA fragments adjacent to mapped H1 genes were isolated and used to probe the λH1 clones. Locations of the DNA fragments are indicated in Figures 4.4 and 4.5. Fragment descriptions are given in Figure 4.6. Most DNA fragments were isolated from LGT agarose gels (Section 2.3.1(b)), after restriction enzyme digestion of appropriate recombinant molecules (Figs. 4.4 and 4.5), and then ^{32}P-labelled by nick-translation (Section 2.3.7(a)). The two non-coding DNA fragments, derived from λCH01, in M13 vectors (see Fig. 4.6), were ^{32}P-labelled as described in Section 2.3.7(c), using E. coli DNA polymerase I, Klenow fragment. Labelled inserts were purified from a 6% polyacrylamide gel (Section 2.3.1(c)).

DNA from each of the λH1 recombinants was denatured, dotted in duplicate on to nitrocellulose filters and probed with the labelled non-coding DNA fragments discussed above, as described in Section 2.3.1(d). Final filter washes were in 0.5 x SSC, 0.1% (w/v) SDS. The results of probings are shown in Figure 4.6 and are discussed below:

Four clones, λH1.1, 2, 4 and 7, were positive for probes adjacent to the λCH02 H1 gene (derived from the subclone pCH4.7E, Fig. 4.5) and one clone, λH1.3, was positive to the 1.26 kb SmaI/EcoRI fragment from the pCH11.5E subclone (Fig. 4.4). The remaining two clones, λH1.9 and λH1.10, were negative for all probes.

λH1.1 was only positive for the probe from the left-hand end of λCH02

FIGURE 4.4

Organization of genomal DNA surrounding λCH01

The organization of histone genes in DNA overlapping that within the genomal clone λCH01 is shown. Organization was determined by analysis of overlapping λ clones, from the chicken genomal library constructed by Dodgson et al. (1979), and by analysis of a single cosmid clone (cosmid 6.3C). Regions of DNA covered by clone inserts are indicated.

λ CH10a and λCH3d were described by Sugarman et al. (1983). Analysis of remaining clones was by R. D'Andrea (D'Andrea et al., 1985).

The locations of histone genes are indicated by blocks or horizontal arrows. Directions of arrows show directions of gene transcription as determined by DNA sequencing (Wang et al., 1985; Sturm, 1985; D'Andrea, 1985; Harvey et al., 1982; Sugarman et al., 1983; see also Chapter 5). H1 genes are emphasized with black fields. The number of H4 genes within the 4 kb EcoRI/HindIII fragment of λCH10a and λCH3d has not been determined (Sugarman et al., 1983).

Detailed maps of the inserts of subclones used in this thesis, viz., pCH5.0E, pCH3.3E and pCH11.5E, are shown and locations of probe fragments used for screening H1 positive genomal clones (section 4.2.2(b)) are indicated (numbered 1 to 5). For fragment descriptions see Figure 4.6.

✸ indicates a 2.5 kb PstI fragment used for analysis of chicken genomal DNA and * indicates polymorphic HindIII and BamHI sites discussed in Section 4.2.6.

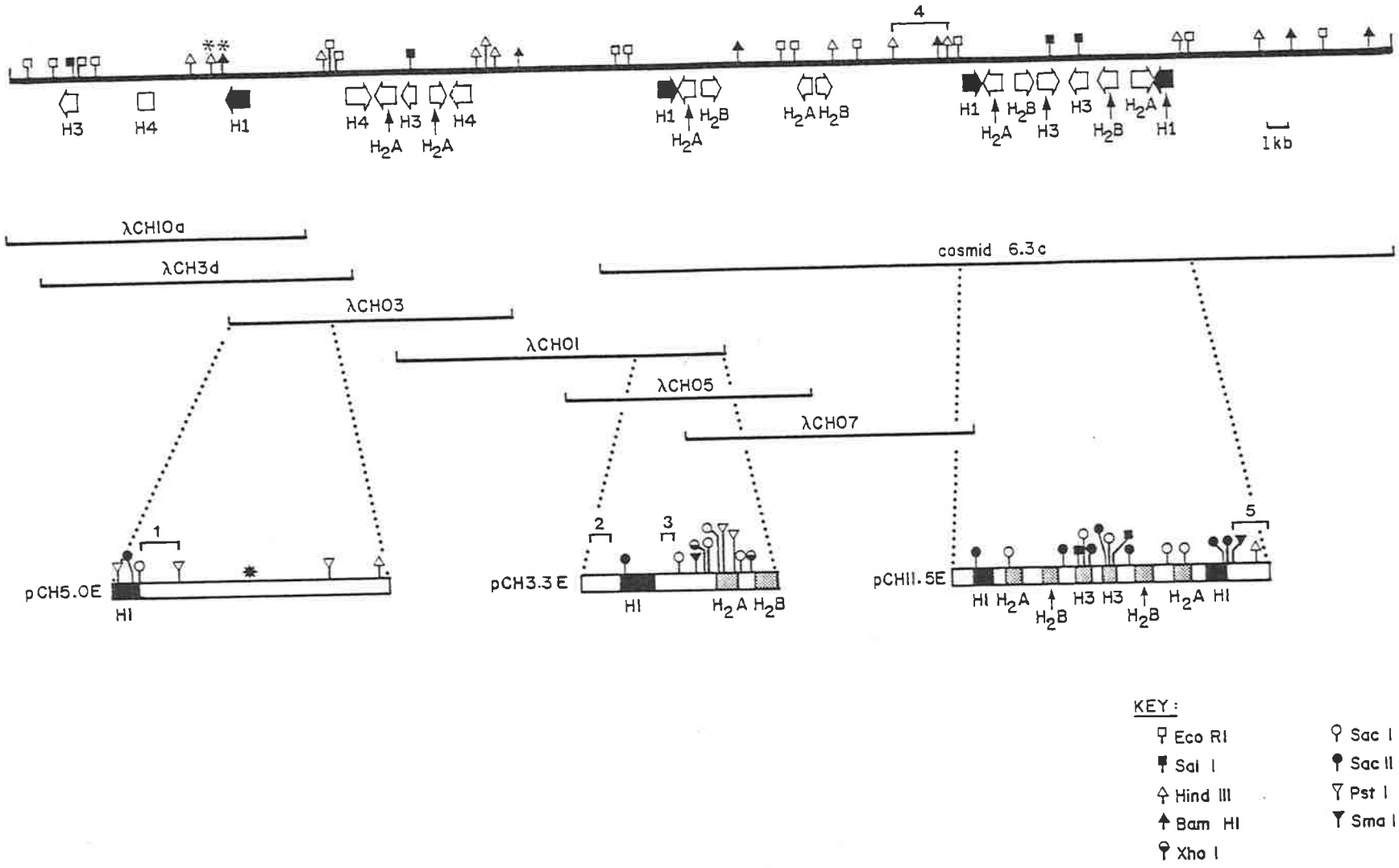

4
H3
H4
H1
H4 H3 H4
H2A H2A
H1 H2B
H2A
H2A H2B
H1 H2B H3 H2A
H2A H3 H2B H1
1kb
λCH10a
λCH3d
λCH03
λCH01
λCH05
λCH07
cosmid 6.3c
1
pCH5.0E
H1
2
3
pCH3.3E
H1
H2A H2B
5
pCH11.5E
H1 H2A H3 H3 H2A H1
H2B
H2B
KEY:
Eco RI
Sal I
Hind III
Bam HI
Xho I
Sac I
Sac II
Pst I
Sma I

FIGURE 4.5

Organization of DNA extending right of λCHO2

The organization of histone genes in DNA extending right of λCHO2 insert DNA was determined by analysis of a cosmid clone 6.1C. Data is from C. Lesnikowski (Lesnikowski, 1983; D'Andrea et al., 1985). Regions of DNA covered by λCHO2 and 6.1C inserts are shown. Gene locations are indicated by blocks or horizontal arrows. The single H1 gene is emphasized by a black field. Direction of transcription of genes (indicated by arrow directions) was determined by DNA sequencing (Harvey et al., 1981, 1982; Sturm, 1985; A. Robins and S. Clark, unpublished).

Characterization of DNA extending left of λCHO2 is discussed in Sections 4.2.3 and 4.2.4.

A map of the insert of the subclone pCH4.7E (Harvey et al., 1981; Clark, 1982), from which probes were isolated (numbered 6 and 7) for screening H1 positive genomal clones (Section 4.2.2(b); Fig. 4.6), is also given.

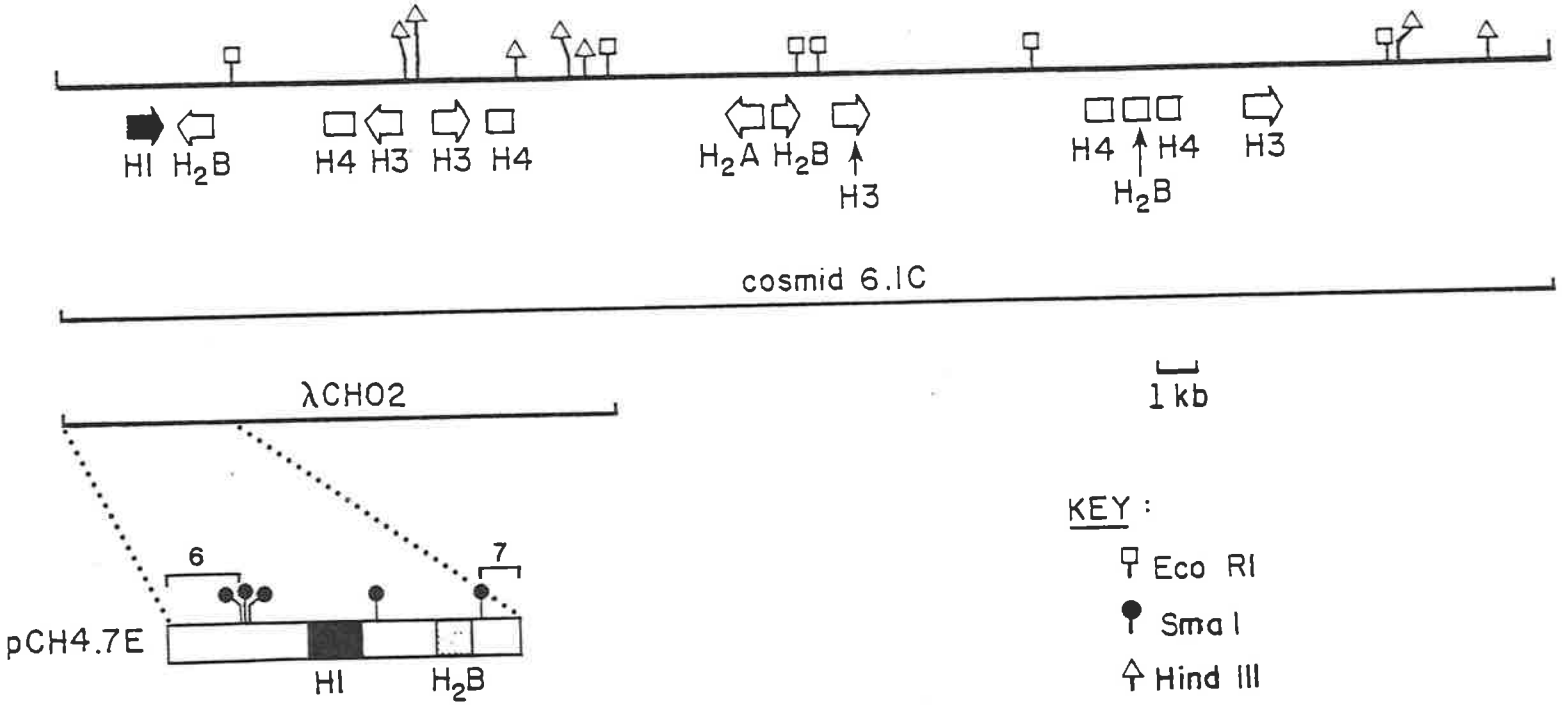

H1 H2B
H4 H3 H3 H4
H2A H2B
H3
H4 H4
H2B
H3
cosmid 6.1C
λCH02
1 kb
6
7
pCH4.7E
H1
H2B
KEY :
Eco RI
Sma I
Hind III

FIGURE 4.6

Screening Hl positive genomal clones

DNA from Hl positive genomal clones (and control DNAs) was denatured and dotted (approx. 100 ng/dot) in duplicate on to nitrocellulose filters, using the dot-blot procedure (Section 2.3.1(d)). Filters were hybridized with various ^{32}P-labelled non-coding probes (preparation described in Section 4.2.2(b)), washed and autoradiographed. Resultant autoradiograms are shown.

Hl clones, λHl.1, λHl.2, λHl.3, λHl.4, λHl.7, λHl.9 and λHl.10 are numbered respectively 1, 2, 3, 4, 7, 9 and 10. The non-coding probe fragments, numbered 1 to 7, are described below:

Probe: 1. 0.7 kb SacI/PstI fragment from pCH5.0E.
2. HaeIII fragment from pCH3.3E cloned in M13 mp93.
3. FnuDII fragment from pCH3.3E cloned in M13 mp93.
4. 2.6 kb HindIII fragment from 6.3C.
5. 1.26 kb SmaI/EcoRI fragment from pCHl1.5E.
6. 1.0 kb SmaI/EcoRI fragment from pCH4.7E.
7. 0.5 kb SmaI/EcoRI fragment from pCH4.7E.

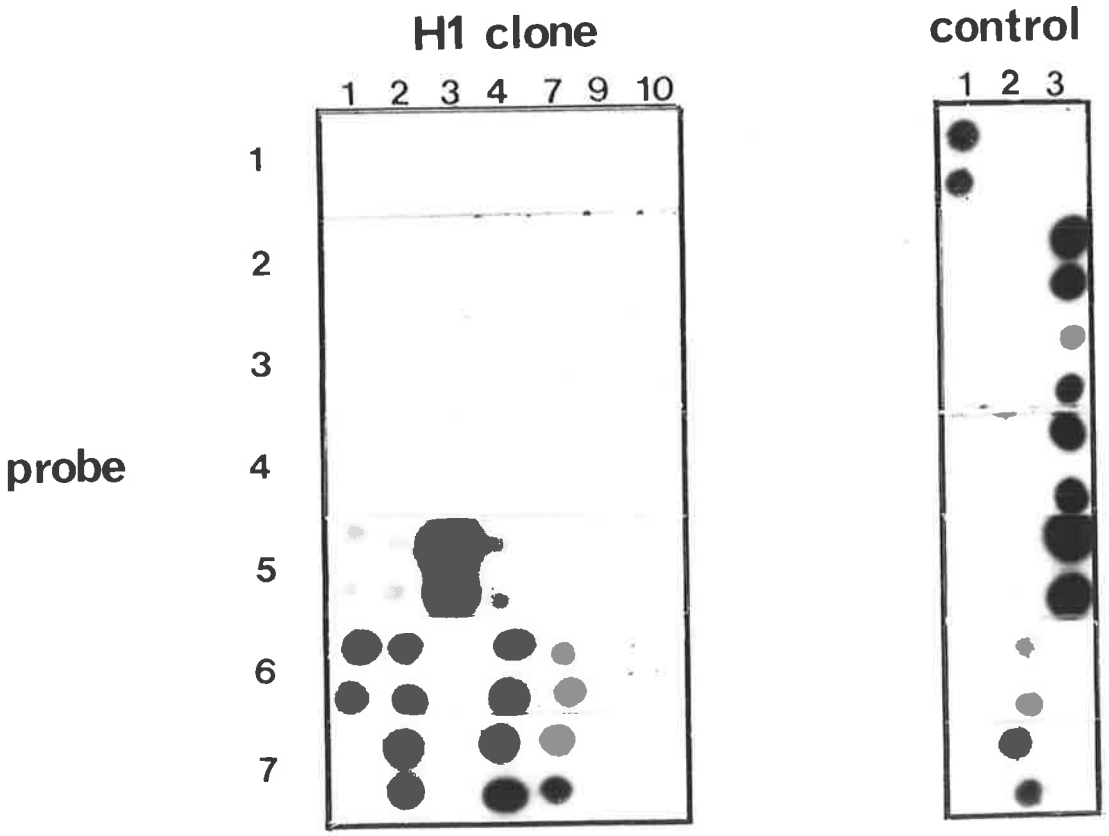
Fragment locations are indicated in Figures 4.4 and 4.5.

Control DNAs 1, 2 and 3 are respectively pCH5.0E, pCH4.7E and cosmid 6.3C.

Positive (+) and negative (-) responses to probes are as follows:

		Hl Clones:						Controls:		
		1	2	3	4	7	9	10	1	2	3
Probe:	1	-	-	-	-	-	-	-	+	-	-
	2	-	-	-	-	-	-	-	-	-	+
	3	-	-	-	-	-	-	-	-	-	+
	4	-	-	-	-	-	-	-	-	-	+
	5	-	-	+	-	-	-	-	-	-	+
	6	+	+	-	+	+	-	-	-	+	-
	7	-	+	-	+	+	-	-	-	+	-

Note: To confirm negative results, it was determined that there was hybridizable DNA on all filters by re-probing filters with appropriate positive probes (data not shown).


H1 clone
control
1 2 3 4 7 9 10
1 2 3
probe
1
2
3
4
5
6
7

(the 1 kb SmaI/EcoRI fragment; Fig. 4.5) and hence probably extended λCHO2 in a left-ward direction. As the clone was negative with the 0.5 kb SmaI/EcoRI fragment from λCHO2 it could not be determined whether λH1.1 contained the λCHO2 H1 gene (see Fig. 4.5). On the other hand, λH1.2, 4 and 7 were positive with both λCHO2 probes suggesting that they contained the λCHO2 H1 gene. As λCHO2 contains both H4 and H3 genes to the right of the 0.5 kb SmaI/EcoRI probe, the lack of cross-reaction of λH1.2 to an H3 gene probe and λH1.7 to both H3 and H4 gene probes suggested that both these clones extended λCHO2 in a left-ward direction. Cross-reaction of λH1.4 with an H2A gene probe also suggested that λH1.4 extended left of λCHO2. λCHO2 had not been shown to contain an H2A gene and the nearest H2A gene to the right of λCHO2 was approximately 3 kb away (Fig. 4.5). The λH1.4 insert (approx. 15.5 kb) was not large enough to be able to cross-react with the two SmaI/EcoRI probes and contain the above discussed H2A gene.

As λH1.1, 2, 4, and 7 all appeared to contain DNA extending λCHO2 that had not previously been mapped, it was possible that these clones contained additional H1 genes. For this reason and also as part of an effort in this laboratory to map the entire chicken histone gene cluster, these clones were further investigated (see below and Sections 4.2.3 and 4.2.4).

The λH1.3 clone, which was positive with the 1.26 kb SmaI/EcoRI fragment from the right-hand end of the pCH11.5E subclone, could have contained DNA extending beyond that already mapped, as the probe fragment is only 8.4 kb from the right-hand end of DNA characterized in Figure 4.4. However, further investigation of λH1.3 by restriction enzyme and hybridization analysis (data not shown) revealed that it did not contain any previously uncharacterized DNA. This clone, therefore, could not have contained any "new" H1 genes and hence was not investigated further.

λH1.9 and λH1.10 being negative for all probes used to screen the λH1

clones, could have contained the λCHO3 H1 gene (Fig. 4.4) if insert DNA from these clones commenced within this H1 gene and extended in a left-ward direction. This was unlikely as neither clone insert contained internal EcoRI sites (the approx. 11 kb of DNA to the left of the λCHO3 H1 gene contains four EcoRI sites, Fig. 4.4) and λH1.10 was not positive for an H3 gene probe (there is an H3 gene in DNA to the left of the λCHO3 H1 gene). Hence it appeared that λH1.9 and λH1.10 contained previously unidentified H1 genes, and were therefore further investigated.

DNA was prepared from the recombinants λH1.9 and λH1.10 and also from λH1.1, λH1.2, λH1.4 and λH1.7 (see above) by the CsCl method (Section 2.3.6(b)). Characterization of λH1.2 and λH1.7 revealed that these two clones probably resulted from DNA rearrangements that occurred during the construction of the chicken library from which they were isolated. An additional "aberrant" histone gene-containing genomal clone has also been isolated from the same library (L. Tabe, this laboratory). λH1.2 and λH1.7 will not be discussed further.

The characterization of clones that potentially overlapped λCHO2, viz. λH1.1 and λH1.4, is discussed in Sections 4.2.3 and 4.2.4. The characterization of λH1.9 and λH1.10 is discussed in Section 4.2.5.

4.2.3 Characterization of Clones Overlapping λCHO2

The H1 gene-containing genomal clones, λH1.1 and λH1.4, were analyzed by both restriction enzyme and hybridization analysis.

(a) Characterization of λH1.4

Restriction enzyme analysis

λH1.4 DNA was digested with the restriction enzymes EcoRI, HindIII,

KpnI and BamHI and with pairwise combinations of these enzymes (Section 2.3.1(a)). Digestions were phenol/chloroform extracted, ethanol precipitated and analyzed on 1% vertical agarose gels (Section 2.3.1(b); Fig. 4.7). Bacteriophage λ DNA digested with HindIII and pBR322 DNA digested with HinfI were also run on gels as markers. Restriction enzyme fragments generated solely from the Charon 4A vector sequences of λH1.4 were also used as size markers. The sizes and relative positions of fragments generated from the various digestions of λH1.4, and the final restriction map of the λH1.4 insert are shown in Figure 4.8.

Coding potential

DNA fragments from all digestions were probed with H1, H2A, H2B, H3 and H4 coding sequences, after transfer of fragments to nitrocellulose filters using the bidirectional transfer procedure (Section 2.3.1(d)). Probes were prepared as previously described in Section 4.2.2(a). Filters were treated by boiling, as described in Section 2.3.1(d), between consecutive probings. Consistent with previous hybridization analysis (Section 4.2.2(a)) λH1.4 insert sequences were detected by all five histone gene probes. From the hybridization data histone coding regions could be localized to specific insert DNA fragments (Fig. 4.9). H3 and H4 gene probes both detected the 2.7 kb HindIII and 4.5 kb EcoRI/HindIII fragments and the H2B gene probe detected the 1.4 kb KpnI/EcoRI fragment. H1 and H2A coding sequences could be localized to the 2.1 kb BamHI/KpnI and 1.7 kb BamHI/KpnI fragments respectively (Fig. 4.9). These fragments are indicated on the restriction enzyme map of the λH1.4 insert (Fig. 4.8).

The relationship of the λH1.4 clone to λCHO2 is discussed later in Section 4.2.4.

FIGURE 4.7

Restriction enzyme analysis of λH1.4

λH1.4 DNA (approx. 1 ug/digestion) was digested with the restriction enzymes EcoRI, HindIII, KpnI and BamHI alone or in pairwise combinations. Digestions were electrophoresed on 1% vertical agarose gels and DNA was visualized with UV light after ethidium bromide staining.

The figure shows λH1.4 digested with the following enzymes:

Track:
a. KpnI
b. KpnI and EcoRI
c. KpnI and HindIII
d. EcoRI
e. EcoRI and HindIII
f. HindIII
g. BamHI and EcoRI
h. BamHI
i. BamHI and HindIII
j. BamHI and KpnI

Sizes of restriction fragments are shown in kilobases. * indicates bands containing two different unresolvable DNA fragments. Horizontal arrows indicate a DNA fragment generated due to the annealing of terminal vector sequences (Maniatis et al., 1982). DNA fragments generated from the ends of vector arms (see Figure 4.8) are often of decreased intensity relative to other DNA fragments of similar size due to such annealing of vector arm termini.

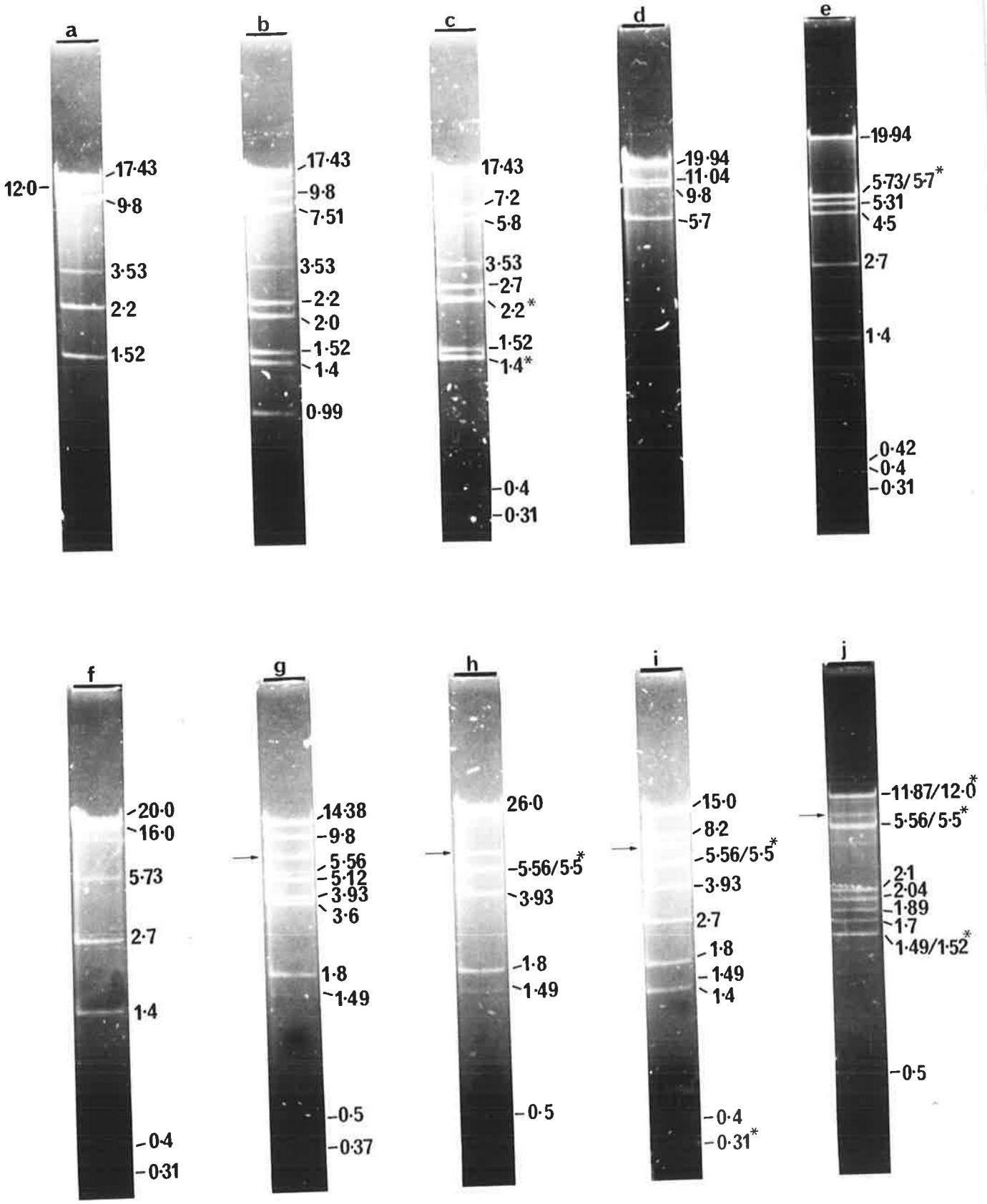

a
12·0
17·43
9·8
3·53
2·2
1·52
b
17·43
9·8
7·51
3·53
2·2
2·0
1·52
1·4
0·99
c
17·43
7·2
5·8
3·53
2·7
2·2*
1·52
1·4*
0·4
0·31
d
19·94
11·04
9·8
5·7
e
19·94
5·73/5·7*
5·31
4·5
2·7
1·4
0·42
0·4
0·31
f
20·0
16·0
5·73
2·7
1·4
0·4
0·31
g
14·38
9·8
5·56
5·12
3·93
3·6
1·8
1·49
0·5
0·37
h
26·0
5·56/5·5*
3·93
1·8
1·49
0·5
i
15·0
8·2
5·56/5·5*
3·93
2·7
1·8
1·49
1·4
0·4
0·31*
j
11·87/12·0*
5·56/5·5*
2·1
2·04
1·89
1·7
1·49/1·52*
0·5

FIGURE 4.8

Restriction enzyme map of λH1.4

The maps of restriction fragments generated from various digestions of λH1.4 (Fig. 4.7) are shown. Sizes of fragments are given in kilobases. A final map of restriction enzyme sites within the λH1.4 insert is given below individual restriction maps. E, H, B and K represent respectively EcoRI, HindIII, BamHI and KpnI restriction enzyme sites. E^L represents sites of EcoRI linkers used in clone construction (Dodgson et al., 1979). HindIII fragments marked * were ordered via alignment with λCH02 (figure 4.17). DNA fragments hybridizing to core and H1 histone gene probes are indicated with open boxes.

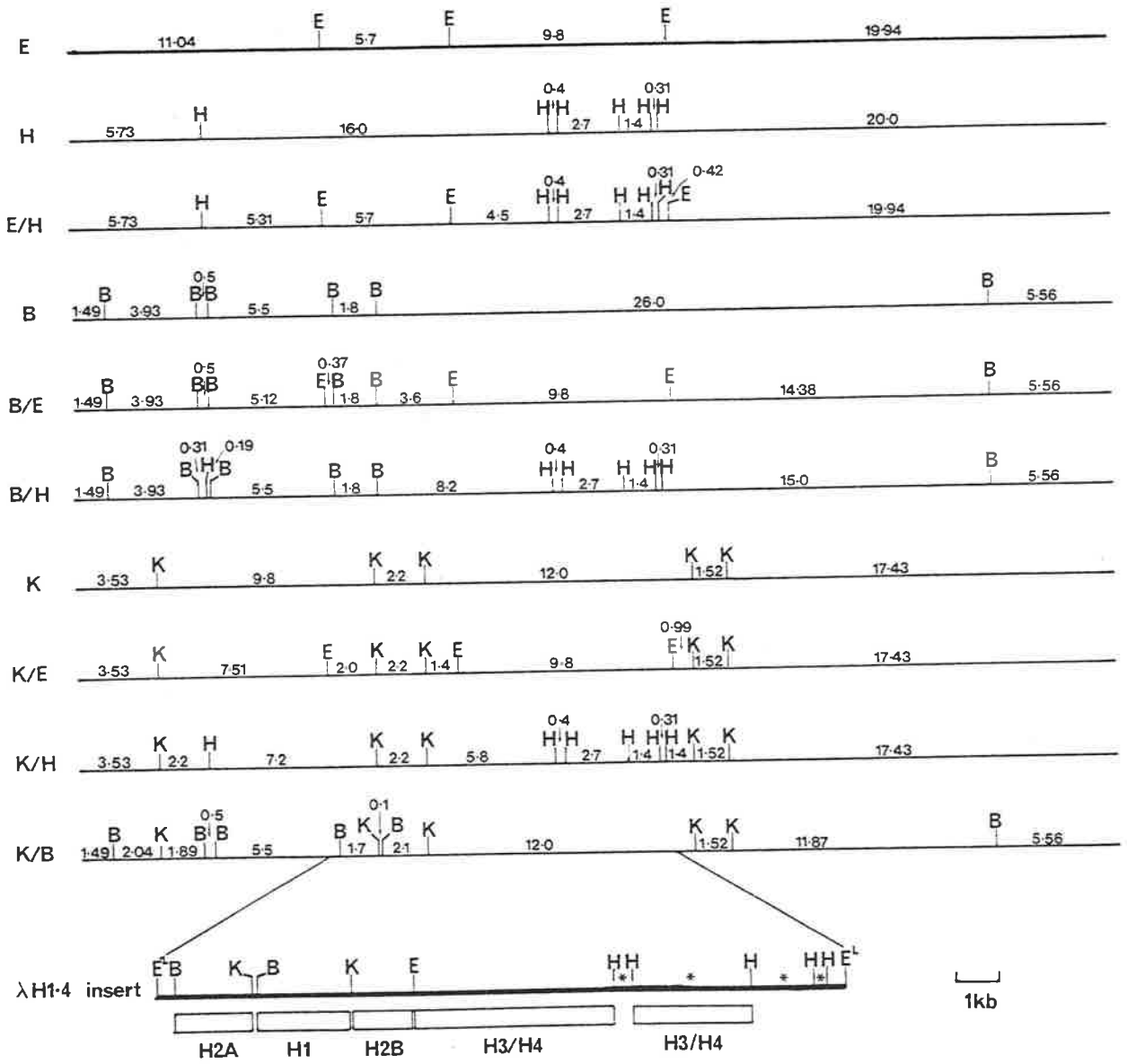
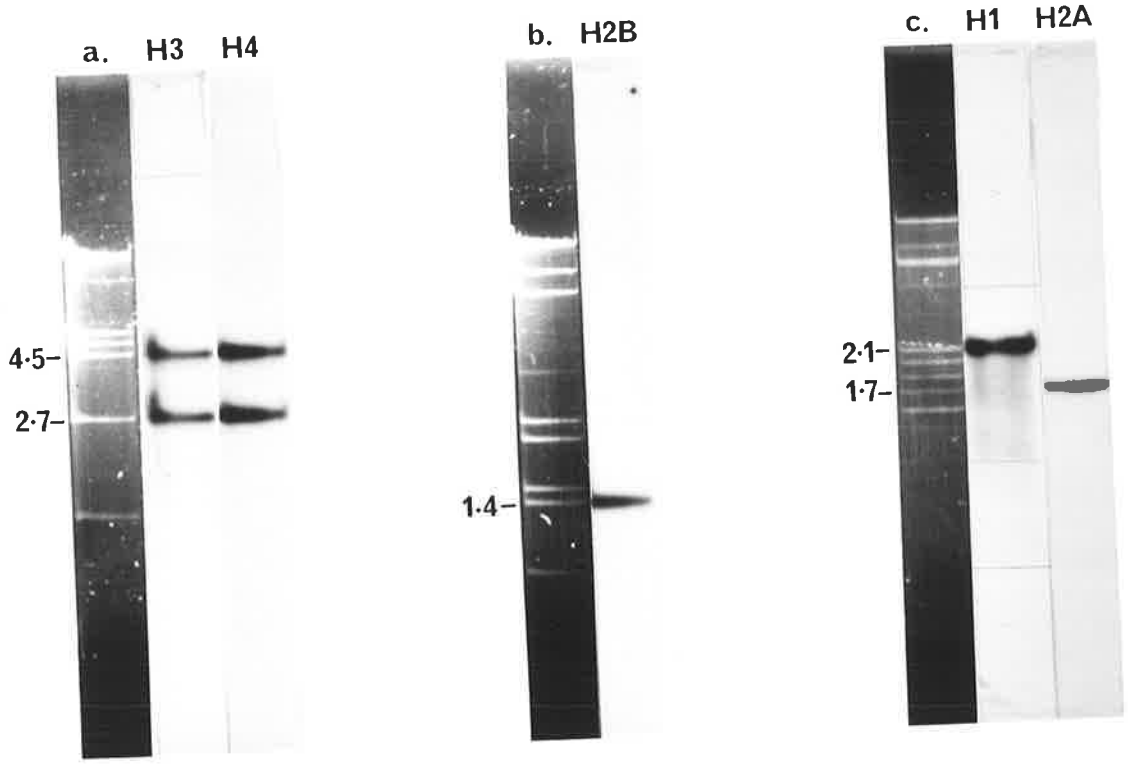

E
11·04
E
5·7
E
9·8
E
19·94
H
5·73
H
16·0
0·4
H H
2·7
H H H
0·31
1·4
20·0
E/H
5·73
H
5·31
E
5·7
E
4·5
0·4
H H
2·7
H H H
0·31
0·42
E
1·4
19·94
B
1·49
B
3·93
0·5
B B
5·5
B
1·8
B
26·0
B
5·56
B/E
1·49
B
3·93
0·5
B B
5·12
0·37
E B
1·8
B
3·6
E
9·8
E
14·38
B
5·56
B/H
1·49
B
3·93
0·31
0·19
B H B
5·5
B
1·8
B
8·2
0·4
H H
2·7
H H H
0·31
1·4
15·0
B
5·56
K
3·53
K
9·8
K
2·2
K
12·0
K K
1·52
17·43
K/E
3·53
K
7·51
E
2·0
K
2·2
K
1·4
E
9·8
0·99
E K K
1·52
17·43
K/H
3·53
K
2·2
H
7·2
K
2·2
K
5·8
0·4
H H
2·7
H H H K K
0·31
1·4 1·4 1·52
17·43
K/B
1·49
B
2·04
K
1·89
B
0·5
B
5·5
B
1·7
K
0·1
B
2·1
K
12·0
K K
1·52
11·87
B
5·56
λH1·4 insert
E B K B K E H H H H E
1kb
H2A
H1
H2B
H3/H4
H3/H4

FIGURE 4.9

Hybridization analysis of λH1.4

Digestions of λH1.4 were transferred to nitrocellulose and hybridized with histone gene probes as discussed in Section 4.2.3(a). Filters were washed in a final wash of 0.5 x SSC, 0.1% (w/v) SDS (65°C) and autoradiographed (Section 2.3.1(d)). From hybridization data histone coding regions could be localized to specific insert DNA fragments as shown.

Tracks a., b. and c. show respectively EcoRI/HindIII, KpnI/EcoRI and BamHI/KpnI digestions of λH1.4. Resultant autoradiograms, after hybridization analysis, are shown adjacent to digestions. Detected restriction fragment sizes are shown in kilobases.


a.
H3
H4
4·5–
2·7–
b.
H2B
1·4–
c.
H1
H2A
2·1–
1·7–

(b) Characterization of λH1.1

Restriction enzyme analysis

λH1.1 DNA was digested with EcoRI, HindIII, BamHI and KpnI alone or in pairwise combinations and analyzed on agarose gels (Fig. 4.10). The size and location of generated restriction fragments is shown in Figure 4.11. The location of fragments was determined solely from digestion data except in the case of the 0.2 kb EcoRI fragment. The 0.2 kb EcoRI fragment is one of two EcoRI fragments derived from the insert of λH1.1. From the digestion data it could not be determined with complete certainty whether this fragment resided at the right or left-hand end of the λH1.1 insert.

In order to locate the 0.2 kb EcoRI fragment, it was isolated from an LGT agarose gel (after digestion of λH1.1), ^{32}P-labelled by nick-translation and used to probe back to a KpnI/HindIII digestion of λH1.1 that had been transferred to nitrocellulose. If the 0.2 kb EcoRI fragment resided at the left-hand end of the λH1.1 insert it would detect the 6.1 kb HindIII fragment of the KpnI/HindIII digestion, but if it resided at the right-hand end it would detect the 2.6 kb KpnI fragment. The result of the hybridization is shown in Figure 4.12. Only the 6.1 kb HindIII fragment is detected, hence locating the small EcoRI fragment as shown in Figure 4.11. The probe was also shown to hybridize back to itself in a BamHI/EcoRI digestion of λH1.1 present on the same filter as the KpnI/HindIII digestion (Fig. 4.12). The final restriction enzyme map of the λH1.1 insert is presented in Figure 4.11.

Coding potential

DNA fragments from digestions of λH1.1 were transferred to nitrocellulose and probed with histone gene probes as described for λH1.4

FIGURE 4.10

Restriction enzyme analysis of λH1.1

λH1.1 DNA was digested with various restriction enzymes and analyzed on 1% vertical agarose gels.
λH1.1 was digested with:

Track: a. EcoRI
b. HindIII and EcoRI
c. HindIII
d. BamHI and HindIII
e. BamHI
f. BamHI and EcoRI
g. KpnI and HindIII
h. KpnI
i. KpnI and BamHI
j. KpnI and EcoRI

DNA fragment sizes are shown in kilobases. Some of the larger DNA fragments (greater than 5 kb) were sized on lower percentage gels using reduced amounts of DNA (for improved resolution - data not shown). * and arrow symbols are as for Figure 4.7. ☆ indicates bands containing three different unresolvable DNA fragments. DNA fragments generated from vector arm termini are often of reduced intensity as previously discussed (Fig. 4.7).

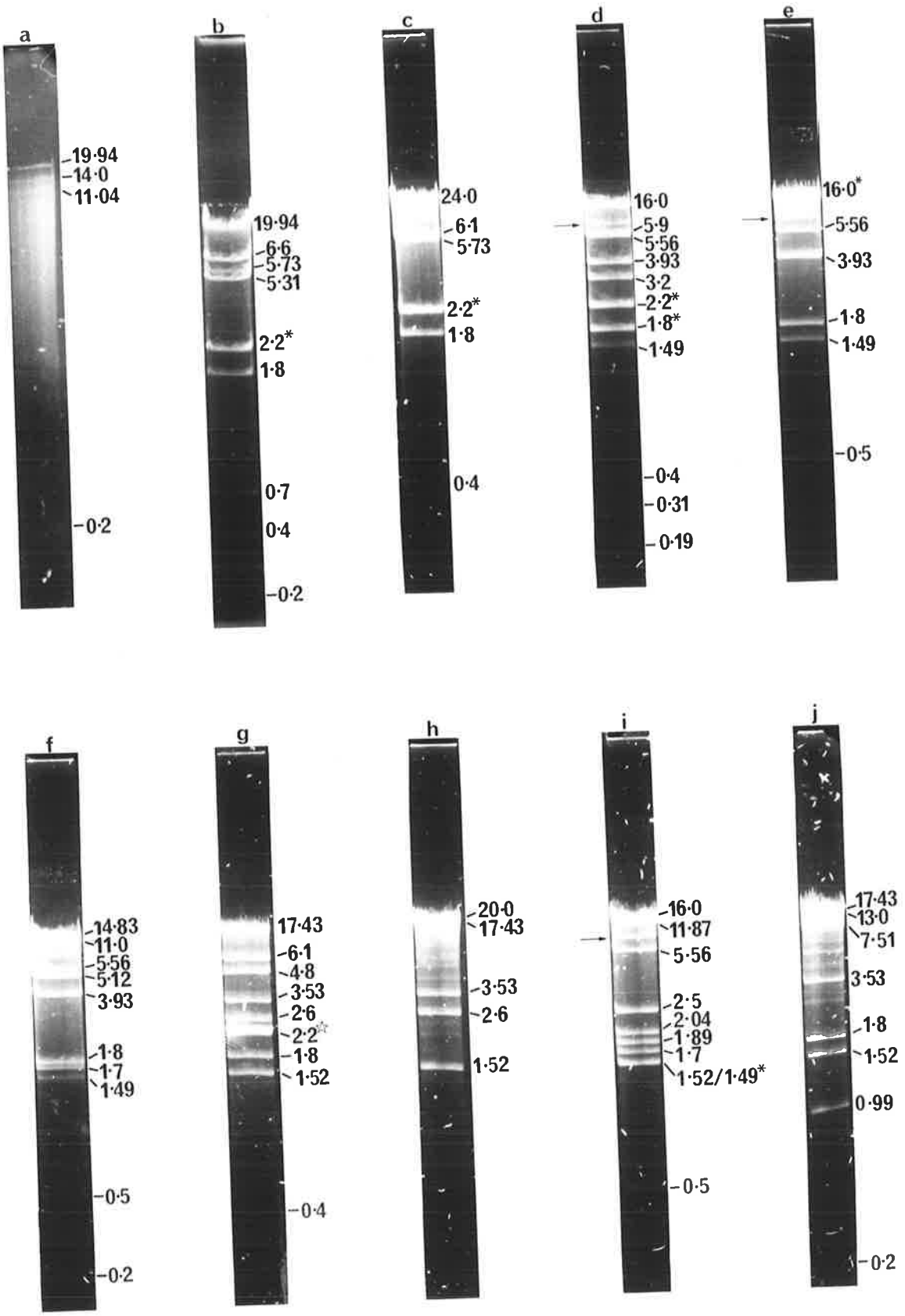

a
19·94
14·0
11·04
0·2
b
19·94
6·6
5·73
5·31
2·2*
1·8
0·7
0·4
0·2
c
24·0
6·1
5·73
2·2*
1·8
0·4
d
16·0
5·9
5·56
3·93
3·2
2·2*
1·8*
1·49
0·4
0·31
0·19
e
16·0*
5·56
3·93
1·8
1·49
0·5
f
14·83
11·0
5·56
5·12
3·93
1·8
1·7
1·49
0·5
0·2
g
17·43
6·1
4·8
3·53
2·6
2·2☆
1·8
1·52
0·4
h
20·0
17·43
3·53
2·6
1·52
i
16·0
11·87
5·56
2·5
2·04
1·89
1·7
1·52/1·49*
0·5
j
17·43
13·0
7·51
3·53
1·8
1·52
0·99
0·2

FIGURE 4.11

Restriction enzyme map of λH1.1

The final map of restriction enzyme sites in the insert of λH1.1 is shown below the maps of restriction fragments generated from individual digestions of λH1.1. Fragment sizes are given in kilobases. E, H, B and K represent EcoRI, HindIII, BamHI and KpnI sites respectively. E^L represents EcoRI linkers. HindIII fragments marked with * have not been ordered relative to each other. DNA fragments hybridizing to histone gene probes are indicated with open boxes.

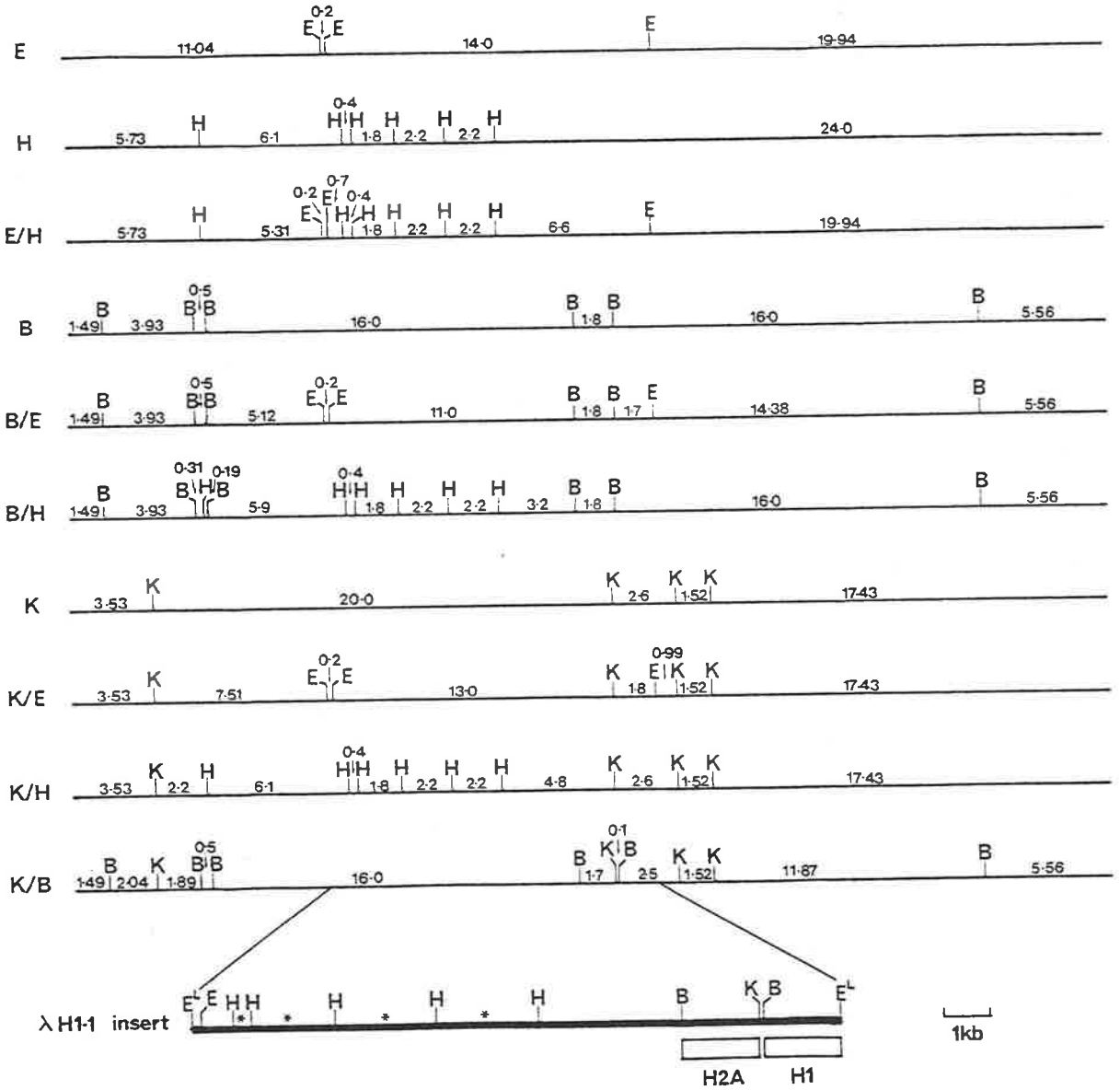
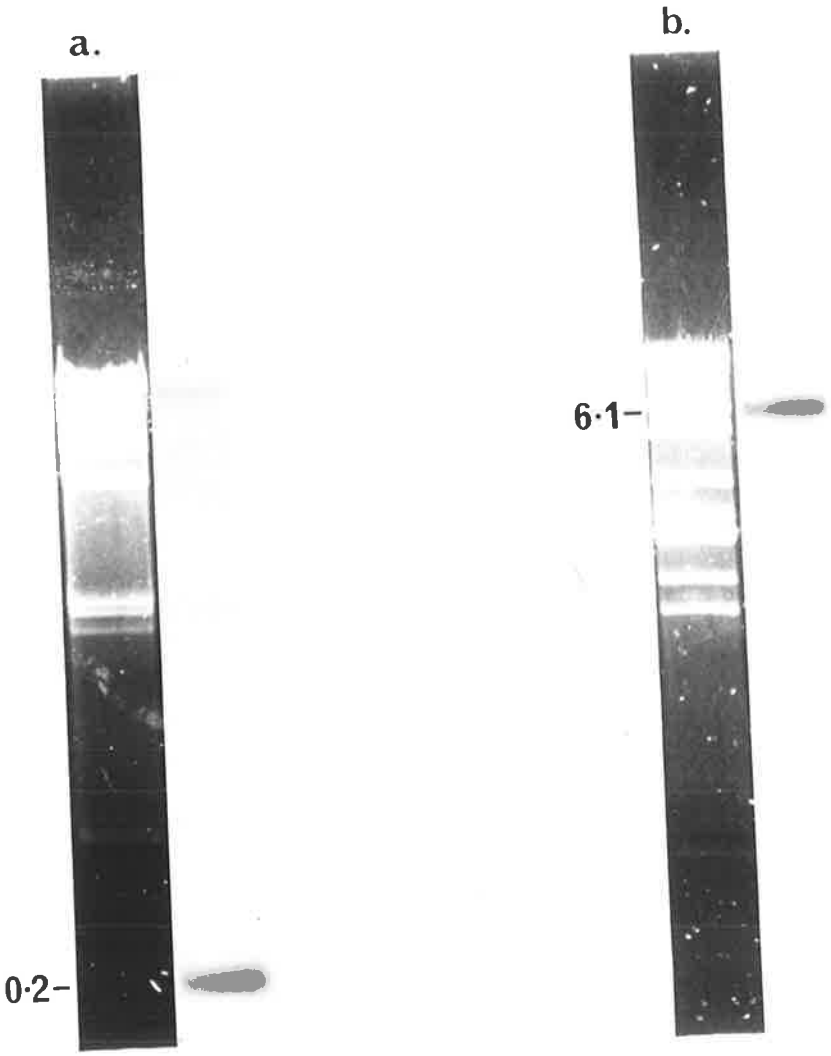

E
11·04
0·2
E E
14·0
E
19·94
H
5·73
H
6·1
0·4
H H
1·8
H
2·2
H
2·2
H
24·0
E/H
5·73
H
5·31
0·2
E
0·7
E
0·4
H H
1·8
H
2·2
H
2·2
H
6·6
E
19·94
B
1·49
B
3·93
0·5
B B
16·0
B
1·8
B
16·0
B
5·56
B/E
1·49
B
3·93
0·5
B B
5·12
0·2
E E
11·0
B
1·8
B
1·7
E
14·38
B
5·56
B/H
1·49
B
3·93
0·31
B
0·19
H B
5·9
0·4
H H
1·8
H
2·2
H
2·2
H
3·2
B
1·8
B
16·0
B
5·56
K
3·53
K
20·0
K
2·6
K
1·52
K
17·43
K/E
3·53
K
7·51
0·2
E E
13·0
0·99
K
1·8
E K
1·52
K
17·43
K/H
3·53
K
2·2
H
6·1
0·4
H H
1·8
H
2·2
H
2·2
H
4·8
K
2·6
K
1·52
K
17·43
K/B
1·49
B
2·04
K
1·89
0·5
B B
16·0
0·1
B
1·7
K B
2·5
K
1·52
K
11·87
B
5·56
λ H1-1 insert
E L E H H
H
H
H
B
K B
E L
1kb
H2A
H1

FIGURE 4.12

Location of the 0.2 kb λH1.1 EcoRI fragment

The 0.2 kb EcoRI fragment from the λH1.1 insert (Fig. 4.11) was isolated and ^{32}P-labelled as described in Section 4.2.3(b). The fragment was used to probe digestions of λH1.1. Results of hybridization enabled the location of the 0.2 kb EcoRI fragment to be determined.

