ANALYSIS OF THE CHICKEN ERYTHROID-SPECIFIC H5 HISTONE GENE

By

Paul Llewellyn Boyer

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Biochemistry

1988

ABSTRACT

ANALYSIS OF THE CHICKEN ERYTHROID-SPECIFIC H5 HISTONE GENE

By

Paul Llewellyn Boyer

The chicken erythroid-specific H5 histone gene has been isolated from a phage library and analyzed. It was found that the chicken H5 histone gene is not closely linked (within 20-30 kb) to any other known histone gene and that only one H5 histone gene exists per haploid chicken genome.

The H5 histone gene was further studied by in vitro mutagenesis. The linker scanner mutagenesis procedure of McKnight and Kingsbury was utilized to construct both linker scanner mutants and deletion mutants within the putative chicken H5 histone promoter region. The deletion mutants were studied by transfection into QT6 quail fibroblasts. RNA was isolated from the transfected cells 36-48 hours after the transfection was completed and subjected to S1 analysis. Deletions within the promoter region have identified a number of possible H5 promoter elements. The H5 transcript steady state levels for each mutant was normalized to the steady state levels of the wild type H5 gene by utilizing an internal control mutant H5 gene. As a number of putative promoter elements were deleted, the H5 tran-

script steady state levels decreased, relative to the steady state levels for the normal H5 gene.

A mutant which deletes the 3' flanking region of the H5 gene was also found to have an effect on the steady state levels of the H5 transcript. Since this mutant has an intact promoter region, it is felt that this mutant has altered the processing of the H5 transcript.

To Mom and Dad

Acknowledgements

I would like to thank "the Boss", Jerry Dodgson, who put up with such minor details as rose bushes on the window-sill, ship models on my desk, and a grad student who only worked at night.

Thanks to Moriko Ito and Sue Kalvonjian for being friends without equal, and who let me ramble on about my hobbies without telling me to shut up.

Thanks to Jiing Dwan Lee, truly someone stranger than I

Thanks to Dr. Richard Schwartz who proofread this thesis in a record time.

And finally, thanks to JoEllyn McGrath who made the last few months here much more enjoyable. I will miss you more than you know.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

bp..... base pair

kb..... kilobase

LTR..... Long Terminal Repeat

A..... Adenine

C..... Cytosine

G..... Guanine

T..... Thymidine

Introduction

The chicken erythroid-specific H5 histone gene has a number of unusual characteristics. The H5 histone gene is not closely linked to any other known histone gene, and the H5 transcript is polyadenylated. While the H5 histone gene is considered to be a member of the H1 class of histones, many of the H1 histone promoter elements have been replaced in the H5 histone promoter. What regulatory elements are responsible for the expression of the H5 histone gene are uncertain.

The isolation of phage clones containing the chicken genomic H5 gene is described in Chapter 3. Study of the isolated phage clones indicated that all were isolates of the same H5 gene, and that the H5 histone gene was unlinked to other known histone genes. The H5 gene was subcloned into the plasmid pBR322. The promoter region of the H5 gene was studied by *in vitro* mutagenesis. Several hundred deletion clones were obtained from the Bal 31 digestions. The deletions' boundaries were roughly determined by restriction enzyme digestion. The exact deletion boundaries for 106 mutants have been determined by sequencing of the clones. The sequencing data allowed the construction of deletion mutants whose boundaries were known to one base pair.

The various mutants that were constructed were analyzed by transfection into QT6 quail fibroblasts (Chapter 4). RNA was isolated from the cells 36-48 hours after

transfection. The level of H5 transcripts within the assay was determined by S1 analysis. A mutant gene, which has a small deletion 3′ of the transcriptional start sites, was used as an internal control. The H5 transcript level of the various mutants were normalized to the expression level of this internal control gene. The effects of the various mutants are described in this chapter.

The last chapter (Chapter 5) includes two other studies of the H5 histone gene. The first part of the chapter studies the degree of DNA methylation at various stages of erythropoeisis. DNA from normal and RAV-1 infected erythroblast and erythrocyte cells were digested with the restriction enzymes CfoI, MspI, and HpaII. These enzymes differ in their sensitivity to DNA methylation in the form $C^{me}G$.

The second portion of Chapter 4 covers SacI restriction enzyme site polymorphisms for both the H5 histone gene and the α^{A}-globin gene. DNA from 14 inbred chicken lines as well as DNA from the domesticated turkey, the Japanese Quail, and the ring-necked pheasant were tested.

Chapter 1

Literature Review

Histones

Histones are a group of small basic proteins that interact with DNA within the nucleus to form the elemental subunit of chromatin structure, the nucleosome (97). There are five major classes of histones based upon their electrophoretic mobility: the core histones (H2A, H2B, H3, and H4), and the linker histone, H1 (73). An unusual histone variant, H5, is considered to be a member of the H1 class of histones and will be discussed in more detail later.

All of the core histones show a similar organization in their protein sequence. Most of the charged amino acids are present in the amino-terminal end of the protein while the carboxy-terminal end is hydrophobic in nature (73). This may indicate that the core histones share a common ancestral gene (55). The H1 histones have charged amino acids (mostly lysine) in both the amino- and carboxyl-terminals while the central region is apolar (73).

By protein sequence analysis, the histones are evolutionarily well conserved across species barriers (42,118,147). H2A and H2B (the slightly lysine rich histones) show the most variability among the core histones while H3 and H4 (the arginine rich histones) show little variation. The changes that do occur in the core histones are usually seen

in the hydrophilic amino-terminus rather than in the hydrophobic carboxyl-terminus (73). This is probably due to the function of the core histones within the nucleosome, and will be discussed below.

H1 histones (including H5) are very lysine rich and show the most variation of all the histones. Even so, H1 histones are well conserved evolutionarily. H1 histones have charged amino acids in both the amino- and carboxyl-terminal ends and have an apolar central region. Most changes in the protein sequence occur in one of the two hydrophilic termini rather than in the central region (73).

The histones were known to complex with DNA but the exact structural role of the histones was unclear until the discovery of the nucleosome (83,97), which is the basic subunit of chromatin structure. By various techniques (58, 89,105), the nucleosome was found to consist of a histone octamer core, containing an H3:H4 tetramer and two H2A:H2B dimers (Fig. 1). The basic amino-terminal ends of the core histones form the external surface of the nucleosome while the apolar carboxyl-terminal ends interact with each other inside the core (97). Wrapped twice around this histone core is 140 base pairs (bp) of DNA. Between two adjoining nucleosomes is anywhere from twenty to eighty base pairs of DNA. H1 (and H5) histones appear to bind in this "linker" region. With their two basic ends, they are believed to

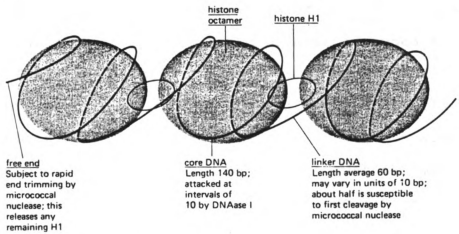


Figure 1 Structure of the nucleosome. The histone octamer consists of an H3:H4 tetramer and two H2A:H2B dimers. 140 base pairs of DNA is wrapped twice around this core. The H1 histone family (including histone H5) binds to the linker region between two adjoining nucleosomes.

Figure from Lewin (93)

"seal" the nucleosome ends where the DNA enters and leaves. A second function assigned to the H1 histones is in the formation of higher order chromatin structure. The H1 histone protein appears to be essential for forming the 30nm fiber, which consists of six nucleosomes arranged in a coil (73). This fiber itself can be further compacted to form even more complex structures whose molecular details are as yet unclear. As discussed below, the H5 histone also is responsible for the assembly of higher order chromatin structures. The replacement of H1 histone with the H5 histone protein within the erythroid chromatin condenses the chromatin, and may play a role in rendering the mature erythrocyte transcriptionally inactive.

Though the histones are usually grouped within the five major classes (H1, H2A, H2B, H3, H4), there exist variant subtypes of histone proteins within these classes. One type of histone variant is that of the modified histones. Modified histones are histones that have undergone post-translational chemical modifications. The most common such modifications are listed below.

Methylation is usually found on certain lysine groups for both H3 and H4 and appears to be reversible (73). Increased levels of methylation may be detected during late S and G2 phases of the cell cycle (165). The function of this

modification is unknown.

Acetylation may occur both on the N-terminal serine residue of the histones H1, H2A, and H4 and on lysine residues within the core histones (H2A, H2B, H3, and H4). The terminal N-acetyl-serine group appears to be a stable, irreversible modification but its function is unclear. In contrast, the internal acetylations are reversible through the action of the enzyme deacetylase (18). H2B, H3, and H4 have four modification sites while H2A has only one site (73). The core histones are acetylated soon after they are synthesized but not all possible sites are modified. In trout H4 for example, only two of the four possible sites are modified on the newly synthesized protein. The addition of the third and fourth acetyl groups procedes more slowly. Shortly after the trout H4 protein has been fully acetylated, removal of these groups begins by the action of the enzyme deacetylase until the protein is completely unmodified or has only one acetyl group remaining. This cycle takes around one day to complete (24,95). A somewhat similar series of acetylation/deacetylation reactions occurs for the other core histones (25,26). This cycle of acetyl group addition and removal creates heterogeneity in the internal extent of modification. Calf thymus histone H4 for example, has one of its modification sites (lysine 16) acetylated only 60% of the time and HeLa cell H4 histones are

acetylated only 40% of the time (93).

There is evidence that hyperacetylated histones congregate in certain discrete regions of the genome (32). A number of these regions are DNase I sensitive (114,148). This is of interest since DNase I sensitivity has been shown to occur in those areas of the genome that are transcriptionally active or potentially active (152,155). It is thought that the hyperacetylation of the histones affects their interactions with the DNA and allows the chromatin to relax, but this is conjecture (2).

Phosphorylation may occur for all the histones and usually is found on specific threonine groups as well as on histidine (H4) and lysine (H1 and H5). This modification is found mostly in the S phase of the cell cycle but may also occur under hormonal induction. The modification is usually lost after the anaphase stage (73). The phosphorylation of H5 does not appear to be cell cycle related, but is related to the stage of maturation of the erythroid cell (140,141).

Addition of a poly(ADP-ribose) moiety to the glutamate or aspartate amino acids of H2A, H2B, H3, and H1 is thought to affect the interactions between the histones but this is uncertain (73).

Attachment of the protein ubiquitin to lysine #119 of

H2A has also been found. Ubiquitin, as the name implies, is found both in prokaryotes and all eukaryotes (61). The ubiquitin:H2A complex is designated A24 and the cellular levels of this complex is thought to be related to the cell's mitotic activity (55).

Histone variants may also be grouped according to the timing of their synthesis (162,163). The major class (for which most histone genes isolated to date code) is that of the replication-dependent histones. The transcripts coding for these histones are most abundant in rapidly dividing tissues. These histone gene transcripts are absolutely dependent on DNA synthesis for their production and they degrade rapidly once DNA synthesis has halted. These histones are produced only during the S phase of the eukaryotic cell cycle.

A second group of histone variants is that of the partially replication-dependent histones. These were discovered during studies on regenerating mouse liver. These histones are produced during the initial stages of S phase but unlike the first class of histones, these continue to be synthesized after DNA replication has ceased.

A third group of histone variants is that of the replication-independent histones, also known as the replacement histones. These histones gradually accumulate during cell

maturation and replace a portion of the replication dependent histones within the chromatin of non-replicating cells such as liver, kidney, and erythrocytes (162,163). The best known examples are the H3.3 variants (23,53), histone H5, $H2A_F$ (70), and $H1^o$ from mammals (137). Messenger RNA coding for these histones is synthesized constitutively at a low level regardless of the replication state of the cell.

Histones may also be classified as to the stage at which they appear in the organism's maturation or development. This classification would include such stages as embryogenesis, spermatogenesis/oogenesis, and tissue specific maturation. One example is the development of the sea urchin embryo which undergoes rapid cell division. The first several cell divisions use histones translated from stored maternal mRNA (19,136), but after this the histones are synthesised from the early histone genes (128,142). As the embryo matures, expression from the early genes is repressed and instead, the late histone genes are used (33,116). There is no evidence however, for such a developmental use of different histone gene sets in avians or mammals.

Organization of histone genes

Since sea urchins produce large quantities of histone mRNAs during their maturation, it was likely that there were many histone genes within their genome. This made them a likely target for isolating the first histone genes. Sea urchin species that had evolved separately for millions of years had very similar organizational patterns for their histone genes (81). It was originally believed that most other organisms would share this arrangement or something similar. As more species have been studied however, it has become apparent that the organizational pattern of the sea urchin histone genes is unique. It reflects their need for large amounts of histone synthesis in a short period of development rather than a standard motif (Fig. 2).

Sea urchins have organized the five major early histone genes into a quintet that is tandemly repeated several hundred times (81). The order of the histones within the quintet (5′ to 3′ relative to transcription) is H4-H2B-H3-H2A-H1. They are all transcribed from the same strand of DNA, but there is no evidence that a polycistronic message is made. Each early histone gene appears to have its own promoter elements (76). Each gene within the quintet is separated from the flanking genes by a stretch of DNA that is AT-rich which appears to contain the needed regulatory elements.

Figure 2 Histone gene arrangements in selected organisms. The identity of each histone gene is shown, as well as the direction of transcription (if known). The map for the chicken histone genes identifies 35 out of the 42 histone genes mapped to date. The histone genes were isolated from two separate chromosome regions. The scale for the chicken histone gene map is half that of sea urchin and Drosophila histone gene maps.

The histone gene quintet of the Drosophila is roughly 5.0 kilo base pairs in length and is tandemly repeated. The histone gene quintet of the sea urchin covers a span roughly 6-7 kb in length and is also tandemly repeated.

Figures from S. Dalton, Ph.D. thesis, University of Adelaide, Adelaide, S. Australia

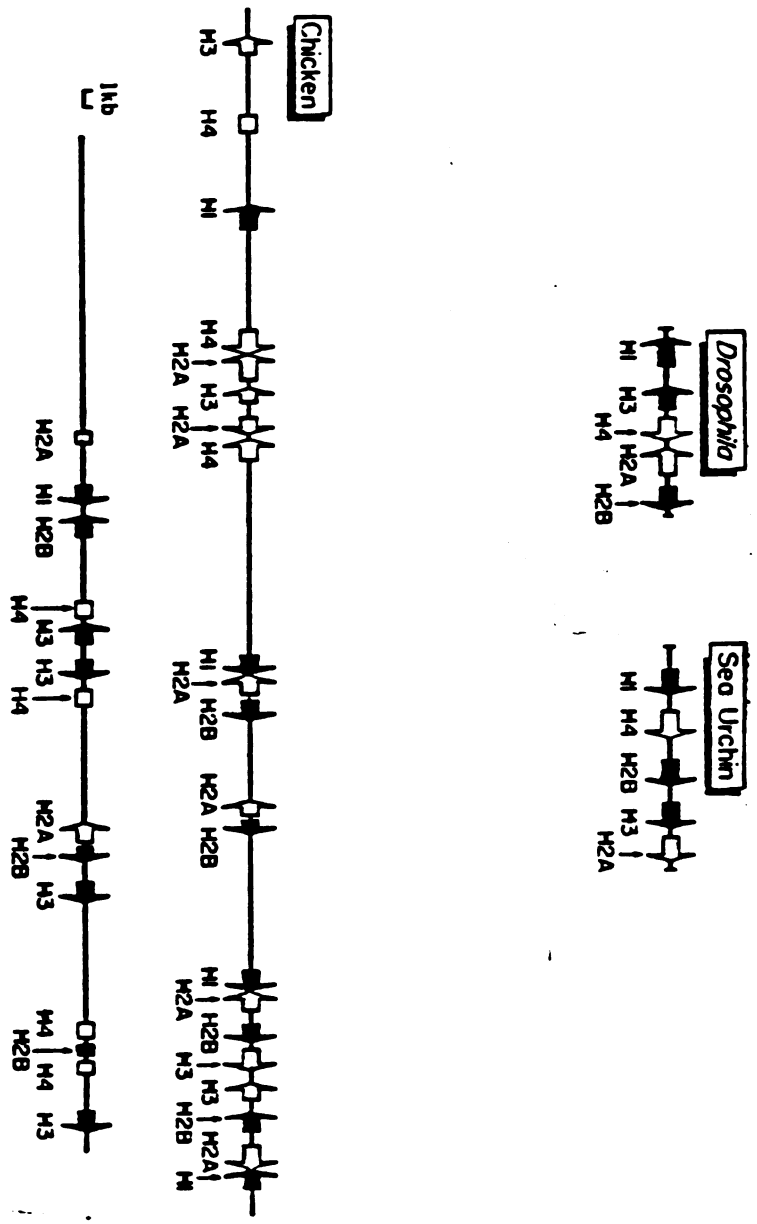

Chicken
H3
H4
H1
H4
H3
H4
H2A
H2A
H2A
H1
H2B
H4
H3
H3
H4
H2A
H1
H2B
H2A
H3
H2B
H4
H4
H3
H2B
H1
H4
H2B
H3
H2A
Drosophila
H1
H3
H4
H2A
H2B
Sea Urchin
H1
H4
H2B
H3
H2A
1kb

Figure 2

The sea urchin late histone genes have been found to be present in only 5-12 copies per haploid genome, and they are organized in clusters of irregular pattern similar to that of the vertebrate histone genes discussed below (96).

It was also discovered that a third class of sea urchin histone genes existed. These genes, termed orphons, appear to be solitary genes unlinked to any of the other histone genes. It is uncertain if some of these are transcriptionally active or if they are all psuedogenes (30).

Using probes made from the sea urchin early histone genes, the histone genes for Drosophila melanogaster were isolated (80). The organizational pattern of these genes shows similarities to that of the sea urchin but with distinct differences (Fig. 2). The order of the Drosophila histone genes is H1-H3-H4-H2A-H2B, and the genes are not all transcribed from the same DNA strand. This quintet is, like the sea urchin quintet, tandemly repeated but Drosophila has two distinct repeat units instead of one (94). Flies have been bred that lack one or the other repeat unit with no apparent deleterious effects (108,156). It appears as though either type of repeat unit is sufficient to maintain the required level of histone. Why two sets are maintained and whether they actually are completely equivalent

is uncertain.

In higher organisms that have a reduced number of histone genes (compared to either the sea urchin or Drosophila), the regularity of the organization pattern decreases. Both birds (39,51,69,139) and mammals (28,134, 133,163) have irregular (i.e.no tandem repeat unit) clusters of histone genes. In the chicken there are two such clusters which vary in both content and in organization (51,139). There is no common order to the closely clustered chicken histone genes, although several examples of a few genes linked in an inverted repeat fashion can be noted. Furthermore an H2A gene is often but not always paired with an H2B gene in such a way that they are transcribed in opposite directions. It also appears that H3 genes tend to pair with H4 genes though this is not as common. The two clusters in chicken contain replication dependent and partially replication dependent histone genes. The replication-independent histone genes (mentioned above) appear to be unlinked to any other histone gene. This may be due to the differences in the way the replication-independent histone genes are expressed.

Eukaryotic Gene Regulatory Elements

Gene expression in eukaryotic cells may be modulated at a number of points. Transcriptional control is the most common form of regulation but modulation may also occur during the course of mRNA processing (removal of intervening sequences and addition of the CAP structure and the polyadenylate tail), in the transport of the mRNA from the nucleus to the cytoplasm, in the stability levels of the mRNA, during the translation of the mRNA into the protein product, and during post-translational modification of the protein (40,115).

Transcriptional regulation controls both the frequency and the timing (i.e., tissue specificity, hormone inducibility, etc.) of transcription from the gene. This regulation results from the interactions between specific promoter elements on the DNA with cellular factors.

One of the first promoter elements identified was the "TATA" box. This sequence (consensus sequence 5'-TATA$\overset{T\ T}{A}$AA-3') is highly conserved and is usually found 25 to 30 base pairs upstream from the transcriptional start site (36). This sequence appears to regulate the location of the mRNA start site since removing or mutating the "TATA" sequence reduces the number of correctly initiated transcripts (65, 149). Deletion mutations have indicated that the spacing

between the "TATA" box and the transcriptional start site is important (20). The "TATA" sequence also appears to be needed for maximal mRNA transcription (65,99).

A second common promoter element is the "CCAAT" box. This sequence (consensus 5'-GGPyCAATCT-3') is generally located in the region between 40 and 90 base pairs upstream of the mRNA start site (12) and has been shown to interact with other promoter elements (15,66). Unlike the "TATA" box, the "CCAAT" sequence seems to be involved only in regulating the levels of transcription, not in the selection of the transcriptional start site (44,100,106).

A third common promoter sequence is the G/C box (consensus 5'-GGGGCGGGGG-3') (79). This sequence is usually found in the region between 40 and 200 base pairs upstream from the transcriptional start site and is often present in multiple copies. This element does not have any obvious sequence symmetry but nevertheless, this element can function in either orientation (47,59). Like the "CCAAT" sequence, the G/C box is involved in the efficient transcription of the gene (12,29,56).

Proteins from cellular extracts have been found that bind to the "TATA" box (127,131), the "CCAAT" box (74,75), and the G/C box (22,47). While little is known about the mechanisms involved, it is known that interactions between

the "TATA" box and its factor(s) (41), between the "CCAAT" sequence and its factor(s) (e.g., "CCAAT" transcription factor or CTF) (75,101), and between the G/C sequence and its factor (designated SP1) (101) are required for either accurate initiation of the transcript or for efficient transcription.

Along with these general promoter elements are gene specific regulatory sequences and enhancers. Gene specific elements are usually detected by mutational analysis, gene fusions, or simply by comparing the DNA sequences of related genes. Examples of this would be the detection of an H1 specific sequence found in all H1 genes studied to date (34), and regulatory elements common to cAMP-regulated genes (35). These gene specific elements combine with the general promoter elements already described to provide correct regulation of gene expression. Each class or type of histone gene studied to date, usually has highly conserved, type-specific elements in the regulatory region of the gene.

Along with these other elements are enhancers. An enhancer was first discovered in the SV40 virus (13,67). Enhancers have several unusual features that differ from most other control elements. Enhancers are usually orientation and position independent (109) and may exert their influence over long distances (91). Enhancers may activate genes with heterologous promoters and as the name suggests,

expression of a gene is higher in the presence of an enhancer than in its absence (8,27). Enhancers were found in many other eukaryotic viruses besides SV40. While large stretches of sequence homology have not been detected (72), the enhancers are often functionally equivalent (92). A common feature of viral enhancers is the presence of tandemly repeated elements (43,120,150), though not all viruses, notably polyoma, have these (72).

Viral enhancers also tend to show host cell specificity (43,88). For example, replacing the natural enhancer of polyoma (which infects mouse cells) with the SV40 enhancer (which infects primate cells) changes the host cell preference of polyoma from mouse cells to primate cells (43). Similar cell type preferences have been noted for other viruses (110,120,132).

Since enhancers are widely used in viruses that infect eukaryotic cells, cellular genes were examined to determine if eukaryotic cells also use enhancers as a means gene regulation. A number of cellular enhancers have indeed been identified. The first cellular enhancer detected was the enhancer from the mouse immunoglobulin heavy chain locus. This enhancer has structural and functional similarities to the viral enhancers (9,60,102), and stimulates transcription when fused to heterologous genes (60). Interest-

ingly, this stimulation is strictly lymphoid-specific (9, 60).

Several more enhancer sequences have since been discovered and many more probably exist. Examples include the pancreas-specific enhancer from the rat elastase I gene (117), the pancreas specific enhancer from the rat insulin II gene (68), the rat prolactin and growth hormone enhancer (113), the chicken β-globin enhancer (which is located in the 3′ flanking region of the gene) (31,78), and the chicken histone H5 enhancer which also is present in the 3′ flanking region of the gene (146).

Erythropoiesis

Erythropoiesis is the process by which an erythroid cell differentiates from a pluripotent stem cell to a non-dividing, transcriptionally inactive erythrocyte (Fig. 3). During this process, various tissue specific genes must be activated so that the erythrocyte can function as the oxygen carrying cell of the organism. These include those genes that encode enzymes for the manufacture of the heme ring (62,159), the genes for the α- and β-globin protein subunits of hemoglobin (45,50,52), the carbonic anhydrase II gene (57,161), various erythroid antigens and membrane proteins (82,112,125,153), cytoskeletal proteins (54,90, 160), iron metabolism proteins, and the erythroid specific histone H5.

After the precursor stem cell, the first identifiable erythroid cell type are the colony forming unit-marrow cells (CFU-M). These cells are self-renewing and can colonize the bone marrow of irradiated chickens, hence the name (124,125). Cells of this type continue to to replicate indefinitely in the self-renewal mode or they may irreversibly commit to terminal hematopoietic differentiation.

After the CFU-M stage, the cell enters the burst forming unit-erythroid (BFU-E), which is then followed by the

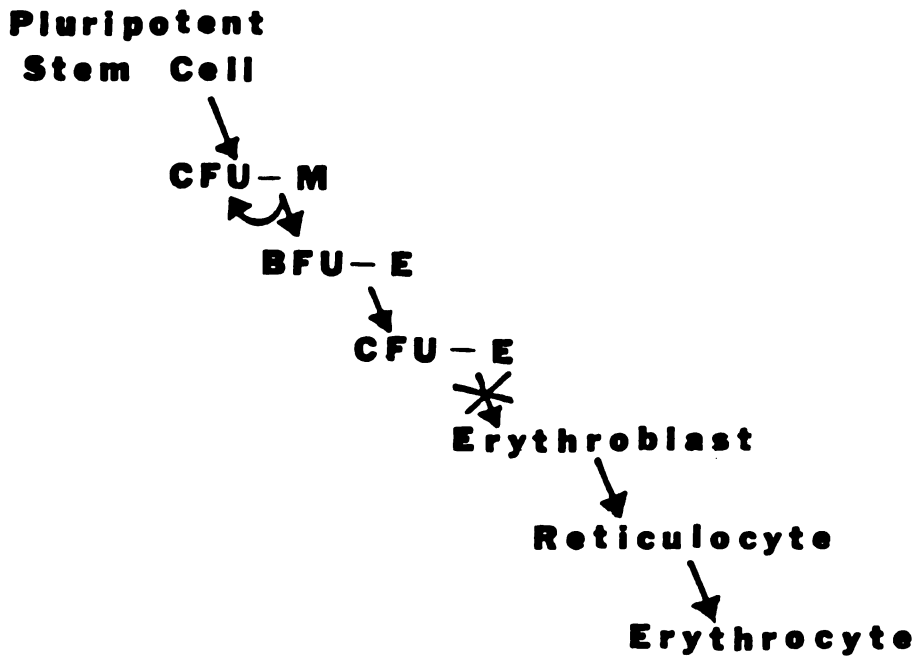


Figure 3 Pathway of chicken erythropoiesis. The CFU-M stage may either continue to self-replicate itself or irreversibly commit to further differentiation. The "X" between the CFU-E and erythroblast stages indicates the differentiation step blocked by the Avian Erythroblastosis Virus (AEV). Abbreviations: CFU-M→Colony forming unit-Marrow, BFU-E→ Burst forming unit-erythroid, CFU-E→ Colony forming unit-erythroid

(124,125,126,164)

colony forming unit-erythroid (CFU-E) stage (126). These cells are defined by their potential to produce different numbers of cells when cultured *in vitro*. Bursts may be distinquished from colonies in that bursts have a large number of cell groups clustered together (usually 3-20 groups) with each group containing 8-60 erythroid cells. Colonies on the other hand are a single group with 8-150 cells within the group. The two cell types may also be identified by their antigen expression and their response to growth factors. CFU-E cells show an absolute requirement for erythropoietin before they will continue to differentiate while BFU-E cells do not have this requirement.

After the CFU-E stage is the erythroblast stage. These cells are the last erythroid cells capable of cell division, and hemoglobin production begins at this stage. The cell size begins to decrease and the nucleus begins to condense as the erythroblast matures through the reticulocyte stage and finally into the erythrocyte stage. Reticulocytes retain mRNA and protein synthetic activities while little or no synthetic activity occurs in the mature erythrocytes. As described below, it has been hypothezised that H5 histone may play a role in the differentiation process.

H5 Histone Protein

H5 histone (previously designated V or f2c) was originally detected as an additional band during gel electrophoresis of histone proteins. It was only seen in histone extracts from mature avian erythrocytes and not from mammalian histone extracts. H5 was found to be in all nucleated erythrocytes tested, i.e., reptiles, fish, amphibians, and birds (103,111). The presence of H5 can be detected in the earliest testable precursors including both adult and embryonic erythroid cells arrested at the CFU-E stage by a temperature-sensitive avian erythroblastosis virus (14,unpublished data). It is also the only histone produced by the nearly mature, non-dividing reticulocyte (5). The levels of H5 are low in the precursor cells, lower than the levels for H1, but H5 gradually accumulates as the cell matures. As the concentration of H5 increases, H5 tends to displace H1 from the chromatin (6,121), until the concentration ratio of H5 to H1 in the chromatin is roughly 2:1 (10). This replacement correlates with the gradual compaction of the erythroid chromatin (143,144) and probably plays a role in the decreased levels of transcription seen in the reticulocyte and erythrocyte (151). In this respect it resembles the mammalian H1 histone variant $H1^{o}$. The mammalian histone however, is not tissue specific like H5

and is found in most adult non-dividing cells (137).

H5 histone is thought to function in rendering the nucleus transcriptionally inactive but uncertainty remains as to why it can be detected so early in the differentiation process when such a function would not seem to be needed until the late reticulocyte stage. It is thought that the concentration of the H5 protein in the earlier dividing cells remains below a certain threshold level which allows transcription to proceed. It appears as though the concentration of H5 protein within the erythroid cell does not markedly increase until the late erythroblast stage (1). The level of post-translational modification may also play a role. Five sites have been found in the protein that may be phosphorylated. Two of these sites are in the amino terminal end, while the remaining three are in the carboxyl-terminal end. All predicted phosphopeptides may be detected in a tryptic digest of monophosphorylated H5 protein, indicating that phosphorylation at any given site is a random event. The fact that the phosphorylation sites are present in the basic amino- and carboxyl-terminal ends suggests that the phosphorylation event may interfere with the DNA binding potential of the H5 histone (141). It has been determined that the phosphorylation levels of the H5 histone decreases as the erythroid cell matures. Early ery-

throid precursor cells isolated from chicken bone marrow have 70% of their H5 histone protein phosphorylated while erythroblasts and reticulocytes isolated from anemic chicken blood have only 50% of their H5 histone proteins modified. In comparison, mature erythrocytes isolated from normal chicken blood have little or no detectable phosphorylation. It has been hypothesized that the H5 phosphorylation present in the early precursor cells prevents the H5 histones from binding to the chromatin and displacing the H1 histones. As the cells mature, this modification is gradually removed and the H5 histone may bind to the chromatin (140).

From the protein sequence, H5 is usually considered to be a member of the H1 family of histones. The H5 protein has partial sequence homology with the H1 protein and is similar in its primary structure (i.e. charged amino acids in the amino- and carboxyl-terminal regions and an apolar central region). Like H1, H5 binds to the linker region between adjacent nucleosomes (7,145,157). When the protein sequences for the chicken H5 histone (21) and the goose H5 histone (158) are compared to the protein sequences of various H1 proteins (rabbit, trout, calf and sea urchin), the similarities suggests that these two proteins were derived from a common ancestor. It is apparent though, that these

two proteins have been evolving separately for a lengthy period of time. It is uncertain when the evolutionary divergence began, but since fish, reptiles, amphibians, and birds all have been shown to contain H5 histone proteins, it is probable that the separation occured relatively early in animal evolution (3,138,158).

H5 consists of two variant forms that differ only at amino acid #15. This amino acid may either be glutamine or arginine (21). Since one of the codons for glutamine is CAG and one of the codons for arginine is CGG, it was felt that a single base pair change was responsible for the two forms of H5 and that the two H5 proteins are different alleles of the same H5 gene. This was shown to be the case when the H5 gene was sequenced (discussed below).

H5 is not as well conserved between species as most other histone proteins. There is only 84% homology between the amino acid sequence of chicken H5 histone and the sequence of either the duck H5 histone or the goose H5 histone (21, unpublished observation). Besides these three complete H5 protein sequences, fragments of the pigeon and quail H5 proteins have also been sequenced (130,157). Comparisons between these various H5 proteins indicate that most changes occur in the amino-terminal end with relatively few changes in the carboxyl-terminal or central regions

(21). Unfortunately, only avian H5 proteins have been sequenced to date. Analysis of H5 histone proteins from fish and frogs however, indicates that these proteins have amino acid compositions similar to that of the chicken H5 protein (48,103).