Track a. is λH1.1 digested with BamHI and EcoRI and track b. is λH1.1 digested with KpnI and HindIII. Autoradiograms resulting from hybridization analysis are shown adjacent to digestions. Sizes of detected fragments are given in kilobases.


a.
0·2–
b.
6·1–

(Section 4.2.3(a)). As predicted from previous hybridization analysis (Section 4.2.2(a)) λH1.1 was only detected by H1 and H2A gene probes. H1 and H2A coding sequences could be localized respectively to a 1.7 kb BamHI/EcoRI fragment and a 1.7 kb BamHI/KpnI fragment (Fig. 4.13). These fragments are located at the right-hand end of the λH1.1 insert, leaving the remaining 10.7 kb of the insert free of histone coding sequences. The hybridizing fragments are indicated on the final restriction map in Figure 4.11.

The relationship of λH1.1, λH1.4 and λCHO2 is discussed in Section 4.2.4. Of the restriction enzymes used to map λH1.1 and λH1.4, λCHO2 had been mapped with EcoRI and HindIII but not with KpnI and BamHI. As part of the comparative analysis of the three clones, the pCH4.7E subclone from λCHO2 (see Fig. 4.5) was mapped with BamHI and KpnI. λH1.1 and λH1.4 were positive with probes from pCH4.7E (Section 4.2.2(b)).

(c) Analysis of pCH4.7E

The insert of λCHO2 consists of two EcoRI fragments. pCH4.7E was constructed by cloning one of these fragments, a 4.7 kb EcoRI fragment from the left-hand end of λCHO2 (as oriented in Fig. 4.5), into the EcoRI site of the plasmid vector pBR325 (Bolivar, 1978). The fragment was cloned in this laboratory and contains no HindIII sites (Harvey et al., 1981).

DNA prepared from pCH4.7E (as described in Section 2.3.5) was digested with HindIII, EcoRI, BamHI, KpnI and pairwise combinations of these enzymes and analyzed on agarose gels (Fig. 4.14). Fragment locations and the final restriction map of the insert are shown in Figure 4.15. Digestions were probed with histone gene probes after transfer to nitrocellulose. In agreement with previous observations (see Fig. 4.5) the insert was positive with H1 and H2B gene probes. H1 and H2B sequences can be localized to a 2.1 kb BamHI/KpnI fragment and a 1.4 kb KpnI/EcoRI fragment respectively (Figs. 4.15 and 4.16). Each hybridizing region contains a single histone

FIGURE 4.13

Hybridization analysis of λH1.1

DNA fragments from digestions of λH1.1 were transferred to nitrocellulose filters and hybridized with histone gene probes (Section 4.2.3(b)). Filters were washed and autoradiographed as previously described (Fig. 4.9). λH1.1 was detected only by H1 and H2A gene probes. Specific insert fragments, to which hybridizing sequences could be localized, are shown.

Track a. shows λH1.1 digested with BamHI and EcoRI and track b. shows λH1.1 digested with BamHI and KpnI. Autoradiograms resulting from hybridization analysis are adjacent to digestions. Sizes of detected DNA fragments are given in kilobases.

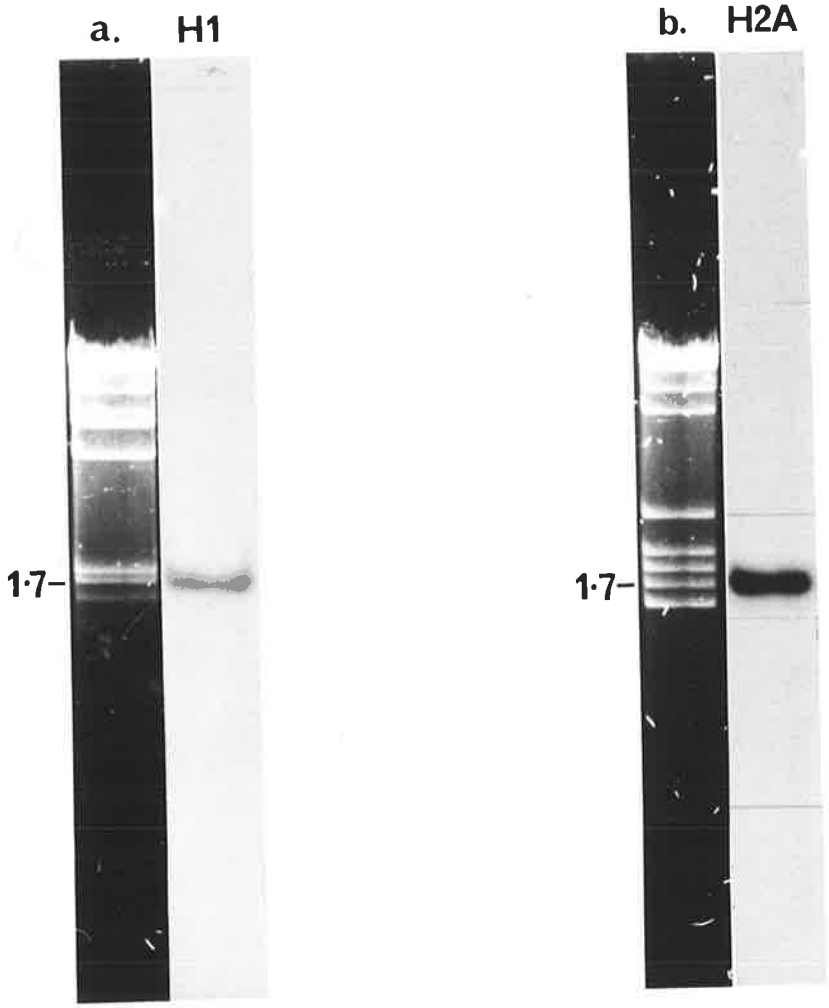

a.
H1
1·7–
b.
H2A
1·7–

FIGURE 4.14

Restriction enzyme analysis of pCH4.7E.

pCH4.7E DNA was digested with various restriction enzymes and analyzed on 1% vertical agarose gels.

Digestions shown are with:

Track:
a. KpnI
b. BamHI and KpnI
c. EcoRI and KpnI
d. BamHI and EcoRI
e. BamHI
f. HindIII and KpnI
g. HindIII and BamHI

DNA fragment sizes are given in kilobases.

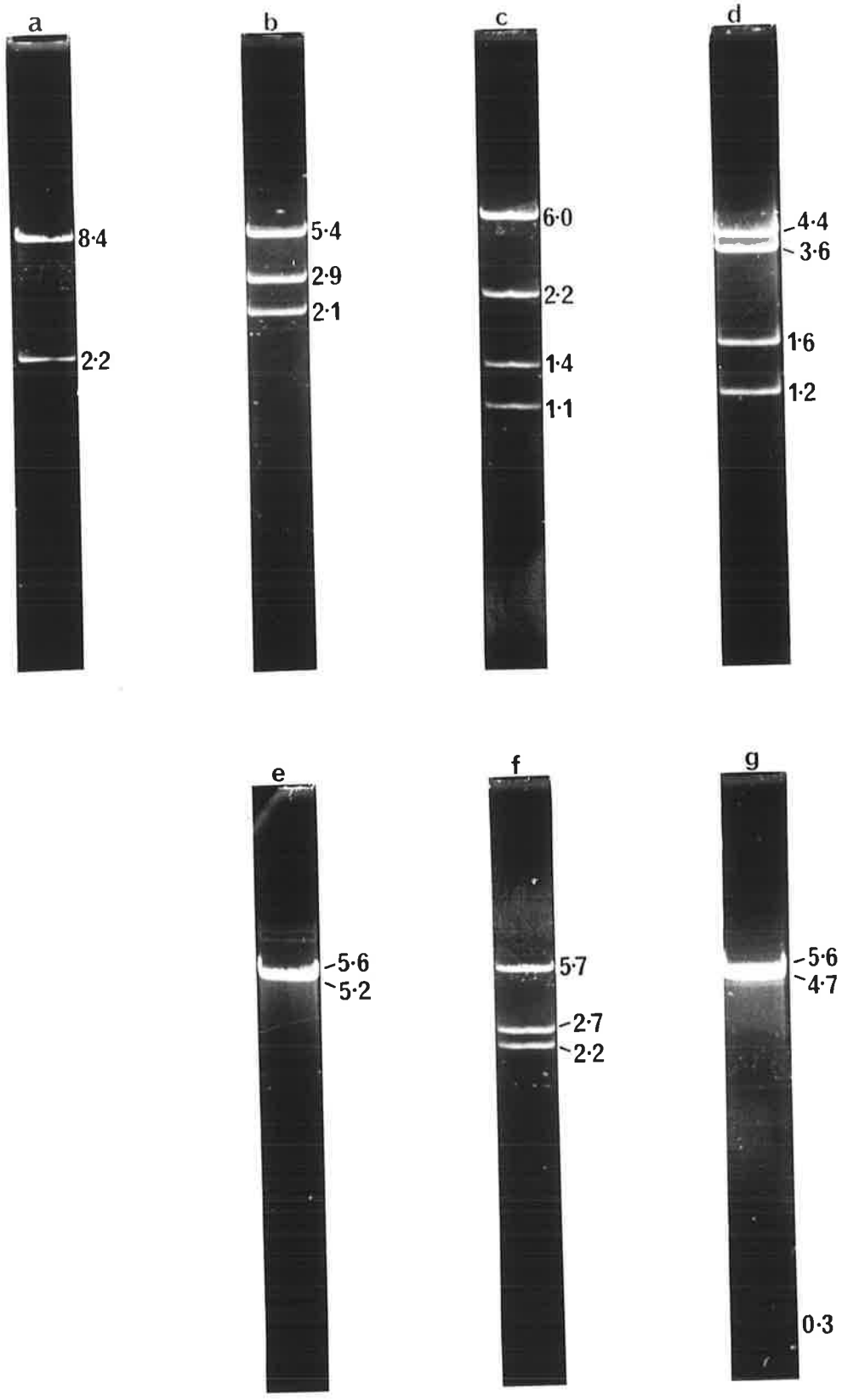

a
8·4
2·2
b
5·4
2·9
2·1
c
6·0
2·2
1·4
1·1
d
4·4
3·6
1·6
1·2
e
5·6
5·2
f
5·7
2·7
2·2
g
5·6
4·7
0·3

FIGURE 4.15

Restriction enzyme map of pCH4.7E

The final map of restriction enzyme sites in the insert of pCH4.7E is shown below the maps of restriction fragments generated from individual digestions of pCH4.7E. Fragment sizes are given in kilobases. E, H, B and K represent EcoRI, HindIII, BamHI and KpnI restriction enzyme sites respectively. The 2.1 kb BamHI/KpnI fragment and the 1.4 kb KpnI/EcoRI fragment, found to be positive with H1 and H2B gene probes respectively (Fig. 4.16), are marked with a ★. The exact locations (indicated by open boxes) and direction of transcription (indicated by horizontal arrows) of H1 and H2B genes is shown and was determined by DNA sequencing (Harvey et al., 1982; A. Robins and S. Clark, unpublished).

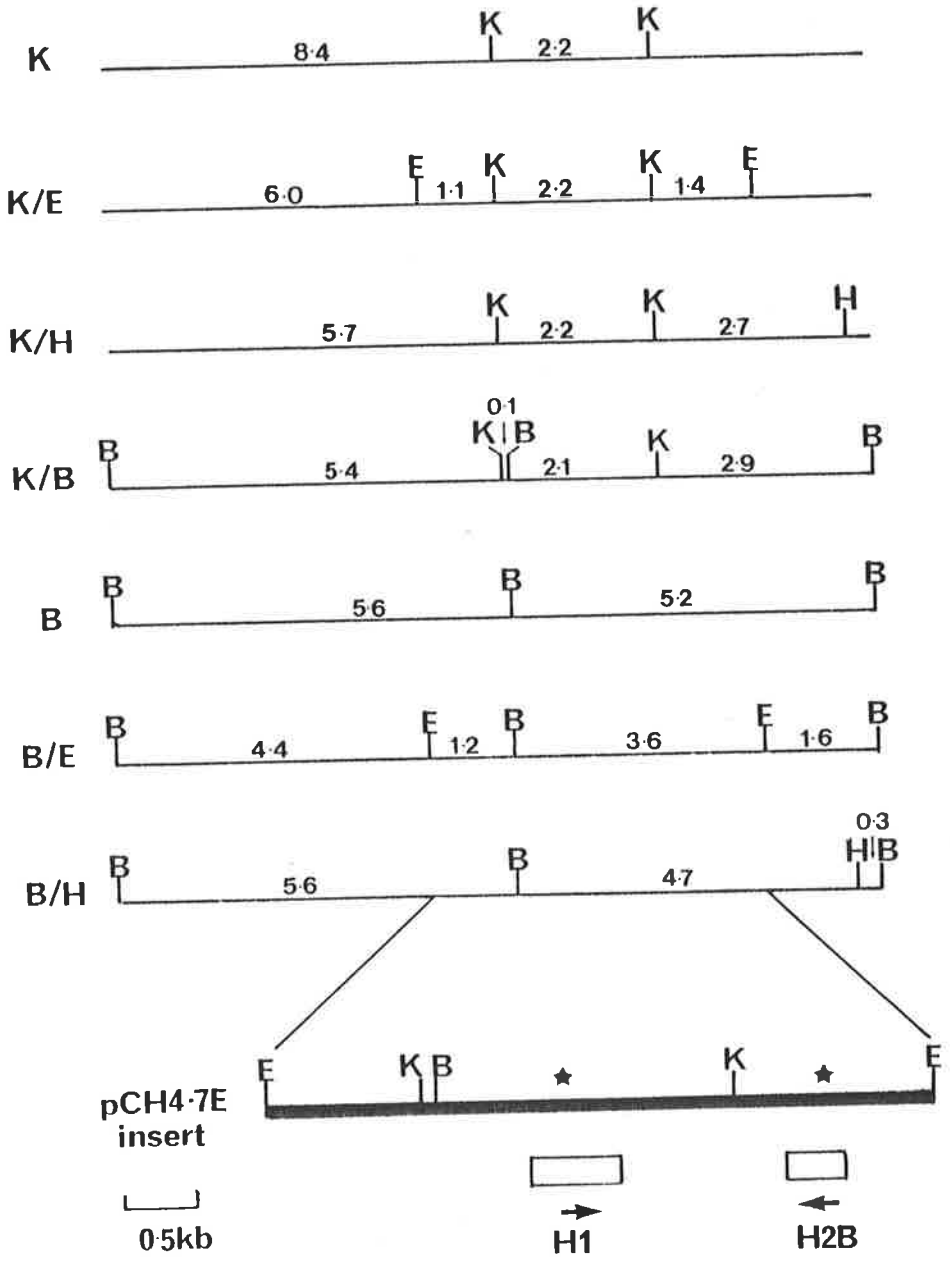
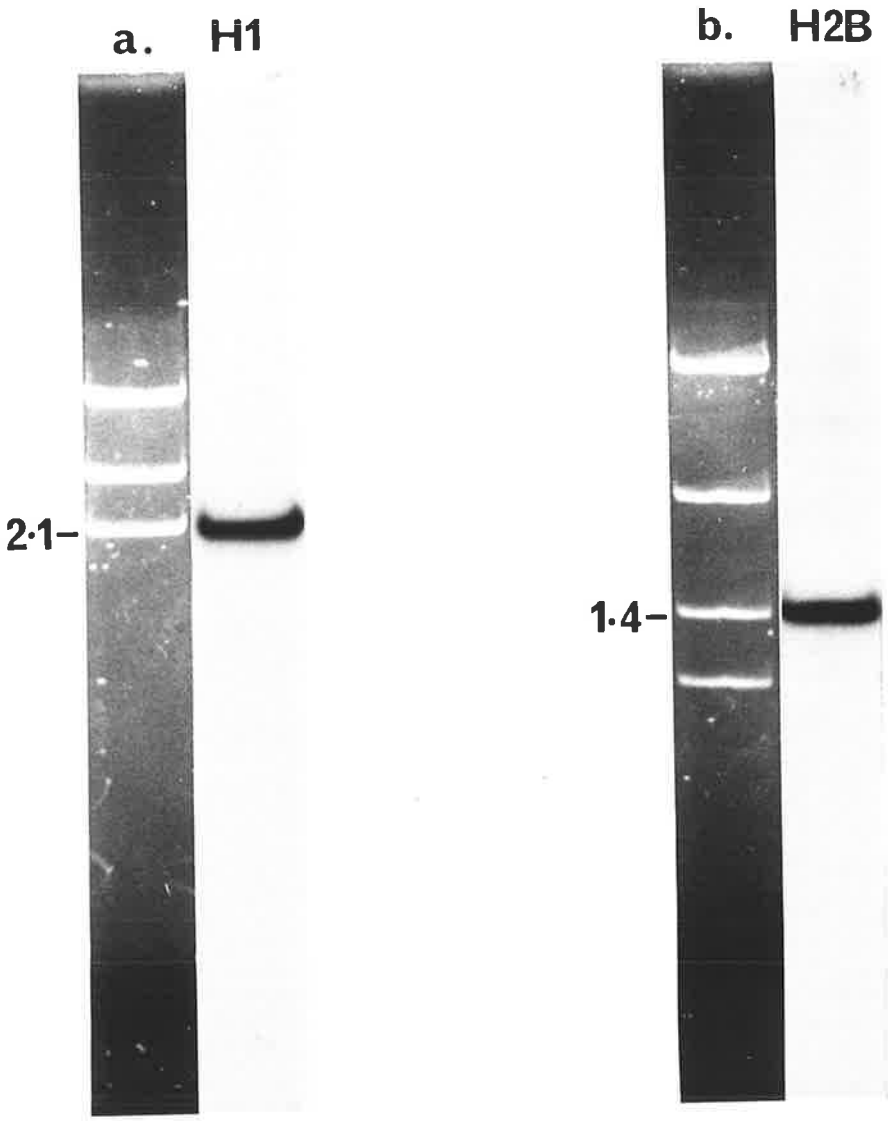

K
8·4
K
2·2
K
K/E
6·0
E
1·1
K
2·2
K
1·4
E
K/H
5·7
K
2·2
K
2·7
H
K/B
B
5·4
0·1
K
B
2·1
K
2·9
B
B
B
5·6
B
5·2
B
B/E
B
4·4
E
1·2
B
3·6
E
1·6
B
B/H
B
5·6
B
4·7
0·3
H
B
E
K
B
K
E
pCH4·7E
insert
0·5kb
H1
H2B

FIGURE 4.16

Hybridization analysis of pCH4.7E

Digestions of pCH4.7E were transferred to nitrocellulose filters and probed with histone gene probes (Section 4.2.3(c)). Filters were washed and autoradiographed as previously described (Fig. 4.9.). pCH4.7E was positive with H1 and H2B gene probes. Hybridizing fragments could be localized to specific insert fragments as shown.

Tracks a. and b. are respectively a BamHI/KpnI and a KpnI/EcoRI digestion of pCH4.7E. Autoradiograms resulting from hybridization analysis are shown adjacent to digestions.


a.
H1
2·1–
b.
H2B
1·4–

gene as previously determined by DNA sequencing (Harvey et al., 1982; A. Robins and S. Clark, unpublished). The exact location and direction of transcription of both the H2B and H1 genes is shown in Figure 4.15. From the hybridization data it can therefore be determined that the left-hand end of the pCH4.7E insert, as shown in Figure 4.15, represents the left-hand end of the λCHO2 insert.

In addition to the above hybridization results, digestions of pCH4.7E on nitrocellulose filters hybridized very weakly to the H2A gene probe, in comparison to the degree of hybridization of digestions of λH1.4 (similar amounts of DNA as used for pCH4.7E) to the H2A gene probe, on the same filters (data not shown). Such weak hybridization of pCH4.7E with the same H2A probe has previously been reported by S. Clark (1982, this laboratory), but DNA sequencing did not reveal any H2A coding sequences (Clark, 1982; A. Robins and S. Clark, unpublished).

4.2.4 Alignment of λCHO2, λH1.1 and λH1.4

Comparison of restriction enzyme and hybridization data allowed the inserts of λCHO2, λH1.1 and λH1.4 to be aligned as shown in Figure 4.17. Both the EcoRI/HindIII restriction map of the λCHO2 insert and the EcoRI/KpnI/BamHI restriction map of the pCH4.7E subclone are shown. All genes in λCHO2 have been positively identified by DNA sequencing (Harvey et al., 1982, 1981; A. Robins and S. Clark, unpublished). The location and direction of transcription of these genes (where determined) are shown.

From the data presented in Figure 4.17 it can be seen that the right-hand end of the λH1.4 insert lines up with the right-hand end of the λCHO2 insert, but that the left-hand end extends past that of λCHO2 by approximately 1 kb. The alignment of λCHO2 and λH1.4 is consistent with the detection of λH1.4 with probes derived from both the right-hand and left-hand ends of the pCH4.7E insert (Section 4.2.2(b); Fig. 4.5). λH1.4 appears to cover a similar region of genomal DNA as a genomal clone, named

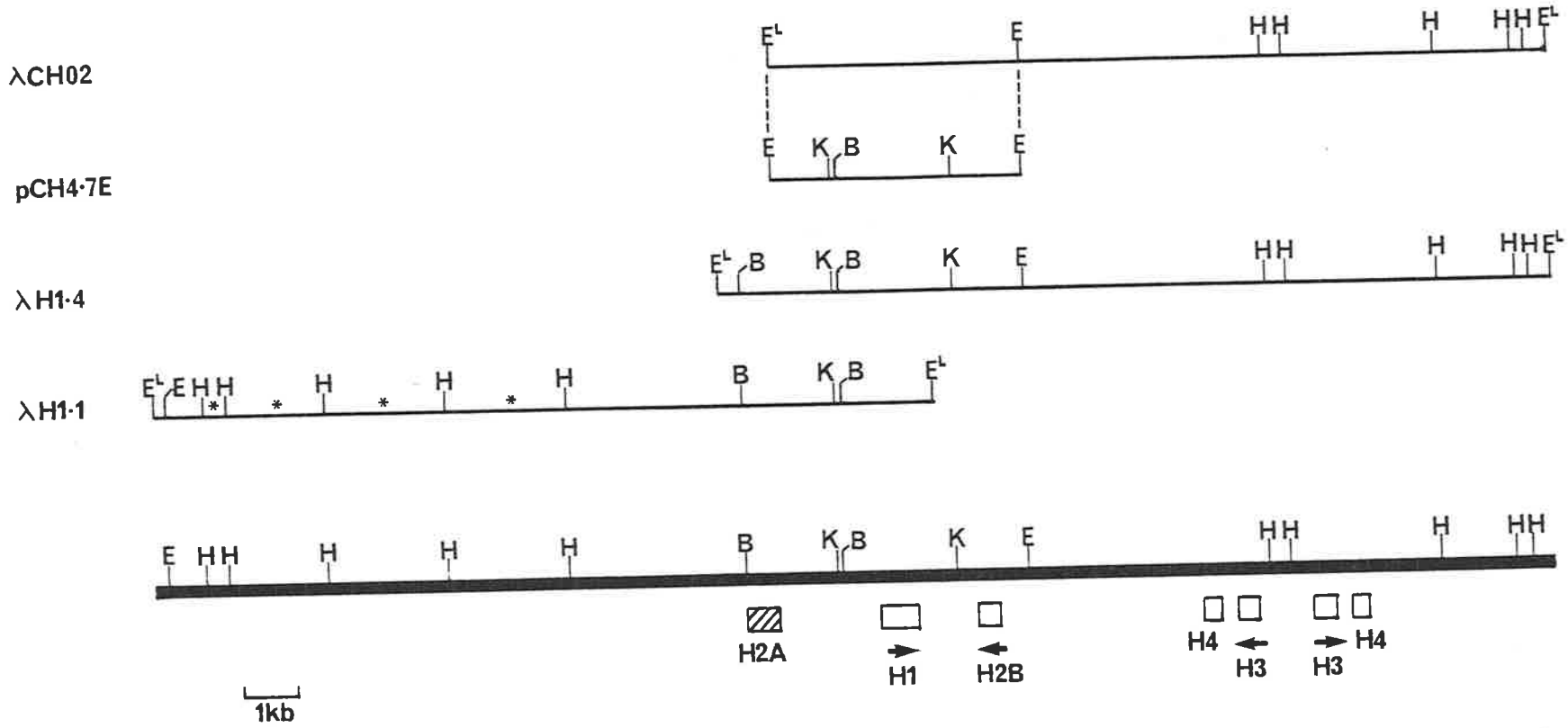
FIGURE 4.17

Alignment of λCH02, λH1.4 and λH1.1

The overall restriction map of the region of genomal DNA covered by inserts of λCH02, λH1.4 and λH1.1 is shown below the maps of individual clones. Data for the EcoRI/HindIII map of the λCH02 insert is from S. Clark (1982)[a]. Data for restriction maps of pCH4.7E (subclone of λCH02), λH1.4 and λH1.1 inserts are taken from Figures 4.15, 4.8 and 4.11. E, H, B and K represent EcoRI, HindIII, BamHI and KpnI restriction enzyme sites respectively. E^L represents EcoRI linkers. HindIII bands in λH1.1 marked with * have not been ordered relative to each other.

The exact locations of H1, H2B, H3 and H4 genes (indicated with open boxes) have been determined by DNA sequencing (Harvey et al., 1982, 1981; A. Robins and S. Clark, unpublished). Direction of transcription of genes, where determined, is indicated by horizontal arrows. Sequences hybridizing to the H2A gene probe can be localized to a 0.6 kb region of DNA (Section 4.2.4), indicated by the striped box. The exact location of the H2A coding sequences within the 0.6 kb region has not been determined.

[a]Note: Data for λCH02 was originally published by Harvey et al. (1981). Some of this data was later extended and modified by S. Clark (1982).


λCH02
pCH4·7E
λH1·4
λH1·1
H2A
H1
H2B
H4
H3
H3
H4
1kb

λCH1a, isolated by Sugarman et al. (1983) from the same library as clones described here.

In contrast to λH1.4, λH1.1 extends λCHO2 by 11.3 kb. The right-hand end of the λH1.1 insert is located 1.8 kb left of the right-hand end of the pCH4.7E insert. This is consistent with λH1.1 being positive with the probe from the left-hand end of pCH4.7E (Section 4.2.2(b)) but not with the probe from the right.

As pCH4.7E contains no H2A genes, the putative H2A gene in λH1.1 and λH1.4 (contained within a 1.7 kb BamHI/KpnI fragment) must be located in the 0.6 kb of DNA between the left-hand end of the λCHO2 insert and the BamHI end of the 1.7 kb BamHI/KpnI fragment. This region probably only contains a single H2A gene (H2A coding regions are approx. 400 bp). In the DNA left of the region positive with the H2A probe no further histone-coding regions were detected.

The overall map of the region of chicken DNA covered by the three clones, λCHO2, λH1.1 and λH1.4, is shown below individual restriction maps in Figure 4.17.

To confirm that the DNA to the left of the λCHO2 insert in λH1.1/λH1.4 was a true extension of λCHO2, the 1.1 kb EcoRI/KpnI fragment from the left-hand end of λCHO2 insert sequences was isolated from LGT agarose, ^{32}P-labelled by nick-translation and used to probe EcoRI and HindIII digested chicken genomal DNA that had been fractionated on a 0.8% agarose gel and transferred to nitrocellulose. Genomal DNA was a gift from R. Harvey, and was prepared from chicken erythrocytes using the method of Gross-Bellard et al. (1973). From the data presented in Figure 4.17 it was predicted that the probe would detect a 15.8 kb EcoRI and a 13 kb HindIII fragment in the chicken genome. Fragments of these approximate sizes were in fact detected (Fig. 4.18).

No previously unidentified H1 genes were detected in the DNA characterized above but the analysis allowed extension of our data on histone coding DNA. The overall map of DNA overlapping that within λCHO2, in both a left-ward direction (DNA characterized here) and a right-ward

FIGURE 4.18

Analysis of chicken genomal DNA

Chicken genomal DNA (gift from R. Harvey) was digested (8 ug/digestion) with either EcoRI or HindIII and fractionated on a 0.8% vertical agarose gel with HindIII digested λ DNA as size markers. DNA was transferred to a single nitrocellulose filter and hybridized (Section 2.3.1(d)) with the 1.1 kb EcoRI/KpnI fragment from pCH4.7E (prepared and ^{32}P-labelled as described in Section 4.2.4). The filter was washed (final wash in 0.1 x SSC, 0.1% (w/v) SDS at 65°C) and autoradiographed as previously described (Section 2.3.1(d)). The resultant autoradiogram is shown. E and H are EcoRI and HindIII digested DNAs respectively. Sizes of marker DNA fragments are in kilobases.

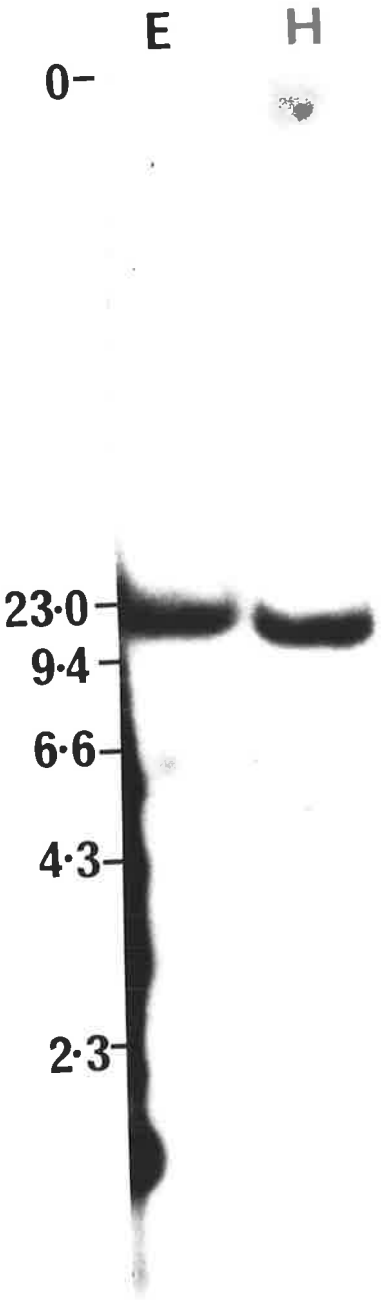
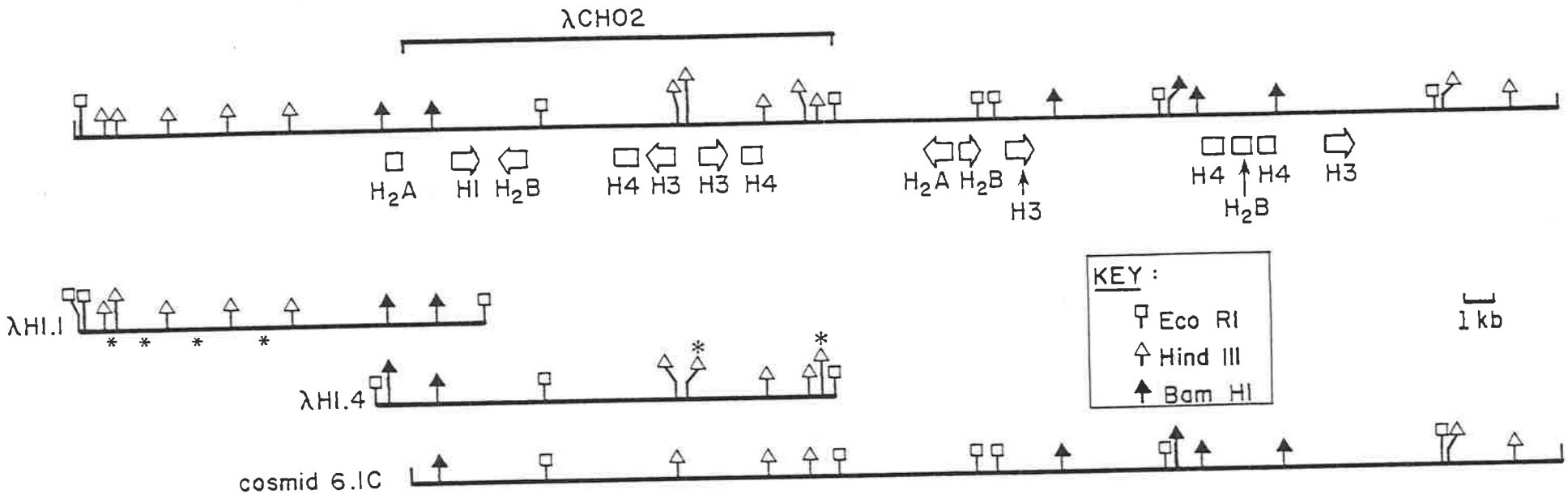

E
H
0-
23·0
9·4
6·6
4·3
2·3

FIGURE 4.19

Organization of DNA surrounding λCHO2

The top line shows the overall organization of genomal DNA surrounding that within λCHO2. The relative location of λCHO2 insert DNA is indicated. Maps of the inserts of λ clones (λH1.1 and λH1.4, Figs. 4.8 and 4.11) and of a single cosmid clone (6.1C, mapped by C. Lesnikowski, 1983), which extend λCHO2 insert DNA in both a left-ward and right-ward direction, are given below the top map. The small asterisks mark HindIII fragments that have not yet been ordered and large asterisks mark polymorphic HindIII sites discussed in Section 4.2.4.

Histone gene locations are indicated by horizontal arrows or blocks. Arrow directions indicate direction of gene transcription (determined by DNA sequencing; Harvey et al., 1981, 1982; Sturm, 1985; A. Robins and S. Clark, unpublished).


λCH02
H2A
H1
H2B
H4
H3
H3
H4
H2A
H2B
H3
H4
H4
H2B
H3
λH1.1
λH1.4
cosmid 6.1C
KEY :
Eco RI
Hind III
Bam HI
1 kb

direction (mapped by C. Lesnikowski, 1983), is presented in Figure 4.19. This data has also been presented by D'Andrea et al. (1985). The DNA mapped by C. Lesnikowski was derived from a cosmid clone (6.1C) isolated from a library prepared by R. D'Andrea and L. Tabe (D'Andrea et al., 1985). 6.1C DNA lacks two HindIII sites which are present in λH1.4/λCHO2 DNA as shown in Figure 4.19. As the cosmid and lambda libraries were prepared from DNA of different individuals, these sites represent polymorphic restriction enzyme sites. Another such site was identified in comparison of the cosmid clone (6.3C) and genomal clones overlapping λCHO1 (D'Andrea, 1985).

The characterization of the genomal clones λH1.10 and λH1.9 is discussed below. As discussed in Section 4.2.2(b) these two clones potentially contained previously unidentified H1 genes.

4.2.5 Characterization of λH1.9 and λH1.10

(a) Restriction enzyme and hybridization analysis

DNA prepared from λH1.9 and λH1.10 was digested with the restriction enzymes EcoRI, KpnI, HindIII and BamHI and analyzed on 1% agarose gels (figs. 4.20 and 4.21). Restriction maps for these digestions and the final maps of the clone inserts are shown in Figures 4.22 and 4.23.

All digestions were probed with the five histone gene probes. As expected (see Section 4.2.2(a)) λH1.9 was positive with H1, H4 and H3 gene probes (Fig. 4.24) and λH1.10 was positive with H2A, H2B, H1 and H4 gene probes (Fig. 4.25). H1, H4 and H3 sequences of λH1.9 were localized within the 0.7 kb EcoRI/HindIII, the 4.15 kb BamHI/KpnI and the 0.55 kb KpnI/EcoRI insert fragments respectively. H2A and H2B sequences of λH1.10 were located within the 1.2 kb EcoRI/HindIII insert fragment while H1 and H4 sequences were located respectively within the 2.1 kb HindIII and the 4.15 kb BamHI/KpnI fragment. Hybridizing fragments are indicated on restriction

FIGURE 4.20

Restriction enzyme analysis of λH1.9

DNA from λH1.9 was digested with restriction enzymes and analyzed on 1% vertical agarose gels.

The figure shows λH1.9 digested with:

Track:
a. BamHI
b. BamHI and EcoRI
c. BamHI and HindIII
d. EcoRI
e. HindIII
f. EcoRI and HindIII
g. KpnI and EcoRI
h. KpnI and HindIII
i. KpnI and BamHI
j. KpnI

Sizes of DNA fragments are in kilobases. Arrows and * mark bands as in Figure 4.7. DNA fragments from vector termini are often of decreased relative intensity as previously discussed (Fig. 4.7).

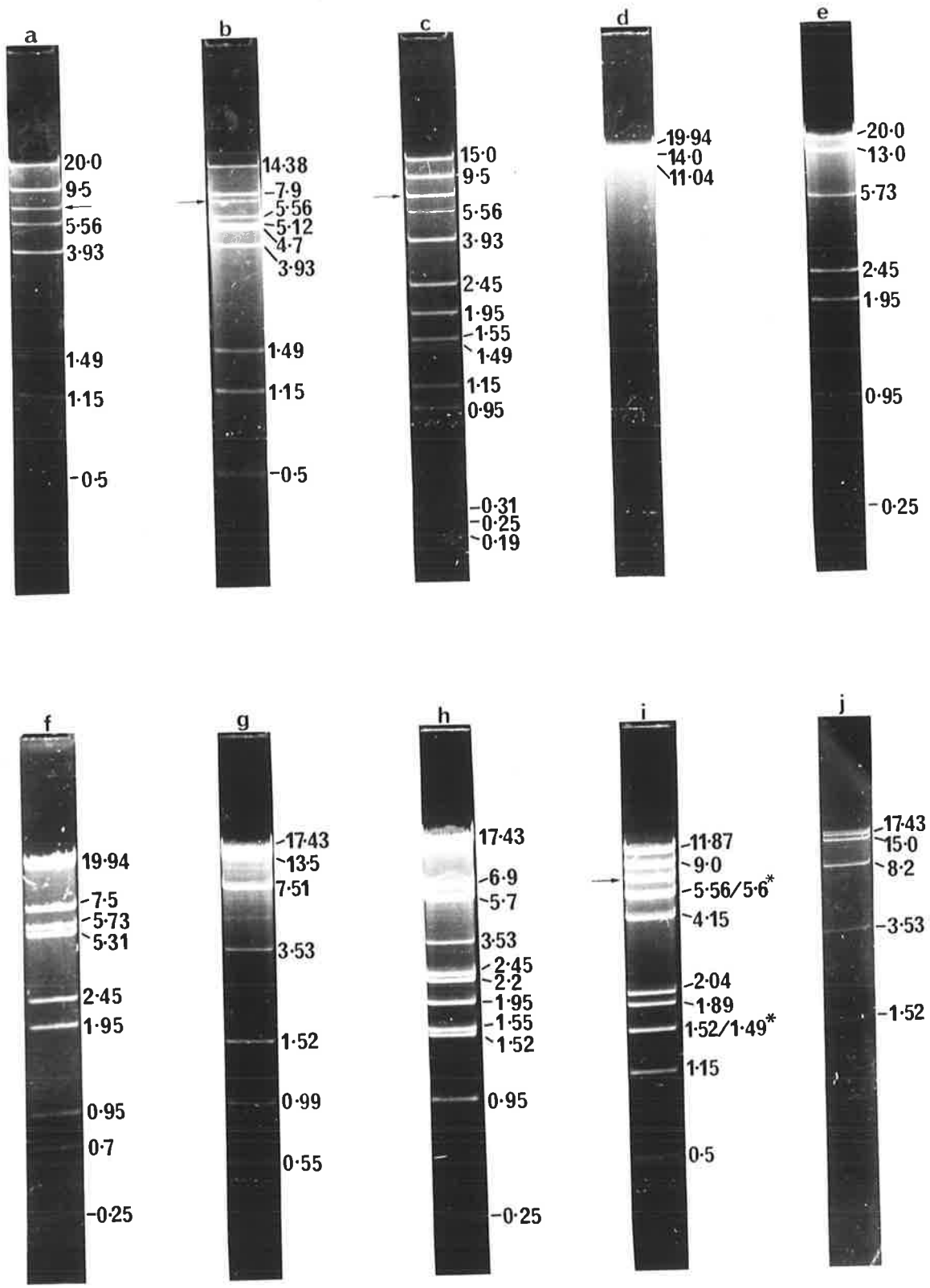

a
20·0
9·5
5·56
3·93
1·49
1·15
0·5
b
14·38
7·9
5·56
5·12
4·7
3·93
1·49
1·15
0·5
c
15·0
9·5
5·56
3·93
2·45
1·95
1·55
1·49
1·15
0·95
0·31
0·25
0·19
d
19·94
14·0
11·04
e
20·0
13·0
5·73
2·45
1·95
0·95
0·25
f
19·94
7·5
5·73
5·31
2·45
1·95
0·95
0·7
0·25
g
17·43
13·5
7·51
3·53
1·52
0·99
0·55
h
17·43
6·9
5·7
3·53
2·45
2·2
1·95
1·55
1·52
0·95
0·25
i
11·87
9·0
5·56/5·6*
4·15
2·04
1·89
1·52/1·49*
1·15
0·5
j
17·43
15·0
8·2
3·53
1·52

FIGURE 4.21

Restriction enzyme analysis of λH1.10

λH1.10 DNA was digested with various restriction enzymes and analyzed on 1% vertical agarose gels.

The figure shows λH1.10 digested with:

Track:
a. BamHI
b. BamHI and EcoRI
c. BamHI and HindIII
d. EcoRI
e. HindIII
f. EcoRI and HindIII
g. KpnI and EcoRI
h. KpnI and HindIII
i. KpnI and BamHI
j. KpnI

Sizes of DNA fragments are given in kilobases. Arrows and * are as for Figure 4.7. DNA fragments from vector termini are often of decreased relative intensity as previously discussed (Fig. 4.7).

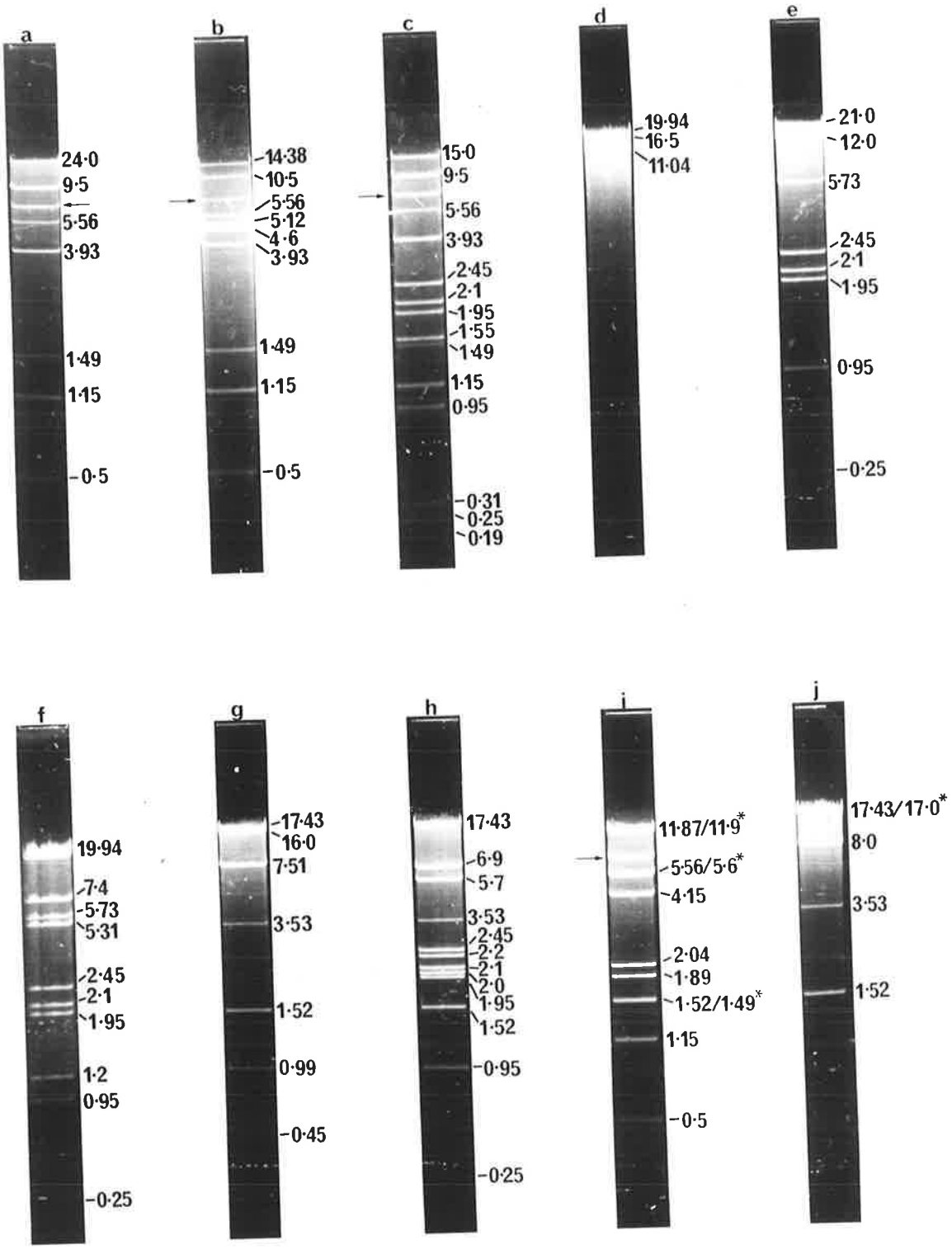

a
24·0
9·5
5·56
3·93
1·49
1·15
0·5
b
14·38
10·5
5·56
5·12
4·6
3·93
1·49
1·15
0·5
c
15·0
9·5
5·56
3·93
2·45
2·1
1·95
1·55
1·49
1·15
0·95
0·31
0·25
0·19
d
19·94
16·5
11·04
e
21·0
12·0
5·73
2·45
2·1
1·95
0·95
0·25
f
19·94
7·4
5·73
5·31
2·45
2·1
1·95
1·2
0·95
0·25
g
17·43
16·0
7·51
3·53
1·52
0·99
0·45
h
17·43
6·9
5·7
3·53
2·45
2·2
2·1
2·0
1·95
1·52
0·95
0·25
i
11·87/11·9*
5·56/5·6*
4·15
2·04
1·89
1·52/1·49*
1·15
0·5
j
17·43/17·0*
8·0
3·53
1·52

FIGURE 4.22

Restriction enzyme map of λH1.9

The final restriction enzyme map of the λH1.9 insert is shown below restriction maps of individual digestions of λH1.9. Sizes of DNA fragments are in kilobases. E, H, K and B represent EcoRI, HindIII, KpnI and BamHI restriction sites respectively. E^L represents EcoRI linkers. HindIII fragments marked with an * have not been ordered relative to each other. Fragments hybridizing to histone gene probes (Fig. 4.24) are indicated with blocks. Precise locations of H1 and H3 genes, as determined by DNA sequencing (Section 4.2.5(b)), are shown in Figure 4.26.