H5 at the mRNA Level

All chicken replication dependent and partially replication dependent histone genes lack introns and are not polyadenylated (17,87,119). These genes are transcribed into a precursor RNA which is then processed to generate the correct 3′ terminal end. This process utilizes a conserved stem loop motif (77) and a U7 snRNP complex (129) to remove the unneeded 3′ portion of the RNA. The H5 histone gene does not have introns, and the H5 transcript was found to undergo the 3′ terminal polyadenylation common to most other cellular mRNAs (107). Other non-replication dependent chicken histone transcripts, including transcripts from the two known H3.3 variant genes (23,53), and from the $H2A_F$ gene (70), are also polyadenylated rather than contain the stem loop sequence.

H5, as described above, is one of the replication independent histones. When the cell enters the S phase of the cell cycle, the mRNA levels of replication dependent histones increase 10-20 fold. As the cell leaves the S phase and enters the G2 phase, these mRNA levels rapidly drop off. The increase in mRNA levels can be blocked by use of DNA synthesis inhibitors, such as cytosine arabinoside, hydroxyurea, and aphidicolin (11,63). The increase in mRNA levels is due both to an increase in the mRNA stability

and to an increase in the rate of transcription (4,71,135). The H5 histone gene appears to be constitutively expressed. How independent H5 is from the cell cycle is still being debated. Two groups used aphidicolin synchronized cells to arrive at different conclusions. One group (37) used both Northern blotting and *in vitro* pulse labeling. Erythroid cells were blocked in the cell cycle by use of aphidicolin. Once the block was released, the cells entered the cell cycle synchronously. At various time points both nuclei and RNA were isolated. By Northern blot analysis of the RNA, the steady state levels of the H5 mRNA appeared to remain constant while the other histone messages went up 15 fold at S phase and faded out at the G2 phase. By *in vitro* pulse labeling of the isolated nuclei, it also appeared that the H5 message was being transcribed at all points in the cell cycle. Their conclusion was that H5 is completely cell cycle independent.

Another group (1) also used aphidicolin blocked erythroid cells. They report that while H5 mRNA is transcribed at all points of the cell cycle, the levels fluctuate. In their studies, the rate of transcription during the G1 phase is only 65% of the levels found in the S or G2 phases. The differences between these two studies cannot be easily resolved but may be related to the respective sensitivity levels of the measurements.

The level of transcription of the H5 gene is low in erythroid precursor cells. Before the cells lose the ability to replicate (the erythroblast stage), the rate of transcription increases 6-fold compared to that of the other histone genes. As the cells enter the reticulocyte stage, the H5 gene is the only histone gene being transcribed. The level of H5 mRNA starts to decline as the cells near the end of their differentiation process even though the transcription rate is still high at that time. Transcription of the H5 gene probably ceases entirely as the cells enter the inactive erythrocyte stage (1).

H5 at the DNA Level

As the first step towards isolating the H5 histone gene, cDNA clones were constructed and identified. Two slightly different strategies were employed to isolate H5 cDNA clones. One group (85) utilized the known chicken H5 protein sequence (21) to construct a synthetic DNA primer complimentary to a portion of the H5 mRNA. The synthetic DNA primer was hybridized to chicken reticulocyte poly A^+ mRNA which was enriched for H5 mRNA. A cDNA library was constructed from this hybridization and screened with the synthetic DNA primer. Ten clones containing H5 histone sequences were isolated from the library. From this experiment, it was estimated that H5 mRNA represents only 0.2% of the chicken reticulocyte poly A^+ population. One of the cDNA clones was found to encode the arginine H5 protein variant, while two other cDNA clones encoded the glutamine variant. As predicted, the only DNA sequence difference between the two variants is the single base pair change required to change the arginine codon CGG to the glutamine codon CAG.

The second group (122) constructed a cDNA library from reticulocyte poly A^+ mRNA that was enriched for H5 histone sequences. The cDNA library was screened with an anti-H5 antibody:^{125}I-protein A complex, which would detect

any bacterial colonies that were producing the H5 protein. Three clones, including p541 which will be discussed later, were isolated. From this data, it was calculated that the H5 mRNA represented 0.1-0.2% of the total reticulocyte mRNA population, a figure in agreement with the figure mentioned above.

Using the cDNA clones as probes, the H5 gene has been isolated from a chicken phage library by several groups including our own (86,123,Chapter 3). It has been determined that the H5 gene is present only once per chicken haploid genome, compared to six known, unique H1 genes. Unlike most other known histone genes, the H5 gene is not closely linked to any other known histone gene. The DNA sequence of the H5 gene (Fig. 4) reveals many differences from the H1 genes. The H5 gene was found to be missing the H1 gene-specific 5′ element 5′-AAACACA-3′ which is present in all H1 genes sequenced to date (123). Instead, the H5 gene has replaced this A-rich sequence with the C-rich element 5′-CCGCCC-3′ (34). The H5 gene does not have a "CCAAT" box in its promoter region nor does it contain a consensus sequence "TATA" box. Instead of the canonical sequence "TATATAT", it has the related sequence "TTAAAT", which does not contain the "ATA" motif. This sequence ("TTAAAT") does occur in the region expected of a "TATA" box (see

Figure 4 DNA sequence of the chicken histone H5 gene. Sequence is from the group of Renaud (123). The gene shown here encodes for the arginine variant of the H5 protein. Putative control elements are underlined or enclosed by boxes. The two open arrows indicate the two transcription start sites, while the solid arrow at +726 indicates the site of polyadenylation. Direct and inverted repeat elements are underlined by horizontal arrows. Differences between this sequence and the H5 DNA sequence reported by Krieg et al (86) are shown above the sequence. Characters on top of the sequence and between nucleotides indicate base difreplacements and insertions in the gene sequenced by Krieg et al while hyphens represent deletions. The large open triangle indicates the position where the H5 cDNA clone p541 has a 9 base pair insertion relative to the H5 genomic clone shown here.

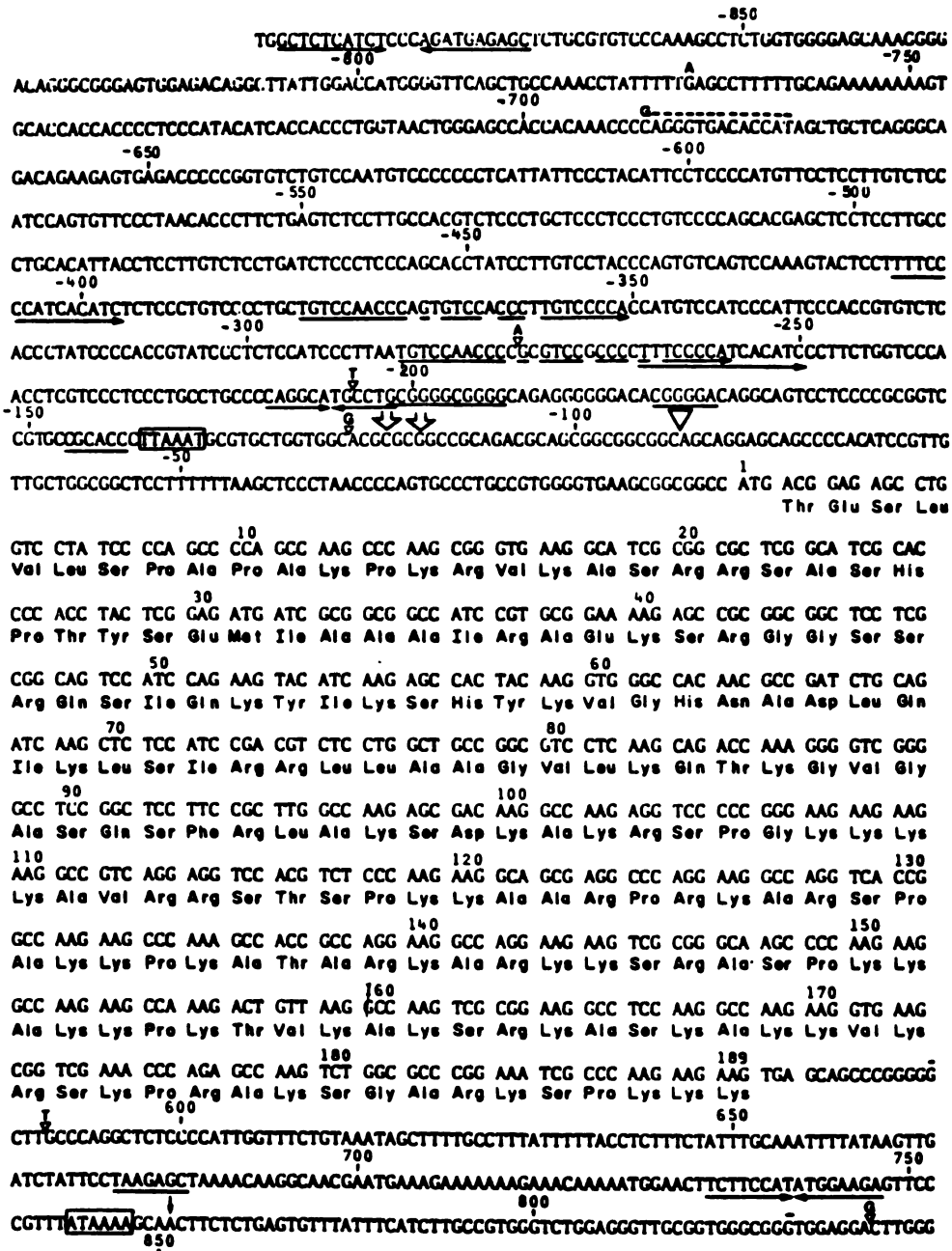


Figure 4

Promoter elements section), in this case 20 base pairs 5' of the transcriptional start site. There are no other AT-rich areas near the transcription start site that could act as a "TATA" box except for this sequence. It is interesting to note that the duck H5 gene also has the "TTAAAT" sequence at the same location as that of the chicken H5 "TTAAAT" sequence and that perfect homology exists between the two genes in the region surrounding this element, indicating selective pressure to keep this sequence (46). Other genes besides these two H5 genes have been found to contain noncanonical "TATA" boxes, including the "TTAAAA" sequence of the chicken lysozyme gene (64), the "ATTTAAA" element for the human intestinal alkaline phosphatase gene (104), and the "TTTAAAA" sequence of the Dictyostelium actin gene (98).

The body of the H5 histone gene is similar to that of most other histone genes in that it does not contain introns. Despite being a member of the H1 family of histones, the coding sequence for H5 has less homology with the coding sequence of H1 then might be expected (123). Computer analysis of the two coding sequences indicates that there is only 40% homology between an H1 histone gene and the H5 histone gene (unpublished data).

As mentioned before, most other histones utilize a stem

loop structure to produce the 3′ mRNA terminal end. The H5 gene does not contain this dyad repeat but may retain a remnant of it (34,123). Since the H5 mRNA is polyadenylated, the expected signal sequence, "AATAAA", was searched for but not found. A similar sequence, "TATAAA", was detected in the sequence of the cDNA clones (85), but it is only four base pairs away from the polyadenylation site. This appears to be too close; usually the distance between the signal sequence and the site of polyadenylation is 11-30 base pairs. The sequence, "TATAAA", may be present only by chance since this area is very AT-rich (84). It is known that the duck H5 gene does not have either the "AATAAA" or the "TATAAA" sequence present in the equivalent position. The 3′ untranslated region and the 3′ flanking region for both the duck and the chicken H5 histone genes contain two inverted repeats that may be involved in the processing of the mRNA 3′ terminal end. One possible stem loop structure is present near the 3′ end of the coding region while the second possible stem loop is near the polyadenylation site (46). Whether these stem loop elements are actually involved in the mRNA 3′ terminal processing is uncertain, but as far as is known, the H5 histone gene does not use the standard polyadenylation signal sequences.

The sequence analysis then, confirms that the H5 and H1

histones are related but only distantly. The H5 histone gene lacks many of the consensus 5′ and 3′ elements of the H1 histone genes and contains other regulatory elements in their place. One of these elements is an erythroid-specific enhancer (146). This H5 gene enhancer has structural and positional similarities to the chicken β-globin enhancer (31,78). Like the β-globin enhancer, the H5 gene enhancer is located in the 3′ flanking region of the gene. The enhancer position was identified by use of an "enhancer trap" plasmid (150) which allows the production of the SV40 large T antigen only when an enhancer sequence is inserted into the plasmid. Enhancer activity was detected in a restriction enzyme fragment isolated from the gene's 3′ flanking region. This activity, however, was detected only when the plasmid was introduced into erythroid-lineage chicken HD3 cells and not when introduced into non-erythroid chicken embryo fibroblasts. In addition, the level of the large T antigen production increased when the HD3 cells (which are arrested at the CFU-E stage by a temperature sensitive avian erythroblastosis virus) were induced to re-enter the erythroid differentiation pathway. These results indicate that the H5 enhancer is erythroid-specific, and that the enhancer may mediate the temporal induction of H5 gene transcription during erythropoiesis.

Smaller restriction enzyme fragments were tested in the "enhancer trap" plasmid to locate the position of the enhancer. An Xmn I- Xma III fragment near the 3′ terminal end of the H5 gene still retained enhancer activity. Within this fragment, a sequence was identified that had a large degree of homology with the β-globin enhancer sequence. When the two sequences were aligned, it was determined that there were 25 out of 34 correct matches. Such a high degree of similarity between the two sequences is unlikely to have arisen by chance. This may represent a 'core' enhancer sequence, but this is uncertain.

H5 enhancer	5′-GGAGGAGAGGGGACTCCTTCTTGTCCATAGGAGT-3′
	*** ******** * * * **** **** ***
β-globin enhancer	5′-GGAAGAGAGGGGGTTAATCC-TGTCAATAGTAGT-3′

A region that exhibits H5 gene-specific trans-activation has been detected on the 5′ side of the H5 gene (154). Xenopus oocytes were either injected with chicken histone genes or coinjected with chicken histone genes and chromatin salt wash fractions (CSWFs) isolated from chicken erythroid cells. Oocytes coinjected with H1 and H2B histone genes plus the erythroid CSWFs showed the same level of transcription from the histone genes as oocytes injected with the histone genes alone. Coinjection of the H5 gene

and erythroid CSWFs however, showed a 10-fold increase in the level of transcription from the H5 gene relative to that observed in oocytes injected only with the H5 gene. If the H2B gene was physically linked to the H5 gene, the H2B gene would also show an increase in its levels of transcription when coinjected with the erythroid CSWFs. Therefore, there appears to be a sequence (or sequences) within the H5 DNA that recognizes a factor (or factors) within the erythroid chromatin extracts. The binding of this factor(s) increases the level of H5 transcription in this system and may activate genes that are linked near it.

The H5 histone gene is not associated with the nuclear matrix in the same fashion as other histone genes. While the other histone genes appear to be associated with the nuclear matrix regardless of their transcriptional activity, the H5 gene appears to be associated with the nuclear matrix only in erythroid cells where it is being transcribed. It was not associated with the matrix in non-erythroid T-cell line. The region of association to the nuclear matrix falls within a 780 base pair area, spanning part of the coding region and all of the 5′ untranslated region. This apparently erythroid-specific association with the nuclear matrix suggests that the H5 gene is activated much more selectively than are the H1 histone genes (38).

BIBLIOGRAPHY

1) Affloter, M., Cote, J., Renaud, J., Ruiz-Carrillo, A., Mol. Cell. Biol. 7 pp. 3663-3672 (1987)

2) Allfrey, V., Faulkner, R., Mirsky, A. PNAS 51 pp. 786-794 (1964)

3) Allan, X., Hartman, P., Crane-Robinson, C., Aviles, F. Nature (London) 288 pp. 675-679 (1980)

4) Alterman, R., Ganguly, S., Schulze, D., Marzluff, W., Schildkraut, L., Skoultchi, A. Mol. Cell. Biol. 4 pp. 123-132 (1984)

5) Appels, R., Wells, J.R.E. J. Mol. Biol. 70 pp. 425-434 (1972)

6) Appels, R., Wells, J.R.E., Williams, A. J. Cell. Sci. 10 pp. 47-59 (1972)

7) Aviles, F., Chapman, G., Kneale, G., Crane-Robinson, C., Bradbury, C. Eur. J. Bioch. 88 pp. 363-371 (1971)

8) Banerji, J., Rusconi, S., Schaffner, W. Cell 27 pp. 299-308, (1981)

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

10) Bates, D., Thomas, J. NAR 9 pp. 5883-5894 (1981)

11) Baumbach, L., Marashi, F., Plumb, M., Stein, G., Stein, J. Biochemistry 23 pp. 1618-1625 (1984)

12) Benoist, C., O'Hare, K., Breathnach, R., Chambon, P. NAR 8 pp. 127-142 (1980)

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

14) Beug, H., Palmieri, S., Zentgraf, C., Graf, T. Cell 28 pp. 907-919 (1982)

15) Bienz, M. Cell 46 pp. 1037-1042 (1986)

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

17) Birchmeier, C., Schümperli, D., Sconzo, G., Birnstiel, M. PNAS 81 pp. 1057-1061 (1984)

18) Boffa, L., Vidali, G., Mann, P., Allfrey, V. J. Bio. Chem. 253 pp. 3364-3366 (1978)

19) Brandhorst, B. Dev. Bio. 79 pp. 139-148 (1980)

20) Breathnach, R., Chambon, P. Ann. Rev. Bioch. 50 pp. 349-383 (1981)

21) Briand, G., Kmiecek, D., Sautiere, P., Wouters, D., Borie-Loy, O., Biserte, G., Mazen, A., Champagne, M. FEBS 112 pp. 147-151 (1980)

22) Briggs, M., Kadonga, J., Bell, S., Tjian, R. Science 234 pp. 47-52 (1986)

23) Brush, D., Dodgson, J., Choi, O., Stevens, P., Engel, J. Mol. Cell. Bio. 5 pp. 1307-1317 (1985)

24) Candido, E., Dixon, G. J. Biol. Chem. 246 pp. 3182-3188 (1971)

25) Candido, E., Dixon, G. J. Biol. Chem. 247 pp. 3868-3873 (1972)

26) Candido, E., Dixon, G. PNAS 69 pp. 2015-2019 (1972)

27) Capecci, M. Cell 22 pp. 479-488 (1980)

28) Carozzi, N., Marashi, F., Plumb, M., Zimmerman, S., Zimmerman, A., Coles, L., Wells, J.R.E., Stein, G., Stein, J. Science 224 pp. 1115-1117 (1984)

29) Charnay, P., Mellon, P., Maniatis, T. Mol. Cell. Biol. 5 pp. 3261-3269 (1985)

30) Childs, G., Maxson, R., Cohen, R., Kedes, L. Cell 23 pp. 651-663 (1981)

31) Choi, O., Engel, J. Nature 323 pp. 731-734 (1986)

32) Christensen, M., Rattner, J., Dixon, G. NAR 12 pp. 4575-4592 (1984)

33) Cohen, I., Lowry, J., Kedes, L. Science 190 pp. 994-997 (1975)

34) Coles, L., Wells, J.R.E. NAR 13 pp. 585-594 (1985)

35) Comb, M., Birnberg, N., Seasholtz, A., Herbert, E., Goodman, H., Nature 323 pp. 353-356 (1986)

36) Corden, J., Wasylyk, B., Buchwalder, A., Chambon, P. Science 209 pp. 1406-1414 (1980)

37) Dalton, S., Coleman, J., Wells, J.R.E. Mol. Cell. Biol. 6 pp. 601-606 (1986)

38) Dalton, S., Youngblood, H., Wells, J.R.E. NAR 14 pp. 6507-6523 (1986)

39) D'Andrea, R., Coles, L., Wells, J.R.E. Mol. Cell. Biol. 5 pp. 3108-3115 (1985)

40) Darnell, J. Nature 297 pp. 365-371 (1982)

41) Davison, B., Egly, J., Mulvihill, E., Chambon, P. Nature 301 pp. 680-686 (1983)

42) DeLange, R., Fambrough, D., Smith, E., Bonner, J. J. Biol. Chem. 244 pp. 5669-5679 (1969)

43) deVilliers, J., Schaffner, W. Nature 312 pp. 242-246 (1984)

44) Dierks, P., VanOoyen, A., Cochran, M., Dobkin, C., Reiser, J., Weissman, C. Cell 32 pp. 695-706 (1983)

45) Dodgson, J., Strommer, J., Engel, J. Cell 17 pp. 879-887 (1979)

46) Doenecke, D., Tönjes, P. J. Mol. Biol. 178 pp. 121-135 (1984)

47) Dynan, W., Tjian, R. Nature 316 pp. 774-778 (1985)

48) Edwards, L., Hnilca, L. Experimentia 24 pp. 228-229 (1968)

50) Engel, J., Dodgson, J.B. PNAS 77 pp. 2596-2600 (1980)

51) Engel, J., Dodgson, J.B. PNAS 78 pp. 2856-2860 (1981)

52) Engel, J., Dodgson, J.B. J. Biol. Chem. 255 pp. 8239-8246 (1981)

53) Engel, J., Sugarman, B., Dodgson, J.B. Nature 297 pp. 434-436 (1982)

54) Elgsalter, A., Stokke, B., Mikkelson, A., Branton, D. Science 234 pp. 1217-1223 (1986)

55) Elgin, S., Weintraub, H. Ann. Rev. Bioch. 46 pp. 725-774 (1975)

56) Everett, R. NAR 11 pp. 6647-6666 (1983)

57) Ferrell, R., Stroup, S., Tanis, R., Tashian, R. J. Biol. Chem. 258 pp. 4459-4463 (1978)

58) Finch, J., Brown, R., Rhodes, D., Richmond, T., Rushton, B., Lutter, L., Klug, A. J. Mol. Biol. 145 pp. 757-769 (1981)

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

60) Gillies, S., Morrison, S., Oi, V., Tonegawa, S. Cell 33 pp. 717-728 (1983)

61) Goldstein, G., Scheid, H., Hammerling, U., Boyse, E., Schlesinger, D., Nioll, H. PNAS 72 pp. 11-15 (1975)

62) Granick, S., Beale, S. Advan. Embryol. 46 pp. 33-203 (1978)

63) Graves, R., Marzluff, W. Mol. Cell. Biol. 4 pp. 351-355 (1984)

64) Grez, M., Land, H., Giesecke, K., Schutz, G., Jung, A., Sippel, A. Cell 25 pp. 743-752 (1981)

65) Grosschedl, R., Birnstiel, M. PNAS 77 pp. 1432-1436 (1980)

66) Grosveld, G., deBoer,, E., Shewmaker, C., Flavell, R. Nature 295 pp. 120-126 (1982)

67) Gruss, P., Dhar, R., Khoury, G. PNAS 78 pp. 943-947 (1981)

68) Hanahan, D. Nature 315 pp. 115-122 (1985)

69) Harvey, R., Krieg, P., Robins, A., Coles, L., Wells, J.R.E. Nature 294 pp. 49-51 (1981)

70) Harvey, R., Whiting, J., Coles, L., Krieg, P., Wells, J.R.E. PNAS 80 pp. 2819-2823 (1983)

71) Heintz, N., Sive, H., Roeder, R. Mol. Cell. Biol. 3 pp. 539-550 (1983)

72) Hen, R., Borrelli, E., Chambon, P. NAR 11 pp. 8747-8760 (1983)

73) Isenberg, L. Ann. Rev. Bioch. 48 pp. 159-191 (1979)

74) Jones, K., Yamamoto, K., Tjian, R. Cell 42 pp. 559-572 (1985)

75) Jones, K., Kadonoga, J., Rosenfeld, P., Kelly, T., Tjian, R. Cell 48 pp. 79-89 (1987)

76) Hentschel, C., Birnstiel, M. Cell 25 pp. 301-313 (1981)

77) Hentschel, C., Irminger, J.-C., Bucher, P., Birnstiel, M. Nature 285 pp. 147-151 (1981)

78) Hesse, J., Nikol, J., Lieber, M., Felsenfeld, G. PNAS 83 pp. 4312-4316 (1986)

79) Kadonaga, J., Tjian, R. PNAS 83 pp. 5885-5893 (1986)

80) Karp, R., Hogness, D. Fed. Proc. 35 pp. 1623-1635 (1976)

81) Kedes, L. Ann. Rev. Bioch. 48 pp. 837-870 (1979)

82) Kline, K., Allison, J., Sanders, B. Develop. Biol. 72 pp. 389-396 (1982)

83) Kornberg, R. Ann. Rev. Bioch. 46 pp. 931-954 (1977)

84) Krieg, P., Robins, A., Colman, A., Wells, J.R.E. NAR 10 pp. 6777-6785 (1982)

85) Krieg, P., Robins, A., Gait, M., Titmas, R., Wells, J.R.E. NAR 10 pp. 1495-1502 (1982)

86) Krieg, P., Robins, A., D'Andrea, R., Wells, J.R.E. NAR 11 pp. 619-627 (1983)

87) Krieg, P., Melton, D. Nature 308 pp. 203-206 (1984)

88) Laimins, L., Gruss, P. PNAS 79 pp. 6453-6457 (1982)

89) Laskey, R., Earnshaw, W. Nature 286 pp. 763-767 (1980)

90) Lazarido, E., Nelson, W. Cell 31 pp. 505-508 (1982)

91) Lee, F., Mulligen, R., Berg, P., Ringold, G., Nature 294 pp. 228-232 (1981)

92) Levinson, B., Khoury, G., Van deWoulde, G., Gruss, P. Nature 295 pp. 568-572 (1982)

93) Lewin, B. Gene Expression 2 (Wiley Interscience, N.Y.) pp. 299-321 (1980)

94) Lifton, R., Goldberg, M., Karp, R., Hogness, D. CSH Symp. Quant. Biol. 42 pp. 1047-1051 (1977)

95) Louie, A., Dixon, G. PNAS 69 pp. 1975-1979 (1972)

96) Maxson, R., Mohun, T., Gormerzano, G., Childs, G., Kedes, L. Nature 301 pp. 120-125 (1983)

97) McGhee, J., Felsenfeld, G. Ann. Rev. Bioch. 49 pp. 1115-1153 (1980)

98) McKeown, M., Firtel, R. Cell 24 pp. 797-807 (1981)

99) McKnight, S., Kingsbury, R. Science 217 pp. 316-324 (1982)

100) McKnight, S. Cell 37 pp. 403-408 (1984)

101) McKnight, S., Tjian, R. Cell 46 pp. 795-805 (1986)

102) Mercola, M., Wang, X., Olsen,J., Calame,K. Science 221 pp. 663-665 (1983)

103) Miki, B., Neelin, J. Can. J. Bioch. 53 pp. 1158-1169 (1975)

104) Millan, J. NAR 15 pp. 1059-1065 (1987)

105) Mirzabekov, A., Belyovsky, A., Bauykin, S., Shick, V., Biosystems 12 pp. 265-271 (1980)

106) Miyamoto, N., Moncollin, V., Wintzerith, M., Hen, R., Egly, J., Chambon, P. NAR 12 pp. 8779-8799 (1984)

107) Molgaard, H., Perucho, M., Ruiz-Carrillo, A. Nature 283 pp. 502-504 (1980)

108) Moore, G., Procunier, J., Cross, D., Grigliatti, A. Nature pp. 312-314 (1979)

109) Moreau, P., Hen, R., Everett, R., Gaub, M., Chambon, P. NAR 9 pp. 6047-6068 (1981)

110) Mosthof, L., Pawlito, M., Gruss, P. Nature pp. 597-600 (1985)

111) Neelin, J., Callahan, P., Lamb, D., Murray, K. Can. J. Bioch. 42 pp. 1743-1752 (1964)

112) Nelson, C., Allison, J., Kline, K., Sanders, B. Can. Res. 42 pp. 4625-4630 (1982)

113) Nelson, C., Crenshaw, E., Franco, R., Rosenfeld, M. Nature 322 pp. 557-562 (1986)

114) Nelson, D., Ferris, P., Zhang, D., Fereny, C. NAR 14 pp. 1667-1682 (1986)

115) Nevins, J. Ann. Rev. Bioch. 52 pp. 441-446 (1983)

116) Newrock, K., Cohen, C., Hendricks, M., Donnelly, R., Weinburg, E. Cell 14 pp. 324-336 (1978)

117) Palmiter, R., Brinster, R. Ann. Rev. Genet. 20 pp. 465-499 (1986)

118) Patthy, L., Smith, E., Johnson, J. J. Biol. Chem. 248 pp. 6834-6840 (1973)

119) Price, D., Parker, C. Cell 38 pp. 423-429 (1984)

120) Rogers, B., Saunders, G. BioEssays 4 pp. 62-65 (1985)

121) Ruiz-Carrillo, A., Wangh, L., Littou, V., Allfrey, V. J. Biol. Chem. 249 pp. 7358-7368 (1974)

122) Ruiz-Vazquez, R., Ruiz-Carrillo, A. NAR 10 pp. 2093-2108 (1982)

123) Ruiz-Carrillo, A., Affolter, M., Renaud, J. J. Mol. Biol. 170 pp. 843-859 (1983)

124) Samarut, J., Jurdic, J., Nigon, V. J. Embryol. Exp. Morph. 50 pp. 1-20 (1979)

125) Samarut, J., Blanchet, J., Nigon, V. Develop. Biol. 72 pp. 155-166 (1979)

126) Samarut, J., Bouabdelli, M. J. Cell. Physio. 105 pp. 553-563 (1980)

127) Sawadago, M., Roeder, R. Cell 43 pp. 165-175 (1985)

128) Schaffer, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H., Birnstiel, L. Cell 14 pp. 655-671 (1978)

129) Schaufele, F., Gilmartin, G., Bannworth, W., Birnstiel, M. Nature 323 pp. 777-781 (1986)

130) Seligy, V., Roy, C., Dove, M., Yaguchi, M. Bioch. Bioph. Res. Comm. 71 pp. 196-202 (1976)

131) Selleck, S., Majors, J. Nature 325 pp. 173-177 (1987)

132) Shaul, Y., Rutter, W., Laub, O. EMBO 4 pp. 427-430 (1985)

133) Sierra, F., Lichtler, A., Marashi, F., Rickles,R., VanDyke, T., Clark, S., Wells, J., Stein, G., Stein, J. PNAS 79 pp. 1795-1799 (1982)

134) Sittman, D., Chiu, I., Pan, C., Cohn, R., Marzluff, W. PNAS 78 pp. 4078-4082 (1981)

135) Sittman, D., Graves, R., Marzluff, W. PNAS 80 pp. 1849-1853 (1983)

136) Skoulttchi, A., Gross, P. PNAS 70 pp. 2840-2844 (1973)

137) Smith, B., Harris, M., Sigourney, L., Mayes, E., Bustin, M. Eur. J. Bioch. 138 pp. 309-317 (1984)

138) Strickland, W., Strickland, M., Brandt, W., VonHolt, C., Lehmann, A., Whittman, B. Eur. J. Bioch. 104 pp. 567-578 (1980)

139) Sugarman, B., Dodgson, J.B., Engel, J. J. Biol. Chem. 258 pp. 9005-9016 (1983)

140) Sung, M. Biochemistry 16 pp. 286-290 (1977)

141) Sung, M., Freedlender, E. Biochemistry 17 pp. 1884-1890 (1978)

142) Sures,I., Lowry, J., Kedes, L. Cell 15 pp. 1033-1044 (1979)

143) Thoma, F., Koller, T., Klug, A. J. Cell. Biol. 83 pp. 403-427 (1979)

144) Thomas, J., Rees, C., Pearson, E. Eur. J. Bioch 147 pp. 143-151 (1985)

145) Torres-Martinez, S., Ruiz-Carrillo, A. NAR 10 pp. 2323-2335 (1982)

146) Trainor, C., Stamler, S., Engel, J. Nature 328 pp. 827-830 (1987).