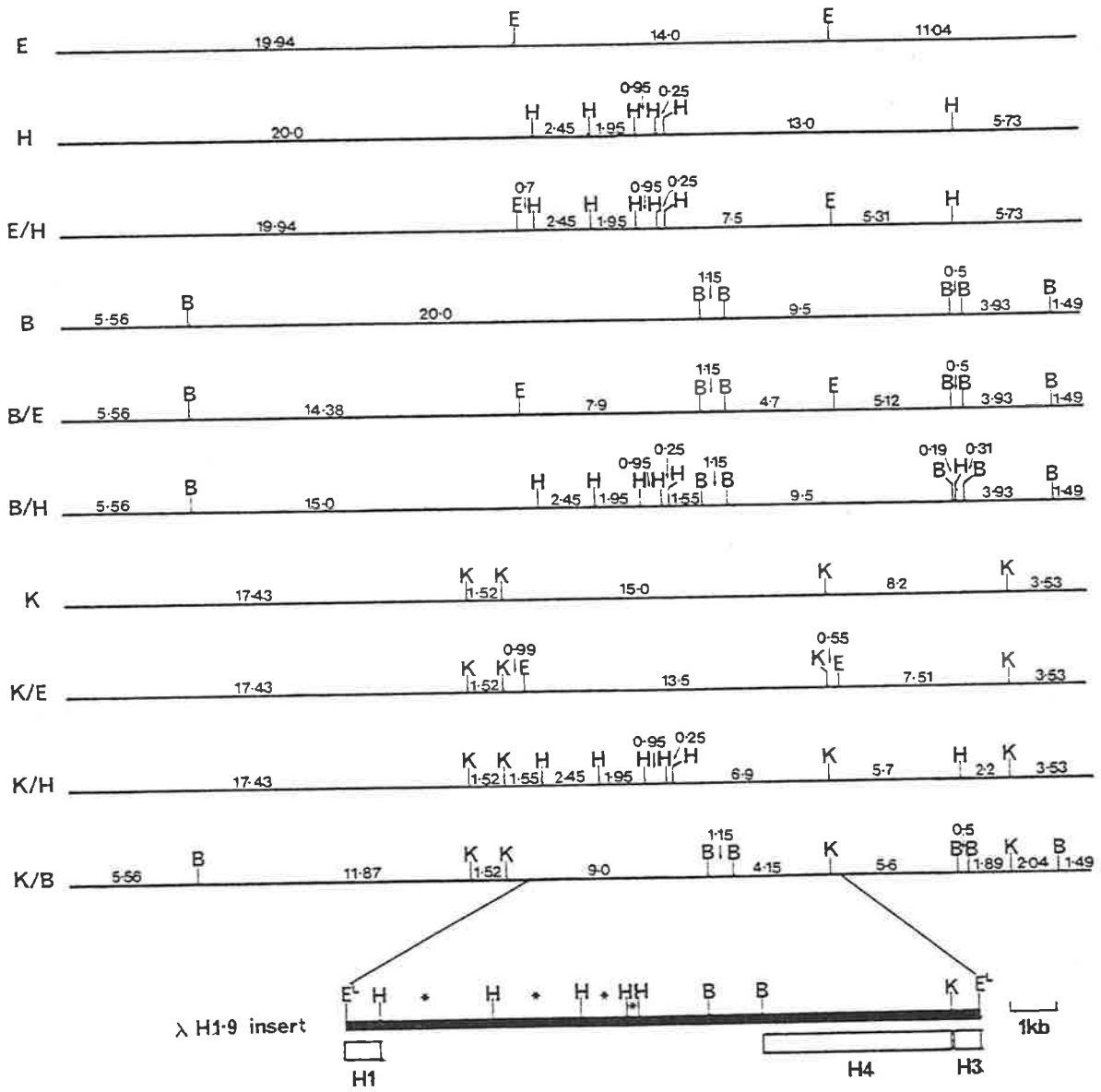

E
19·94
E
14·0
E
11·04
H
20·0
0·95
0·25
H
2·45
H
1·95
H
H
H
13·0
H
5·73
E/H
19·94
0·7
E
H
2·45
H
1·95
0·95
H
H
0·25
H
7·5
E
5·31
H
5·73
B
5·56
B
20·0
1·15
B
B
9·5
0·5
B
B
3·93
B
1·49
B/E
5·56
B
14·38
E
7·9
1·15
B
B
4·7
E
5·12
0·5
B
B
3·93
B
1·49
B/H
5·56
B
15·0
H
2·45
H
1·95
0·95
H
H
0·25
H
1·55
1·15
B
B
9·5
0·19
B
0·31
H
B
3·93
B
1·49
K
17·43
K
K
1·52
15·0
K
8·2
K
3·53
K/E
17·43
K
K
1·52
0·99
E
13·5
0·55
K
E
7·51
K
3·53
K/H
17·43
K
1·52
K
1·55
H
2·45
H
1·95
0·95
H
H
0·25
H
6·9
K
5·7
H
2·2
K
3·53
K/B
5·56
B
11·87
K
K
1·52
9·0
1·15
B
B
4·15
K
5·6
0·5
B
B
1·89
K
2·04
B
1·49
λ H1·9 insert
E
H
H
H
H
H
B
B
K
E
1kb
H1
H4
H3

FIGURE 4.23

Restriction enzyme map of λH1.10

The final restriction enzyme map of the λH1.10 insert is shown below restriction maps of individual digestions of λH1.10. Sizes of DNA fragments are in kilobases. E, H, B and K represent EcoRI, HindIII, BamHI and KpnI restriction sites respectively. E^L represents EcoRI linkers. The 2.1 kb HindIII fragment was positioned by alignment with the insert of λH1.9 (see Fig. 4.26). Remaining HindIII fragments, marked with an asterisk, have not been ordered relative to each other. Fragments hybridizing to histone gene probes are indicated with blocks. Precise locations of H2A, H2B and H1 genes, as determined by DNA sequencing (section 4.2.5(b)), are shown in Figure 4.26.

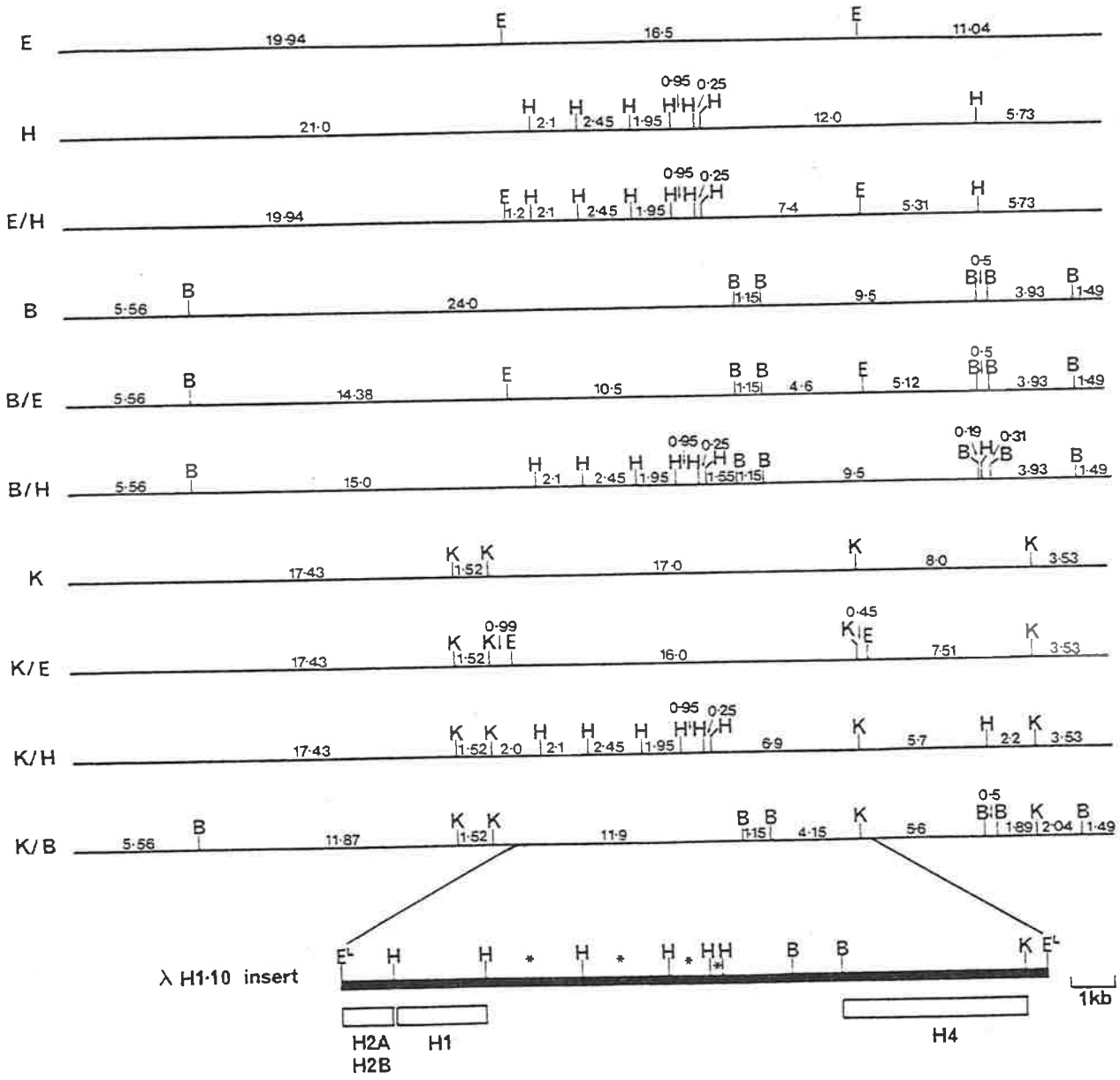

E
19·94
E
16·5
E
11·04
H
21·0
H 2·1 H 2·45 H 1·95 H H H
0·95 0·25
12·0
H
5·73
E/H
19·94
E 1·2 H 2·1 H 2·45 H 1·95 H H H
0·95 0·25
7·4
E
5·31
H
5·73
B
5·56
B
24·0
B B
1·15
9·5
0·5
B B
3·93
B
1·49
B/E
5·56
B
14·38
E
10·5
B B
1·15
4·6
E
5·12
0·5
B B
3·93
B
1·49
B/H
5·56
B
15·0
H 2·1 H 2·45 H 1·95 H H H
0·95 0·25
B B
1·55 1·15
9·5
0·19 0·31
B H B
3·93
B
1·49
K
17·43
K K
1·52
17·0
K
8·0
K
3·53
K/E
17·43
K K E
0·99
1·52
16·0
0·45
K E
7·51
K
3·53
K/H
17·43
K 1·52 K 2·0 H 2·1 H 2·45 H 1·95 H H H
0·95 0·25
6·9
K
5·7
H 2·2 K
3·53
K/B
5·56
B
11·87
K K
1·52
11·9
B B
1·15
4·15
K
5·6
0·5
B B
K
1·89
2·04
B
1·49
λ H1·10 insert
E L H H * H * H * H H B B K E L
1kb
H2A
H2B
H1
H4

FIGURE 4.24

Hybridization analysis of λH1.9

Digestions of λH1.9 were transferred to nitrocellulose and probed with histone gene probes (Section 4.2.5(a)). Filters were washed (final wash, 0.5 x SSC, 0.1% (w/v) SDS at 65°C) and autoradiographed (Section 2.3.1(d)). λH1.9 insert sequences hybridized to H1, H4 and H3 gene probes. Hybridizing sequences could be localized to specific restriction fragments as shown.

Tracks a., b. and c. are respectively EcoRI/HindIII, BamHI/KpnI, and KpnI/EcoRI digested λH1.9 DNA. Autoradiograms resulting from hybridization analysis are shown adjacent to digestions. Sizes of hybridizing fragments are given in kilobases.

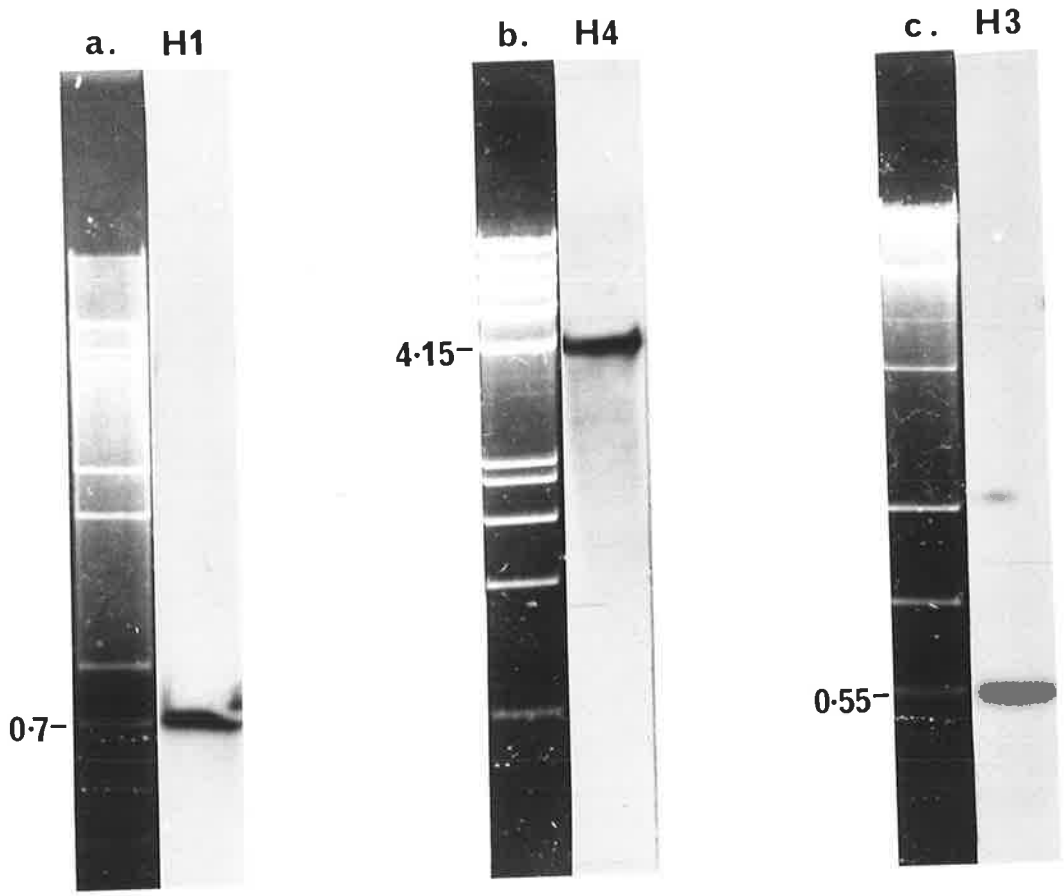
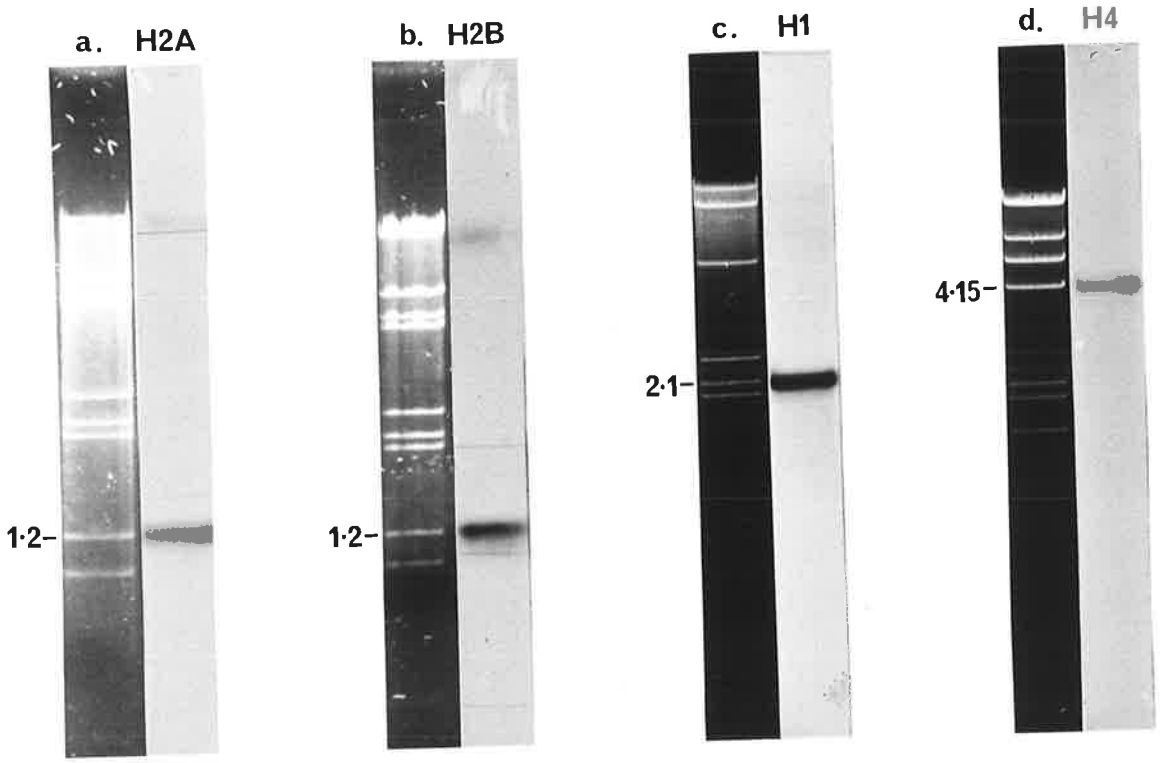

a.
H1
0·7–
b.
H4
4·15–
c.
H3
0·55–

FIGURE 4.25

Hybridization analysis of λH1.10

Digestions of λH1.10 were transferred to nitrocellulose filters and hybridized with the five histone gene probes (Section 4.2.5(a)). Filters were washed and autoradiographed as previously described (Fig. 4.24). λH1.10 insert sequences were detected by H2A, H2B, H1 and H4 gene probes. Hybridizing sequences were localized to specific insert fragments as shown.

Tracks show λH1.10 DNA digested with EcoRI and HindIII (a. and b.), with HindIII (c.) and with BamHI and KpnI (d.). Autoradiograms resulting from hybridization analysis are shown adjacent to digestions. Sizes of detected fragments are given in kilobases.


a. H2A
1·2–
b. H2B
1·2–
c. H1
2·1–
d. H4
4·15–

maps of the clone inserts (Figs. 4.22 and 4.23).

From comparison of restriction enzyme and hybridization data generated for λH1.9 and λH1.10 the inserts of the two clones can be aligned as shown in Figure 4.26. Such alignment enabled the positioning of the 2.1 kb H1 positive internal (viz., containing only insert sequences) HindIII fragment of λH1.10. The remaining four internal HindIII fragments common to both λH1.9 and λH1.10 (indicated in Fig. 4.26) have not been ordered relative to each other. These fragments do not contain histone-coding sequences. The overlapping nature of λH1.9 and λH1.10 was confirmed by comparison of DNA sequences generated from restriction fragments containing DNA common to both clones (data not shown). The insert of λH1.10 contains DNA that extends 2.6 kb to the left of that contained within λH1.9. This additional DNA contains H2A and H2B coding sequences. The λH1.9 insert on the other hand, contains DNA extending in a right-ward direction past that contained within λH1.10 by only 0.1 kb. This 0.1 kb of DNA contains the 5' end of an H3 gene (see below).

An overall map of the region of DNA covered by λH1.9 and λH1.10 is shown in Figure 4.26. Genes within fragments hybridizing to histone gene probes were located and oriented by DNA sequencing (except in the case of H4 sequences) as described below.

(b) Sequence analysis

The H3 coding region

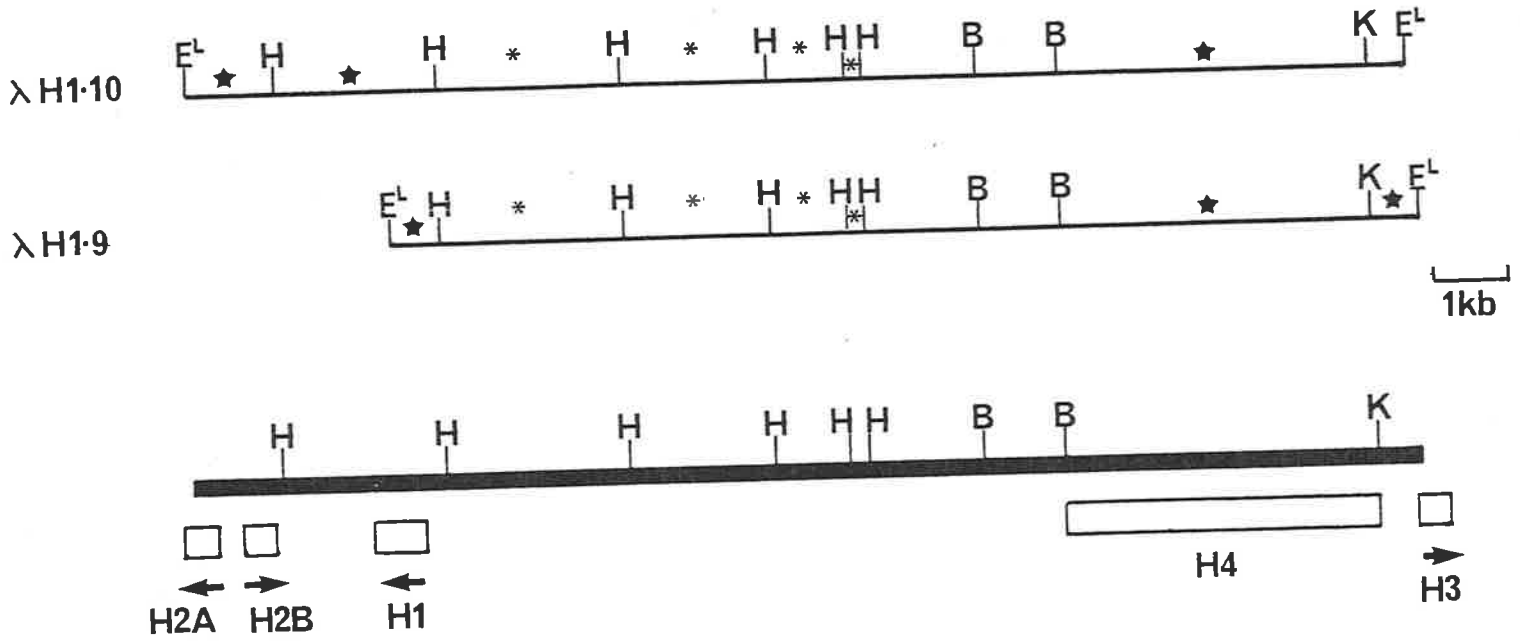
The 0.55 kb KpnI/EcoRI fragment from the right-hand end of λH1.9 was cloned into an EcoRI/KpnI digested mp19 M13 vector (Norrander et al., 1983), as described in Section 2.3.8, and sequenced using the di-deoxy sequencing procedure (Section 2.3.10). Generated sequence read from the EcoRI site towards the KpnI site. The EcoRI end of the 0.55 kb fragment

FIGURE 4.26

Organization of DNA covered by λH1.9 and λH1.10

The overall organization of genomal DNA covered by the inserts of the overlapping genomal clones λH1.9 and λH1.10 is shown below the restriction maps of the individual clone inserts (data from Figs. 4.22 and 4.23). E, H, B and K represent EcoRI, HindIII, BamHI and KpnI restriction sites respectively. E^L represents the location of EcoRI linkers used in clone construction (Dodgson et al., 1979). The 2.1 kb H1 positive HindIII fragment of λH1.10 was positioned by alignment with λH1.9. The remaining four HindIII fragments marked with * have not been ordered relative to each other.

Fragments that hybridized with histone gene probes are marked with a ★. The exact locations of H2A, H2B, H1 and H3 genes within hybridizing fragments were determined by DNA sequencing (Section 4.2.5(b)) and are indicated with open boxes. Arrows indicate directions of gene transcription. H4 hybridizing sequences were localized to a 4.15 kb BamHI/KpnI fragment as shown. The number of H4 genes within this fragment has not been determined.


λH1·10
λH1·9
1kb
H2A
H2B
H1
H4
H3

represents an EcoRI linker used in construction of the genomal library (Dodgson et al., 1979). Sequencing revealed the 5' end of an H3 gene adjacent to the EcoRI linker as shown in Figure 4.27, thus orienting the H3 gene. The coding DNA adjacent to the linker represents that coding for the first 20 amino acids of an H3 protein. Potential promoter elements, viz. "TATA" and "CCAAT" sequences (Section 1.6.1(a)), are indicated in Figure 4.27. These sequences are in similar positions, relative to the ATG codon, to elements identified in the H3 gene from the chicken genomal clone λCH03 (Wang et al., 1985).

The location of H3 coding sequences in the 60 bp of DNA at the right-hand end of the λH1.9 insert is consistent with the finding that λH1.10, which is shorter than λH1.9 by 0.1 kb at its right-hand end, is negative with the H3 gene probe (Section 4.2.2(a) and data not shown).

H2A and H2B coding regions

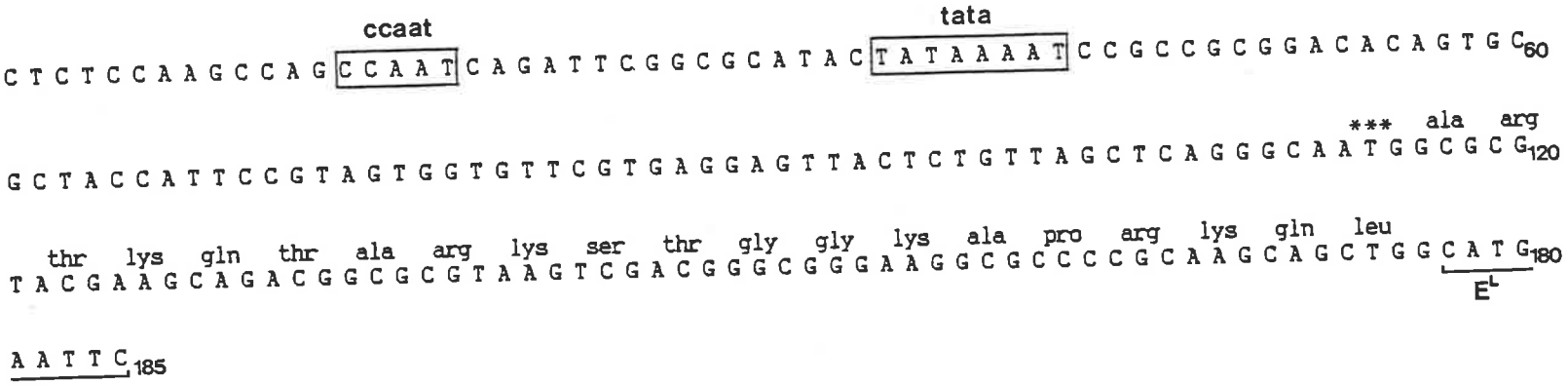
The 1.2 kb EcoRI/HindIII fragment from the left-hand end of λH1.10 (Figure 4.26) was cloned into EcoRI/HindIII digested mp83 and mp93 (Messing and Vieira, 1982) M13 vectors to permit sequencing from both the EcoRI site and the HindIII site of the fragment. The EcoRI end of the fragment represents the left-hand EcoRI linker of λH1.10. Sequencing in from the EcoRI site resulted in the identification of an H2A coding region while sequencing in from the HindIII site identified an H2B coding region. The two genes are divergently transcribed as shown in Figure 4.26. The sequence of the entire EcoRI/HindIII fragment was completed by R. Sturm of this laboratory and has been presented elsewhere (Sturm, 1985). The sequence, from the first base of the EcoRI site to the last base of the HindIII site, is 1113 bp long. The region contains DNA coding for the first 93 amino acids of an H2A protein and the entire coding region of an H2B gene. The distance between the ATG codons of the two genes is 333 bp.

FIGURE 4.27

Sequence of the λH1.9 H3 gene

DNA sequence was generated from the H3 positive, 0.55 kb KpnI/EcoRI fragment from λH1.9 as described in Section 4.2.5(b). 185 bp of generated sequence is shown reading towards the EcoRI end of the 0.55 kb fragment. The bracketed sequence marked E^L, represents that derived from the λH1.9 right-hand EcoRI linker. The amino acid sequence (in three-letter code; see Fig. 5.5) derived from the H3 coding region is shown above the DNA sequence. The initiation codon is marked ***. Potential promoter elements, viz., "TATA" and "CCAAT" sequences (Section 1.6.1 (a)), are indicated.

λH1·9 H3


ccaat
tata
CTCTCCAAGCCAGCCAATCAGATTCGGCGCATACTATAAAATCCGCCGCGGACACAGTGC60
*** ala arg
GCTACCATTCCGTAGTGGTGTTCGTGAGGAGTTACTCTGTTAGCTCAGGGCAATGGCGCG120
thr lys gln thr ala arg lys ser thr gly gly lys ala pro arg lys gln leu
TACGAAGCAGACGGCGCGTAAGTCGACGGGCGGGAAGGCGCCCCGCAAGCAGCTGGCATG180
EL
AATTC185

111 bp 3' to the H2B coding region are also present.

Divergently transcribed H2A/H2B gene pairs have been found to be a common feature among chicken histone genes (Sturm, 1985; Harvey et al., 1982; Figs. 4.4 and 4.19). Coding regions of H2A/H2B gene pairs are separated on average by a distance of 342 bp. Each gene pair contains a series of conserved sequence elements and it has been demonstrated, via analysis of the H2A/H2B gene pair from λCH01 in the Xenopus oocyte transcription system, that the divergent genes contain overlapping promoter elements (Sturm, 1985). It is suggested that the promoter arrangement of the two genes has the potential to co-ordinate expression of these genes at the transcriptional level, resulting in the balanced production of H2A and H2B proteins (Sturm, 1985).

The protein encoded by the H2B gene from λH1.10 corresponds to the H2B.1 chicken variant (Urban et al., 1979; Section 1.2.3) as do the proteins encoded by most chicken H2B genes, where sufficient sequence data is available (Harvey et al., 1982; Sturm, 1985; Wigley, 1982; Section 3.2.4(b)). In contrast, the H2A gene from λH1.10 codes for a protein that corresponds to the H2A.2 chicken variant (Urban et al., 1979; Section 1.2.3), while the proteins coded by all other sequenced chicken H2A genes correspond to the H2A.1 variant (Harvey et al., 1982; Sturm, 1985; Wang et al., 1985). This data is consistent with the observation that H2A.1 and H2A.2 represent the major and minor chicken H2A variants respectively (Urban and Zweidler, 1983).

The H1 coding region

The H1 coding region was located and oriented by directional sequencing of the 0.7 kb EcoRI/HindIII fragment from the left-hand end of λH1.9. The entire sequence of the H1 coding region plus flanking sequences was determined by sequence analysis of the 2.1 kb HindIII fragment from λH1.10. The location and direction of transcription of the H1 gene is shown

in Figure 4.26. A more detailed description of the strategy used to sequence this gene and the final sequence is presented in Section 5.2.2.

(c) Summary of λH1.9/λH1.10 data and the total H1 gene number

Together the inserts of λH1.9 and λH1.10 cover 16.3 kb of DNA as shown in Figure 4.26. The region of DNA was found to contain sequences hybridizing to all five histone gene probes. DNA sequencing subsequently identified a single H2A, H2B, H1 and H3 gene. The fragment hybridizing to the H4 gene probe (a 4.15 kb BamHI/KpnI fragment) was not analyzed by sequencing and hence may contain more than one H4 gene. The direction of transcription of all sequenced genes was determined by directional sequencing from restriction sites.

Comparison of the map of DNA covered by λH1.9 and λH1.10 with that of DNA in Figures 4.4 and 4.19 shows that the λH1.9/λH1.10 region cannot overlap with any of the DNA in the regions spanning λCHO1 (Fig. 4.4) or λCHO2 (Fig. 4.19) and hence cannot contain a previously identified H1 gene. This is consistent with data discussed in Section 4.2.2(b). R. D'Andrea (D'Andrea, 1985; D'Andrea et al., 1985) has independently determined that the three regions of characterized histone DNA (Figs. 4.4, 4.19 and 4.26) are non-overlapping. Preliminary analysis of an additional region of DNA containing a single H2A and H3 gene (contained within a single cosmid 4.2C) shows that this too is independent of the other three unlinked regions (D'Andrea et al., 1985; D'Andrea, 1985).

The four non-overlapping regions discussed above, containing clustered histone genes, are not closely linked to DNA containing the solitary histone genes that code for the replication-independent histone variants H2A.F, H5 and H3.3 (Section 1.5.3(d); D'Andrea et al., 1985; D'Andrea, 1985). As discussed in Section 1.5.3(d) analysis of clustered histone genes has so far only revealed replication-dependent genes.

The data presented here from characterization of λH1.9 and λH1.10 (Fig. 4.26) therefore represents a previously uncharacterized stretch of chicken histone DNA, containing at least five "new" histone genes and of particular relevance to this study, contains a "new" H1 gene. The data in Figure 4.26 completed the characterization of clustered chicken histone genes at the organizational level, as presented by D'Andrea et al. (1985).

Altogether six H1 gene-containing regions have been identified in the characterized chicken histone DNA (D'Andrea, 1985; Lesnikowski, 1983; this chapter; Figures 4.4, 4.19 and 4.26). As already discussed the single H1 positive regions in DNA covered by λH1.9/λH1.10 (Figure 4.26) and in DNA spanning λCH02 (Fig. 4.19) contain single H1 genes (Sections 4.2.5(b) and 4.2.3(c)). The four H1 positive regions, located in DNA spanning λCH01 (Fig. 4.4), also represent single H1 genes (discussed in Chapter 5). A total of six H1 genes have, therefore, been located. Data presented below suggests that these genes represent the full complement of H1 genes in the chicken genome.

4.2.6 Southern Analysis of Chicken Genomal DNA

Chicken genomal DNA from chicken erythrocytes (see Section 4.2.4) was digested with EcoRI, HindIII and BamHI alone or in pairwise combinations and electrophoresed on a 0.65% agarose gel. Digested DNA was then transferred to nitrocellulose (Section 2.3.1(d)) and probed with the ^{32}P-labelled insert of the H1 cDNA clone, cH1.1 (used throughout this chapter for the detection of H1 gene sequences). The resultant autoradiogram is shown in Figure 4.28. This data has also been presented by D'Andrea et al. (1985).

Using the data in Figures 4.4, 4.19 and 4.26 all the genomic DNA fragments detected by the H1 gene probe could be accounted for. The sizes of predicted and observed H1 gene containing DNA fragments are shown in

FIGURE 4.28

Analysis of chicken genomal DNA with an H1 probe

Chicken genomal DNA from chicken erythrocytes (gift from R. Harvey) was digested (8 ug/digestion) with EcoRI, HindIII and BamHI, alone or in pairwise combinations, and electrophoresed on a 0.65% agarose gel. HindIII digested λ DNA was run as size markers. Digested DNA was transferred to a single nitrocellulose filter and hybridized with the ^{32}P-labelled insert of the H1 cDNA clone, cH1.1 (2×10^7 dpm; probe prepared as described in Section 4.2.1). The nitrocellulose filter was washed (final wash in 0.1 x SSC, 0.1% (w/v) SDS at 65°C) and autoradiographed for several days (Section 2.3.1(d)).

The figure shows the autoradiogram resulting from hybridization analysis. Genomal DNA was digested with:

Track:
a. BamHI and HindIII
b. BamHI
c. BamHI and EcoRI
d. EcoRI
e. EcoRI and HindIII
f. HindIII

Sizes of marker DNA fragments are given in kilobases.

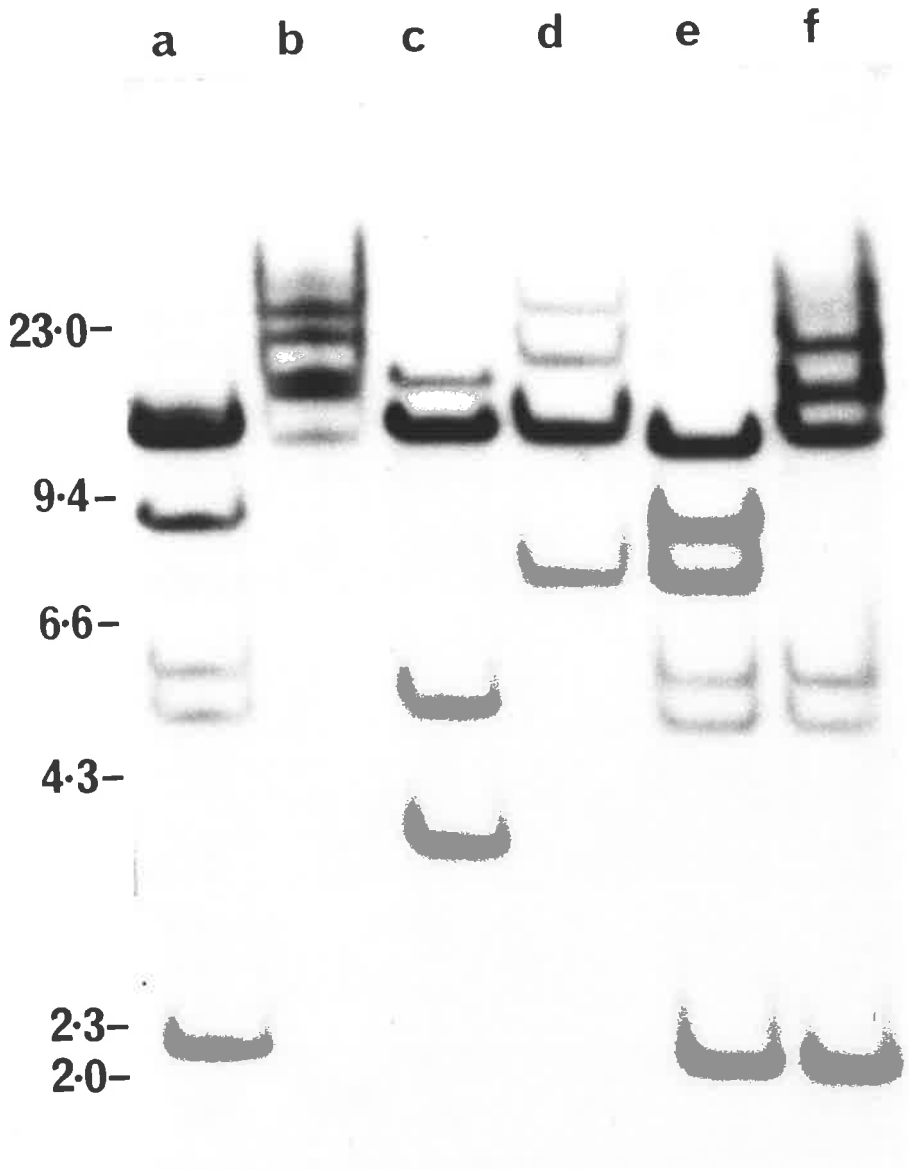

a
b
c
d
e
f
23·0–
9·4–
6·6–
4·3–
2·3–
2·0–

Table 4.1. The DNA of the individual used for the analysis lacked the BamHI site, in Figure 4.4, that is 3' to the H1 gene from λCH03. This individual was also heterozygous for the presence or absence of the HindIII site immediately adjacent to the above mentioned BamHI site, hence leading to the generation of 6.4 kb and 5.4 kb HindIII fragments which contain the same H1 gene (viz. the λCH03 H1 gene) and haploid amounts of DNA. The BamHI and HindIII sites discussed above are marked on Figure 4.4.

The allelic nature of the 6.4 kb and 5.4 kb HindIII fragments was confirmed via probing HindIII digested chicken genomal DNA with a non-coding probe (a 2.5 kb PstI fragment adjacent to the λCH03 H1 gene; indicated in Fig. 4.4) selected as being common to both HindIII fragments. The results of the probing (Fig.4.29) show that both the 6.4 kb and 5.4 kb H1 positive HindIII fragments were detected by the probe and hence represent allelic DNA. Ruiz-Carrillo et al. (1983) have also analyzed restriction enzyme digested chicken genomal DNA with an homologous H1 probe. In their experiments a DNA fragment corresponding to the 6.4 kb HindIII fragment, but not the 5.4 kb HindIII fragment, was detected. The DNA of the individual they used, therefore, lacks the HindIII site (marked in Fig. 4.4) for which the individual described above is heterozygous.

The analysis of chicken genomal DNA described here suggests that the six H1 genes, presented in Figures 4.4, 4.19 and 4.26, represent the full complement of H1 genes in the chicken genome. It appears that the full complement of clustered core histone genes have also been isolated among our characterized regions of histone DNA (D'Andrea et al., 1985). In addition to the six H1 genes, ten H2A, eight H2B, ten H3 and eight H4 genes have been located (contained in DNA presented in Figures 4.4, 4.19 and 4.26, and within the cosmid 4.2C; see Section 4.2.5(c)).

TABLE 4.1

Results of genomal DNA analysis

The predicted and observed sizes of H1 gene-containing genomal DNA fragments, in kilobases (kb), are given in the table. Predicted fragment sizes were determined from data in Figures 4.4, 4.19 and 4.26. The particular H1 gene(s) present on each predicted DNA fragment is indicated. 03, 01, p11 , .10 and 02 represent λCH03, λCH01, pCH11.5E, λH1.9/λH1.10 and λCH02 H1 genes respectively.

The allelic HindIII fragments discussed in Section 4.2.6 are marked with an *.

BamHI/HindIII

Predicted	Observed
11.5 kb (p11)	12.0 kb
11.0 kb (01)	12.0 kb
8.0 kb (02)	8.5 kb
*6.4 kb (03)	5.8 kb
*5.4 kb (03)	4.8 kb
2.1 kb (.10)	2.1 kb

BamHI

Predicted	Observed
> 24.5 kb (03)	~26.0 kb
21.5 kb (02)	21.5 kb
> 10.45 kb (.10)	17.5 kb
17.2 kb (p11)	16.5 kb
11.0 kb (01)	12.0 kb

BamHI/EcoRI

Predicted	Observed
> 10.45 kb (.10)	17.5 kb
11.5 kb (p11)	12.0 kb
11.2 kb (03)	12.0 kb
5.4 kb (01)	5.0 kb
3.5 kb (02)	3.5 kb

EcoRI

Predicted	Observed
> 16.3 kb (.10)	~26.0 kb
15.8 kb (02)	18.0 kb
11.5 kb (p11)	12.0 kb
11.2 kb (03)	12.0 kb
7.2 kb (01)	7.2 kb

EcoRI/HindIII

Predicted	Observed
11.0 kb (p11)	11.0 kb
8.5 kb (02)	8.5 kb
7.2 kb (01)	7.2 kb
*6.4 kb (03)	5.8 kb
*5.4 kb (03)	4.8 kb
2.1 kb (.10)	2.1 kb

HindIII

Predicted	Observed
17.0 kb (01)	19.0 kb
13.0 kb (02)	15.0 kb
11.5 kb (p11)	12.0 kb
*6.4 kb (03)	5.8 kb
*5.4 kb (03)	4.8 kb
2.1 kb (.10)	2.1 kb

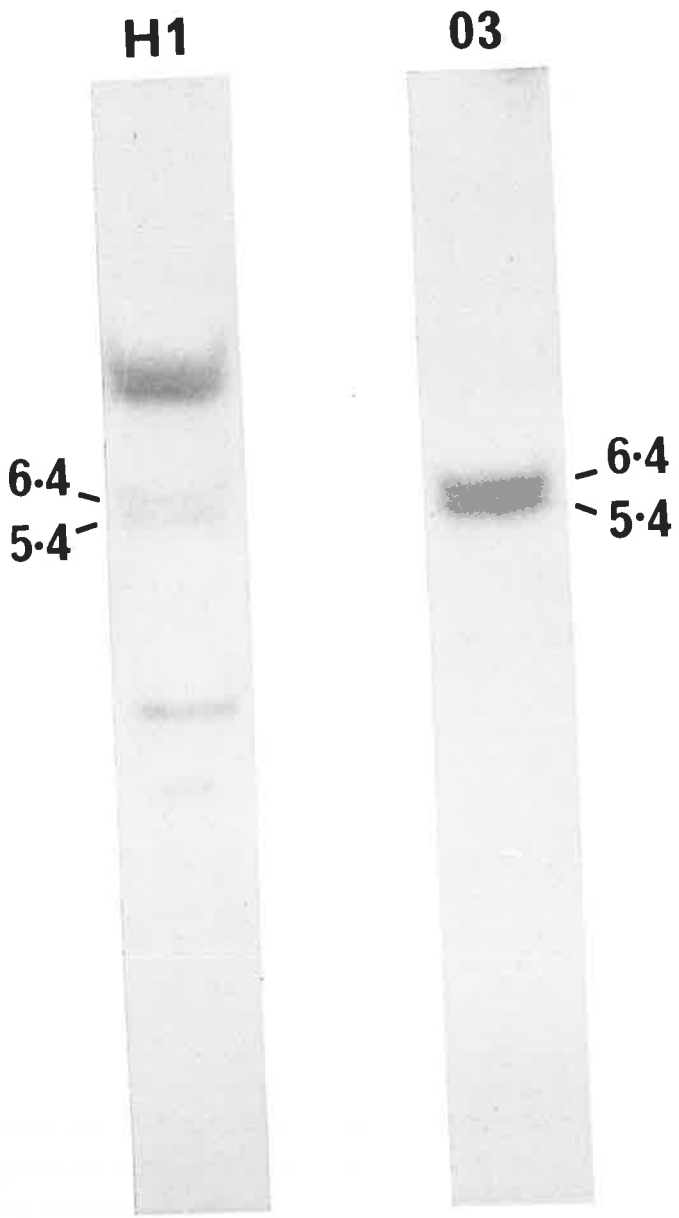
FIGURE 4.29

Detection of allelic HindIII fragments in genomal DNA

15 ug of genomal DNA from chicken erythrocytes (used in previous analysis; Fig. 4.28) was digested with HindIII, split into two samples and electrophoresed on a 0.8% agarose gel. Digested DNA was transferred to a nitrocellulose filter.

One track of HindIII digested DNA was hybridized with the ^{32}P-labelled insert of the H1 cDNA clone cH1.1 (prepared as described previously; Fig. 4.28), while, the other track was hybridized with a non-coding 2.5 kb PstI fragment from λCHO3 (fragment is indicated in Fig. 4.4). The 2.5 kb fragment was isolated from LGT agarose (Section 2.3.1(b)), after PstI digestion of the λCHO3 subclone, pCH5.0E (Fig 4.4), and ^{32}P-labelled by nick-translation (Section 2.3.7(a)). Filters were washed (final wash, 0.1 x SSC, 0.1% (w/v) SDS at 65°C) and autoradiographed.

Autoradiograms resulting from hybridization analysis are shown. H1 represents digested genomal DNA probed with the cH1.1 probe, and O3 represents digested genomal DNA probed with the non-coding probe from λCHO3. Sizes of allelic HindIII fragments discussed in Section 4.2.6 are given in kilobases.


H1
03
6·4
5·4
6·4
5·4

4.2.7 Discussion

Analysis of H1 positive recombinants from a chicken genomal library, resulted in the identification of four clones containing previously uncharacterized histone-coding DNA. Two of the recombinants, λH1.1 and λH1.4, overlapped and extended DNA within the recombinant λCHO2. Another two recombinants, λH1.9 and λH1.10, overlapped each other but were independent of other mapped regions of histone DNA. These two clones contained a previously unidentified H1 gene.

Characterization of the above four clones completed our analysis of clustered chicken histone genes at the organizational level (Section 4.2.5(c); D'Andrea et al., 1985). Of the forty-two clustered histone genes located, six of these are H1 genes. Data presented here suggests that this represents the full complement of H1 genes in the chicken genome.

Five of the six chicken H1 genes are found to be associated with both an H2A and an H2B gene (Figs. 4.4, 4.19 and 4.26). The three sets of associated genes in DNA spanning λCHO1 (Fig. 4.4) have the same gene organization, viz., $\overrightarrow{H1}$ $\overleftarrow{H2A}$ $\overrightarrow{H2B}$ (arrows indicate direction of transcription), while the genes in DNA covered by λH1.9/λH1.10 (Fig. 4.26) and in DNA spanning λCHO2 (Fig. 4.19) have different organizations, of $\overleftarrow{H2A}$ $\overrightarrow{H2B}$ $\overleftarrow{H1}$ and H2A $\overrightarrow{H1}$ $\overleftarrow{H2B}$ respectively. The various clusters may have arisen from a common ancestral cluster.

Preferred arrangements of genes,such as those described above, are a common feature of chicken histone DNA (Section 1.5.3(d)), even though the chicken genes lack the high degree of order exhibited in organisms with a high histone gene copy number (Section 1.5.5). The preferred arrangements of histone genes seen in the chicken may be relevant to the expression of these genes, as appears to be the case for H2A/H2B gene pairs (discussed in Section 4.2.5(b); Sturm, 1985). It is suggested that associated histone genes may be co-ordinately expressed (D'Andrea et al., 1985; Wang et al., 1985; Sturm, 1985).

CHAPTER 5

ANALYSIS OF THE CHICKEN H1 GENES

5.1 INTRODUCTION

As discussed in Chapter 4 a total of six H1 genes were located in three non-overlapping regions of chicken histone DNA. Four H1 genes reside in DNA spanning λCHO1 (Fig. 4.4), while the regions of DNA spanning λCHO2 (Fig. 4.19) and covered by λH1.9 and λH1.10 (Fig. 4.26) each contain a single H1 gene. The six located H1 genes appear to represent the full complement of such genes in the chicken genome (Section 4.2.6). All located H1 genes are clustered with core histone genes and are preferentially associated with H2A and H2B genes. None of the H1 genes discussed here, therefore, exist as "solitary" histone genes as does the gene for the H1 related, H5 protein.

The characterization of the clustered chicken H1 genes is discussed below.

5.2 RESULTS

Each of the six H1 genes was positively identified by DNA sequencing (for this thesis and by others; J. Powell, 1984; A. Robins and S. Clark, unpublished; Sugarman et al., 1983) which also enabled the precise location and orientation of each gene. The complete coding region sequence and extensive flanking sequences were determined for the λCHO1, λH1.9/λH1.10 and λCHO2 H1 genes and for the H1 gene from the left-hand end of the pCH11.5E subclone (from the cosmid clone 6.3C, Fig 4.4). Partial sequences have been determined for the λCHO3 H1 gene and for the H1 gene from the right-hand end of pCH11.5E. The complete sequence of the λCHO2 H1 gene was determined by Sugarman et al. (1983) while the partial sequence of the λCHO3 H1 gene was determined by J. Powell (1984).

Sequence analysis of the six H1 genes along with some preliminary analysis of the H1 gene transcripts is discussed in this Chapter.

Characterization of the λCH01 H1 gene is discussed first. This gene is used as the "standard" for sequence comparisons in Chapter 6.

5.2.1 Characterization of the λCH01 H1 Gene

H1 coding sequences within the genomal clone λCH01 were localized by R. D'Andrea (1985) to a 1 kb SacII/SacI restriction fragment by hybridization analysis, using the insert of the H1 cDNA clone, cH1.1 (Section 3.2.5), as a probe. This region of DNA is contained within a subclone, pCH3.3E, derived from λCH01. This subclone was constructed by cloning the right-hand 3.3 kb EcoRI fragment from λCH01 (Fig. 4.4) into the plasmid vector pBR325 (Bolivar, 1978) as previously described (Harvey et al., 1981). A restriction map of the pCH3.3E insert and location of coding regions, viz. H1, H2A and H2B gene coding regions, is shown in Figure 5.1. The right-hand end of the pCH3.3E insert represents the right-hand EcoRI linker of λCH01.

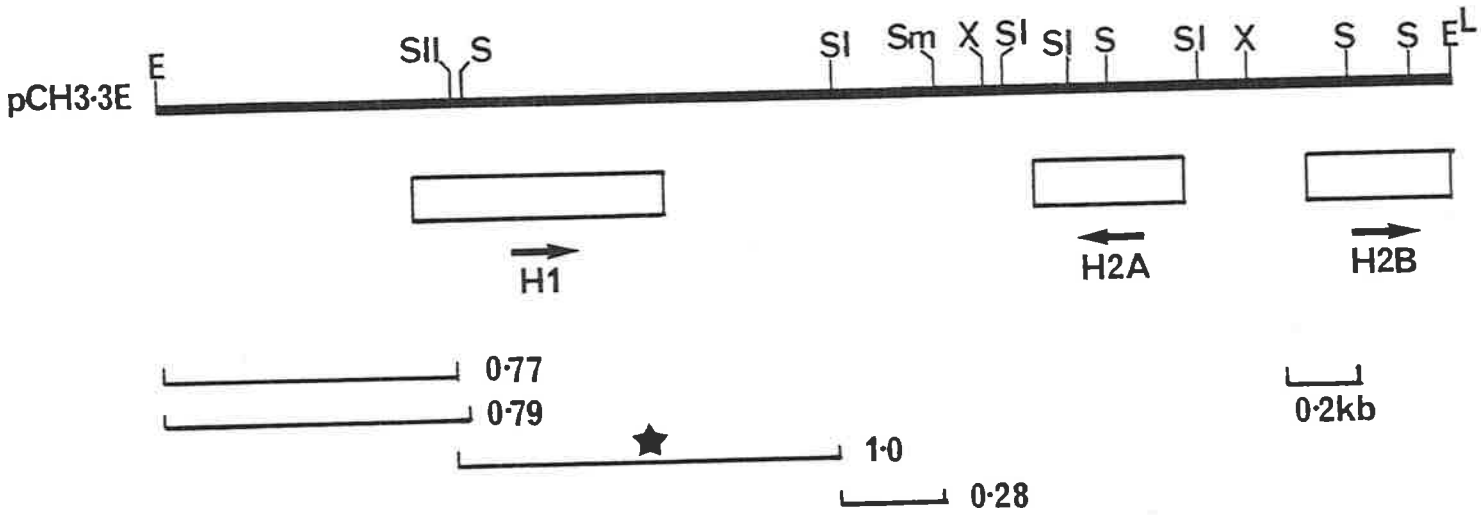
The 1.3 kb SmaI/EcoRI fragment from pCH3.3E has been completely sequenced in this laboratory (Harvey et al., 1982) and contains a single H2A and H2B gene which are divergently transcribed (Fig. 5.1). The EcoRI linker end of the clone insert is a few base pairs 3' to the H2B coding region. For analysis of the H1 coding sequences it was decided to sequence a 2.2 kb EcoRI/XhoI fragment which contained within it the 1 kb SacII/SacI fragment positive to the H1 probe. This would enable the completion of the sequence of the entire pCH3.3E insert and ensure the detection of the entire H1 gene and its flanking sequences. It was possible that the SacII/SacI fragment did not contain a complete H1 gene as coding sequences in the probe only represent those encoding amino acid 49 to the terminal amino acid (Section 3.2.5). The determination of the sequence of the 2.2 kb EcoRI/XhoI fragment is discussed below.

FIGURE 5.1

Restriction map of the pCH3.3E insert

A restriction map of the insert of pCH3.3E is shown. Data is from R. D'Andrea (1985) and from DNA sequence information (Harvey et al., 1982; and described here). E, SII, S, Sm, X and SI represent respectively EcoRI, SacII, Sau3AI, SmaI, XhoI and SacI restriction enzyme sites. E^L represents the right-hand EcoRI linker of λCHO1.

The exact location (indicated with boxes) and direction of transcription of histone genes (indicated by horizontal arrows) has been determined by DNA sequencing. H2A and H2B genes were sequenced by Harvey et al. (1982). Sequencing of the 2.2 kb EcoRI/XhoI fragment containing the H1 gene is described in Section 5.2.1(a). The location of restriction fragments discussed in the text are indicated. Sizes of fragments are in kilobases. ★ marks the H1 positive 1.0 kb SacII/SacI fragment indentified by R. D'Andrea (1985).


pCH3·3E
E
SII
S
SI
Sm
X
SI
SI
S
SI
X
S
S
EL
H1
H2A
H2B
0·77
0·79
1·0
0·28
0·2kb

(a) Sequence analysis

The 2.2 kb EcoRI/XhoI fragment from pCH3.3E (gift from R. Harvey) was isolated from an LGT agarose gel (Section 2.3.1(b)) and digested with either HaeIII or FnuDII. Restricted DNA was analyzed on a 5% polyacrylamide gel as shown in Figure 5.2. All fragments greater than approximately 60 bp were recovered from the gel (Section 2.3.1(c)) and cloned into the SmaI site of M13 mp83 or mp93 vectors (Messing and Vieira, 1982; Section 2.3.8). Single-stranded recombinants containing cloned fragments in both orientations were obtained by complementarity testing (c-testing, Section 2.3.9(a)) and then subjected to di-deoxy sequence analysis (Section 2.3.10).

Numerous GC-rich stretches within the λCH01 H1 gene coding region and within the coding regions of the other chicken H1 genes were difficult to resolve on standard sequencing gels, resulting in sequencing "compressions". Where this occurred sequencing reactions were run on gels containing 25 to 40% formamide (Section 2.3.10(c)). An example of the use of a formamide sequencing gel in resolving a compression is shown in Figure 5.3.

A map of restriction enzyme sites in the 2.2 kb fragment and the regions of the fragment sequenced from HaeIII and FnuDII restriction sites is shown in Figure 5.4. The data generated enabled the determination of a large portion of the sequence of the 2.2 kb fragment. Sequencing of additional restriction enzyme generated and BAL-31 generated DNA fragments resulted in the completion of the sequence (see Fig. 5.4).

The additional restriction fragments from which sequence were generated were a 0.77 kb EcoRI/SacII, a 1 kb SacII/SacI, a 0.28 kb SacI/SmaI, a 0.79 kb EcoRI/Sau3AI and 168 bp and 162 bp HpaII fragments. The former four fragments (shown on Fig. 5.1) were isolated from LGT agarose or a 6% polyacrylamide gel (the 0.28 kb SacI/SmaI fragment) after appropriate digestion of pCH3.3E. The former three fragments were then

FIGURE 5.2

Digestion of the 2.2 kb EcoRI/XhoI fragment

The 2.2 kb EcoRI/XhoI fragment from pCH3.3E was digested (1 ug/digestion) with either FnuDII (F) or HaeIII (H) and electrophoresed on a 5% polyacrylamide gel as shown. The sizes of marker DNA fragments (pBR322 digested with HinfI) are shown in base pairs. Fragments greater than 60 bp were recovered from the gel (Section 2.3.1(c)) for cloning into M13 vectors. The arrow indicates a FnuDII fragment used in the generation of BAL-31 fragments (Section 5.2.1(a)).

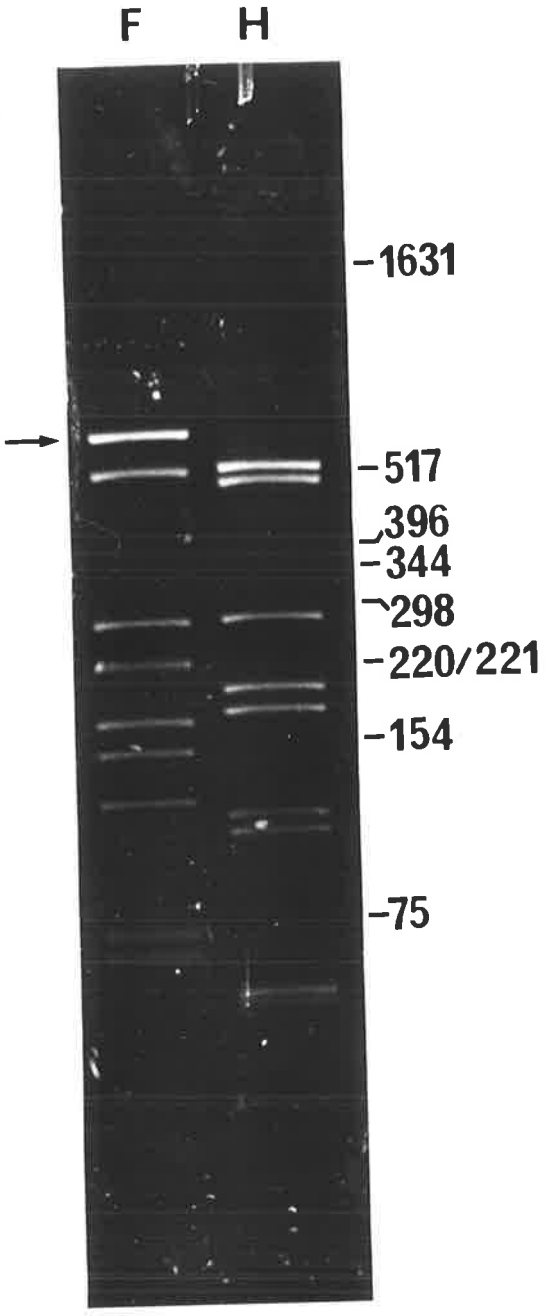

F
H
1631
517
396
344
298
220/221
154
75

FIGURE 5.3

Resolution of sequence compressions

GC-rich stretches in DNA sequences were often difficult to resolve on normal sequencing gels, resulting in sequencing "compressions". Where this occurred sequencing reactions were run on gels containing 25 to 40% formamide (Section 2.3.10(c)).

The figure shows a portion of an autoradiogram of sequencing reactions that were electrophoresed on a normal 6% polyacrylamide sequencing gel (Section 2.3.10(c)). The bracketed region shows an example of a sequencing "compression". The same sequencing reactions were then run on a sequencing gel containing 40% formamide (Section 2.3.10(c)), resulting in the resolution of the "compressed" region. The sequence in the bracketed region reads 5' CGGCGGGC 3'. This sequence represents residue number 80 to residue number 87 of the λCHO1 H1 gene coding region (Fig. 5.5).

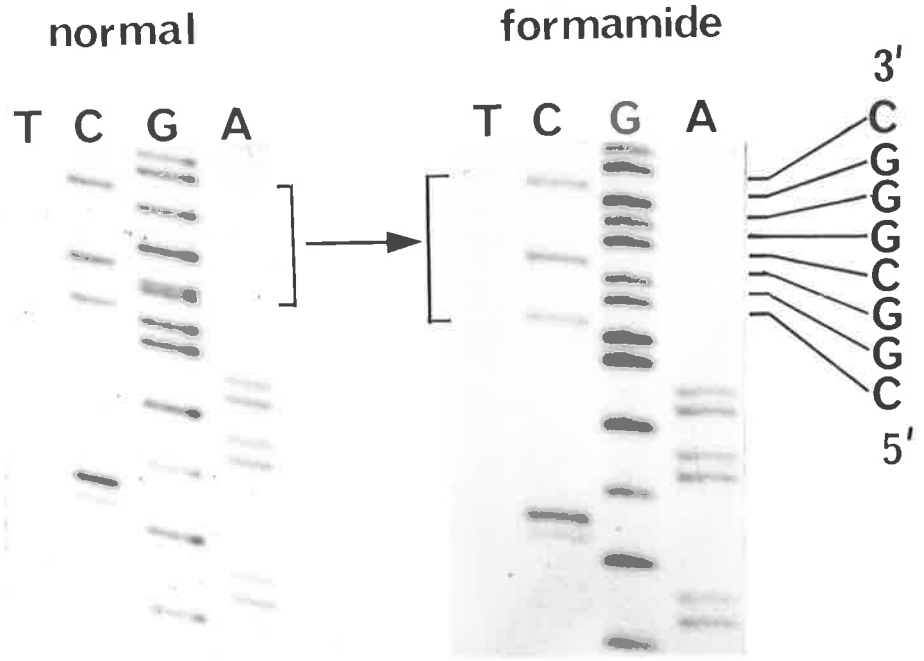

normal
formamide
T C G A
T C G A
3′
C
G
G
G
C
G
G
C
5′

FIGURE 5.4

Strategy for sequencing the λCH01 H1 gene

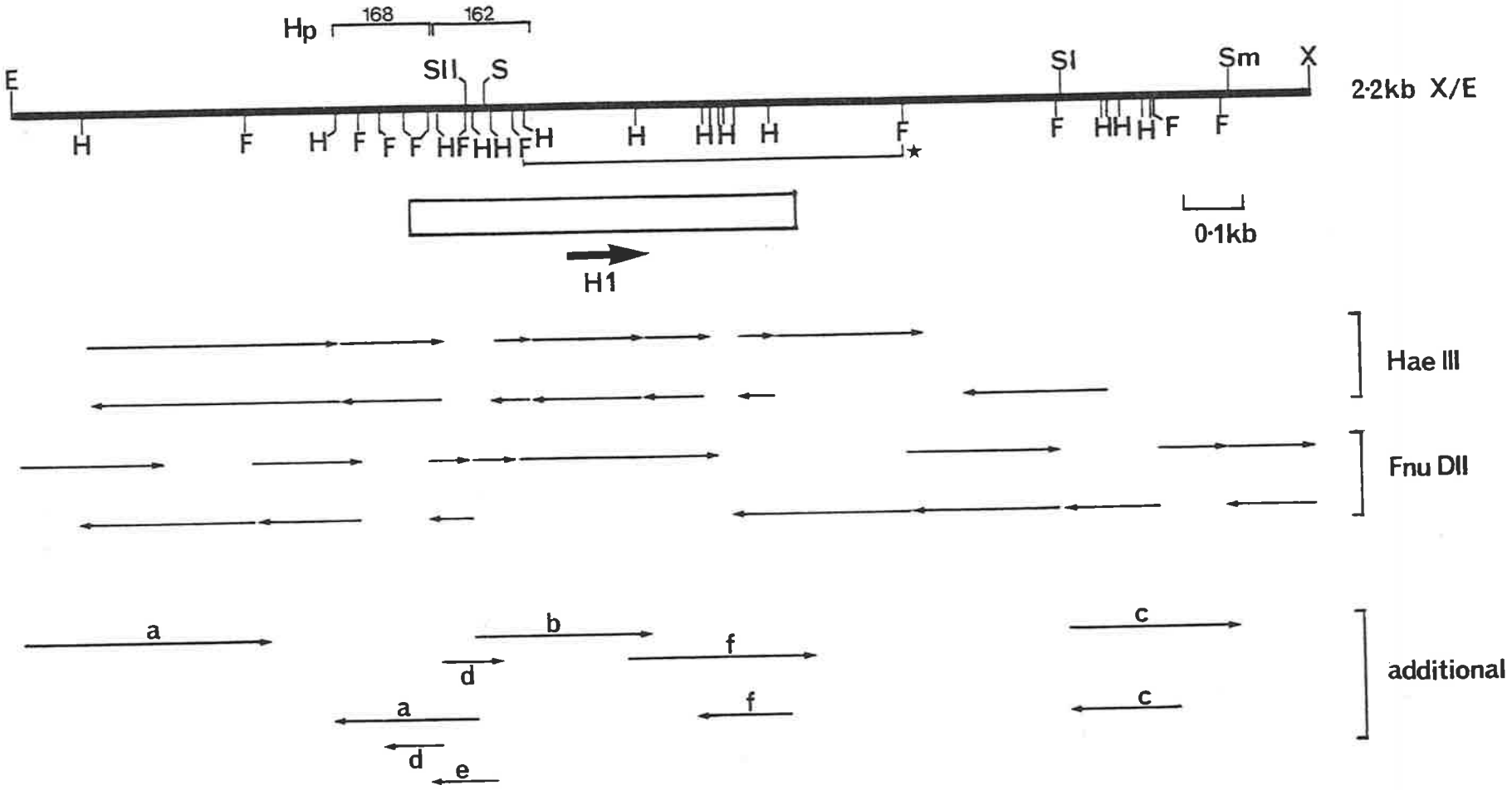
A restriction map of the 2.2 kb EcoRI/XhoI fragment from pCH3.3E is shown. Restriction sites used in the generation of DNA sequence are indicated, viz., SacII (SII), Sau3AI (S), SacI (SI), SmaI (Sm), HaeIII (H) and FnuDII (F) sites. E and X represent EcoRI and XhoI sites respectively. The location of 168 and 162 bp HpaII fragments (Hp) used in sequence determination are also shown. The FnuDII fragment, from which BAL-31 fragments were generated (Section 5.2.1(a)), is marked with a ★.

The location of the H1 gene (indicated by the open block) is shown and the large vertical arrow shows the direction of gene transcription.

Small vertical arrows indicate regions of sequence determined from HaeIII, FnuDII and additional DNA fragments. Arrow directions indicate the 5' to 3' direction of generated sequence.

Additional sequences are from:

a. 0.77 kb EcoRI/SacII fragment
b. 1 kb SacII/SacI fragment
c. 0.28 kb SacI/SmaI fragment
d. 168 and 162 bp HpaII fragments
e. 0.79 kb EcoRI/Sau3AI fragment
f. BAL-31 generated fragments (Section 5.2.1(a)).


Hp
168
162
E
SII
S
SI
Sm
X
2·2kb X/E
H
F
H
F
F
F
HF
HH
H
F
H
H
HH
H
F
F
HHH
F
F
0·1kb
H1
Hae III
Fnu DII
a
b
f
c
d
a
f
c
d
e
additional

made blunt-ended (Section 2.3.4(a)), cloned into a SmaI digested M13 vector, and c-tested, while the 0.79 kb EcoRI/Sau3AI fragment was cloned into a BamHI/EcoRI mp83 directional M13 vector. The regions of DNA sequenced from cloned fragments are shown in Figure 5.4. The two HpaII fragments were isolated from a polyacrylamide gel, after digestion of the 2.2 kb EcoRI/XhoI fragment with HpaII, and cloned into an AccI digested M13 vector. The locations of the two HpaII fragments are shown in Figure 5.4. For simplicity the entire HpaII restriction map of the EcoRI/XhoI fragment is not shown.

The BAL-31 generated fragments, used for completion of the 2.2 kb sequence (see above), were derived from the 672 bp FnuDII fragment indicated on Figures 5.2 and 5.4. The 672 bp fragment, as for the other FnuDII fragments, had been cloned into the SmaI site of an M13 vector. The insert of the generated clone was isolated from LGT agarose after double-digestion of the clone (double-stranded form, Section 2.3.8(a)) with BamHI and EcoRI (sites flank the SmaI site). The insert (approx. 2 ug) was digested with BAL-31 for 2 minutes, blunt-ended and fractionated on an LGT agarose gel as described in Section 2.3.8(b). DNA of approximately 250-300 bp was isolated, cloned into SmaI digested M13 mp83, c-tested and sequenced. The region of the 2.2 kb EcoRI/XhoI fragment sequenced from the generated clones is shown in Figure 5.4.

The complete sequence of the 2.2 kb EcoRI/XhoI fragment from pCH3.3E is shown in Figure 5.5 and has recently been presented elsewhere (Coles and Wells, 1985). The fragment contains a single H1 coding region of 660 bp (ATG and TAA inclusive), which codes for an H1 protein of 218 amino acids, 658 bp of 5' non-coding DNA and 883 bp of 3' non-coding DNA.

There is no evidence for introns within the λCHO1 H1 gene, a feature common to most sequenced chicken histone genes (Section 1.5.3(d)) and histone genes from other species (Maxson et al., 1983a, 1983b; Hentschel and Birnstiel, 1981; Old and Woodland, 1984). The only chicken histone

FIGURE 5.5

DNA sequence of the λCHO1 H1 gene

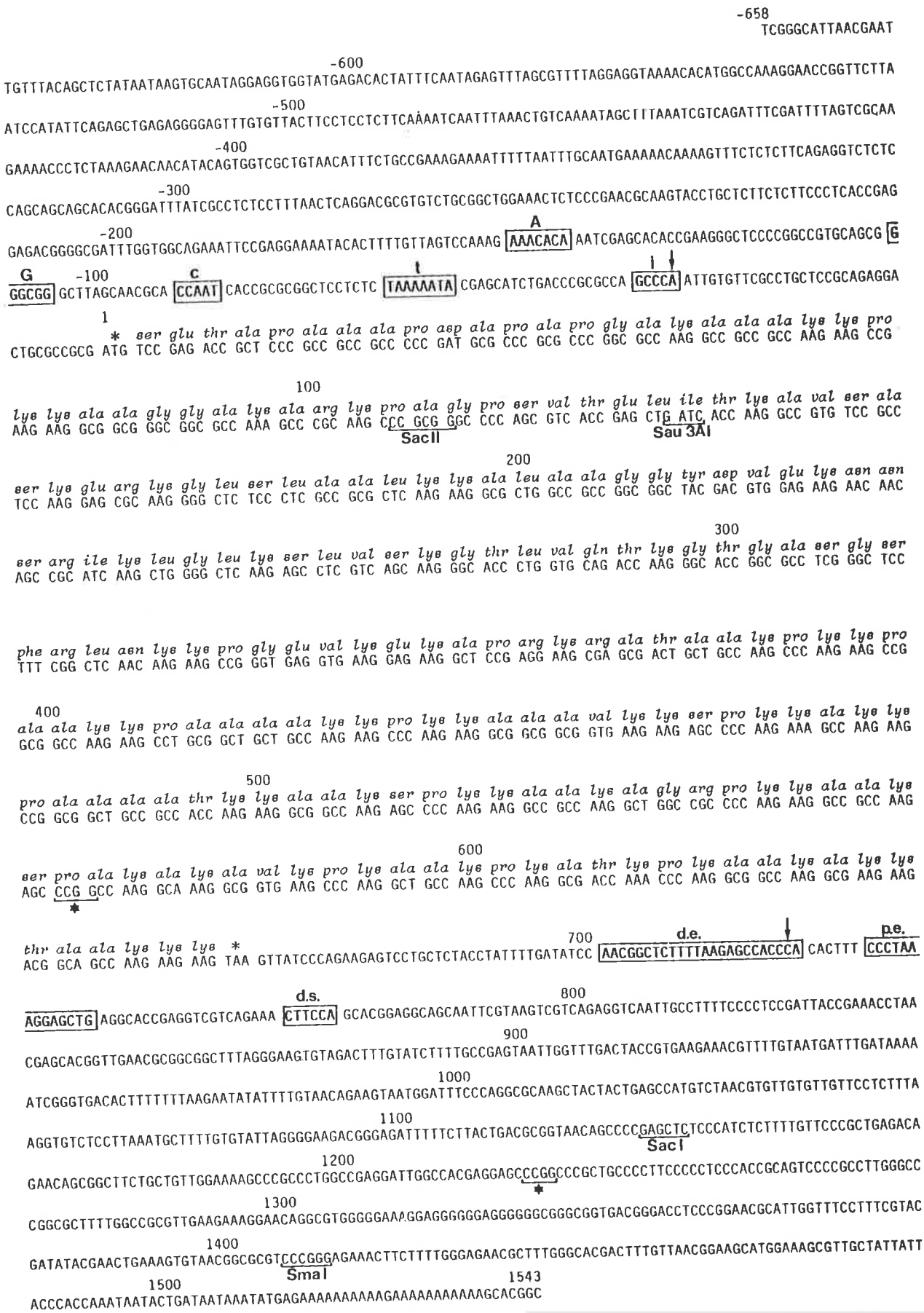
The nucleotide sequence of the 2.2 kb EcoRI/XhoI fragment from pCH3.3E is shown. The H1 protein coding region is represented as triplets and the derived amino acid sequence is given above the DNA sequence [a]. Initiation (ATG) and termination (TAA) codons are marked with an *. Restriction sites used in the generation of DNA sequence are indicated. HpaII sites flanking a 659 bp restriction fragment, used in an S1 nuclease protection experiment, are also indicated and marked with a ✱.

Conserved sequences in 5' and 3' flanking regions (see Figs. 6.4, 6.5, and 6.6) are labelled as follows:

A.	A-rich H1-specific element
G.	G-rich element
c.	"CCAAT" sequence
t.	"TATA" sequence
i.	sequence containing site for initiation of transcription
d.e.	3' histone gene-specific dyad symmetry element
p.e.	purine-rich element
d.s.	conserved downstream sequence

The locations of sites representing mRNA 5' and 3' termini are indicated by vertical arrows (Section 5.2.1(b)).

[a] Amino acids are abbreviated by the three-letter code, viz., asp, aspartic acid; asn, asparagine; thr, threonine; ser, serine; glu, glutamic acid; gln, glutamine; pro, proline; gly, glycine; ala, alanine; val, valine; ile, isoleucine; leu, leucine; tyr, tyrosine; phe, phenylalanine; lys, lysine; arg, arginine.


-658
TCGGGCATTAACGAAT
-600
TGTTTACAGCTCTATAATAAGTGCAATAGGAGGTGGTATGAGACACTATTTCAATAGAGTTTAGCGTTTTAGGAGGTAAAACACATGGCCAAAGGAACCGGTTCTTA
-500
ATCCATATTCAGAGCTGAGAGGGGAGTTTGTGTTACTTCCTCCTCTTCAAAATCAATTTAAACTGTCAAAATAGCTTTAAATCGTCAGATTTCGATTTTAGTCGCAA
-400
GAAAACCCTCTAAAGAACAACATACAGTGGTCGCTGTAACATTTCTGCCGAAAGAAAATTTTTAATTTGCAATGAAAAACAAAAGTTTCTCTCTTCAGAGGTCTCTC
-300
CAGCAGCAGCACACGGGATTTATCGCCTCTCCTTTAACTCAGGACGCGTGTCTGCGGCTGGAAACTCTCCCGAACGCAAGTACCTGCTCTTCTCTTCCCTCACCGAG
-200
GAGACGGGGCGATTTGGTGGCAGAAATTCCGAGGAAAATACACTTTTGTTAGTCCAAAG AAACACA AATCGAGCACACCGAAGGGCTCCCCGGCCGTGCAGCG G
G GGCGG -100 GCTTAGCAACGCA c CCAAT CACCGCGCGGCTCCTCTC t TAAAAATA CGAGCATCTGACCCGCGCCA GCCCA ATTGTGTTCGCCTGCTCCGCAGAGGA
1
* ser glu thr ala pro ala ala ala pro asp ala pro ala pro gly ala lys ala ala ala lys lys pro
CTGCGCCGCG ATG TCC GAG ACC GCT CCC GCC GCC GCC CCC GAT GCG CCC GCG CCC GGC GCC AAG GCC GCC GCC AAG AAG CCG
100
lys lys ala ala gly gly ala lys ala arg lys pro ala gly pro ser val thr glu leu ile thr lys ala val ser ala
AAG AAG GCG GCG GGC GGC GCC AAA GCC CGC AAG CCC GCG GGC CCC AGC GTC ACC GAG CTG ATC ACC AAG GCC GTG TCC GCC
SacII Sau 3AI
200
ser lys glu arg lys gly leu ser leu ala ala leu lys lys ala leu ala ala gly gly tyr asp val glu lys asn asn
TCC AAG GAG CGC AAG GGG CTC TCC CTC GCC GCG CTC AAG AAG GCG CTG GCC GCC GGC GGC TAC GAC GTG GAG AAG AAC AAC
300
ser arg ile lys leu gly leu lys ser leu val ser lys gly thr leu val gln thr lys gly thr gly ala ser gly ser
AGC CGC ATC AAG CTG GGG CTC AAG AGC CTC GTC AGC AAG GGC ACC CTG GTG CAG ACC AAG GGC ACC GGC GCC TCG GGC TCC
phe arg leu asn lys lys pro gly glu val lys glu lys ala pro arg lys arg ala thr ala ala lys pro lys lys pro
TTT CGG CTC AAC AAG AAG CCG GGT GAG GTG AAG GAG AAG GCT CCG AGG AAG CGA GCG ACT GCT GCC AAG CCC AAG AAG CCG
400
ala ala lys lys pro ala ala ala ala lys lys pro lys lys ala ala ala val lys lys ser pro lys lys ala lys lys
GCG GCC AAG AAG CCT GCG GCT GCT GCC AAG AAG CCC AAG AAG GCG GCG GCG GTG AAG AAG AGC CCC AAG AAA GCC AAG AAG
500
pro ala ala ala ala thr lys lys ala ala lys ser pro lys lys ala ala lys ala gly arg pro lys lys ala ala lys
CCG GCG GCT GCC GCC ACC AAG AAG GCG GCC AAG AGC CCC AAG AAG GCC GCC AAG GCT GGC CGC CCC AAG AAG GCC GCC AAG
600
ser pro ala lys ala lys ala val lys pro lys ala ala lys pro lys ala thr lys pro lys ala ala lys ala lys lys
AGC CCG GCC AAG GCA AAG GCG GTG AAG CCC AAG GCT GCC AAG CCC AAG GCG ACC AAA CCC AAG GCG GCC AAG GCG AAG AAG
*
thr ala ala lys lys lys *
ACG GCA GCC AAG AAG AAG TAA GTTATCCCAGAAGAGTCCTGCTCTACCTATTTTGATATCC
700 d.e. AACGGCTCTTTTAAGAGCCACCCA CACTTT p.e. CCCTAA
d.s. 800
AGGAGCTG AGGCACCGAGGTCGTCAGAAA CTTCCA GCACGGAGGCAGCAATTCGTAAGTCGTCAGAGGTCAATTGCCTTTTCCCCTCCGATTACCGAAACCTAA
900
CGAGCACGGTTGAACGCGGCGGCTTTAGGGAAGTGTAGACTTTGTATCTTTTGCCGAGTAATTGGTTTGACTACCGTGAAGAAACGTTTTGTAATGATTTGATAAAA
1000
ATCGGGTGACACTTTTTTTAAGAATATATTTTGTAACAGAAGTAATGGATTTCCCAGGCGCAAGCTACTACTGAGCCATGTCTAACGTGTTGTGTTGTTCCTCTTTA
1100
AGGTGTCTCCTTAAATGCTTTTGTGTATTAGGGGAAGACGGGAGATTTTTCTTACTGACGCGGTAACAGCCCCGAGCTCTCCCATCTCTTTTGTTCCCGCTGAGACA
SacI
1200
GAACAGCGGCTTCTGCTGTTGGAAAAGCCCGCCCTGGCCGAGGATTGGCCACGAGGAGCCCGGCCCGCTGCCCCTTCCCCCTCCCACCGCAGTCCCCGCCTTGGGCC
*
1300
CGGCGCTTTTGGCCGCGTTGAAGAAAGGAACAGGCGTGGGGGAAAGGAGGGGGGAGGGGGGCGGGCGGTGACGGGACCTCCCGGAACGCATTGGTTTCCTTTCGTAC
1400
GATATACGAACTGAAAGTGTAACGGCGCGTCCCGGGAGAAACTTCTTTTGGGAGAACGCTTTGGGCACGACTTTGTTAACGGAAGCATGGAAAGCGTTGCTATTATT
SmaI
1500
ACCCACCAAATAATACTGATAATAAATATGAGAAAAAAAAAAAGAAAAAAAAAAAGCACGGC
1543

genes found to contain introns are the replication-independent H2A.F and H3.3 genes (Robins et al., 1985; Brush et al., 1985; Section 1.5.3(d)).

"TATA" and "CCAAT" box sequences, common to histone genes and other RNA polymerase II transcribed genes (Section 1.6.1(a)) are indicated in Figure 5.5, in the DNA 5' to the H1 coding region. In DNA 3' to the H1 coding region, sequences representing histone gene-specific conserved 3' elements (Section 1.6.2(a)) are shown. These sequences are necessary for efficient and accurate 3' end formation of non-polyadenylated histone mRNAs (Birchmeier et al., 1982, 1983, 1984; Georgiev and Birnstiel, 1985; Section 1.6.2(a)). Other sequences discussed in Chapter 6 are also indicated. Vertical arrows represent mRNA 5' and 3' termini. The location of these sites is discussed below.

(b) Mapping mRNA termini

The 3' terminus of the λCHO1 H1 gene mRNA was mapped in an S1 nuclease protection experiment (Section 2.3.11(b)). The RNA used was from 5 day old chicken embryos. As previously discussed, such RNA is a good source of histone mRNA (Sections 3.1 and 3.2.1; Crawford et al., 1979) and was used to construct the cDNA library described in Chapter 3. The embryo RNA used in the S1 experiment was a gift from R. Sturm (5-16S fraction, prepared by the method of Brooker et al., 1980).

For probe preparation, the 2.2 kb EcoRI/XhoI fragment containing the λCHO1 H1 gene was digested with HpaII and a 659 bp HpaII fragment was isolated from a 6% polyacrylamide gel. This fragment flanks the 3' end of the λCHO1 H1 gene coding region. Its location is indicated on the sequence in Figure 5.5. The fragment was then 3' end-labelled by end-fill labelling (Section 2.3.7(b)) using α-^{32}P-dCTP, and hybridized to a single-stranded M13 recombinant containing sequences complementary to the λCHO1 H1 gene coding strand as described in Section 2.3.11(b). The M13 clone used was

a gift from R. Sturm and contains DNA covering the entire 3.3 kb of DNA represented in pCH3.3E. The required unhybridized labelled DNA strand was isolated from a 6% polyacrylamide gel, hybridized to the fractionated 5 day embryo RNA, treated with SI nuclease and analyzed on a 6% polyacrylamide sequencing gel as previously described (Section 2.3.11(b)). The results of the SI analysis are shown in Figure 5.6.

The single major protected DNA fragment of 160 bases maps the 3' terminus of the λCH01 H1 gene mRNA within the 5' ACCCA 3' sequence present at the 3' end of the highly conserved histone-specific dyad symmetry element (Section 1.6.2(a)) as shown in Figure 5.5. The 3' termini of the λCH01 H2A and H2B gene mRNAs have been mapped to a similar position by SI analysis of 5 day embryo RNA (Sturm, 1985). This location of histone mRNA 3' termini is consistent with observations by others, that such termini in other species generally map at or near the 3' end of the dyad element (Hentschel and Birnstiel, 1981; Maxson et al., 1983b; Section 1.6.2(a)).

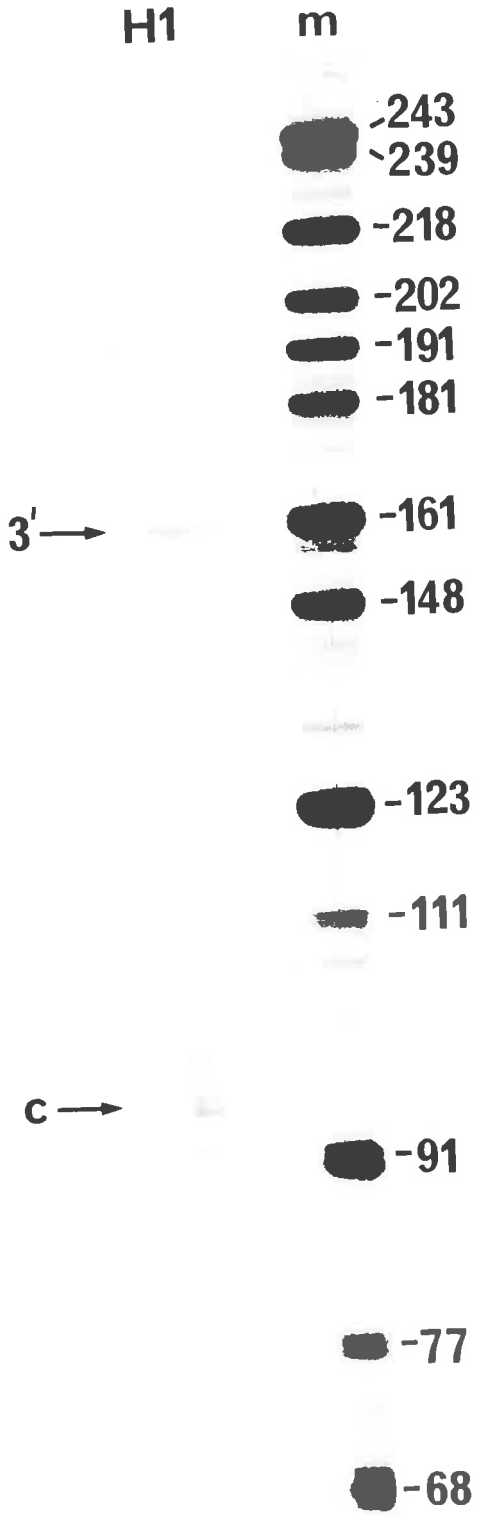
The additional protected DNA fragments, indicated in Figure 5.6, which range in size from 82 to 97 bases, represent those protected by the protein coding regions of mRNAs derived from other H1 genes expressed in the 5 day embryo. Immediately 3' to coding regions there is little homology between the chicken H1 gene sequences (data available for the λCH01, λCH02, λH1.9/H1.10 and pCH11.5E left-hand H1 genes; Sections 5.2.2, 5.2.3 and 5.2.5).

The 5' terminus of mRNA from the λCH01 H1 gene was mapped by primer extension analysis (McKnight et al., 1981) using a gene-specific 26-base synthetic DNA primer and the same fractionated 5 day embryo RNA as used above. This work was performed by R. Sturm (1985). The terminus indicated by a vertical arrow in Figure 5.5, was mapped to the residue 37 bp 5' to the initiation codon, and is located at the expected distance downstream from the "TATA" box (Hentschel and Birnstiel, 1981; Section 1.6.1(a)). The 5' terminus of the λCH02 H1 gene mRNA has been mapped, by Sugarman et al.

FIGURE 5.6

Mapping λCH01 H1 gene 3' mRNA termini

The 3' terminus of the λCH01 H1 gene mRNA was mapped in an $S1$ nuclease protection experiment as described in Section 5.2.1(b). RNA was from 5 day old chicken embryos (gift of R. Sturm). The location of the 659 bp HpaII fragment used in the protection experiment is indicated in Figure 5.5. The $S1$ reaction mixture was electrophoresed on a 6% polyacrylamide sequencing gel and autoradiographed as previously described (Section 2.3.11(b)). HpaII digested pBR322, ^{32}P-labelled by end-fill labelling (Section 2.3.7(b)) using α^{32}P-dCTP, were run on the gel as size markers. The resultant autoradiogram of the $S1$ experiment is shown. The protected DNA fragment mapping the λCH01 H1 mRNA 3' terminus (3') is indicated. Fragments protected by the protein coding regions of mRNAs derived from other H1 genes are also indicated (c). Sizes of markers (m) are in bases.


H1
m
243
239
218
202
191
181
3'
161
148
123
111
c
91
77
68

(1983; Section 5.2.5) in 4.5 day chicken embryos using a 5' protection assay. The 5' termini of mRNAs from both chicken H1 genes map to the same sequence, viz. to the sequence 5' GCCCA 3', the A of which represents the terminus.

5.2.2 Sequence Analysis of the λH1.9/λH1.10 H1 Gene

Within DNA covered by the overlapping genomal clones λH1.9 and λH1.10 (Fig. 4.26) a single H1 positive region was identified (Section 4.2.5(a)). λH1.9/λH1.10 DNA also contains H2A, H2B, H3 and H4 coding sequences, the former three being positively identified by DNA sequencing (Section 4.2.5(b)). The overlapping clones are presented again in Figure 5.7.

The H1 positive regions in λH1.10 and λH1.9 are respectively a 2.1 kb HindIII and a 0.7 kb EcoRI/HindIII fragment. The EcoRI end of the 0.7 kb fragment represents the left-hand EcoRI linker of λH1.9.

(a) Gene orientation

The 0.7 kb EcoRI/HindIII fragment from λH1.9 was cloned into EcoRI/HindIII double-digested mp83 and mp93 M13 vectors so that sequence could be generated from both the EcoRI (linker) and HindIII ends of the fragment as shown in Figure 5.7. Such sequencing resulted in the identification of a 5' portion of an H1 coding region (coding for the first 132 amino acids of an H1 protein) adjacent to the EcoRI linker end of the fragment, hence orienting the H1 gene.

The complete sequence of the λH1.9/λH1.10 H1 gene was determined by sequence analysis of the 2.1 kb HindIII fragment from λH1.10.

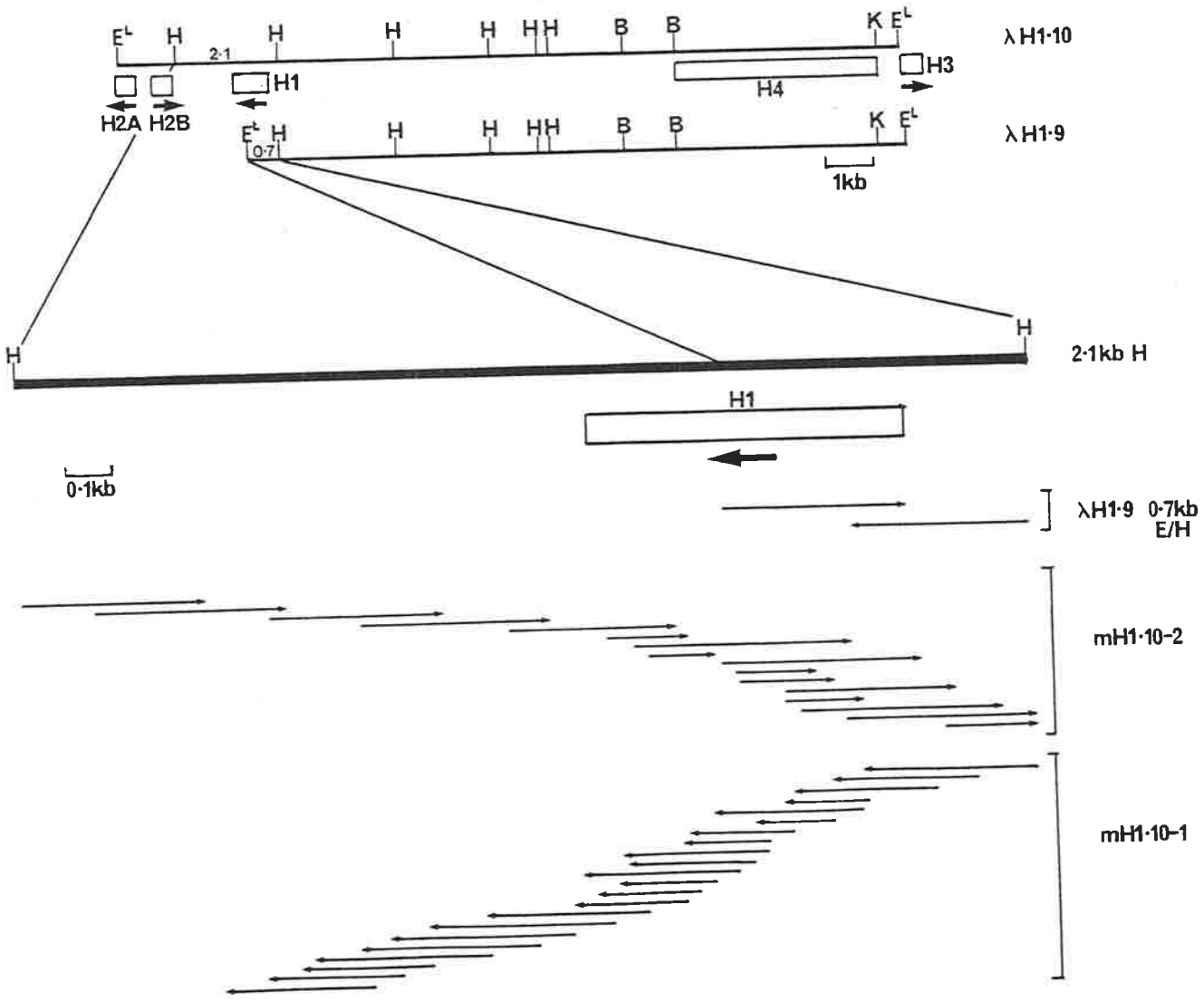
FIGURE 5.7

Strategy for sequencing the λH1.9/λH1.10 H1 gene

Restriction maps of the inserts of the overlapping genomal clones λH1.9 and λH1.10 (Fig. 4.26) are shown at the top of the diagram. H, B and K represent HindIII, BamHI and KpnI restriction enzyme sites respectively. E^L represents EcoRI linkers. HindIII fragments (except the 2.1 kb fragment) have not been ordered relative to each other.

H2A, H2B and H3 histone genes were located by DNA sequencing, as described in Section 4.2.5(b). Arrows indicate directions of gene transcription. The number of H4 genes within the 4.15 kb H4 positive BamHI/KpnI fragment has not been determined.

The 2.1 kb HindIII and 0.7 kb EcoRI/HindIII fragments, used for determination of the H1 gene sequence, are indicated (sizes in kilobases). Sequence from the 0.7 kb EcoRI/HindIII fragment was used to orient the H1 gene (Section 5.2.2(a)). For complete sequence determination the λH1.10 2.1 kb HindIII fragment was cloned into M13 in both orientations (generating mH1.10-2 and mH1.10-1) and BAL-31 fragments were then generated as described in Section 5.2.2(b). Regions of DNA sequenced from the 0.7 kb fragment and from the mH1.10-2 and mH1.10-1 BAL-31 fragments, are indicated by horizontal arrows. Directions of arrows represent the 5' to 3' direction of generated sequence.


λH1·10
λH1·9
2·1kb H
λH1·9 0·7kb E/H
mH1·10-2
mH1·10-1
H1
H2A
H2B
H3
H4
1kb
0·1kb

(b) Complete sequence determination

The 2.1 kb HindIII fragment from λH1.10 was isolated from LGT agarose and cloned into the HindIII site of M13 mp83. Constructed clones were then c-tested (Section 2.3.9(a)). Clones containing DNA in both orientations, named mH1.10-1 and mH1.10-2, were sequenced (Section 2.3.10), hence generating sequence from both ends of the 2.1 kb fragment. Sequence from the right-hand HindIII site was identical to that generated by sequencing from the HindIII site of the λH1.9 0.7 kb fragment.