147) Van Helden, P., Strickland, W., Strickland, M., Von Holt, C. Bioch. Bioph. Acta 703 pp. 17-20 (1982)

148) Vidali, G., Boffa, L., Bradbury, E., Allfrey, V. PNAS 75 pp. 2239-2243 (1978)

149) Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teule, R., Chambon, P. PNAS 77 pp. 7024-7028 (1980)

150) Weber, F., deVilliars, J., Schaffner, W. Cell 36 pp. 983-992 (1984)

151) Weintraub, H., Cambell, G., Holtzer, H. J. Cell. Biol. 50 pp. 652-668 (1971)

152) Weintraub, H., Groudine, M. Science 193 pp. 848-856 (1976)

153) Weise, M., Chan, L. J. Biol. Chem. 253 pp. 1892-1897 (1978)

154) Wigley, P., Wells, J.R.E. Mol. Cell. Biol. 7 pp. 3853-3856 (1987)

155) Wood, W., Felsenfeld, G., J. Biol. Chem. 257 pp. 7730-7736 (1982)

156) Wright, T., Hodgetts, R., Sherald, A. Genetics 84 pp. 267-285 (1976)

157) Yaguchi, M., Roy, C., Dove, M., Seligy, V. Bioch. Bioph.Res. Comm. 76 pp. 100-106 (1977)

158) Yaguchi, M., Roy, C., Seligy, V. Bioch. Bioph. Res. Comm. 90 pp. 1400-1406 (1979)

159) Yamamoto, M., Yew, N., Dodgson, J.B., Hayashi, N., Engel, J. PNAS 82 pp. 3702-3706 (1985)

160) Yew, N., Choi, H., Gallarda, J., Engel, J. PNAS 84 pp. 1035-1039 (1987)

161) Yoshihara, C., Dodgson, J.B. Annals of the N.Y. Academy of Science-Biology and Chemistry of the Carbonic Anhydrases 149 pp. 332-334 (1983)

162) Zweidler, A. Dev. Bioch. 15 pp. 47-56 (1980)

163) Zweidler, A. Histone Gene Structure, Organization and Regulation Stein, G., Stein, J., Marzluff, W., eds. J. Wiley and Sons, New York. pp. 339-375 (1984)

164) Gazzolo, L., Samurut, J., Boubdelli, M., Blanchet, J. Cell 22 pp. 683-691 (1980)

165) Borun, T., Pearson, D., Paik, W. J. Biol. Chem.
247 pp. 4288-4298 (1973)

Chapter 2

Materials and Methods

Materials

Restriction enzymes, calf alkaline phosphatase, T4 DNA ligase, T4 polynucleotide kinase, and E. coli DNA polymerase I were obtained from the following sources: Bethesda Research Laboratories, IBI, Promega Biotec, New England Biolabs, or Boeringer Mannheim. The nuclease Bal 31 (mixed form) was obtained from Bethesda Research Laboratories.

The plasmids pH5-BR 2.4, pH5-HR 1.5 and pH5-BR delta Cla I were constructed as described in the Results section. The H5 cDNA clone, p541 was obtained from Dr. Ruiz-Carrillo (University of Toronto)

QT6 and HD3 cells were both obtained from Dr. Hsing-Jien Kung (Case Western Reserve University).

Methods

Most of the cloning procedures listed below generally follow the protocols outlined by the Cloning Manual of Maniatis, Fritsch, and Sambrook (1).

Screening of the Chicken Phage Library

The library of Dodgson, Strommer and Engel (2) was screened for the presence of phage containing the H5 histone gene. The phage library was plated at a density of

50,000-100,000 plaque forming units (pfu) per 150 mm petri dish, using the E. coli strain K803 SupF as the host. The phage plaques were transferred to nitrocellulose filters and were treated as described by Maniatis et al. (1). The H5 cDNA clone, p541, was used to probe the resulting filters. To decrease the possibility of background problems, the p541 plasmid was digested with Pst I and the cDNA insert was isolated from the pBR322 vector sequences.

Any bacteriophage plaques from the initial screening which appeared to have hybridized to the p541 probe were rescreened at a lower density (100-200 pfu/100mm petri dish). Those bacteriophage plaques which tested positive in this second screening were isolated and a stock culture of the bacteriophage was obtained.

Bacteriophage DNA was isolated from bacteriophage purified on a CsCl step gradient (1). The DNA was then digested with various restriction enzymes and subjected to agarose gel electophoresis. The DNA was transferred to a nitrocellulose filter by Southern blotting. The filter was probed with the H5 cDNA clone, p541.

Bal 31 Digestion

The Bal 31 exonuclease digestions were generally performed as described by Maniatis et al (1). Five micrograms of the target plasmid were linearized with either Sac I (for pH5-BR 2.4) or Cla I (for pH5-BR delta Cla I). The DNA was then digested with 1-5 units of Bal 31 at 30° for the lengths of time indicated (3 to 5 minutes). The digestion was halted by the addition of EGTA (which chelates the necessary calcium ions) and cooling of the mixture to 4°. The samples were deproteinized by extraction with an equal volume of 1:1 phenol to chloroform. The DNA was ethanol precipitated, then made blunt-ended by treatment with DNA polymerase Klenow fragment. HindIII linkers were ligated to the DNA using T4 DNA ligase. The linkers were then digested with Hind III, and the excess Hind III linkers removed by ethanol precipitation. The plasmids were afterwards recircularized by ligation under dilute conditions and transformed into the E. coli HB101.

Individual colonies were picked and analyzed. DNA was isolated from the bacteria (3) and digested with either Pst I + Hind III (for the pH5-BR 2.4 derived clones, the "S" series) or Sac I + Hind III (for the pH5-BR delta Cla I derived clones, or the "C" series). The digestions were

subjected to gel electrophoresis on a 5% polyacrylamide gel, then stained with ethidium bromide. Those bacterial clones which were identified as having a HindIII linker within the target region were stored as a glycerol stock.

Sequencing

Clones were sequenced by the chemical degradation method of Maxam and Gilbert (4) as modified by Smith and Calvo (5). Individual clones were digested with Hind III to linearize the plasmid at the linker site. The DNA was treated with the enzyme calf alkaline phosphatase to remove the end terminal phosphate groups. The DNA was then radiotively labelled by incubating the DNA with T4 polynucleotide kinase and γ-^{32}P ATP. The unneeded radioactively labelled end was removed by digesting the DNA with either BamHI or Sal I.

Since the sequence of the H5 gene is known, only three of the five standard Maxam-Gilbert reactions were used. These three reactions (C+T, A+G, G) were sufficient to determine the portion of the 5' flanking region being studied. Only the terminal 20-30 base pairs of the clone's sequence needed to be determined. The chemical reactions

therefore, were allowed to proceed for twice the usual time. The reaction products were run on a 20% denaturing polyacrylamide gel and exposed to X-Ray film.

Cell Culture

QT6 cells (6) are chemically transformed quail fibroblasts. These cells are cultured in Dulbecco Modified Eagle's Medium (Grand Island Biological Co., Grand Island, NY) supplemented with 4% fetal calf serum, 1% chicken serum, and 1% v/v DMSO. HD3 cells (7,8) are chicken CFU-E erythroid cells differentially arrested by a temperature sensitive Avian Erythroblastosis Virus (ts-AEV). These cells are cultured in Dulbecco Modified Eagle's Medium supplemented with 8% fetal calf serum, 2% chicken serum, and 10 mM HEPES pH 7.4. Both cell types are grown at 37^{o} and 5% CO_2.

DNA Transfection

The transformation protocol is that of Wigler et al. (9). Twenty four hours before transformation, QT6 cells were plated to a density of 1-2 x 10^6 cells per 100mm tissue culture plate. Four to five hours prior to the

transformation, the old media was removed from the cells and replaced with 7 mls of fresh media.

25 to 50 micrograms of each individual clone's DNA was ethanol precipitated. If more than one clone was to be transformed into the same plate of cells, the DNA for both clones were co-precipitated. The DNA was resuspended in 0.62 mls of ddH_2O and adjusted to a final concentration of 250 mM $CaCl_2$ by addition of 0.08 mls of 2.5 M $CaCl_2$. The DNA/$CaCl_2$ mixture was rapidly added to an equal volume of 2X Hepes buffered saline (280 mM NaCl, 50 mM Hepes, 1.5 mM Na_2HPO_4, pH 7.10 ± 0.05). The DNA-calcium phosphate precipitate was allowed to form for 15-30 minutes at room temperature. The DNA-calcium phosphate mixture was then added to the QT6 cells. After 8-12 hours, this mixture was removed and replaced with fresh QT6 media. RNA was isolated from the cells 36-40 hours after the DNA-calcium phosphate mixture was removed.

RNA Isolation

For total cellular RNA, two 100 mm plates of QT6 cells were lysed with 5 mls of guanidinium isothiocyanate buffer (4 M guanidinium isothiocyanate, 50 mM Tris pH 7.5, 0.14 β-mercaptoethanol, 2% Sarcosyl, and 10mM EDTA). The cellular

homogenate was then passed five times through a 22 gauge needle to shear the high molecular weight DNA. Two grams of RNase free CsCl was added to every 5 mls of homogenate. This mixture was then layered over 4 mls of RNase free 5.7 M CsCl (in 0.03 M NaOAc, pH 5.2) in a polyallomer SW 41 tube. The samples were spun at 28,000 rpm for 16 hours in a SW 41 rotor. After the run was complete, the supernatant was carefully removed with an RNase free pipet. The RNA pellet was resuspended in 300 microliters of RNase free ddH_2O and extracted with an equal volume of 4:1 Chloroform to n-Butanol solution. NaOAc was added to a final concentration of 300 mM and the RNA precipitated with 2.5 volumes of ethanol.

For Poly A^+ RNA, two 100 mm plates of QT6 cells were lysed with 3 mls of Proteinase K buffer (0.5 M NaCl, 10 mM Tris pH 7.5, 1mM EDTA, 1.0% SDS, 200 micrograms/ml proteinase K) and the high molecular weight DNA released by the lysis sheared by passage through a 22 gauge needle. Fresh proteinase K (100 micrograms/ml) was added to the homogenate which was then incubated at 37^{o} for one hour.

Oligo dT cellulose was swelled in RNase free ddH_2O, then transferred to proteinase K buffer and proteinase K treated to remove any residual RNases. The oligo dT cellulose was added to the cell homogenates and the homogenates were

rocked at room temperature for one hour. The oligo dT cellulose was spun out of the solution and was washed twice with a high salt buffer (0.5 M NaCl, 10 mM Tris, pH 7.5, 1mM EDTA), then twice with a low salt buffer (0.1 M NaCl, 10 mM Tris pH 7.5, 1 mM EDTA). The poly A^+ RNA was removed from the cellulose by addition of two mls of RNase free ddH_2O. 50 micrograms of RNase free tRNA was added to the supernatant and the RNA precipitated with 2.5 volumes of ethanol.

S1 Analysis

Two DNA probes were used to analyze the RNA obtained from the QT6 transformations. Early studies utilized a BstEII restriction enzyme site present at +500 while later studies utilized a Eco0109 site present at +380. The procedure used to label this DNA is similar to that of Maniatis *et al*. Briefly, the plasmid pH5-BR 2.4 was digested with either BstEII or Eco0109, then the terminal phosphates were removed by the enzyme calf alkaline phosphatase. The DNA was separated on a 6% agarose gel, and the desired fragment isolated. This fragment was radioactively labelled by treatment with γ-^{32}P ATP and T4 polynucleotide kinase. The DNA was then digested with the restriction en-

zyme HinfI to remove the unneeded labelled end. The DNA was separated on a 6% agarose gel and the desired fragment was isolated.

The labelled DNA fragment was mixed with the RNA being studied and both were ethanol precipitated. The pellet was resuspended in 10 microliters of hybridization buffer (80% formamide, 0.4 M NaCl, 0.04 M Pipes, pH 7.25). The sample was heated at 90^{o} for 4 minutes to denature both the RNA and the DNA. The sample was allowed to hybridize at 55^{o} for 12 hours. After the hybridization was completed, 200 microliters of 1X S1 buffer (0.03 M NaOAc, pH 4.5, 0.25 M NaCl, 4mM ZnOAc, 100 micrograms/ml denatured, sheared salmon sperm DNA) and 200 units of S1 nuclease was added to the sample. The reaction was allowed to proceed for 15 minutes at room temperature, then was stopped by deproteination by extraction with an equal volume of a 1:1 phenol: chloroform mixture. The supernatant was ethanol precipitated. The reaction products were analyzed on a 6% denaturing polyacrylamide gel.

References

1) Maniatis, T., Fritsch, E., Sambrook, J. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, NY, 1982)

2) Dodgson, J., Strommer, J., Engel, J. Cell 17 pp. 879-887 (1979)

3) Ish-Horowicz, D., Burke, C. NAR 9 pp. 2989-2998 (1981)

4) Maxam, A., Gilbert, W. Methods Enzymol. 65 pp. 499-560 (1980)

5) Smith, D., Calvo, J. NAR 8 pp. 2255-2274 (1980)

6) Moscovici, C., Moscovici, G., Jimenez, H., Lai, M., Hayman, M., Vogt, P. Cell 11 pp. 95-103 (1977)

7) Beug, H., Palmieri, S., Freudenstein, C., Zentgraf, H., Graf, T. Cell 28 pp. 907-919 (1982)

8) Weintraub, H., Beug, H., Groudine, M., Graf, T. Cell 28 pp. 931-940 (1982)

9) Wigler, M., Pellicer, A., Silverstein, S., Axel, R. Cell 14 pp. 725-731 (1978)

Chapter 3

Results

Isolation of H5 Lambda Phage Clones

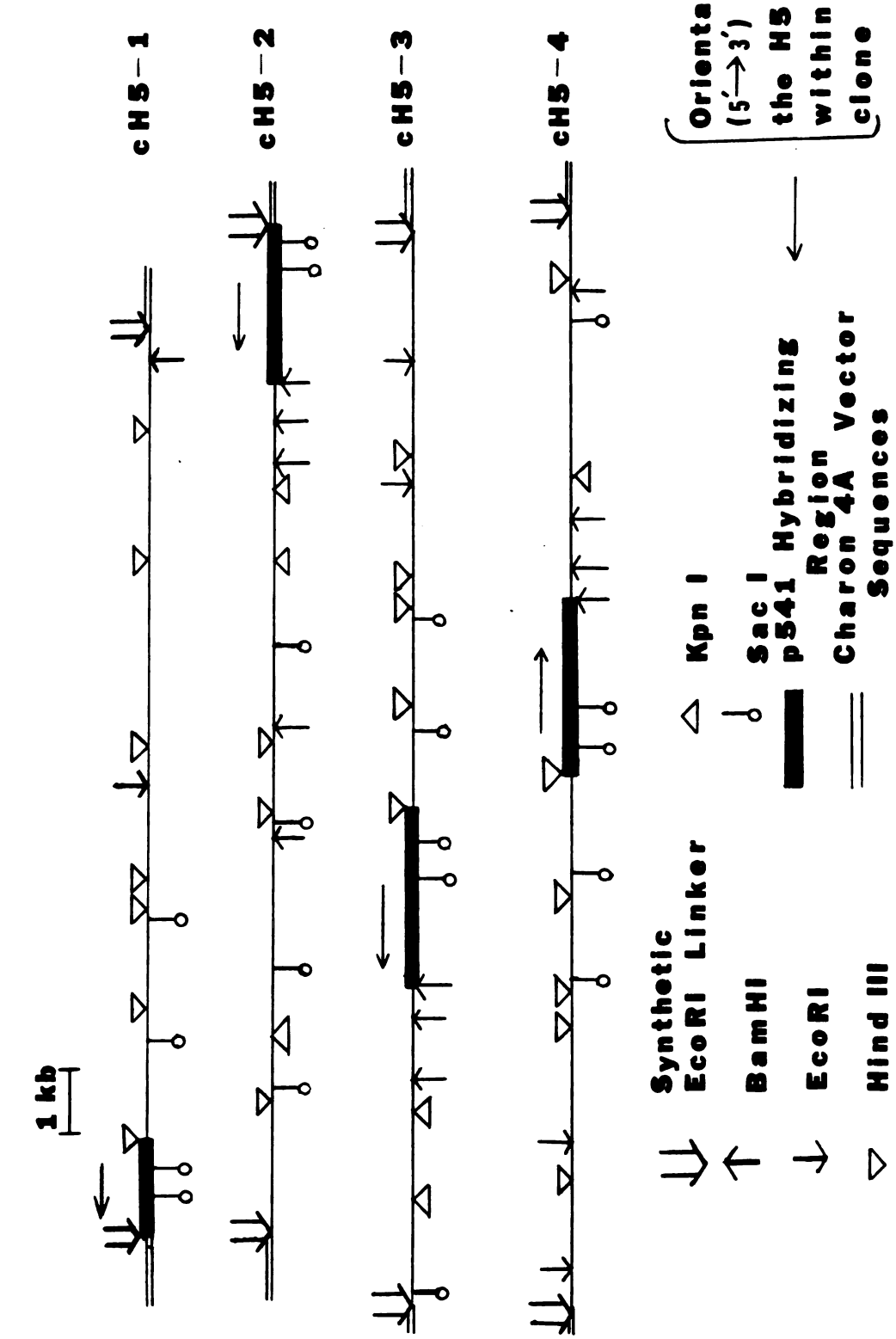
The first attempt at isolating lambda phage clones containing the chicken H5 histone gene made use of the fact that both the chicken replication-dependent and partially replication-dependent histone genes are loosely clustered together within the chicken genome. It was hoped that the H5 gene would be present within one of these histone gene clusters, even though it is a replication-independent histone gene. Fifty phage clones which were known to contain various histone genes were screened for the presence of an H5 gene with an H5 cDNA clone, p541 (1) which was obtained from Dr. Ruiz-Carrillo. The cDNA insert of p541 is 250 base pairs in length, and is cloned into the Pst I site of pBR322. None of the phage clones hybridized to the p541 probe, indicating that the H5 gene is not closely linked to any other known histone gene. Two other groups (4,5) have also determined that the H5 gene is a solitary histone gene. This is possibly due to the fact that the H5 gene is a replication-independent histone as well as a tissue-specific histone. Interestingly, the H5 probe did not cross-hybidize to any histone H1 genes, even under low stringency conditions. This is further evidence that the H5 gene has

evolved separately from the H1 genes for a lengthy period of time.

The chicken phage library of Dodgson, Strommer, and Engel (2) was then screened for lambda phage clones containing all or part of the H5 histone gene. Screening the phage library with the p541 probe identified two different phage clones, designated cH5-1 and cH5-2, which hybridized strongly to the probe. The restriction enzyme maps of the two phage clones and the p541 hybridizing region are shown in Fig. 5. It is not apparent from the maps whether these two clones are overlapping each other, and therefore contain the same H5 gene. While there is a 350 base pair SacI fragment in both clones within the region of H5 hybridization, this alone was not enough to prove that both phage clones contained the same H5 gene. At the time this work was being done, it had not been established that only one H5 gene/haploid genome was present in the chicken. On the contrary, an early study involving hybridization kinetics (3) indicated that the H5 gene might be present at up to 10 copies/haploid genome. While this has since been shown to be incorrect, it was not possible at the time these clones were being studied to state with certainty that both clones contained the same H5 gene. If several H5 genes did exist, more phage clones containing an H5 gene should have

Figure 5 Restriction enzyme maps of the four phage clones containing the H5 gene. The region which hybridized to the p541 H5 cDNA clone is shown as a thick line. The transcriptional direction of the H5 gene is shown above this region.

Figure 5


cH5-1
cH5-2
cH5-3
cH5-4
1 kb
Orientation (5'→3') of the H5 gene within the clone
Kpn I
Sac I
pS41 Hybridizing Region
Charon 4A Vector Sequences
Synthetic EcoRI Linker
BamHI
EcoRI
Hind III

been present in the phage library. The 2.4 kilobase EcoRI-BamHI fragment from cH5-2 which hybridized to the p541 probe was selected as a probe for rescreening the phage library. Two additional phage clones, designated cH5-3 and cH5-4 (Fig. 5), were isolated and analyzed. Comparisons between the four restriction enzyme maps indicate that not only are cH5-3 and cH5-4 similar to one another, but that they overlap both cH5-1 and cH5-2. All four phage clones appear to span the same region of the genome and contain the same H5 gene. This would indicate that there is only a single H5 gene/haploid genome rather than multiple copies. To test this further, DNA was isolated from chicken reticulocytes and digested with the restriction enzyme HindIII. After Southern blotting, the blot was probed with the 2.4 kb EcoRI-BamHI fragment from cH5-2. The H5 gene analyzed in our lab lies within a 7 kb HindIII fragment (cH5-4, Fig. 5). If there were other H5 histone genes, it was likely that they would be surrounded by different flanking regions and would appear as different sized bands on the Southern blot. As shown in Fig. 6, only a single hybridizing band, roughly 7 kb in size as predicted, was detected. This blot however, would not rule out a reiterated H5 gene.

Two other laboratories (4,5) have also isolated H5 histone phage clones. Several of these clones are similar to

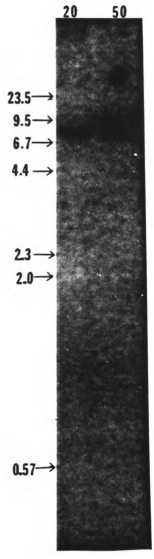


Figure 6 Chicken reticulocyte DNA was digested with HindIII and transferred to nitrocellulose paper by Southern blotting. 20 and 50 micrograms of DNA was tested in this gel. The blot was probed with the pH5-BR 2.4 subclone described in Figure 6. Predicted size of the hybridizing band is 7 kb.

clones isolated in our laboratory, and the restriction enzyme maps of the region surrounding their H5 genes are very similar or identical to the cloned region described above. That all three groups have isolated identical H5 genes is further evidence that only one H5 gene is present in the chicken genome.

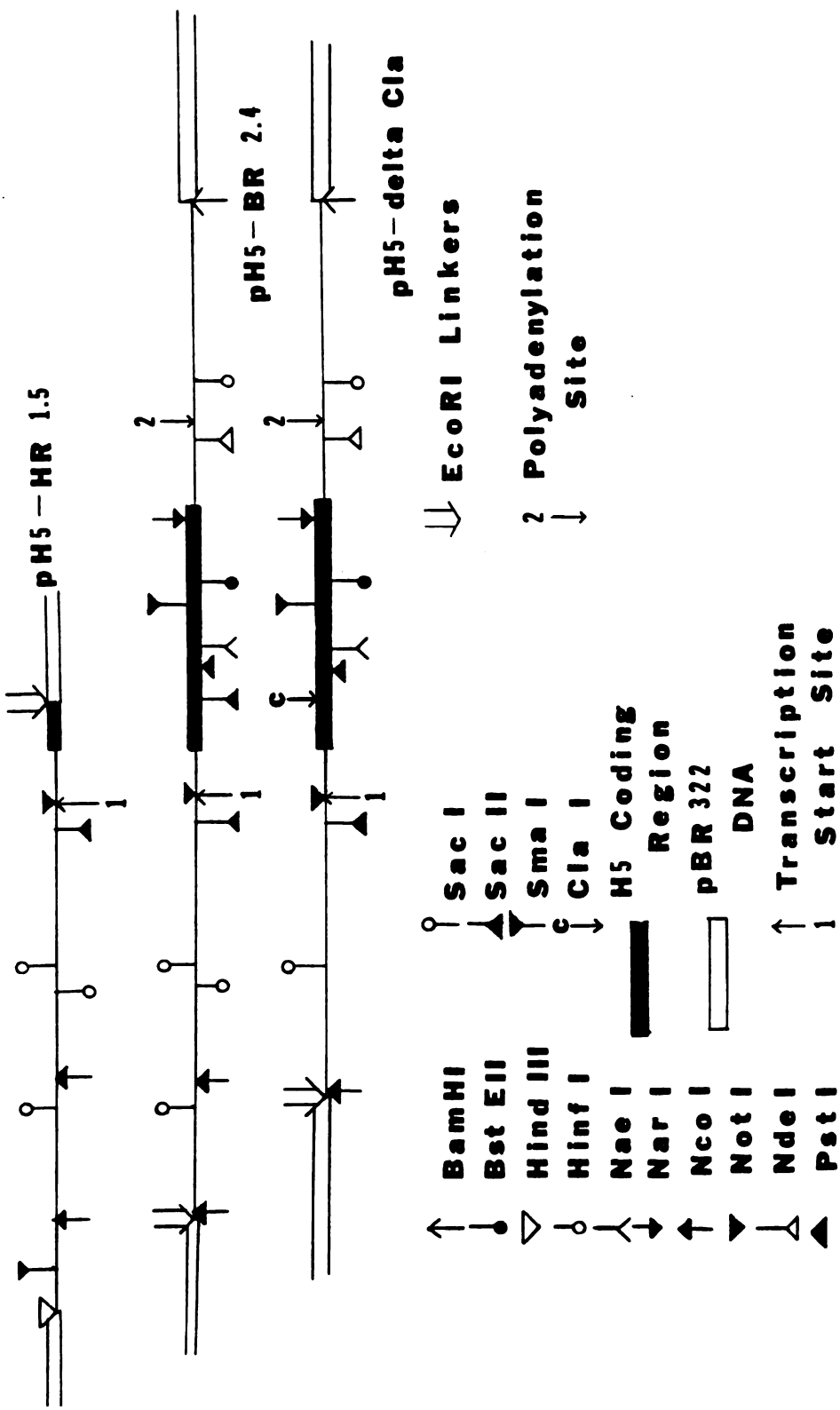
H5 Plasmid Subclones

A number of restriction enzyme fragments have been isolated from the phage clones and subcloned into the plasmid pBR322. The two subclones that have been analyzed in detail are pH5-HR 1.5 and pH5-BR 2.4 (Fig. 7). pH5-HR 1.5 contains the 1.5 kb HindIII-EcoRI fragment from cH5-1 which hybridizes to the p541 probe, while pH5-BR 2.4 contains the 2.4 kb BamHI-EcoRI fragment from cH5-2 which likewise hybridizes to the p541 probe. It should be noted that the EcoRI sites in both cases are not present in the normal H5 genomic sequence (compare the maps of cH5-3, and cH5-4 versus the maps of cH5-1 and cH5-4), but were introduced by use of synthetic EcoRI linkers during the construction of the chicken phage library (2).

Both subclones were mapped by restriction enzymes (Fig. 7) and compared to the known restriction enzyme map of the H5 gene (4,5). This allowed both the position of the gene

Figure 7 **Restriction enzyme maps of the H5 gene subclones described in the text. The transcriptional direction of the H5 gene is in all cases left to right. The location of the transcription start site and the polyadenylation site are shown.**

Figure 7


pH5-HR 1.5
pH5-BR 2.4
pH5-delta Cla
EcoRI Linkers
2 Polyadenylation Site
Sac I
Sac II
Sma I
Cla I
H5 Coding Region
pBR 322 DNA
Transcription Start Site
BamHI
Bst EII
Hind III
Hinf I
Nae I
Nar I
Nco I
Not I
Nde I
Pst I

within the clone as well as the orientation of the gene to be determined. As shown in Fig. 7, pH5-BR 2.4 contains the entire H5 gene as well as 1.2 kb of the 5′ flanking region and 0.7 kb of the 3′ flanking region. The pH5-HR 1.5 subclone contains a truncated H5 gene. Only about 100 bases of the H5 coding region is present within the insert, which also includes 1.4 kb of the 5′ flanking region.

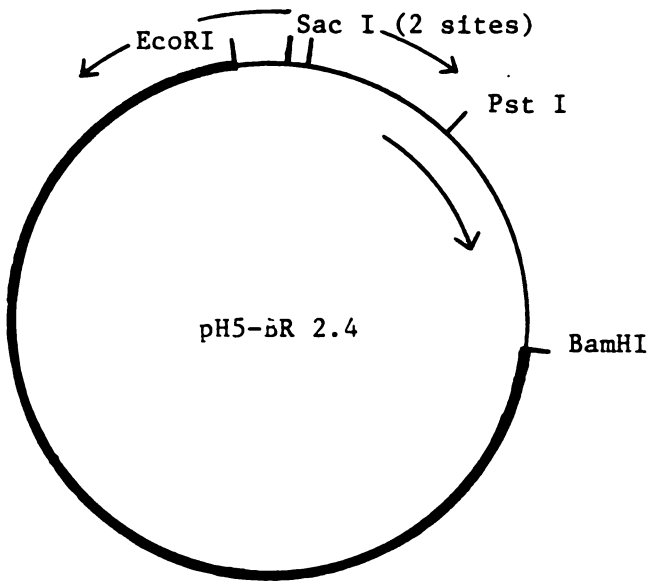
Construction of Linker Scanner Mutants

Linker scanner mutagenesis (6) uses synthetic oligonucleotide linkers to generate a cluster of point mutations within the region of DNA being studied. The region being studied for the H5 histone gene is the region extending from the CAP site to roughly -200, which appears to contain most of the putative promoter elements.

Before the linker scanner mutants can be constructed, two sets of deletions must be made. One set deletes progressively larger amounts of the region being analyzed in a 5′-3′ manner, while the second set progressively deletes over the same region in a 3′-5′ manner. The plasmid pH5-BR 2.4 was used to construct the first set of deletion mutants (Fig. 8). The plasmid was first digested with the restriction enzyme SacI. There are only two SacI sites,

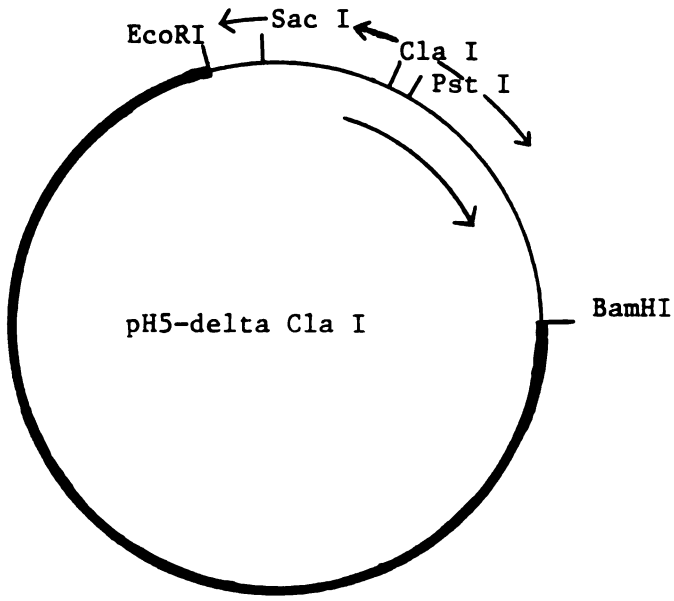
5′ of the promoter region (Fig. 7). After the plasmid was digested with SacI, the DNA was incubated with the double-stranded exonuclease, Bal-31. At various time points during the Bal-31 digestion, aliquots were taken and the Bal-31 reaction stopped. This procedure allowed a series of progressively larger deletions to be obtained. Synthetic HindIII linkers were ligated to the Bal-31 treated pH5-BR 2.4. HindIII linkers were chosen because no HindIII sites are present within the pH5-BR 2.4 plasmid. The linkers are ten nucleotides in length, with the sequence GCAAG CTTGC. After the ligation was completed, the linkers were digested with HindIII and the plasmid recircularized and ligated. The result of this series of reactions is a set of deletions which removes progressively larger amounts of the putative promoter region in a 5′ to 3′ direction. This set of deletions has been designated the "S" series.

A similar set of reactions was done on the 3′ side of the putative promoter region (Fig. 9), using the plasmid pH5-BR delta ClaI. pH5-BR delta ClaI (Fig. 7) is a derivative of pH5-BR 2.4, and differs from the parent plasmid in two respects. The first difference is that the 350 base pair SacI fragment has been deleted and only one SacI site remains, instead of two. The second difference is that a synthetic ClaI linker was inserted into one of the



1. Digest plasmid with Sac I
2. Incubate for various lengths of time with Bal 31
3. End fill with Klenow polymerase and ligate to synthetic Hind III linkers
4. Digest linkers with Hind III and recircularize plasmids by ligation in a dilute solution
5. Transform E. coli and screen clones by Hind III + Pst I digestion

Figure 8 Scheme for construction of the 5' to 3' set of deletions. The location and transcriptional direction of the H5 gene is shown by the arrow within the circle. The area deleted is shown by the arrows on the outside of the circle. The pBR322 vector sequences are shown as a thick line while the H5 insert is shown as a thin line.



1. Digest plasmid with Cla I
2. Incubate with Bal 31 for various lengths of time
3. End fill with Klenow polymerase and ligate to synthetic Hind III linkers
4. Digest linkers with Hind III and recircularize plasmids by ligation in a dilute solution
5. Transform E. coli and screen clones by Hind III + Sac I digestion

Figure 9 Scheme for construction of the 3' to 5' set of deletions. Position and transcriptional direction of the H5 gene is shown by the internal arrow while the area of deletion is shown by the external arrow. The thin line indicates the H5 insert while the thick line represents the pBR322 vector sequences.

SacII recognition sites. This ClaI linker is within the coding region of the H5 gene and is 3′ of the putative promoter region. The pH5-BR delta ClaI plasmid was linearized by digestion with ClaI, and afterwards treated as described above and in Figure 9. The deletion mutants that resulted from this series of reactions progressively delete the putative promoter region in a 3′ to 5′ manner. This set of deletions has been designated the "C" series.