For use in the construction of directional BAL-31 fragment containing clones, double-stranded DNA from mH1.10-1 and mH1.10-2 was prepared (Section 2.3.8(a)). DNA of each clone was then cut with EcoRI (EcoRI site is 3', relative to the direction of sequencing, to the HindIII site in the mp83 linker) and treated with BAL-31. After various lengths of time aliquots were taken from the reaction mix and blunt-ended as described in Section 2.3.8(b). Treated DNA was then cut with HindIII, to separate insert from vector sequences, and fractionated on a 1% LGT agarose gel (Fig. 5.8). Approximately ten size fractions, between 0.2 kb and 2 kb, of insert DNA from each of the treated mH1.10 clones were isolated. DNA from each size fraction was then cloned into M13 mp19 (Norrander et al., 1983) double-digested with HincII (generates blunt-ends) and HindIII. Hence sequence generated read from the BAL-31 produced ends of the cloned fragments (blunt-ends) towards their HindIII ends. HindIII ends of fragments represented either the right-hand or left-hand HindIII ends of the "parental" 2.1 kb HindIII fragment, depending on whether the fragments were derived from mH1.10-2 or mH1.10-1.

The regions of the 2.1 kb HindIII fragment sequenced from the constructed clones are shown in Figure 5.7. Sequence was determined for the entire 2.1 kb DNA fragment and was found to contain only a single H1 gene, viz., the one identified by sequencing the 0.7 kb EcoRI/HindIII fragment of λH1.9.

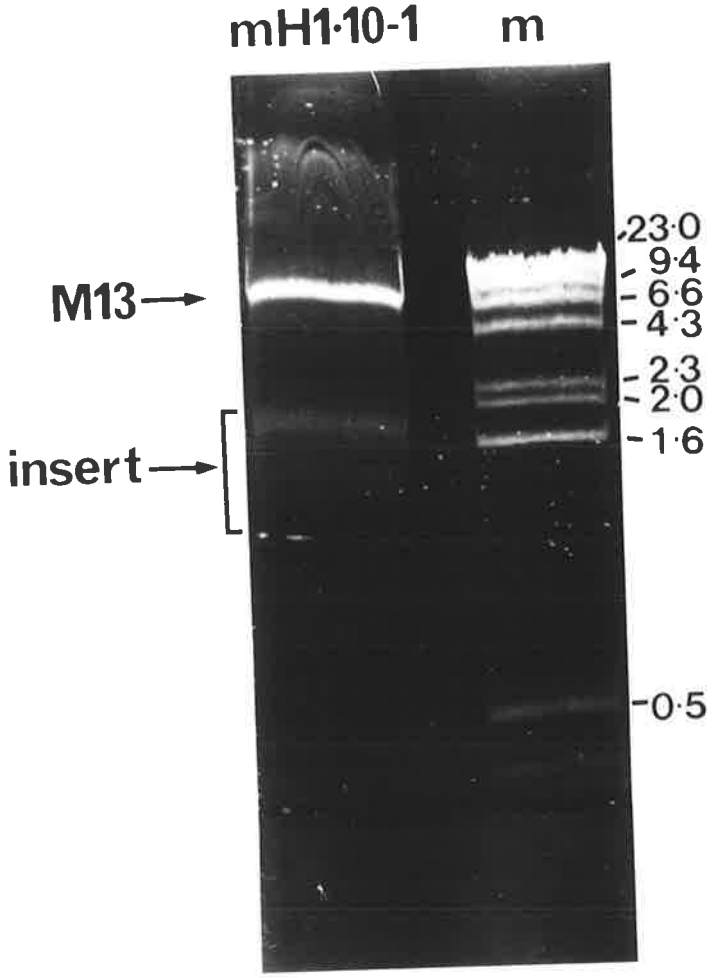
FIGURE 5.8

Generation of λH1.9/λH1.10 H1 gene BAL-31 fragments

Double-stranded DNA of the M13 clones, mH1.10-1 and mH1.10-2, was used for the generation of BAL-31 fragments for sequence determination (Section 5.2.2(b)).

Shown, is an example of BAL-31 fragments generated from mH1.10-1. 1 ug mH1.10-1 DNA was digested with EcoRI and treated with BAL-31. 2 and 5 minute time samples were taken and the DNA was made blunt-ended (Section 2.3.8(b)). Treated DNA was then cut with HindIII, to separate vector from insert sequences, and fractionated on a 1% horizontal LGT agarose gel.

The gel track containing BAL-31 treated DNA is labelled mH1.10-1 and the track labelled m is marker DNA (sizes are in kilobases). M13 DNA and insert DNA (present as a smear of BAL-31 fragments) are indicated. Additional BAL-31 treatments were also performed for varying lengths of time in order to acquire a complete range of insert size fractions (Section 5.2.2(b)).


mH1·10-1
m
M13
insert
23·0
9·4
6·6
4·3
2·3
2·0
1·6
0·5

1497 bp of the 2.1 kb fragment are presented in Figure 5.9. This sequence reads from and includes the right-hand HindIII site of the 2.1 kb fragment as oriented in Figure 5.7. The sequence represents 247 bp of 5' non-coding DNA, 663 bp of coding DNA, including ATG and TGA codons, and 587 bp of 3' non-coding DNA. The gene codes for an H1 protein of 219 amino acids. 5' and 3' sequence elements are indicated - for example, 5' conserved "TATA" and "CCAAT" sequences and 3' conserved histone gene-specific sequences. From comparison to the λCHO1 H1 gene (Section 5.2.1(b)), the location of the λH1.9/H1.10 H1 gene mRNA 5' and 3' termini can be predicted as shown by vertical arrows (Fig. 5.9).

5.2.3 Sequence Analysis of the pCH11.5E H1 Genes

An 11.5 kb EcoRI restriction fragment from the cosmid clone 6.3C (overlaps λCHO1, Fig. 4.4) was found by R. D'Andrea (1985) to contain two separate regions that were positive to an H1 gene probe. The probe used in this case was the 1.3 kb insert of the M13 recombinant CHE-7R, which contains the entire coding region of the λCHO2 H1 gene (Section 3.2.3(a)). This probe was used to isolate the H1 cDNA clone, cH1.1 (Section 3.2.5). The 11.5 kb EcoRI fragment was subcloned into the plasmid vector pBR325 (D'Andrea, 1985). The fragment contains a cluster of histone genes which represents one of three clusters whose genes are symmetrically ordered around central H3 genes (Section 1.5.3(d); D'Andrea et al., 1985; Figs. 4.4 and 4.19). The recombinant generated from cloning the fragment was named pCH11.5E. A restriction map of the pCH11.5E insert is shown in Figure 4.4 and is presented again in Figure 5.10. Restriction fragments which hybridized with the H1 gene probe (D'Andrea, 1985) were the 0.8 kb EcoRI/SacII and 1.2 kb SacII/SacI fragments from the left-hand end of pCH11.5E and the 1.3 kb SacI/SacII fragment from the right-hand end of pCH11.5E. These fragments are indicated in Figure 5.10.

FIGURE 5.9

DNA sequence of the λH1.9/λH1.10 H1 gene

1497 bp of DNA sequence containing the λH1.9/λH1.10 H1 gene are shown. This sequence reads from and includes the right-hand HindIII site of the 2.1 kb HindIII fragment of λH1.10 (Fig. 5.7). The HindIII site is indicated. The amino acid sequence (in three-letter code; see Fig. 5.5) derived from the H1 protein coding region is given above the DNA sequence. Initiation (ATG) and termination codons (TGA) are marked ***. X represents an undetermined amino acid and - an undetermined nucleotide. The location of the left-hand EcoRI linker (E^L) of λH1.9 is indicated.

Conserved sequences in 5' and 3' flanking regions (Figs. 6.4, 6.5 and 6.6) are labelled as described previously (Fig. 5.5). Proposed locations of mRNA 5' and 3' termini are indicated by vertical arrows.

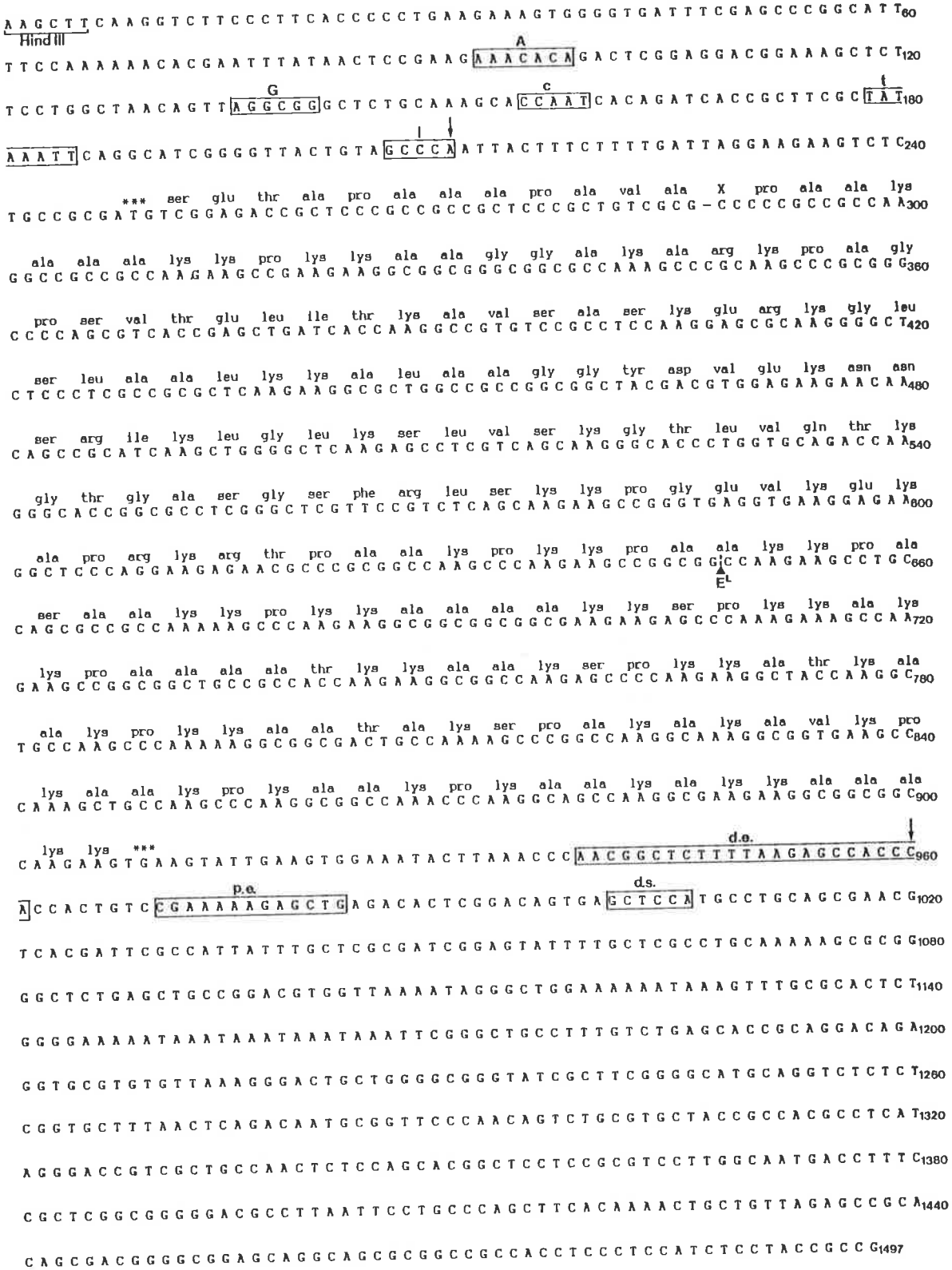
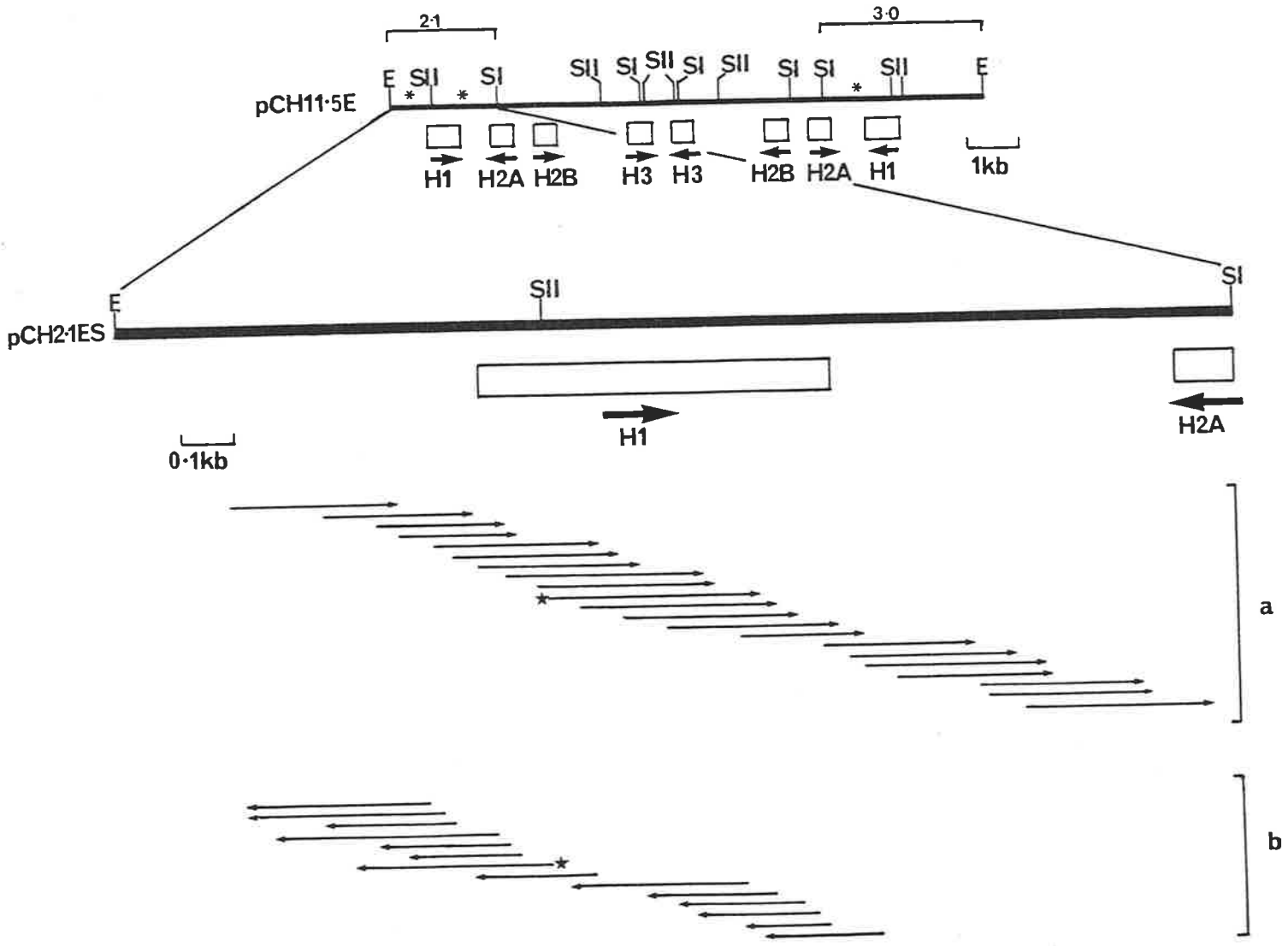

Hind III
A
G
C
t
I
***
ser glu thr ala pro ala ala ala pro ala val ala X pro ala ala lys
ala ala ala lys lys pro lys lys ala ala gly gly ala lys ala arg lys pro ala gly
pro ser val thr glu leu ile thr lys ala val ser ala ser lys glu arg lys gly leu
ser leu ala ala leu lys lys ala leu ala ala gly gly tyr asp val glu lys asn asn
ser arg ile lys leu gly leu lys ser leu val ser lys gly thr leu val gln thr lys
gly thr gly ala ser gly ser phe arg leu ser lys lys pro gly glu val lys glu lys
ala pro arg lys arg thr pro ala ala lys pro lys lys pro ala ala lys lys pro ala
E
ser ala ala lys lys pro lys lys ala ala ala ala lys lys ser pro lys lys ala lys
lys pro ala ala ala ala thr lys lys ala ala lys ser pro lys lys ala thr lys ala
ala lys pro lys lys ala ala thr ala lys ser pro ala lys ala lys ala val lys pro
lys ala ala lys pro lys ala ala lys pro lys ala ala lys ala lys lys ala ala ala
d.e.
lys lys ***
p.e.
d.s.

FIGURE 5.10

Restriction map of the pCH11.5E insert and strategy for sequencing the left-hand pCH11.5E H1 gene

A restriction enzyme map of the insert of the subclone pCH11.5E (Fig. 4.4) is shown at the top of the diagram (data is from D'Andrea, 1985). E, SII and SI represent EcoRI, SacII and SacI restriction sites respectively. H2A, H2B and H3 gene locations were determined by DNA sequencing (Sturm, 1985; D'Andrea, 1985). H1 positive fragments discussed in Section 5.2.3 are indicated with * and the 2.1 kb and 3 kb SacI/EcoRI fragments, subcloned to generate pCH2.1ES and pCH3.0ES respectively (Section 5.2.3(a)), are indicated (sizes in kilobases).

A map of the pCH2.1ES insert is shown below the pCH11.5E map. The locations of H1 and H2A histone genes (indicated by open boxes) are shown and large horizontal arrows show the direction of gene transcription. Sequence generated from directional BAL-31 fragments, reading towards the SacI end of the 2.1 kb fragment (a) and towards the EcoRI end of the fragment (b), is indicated by horizontal arrows. ★ represents sequence generated from 0.8 kb EcoRI/SacII and 1.2 kb SacII/SacI fragments.


2·1
3·0
pCH11·5E
E
SII
SI
SII
SI
SII
SI
SII
SI
SI
SII
E
H1
H2A
H2B
H3
H3
H2B
H2A
H1
1kb
pCH21ES
E
SII
SI
H1
H2A
0·1kb
a
b

(a) Subcloning pCH11.5E right-hand and left-hand H1 genes

pCH11.5E DNA (gift from R. D'Andrea) was digested with SacI and EcoRI and fractionated on a 1% LGT agarose gel (Fig. 5.11(a)). The 2.1 and 3 kb EcoRI/SacI fragments, containing respectively the left and right-hand H1 positive regions of pCH11.5E (Fig.5.10) were isolated and cloned into the plasmid vector pUC18 (Yanisch-Perron et al., 1985; Norrander et al., 1983) that had been double-digested with SacI and EcoRI (Section 2.3.4(b)). Ampicillin resistant colonies were analyzed using the miniscreen procedure (Section 2.3.3(b)). Large-scale DNA preparations (Section 2.3.5) were performed on recombinants containing appropriate inserts. Recombinants containing respectively the 2.1 kb and 3 kb EcoRI/SacI fragments were named pCH2.1ES and pCH3.0ES. Restriction maps of subclone inserts are shown in Figures 5.10 and 5.14.

(b) Sequence analysis of the left-hand H1 gene of pCH11.5E

Gene Location

pCH2.1ES (Fig. 5.10) DNA was digested with SacII, treated to generate blunt-ends (Section 2.3.4(a)) and then cut with either EcoRI or SacI. Digested DNA was fractionated on an LGT agarose gel (Fig. 5.11(b)). The 0.8 kb EcoRI/SacII and the 1.2 kb SacII/SacI fragments were isolated and cloned respectively into SmaI/EcoRI and SmaI/SacI digested mp18 (Norrander et al., 1983) M13 vectors. Sequence generated from these clones (Section 2.3.10) read from the SacII ends of the restriction fragments, and enabled the location of H1 coding sequences as shown in Figure 5.10. The SacII site is located within the 5' portion of the H1 gene coding region.

The complete sequence of the pCH11.5E left-hand H1 gene was determined from BAL-31 generated DNA fragments.

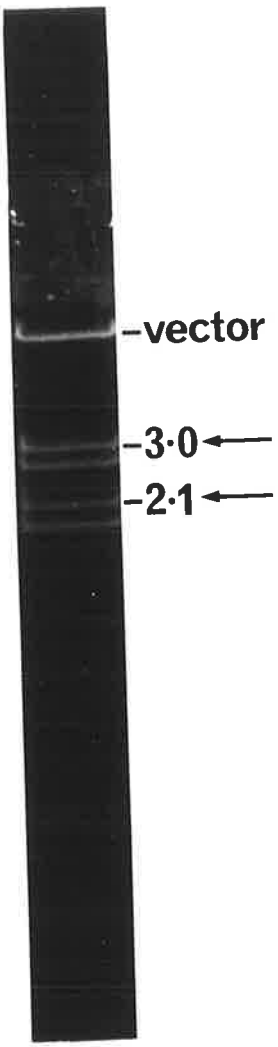
FIGURE 5.11

Digestion of pCH11.5E and pCH2.1ES

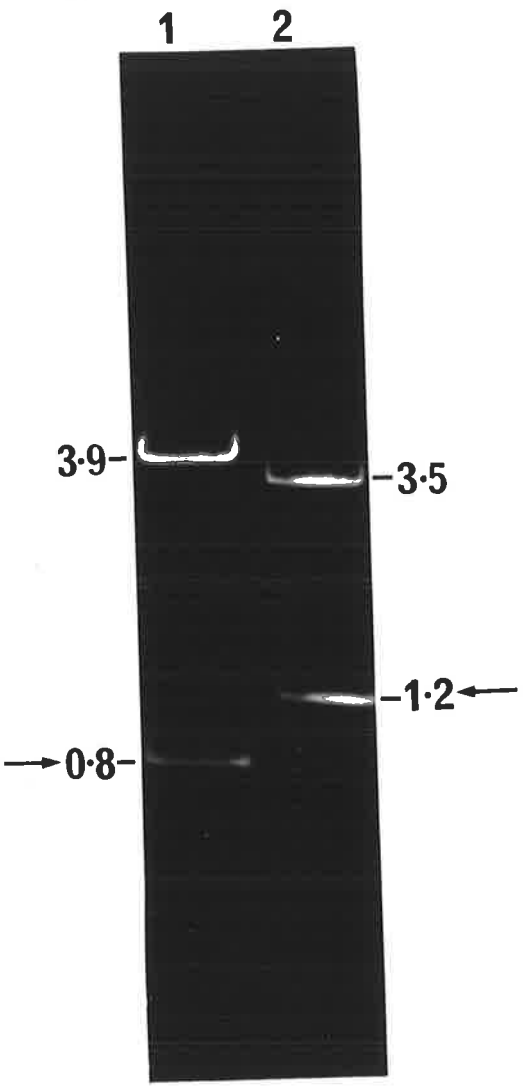
a) pCH11.5E DNA was digested with SacI and EcoRI and fractionated on a 1% vertical LGT agarose gel. The 3 kb and 2.1 kb fragments (marked with horizontal arrows; sizes are in kilobases) were isolated from the gel and subcloned into pUC18 (Section 5.2.3(a)) to generate pCH3.0ES and pCH2.1ES.

b) pCH2.1ES DNA was digested with SacII, treated to generate blunt ends, and then cut with either EcoRI or SacI. DNA was fractionated on a 1% vertical LGT agarose gel. Tracks 1 and 2 represent EcoRI/SacII digested DNA and SacI/SacII digested DNA respectively. The 0.8 kb EcoRI/SacII and 1.2 kb SacII/SacI fragments (indicated by horizontal arrows) were isolated for DNA sequencing (Section 5.2.3(b)).

a.



b.



The complete H1 gene sequence

pCH2.1ES DNA was digested with EcoRI, treated with BAL-31 for various lengths of time (Section 2.3.8(b)), then cut with HindIII (HindIII site is adjacent to the SacI site in the pUC polylinker) and fractionated on a 1% LGT agarose gel (Fig. 5.12). Ten size fractions of treated DNA, between 0.2 and 2 kb, were isolated and cloned into HincII/HindIII double-digested M13 mp19. Due to the similarity in size of the pCH2.1ES insert and vector (2.1 kb and 2.7 kb) there was little separation of insert and vector derived sequences fractionated on the LGT agarose gel (Fig. 5.12). The constructed M13 recombinants were screened with pBR322 DNA (^{32}P-labelled by nick-translation (Section 2.3.7(a)) as described in Section 2.3.9(b), in order to screen out clones containing plasmid vector sequences. Clones negative with the vector probe were sequenced (Section 2.3.10). DNA sequence generated read towards the SacI end of the 2.1 kb EcoRI/SacI insert of pCH2.1ES. Regions of the 2.1 kb fragment sequenced are shown in Figure 5.10.

To produce sequence in the opposite direction, pCH2.1ES DNA was digested first with HindIII, BAL-31 treated and secondary cut with EcoRI. DNA was size fractionated as described above. Isolated DNA was cloned into SmaI (blunt-ended)/EcoRI double-digested M13 mp18, and resultant recombinants were screened for plasmid vector sequences as above. The regions of the 2.1 kb fragment sequenced from vector-negative recombinants are shown in Figure 5.10.

As indicated in Figure 5.10 sequence was generated from most of the 2.1 kb EcoRI/SacI fragment. This DNA contained a single H1 gene and the 3' end of the adjacent H2A gene. The H2A gene had been previously identified and partially sequenced by R. Sturm (1985).

1587 bp of the 2.1 kb EcoRI/SacI fragment from pCH2.1ES are presented in Figure 5.13. The sequence starts approximately 200 bp from the EcoRI

FIGURE 5.12

Generation of BAL-31 fragments from pCH2.1ES

pCH2.1ES DNA was digested with EcoRI, treated with BAL-31 for various lengths of time, made blunt-ended (Section 2.3.8(b)) and then cut with HindIII. Treated DNA was fractionated on a 1% horizontal LGT agarose gel.

The figure shows treated DNA from four BAL-31 time samples. Tracks a., b., c. and d. represent 10, 7, 5 and 2 minute time samples respectively. Each track contains approximately 1 ug of DNA. Sizes of markers are in kilobases.

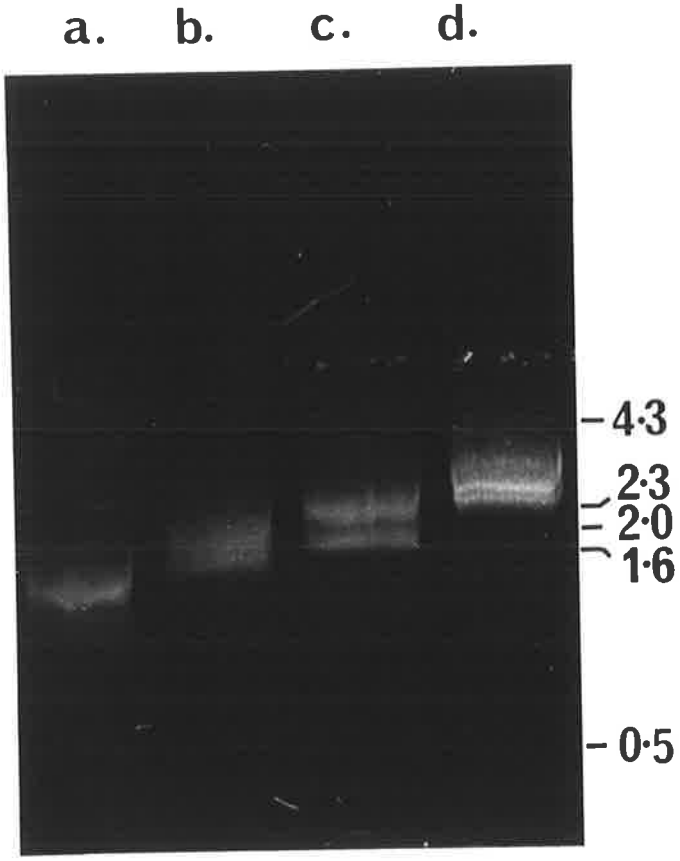
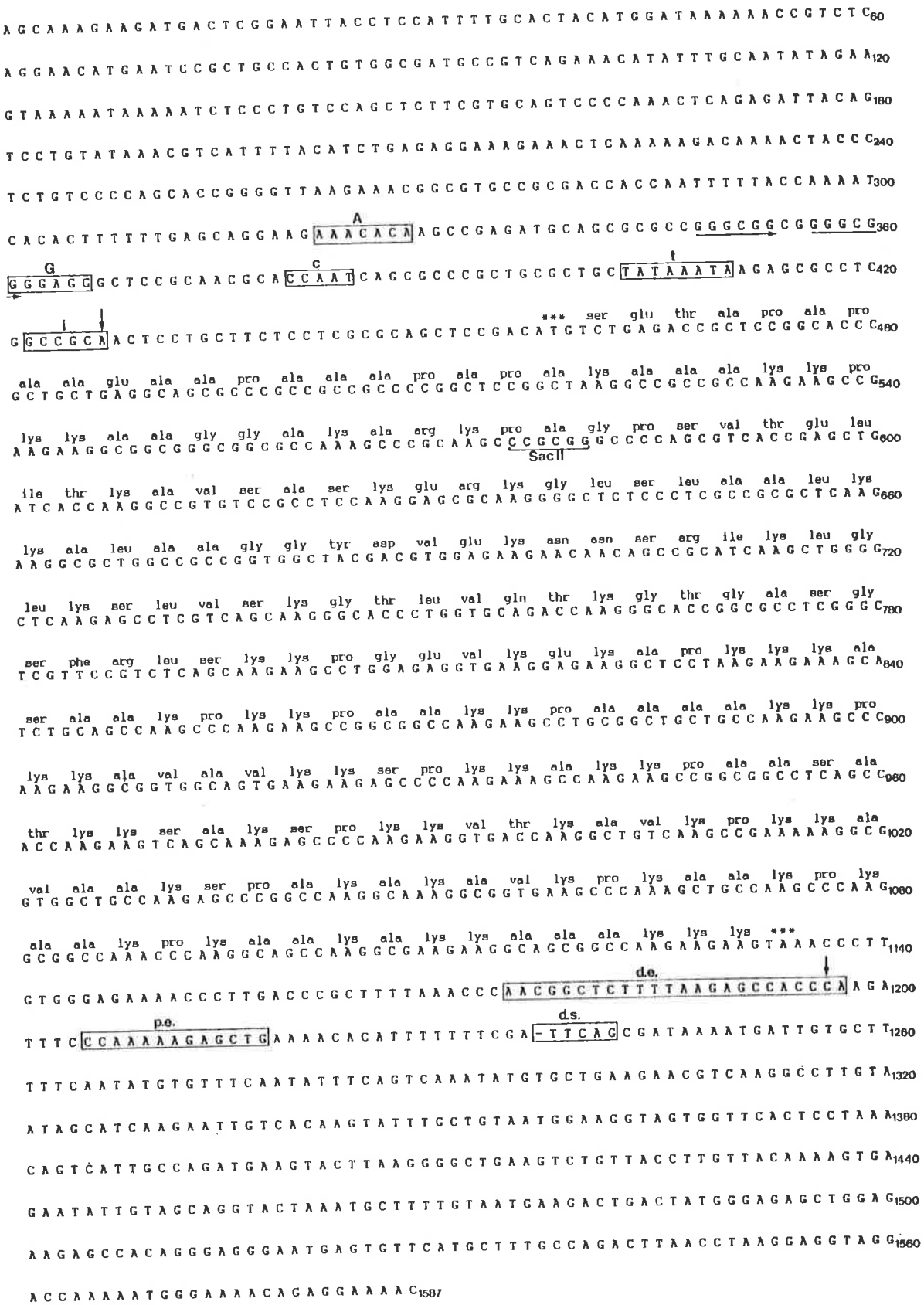

a.
b.
c.
d.
4·3
2·3
2·0
1·6
0·5

FIGURE 5.13

DNA sequence of the left-hand pCH11.5E H1 gene

1587 bp of DNA sequence containing the left-hand pCH11.5E H1 gene are presented. The sequence commences approximately 200 bp from the EcoRI end of the 2.1 kb EcoRI/SacI fragment, used to generate DNA sequence (insert of pCH2.1ES), and reads towards the SacI end (Fig. 5.10). The amino acid sequence (amino acids abbreviated by three-letter code; see Fig. 5.5) derived from the H1 protein coding region is given above the DNA sequence. Initiation (ATG) and termination (TAA) codons are marked ***. - represents an undetermined nucleotide. The SacII site used in sequence determination is indicated.

Conserved sequences in 5' and 3' flanking regions (Figs. 6.4, 6.5 and 6.6) are labelled as described previously (Fig. 5.5). Two G-rich sequences discussed in Section 6.4.2 are indicated with horizontal arrows. Proposed locations of mRNA 5' and 3' termini are indicated with vertical arrows.


A
G
c
t
i
*** ser glu thr ala pro ala pro
ala ala glu ala ala pro ala ala ala pro ala pro ala lys ala ala ala lys lys pro
lys lys ala ala gly gly ala lys ala arg lys pro ala gly pro ser val thr glu leu
SacII
ile thr lys ala val ser ala ser lys glu arg lys gly leu ser leu ala ala leu lys
lys ala leu ala ala gly gly tyr asp val glu lys asn asn ser arg ile lys leu gly
leu lys ser leu val ser lys gly thr leu val gln thr lys gly thr gly ala ser gly
ser phe arg leu ser lys lys pro gly glu val lys glu lys ala pro lys lys lys ala
ser ala ala lys pro lys lys pro ala ala lys lys pro ala ala ala ala lys lys pro
lys lys ala val ala val lys lys ser pro lys lys ala lys lys pro ala ala ser ala
thr lys lys ser ala lys ser pro lys lys val thr lys ala val lys pro lys lys ala
val ala ala lys ser pro ala lys ala lys ala val lys pro lys ala ala lys pro lys
ala ala lys pro lys ala ala lys ala lys lys ala ala ala lys lys lys ***
d.e.
p.e.
d.s.

end of the fragment and reads towards the SacI end. The sequence contains 456 bp of 5' non-coding DNA, 678 bp of coding DNA (including ATG and TAA codons) and 453 bp of 3' non-coding DNA. The coding region codes for an H1 protein of 224 amino acids. 5' and 3' sequence elements and predicted locations of mRNA termini are indicated.

(c) Sequence analysis of the right-hand H1 gene from pCH11.5E

pCH3.0ES (Fig. 5.14) which contains H1 positive sequences from the right-hand end of pCH11.5E, was digested with SacII, blunt-ended (Section 2.3.4(a)) and then cut with either EcoRI or SacI. The 1.3 kb SacI/SacII fragment, the 0.15 kb SacII fragment and the 1.55 kb SacII/EcoRI fragment were isolated from LGT agarose and cloned respectively into SmaI/SacI digested mp18, SmaI digested mp19 and SmaI/EcoRI digested mp18 M13 vectors. Directional sequencing of cloned fragments enabled the location of an H1 coding region. The regions of the pCH3.0ES insert sequenced are indicated in Figure 5.14.

560 bp of generated sequence are presented in Figure 5.15. The sequence reads towards the SacI end of the 3 kb insert of pCH3.0ES and contains 77 bp of 5' non-coding DNA and 483 bp of coding DNA (including ATG), which codes for the first 160 amino acids of an H1 protein. The two SacII sites in the pCH3.0ES insert (Fig. 5.14) are indicated on the DNA sequence. 5' sequence elements are also shown.

Given the location of the pCH3.0ES H1 gene and the size of coding regions for other sequenced chicken H1 genes (657 - 678 bp) it can be estimated that the 3' end of the pCH3.0ES H1 gene is separated from the 3' end of the adjacent H2A gene (Sturm, 1985; Figs. 5.10 and 5.14) by approximately 650 bp. It is therefore unlikely that the 1.3 kb H1 positive SacI/SacII fragment contains H1 coding sequences in addition to those already described.

FIGURE 5.14

Sequencing strategy of the right-hand pCH11.5E H1 gene

A restriction map of the insert of pCH3.0ES, containing the right-hand H1 gene of pCH11.5E, is shown. SI, SII and E represent SacI, SacII and EcoRI restriction sites respectively. The location of H1 and H2A genes are shown (indicated by open boxes) and directions of transcription indicated (by horizontal arrows).

Arrows marked a, b and c indicate regions of insert DNA sequenced from the 1.3 kb SacI/SacII, the 0.15 kb SacII and the 1.55 kb SacII/EcoRI restriction fragments respectively (Section 5.2.3(c)).

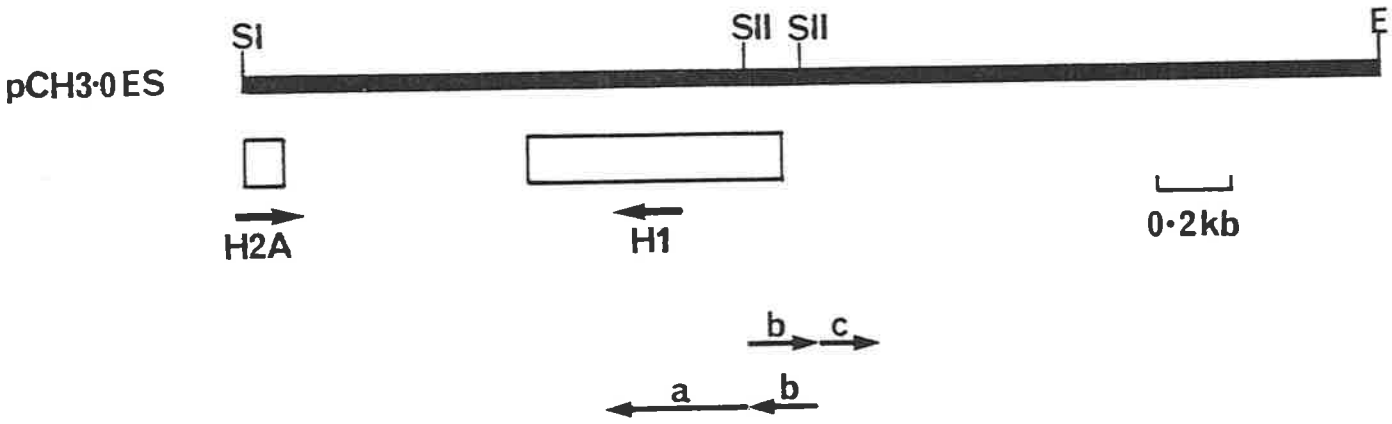

pCH3·0 ES
SI
SII SII
E
H2A
H1
0·2kb
b
c
a
b

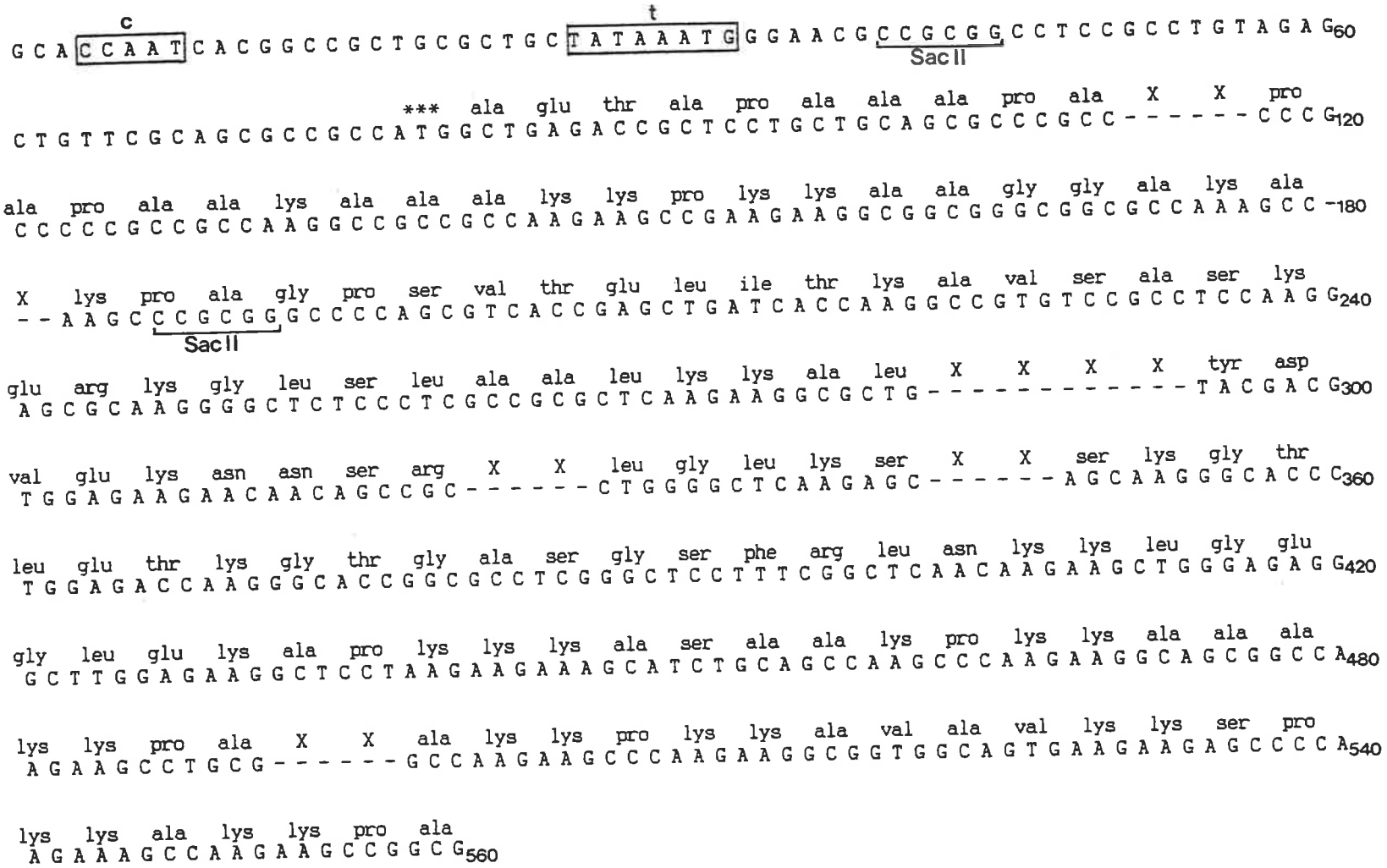
FIGURE 5.15

DNA sequence of the right-hand pCH11.5E H1 gene

The 3 kb EcoRI/SacI fragment containing the right-hand pCH11.5E H1 gene was subcloned into a pUC vector (generating pCH3.0ES). DNA fragments were isolated from this subclone for sequence determination (Fig. 5.14).

560 bp of generated sequence containing a portion of the right-hand pCH11.5E H1 gene are shown. The sequence reads through both SacII sites of the 3 kb EcoRI/SacI fragment, towards the SacI end of the fragment. - represents undetermined nucleotides.

The amino acid sequence (in three-letter code; Fig. 5.5) derived from the H1 protein coding region is shown above the DNA sequence. *** marks the initiation codon (ATG). X represents an undetermined amino acid due to lack of DNA sequence. "CCAAT" (c) and "TATA" (t) sequences are indicated.


c
t
SacII
SacII

5.2.4 The λCHO3 H1 Gene

DNA from λCHO3, positive to an H1 probe (D'Andrea, 1985), is also contained within the overlapping genomal glones λCH10a and λCH3d (Sugarman et al., 1983; Fig. 4.4). The overlapping H1 gene-containing regions of λCHO3 and λCH10a are shown in Figure 5.16. H1 positive sequences in λCHO3 were localized to the 0.5 kb EcoRI/SacI fragment, at the left-hand end of the clone, by R. D'Andrea (1985). Sequence analysis performed by J. Powell (1984) of the pCH5.0E subclone (Figs. 4.4 and 5.16) from λCHO3 located the H1 coding sequences as shown in Figure 5.16.

Sugarman et al. (1983) have found no H1 positive sequences to the left of the BamHI site, in λCH10a, that is 3' to the identified H1 gene (see Fig. 5.16). Given the size of chicken H1 gene coding regions, this localization of H1 sequences therefore prohibits the existence of additional H1 genes 3' to the one located by J. Powell.

289 bp of sequence generated are presented in Figure 5.17. This sequence reads through the SacI site shown in Figure 5.16 towards the left-hand end of the pCH5.0E insert (the left-hand EcoRI linker end of λCHO3). The DNA contains 85 bp of 5' non-coding DNA and 204 bp of coding DNA (including ATG), which codes for the first 67 amino acids of an H1 protein. The SacI restriction site and 5' sequence elements are indicated.

5.2.5 The λCHO2 H1 Gene

The single H1 gene from λCHO2 was initially identified by DNA sequencing in this laboratory (A. Robins and S. Clark, unpublished). The complete sequence of this gene was later determined by Sugarman et al. (1983). These workers isolated the gene from the chicken genomal clone λCH1a whose insert overlaps that of λCHO2 (both clones are derived from the library constructed by Dodgson et al., 1979). Comparison of sequences of the H1 and H2B genes and their flanking sequences, from λCHO2 and λCH1a

FIGURE 5.16

Location of the λCH03 H1 gene

The overlapping H1 gene-containing regions of the genomal clones λCH03 and λCH10a (Fig. 4.4) are shown. A map of the insert of the λCH03 subclone, pCH5.0E (Fig. 4.4), is also shown. E, H, B and SI represent EcoRI, HindIII, BamHI and SacI restriction sites respectively. E^L represents sites of EcoRI linkers. The location of the H1 gene (indicated with an open block) was determined by J. Powell (1984) by sequence analysis of pCH5.0E. The direction of transcription of the gene is indicated by the horizontal arrow. The BamHI site discussed in Section 5.2.4 is marked with an *.

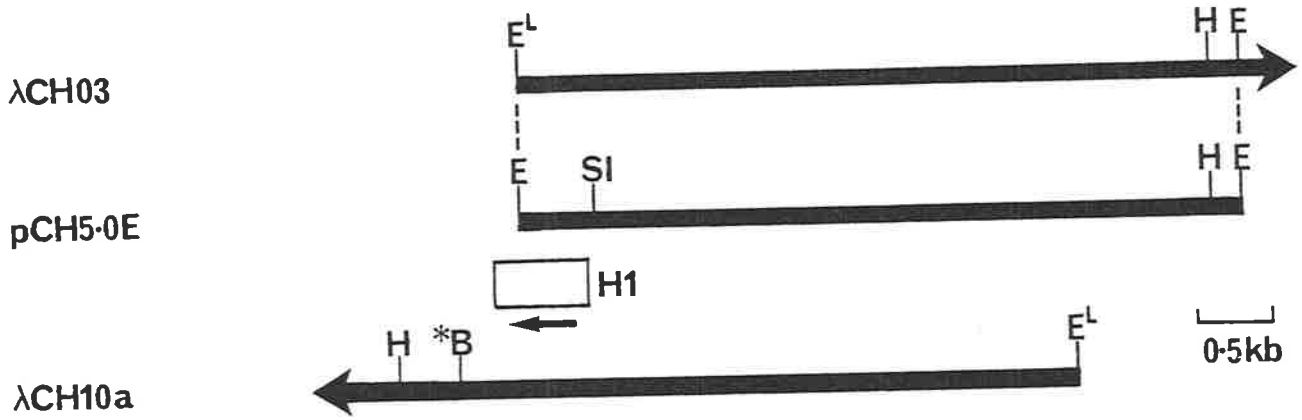
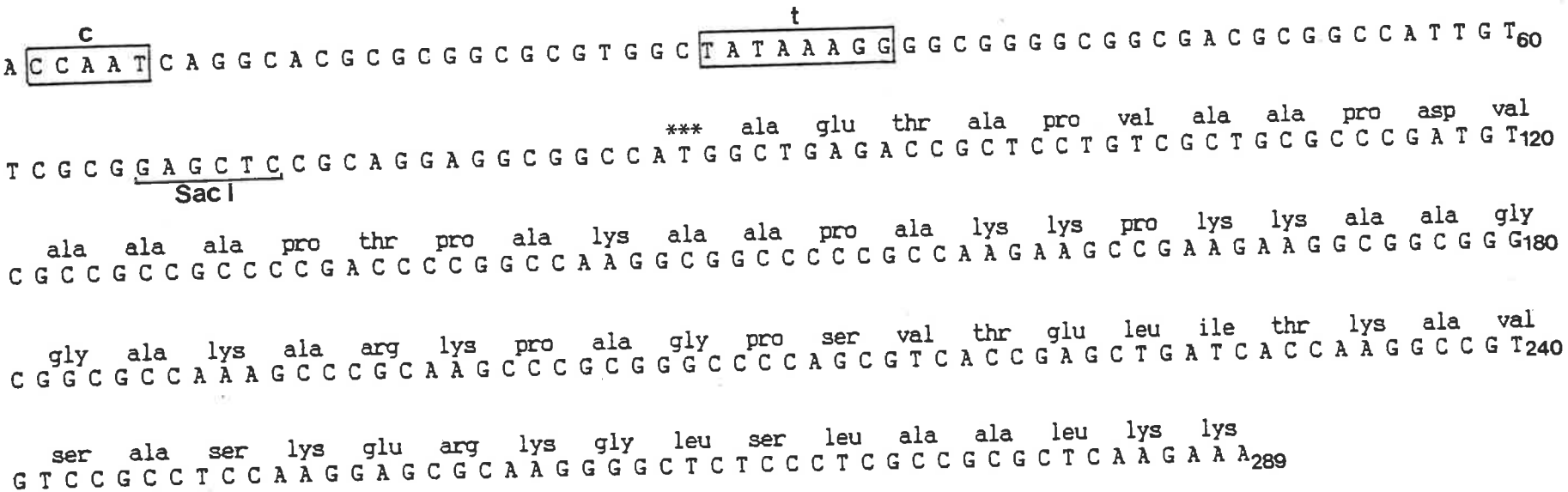

λCH03
pCH5·0E
λCH10a
E^L
H
E
E
SI
H
E
H1
H
*B
E^L
0·5kb

FIGURE 5.17

DNA sequence of the λCH03 H1 gene

289 bp of DNA sequence containing a portion of the λCH03 H1 gene are shown. The sequence was generated by J. Powell (1984). This sequence reads through the SacI site, shown in Figure 5.16, towards the left-hand EcoRI linker of λCH03.

The amino acid sequence (in three-letter code; see Fig. 5.5) derived from the H1 protein coding region is shown above the DNA sequence. *** marks the initiation codon (ATG). The SacI site and "CCAAT" (c) and "TATA" (t) sequences are indicated.


c
t
A C C A A T C A G G C A C G C G C G G C G C G T G G C T A T A A A G G G G C G G G G C G G C G A C G C G G C C A T T G T 60
*** ala glu thr ala pro val ala ala pro asp val
T C G C G G A G C T C C G C A G G A G G C G G C C A T G G C T G A G A C C G C T C C T G T C G C T G C G C C C G A T G T 120
Sac I
ala ala ala pro thr pro ala lys ala ala pro ala lys lys pro lys lys ala ala gly
C G C C G C C G C C C C G A C C C C G G C C A A G G C G G C C C C C G C C A A G A A G C C G A A G A A G G C G G C G G G 180
gly ala lys ala arg lys pro ala gly pro ser val thr glu leu ile thr lys ala val
C G G C G C C A A A G C C C G C A A G C C C G C G G G C C C C A G C G T C A C C G A G C T G A T C A C C A A G G C C G T 240
ser ala ser lys glu arg lys gly leu ser leu ala ala leu lys lys
G T C C G C C T C C A A G G A G C G C A A G G G G C T C T C C C T C G C C G C G C T C A A G A A A 289

(A. Robins and S. Clark, unpublished; Sugarman et al., 1983; Grandy et al., 1982; Harvey et al., 1982) confirmed that these two clones are true overlaps. The exact location of the λCH1a/λCH02 H1 gene is indicated in DNA presented in Figures 4.17 and 4.19.

The published λCH1a/λCH02 H1 gene sequence is presented in Figure 5.18, for comparison to the H1 gene sequences described in Sections 5.2.1 to 5.2.4. The 1098 bp of DNA sequence contains 208 bp of 5' non-coding DNA, 657 bp of coding DNA (including ATG and TAA codons) and 233 bp of 3' non-coding DNA. The coding region codes for an H1 protein of 217 amino acids.

As discussed in Section 3.2.5(b) the sequence of the 536 bp insert of the H1 cDNA clone, cH1.1 (contains 510 bp of coding DNA, including termination codon, and 26 bp of 3' non-coding DNA), is identical to the sequence of the equivalent region of the λCH1a/λCH02 H1 gene, except for a single silent third base change within the coding region (Fig. 3.7). The relative location of the cH1.1 insert sequence on the λCH1a/λCH02 sequence is indicated in Figure 5.18. The base change is also indicated. The coding sequences in cH1.1 represent those coding for amino acids 49 to 217. It was suggested (Section 3.2.5(b)) that cH1.1 was derived from a transcript of the λCH1a/λCH02 H1 gene.

5' and 3' sequence elements are also indicated in Figure 5.18 as well as the location of 5' mRNA termini, mapped using RNA from 4.5 day chicken embryos (Sugarman et al., 1983; discussed in Section 5.2.1(b)), and the proposed location of 3' mRNA termini.

5.2.6 Northern Analysis of Chicken H1 mRNA

Histone mRNAs are generally found to be non-polyadenylated (Section 1.6.2(a)) in contrast to the mRNAs derived from most other RNA polymerase II transcribed genes (Nevins, 1983; Proudfoot, 1984). To investigate the

FIGURE 5.18

DNA sequence of the λCHO2 H1 gene

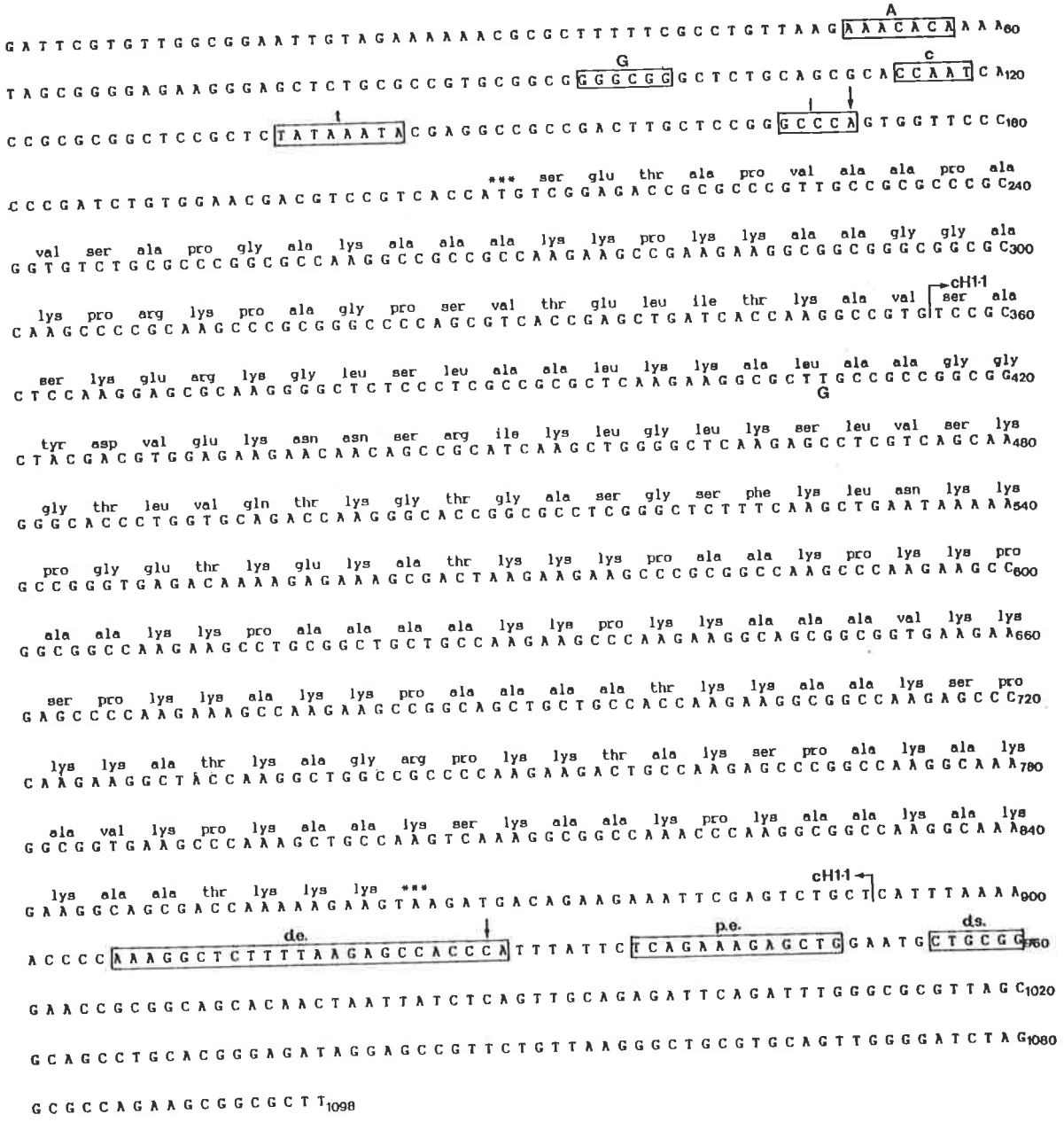
The H1 gene from λCHO2 (Fig. 4.19) was identified initially by DNA sequencing in this laboratory (A. Robins and S. Clark, unpublished). The complete sequence of this gene was later determined by Sugarman *et al.* (1983 ; gene was isolated from the insert of the genomal clone λCH1a which overlaps that of λCHO2; discussed in Section 5.2.5).

The 1098 bp sequence, generated by Sugarman *et al.* (1983),containing the λCHO2 H1 gene, is shown.

The amino acid sequence (in three-letter code; Fig. 5.5) derived from the H1 protein coding region is shown above the DNA sequence. *** marks initiation (ATG) and termination (TAA) codons.

The relative location of the insert of the H1 cDNA clone (cH1.1) is shown on the DNA sequence. The cH1.1 sequence differs from the λCHO2 by a single base change shown below the λCHO2 DNA sequence (Section 3.2.5(b); Fig. 3.7).

Conserved 5' and 3' sequences are shown labelled as described in Figure 5.5. The location of 5' mRNA termini (as determined by Sugarman *et al.* (1983)) and proposed 3' termini are indicated by vertical arrows.


A
G
c
t
*** ser glu thr ala pro val ala ala pro ala
val ser ala pro gly ala lys ala ala ala lys lys pro lys lys ala ala gly gly ala
lys pro arg lys pro ala gly pro ser val thr glu leu ile thr lys ala val ser ala
cH1-1
ser lys glu arg lys gly leu ser leu ala ala leu lys lys ala leu ala ala gly gly
G
tyr asp val glu lys asn asn ser arg ile lys leu gly leu lys ser leu val ser lys
gly thr leu val gln thr lys gly thr gly ala ser gly ser phe lys leu asn lys lys
pro gly glu thr lys glu lys ala thr lys lys lys pro ala ala lys pro lys lys pro
ala ala lys lys pro ala ala ala ala lys lys pro lys lys ala ala ala val lys lys
ser pro lys lys ala lys lys pro ala ala ala ala thr lys lys ala ala lys ser pro
lys lys ala thr lys ala gly arg pro lys lys thr ala lys ser pro ala lys ala lys
ala val lys pro lys ala ala lys ser lys ala ala lys pro lys ala ala lys ala lys
cH1-1
lys ala ala thr lys lys lys ***
de.
p.e.
ds.

nature of chicken H1 mRNAs, polyadenylated (A+) and non-polyadenylated (A-) RNAs from chicken embryos and from an adult tissue were analyzed by Northern analysis (Section 2.3.11(a)).

A+ and A- RNAs from 3 day chicken embryos, 4 day chicken embryos and adult anaemic bone marrow (Williams, 1970) were electrophoresed on 1.5% vertical agarose gels (in NaH_2PO_4) and transferred to nitrocellulose as described in Section 2.3.11(a). RNA preparations were a gift from J. Whiting. Embryo RNA was prepared by the method of Chirgwin et al. (1979), while the adult RNA was prepared by the method of Brooker et al. (1980). RNA was fractionated by passaging twice through an oligo(dT)-cellulose column.

It is not known how many of the six characterized chicken H1 genes are expressed in the embryo and adult tissues used here. In the embryo it is known, however, that at least the λCHO1 and λCHO2 H1 genes are expressed, as discussed previously (Sections 5.2.1(b) and 5.2.5). Other H1 genes are probably also expressed in the embryo as indicated by S1 analysis data presented in Figure 5.6 (see Section 5.2.1(b)).

Nitrocellulose filters containing A+ and A- RNA samples, were probed with the insert (^{32}P-labelled by nick-translation) of the H1 cDNA clone cH1.1 (Section 3.2.5) as described in Section 2.3.11(a). Resultant autoradiograms are shown in Figure 5.19. As is indicated, a single band of approximately 760 bases was detected in each of the A- RNA samples. No hybridization was detected in tracks containing A+ RNA. The size of the detected RNA is consistent with the predicted sizes of transcripts from the four H1 genes, which have been completely sequenced. The predicted sizes of transcripts from the λCHO1, λCHO2, λH1.9/λH1.10 and pCH11.5E (left-hand) H1 genes are respectively 760, 758, 751 and 770 bases. Transcripts of these sizes would not be resolved on the gel system used. Sugarman et al. (1983) also detected a single band of RNA of similar size to that detected here when analyzing RNA from 5 day chicken red blood cells with an homologous H1 probe.

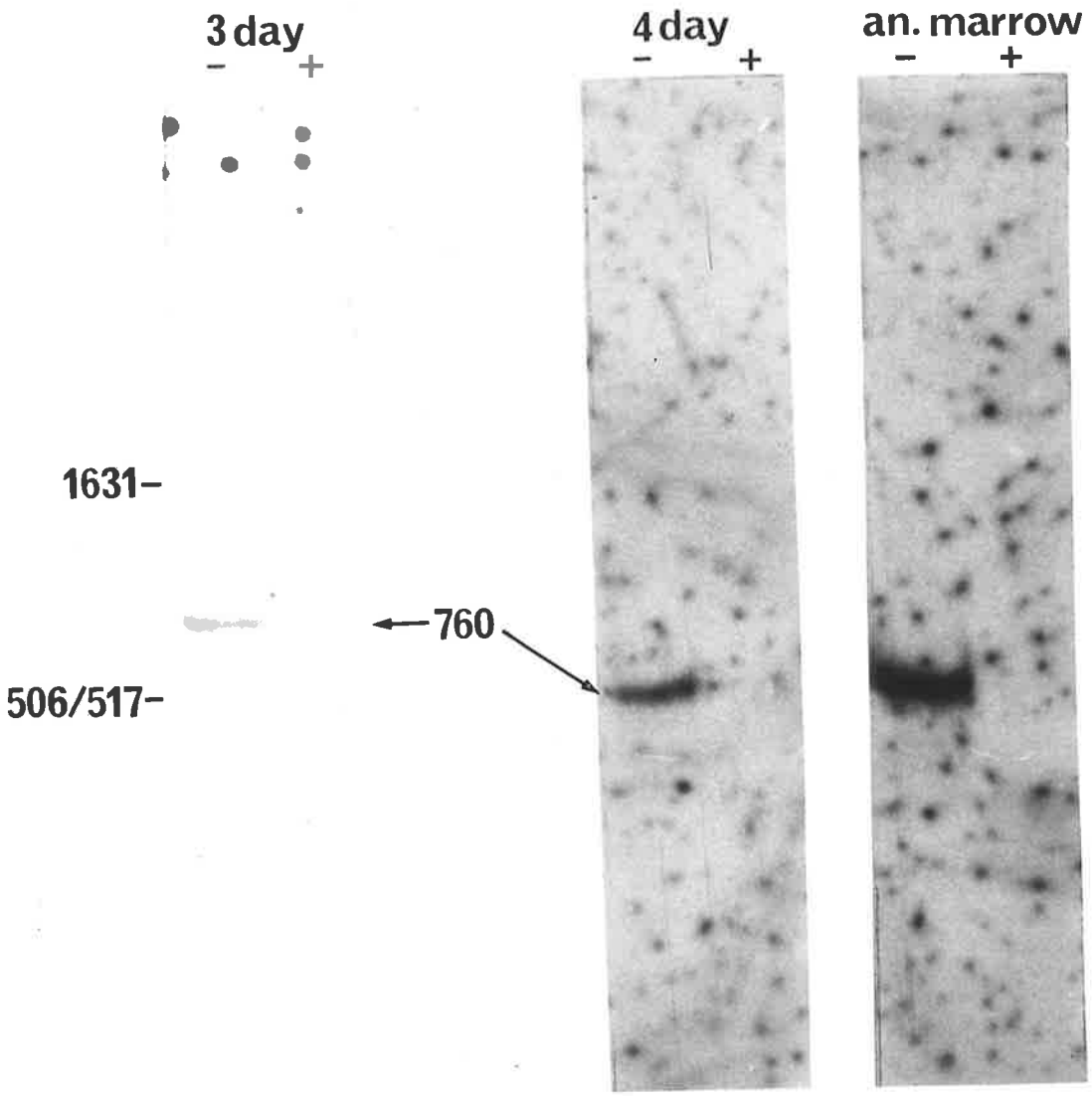
FIGURE 5.19

Northern analysis of chicken H1 mRNA

Polyadenylated (A+; 1 ug/sample) and non-polyadenylated (A-; 10 ug/sample) RNAs from, 3 day chicken embryos, 4 day chicken embryos and adult chicken anaemic bone marrow (gift from J. Whiting, prepared as described in Section 5.2.6) were glyoxylated, electrophoresed on 1.5% vertical agarose gels (in NaH_2PO_4), and transferred to nitrocellulose as described previously (Section 2.3.11(a)). Markers (glyoxylated HinfI digested pBR322) were also run on gels.

Filters were probed with the insert of the H1 cDNA clone, cH1.1 (^{32}P-labelled by nick-translation; approx. 10^6 dpm/gel track), as described in Section 2.3.11(a) . Filters were washed (final wash in 0.5 x SSC;0.1% (w/v) SDS at 65°C) and autoradiographed for several days.

Resultant autoradiograms are shown. Tracks marked 3 day and 4 day represent embryo RNAs. Tracks marked an. marrow represent adult anaemic bone marrow RNAs. - and + represent respectively A- and A+ RNA samples. Marker and RNA sizes are in bases.


3 day
–
+
4 day
–
+
an. marrow
–
+
1631–
760
506/517–

It has recently been reported that H1 mRNAs are also non-polyadenylated in dividing and non-dividing chicken myogenic cells (Winter, 1984; Winter et al., 1985b). The production of non-polyadenylated transcripts from the chicken H1 genes (at least those being expressed in the tissues examined), observed here and in myogenic cells, is consistent with the presence of histone gene-specific conserved elements (Section 1.6.2(a))in the 3' non-coding DNA of the chicken H1 genes (where data available, Figs. 5.5, 5.9, 5.13 and 5.18). Histone genes, such as the replication-independent chicken H2A.F, H5 and H3.3 genes that produce polyadenylated mRNAs, lack such 3' conserved elements. In addition to the H1 genes, clustered chicken core histone genes (Figs. 4.4, 4.19 and 4.26) also contain the histone gene-specific 3' elements and produce non-polyadenylated transcripts (Harvey et al., 1983; R. Harvey and L. Coles, unpublished).

Sugarman et al. (1983) have previously suggested that the clustered chicken H1 and core histone genes may be "embryo-specific". The detection of H1 sequences in adult RNA (Fig. 5.19) with the cH1.1 H1 probe (used to isolate the "full complement" of clustered chicken H1 genes; Chapter 4) suggests that this is not the case. Clustered chicken core histone genes have also been shown, in this laboratory to be expressed in an adult tissue-derived cell line (Harvey et al., 1983; Dalton et al., 1985; S. Dalton and J. Coleman, unpublished). The amount of histone mRNA detected in any particular RNA sample is, no doubt, related to the rate of cell division in the tissue from which the RNA was derived (Section 1.6.3(a)).

5.2.7 Discussion

Sequences have been derived from the six chicken H1 genes, located in DNA shown in Figures 4.4, 4.19and 4.26. Four of these genes have been completely sequenced (λCHO1, λH1.9/λH1.10, pCH11.5E left-hand and λCHO2 H1

genes), while the remaining two genes (λCHO3 and pCH11.5E right-hand H1 genes) have been partially sequenced.

The chicken H1 genes have features typical of most sequenced histone genes, including the core histone genes with which they are clustered, viz., there is no evidence for introns in the H1 genes, they contain the 5' conserved sequences of RNA polymerase II transcribed genes, and the 3' conserved histone gene-specific sequences.

The 5' and 3' termini of mRNAs from the λCHO1 H1 gene were determined. 5' termini map to a sequence 5' GCCCA 3' (Sturm, 1985), while 3' termini map to within the 5' ACCCA 3'sequence found at the 3' end of the histone gene-specific dyad element, and hence in a similar position to other mapped histone mRNA 3' termini.

Also, a feature common to most histone genes, work described here and by others (Winter, 1984; Winter et al., 1985b) suggests that the chicken H1 gene mRNAs are not polyadenylated.

CHAPTER 6

FINAL DISCUSSION

6.1 H1 GENE NUMBER AND ORGANIZATION

The characterization of chicken histone DNA, described in this thesis and by others (D'Andrea et al., 1985), resulted in the identification of six regions of DNA containing H1 coding sequences. Subsequent sequence analysis, described in Chapter 5, revealed that each region of DNA contained a single H1 gene. Four of the six chicken H1 genes are contained within DNA spanning the genomal clone λCH01 (Fig.4.4), one is contained within DNA spanning λCH02 (Fig. 4.19), while the sixth H1 gene is contained within DNA covered by λH1.9 and λH1.10 (Fig. 4.26). Data from Southern analysis of chicken genomal DNA suggests that the located genes represent the full complement of H1 genes in the chicken genome (Section 4.2.6).

All the chicken H1 genes identified are intermingled with core histone genes, in contrast to the gene for the H1-related, chicken H5 protein which exists as a "solitary" histone gene (Krieg et al., 1983; Ruiz-Carrillo et al., 1983; Section 1.5.3(d)). Other solitary chicken histone genes are the H2A.F (Robins et al., 1985) and H3.3 genes (Brush et al., 1985; Engel et al., 1982). None of these genes appear to be linked to each other or to four regions of non-overlapping DNA containing clustered chicken histone genes (D'Andrea et al., 1985; D'Andrea, 1985; discussed in Sections 1.5.3(d) and 4.2.5(c)). The isolation of H2A.F cDNA sequences from a chicken embryo cDNA library was discussed in Chapter 3.

The two types of chicken histone gene, the solitary and clustered genes, differ both in their structure and expression (Section 1.5.3(d)). Segregation of differentially expressed histone genes is also observed in the sea urchin, viz., as observed for the genes coding for the early and late sea urchin histone variants (Section 1.5.2(a)). The generation of solitary histone genes in the chicken may have occurred by mechanisms similar to those involved in the generation of the isolated orphon histone genes found in sea urchin and Drosophila (Maxson et al., 1983b;

Childs et al., 1981; Liebermann et al., 1983; Sections 1.5.2(a) and 1.5.2(b)).

The characterized regions of DNA containing clustered chicken histone genes probably also contain the full complement of clustered core histone genes (D'Andrea et al., 1985) as determined from data generated from Southern analysis of genomal DNA (Ruiz-Carrillo et al., 1983; D'Andrea, 1985; Harvey, 1982). In addition to the six H1 genes, ten H2A, eight H2B, ten H3 and eight H4 genes have been located, giving a total of forty-two histone genes in 175 kb of DNA. The numbers of the histone genes are hence maintained in balanced proportions, even though they are not organized into a quintet structure as is observed in organisms with a high histone gene copy number (Section 1.5.5).

Despite the lack of order of chicken histone genes, compared to that observed in organisms such as the sea urchin, Drosophila, newt and Xenopus (Section 1.5), chicken histone genes do exhibit preferred arrangements of genes. One of these preferred arrangements involves the H1 genes, viz., five out of the six genes are associated with an H2A and an H2B gene (discussed in Section 4.2.7). Other features of chicken histone DNA are divergently transcribed H2A/H2B gene pairs, H3/H4 gene associations and the presence of three clusters of genes symmetrically ordered around central H3 genes (D'Andrea et al., 1985; Section 1.5.3(d)). The H2A/H2B gene pairs have been shown to contain overlapping promoters (Sturm, 1985). The relevance of other preferred gene arrangements to the expression of the associated genes is yet to be determined.

6.2 EACH H1 GENE CODES FOR A UNIQUE H1 VARIANT

Four of the six H1 genes have been completely sequenced as discussed in Chapter 5. These genes are the λCHO1, λH1.9/λH1.10, pCH11.5E left-hand and λCHO2 H1 genes. The locations of these genes are shown in Figures 4.4, 4.19

and 4.26 and their DNA sequences are presented respectively in Figures 5.5, 5.9, 5.13 and 5.18. Partial DNA sequences were determined for the pCH11.5E right-hand and λCH03 H1 genes as presented in Figures 5.15 and 5.17. The complete λCH02 H1 gene sequence was determined by Sugarman et al. (1983) while the λCH03 H1 gene sequence was determined by J. Powell (1984). For convenience the λCH01, λH1.9/λH1.10, pCH11.5E left-hand, λCH02, pCH11.5E right-hand and λCH03 H1 genes will be called respectively the 01, .10, 11L, 02, 11R and 03 H1 genes. There was no evidence for introns within any of these H1 genes.

The sequence of the H1 cDNA clone, cH1.1 (Section 3.2.5), used throughout this thesis as an H1 gene-specific probe, differs from the λCH02 H1 gene sequence by only a single silent third base change and hence will not be discussed further. The cH1.1 insert is most probably derived from a transcript of the λCH02 H1 gene.

Each chicken H1 gene codes for a different H1 protein. The 01, 02, .10 and 11L genes code respectively for proteins of 218, 217, 219 and 224 amino acids. The partial sequences of the 11R and 03 genes codes respectively, for the first 160 and the first 67 amino acids of an H1 protein. Both the 11R and 03 gene sequences are tentative, but enough of these sequences have been verified to determine that they code for unique H1 variants. Due to the tentative nature of the 11R and 03 gene sequences only the 01, 02, .10 and 11L proteins will be used in comparative analyses. The 11R and 03 proteins are similar in sequence to the proteins derived from the completed H1 genes.

Figure 6.1 shows the 02, .10 and 11L H1 protein sequences compared to the 01 H1 protein. The 02 protein differs from the 01 by 17 amino acids and has 1 amino acid deletion. The .10 protein differs by 14 amino acids, has 2 amino acid insertions and 1 deletion. The .10 protein also has an undetermined amino acid marked as an X (see Fig. 5.9). The 11L protein has 14 amino acid changes, 7 insertions and 1 deletion. Nucleotide sequences

FIGURE 6.1

Chicken H1 protein sequences

The amino acid sequences derived from the H1 protein coding regions of the λCHO2 (O2) H1 gene, the λH1.9/λH1.10 (.10) H1 gene and the pCH11.5E left-hand (11L) H1 gene (Figs. 5.18, 5.9 and 5.13) are shown compared to the amino acid sequence derived from the λCHO1 (O1) H1 gene (Fig. 5.5).

The λCHO1 H1 protein sequence is presented above the other protein sequences, and amino acids are abbreviated by the one-letter code (D, aspartic acid; N, asparagine; T, threonine; S, serine; E, glutamic acid; Q, glutamine; P, proline; G, glycine; A, alanine; V, valine; I, isoleucine; L, leucine; Y, tyrosine; F, phenylalanine; K, lysine; R, arginine).

Where the λCHO2, λH1.9/λH1.10 and pCH11.5E left H1 gene-derived protein sequences differ from the λCHO1 sequence the amino acid change is indicated, but, where sequences are the same * is presented. Gaps (-) are introduced into sequences for maximum alignment. X in the λH1.9/λH1.10 sequence represents an undetermined amino acid (Fig. 5.9). Numbering is for the λCHO1 sequence. Total protein lengths are indicated at the end of sequences.

The start and end of H1 protein globular domains, as defined by Allan *et al*. (1980), are indicated with arrowheads.

```
                                                                                      ▼                   50
01   S E T A P A - - - - A A P - Ḋ A P A P G A K A A Ȧ K K P K K A A G G Ȧ K A R K P A G P S V̇ T E L I T K A V S Ȧ S K E R K
02   * * * * * V - - - - * * * - A V S * * * * * * * * * * * * * * * * * * * P * * * * * * * * * * * * * * * * * * * * * * *
·10  * * * * * * - - - - * * * - A V A X * A * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
11L  * * * * * * P A A E * * * A A * * * * - * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

                                                                                  100           ▼
     G L S L Ȧ A L K K A L A A G Ġ Y D V E K N N S R İ K L G L K S L V S K̇ G T L V Q T K G T Ġ A S G S F R L N K K̇ P G E V K
     * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * K * * * * * * * T *
     * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * S * * * * * * *
     * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * S * * * * * * *

                                                                    150
     E K A P Ṙ K R A T A A K P K K̇ P A A K K P A A A Ȧ K K P K K A A A V K̇ K S P K K A K K P Ȧ A A A T K K A A K Ṡ P K K A A
     * * * T K * K - P * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * T
     * * * * * * * T P * * * * * * * * * * * * * S * * * * * * * * * * A * * * * * * * * * * * * * * * * * * * * * * * * * T
     * * * * K * K * S * * * * * * * * * * * * * * * * * * * * * * V * * * * * * * * * * * * * * S * * * * S * * * * * * V T

                                                     200
     K A G R Ṗ K K A - - A K S P A K Ȧ K A V K P K A A K Ṗ K A T K P K A A K Ȧ K K T A A K K K 218
     * * * * * * * T - - * * * * * * * * * * * * * * * * S * * A * * * * * * * * * A * T * * * 217
     * * A K * * * * A T * * * * * * * * * * * * * * * * * * * A * * * * * * * * * A * * * * - 219
     * * V K * * * * V A * * * * * * * * * * * * * * * * * * * A * * * * * * * * * A * * * * * 224
```

of the 02, .10 and 11L H1 protein coding regions differ from those of the 01 gene by approximately 8%, 6.3% and 7.7% respectively.

The H1 proteins coded by the six chicken H1 genes are quite distinct in sequence from the H1-related chicken H5 protein (Briand et al., 1980) which is expressed only in erythroid cells, largely replacing other H1 molecules. The accumulation of H5 correlates with the increased compaction of chromatin during erythropoiesis (Section 1.3.3(c)). H5 also appears to bind more tightly to chromatin than other H1 molecules (Thomas et al., 1983; Thomas, 1984). H5 is hence probably a functionally distinct H1 variant. The gene for this protein has been isolated and investigated in this laboratory and by others (Krieg et al., 1982a, 1982b, 1983; Wigley et al., 1985; Dalton et al., 1985; Ruiz-Carrillo et al., 1983; Section 1.5.3(d)).

Separation of chicken H1 proteins from erythrocytes on one and two-dimensional gel systems and by column chromatography (Powell, 1984; this laboratory) suggested that there may be six different chicken H1 subtypes, excluding H5. This estimate of chicken H1 subtype number is consistent with work done by others (Winter, 1984; Winter et al., 1985a; Dupressoir and Sautiere, 1984). As is observed in other vertebrate species it appears that the complement of chicken H1 proteins is expressed in all tissues but that the ratio of H1 subtypes varies between tissues (Kinkade, 1969; Panyim et al., 1971; Berdikov et al., 1975; Winter, 1984; Section 1.3.3(b)). The proportions of the chicken H1 subtypes have also been observed to change during the differentiation of myogenic cells in vitro and during erythropoiesis (Winter, 1984; Winter et al., 1985a; Appels et al., 1972; Section 1.3.3(b)).

The chicken H1 protein sequences display the three domain structure typical of H1 proteins (reviewed in Von Holt et al., 1979; Isenberg, 1979; Cole, 1984; Section 1.3.1), viz., a central hydrophobic globular domain of approximately 80 amino acids, flanked by basic amino-terminal and carboxyl-

terminal domains. The globular region usually commences at residue 30-40 in H1 protein sequences and is the most highly conserved region of the protein between species (see Isenberg, 1979; Von Holt et al., 1979, 1984; Allan et al., 1980; Cole et al., 1984; Levy et al., 1982; Fig. 6.2). H1 molecules bind to the nucleosome, at the exit and entry points of DNA, through their globular domains. The whole H1 molecule, however, is required for chromatin condensation (Section 1.3.1).

The location of domain boundaries as defined by Allan et al. (1980) in the four complete chicken H1 protein sequences is shown in Figure 6.1. The central globular domains of these proteins are presented again in Figure 6.2 with the 01 H1 sequence being used as the standard. The globular domains of other vertebrate H1 proteins are also compared to the 01 H1 sequence. Sequences are available for mammalian, amphibian and fish species. The sequences for Xenopus (Turner et al., 1983), newt (Stephenson et al., 1981b) and human H1 (Carozzi et al., 1984) are derived from nucleotide sequences.

The globular domains of the chicken proteins (Fig. 6.2) are almost identical. This region is 74 amino acids long and commences at amino acid 36 in the 01, 02 and .10 H1 proteins while the globular domain of the 11L protein commences at amino acid 40. It is obvious from Figure 6.2 that the globular domains of the chicken H1 proteins show greater homology to the equivalent regions of H1 proteins from other vertebrate species than they do to the globular domain of the chicken H5 protein.

Analysis of mammalian H1 proteins has enabled the division of these proteins into three types (Cole et al., 1984). These are

(1) the standard somatic H1 variants,

(2) the H1° variants, which are enriched in most non-proliferative tissues (Section 1.3.3(b); Lennox and Cohen 1984a, 1983; Gjerset et al., 1982; Harris and Smith, 1983; Smith et al., 1984a), and

(3) the testis-specific H1t variants (Section 1.3.3(c) and refs. therein).

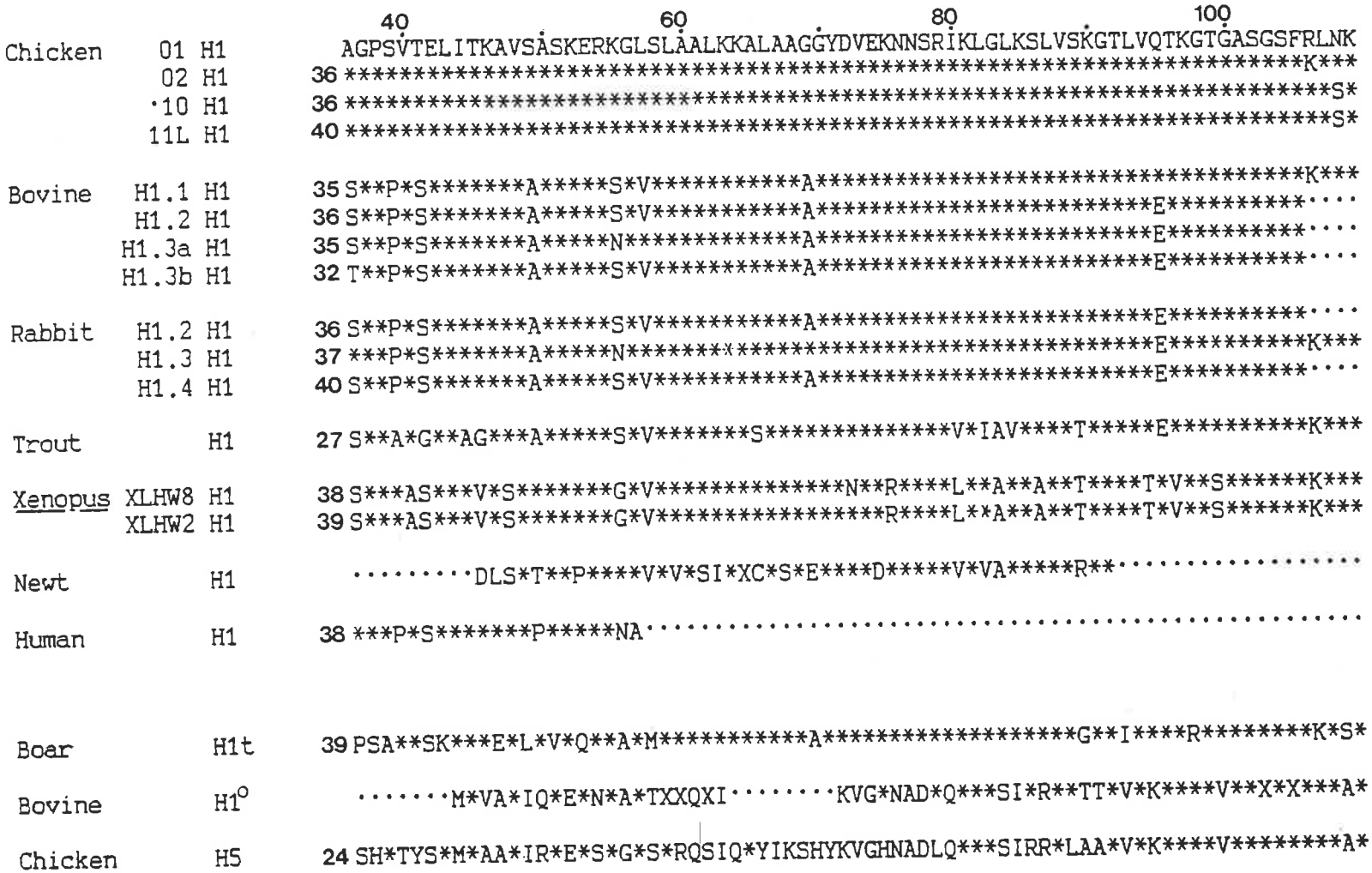
FIGURE 6.2

Comparison of vertebrate H1 protein globular domains

A comparison of the amino acid sequences of vertebrate H1 protein globular domains is shown. The λCH01 (01) chicken H1 sequence is shown on the top line and amino acids are abbreviated by the one-letter code (as for Fig. 6.1; in addition M, H and C represent methionine, histidine and cysteine respectively). Amino acid changes in other H1 protein sequences are indicated. A * is presented where amino acid sequences are the same as the λCH01 sequence. An X represents an undetermined amino acid.

The chicken sequences shown, viz., the λCH01 (01), λCH02 (02), λH1.9/λH1.10 (.10) and pCH11.5E left (11L), are taken from Figure 6.1. Other sequences shown are of "standard" mammalian H1s (bovine H1 s: Liao and Cole, 1981d; L. Liao, M.W. Hsiang and R.D. Cole, cited in Pehrson and Cole, 1981; M.W. Hsiang and R.D. Cole, cited in Cole et al., 1984; rabbit H1s: Von Holt et al., 1979; M.W. Hsiang, C. Largman and R.D. Cole, cited in Cole et al., 1984; human H1: Carozzi et al., 1984), trout testis H1 (Macleod et al., 1977), Xenopus laevis H1s (XLHW8 and XLHW2; Turner et al., 1983), newt H1 (Stephenson et al., 1981b), boar H1 t (Cole et al., 1984), bovine H1° (Smith et al., 1980; Pehrson and Cole, 1981) and chicken H5 (Briand et al., 1980). Chicken, Xenopus, newt and human sequences are derived from gene sequences. Regions marked ... are those for which there is no available protein sequence data.

Numbering on the top line is for the λCH01 sequence. The residue number of the first amino acid of each sequence shown is also indicated.


40
60
80
100
Chicken 01 H1 AGPSVTELITKAVSASKERKGLSLAALKKALAAGGYDVEKNNSRIKLGLKSLVSKGTLVQTKGTGASGSFRLNK
02 H1 36 ********************K***
·10 H1 36 ********************S*
11L H1 40 ********************S*
Bovine H1.1 H1 35 S**P*S*******A*****S*V**********A**********K***
H1.2 H1 36 S**P*S*******A*****S*V**********A**********E**********····
H1.3a H1 35 S**P*S*******A*****N**********A**********E**********····
H1.3b H1 32 T**P*S*******A*****S*V**********A**********E**********····
Rabbit H1.2 H1 36 S**P*S*******A*****S*V**********A**********E**********····
H1.3 H1 37 ***P*S*******A*****N**********E**********K***
H1.4 H1 40 S**P*S*******A*****S*V**********A**********E**********····
Trout H1 27 S**A*G**AG***A*****S*V*******S**********V*IAV****T*****E**********K***
Xenopus XLHW8 H1 38 S***AS***V*S*******G*V**********N**R****L**A**A**T****T*V**S******K***
XLHW2 H1 39 S***AS***V*S*******G*V**********R****L**A**A**T****T*V**S******K***
Newt H1 ·····DLS*T**P****V*V*SI*XC*S*E****D*****V*VA*****R**·····
Human H1 38 ***P*S*******P*****NA·····
Boar H1t 39 PSA**SK***E*L*V*Q**A*M**********A**********G**I****R********K*S*
Bovine H1° ·····M*VA*IQ*E*N*A*TXXQXI·····KVG*NAD*Q***SI*R**TT*V*K****V**X*X***A*
Chicken H5 24 SH*TYS*M*AA*IR*E*S*G*S*RQSIQ*YIKSHYKVGHNADLQ***SIRR*LAA*V*K****V********A*

Each mammalian H1 type has a distinct version of the globular domain as can be seen in Figure 6.2. The globular domains of the chicken H1 proteins are most similar to those of the standard somatic mammalian H1 subtypes, showing nearly 90% homology with these sequences. The H5 protein, on the other hand, shows considerable homology to the globular domain of mammalian H1°. The division of H1 protein types in mammals hence also applies to the chicken proteins, with the globular domains of the various H1 types being well conserved through evolution.

The similarity in globular domains of the chicken H1s and the standard somatic mammalian proteins is also consistent with similarities in amino acid composition and total protein length (Fig. 6.3). A completely sequenced standard rabbit H1, H1.3 (M. W. Hsiang, C. Largman and R. D. Cole, cited in Cole et al., 1984), has 224 residues. The chicken H5 protein in contrast is only 189 amino acids in length (Briand et al., 1980) and has a distinct amino acid composition (Fig. 6.3). Of particular note is the high arginine content of H5 compared to the chicken H1 proteins and standard mammalian proteins. Such high arginine contents are also observed in other H1 proteins associated with highly condensed chromatin, viz., H1t (Cole et al., 1984) and sea urchin sperm H1s (Von Holt et al., 1984; Section 1.3.3(c)).

The differences between the chicken H1 and H5 proteins pointed out above, such as differences in globular domain, protein length and amino acid composition, no doubt play a role in determining functional differences between the two types of protein, but this is yet to be determined.

Recent data suggests that the standard somatic mammalian H1 subtypes, to which the chicken H1 subtypes described here are probably analogous, are non-randomly distributed between active and inactive chromatin (Huang and Cole, 1984; see Section 1.3.4(b)). Correlation between the ability of particular H1 subtypes to aggregate DNA and the degree of condensation of

FIGURE 6.3

Chicken H1 protein amino acid compositions

The amino acid compositions of chicken H1 protein sequences derived from λCHO1 (01), λCHO2 (02), λH1.9/λH1.10 (.10) and pCH11.5E left (11L) H1 genes (protein sequences are given in Fig. 6.1) are shown compared to the amino acid compositions of the completely sequenced chicken H5 protein (Briand et al., 1980) and a completely sequenced "standard" somatic mammalian H1 protein, rabbit H1.3 (M.W. Hsiang, C. Largman and R.D. Cole, cited in Cole et al., 1984). Amino acids are given in three-letter code (asp, aspartic acid; asn, asparagine; thr, threonine; ser, serine; glu, glutamic acid; gln, glutamine; pro, proline; gly, glycine; ala, alanine; val, valine; met, methionine; ile, isoleucine; leu, leucine; tyr, tryosine; phe, phenylalanine; his, histidine; lys, lysine; arg, arginine). The number of each type of amino acid present in each protein sequence are shown. X represents an undetermined amino acid in the λH1.9/λH1.10 (.10) protein (Fig. 6.1). Residue numbers for H5 are for the arginine variant (allelic H5 variants have either arginine or glutamine at position 15; Briand et al., 1980).

Total protein lengths of sequences are also indicated.

PROTEIN:

AMINO ACID:	Chicken H1 01	Chicken H1 02	Chicken H1 10	Chicken H1 11L	Chicken H5	Rabbit H1,3
asp	2	1	1	1	2	1
asn	3	3	2	2	1	4
thr	10	12	10	8	6	8
ser	13	15	15	17	26	13
glu	6	6	6	7	3	8
gln	1	1	1	1	4	0
pro	21	20	21	22	13	26
gly	14	14	12	12	9	15
ala	59	55	62	61	29	58
val	8	9	8	12	8	8
met	0	0	0	0	1	0
ile	2	2	2	2	6	2
leu	10	10	10	10	8	10
tyr	1	1	1	1	3	1
phe	1	1	1	1	1	1
his	0	0	0	0	3	0
lys	60	63	60	63	44	66
arg	7	4	6 +X	4	22	3
RESIDUE NUMBER:	218	217	219	224	189	224

the chromatin fraction in which the subtype is enriched, suggested that the different subtypes may be involved in the production of different *in vivo* chromatin states and hence may play a role in gene expression (discussed in Section 1.3.4(b)).

Core histone variants also appear to be non-randomly distributed in chromatin (Section 1.2.4(a)). H2A.F/H2A.Z-like proteins, for example, are probably preferentially associated with active chromatin and, as for different H1 variants, may play a role in gene expression via effects on nucleosome structure and higher order chromatin conformation (discussed in Sections 1.2.4(b) and 3.2.4(b)).

As the globular domains of the standard H1 subtypes are essentially identical (Fig. 6.2), functional differences between proteins are most probably attributed to differences in their amino-terminal and carboxyl-terminal regions (Cole, 1984; Cole *et al.*, 1984; Von Holt *et al.*, 1979).

Among the chicken H1 proteins, variation in arginine content (Fig. 6.3), for example, could contribute to the proteins having different effects on chromatin structure. The 01 H1 has the highest arginine content (3.2%), being very similar to that for H1° (Smith *et al.*, 1980) which appears to be enriched in repressed chromatin fractions (Roche *et al.*, 1985; Section 1.3.4(b)). Differences in post-translational modification sites could also contribute to the effect that different H1 subtypes have on chromatin conformation (Section 1.3.4(c); Hohmann, 1983; Cole, 1984).

6.3 H1 GENE 5' NON-CODING REGIONS

6.3.1 Comparison of Chicken H1 Genes

Sequences conserved in the DNA 5' to the ATG codons of the completely sequenced 01, 02, .10 and 11L chicken H1 genes are presented in Figure 6.4. Only elements conserved between all four H1 genes, where data is available,

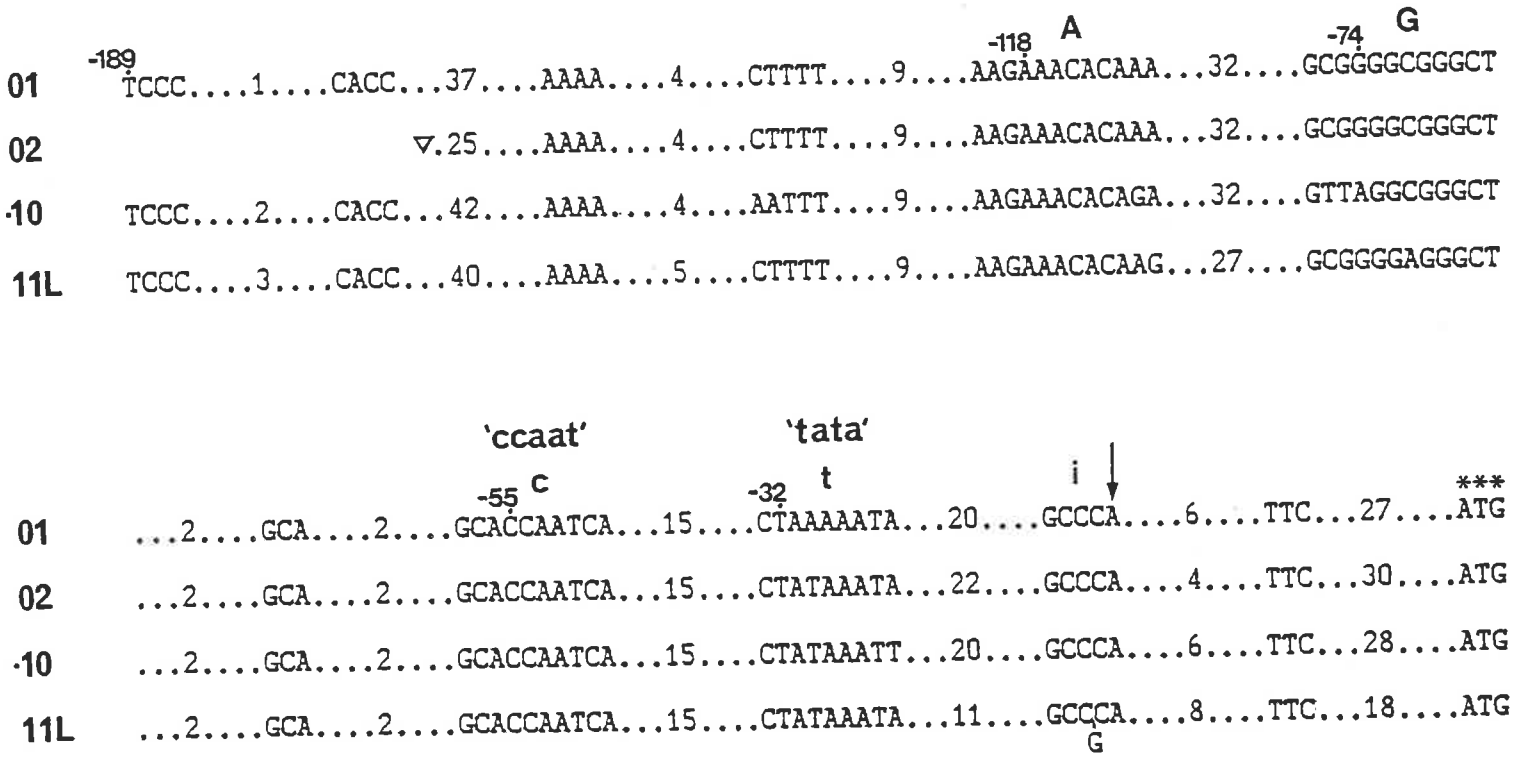
FIGURE 6.4

Chicken H1 gene 5' regions

Sequences conserved in DNA 5' to the initiation codons (marked ***) of the λCHO1 (O1), λCHO2 (O2), λH1.9/λH1.10 (.10) and pCH11.5E left (11L) chicken H1 genes are presented (discussed in Section 6.3.1). The λCHO1, λH1.9/λH1.10 and pCH11.5E left H1 sequences were determined for this thesis. The λCHO2 sequence was determined by Sugarman et al. (1983).