After the plasmids were recircularized, the DNA was transformed into the E. coli strain, HB101. Plasmid DNA was isolated from the resulting colonies, and analyzed by digestion with restriction enzymes. pH5-BR 2.4 deletion mutants ("S" series) were digested with both HindIII and PstI, while the pH5-BR delta ClaI derived mutants ("C" series) were digested with HindIII and SacI. The purpose of this digestion is twofold: first, it determines whether a HindIII linker is present within the mutant, and second, if a HindIII linker is present, its location within the promoter region may be established within a roughly 50 base pair area. Over 1,000 bacterial colonies have been tested in this manner. Not all of the mutants tested had a Hind III linker present within their DNA. Efficiencies varied from 80% Hind III^+ to a low of only 40% Hind III^+. The reasons behind these different efficiencies is unclear.

The majority of the clones which have the HindIII linker, have it within the target area, i.e. between -200 and the CAP site.

Measuring the position of the linker by restriction enzyme digestion gives only a rough estimate of the linker's location. To make the linker scanner mutants, the exact location of the linker must be determined by DNA sequencing. A modified Maxam and Gilbert sequencing protocol (7) was used to sequence selected mutants. Unfortunately, no usable restriction enzyme sites were present near the region to be sequenced. Therefore, the mutants were digested with HindIII and opened directly at the linker. The end of the DNA was labeled with gamma ^{32}P ATP as described in the Materials and Methods section. Because the sequence of the H5 gene was known, only three out of the standard five reactions were done. These three reactions ("C+T", "A+G", and "G"), were sufficient to determine the location of the linker. Another modification to the standard protocol was in the length of time the reactions were allowed to procede. Only the terminal 20-30 base pairs of the sequence needed to be determined. For this reason, the chemical reactions were allowed to proceed twice as long as normally required. Due to the small size of the fragments, the reaction products were electrophoresed on a 20% denaturing

Figure 10 Maxam-Gilbert sequencing of the "S" series clone S'4-6. This particular clone was studied with five chemical reactions (C+T, C, A+G, G, and T>G). Most clones were sequenced with only three of the reactions (C+T, A+G,and G).

The bottom five nucleotides (below the arrow) are from the artificial Hind III linker. The first nucleotide above the arrow, represents the beginning of the chicken H5 histone DNA sequence. Roughly 25 to 30 base pairs of sequence may be determined from the gel. Since the sequence of the H5 flanking region is known, the exact location of the artificial HindIII linker may then be determined.

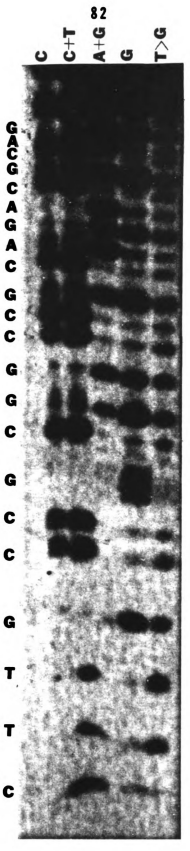


Figure 10

polyacrylamide gel. As shown in Figs. 11 through 14, the exact location of the HindIII linker has been determined for 106 mutants: 53 from the "S" series and 53 from the "C" series.

The construction of a linker scanner mutant requires an exact match between the location of the linker in the "S" series clone and in the "C" series clone. When these two clones are digested with HindIII and BamHI and are ligated together (Fig. 15), a mutant is created that maintains the original spacing of the H5 gene promoter, but has replaced 10 base pairs of the original promoter sequence with 10 base pairs of synthetic HindIII linker sequence. By using the same technique with "S" series and "C" series clones that are not an exact match, deletions or duplications of various sizes within the H5 histone promoter may be obtained. The various mutants that were constructed for studying the H5 histone promoter will be discussed in more detail in the next chapter.

Figure 11 Location of the Hind III linker for "S" series clones between -130 and -40. The sequence of the H5 flanking region is at the top. The thin line to the right of the Hind III linker sequence indicates unchanged H5 sequences. The designation for each mutant is on the far right.

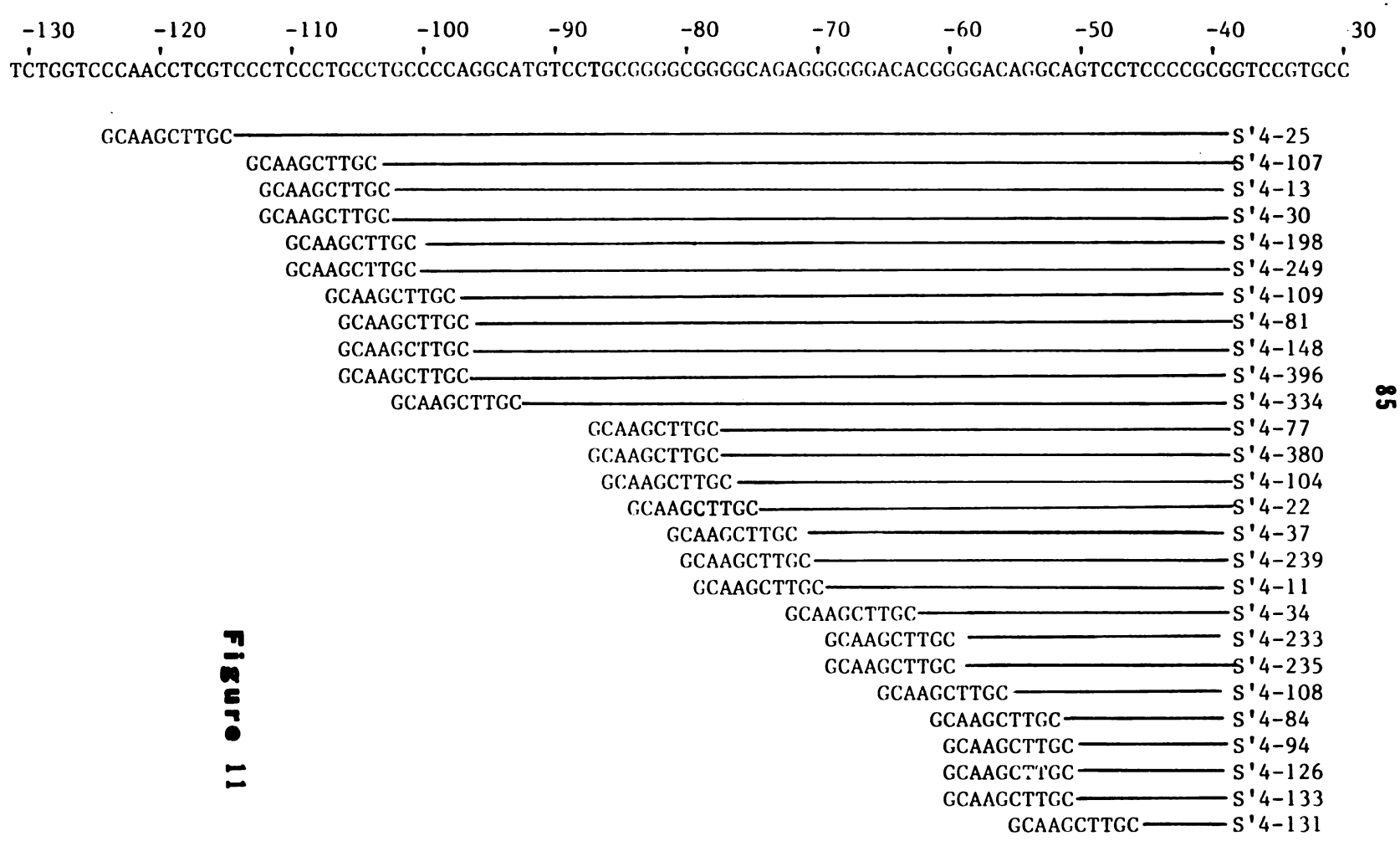

-130
-120
-110
-100
-90
-80
-70
-60
-50
-40
30
TCTGGTCCCAACCTCGTCCCTCCCTGCCTGCCCCAGGCATGTCCTGCGGGGCGGGGCAGAGGGGGGACACGGGGACAGGCAGTCCTCCCCGCGGTCCGTGCC
GCAAGCTTGC S'4-25
GCAAGCTTGC S'4-107
GCAAGCTTGC S'4-13
GCAAGCTTGC S'4-30
GCAAGCTTGC S'4-198
GCAAGCTTGC S'4-249
GCAAGCTTGC S'4-109
GCAAGCTTGC S'4-81
GCAAGCTTGC S'4-148
GCAAGCTTGC S'4-396
GCAAGCTTGC S'4-334
GCAAGCTTGC S'4-77
GCAAGCTTGC S'4-380
GCAAGCTTGC S'4-104
GCAAGCTTGC S'4-22
GCAAGCTTGC S'4-37
GCAAGCTTGC S'4-239
GCAAGCTTGC S'4-11
GCAAGCTTGC S'4-34
GCAAGCTTGC S'4-233
GCAAGCTTGC S'4-235
GCAAGCTTGC S'4-108
GCAAGCTTGC S'4-84
GCAAGCTTGC S'4-94
GCAAGCTTGC S'4-126
GCAAGCTTGC S'4-133
GCAAGCTTGC S'4-131

Figure 11

Figure 12 **Location of the Hind III linker for "S" series clones between -50 and +40.**

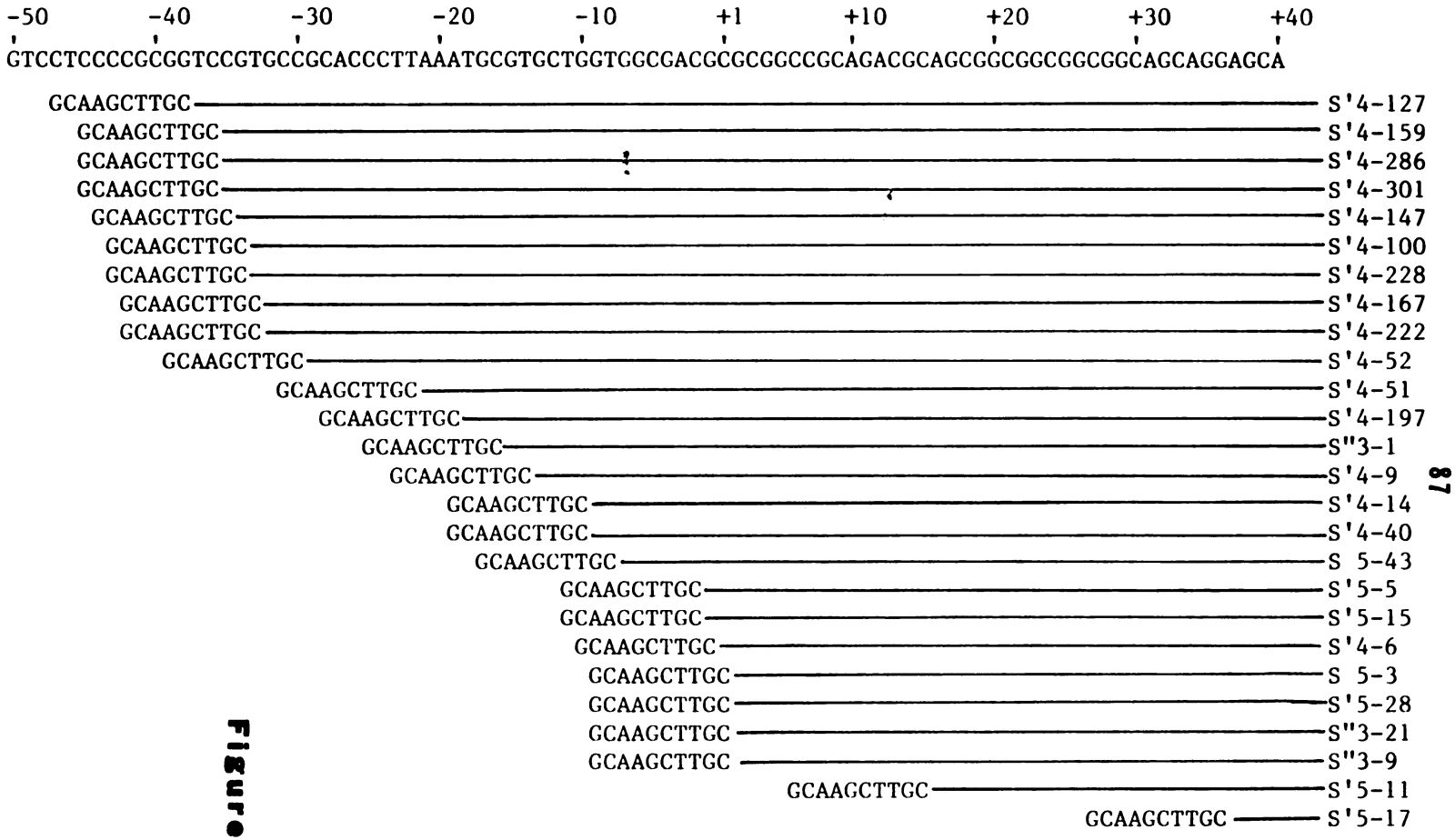

-50
-40
-30
-20
-10
+1
+10
+20
+30
+40
GTCCTCCCCGCGGTCCGTGCCGCACCCTTAAATGCGTGCTGGTGGCGACGCGCGGCCGCAGACGCAGCGGCGGCGGCGGCAGCAGGAGCA
GCAAGCTTGC S'4-127
GCAAGCTTGC S'4-159
GCAAGCTTGC S'4-286
GCAAGCTTGC S'4-301
GCAAGCTTGC S'4-147
GCAAGCTTGC S'4-100
GCAAGCTTGC S'4-228
GCAAGCTTGC S'4-167
GCAAGCTTGC S'4-222
GCAAGCTTGC S'4-52
GCAAGCTTGC S'4-51
GCAAGCTTGC S'4-197
GCAAGCTTGC S"3-1
GCAAGCTTGC S'4-9
GCAAGCTTGC S'4-14
GCAAGCTTGC S'4-40
GCAAGCTTGC S 5-43
GCAAGCTTGC S'5-5
GCAAGCTTGC S'5-15
GCAAGCTTGC S'4-6
GCAAGCTTGC S 5-3
GCAAGCTTGC S'5-28
GCAAGCTTGC S"3-21
GCAAGCTTGC S"3-9
GCAAGCTTGC S'5-11
GCAAGCTTGC S'5-17

Figure 12

Figure 13 **Location of the Hind III linker for "C" series clones between -150 and -50. The normal H5 DNA sequence is represented as a thin line to the right of the Hind III linker.**