Distances between conserved sequences are given in base pairs. The vertical arrow indicates the location of the 5' termini of mRNAs derived from λCHO1 and λCHO2 H1 genes (Section 5.2.1(b) and 5.2.5; Sturm, 1985; Sugarman et al., 1983). Nucleotide numbers are for the λCHO1 gene sequence. Nucleotides are numbered relative to the location of the mRNA 5' terminus which is designated as +1. ∇ denotes the end of available sequence data for the λCHO2 H1 gene.

Conserved sequences marked A, G, c ("CCAAT" sequence), t ("TATA" sequence) and i are indicated on DNA sequences in Figures 5.5, 5.18, 5.9 and 5.13 (see also Fig. 6.5).


A
G
-189
-118
-74
01 TCCC....1....CACC...37....AAAA....4....CTTTT....9....AAGAAACACAAA...32....GCGGGGCGGGCT
02 ∇.25....AAAA....4....CTTTT....9....AAGAAACACAAA...32....GCGGGGCGGGCT
10 TCCC....2....CACC...42....AAAA....4....AATTT....9....AAGAAACACAGA...32....GTTAGGCGGGCT
11L TCCC....3....CACC...40....AAAA....5....CTTTT....9....AAGAAACACAAG...27....GCGGGGAGGGCT
'ccaat'
'tata'
c
t
i
***
-55
-32
01 ...2....GCA....2....GCACCAATCA...15....CTAAAAATA...20....GCCCA....6....TTC...27....ATG
02 ...2....GCA....2....GCACCAATCA...15....CTATAAATA...22....GCCCA....4....TTC...30....ATG
10 ...2....GCA....2....GCACCAATCA...15....CTATAAATT...20....GCCCA....6....TTC...28....ATG
11L ...2....GCA....2....GCACCAATCA...15....CTATAAATA...11....GCCCA....8....TTC...18....ATG
G

are shown. Due to the tentative nature of the 03 and 11R H1 gene sequences, data for these genes are not included. Preliminary analysis of these genes, however (Sections 5.2.3(c) and 5.2.4; Powell, 1984), suggests that the basic structure of their 5' ends is the same as the other four genes.

As discussed in Chapter 5 (Sections 5.2.1(b) and 5.2.5; Sugarman et al., 1983) the 5' termini of mRNAs derived from the 01 and 02 H1 genes mapped to the A residue of the sequence 5' GCCCA 3'. This sequence is found in a similar position in the .10 H1 gene and a derivative of this sequence, 5' GCCGCA 3', is found in the 11L gene. These sequences in the .10 and 11L genes probably also represent sites of mRNA termini. 5' GCCGCA 3' has been found to be the cap site of the chicken erythrocyte-specific H5 gene with the most 3' C residue representing the mRNA 5' end (Krieg et al., 1983; Wigley et al., 1985).

There is little homology between the sequences coding for the chicken H1 gene mRNA leaders except for a 5' TTC 3' motif 4 to 8 bp downstream from the site of transcription initiation. It is generally found that there is little conservation of the sequences in leader regions of histone mRNAs either between or within species (Maxson et al., 1983b). Turner et al. (1983) previously noted the conservation of the sequence 5'TTTT(A)GTT 3' in the leader regions of Xenopus and sea urchin H1 genes, but this sequence is not conserved in the chicken H1 genes.

The length of the 01, 02, .10 and 11L H1 gene leader regions are respectively 37, 38, 38 and 30 bp. Short leaders are common to most histone genes (Maxson et al., 1983b), including the clustered chicken core histone genes (Section 1.5.3(d)). The chicken H5 and H2A.F genes in contrast have leader sequences of greater than 100 bp (Krieg et al., 1983; Wigley et al., 1985; Robins et al., 1985). The chicken and human H3.3 genes also code for unusually long untranslated regions (Brush et al., 1985; Wells and Kedes, 1985).

As discussed in Chapter 5 (Section 5.2.7) the chicken H1 genes have the "TATA" and "CCAAT" motifs typical of most histone genes and other genes transcribed by RNA polymerase II (Section 1.6.1(a); Hentschel and Birnstiel, 1981). The chicken H1 genes show an extended homology around both these motifs. "CCAAT" extends to 5' GCACCAATCA 3' while the "TATA" sequence extends to a consensus of 5'CTATAAATA 3'. Extended "CCAAT" box homologies have also been observed for chicken H2B genes (Sturm, 1985) and H3 genes (Wang et al., 1985).

In addition to the already discussed sequence motifs, the chicken H1 genes also show sequence conservation in further 5' sequences, as illustrated in Figure 6.4. The most extensive regions of homology are an A-rich and a G-rich sequence at approximately -120 and -75 bp, respectively, from the sites of mRNA 5' ends. As well as conservation of sequence there is also conservation of distances between motifs.

6.3.2 Identification of an H1 Gene-Specific Sequence and a G-Rich Promoter Element

DNA sequence conserved between the 5' regions of the chicken, Xenopus laevis (Turner et al., 1983) and sea urchin (P. miliaris, Birnstiel et al., 1978; S. purpuratus, Levy et al., 1982) H1 genes is presented in Figure 6.5. In addition to "TATA" and "CCAAT" motifs the H1 genes from Xenopus and sea urchin have sequences sharing homology with the previously mentioned upstream A-rich (at approx. -120) and G-rich (at approx. -75) sequences observed in the chicken H1 genes.

The sequence 5' AAACACA 3' within the A-rich region is present in all the H1 genes in a similar relative position. A search through published histone gene sequences, revealed that this sequence is unique to H1 genes and as far as we are aware is not present in the promoter region of any other non-histone gene. This element is the first example of an H1 gene-specific sequence. In the chicken genes the sequence is extended to 5' AAGAAACACA $\begin{matrix}AA\\GG\end{matrix}$ 3'.

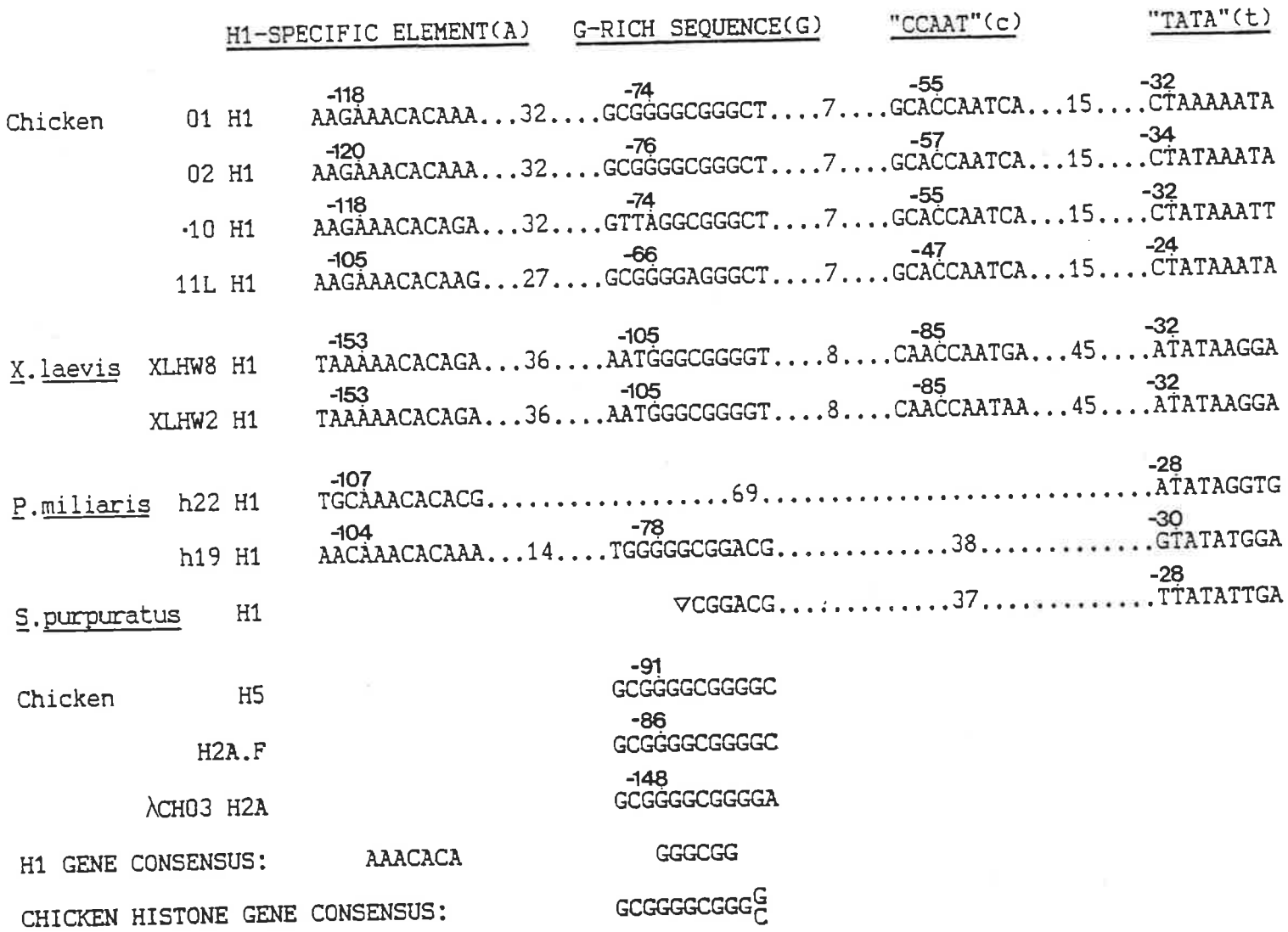
FIGURE 6.5

H1 gene 5' non-coding regions

DNA sequences conserved between the 5' non-coding regions of chicken H1 genes (λCH01 (01), λCH02 (02; Sugarman et al., 1983), λH1.9/λH1.10 (.10) and pCH11.5E left (11L) genes; data from Figs. 5.5, 5.18, 5.9 and 5.13), Xenopus laevis (Turner et al., 1983) and sea urchin (P. miliaris, Birnstiel et al., 1978; S. purpuratus, Levy et al., 1982) H1 genes are presented. Nucleotides are numbered relative to locations of 5' mRNA termini (designated as +1). Termini have been experimentally determined for the chicken λCH01 (Section 5.2.1(b); Sturm, 1985) and λCH02 genes (Section 5.2.5; Sugarman et al., 1983; Fig. 6.4), the P. miliaris h22 (Hentschel et al., 1980) and the S. purpuratus (Sures et al., 1980) H1 genes. Termini locations for other H1 genes are predicted from sequence homology. Distances between conserved sequences are in base pairs. X. laevis H1 genes have additional "TATA" and "CCAAT" sequences, at positions -70 and -58 respectively, which are not shown here. ∇ denotes the end of available sequence data for the S. purpuratus H1 gene.

The conserved sequences from left to right are marked A, G, c and t (as shown in brackets) on chicken H1 DNA sequences in Figures 5.5, 5.18, 5.9 and 5.13 (see also Fig. 6.4).

Sequences representing G-rich elements in the chicken H5 (Krieg et al., 1983), H2A.F (Robins et al., 1985) and λCH03 H2A (Wang et al., 1985) genes are shown below H1 sequences (discussed in Section 6.3.2). Consensus sequences are also given.


H1-SPECIFIC ELEMENT(A)
G-RICH SEQUENCE(G)
"CCAAT"(c)
"TATA"(t)
Chicken 01 H1 -118 AAGAAACACAAA...32....-74 GCGGGGCGGGCT....7....-55 GCACCAATCA...15....-32 CTAAAAATA
02 H1 -120 AAGAAACACAAA...32....-76 GCGGGGCGGGCT....7....-57 GCACCAATCA...15....-34 CTATAAATA
·10 H1 -118 AAGAAACACAGA...32....-74 GTTAGGCGGGCT....7....-55 GCACCAATCA...15....-32 CTATAAATT
11L H1 -105 AAGAAACACAAG...27....-66 GCGGGGAGGGCT....7....-47 GCACCAATCA...15....-24 CTATAAATA
X.laevis XLHW8 H1 -153 TAAAAACACAGA...36....-105 AATGGGCGGGGT....8....-85 CAACCAATGA...45....-32 ATATAAGGA
XLHW2 H1 -153 TAAAAACACAGA...36....-105 AATGGGCGGGGT....8....-85 CAACCAATAA...45....-32 ATATAAGGA
P.miliaris h22 H1 -107 TGCAAACACACG.....69.....-28 ATATAGGTG
h19 H1 -104 AACAAACACAAA...14....-78 TGGGGGCGGACG.....38.....-30 GTATATGGA
S.purpuratus H1 ▽CGGACG...:.....37.....-28 TTATATTGA
Chicken H5 -91 GCGGGGCGGGGC
H2A.F -86 GCGGGGCGGGGC
λCH03 H2A -148 GCGGGGCGGGGA
H1 GENE CONSENSUS: AAACACA GGGCGG
CHICKEN HISTONE GENE CONSENSUS: GCGGGGCGGG G C

Ubiquitous 5' elements have also been observed in H4 genes (Clerc et al., 1983) and H2B genes (Harvey et al., 1982). The H4 gene-conserved element has been shown to be important for maximal expression by analysis of a Xenopus H4 gene in an homologous oocyte transcription system (Clerc et al., 1983). It also appears that the H2B gene-conserved sequence represents a promoter element (J. Mous and H. Stunnenberg, cited in Birnstiel, 1984). In contrast to the H1 gene-specific element the H2B gene sequence can be found in a similar position in numerous other non-histone genes (Sturm, 1985), in particular, this sequence has homology to the conserved upstream decanucleotide (dc) element of immunoglobulin genes (Falkner and Zachau, 1984). It is postulated that this sequence may be part of a general stress response mechanism (Sturm, 1985).

The location of the H1 gene-specific element is different from that of H4 and H2B gene elements, which are located at approximately -50 bp from sites of 5' mRNA termini, but is similar to the location of a sea urchin H2A gene promoter element (Grosschedl et al., 1983; Grosschedl and Birnstiel, 1980b; Section 1.6.1(a)). This sequence has also been observed upstream of some chicken H2A genes (Wang et al., 1985).

For the second upstream region conserved between H1 genes, viz., the G-rich region in Figure 6.5, a cross-species consensus of 5' GGGCGG 3' can be derived. Unlike the H1 gene-specific sequence, the G-rich motif is not specific to H1 histone genes and is found in a similar position in the chicken H5 (Krieg et al., 1983) and H2A.F genes (Robins et al., 1985) and at a more 5' position in the H2A genes from the chicken genomal clone λCH03 (Wang et al., 1985) as shown in Figure 6.5. Of particular note the chicken H5 gene, which codes for an H1-related protein, lacks the H1 gene-specific 5' AAACACA 3' sequence upstream from the G-rich motif but has instead a complementary C-rich motif 5' CCGCCC 3', the first C of which is located at position -164. Upstream C-rich sequences are also observed in the H2A.F and λCH03 H2A genes. The G-rich sequence of the chicken genes can be extended to 5' GCGGGGCGGG$\begin{matrix}C\\G\end{matrix}$ 3'.

The 6 bp 5' GGGCGG 3' sequence, or its inverted complement, have also been noted in the 5' promoter regions of numerous non-histone genes, viz., in herpesvirus thymidine kinase (tk, McKnight and Kingsbury, 1982) and IE genes (Cordingley et al., 1983; Whitton et al., 1983; Preston et al., 1984; Kristie and Roizman, 1984), in an IE gene of human cytomegalovirus (Boshart et al., 1985), contained within the 21 bp repeat sequences of SV40 (Everett et al., 1983), in the "house-keeping" genes for hypoxanthine phosphoribosyl transferase (HPRT, Melton et al., 1984), 3-hydroxy-3-methylglutaryl coenzyme A (HMG CoA) reductase (Reynolds et al., 1984) and adenosine deaminase (ADA, Valerio et al., 1985), and in metallothionein genes (Carter et al., 1984). The G-rich/C-rich motifs have been shown to be important for efficient transcription of the herpesvirus tk gene (McKnight and Kingsbury, 1982; McKnight, 1982; McKnight et al., 1984) and SV40 early and late genes (Everett et al., 1983; Baty et al., 1984; Hartzell et al., 1984; Vigneron et al., 1984).

B. Younghusband, this laboratory (unpublished data), has investigated the role of the G-rich sequence in the chicken H1 gene from λCH01. Deletion of the 5' GGGCGG 3' motif using oligo-nucleotide directed mutagenesis, caused an approximately ten-fold reduction in the level of H1 gene expression when analyzed in both Xenopus ooctyes and HeLa cells. As judged by primer extension analysis, the deletion had no effect on the location of λCH01 H1 gene 5' mRNA termini. The G-rich sequence therefore appears to be an important promoter element, not only for the viral genes mentioned above, but also for H1 genes and presumably other genes containing this sequence. Preliminary experiments using gross deletions in the H2A.F and H5 gene 5'regions also suggests that the 5' GGGCGG 3' sequences of these genes are important for maximal levels of transcription when assayed in the Xenopus oocyte transcription system (Robins et al., 1985; P. Wigley, unpublished data). "Mutant" G-rich sequences occur in the .10 and 11L chicken H1 genes (Figs. 6.4 and 6.5) and, as discussed below, may play a role in the differential expression of these genes (Section 6.4.2).

5' GGGCGG 3' motif repeats in SV40 and in a related monkey promoter have been found to be binding sites for a human transcription factor Sp1 (Gidoni et al., 1984; Dynan and Tjian, 1983a, 1983b). G-rich motifs in other genes may similarly be binding sites for transcription factors like Sp1. The extended homology around the G-rich motif in the chicken genes (Figs. 6.4 and 6.5) suggests that sequences around the "core" motif may influence binding of transcription factors. A G-rich motif has also been observed as part of one of the binding sites for a nuclear factor(s) that binds specifically to the 5' flanking region of the chicken adult B-globin gene (Emerson et al., 1985). P. Wigley (this laboratory) has isolated chromatin wash fractions, from an erythroid-derived cell line, that increase the expression of the chicken H5 gene. It will be of interest to determine whether any of these fractions bind to the G-rich sequence.

In contrast to the G-rich motif, specific deletion of the A-rich H1 gene-specific sequence had no effect on the level of expression of the λCHO1 H1 gene when assayed in Xenopus oocytes (B. Younghusband, unpublished data). None-the-less, the strict conservation of this sequence across species barriers suggests that it plays an essential role in some aspect of H1 gene expression. As discussed below, it may be a candidate for a sequence involved in cell cycle regulation of H1 genes.

6.4 REGULATION OF H1 GENE EXPRESSION

6.4.1 Cell Cycle Regulation

The differences in linkage of the synthesis of different H1 variants to the cell cycle, plays a role in determining the changes that occur in the relative levels of H1 variants as cells cease division during differentiation, and hence also plays a role in determining tissue-specific H1 variant ratios (discussed in Section 1.3.3(b); Lennox and Cohen,

1984a, 1983; Winter, 1984; Winter et al., 1985a). Variants such as mammalian H1° and chicken H-1c are found, for instance, to accumulate in non-dividing cells (Section 1.3.3(b)). The synthesis of these variants appears to be replication-independent (Winter, 1984; Winter et al., 1985a; Chabanas et al., 1983; D'Anna et al., 1985). This property, in addition to the relative stability of these variants, plays a major part in their accumulation. The tissue-specifically expressed avian H5 and mammalian H1t variants are also examples of H1 variants whose synthesis is independent of DNA replication (Section 1.3.3(c); Appels and Wells, 1972; Ruiz-Carrilo et al., 1976; Bucci et al., 1982). In contrast to the above proteins, however, the synthesis of most H1 proteins is replication-dependent, with the bulk of H1 synthesis occurring during the S-phase of the cell cycle (Section 1.3.3(b)).

Replication-independent and dependent H1 protein synthesis, as for core histone synthesis, is regulated primarily at the level of mRNA availability (Sections 1.6.3(a) and 1.6.3(b); Plumb et al., 1984; Stein and Stein, 1984; Baumbach et al., 1984; Helms et al., 1984; Winter, 1984; Winter et al., 1985b; Dalton et al., 1985). In the case of replication-dependent protein synthesis, increased transcription rates and lengthening of histone mRNA half-lives results in increased levels of histone mRNA during the S-phase (Section 1.6.3(a)). In contrast, Dalton et al. (Dalton et al., 1985; Robins et al., 1985; this laboratory) found that the genes for the replication-independent chicken variants H5 and H2A.F are transcribed at a constant rate throughout the cell cycle in AEV-transformed erythroid cells (ts34 cells) resulting in relatively constant levels of histone mRNA.

Analysis of chicken H1 mRNA levels during the cell cycle of ts34 cells revealed that these mRNAs, unlike H5 mRNA, were S-phase regulated (Dalton et al., 1985). The levels of mRNAs homologous to the core histone genes, with which the chicken H1 genes identified in this thesis are linked, were

also found to be S-phase regulated by Dalton et al. (1985). Primer extension analysis has confirmed that the 01 H1 gene is expressed in ts34 cells (S. Dalton; unpublished data) but it is not known how many of the other five located H1 genes are expressed. The results suggest however that the H1 genes that are expressed, are also S-phase regulated and hence code for replication-dependent H1 variants.

Features of the chicken H1 genes, such as their linkage to core histone genes, the presence of histone gene-specific conserved 3' elements, the production of non-polyadenylated mRNAs and the lack of introns (Section 5.2.7) supports the notion that these genes are S-phase regulated (Section 1.6.3(a)). The replication-independent chicken genes, in contrast, are solitary genes, lack the conserved 3' elements and produce polyadenylated mRNAs (Section 1.6.3(b)). The chicken H2A.F and H3.3 genes also contain introns (Section 1.5.3(d)). These features also appear to apply to human H3.3 genes (Wells and Kedes, 1985; Section 1.5.3(e)).

If all the chicken H1 genes located so far are S-phase regulated then it may be that an as yet unidentified gene codes for the H-1c chicken H1 variant, described by Winter et al. (1985a). The expression of mRNA coding for the H-1c protein was found to be almost completely independent of DNA replication in chicken myogenic cells (Winter et al., 1985b; Section 1.3.3(b)). The H-1c protein may vary in amino acid sequence from the H1 proteins coded for by the located chicken H1 genes, as does the replication-independent H5 protein (see Section 6.2), and hence not allow cross-reaction between the gene sequences of the six located H1 genes and the hypothetical H-1c gene. The mammalian standard somatic H1 variants are also distinctly different in primary sequence from their replication-independent counterparts H1t and H1° (Section 6.2; Fig. 6.2).

The lack of function of the H1 gene-specific sequence 5'AAACACA 3' in the non-dividing Xenopus oocyte (Section 6.3.2) and its absence from the non-S-phase regulated H5 gene suggests that the conserved sequence may be a

candidate for a sequence involved in S-phase regulation of H1 genes, while the G-rich sequence represents a general RNA polymerase II promoter element present in a certain class of polymerase II transcribed genes.

The H1 gene-specific sequence is in a similar position to a region of DNA, between -70 and -110 bp from the site of transcription initiation of a human H4 gene, that has recently been shown to be required for S-phase specific enhancement of histone gene transcription *in vitro* (Hanley *et al.*, 1985). The sequences appear to be responsive to soluble transcription factor(s) present maximally during the S-phase of the cell cycle (Heintz and Roeder, 1984; Section 1.6.3(a)) and hence may be involved in cell cycle regulation *in vivo*. The activities stimulating H4 gene transcription may be H4 gene-specific (Heintz and Roeder, 1984). The -70 to -110 region does in fact contain an H4 gene-conserved sequence, previously noted by Clerc *et al.* (1983), located at approximately -75. This sequence is distal to the -50 H4 gene-conserved region that is necessary for maximal H4 gene transcription in *Xenopus* oocytes (Clerc *et al.*, 1983; Section 6.3.2).

The H1 gene-specific sequence may therefore represent the binding site of an H1 gene-specific factor involved in S-phase regulation. If so the organization of promoter elements in the H1 genes is similar to that of H4 genes, in that both types of genes have two conserved sequence elements, the most proximal of which probably represent general promoter elements while the distal elements may be involved in S-phase regulation.

An association between the events involved in the initiation of DNA replication and histone gene transcription has been proposed for both yeast (Hereford *et al.*, 1982) and human (Iqbal *et al.*, 1984) histone genes (see Section 1.6.3(a)). In support of this idea a sequence necessary for the cell cycle regulation of yeast histone genes appears to be coincident with a sequence that supports autonomous replication (*ars*) in yeast (Osley and Hereford, 1982; Section 1.6.3(a)). Interestingly it has been noted (S. Dalton, unpublished) that the H1 gene-specific sequence 5'AAACACA 3' bears

striking homology to a yeast ars consensus sequence 5' TAAA$^{C}_{T}$A$^{C}_{T}$AA$^{AA}_{GT}$ 3' (consensus taken from Montiel et al., 1984). A number of chicken histone gene clusters have in fact been shown to have weak ars activity in yeast (G. Pure, this laboratory, unpublished). A region of DNA containing the non-S-phase regulated H5 gene, in contrast, shows no ars activity. An ars-like sequence has also been identified in the 5' region of a chicken H3 gene (Wang et al., 1985).

The true function of the H1 gene-specific sequence must await its manipulation and expression in cycling cells.

6.4.2 Differential H1 Gene Expression

Besides the obvious differences between the genes coding for replication-independent and replication-dependent H1 variants, there are presumably subtle differences between genes coding for the various replication-dependent H1 proteins, within an organism, which result in the synthesis of these proteins at different relative levels within a single cell (Section 1.3.3(b); Lennox and Cohen, 1984a, 1983; Winter, 1984; Winter et al., 1985a).

Inspection of the 5' regions of the set of presumably S-phase regulated chicken H1 genes (Fig. 6.4) reveals several differences between these genes that could result in their protein products being expressed at different relative levels.

Both the .10 and 11L chicken H1 genes contain "mutant" forms of the conserved G-rich sequence. The core hexanucleotide G-rich motifs for the .10 and 11L genes are respectively 5' AGGCGG 3' and 5' GGGAGG 3' , varying from the conserved 5' GGGCGG 3' sequence by a G → A substitution in the .10 H1 gene sequence and a C → A substitution in the 11L H1 gene. Mutations within the G-rich/C-rich hexanucleotides of the herpesvirus tk gene (McKnight and Kingsbury, 1982; McKnight et al., 1984) and SV40 21 bp

repeats (Baty et al., 1984; Everett et al., 1983) can significantly impair the function of these sequences. The lack of perfect G-rich motifs in the .10 and 11L H1 genes may, therefore, result in their being transcribed less efficiently than the other chicken H1 genes. The "mutant" motif in the 11L H1 gene may be compensated, however, by the presence of two alternative 5' GGGCGG 3' motifs immediately 5' to and overlapping the latter. These are indicated on the 11L DNA sequence in Figure 5.13.

The 01 H1 gene has an unusual "TATA" sequence which may decrease its relative level of expression. The gene has an A instead of a T residue at the third position of the "TATA" sequence. The presence of a T residue at this position is highly conserved in RNA polymerase II transcribed genes (Breathnach and Chambon, 1981) and substitution of this residue in classical "TATA" box sequences results in substantial reductions in transcription levels both in vivo and in vitro (Grosschedl et al., 1981; Hirose et al., 1984). Mutation of the third base does not alter the site of RNA chain initiation. Substitutions in the "TATA" sequence may influence the binding of transcription factors such as those isolated by Parker and Topol (1984a; Section 1.6.1(b)) that bind to this region of DNA.

The 11L H1 gene differs from the 01, 02 and .10 H1 genes in the sequence containing its proposed site of transcription initiation. As previously pointed out (Section 6.3.1) the latter three genes have a cap site sequence of 5' GCCCA 3' whereas the 11L gene has the sequence 5' GCCGCA 3'. This may affect the relative expression of the 11L gene as cap site sequences have been shown to be involved in determining the start of transcription (Grosschedl and Birnstiel, 1980a; Hanley et al., 1985). The 11L gene also differs from the other H1 genes in the distances between some of the conserved sequence blocks shown in Figure 6.4.

Differences in the sequences immediately 5' to the ATG codons in the 01, 02, .10 and 11L H1 gene leaders (Figs. 5.5, 5.18, 5.9 and 5.13) may also be involved in differential production of H1 variants via affecting

translation efficiencies. Kozak (1984a, 1984b) has shown that the region spanning eukaryotic initiation codons has a consensus of 5' CC$\begin{matrix}A\\G\end{matrix}$CCATGG 3', and that mutations of this region affect levels of mRNA translation in vivo.

As well as intrinsic differences between H1 genes, such as those described above, the expression of the different genes may also be influenced by the cell type in which it is being expressed, as the relative level of any one H1 protein type in vertebrate species varies from cell type to cell type, even if they have similar proliferative rates (Section 1.3.3(b)). Individual H1 genes may respond differently to tissue-specific signals. Patterns of H1 protein expression have,in fact, been shown to be responsive to hormonal signals (Gjerset et al., 1982; Hohmann and Cole, 1969, 1971).

It could be envisaged that a poor "TATA" box, for instance, such as that found in the chicken 01 H1 gene, could be compensated by the specific binding of regulatory factors in a particular tissue, such that the 01 H1 gene could be expressed at a low level in one tissue and at a high level in another. Given the set of characterized chicken H1 genes, it will now be possible to examine the mechanisms involved in the differential expression of chicken H1 variants.

6.5 H1 GENE 3' NON-CODING REGIONS

DNA sequence is available for the 3' non-coding regions of the 01, 02, .10 and 11L chicken H1 genes. The structure of their 3' ends is shown in Figure 6.6 and compared to the 3' sequences of H1 genes from Xenopus laevis (Turner et al., 1983), Drosophila (Goldberg, 1979) and the sea urchin species, P. miliaris (Busslinger et al., 1979) and S. purpuratus (Levy et al., 1982).

FIGURE 6.6

H1 gene 3' non-coding regions

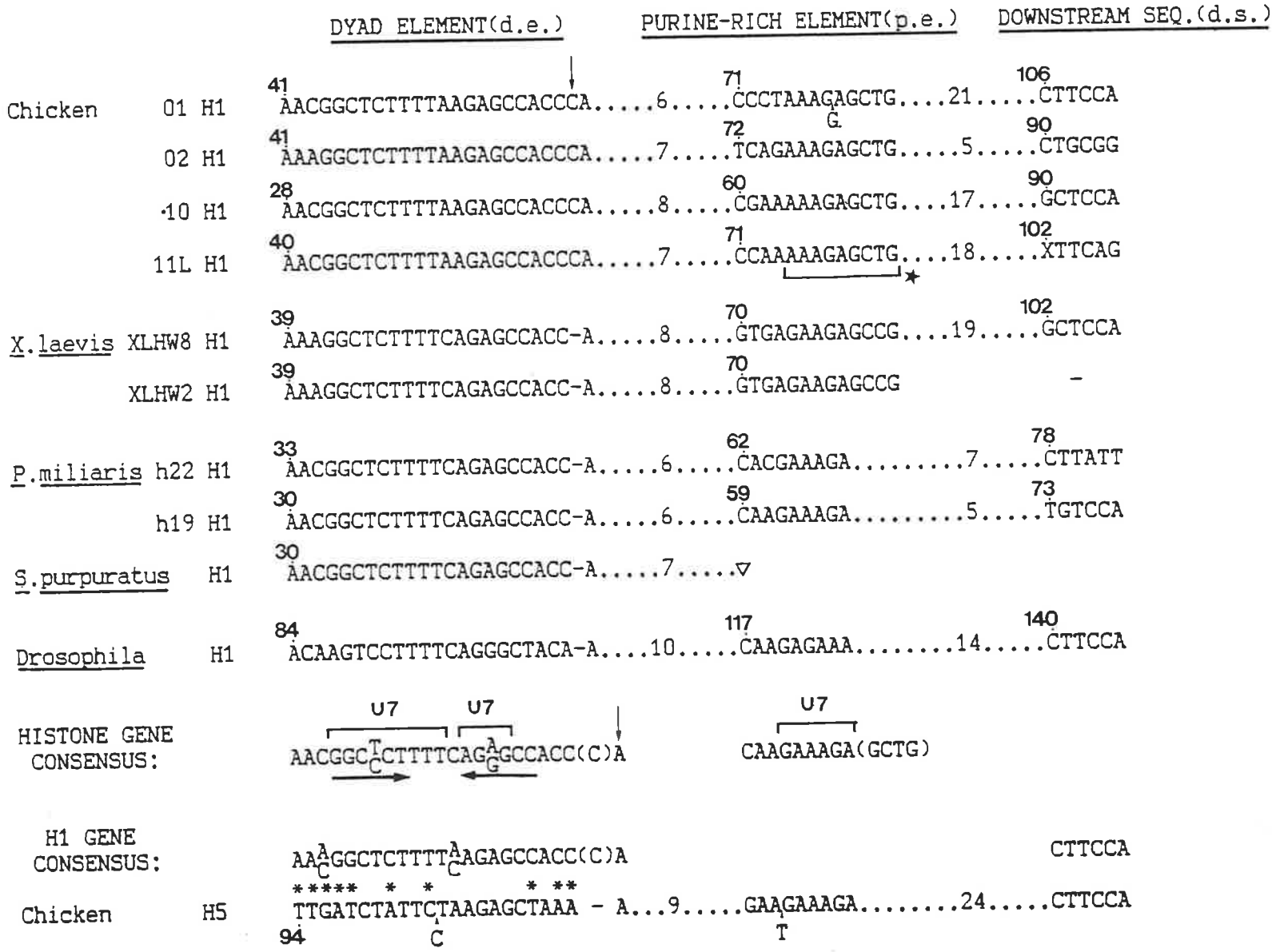
Sequences conserved in DNA 3' to protein coding regions of chicken H1 genes (λCHO1 (O1), λCHO2 (O2), λH1.9/λH1.10 (.10) and pCH11.5E left (11L) genes; data from Figs. 5.5, 5.18, 5.9 and 5.13), Xenopus laevis (Turner et al., 1983), sea urchin (P. miliaris, Busslinger et al., 1979; S. purpuratus, Levy et al., 1982) and Drosophila (Goldberg, 1979) H1 genes are presented. Sequences are numbered from the first nucleotide 3' to termination codons and distances between conserved elements are shown in base pairs. The X represents an undetermined nucleotide in the chicken pCH11.5E left sequence (Fig. 5.13). ∇ denotes the end of available sequence data for the S. purpuratus H1 gene.

Histone gene and H1 gene consensus sequences discussed in Section 6.5 are shown. On the histone gene consensus sequences, a region of dyad symmetry is indicated by horizontal arrows. The consensus location of mRNA 3' termini is indicated by a vertical arrow (Hentschel and Birsnstiel, 1981) and the proposed binding sites for U7 RNA (Strub et al., 1984) are bracketed.

The vertical arrow on the λCHO1 H1 sequence marks the location of λCHO1 H1 gene mRNA 3' termini (Section 5.2.1(b)) and the 5' AAAG(G)AGCTG 3' sequence discussed in Section 6.5 is bracketed and marked with a star.

Relevant 3' sequences in the chicken H5 gene (Krieg et al., 1983) are shown below H1 gene sequences. * represents mismatches between the H5 gene remnant dyad symmetry element and the H1 gene dyad element consensus.

The conserved sequences from left to right are marked d.e., p.e and d.s. on chicken H1 DNA sequences in Figures 5.5, 5.18, 5.9 and 5.13.


DYAD ELEMENT(d.e.)
PURINE-RICH ELEMENT(p.e.)
DOWNSTREAM SEQ.(d.s.)
Chicken 01 H1
41 AACGGCTCTTTTAAGAGCCACCCA.....6.....71 CCCTAAAGAGCTG....21.....106 CTTCCA
02 H1
41 AAAGGCTCTTTTAAGAGCCACCCA.....7.....72 TCAGAAAGAGCTG.....5.....90 CTGCGG
10 H1
28 AACGGCTCTTTTAAGAGCCACCCA.....8.....60 CGAAAAAGAGCTG....17.....90 GCTCCA
11L H1
40 AACGGCTCTTTTAAGAGCCACCCA.....7.....71 CCAAAAAGAGCTG....18.....102 XTTCAG
X.laevis XLHW8 H1
39 AAAGGCTCTTTTCAGAGCCACC-A.....8.....70 GTGAGAAGAGCCG....19.....102 GCTCCA
XLHW2 H1
39 AAAGGCTCTTTTCAGAGCCACC-A.....8.....70 GTGAGAAGAGCCG
P.miliaris h22 H1
33 AACGGCTCTTTTCAGAGCCACC-A.....6.....62 CACGAAAGA.....7.....78 CTTATT
h19 H1
30 AACGGCTCTTTTCAGAGCCACC-A.....6.....59 CAAGAAAGA.....5.....73 TGTCCA
S.purpuratus H1
30 AACGGCTCTTTTCAGAGCCACC-A.....7.....
Drosophila H1
84 ACAAGTCCTTTTCAGGGCTACA-A....10.....117 CAAGAGAAA.....14.....140 CTTCCA
HISTONE GENE CONSENSUS:
U7
U7
AACGGC T/C CTTTTCAG A/G GCCACC(C)A
U7
CAAGAAAGA(GCTG)
H1 GENE CONSENSUS:
AA A/C GGCTCTTTT A/C AGAGCCACC(C)A
CTTCCA
Chicken H5
TTGATCTATTCTAAGAGCTAAA - A...9.....GAAGAAAGA.....24.....CTTCCA
94 C T

As previously discussed (Section 5.2.1(b)) the 3' terminus of the mRNA from the λCH01 (01) H1 gene, in 5 day chicken embryos, maps to the most 3' C residue of the sequence 5' ACCCA 3'. This sequence is located at the 3' end of a highly conserved histone gene-specific 3' element, the dyad element, which contains a region of dyad symmetry (Section 1.6.2(a)). The location of the 01 gene 3' mRNA terminus is indicated in Figure 6.6 and on the gene sequence in Figure 5.5. The 3' termini of mRNAs from clustered chicken core histone genes have been mapped to a similar position (Sturm, 1985). This location of chicken histone mRNA 3' termini is consistent with the location of such termini in other species (Section 1.6.2(a); Hentschel and Birnstiel, 1981; Maxson et al., 1983b).

As for other histone genes, including the clustered chicken core histone genes, containing conserved 3' elements, the chicken H1 gene mRNAs appear to be non-polyadenylated (Section 5.2.6) and S-phase regulated (Section 6.4.1; Dalton et al., 1985). Conserved histone gene-specific 3' elements are absent from non-S-phase regulated (replication-independent) histone genes which also produce polyadenylated mRNAs (Section 1.6.2(b); Krieg et al., 1983; Robins et al., 1985; Brush et al., 1985; Wells and Kedes, 1985).

The dyad elements of the chicken H1 genes and the H1 genes from other species, except for Drosophila (Fig. 6.6), all show a high degree of homology to the consensus sequence for this element, viz., 5' AACGGC$^{T}_{C}$ CTTTTCAG$^{A}_{G}$GCCACCA 3' (Section 1.6.2(a); Hentschel and Birnstiel, 1981; Birchmeier et al., 1983), and contain T and A residues for the optional bases. An H1 gene consensus for the dyad element of the form 5' AA$^{A}_{C}$ GGCTCTTTT$^{A}_{C}$AGAGCCACC(C)A 3' can be derived from the comparative data in Figure 6.6. The presence of an A instead of a C residue in the centre of the dyad elements of the chicken H1 genes extends the symmetry found within the dyad element beyond that found in other histone genes that contain a C residue at this position.

The location of dyad symmetry elements in the chicken H1 genes is in a similar position to most other sequenced histone genes giving rise to 63, 63, 50 and 62 bp respectively, coding for mRNA 3' untranslated regions in the 01, 02, .10 and 11L H1 genes. This is in contrast to the replication-independent chicken histone genes which code for 3' untranslated regions of around 200 bp (Krieg et al., 1983; Robins et al., 1985; Brush et al., 1985). Human cDNA sequences coding for the replication-independent H3.3 variant also have unusually long 3' untranslated regions (Wells and Kedes, 1985).

Upstream from dyad elements there is little homology between the untranslated regions of H1 genes. This may reflect the lack of requirement of these sequences in 3' RNA processing events (Birchmeier et al., 1983; see below).

Between 6 to 8 bp downstream from the location of mRNA 3' termini in the chicken H1 genes is a purine-rich sequence, which is also present in the H1 genes from other species (Fig. 6.6) and in most sequenced core histone genes. This sequence represents a second histone gene-specific conserved 3' element. This sequence was first identified in sea urchin histone genes with a consensus of 5' CAAGAAAGA 3' (Section 1.6.2(a); Hentschel and Birnstiel, 1981). Only the latter part of this element is well conserved in vertebrate genes. In such genes a 4 bp extension of 5' GCTG 3' has been defined (Harvey et al., 1982; Turner et al., 1983; Birnstiel et al., 1985; Fig. 6.6). All four chicken H1 genes contain the element 5' AAAG(G)AGCTG 3' within the conserved region of DNA, with the (G) representing a single base insertion in the 01 H1 gene. This sequence is shown bracketed in Figure 6.6.

That histone mRNA 3' termini are produced by post-transcriptional processing of larger precursor RNAs has been demonstrated for sea urchin (Birchmeier et al., 1984), Drosophila (Price and Parker, 1984), Xenopus (Georgiev et al., 1984), mouse (cited in Lüscher et al., 1985) and chicken

(Krieg and Melton, 1984) histone genes. The presence of both the histone gene-specific conserved 3' elements, viz., the dyad element and the purine-rich sequence, is required in precursor RNAs for processing (Section 1.6.2(a); Birchmeier et al., 1982, 1983, 1984; Georgiev and Birnstiel, 1985). The approximately 60 nucleotide long U7 RNA, found to be necessary for the processing of sea urchin H3 precursor RNA (Section 1.6.2(a); Birchmeier et al., 1984), is hypothesized to hybridize to sequences within both the histone gene-conserved 3' elements in pre-mRNAs, leaving the mRNA 3' site in a single-stranded RNA loopout (Strub et al., 1984). Regions of the conserved elements presumed to be involved in RNA-RNA hybridization are indicated in Figure 6.6. The presence of RNA species in Drosophila histone RNA processing activities (Price and Parker, 1984) and the conservation of the sequences between species presumed to hybridize to the sea urchin U7 RNA, suggests a universal mechanism for histone mRNA 3' end formation.

Processing in sea urchins also requires the presence of approximately 50 to 80 bases of sequence further 3' to the above discussed conserved sequences in precursor RNAs (Birchmeier et al., 1984; Georgiev and Birnstiel, 1985). These sequences are not involved in proposed contacts with U7 RNAs and must therefore play some other as yet undefined role. Although evolutionarily divergent the actual sequence of the downstream region appears to be important (Georgiev and Birnstiel, 1985). Comparison of sequences 3' to the purine-rich region in H1 genes, revealed a small region of homology with a loose consensus of 5' CTTCCA 3' (Fig. 6.6). Such a sequence could be a candidate for a region involved in processing mechanisms.

Another function of histone gene-conserved 3' sequences may be in cell cycle regulation. As previously discussed post-transcriptional mechanisms play a major role in the S-phase regulation of histone gene expression (Section 1.6.3(a)). It has recently been determined that the sequences in the 3' terminal portion of a mouse H4 gene contain the requirements for

its post-transcriptional regulation (Lüscher et al., 1985).

Differences between the 5' regions of chicken H1 genes that could play a role in their differential expression were discussed previously (Section 6.4.2). Similarly differences between these genes in their 3' non-coding regions could also play a part in such expression. For example, the single G residue insertion into the purine-rich chicken H1 gene consensus, 5' AAAG(G)AGCTG 3', in the O1 H1 gene (Fig. 6.6), could affect the binding of a U7-type RNA and consequently the processing efficiency of O1 H1 gene precursor RNAs. Variations in the 5' CTTCCA 3' sequence could also change processing efficiencies. In addition, differences between genes could influence post-transcriptional regulation of transcripts during S-phase, resulting in varying levels of H1 mRNA outside S-phase. Efficiency of transcriptional termination may also vary due to differences in 3' sequence content. In a sea urchin H2A gene the information for transcriptional termination appears to lie within the first 100 - 200 bp 3' to the site for mRNA 3' termini (Birchmeier et al., 1984). The actual sequences involved in transcriptional termination in histone genes or those of other higher eukaryotic non-histone genes are yet to be defined (Birnstiel et al., 1985).

6.6 A REMNANT DYAD SYMMETRY ELEMENT IN THE CHICKEN H5 GENE

The lack of histone gene-specific conserved 3' elements, adjacent to sites of mRNA 3' termini, in the gene for the H1-related chicken H5 protein and other non-S-phase regulated genes, suggests that these genes have developed alternative mechanisms for the generation of mRNA 3' ends. The resultant mRNAs have long 3' untranslated regions, are polyadenylated and generally lack the 5' AAUAAA 3' sequence (Section 1.6.2(b)). A hairpin structure that could potentially form in H5 mRNA, immediately adjacent to 3' termini, has been proposed to be involved in the generation of H5

3'mRNA ends (Krieg et al., 1982a). The use of hybrid RNA structures in 3' termini formation may be a common feature of eukaryotic genes (Birnstiel et al., 1985). Alternative sequences, to those found in most histone genes have also been postulated to be involved in the 3' end formation of yeast histone mRNAs (Smith, 1984; Section 1.6.2(b)).

Although the 3' non-coding sequences of the chicken H5 gene have diverged functionally from those of the chicken H1 genes and other S-phase regulated histone genes, sequence homology to the three elements shown in Figure 6.6, can be found in the H5 gene. As shown (Fig. 6.6) the purine-rich and 5' CTTCCA 3' H5 gene elements are well conserved. The former sequence lacks the vertebrate-specific 5' GCTG 3' extension but retains the 5' GAAAGA 3' sequence proposed to bind to the U7 "processing" RNA (See Fig. 6.6). The dyad element, located 94 bp from the termination codon, in contrast to the other two elements, has accumulated several mutations relative to the consensus. These mutations would affect both, the ability of this region to form a hairpin in RNA, the potential of which has been found to be vital for the processing of sea urchin histone precursor RNAs (Birchmeier et al., 1983, 1984), and the ability to bind a hypothetical U7-type RNA, hence making the H5 dyad sequence non-functional as a processing signal. It is proposed that the nucleotides in the loop of the RNA hairpin, that can be formed from the histone dyad element, may act as a nucleation site for the binding of U7 RNA (Strub et al., 1984).

It appears, therefore, that the H5 gene may have evolved from a typical S-phase regulated H1 gene. Its separation from other histone genes, which would remove it from the homogenizing forces associated with clustered genes (Roberts et al., 1984; Maxson et al., 1983b; Section 1.5.5), has presumably left the H5 gene free to evolve and to come under different controls.

The lack of the H1 gene-specific 5' AAACACA 3' sequence in the 5' region of the H5 gene and the lack of a functional dyad element in the 3'

region may both facilitate the removal of the H5 gene from cell cycle regulation. As discussed previously (Section 6.4.1) the H1 gene-specific sequence may be involved in cell cycle regulation of H1 genes at the transcriptional level. In contrast, sequences in the 3' ends of histone genes are probably involved in post-transcriptional cell cycle regulatory mechanisms (Lüscher et al., 1985). The conserved dyad element is a possible target for such mechanisms. As continued protein synthesis appears to be necessary for S-phase regulation of histone mRNAs (Section 1.6.3(a)), sequences such as the dyad element may be sites of interaction with regulatory proteins. Features of H5 mRNAs and the mRNAs from other non-S-phase regulated genes, such as polyadenylation and long 5' and 3' untranslated regions, may also protect such mRNAs from cell cycle-specific post-transcriptional regulatory processes (Old and Woodland, 1984).

The H5 gene has, in addition, presumably acquired sequences that bring it under tissue-specific control, as this gene is found to be expressed exclusively in erythroid cells (Section 1.3.3(c); Shannon et al., 1985). P. Wigley (this laboratory) has noted that the H5 gene contains a sequence element that is conserved in the 5' region of B-globin genes (Dierks et al., 1983), which is also part of a binding site for erythrocyte-derived nuclear factors (Emerson et al., 1985; Plumb et al., 1985). Such a sequence may be involved in the tissue-specific expression of the H5 gene. As previously discussed (Section 6.3.2) P. Wigley has isolated chromatin factors from an erythroid-derived cell line that increase the expression of the H5 gene. Such factors could be involved in tissue-specific expression.

6.7 FINAL SUMMARY

Six chicken H1 genes have been identified and analyzed and each was found to code for a different H1 variant. The genes appear to be typical histone genes in both conservation of histone gene-specific sequence

elements and in their expression. The genes probably represent the full complement of S-phase regulated H1 genes in the chicken genome. In addition to the H5 gene, further genes may exist that code for replication-independent H1 variants, such as the H-1c protein described by Winter et al. (Winter, 1984; Winter et al., 1985a) or proteins analogous to H1t variants of mammals (Cole et al., 1984). By analogy to findings in mammals (Huang and Cole, 1984), the different H1 variants coded for by the six isolated chicken H1 genes may be different functionally and play a role in gene expression via being non-randomly distributed in chromatin. The chicken H5 protein is, no doubt, another example of a functionally distinct variant (Section 6.2). Work is at present being directed towards determining which of the observed chicken H1 variants each of the characterized chicken H1 genes code for (J. Powell and F. Shannon, this laboratory, unpublished).

Comparison of the 5' regions of chicken H1 genes to the H1 genes from other species has resulted in the identification of an H1 gene-specific sequence, 5' AAACACA 3', and a G-rich promoter element, which is also present in other histone and non-histone genes. The G-rich sequence has been found to be important for efficient transcription of a chicken H1 gene (B. Younghusband, unpublished) and other non-histone genes (Section 6.3.2). A function for the H1 gene-specific sequence is yet to be determined, but a possible role for this sequence could be in the S-phase regulation of H1 genes (Section 6.4.1). In the 3' regions of H1 genes a small region of homology was identified downstream from histone gene-specific 3' elements. Such a sequence could play a role in RNA processing (Section 6.5).

As well as similarities between H1 genes, differences between the 5' and 3' non-coding regions of the set of chicken H1 genes were identified, that could be involved in differential expression of H1 gene protein products. Future analysis of mammalian genes coding for the set of standard somatic H1 variants, to which chicken H1 gene-derived proteins

appear to be analogous (Section 6.2), may enable the identification of DNA sequence elements involved in the expression of different types of H1 variant. At present partial sequence is only available for a single human H1 gene (Carozzi et al., 1984).

The gene for the H1-related chicken H5 protein, in contrast to most histone genes, is both tissue-specifically expressed and non-S-phase regulated. Comparison of this gene to H1 genes revealed, however, the remnants of 3' conserved sequences which suggested that the H5 gene evolved from a standard S-phase regulated gene. Several features of the H5 gene, in particular, the lack of the 5' H1 gene-specific element, may contribute to observed differences in expression between H5 and H1 genes (Section 6.6).

The set of characterized chicken H1 genes provides a starting point for further investigation of H1 gene expression. Sequences involved in cell cycle regulation, for example, could be identified by use of S-phase nuclear extracts, as previously described by Hanley et al. (1985), or by analysis of manipulated genes introduced into tissue culture cells (Lüscher et al., 1985). Given the potential role of ars-type sequences in cell cycle regulation (Section 6.4.1), experiments are in progress to localize such sequences in chicken histone gene clusters (G. Pure, unpublished). Sequences involved in differential H1 gene expression could also be identified by analysis of manipulated genes in appropriate expression systems. In this regard, it now seems to be possible to introduce genes into chicken cells both in vitro and in vivo (Souza et al., 1984).

It is feasible that the chicken H1 genes could also be used to investigate H1 protein function. H1 protein products derived from chicken H1 genes in a bacterial expression system (Guarente et al., 1980) could be used, for example, in chromatin reconstitution experiments to test the effects of different H1 variants. Manipulation of gene coding regions, in these experiments, could permit the determination of essential functional

regions in H1 proteins. Expression of H1 genes in yeast (Hitzeman et al., 1981), which lack H1 proteins (Certa et al., 1984; Section 1.3.1), may also be useful in determining the effects different H1 variants have on chromatin structure. Similar experiments are being undertaken to determine the functional role of the chicken H2A.F variant (G. Pure, J. Whiting, unpublished).

In conclusion, the chicken H1 gene family, characterized here, provides a useful system for the further analysis of mechanisms involved in the general, differential and cell cycle-specific expression of vertebrate H1 genes, and also for the determination of the functional properties of different H1 variants in chromatin.

NOTE ADDED IN PROOF:

Two new *Xenopus* H1 gene sequences (X1h1 and X1h3) have recently been published (10/85) by Perry *et al*. (Perry, M., Thomsen, G.H. and Roeder, R.G. (1985) J. Mol. Biol. *185*, 479-499). In addition, these workers presented a comparative analysis of H1 gene 5' non-coding regions. This comparison revealed the H1 gene-specific and G-rich elements pointed out in the comparative analysis presented in this thesis (Fig. 6.5; data previously published (1/85) in Coles and Wells (1985)). The analysis of Perry *et al*. (1985) also cites data from an unpublished human H1 gene (R. Zhong and N. Heintz, unpublished).

A trout H1 gene sequence has also recently been published (Mezquita, J., Conner, W., Winkfein, R.J. and Dixon, G.H. (1985) J. Mol. Evol. *21*, 209-219); this too contains the 5' elements discussed above.

The conservation of these 5' elements in a wide variety of animal species strengthens the predicted importance of these sequences in the function of H1 genes.

BIBLIOGRAPHY

Alder, D. and Gorovsky, M.A. (1975) J. Cell Biol. 64, 389-397.

Allan, J., Hartman, P.G., Crane-Robinson, C., Aviles, F.X. (1980) Nature 288, 675-679.

Allan, J., Harborne, N., Rau, D.C. and Gould, H. (1982) J. Cell Biol. 93, 285-297.

Allis, C.D., Glover, C.V.C., Bowen, J.K. and Gorovsky, M.A. (1980) Cell 20, 609-617.

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

Allis, C.D., Allen, R.L., Wiggins, J.C., Chicoine, L.G. and Richman, R. (1984) J. Cell Biol. 99, 1669-1677.

Anderson, J.N., Vanderbilt, J.N., Lawson, G.M., Tsai, M.J. and O'Malley, B.W. (1983) Biochemistry 22, 21-30.

Appels, R. and Wells, J.R.E. (1972) J. Mol. Biol. 70, 425-434.

Appels, R., Wells, J.R.E. and Williams, A.F. (1972) J. Cell Sci. 10, 47-59.

Arceci, R.J. and Gross, P.R. (1980) Dev. Biol. 80, 186-209.

Artishevsky, A., Delegeane, A.M. and Lee, A.S. (1984) Mol. Cell. Biol. 4, 2364-2369.

Aubin, R.J., Fréchett, A., DeMurcia, G., Mandel, P., Lord, A., Grondin, G. and Poirier, G.G. (1983) EMBO J. 2, 1685-1693.

Baer, B.W. and Rhodes, D. (1983) Nature 301, 482-488.

Bagshaw, J.C., Andrews, M.T. and Perry, B.A. (1984) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 181-196.

Ball, D.J., Slaughter, C.A., Hensley, P. and Garrard, W.T. (1983) FEBS Lett. 154, 166-170.

Baltimore, D. (1981) Cell 24, 592-594.

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

Bannon, G.A. and Gorovsky, M.A. (1984) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 163-179.

Bannon, G.A., Calzone, F.J., Bowen, J.K., Allis, C.D. and Gorovsky, M.A. (1983) Nucl. Acids Res. 11, 3903-3917.

Bannon, G.A., Bowen, J.K., Yao, M.C. and Gorovsky, M.A. (1984) Nucl. Acids Res. 12, 1961-1975.

Bates, D.L. and Thomas, J.O. (1981) Nucl. Acids Res. 9, 5883-5894.

Baty, D., Barrera-Saldana, H.A., Everett, R.D., Vigneron, M. and Chambon, P. (1984) Nucl. Acids Res. 12, 915-932.

Baumbach, L.L., Marashi, F., Plumb, M., Stein, G. and Stein, J. (1984) Biochemistry 23, 1618-1625.

Belyavsky, A.V., Bavykin, S.G., Goguadze, E.G. and Mirzabekov, A.D. (1980) J. Mol. Biol. 139, 519-536.

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

Berdikov, V.A., Gorel, C., Argutinskaya, S.V., Cherepanova, V.A. and Kileva, E.V. (1975) Molekulyarnaya Biologiya 10, 887-896.

Berent, S.L. and Sevall, J.S. (1984) Biochemistry 23, 2977-2983.

Bhatnagar, Y.M., McCullar, M.K. and Chronister, R.B. (1984) Cell Biol. Int. Rep. 8, 971-979.

Billet, M.A. and Hindley, J. (1972) Eur. J. Biochem. 28, 451-462.

Billings, P.C., Orf, J.W., Palmer, D.K., Talmage, D.A., Pan, C.G. and Blumenfeld, M. (1979) Nucl. Acids Res. 6, 2151-2164.

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

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

Birchmeier, C., Schümperli , D., Sconzo, G. and Birnstiel, M.L. (1984) Proc. Natl. Acad. Sci. USA 81, 1057-1061.

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

Birnstiel, M.L. (1984) Philos. Trans. R. Soc. Lond. B. 307, 293-295.

Birnstiel, M.L., Portmann, R., Busslinger, M., Schaffner, W., Probst, E. and Kressman, A. (1978) Alfred Benzon Symp. 13, 117-129.

Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349-359.

Bolivar, F. (1978) Gene 4, 121-136.

Bolivar, F., Rodriguez, R., Greene, P.J., Betlach, M., Heyneker, H.L., Boyer, H.W., Crosa, J. and Falkow, S. (1977) Gene 2, 95-113.

Boshart, M., Weber, F., Jahn, G., Dorsch-Häsler, K., Fleckenstein, B. and Schaffner, W. (1985) Cell 41, 521-530.

Boss, M.A. (1983) Nature 303, 281-282.

Boulikas, T., Wiseman, J.M. and Garrard, W.T. (1980) Proc. Natl. Acad. Sci. USA 77, 127-131.

Bram, R.J. and Kornberg, R.D. (1985) Proc. Natl. Acad. Sci. USA 82, 43-47.

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

Briand, G., Kmiecik, D., Sautiere, P., Wouters, D., Borie-loy, O., Biserte, G., Mazen, A. and Champagne, M. (1980) FEBS Lett. 112, 147-151.

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

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

Brush, D., Dodgson, J.B., Choi, O.-R., Stevens, P.W. and Engel, J.D. (1985) Mol. Cell. Biol. 5, 1307-1317.

Bucci, L.R., Brock, W.A. and Meistrich, M.L. (1982) Exp. Cell Res. 140, 111-118.

Busslinger, M., Portmann, R. and Birnstiel, M.L. (1979) Nucl. Acids Res. 6, 2997-3008.

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

Bustin, M. and Cole, R.D. (1968) J. Biol. Chem. 243, 4500-4505.

Butler, P.J.G. (1983) CRC Crit. Rev. Biochem. 15, 57-91.

Butler, P.J.G. (1984) EMBOJ. 3, 2599-2604.

Carozzi, N., Marashi, F., Plumb, M., Zimmerman, S., Zimmerman, A., Coles, L.S., Wells, J.R.E., Stein, G. and Stein, J. (1984) Science 224, 1115-1117.

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

Cartwright, I.L., Abmayr, S.M., Fleischmann, G., Lowenhaupt, K., Elgin, S.C.R., Keene, M.A. and Howard, G.C. (1982) CRC Crit. Rev. Biochem. 13, 1-86.

Certa, V., Colavito-Shepanski, M. and Grunstein, M. (1984) Nucl. Acids Res. 12, 7975-7985.

Chabanas, A., Lawrence, J.J., Humbert, J. and Eisen, H. (1983) EMBOJ. 2, 833-837.

Chambers, S.A.M., Vaughn, J.P. and Shaw, B.R. (1983) Biochemistry 22, 5626-5631.

Chang, A.C.Y., Nunberg, J.H., Kaufman, R.J., Erlich, H.A., Schimke, R.T. and Cohen, S.N. (1978) Nature 275, 617-624.

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

Childs, G., Maxson, R., Cohn, R.H. and Kedes, L.H. (1981) Cell 23, 651-663.

Childs, G., Nocente-McGrath, C., Lieber, T., Holt, C. and Knowles, J.A. (1982) Cell 31, 383-393.

Chirgwin, J.M., Przybyla, A.E., Macdonald, R.J., Rutter, W.J. (1979) Biochemistry 24, 5294-5299.

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

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

Clark, S.J. (1982) Ph.D. thesis, University of Adelaide, Adelaide, South Australia.

Clarke, L. and Carbon, J. (1976) Cell 9, 91-99.

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

Coen, E., Strachan, T., Dover, G. (1982) J. Mol. Biol. 158, 17-35.

Cohn, R.H. and Kedes, L.H. (1979a) Cell 18, 843-853.

Cohn, R.H. and Kedes, L.H. (1979b) Cell 18, 855-864.

Cole, K.D., York, R.G. and Kistler, W.S. (1984) J. Biol. Chem. 254, 13695-13702.

Cole, R.D. (1977) in "The Molecular Biology of the Mammalian Genetic Apparatus", Ts'o, P., ed., Elsevier/North-Holland Biomedical Press, Amsterdam, New York, pp. 93-104.

Cole, R.D. (1984) Anal. Biochem. 136, 24-30.

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

Conner, W., Mezquita, J., Winkfein, R.J., States, J.C. and Dixon, G.H. (1984a) J. Mol. Evol. 20, 227-235.

Conner, W., States, J.C., Mezquita, J. and Dixon, G.H. (1984b) J. Mol. Evol. 20, 236-250.

Cooper, E. and Spaulding, S.W. (1985) Mol. Cell. Endocrinol. 39, 1-20.

Cordingley, M.G., Campbell, M.E.M. and Preston, C.M. (1983) Nucl. Acids Res. 11, 2347-2365.

Crane-Robinson, C. and Privalov, P.L. (1983) Biopolymers 22, 113-118.

Crawford, R., Krieg, P., Harvey, R.P., Hewish, D.A. and Wells, J.R.E. (1979) Nature 279, 132-136.

Dalton, S., Coleman, J.R. and Wells, J.R.E. (1985) Mol. Cell. Biol., in press.

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

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

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

D'Anna, J.A., Thayer, M.M., Tobey, R.A. and Gurley, L.R. (1985) Biochemistry 24, 2005-2010.

Davis, R.W., Botstein, D. and Roth, J.R. (1980) in "Advanced Bacterial Genetics", Cold Spring Harbor Laboratory, New York, pp. 227-230.

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

De Groot, P., Strickland, W.N., Brandt, W.F. and Von Holt, C. (1983) Biochim. Biophys. Acta 747, 276-283.

Delegeane, A.M., and Lee, A.S. (1982) Science 215, 79-81.

Destrée, O.H.J., Bendig, M.M., De Laaf, R.T.M. and Koster, J.G. (1984) Biochim. Biophys. Acta 782, 132-141.

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

Dodgson, J.B., Strommer, J. and Engel, J.D. (1979) Cell 17, 879-887.

Doenecke, D. and Gallwitz, D. (1982) Mol. Cell. Biochem. 44, 113-128.

Dover, G. (1982) Nature 299, 111-117.

Dupressoir, T. and Sautiere, P. (1984) Biochem. Biophys. Res. Comm. 122, 1136-1145.

Dynan, W.S. and Tjian, R. (1983a) Cell 32, 669-680.

Dynan, W.S. and Tjian, R. (1983b) Cell 35, 79-87.

Emerson, B.M., Lewis, C.D. and Felsenfeld, G. (1985) Cell 41, 21-30.

Engel, J.D. (1984) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 263-280.

Engel, J.D., Sugarman, B.J. and Dodgson, J.B. (1982) Nature 297, 434-436.

Etkin, L.D. and Di Berardino, M.A. (1983) in "Eukaryotic Genes, their Structure, Activity and Regulation", Maclean, N., Gregory, S.P. and Flavell, R.A., eds., Butterworths and Co., Cambridge U.K., pp. 127-156.

Etkin, L.D. and Maxson, R.E. (1980) Dev. Biol. 75, 13-25.

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

Fahrner, K., Yarger, J. and Hereford, L. (1980) Nucl. Acids Res. 8, 5725-5737.

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

Finch, J.T. and Klug, A. (1976) Proc. Natl. Acad. Sci. USA 73, 1897-1901.

Franklin, S.G. and Zweidler, A. (1977) Nature 266, 273-275.

Franks, R.R. and Davis, F.C. (1983) Dev. Biol. 98, 101-109.

Gabrielli, F., Hancock, R. and Faber, A.J. (1981) Eur. J. Biochem. 120, 363-369.

Gabrielli, F., Alden, D.P., Carrel, S.C., Von Bahr, C., Rane, A., Angeletti, C.A. and Hancock, R. (1985) Mol. Cell. Biochem. 65, 57-66.

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

Gargiulo, G., Razvi, F. and Worcel, A. (1984) Cell 38, 511-521.

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

Georgiev, O. and Birnstiel, M.L. (1985) EMBO J. 4, 481-489.

Georgiev, O., Mous, J. and Birnstiel, M.L. (1984) Nucl. Acids Res. 12, 8539-8551.

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

Giniger, E., Varnum, S.M. and Ptashne, M. (1985) Cell 40, 767-774.

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

Glover, C.V.C., Vavra, K.J., Guttman, S.D. and Gorovsky, M.A. (1981) Cell 23, 73-77.

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

Gorka, C. and Lawrence, J.J. (1979) Nucl. Acids Res. 7, 347-359.

Gould, A.R., Palukaitus, P., Symons, R.H. and Mossop, D.W. (1978) Virology 84, 443-455.

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

Graves, R.A. and Marzluff, W.F. (1984) Mol. Cell. Biol. 4, 351-357.

Graves, R.A., Wellmann, S.E., Chiu, I.-M. and Marzluff, W.F. (1985) J. Mol. Biol. 183, 179-194.

Groner, B., Kennedy, N., Skroch, P., Hynes, N.E. and Ponta, H. (1984) Biochim. Biophys. Acta 781, 1-6.

Groppi, V.E. and Coffino, P. (1980) Cell 21, 195-204.

Gross, K., Schaffner, W., Telford, J. and Birnstiel, M. (1976) Cell 8, 479-484.

Gross-Bellard, M., Oudet, P. and Chambon, P. (1973) Eur. J. Biochem. 36, 32-38.

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

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

Grosschedl, R. and Birnstiel, M.L. (1982) Proc. Natl. Acad. Sci. USA 79, 297-301.

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

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

Grosveld, G.C., Rosenthal, A. and Flavell, R.A. (1982) Nucl. Acids Res. 10, 4951-4972.

Grove, G.W. and Zweidler, A. (1984) Biochemistry 23, 4436-4443.

Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.

Grunstein, M., Rykowski, M., Kolodrubetz, D., Choe, J. and Wallis, J. (1984) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 35-63.

Guarente, L. (1984) Cell 36, 799-800.

Guarente, L., Roberts, T.M. and Ptashne, M. (1980) Science 209, 1428-1430.

Hanley, S.M., Bleecker, G.C. and Heintz, N. (1985) Mol. Cell. Biol. 5, 380-389.

Hannon, R., Bateman, R., Allan, J., Harborne, N. and Gould, H. (1984) J. Mol. Biol. 180, 131-149.

Harborne, N. and Allan, J. (1983) FEBS Lett. 155, 88-92.

Harris, M.R. and Smith, B.J. (1983) Biochem. J. 211, 763-766.

Hartman, P.G., Chapman, G.E., Moss, T. and Bradbury, E.M. (1977) Eur. J. Biochem. 77, 45-51.

Hartzell, S.W., Byrne, B.J. and Subramanian, K.N. (1984) Proc. Natl. Acad. Sci. USA 81, 23-27.

Harvey, R.P. (1982) Ph.D. thesis, University of Adelaide, Adelaide, South Australia.

Harvey, R.P. and Wells, J.R.E. (1979) Nucl. Acids Res. 7, 1787-1798.

Harvey, R.P. and Wells, J.R.E. (1984) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 245-262.

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

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

Harvey, R.P., Whiting, J.A., Coles, L.S., Krieg, P.A. and Wells, J.R.E. (1983) Proc. Natl. Acad. Sci. USA 80, 2819-2823.

Heintz, N. and Roeder, R.G. (1984) Proc. Natl. Acad. Sci. USA 81, 2713-2717.

Heintz, N., Zernick, M. and Roeder, R.G. (1981) Cell 24, 661-668.

Heintz, N., Sive, H.L. and Roeder, R.G. (1983) Mol. Cell. Biol. 3, 539-550.

Helms, S., Baumbach, L., Stein, G. and Stein, J. (1984) FEBS Lett. 168, 65-69.

Hentschel, C., Irminger, J.-C., Bucher, P. and Birnstiel, M.L. (1980) Nature, 285, 147-151.

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

Hentzen, P.C., Rho, J.H. and Bekhor, I. (1984) Proc. Natl. Acad. Sci. USA 81, 304-307.

Hereford, L., Fahrner, K., Woolford, J., Rosbach, M. and Kaback, D.B. (1979) Cell 18, 1261-1271.

Hereford, L.M., Osley, M.A., Ludwig, J.R. and McLaughlin, C.S. (1981) Cell 24, 367-375.

Hereford, L., Bromley, S. and Osley, M.A. (1982) Cell 30, 305-310.

Hirose, S., Takeuchi, K., Hori, H., Hirose, T., Inayama, S. and Suzuki, Y. (1984) Proc. Natl. Acad. Sci. USA 81, 1394-1397.

Hitzeman, R.A., Hagie, F.E., Levie, H.L., Goeddel, D.V., Ammerer, G. and Hall, B.D. (1981) Nature 293, 717-722.

Hogan, B. (1983) Nature 306, 313-314.

Hohmann, P. (1978) Subcellular Biochem. 5, 87-127.

Hohmann, P. (1983) Mol. Cell. Biochem. 57, 81-92.

Hohmann, P. and Cole, R.D. (1969) Nature 223, 1064-1066.

Hohmann, P. and Cole, R.D. (1971) J. Mol. Biol. 58, 533-540.

Holt, C.A. and Childs, G. (1984) Nucl. Acids Res. 12, 6455-6471.

Huang, H.-C. and Cole, R.D. (1984) J. Biol. Chem. 259, 14237-14242.

Huang, P.C., Branes, L.P., Mura, C., Quagliarello, V. and Kropkowski-Bohdan, P. (1977) in "The Molecular Biology of the Mammalian Genetic Apparatus", T'so , P.,ed., Elsevier/North-Holland Biomedical Press, Amsterdam, New York, pp. 105-126.

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

Ikenaga, H. and Saigo, K. (1982) Proc. Natl. Acad. Sci. USA 79, 4143-4147.

Iqbal, M.A., Plumb, M., Stein, J., Stein, G. and Schildkraut, C.L. (1984) Proc. Natl. Acad. Sci. USA 81, 7723-7727.

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

Jones, K.A. and Tjian, R. (1984) Cell 36, 155-162.

Kafatos, F.C., Jones, C.W. and Efstratiadis, A. (1979) Nucl. Acids Res. 7, 1541-1552.

Kaplan, L.J., Bauer, R., Morrison, E., Langan, T.A. and Fasman, G.D. (1984) J. Biol. Chem. 259, 8777-8785.

Karp, R. (1979) Ph.D. thesis, Stanford University, Palo Alto, California.

Karpov, V.L., Preobrazhenskaya, O.V. and Mirzabekov, A.D. (1984) Cell 36, 423-431.

Kedes, L.H. (1976) Cell 8, 321-331.

Kedes, L.H. (1979) Ann. Rev. Biochem. 48, 837-870.

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

Kimura, T., Mills, F.C., Allan, J. and Gould, H. (1983) Nature 306, 709-712.

Kinkade, J.M. (1969) J. Biol. Chem. 244, 3375-3386.

Knowles, J.A. and Childs, G.J. (1984) Proc. Natl. Acad. Sci. USA 81, 2411-2415.

Koropatnick, J., Andrews, G., Duerkson, J.D., Varshney, V. and Gedamu , L. (1983) Nucl. Acids Res. 11, 3255-3267.

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

Kozak, M. (1984b) Nucl. Acids Res. 12, 857-872.

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

Krieg, P.A., Robins, A.J., Colman, A. and Wells, J.R.E. (1982a) Nucl. Acids Res. 10, 6777-6785.

Krieg, P.A., Robins, A.J., Gait, M.J., Titmas, R.C. and Wells, J.R.E. (1982b) Nucl. Acids Res. 10, 1495-1502.

Krieg, P.A., Robins A.J., D'Andrea, R. and Wells, J.R.E. (1983) Nucl. Acids Res. 11, 619-627.

Kristie, T.M. and Roizman, B. (1984) Proc. Natl. Acad. Sci. USA 81, 4065-4069.

Kumaroo, K.K. and Irvin, J.L. (1980) Biochem. Biophys. Res. Comm. 94, 49-54.

Langan, T.A., Rall, S.C. and Cole, R.D. (1971) J. Biol. Chem. 246, 1942-1948.

Lennox, R.W. (1984) J. Biol. Chem. 259, 669-672.

Lennox, R.W. and Abrams, L.A. (1982) J. Cell Biol. 95, 83.

Lennox, R.W. and Cohen, L.H. (1983) J. Biol. Chem. 258, 262-268.

Lennox, R.W. and Cohen, L.H. (1984a) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 373-395.

Lennox, R.W. and Cohen, L.H. (1984b) Dev. Biol. 103, 80-84.

Lennox, R.W., Oshima, R.G. and Cohen, L.H. (1982) J. Biol. Chem. 257, 5183-5189.

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

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

Levy, S., Sures, I. and Kedes, L. (1982) J. Biol. Chem. 257, 9438-9443.

Lewin, R. (1983) Science 219, 1052-1054.

Liao, L.W. and Cole, R.D. (1981a) J. Biol. Chem. 256, 6751-6755.

Liao, L.W. and Cole, R.D. (1981b) J. Biol. Chem. 256, 10124-10128.

Liao, L.W. and Cole, R.D. (1981c) J. Biol. Chem. 256, 11145-11150.

Liao, L.W. and Cole, R.D. (1981d) J. Biol. Chem. 256, 3024-3029.

Liebermann, D., Hoffmann-Liebermann, B., Wienthal, J., Childs, G., Maxson, R., Mauron, A., Cohen, S.N. and Kedes, L. (1983) Nature 306, 342-347.

Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977) Cold Spring Harbor Symp. Quant.Biol. 42, 1047-1051.

Lohr, D. (1983) Nucl. Acids Res. 11, 6755-6773.

Lohr, D. and Hereford, L. (1979) Proc. Natl. Acad. Sci. USA 76, 4285-4288.

Losa, R., Thoma, F. and Koller, Th. (1984) J. Mol. Biol. 175, 529-551.

Lüscher, B., Stauber, C., Schindler, R. and Schümperli, D. (1985) Proc. Natl. Acad. Sci. USA 82, 4389-4393.

Macleod, A.R., Wong, N.C.W. and Gordon, G.H. (1977) Eur. J. Biochem. 78, 281-291.

Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) Proc. Natl. Acad. Sci. USA 72, 1184-1188.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Conell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell 15, 687-701.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) "Molecular Cloning (a Laboratory Manual)", Cold Spring Harbor Laboratory, New York.

Marashi, F., Prokopp, K., Stein, J. and Stein, G. (1984) Proc. Natl. Acad. Sci. USA 81, 1936-1940.

Marzluff, W.F. and Graves, R.A. (1984) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 281-315.

Mauron, A., Kedes, L., Hough-Evans, B.R. and Davidson, E.H. (1982) Dev. Biol. 94, 425-434.

Maxson, R.E. and Wilt, F.H. (1981) Dev. Biol. 83, 380-386.

Maxson, R.E. and Wilt, F.H. (1982) Dev. Biol. 94, 435-440.

Maxson, R., Mohun, T. and Kedes, L. (1983a) in "Eukaryotic Genes, their Structure, Activity and Regulation", Maclean, N., Gregory, S.P. and Flavell, R.A., eds., Butterworths and Co., Cambridge U.K., pp. 277-298.

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

Maxson, R., Mohun, T., Gormezano, G., Childs, G. and Kedes, L. (1983c) Nature 301, 120-125.

McGhee, J.D. and Felsenfeld, G. (1980) Ann. Rev. Biochem. 49, 1115-1156.

McGhee, J.D., Rau, D.C., Charney, E. and Felsenfeld, G. (1980) Cell 22, 87-96.

McGhee, J.D., Nickol, J.M., Felsenfeld, G. and Rau, C.D. (1983) Cell 33, 831-841.

McKnight, S.L. (1982) Cell 31, 355-365.

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

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

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

McMaster, G.K. and Carmichael, G.G. (1977) Proc. Natl. Acad. Sci. USA 74, 4835-4838.

Melton, D.W., Konecki, D.J., Brennand, J. and Caskey, C.T. (1984) Proc. Natl. Acad. Sci. USA 81, 2147-2151.

Messing, J. and Vieira, J. (1982) Gene 19, 269-276.

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

Molgaard, H.V., Perucho, M. and Ruiz-Carrillo, A. (1980) Nature 283, 502-504.

Montiel, J.F., Norbury, C.J., Tuite, M.F., Dobson, M.J., Mills, J.S., Kingsman, A.J. and Kingsman, S.M. (1984) Nucl. Acids Res. 12, 1049-1068.

Neelin, J.M. (1968) Can. J. Biochem. 46, 241-247.

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

Norrander, J., Kempe, T. and Messing, J. (1983) Gene 26, 101-106.

North, G. (1984) Nature 312, 308-309.

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

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

Osborne, H.B. and Chabanas, A. (1984) Exp. Cell Res. 152, 449-458.

Osborne, T.F., Arvidson, D.N., Tyau, E.S., Dunsworth-Browne, M. and Berk, A.J. (1984) Mol. Cell. Biol. 4, 1293-1305.

Osley, M.A. and Hereford, L.M. (1981) Cell 24, 377-384.

Osley, M.A. and Hereford, L. (1982) Proc. Natl. Acad. Sci. USA 79, 7689-7693.

Panyim, S., Bilek, D. and Chalkley, R. (1971) J. Biol. Chem. 246, 4206-4215.

Parker, C.S. and Topol, J. (1984a) Cell 36, 357-369.

Parker, C.S. and Topol, J. (1984b) Cell 37, 273-283.

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

Pehrson, J.R. and Cohen, L.H. (1984) Biochemistry 23, 6761-6764.

Pehrson, J.R. and Cole, R.D. (1981) Biochemistry 20, 2298-2301.

Pehrson, J.R. and Cole, R.D. (1982) Biochemistry 21, 456-460.

Perry, R.P. (1984) Nature 310, 14-15.

Phelan, J.J., Subirana, J.A. and Cole, R.D. (1972) Eur. J. Biochem. 31, 63-68.

Plumb, M., Marashi, F., Green, L., Zimmerman, A., Zimmerman, S., Stein, J. and Stein, G. (1984) Proc. Natl. Acad. Sci. USA 81, 434-438.

Plumb, M.A., Nicolas, R.H., Wright, C.A. and Goodwin, G.H. (1985) Nucl. Acids Res. 13, 4047-4065.

Poccia, D., Salik, J. and Krystal, G. (1981) Dev. Biol. 82, 287-296.

Poirier, G.G., De Murcia, G., Jonstra-Bilen, J., Neidergang, C. and Mandel, P. (1982) Proc. Natl. Acad. Sci. USA 79, 3423-3427.

Powell, J. (1984) Honours thesis, University of Adelaide, Adelaide, South Australia.

Preston, C.M., Cordingley, M.G. and Stow, N.D. (1984) J. Virol. 50, 708-716.

Price, D.H. and Parker, C.S. (1984) Cell 38, 423-429.

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

Proudfoot, N. (1984) Nature 307, 412-413.

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

Rao, B.J., Brahmachari, S.K., Rao, M.R.S. (1983) J. Biol. Chem. 258, 13478-13485.

Reeder, R.H. (1984) Cell 38, 349-351.

Reeves, R. (1984) Biochim. Biophys. Acta. 782, 343-393.

Renz, M. and Day, L.A. (1976) Biochemistry 15, 3220-3228.

Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L. (1984) Cell 38, 275-285.

Richmond
Rushton, T.J., Finch, J.T., Rushton, B., Rhodes, D. and Klug, A. (1984) Nature 311, 532-537.

Rickles, R., Marashi, F., Sierra, F., Clark, S., Wells, J., Stein, J. and Stein, G. (1982) Proc. Natl. Acad. Sci. USA 79, 749-753.

Risley, M.S. and Eckhardt, R.A. (1980) Eur. J. Cell Biol. 22, 80.

Risley, M.S. and Eckhardt, R.A. (1981) Dev. Biol. 84, 79-87.

Roberts, S.B., Weisser, K.E. and Childs, G. (1984) J. Mol. Biol. 174, 647-662.

Robins, A.J., Harvey, R.P., Wang, S.-W., Sturm, R., Dalton, S., Coleman, J.R. and Wells, J.R.E. (1985) Submitted.

Roche, J., Gorka, Cl., Goeltz, P. and Lawrence, J. (1985) Nature 314, 197-198.

Rodman, T.C., Litwin, S.D., Romani, M. and Vidali, G. (1979) J. Cell Biol. 80, 605-620.

Roychoudhury, R., Jay, E. and Wu, R. (1976) Nucl. Acids Res. 3, 863-877.

Ruberti, I., Fragapane, P., Pierandrei-Amaldi, P., Beccari, E., Amaldi, F. and Bozzoni, I. (1982) Nucl. Acids Res. 10, 7543-7559.

Ruiz-Carrillo, A., Wangh, L.J., Littau, V.C. and Allfrey, V.G. (1974) J. Biol. Chem. 249, 7358-7368.

Ruiz-Carrillo, A., Wangh, L.J. and Allfrey, V.G. (1976) Arch. Biochem. Biophys. 174, 273-290.

Ruiz-Carrillo, A., Affolter, M. and Renaud, J. (1983) J. Mol. Biol. 170, 843-859.

Ruiz-Vazquez, R. and Ruiz-Carrillo, A. (1982) Nucl. Acids Res. 10, 2093-2107.

Rutledge, R.G., Neelin, J.M. and Seligy, V.L. (1984) Eur. J. Biochem. 144, 191-198.

Rykowski, M.C., Wallis, J.W., Choe, J. and Grunstein, M. (1981) Cell 25, 477-487.

Ryoji, M. and Worcel, A. (1984) Cell 37, 21-32.

Ryoji, M. and Worcel, A. (1985) Cell 40, 923-932.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Sassone-Corsi, P., Wildeman, A. and Chambon, P. (1985) Nature 313, 458-463.

Schlissel, M.S. and Brown, D.D. (1984) Cell 37, 903-913.

Séguin, C., Felber, B.K., Carter, A.D. and Hamer, D.H. (1984) Nature 312, 781-785.

Seligy, V., Roy, C., Dove, M. and Yaguchi, M. (1976) Biochem. Biophys. Res. Comm. 71, 196-202.

Seyedin, S.M. and Kistler, W.S. (1979) Biochemistry 18, 1371-1375.

Seyedin, S.M. and Kistler, W.S. (1980) J. Biol. Chem. 255, 5949-5954.

Seyedin, S.M. and Kistler, W.S. (1983) Exp. Cell Res. 143, 451-454.

Seyedin, S.M., Cole, R.D. and Kistler, W.S. (1981) Exp. Cell Res. 136, 399-405.

Shannon, M.F., Wigley, P.L. and Wells, J.R.E. (1985) FEBS Lett. 186, 180-186.

Sharp, P.A. (1983) Nature 301, 471-472.

Shen, S., Slightom, J.L. and Smithies, O. (1981) Cell 26, 191-203.

Sheridan, W.F. and Stern, H. (1967) Exp. Cell Res. 45, 323-335.

Shires, A., Carpenter, M.P. and Chalkley, R. (1975) Proc. Natl. Acad. Sci. USA 72, 2714-2718.

Sierra, F., Lichtler, A., Marashi, F., Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G. and Stein, J. (1982a) Proc. Natl. Acad. Sci. USA 79, 1795-1799.

Sierra, F., Leza, A., Marashi, F., Plumb, M., Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G.S. and Stein, J.L. (1983b) Biochem. Biophys. Res. Comm. 104, 785-792.

Simpson, R.T. (1978) Biochemistry 17, 5524-5531.

Simpson, R.T. (1981) Proc. Natl. Acad. Sci. USA 78, 6803-6807.

Simpson, R.T. and Bergman, L.W. (1980) J. Biol. Chem. 255, 10702-10709.

Sittman, D.B., Chiu, I.-M., Pan, C.-J., Cohn, R.H., Kedes, L.H. and Marzluff, W.F. (1981) Proc. Natl. Acad. Sci. USA 78, 4078-4082.

Sittman, D.B., Graves, R.A. and Marzluff, W.F. (1983a) Proc. Natl. Acad. Sci. USA 80, 1849-1853.

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

Sive, H.L., Heintz, N. and Roeder, R.G. (1984) Mol. Cell. Biol. 4, 2723-2734.

Slightom, J.L., Blechl, A.E. and Smithies, O. (1980) Cell 21, 627-638.

Smerdon, M.J. and Isenberg, I. (1976) Biochemistry 15, 4242-4247.

Smith, B.J., Walker, J.M. and Johns, E.W. (1980) FEBS Lett. 112, 42-44.

Smith, B.J., Cook, Y., Johns, E.W. and Weiss, R.A. (1981) FEBS Lett. 135, 77-80.

Smith, B.J., Harris, M.R., Sigournay, C.M., Mayes, E.L.V. and Bustin, M. (1984a) Eur. J. Biochem. 138, 309-317.

Smith,G.E. and Summers, M.D. (1980) Anal. Biochem. 109, 123-129.

Smith, M.M. (1984) in "Histone Genes, Sructure, Organizaton and Regulaton", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 3-33.

Smith, M.M. and Andrésson, O.S. (1983) J. Mol. Biol. 169, 663-690.

Smith, M.M. and Murray, K. (1983) J. Mol. Biol. 169, 641-661.

Smith, R.D., Yu, J. and Seale, R.L. (1984b) Biochemistry 23, 785-790.

Smith, R.D., Yu, J. and Annuziato, A. and Seale, R.L. (1984c) Biochemistry 23, 2970-2976.

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

Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.

Souza, L.M., Boone, T.C., Murdock, D., Langley, K., Wypych, J., Fenton, D., Johnson, S., Lai, P.H., Everett, R., Hsu, R.-Y. and Bosselman, R. (1984) J. Exp. Zool. 232, 465-473.

Srebreva, L. and Zlatanova, J. (1983) Biochim. Biophys. Acta 740, 163-168.

Srebreva, L.N., Andreeva, N.B., Gasargan, K.G., Tsanev, R.G. and Zlatanova, J.S. (1983) Differentiaton 25, 113-120.

Stahl, H. and Gallwitz, D. (1977) Eur. J. Biochem. 72, 385-392.

Stein, G.S. and Stein, J.L. (1984) Mol. Cell. Biochem. 64, 105-110.

Stein, G., Stein, J. and Kleinsmith, L.J. (1983) in "Eukaryotic Genes, their Structure, Activity and Regulation", Maclean, N., Gregory, S.P. and Flavell, R.A., eds., Butterworths and Co., Cambridge U.K., pp. 31-51.

Stein, G.S., Sierra, F., Plumb, M., Marashi, F., Baumbach, L., Stein, J.L., Carozzi, N. and Prokopp, K. (1984) in "Histone Genes, Structure, Organization and Regulaton", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 397-455.

Stephenson, E.C., Erba, H.P. and Gall, J.G. (1981a) Cell 24, 639-647.

Stephenson, E.C., Erba, H.P. and Gall, J.G. (1981b) Nucl. Acids Res. 9, 2281-2295.

Stimac, E., Groppi, V.E. and Coffino, P. (1984) Mol. Cell. Biol. 4, 2082-2090.

Strausbaugh, L.D. and Weinberg, E.S. (1982) Chromosoma 85, 489-505.

Strickland, W.N., Strickland, M., De Groot, P.C., Von Holt, C. and Wittmann-Liebold, B. (1980a) Eur. J. Biochem. 104, 559-566.

Strickland, W.N., Strickland, M., Brandt, W.F., Von Holt, C., Lehmann, A. and Wittmann-Liebold, B. (1980b) Eur. J. Biochem. 104, 567-578.

Strickland, W.N., Strickland, M. and Von Holt, C. (1982a) Biochim. Biophys. Acta 700, 127-129.

Strickland, W.N., Strickland, M., Von Holt, C. and Giancotti, V. (1982b) Biochim. Biophys. Acta 703, 95-100.

Strub, K., Galli, G., Busslinger, M. and Birnstiel, M.L. (1984) EMBO J. 3, 2801-2807.

Stunnenberg, H.G. and Birnstiel, M.L. (1982) Proc. Natl. Acad. Sci. USA 79, 6201-6204.

Sturm, R. (1985) Ph.D. thesis, University of Adelaide, Adelaide, South Australia.

Sugarman, B.J., Dodgson, J.B. and Engel, J.D. (1983) J. Biol. Chem. 258, 9005-9016.

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

Sures, I., Levy, S. and Kedes, L.H. (1980) Proc. Natl. Acad. Sci. USA 77, 1265-1269.

Sutcliffe, J.G. (1978) Nucl. Acids Res. 5, 2721-2728.

Tabata, T. and Iwabuchi, M. (1984) Gene 31, 285-289.

Tabata, T., Sasaki, K. and Iwabuchi, M. (1983) Nucl. Acids Res. 11, 5865-5875.

Tabata, T., Fukasawa, M. and Iwabuchi, M. (1984) Mol. Gen. Genet. 196, 397-400.

Tarnowka, M.A., Baglioni, C. and Basilico, C. (1978) Cell 15, 163-171.

Thoma, F. and Koller, Th. (1977) Cell 12, 101-107.

Thoma, F. and Koller, Th. (1981) J. Mol. Biol. 149, 709-733.

Thoma, F., Koller,Th. and Klug, A. (1979) J. Cell Biol. 83, 403-427.

Thoma, F., Losa, R. and Koller, Th. (1983) J. Mol. Biol. 167, 619-640.

Thomas, G. and Padayatty, J.D. (1983) Nature 306, 82-84.

Thomas, J.O. (1983) in "Eukaryotic Genes, their Structure, Activity and Regulation", Maclean, N., Gregory, S.P. and Flavell, R.A., eds., Butterworths and Co., Cambridge U.K., pp. 9-30.

Thomas, J.O. (1984) J. Cell Sci. Suppl. 1, 1-20.

Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.

Tiktopulo, E.I., Privalov, P.L., Odintsova, T.I., Ermokhina, T.M., Krasheninnikou, I.A., Aviles, F.X., Cary, P.D. and Crane-Robinson, C. (1982) Eur. J. Biochem. 122, 327-331.

Trostle-Weige, P.K., Meistrich, M.L., Brock, W.A., Nishioka, K. and Bremer, J.W. (1982) J. Biol. Chem. 257, 5560-5567.

Trostle-Weige, P.K., Meistrich, M.L., Brock, W.A. and Nishioka, K. (1984) J. Biol. Chem. 259, 8769-8776.

Tsanev, R. (1983) Mol. Biol. Rep. 9, 9-17.

Turner, P. (1985) Nature 316, 105-106.

Turner, P.C. and Woodland, H.R. (1983) Nucl. Acids Res. 11, 971-986.

Turner, P.C., Aldridge, T.C., Woodland, H.R. and Old, R.W. (1983) Nucl. Acids Res. 11, 4093-4107.

Urban, M.K. and Zweidler, A. (1983) Dev. Biol. 85, 421-428.

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

Uzman, J.A. and Wilt, F.H. (1984) Dev. Biol. 106, 174-180.

Valerio, D., Duyvesteyn, M.G.C., Dekker, B.M.M., Weeda, G., Berkvens, Th. M., Van der Voorn, L., Van Ormondt, H. and Van der Eb, A.J. (1985) EMBO J. 4, 437-443.

Van Dongen, W.M.A.M., Moorman, A.F.M. and Destrée, O.H.J. (1984) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 199-224.

Varshavsky, A., Levinger, L., Sundin, O., Barsoum, H., Özkaynak, E., Swerdlow, P. and Finley, D. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 511-528.

Vigneron, M., Barrera-Saldana, H.A., Baty, D., Everett, R.E. and Chambon, P. (1984) EMBO J. 3, 2373-2382.

Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Nuber, S.P., Chick, W.L. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA 75, 3727-3731.

Vitelli, L. and Weinberg, E.S. (1983) Nucl. Acids Res. 11, 2135-2153.

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

Von Holt, C., De Groot, P., Schwager, S. and Brandt, W.F. (1984) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 65-105.

Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687.

Waithe, W.I., Renaud, J., Nadeau, P. and Pallotta, D. (1983) Biochemistry 22, 1778-1783.

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

Wang, S. -W., Robins, A.J., D'Andrea, R. and Wells, J.R.E. (1985) Nucl. Acids Res. 13, 1369-1387.

Weinberg, E.S., Hendricks, M.B., Hemminki, K., Kuwabara, P.E. and Farrelly, L.A. (1983) Dev. Biol. 98, 117-129.

Weintraub, H. (1978) Nucl. Acids Res. 5, 1179-1188.

Weintraub, H. (1984) Cell 38, 17-27.

Weintraub, H., Worcel, A. and Alberts, B. (1976) Cell 9, 409-417.

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

Weiss, E.H., Mellor, A., Golden, L., Fahrner, K., Simpson, E., Hurst, J., Flavell, R.A. (1983) Nature 301, 671-674.

Welch, S.L. and Cole, R.D. (1979) J. Biol. Chem. 254, 662-665.

Welch, S.L. and Cole, R.D. (1980) J. Biol. Chem. 255, 4516-4518.

Wells, D. and Kedes, L. (1985) Proc. Natl. Acad. Sci. USA 82, 2834-2838.

Wells, J.R.E., Coles, L.S., D'Andrea, R., Harvey, R.P., Krieg, P.A., Robins, A. and Whiting J. (1983) in "Manipulation and Expression of Genes in Eukaryotes", Nagley, P., Linnane, A.W., Peacock, W.J. and Pateman, J.A., eds., Academic Press, Sydney, pp. 73-80.

Wenkert, D. and Allis, C.D. (1984) J. Cell Biol. 98, 2107-2117.

Whitton, J.L., Rixon, F.J., Easton, A.J. and Clements, J.B. (1983) Nucl. Acids Res. 11, 6271-6287.

Wigley, P. (1982) Honours thesis, University of Adelaide, Adelaide, South Australia.

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

Williams, A.F. (1970) Ph.D. thesis, University of Adelaide, Adelaide, South Australia.

Winter, E.P. (1984) Ph.D. thesis, State University of New York, Stony Brook.

Winter, E., Levy, D. and Gordon, J.S. (1985a) J. Cell Biol. 101, 167-174.

Winter, E., Palatnik, C.M., Williams, D.L., Coles, L.S., Wells, J.R.E. and Gordon, J.S. (1985b) J. Cell Biol. 101, 175-181.

Winter, G. (1980) EMBO Summer Course Notes.

Woudt, L.P., Pastink, A., Kempers-Veenstra, A.E., Jansen, A.E.M., Hager, W.H. and Planta, R.J. (1983) Nucl. Acids Res. 11, 5347-5360.

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

Wu, R.S., Nishioka, D. and Bonner, W.M. (1982) J. Cell Biol. 93, 426-431.

Wu, R.S., West, M.H.P., Bonner, W.M. (1984) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 457-483.

Yaguchi, M., Roy, C., Dove, M. and Seligy, V. (1977) Biochem. Biophys. Res. Comm. 76, 100-106.

Yaguchi, M., Roy, C. and Seligy, V.L. (1979) Biochem. Biophys, Res. Comm. 90, 1400-1406.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.

Yaniv, M. (1984) Biol. Cell 50, 203-216.

Zernik, M., Heintz, N., Boime, J. and Roeder, R.G. (1980) Cell 22, 807-815.

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

Zweidler, A. (1980) Dev. Biochem. 15, 47-56.

Zweidler, A. (1984) in "Histone Genes, Structure, Organization and Regulation", Stein, G.S., Stein, J.L. and Marzluff, W.F., eds., J. Wiley and Sons, New York, pp. 339-371.

Zweidler, A., Urban, M., Goldman, P. (1978) Miami Winter Symp. 15, 531.

Zwollo, P., Stein, G.S. and Stein, J.L. (1984) Biochem. Biophys. Res. Comm. 124, 988-993.

APPENDIX:

LIST OF PUBLICATIONS

Carozzi, N., Marashi, F., Plumb, M., Zimmerman, S., Zimmerman, A., Coles, L.S., Wells, J.R.E., Stein, G. and Stein, J. (1984). Clustering of human H1 and core histone genes. Science 224, 1115-1117.

Coles, L.S. and Wells, J.R.E. (1985). An H1 histone gene-specific 5' element and evolution of H1 and H5 genes. Nucl. Acids Res. 13, 585-594.

D'Andrea, R.J., Coles, L.S., Lesnikowski, C., Tabe, L. and Wells, J.R.E. (1985). Chromosomal organisation of chicken histone genes: Preferred associations and inverted duplications. Mol. Cell.Biol., in press.

Harvey, R.P., Krieg, P.A., Robins, A.J., Coles, L.S. and Wells, J.R.E. (1981). Non-tandem arrangement and divergent transcription of chicken histone genes. Nature 294, 49-53.

Harvey, R.P., Whiting, J.A., Coles, L.S., Krieg, P.A. and Wells, J.R.E. (1983). H2A.F: An extremely variant histone H2A sequence expressed in the chicken embryo. Proc. Natl. Acad. Sci. USA 80, 2819-2823.

Wells, J.R.E., Coles, L.S., D'Andrea, R.J., Harvey, R.P., Krieg, P.A., Robins, A. and Whiting, J. (1983). Chicken histone genes: Analysis of core, H1, H5 and variant sequences. In "Manipulation and Expression of Genes in Eukaryotes", Nagley, P., Linnane, A.W., Peacock, W.J. and Pateman, J.A., eds., Academic Press, Sydney, pp. 73-80.

Winter, E., Palatnik, C.M., Williams, D.L., Coles, L.S., Wells, J.R.E. and Gordon, J.S. (1985). Changes in the H-1 histone complement during myogenesis. II. Regulation by differential coupling of H-1 variant mRNA accumulation to DNA replication. J. Cell Biol. 101, 175-181.