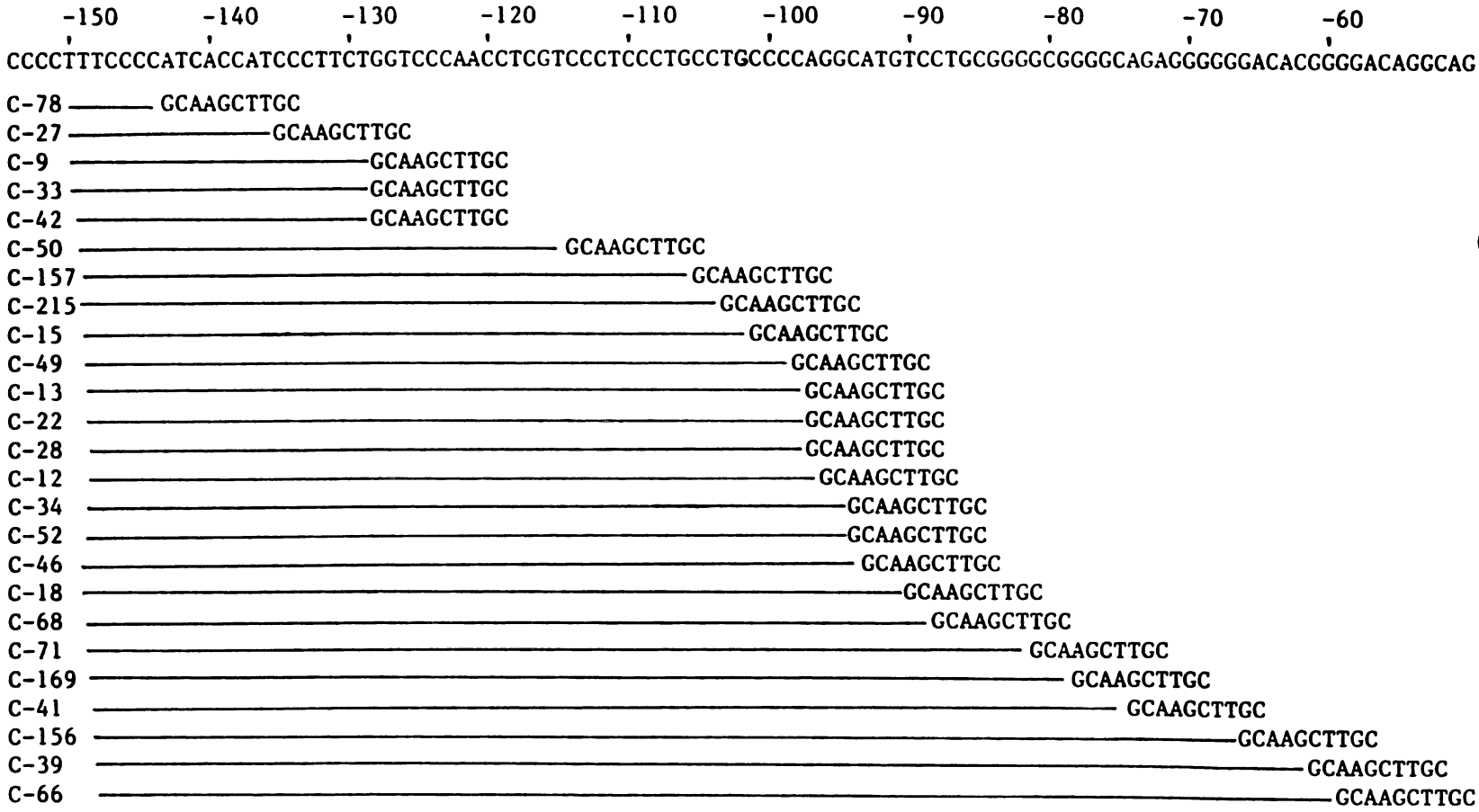

-150
-140
-130
-120
-110
-100
-90
-80
-70
-60
CCCCTTTCCCCATCACCATCCCTTCTGGTCCCAACCTCGTCCCTCCCTGCCTGCCCCAGGCATGTCCTGCGGGGCGGGGCAGAGGGGGGACACGGGGACAGGCAG
C-78 GCAAGCTTGC
C-27 GCAAGCTTGC
C-9 GCAAGCTTGC
C-33 GCAAGCTTGC
C-42 GCAAGCTTGC
C-50 GCAAGCTTGC
C-157 GCAAGCTTGC
C-215 GCAAGCTTGC
C-15 GCAAGCTTGC
C-49 GCAAGCTTGC
C-13 GCAAGCTTGC
C-22 GCAAGCTTGC
C-28 GCAAGCTTGC
C-12 GCAAGCTTGC
C-34 GCAAGCTTGC
C-52 GCAAGCTTGC
C-46 GCAAGCTTGC
C-18 GCAAGCTTGC
C-68 GCAAGCTTGC
C-71 GCAAGCTTGC
C-169 GCAAGCTTGC
C-41 GCAAGCTTGC
C-156 GCAAGCTTGC
C-39 GCAAGCTTGC
C-66 GCAAGCTTGC

Figure 13

Figure 14 **Location of the Hind III linker for "C" series clones between -60 and +30.**

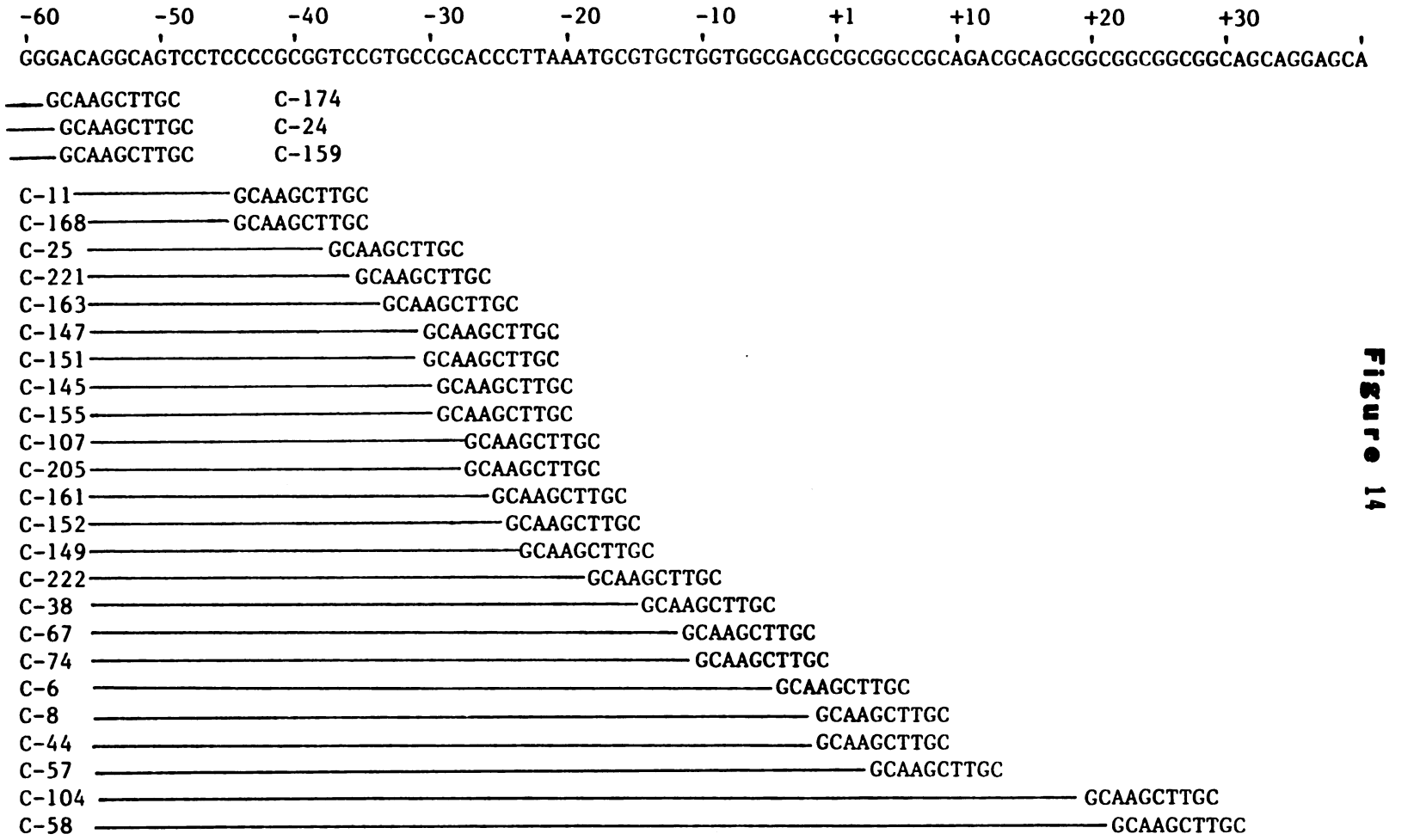
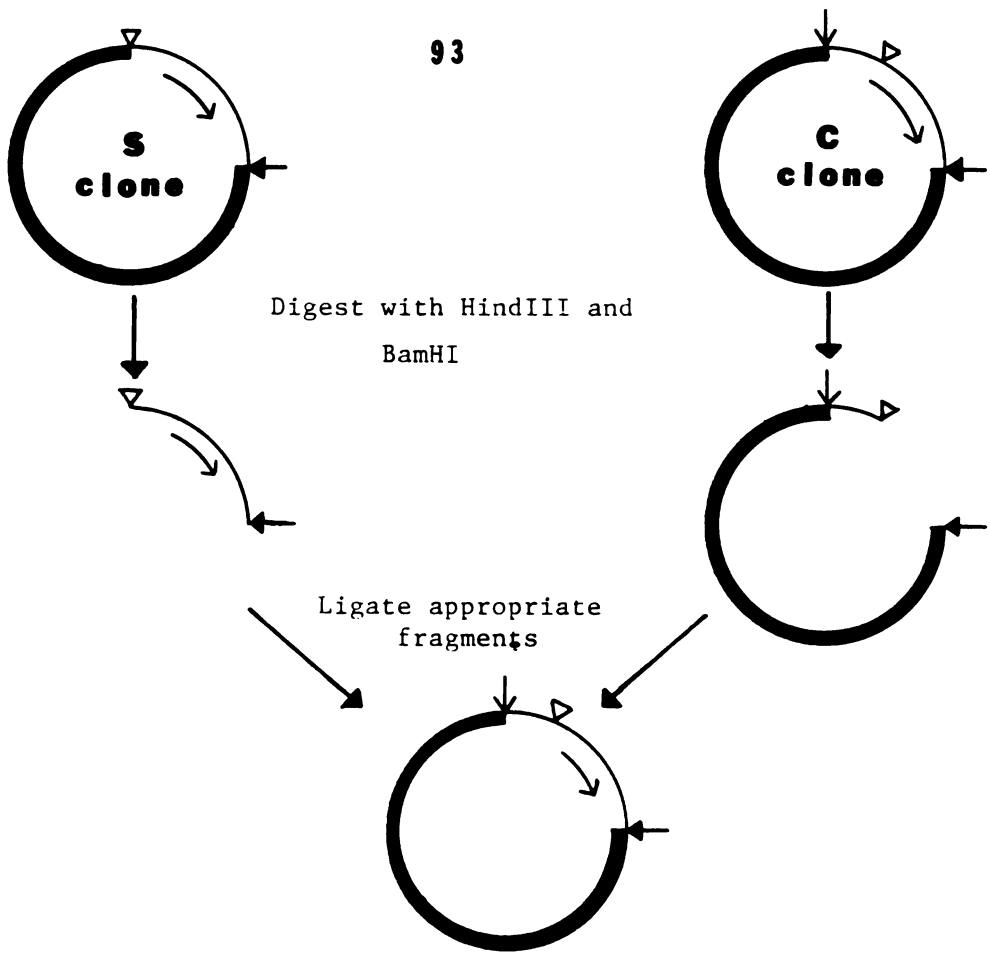

-60 -50 -40 -30 -20 -10 +1 +10 +20 +30
GGGACAGGCAGTCCTCCCCGCGGTCCGTGCCGCACCCTTAAATGCGTGCTGGTGGCGACGCGCGGCCGCAGACGCAGCGGCGGCGGCGGCAGCAGGAGCA
GCAAGCTTGC C-174
GCAAGCTTGC C-24
GCAAGCTTGC C-159
C-11 GCAAGCTTGC
C-168 GCAAGCTTGC
C-25 GCAAGCTTGC
C-221 GCAAGCTTGC
C-163 GCAAGCTTGC
C-147 GCAAGCTTGC
C-151 GCAAGCTTGC
C-145 GCAAGCTTGC
C-155 GCAAGCTTGC
C-107 GCAAGCTTGC
C-205 GCAAGCTTGC
C-161 GCAAGCTTGC
C-152 GCAAGCTTGC
C-149 GCAAGCTTGC
C-222 GCAAGCTTGC
C-38 GCAAGCTTGC
C-67 GCAAGCTTGC
C-74 GCAAGCTTGC
C-6 GCAAGCTTGC
C-8 GCAAGCTTGC
C-44 GCAAGCTTGC
C-57 GCAAGCTTGC
C-104 GCAAGCTTGC
C-58 GCAAGCTTGC

Figure 14

Figure 15 **Scheme for construction of linker scanner mutants. Both the "S" series and the "C" series clone are digested with HindIII and BamHI. The The smaller band from the "S" series clone is isolated. This fragment contains part of the H5 histone promoter as well as the H5 gene itself and the 3' flanking region. The larger fragment from the "C" series clone, which contains the remainder of the H5 promoter region and the prokaryotic vector sequences, is isolated and ligated to the "S" series fragment.**



pBR322

HindIII

BamHI

EcoRI

Location and transcriptional direction of the H5 gene

Fig. 15

References

1) Ruiz-Vazquez, R., Ruiz-Carrillo, A. NAR 10 pp. 2093-2108 (1982)

2) Dodgson, J., Strommer, J., Engel, J. Cell 17 pp. 879-887 (1979)

3) Scott, A., Wells, J.R.E. Nature 259 pp. 635-638 (1976)

4) Ruiz-Carrillo, A., Affloter, M., Renaud, J. J. Mol. Biol. 170 pp. 843-859 (1983)

5) Krieg, P., Robins, A., D'Andrea, R., Wells, J.R.E. NAR 11 pp. 619-627 (1983)

6) McKnight, S., Kingsbury, R. Science 217 pp. 316-324 (1982)

7) Maxam, A., Gilbert, W. Methods Enzymol. 65 pp. 499-560 (1980)

Chapter 4

Mutational Analysis of the H5 Histone Promoter

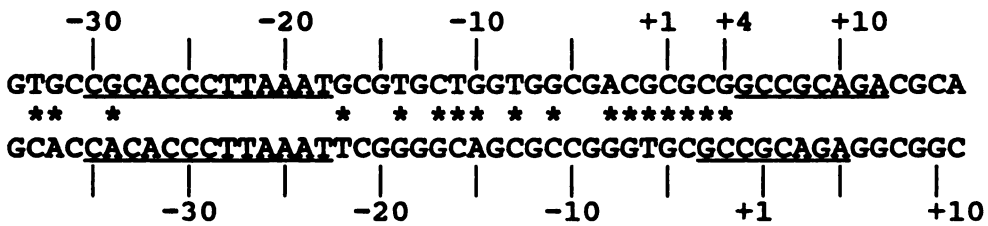
As described in the literature review, the putative chicken H5 histone promoter region lacks many of the chicken H1 histone promoter elements. The putative H5 promoter region lacks a "CCAAT" sequence and unlike all H1 histone promoter regions sequenced to date, the H5 promoter region does not include an "H1 box" (5'-AAACACA-3') (1). There are a number of sequences within the H5 promoter region, however, which may act as promoter elements for the H5 histone gene (Figure 16).

The first of these sequences is an element eight nucleotides in length located near the transcriptional initiation site. This element is also present at a similar location in the duck H5 promoter region. While the purpose of this element is unclear, it may serve as a recognition sequence for determining the location of transcription initiation (6). The chicken H5 histone gene has two transcriptional start sites, which are three nucleotides apart and located immediately 5' of the octanucleotide described above (7,8, Figure 17). The two transcriptional start sites appear to be used at an equal frequency.

Another probable promoter element is the noncanonical "TATA" box (5'-TTAAAT-3') located at -20. As shown in Figure 16, the duck H5 histone gene has the identical sequence in this area of the promoter. It is interesting to note

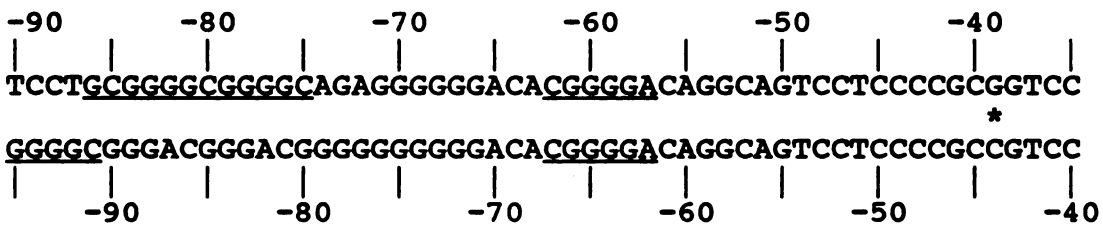
Figure 16 **Comparison of the chicken and duck H5 putative promoter regions. The two sequences have been aligned at the "TTAAAT" element. Asterisks indicate differences between the two sequences (only shown for the region between +10 and -60). The numbering system for the chicken H5 promoter starts from the 5' transcriptional start site. The second start site is at +4. The duck H5 promoter is numbered from the single transcriptional start site. Putative promoter elements (described in the text) are underlined.**

Chicken



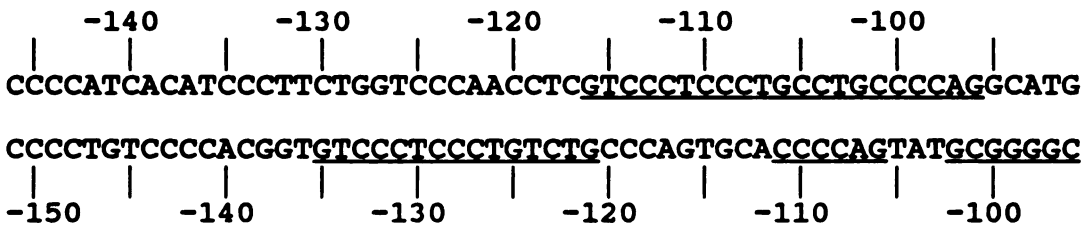
Duck

Chicken



Duck

Chicken



Duck

Figure 16

Figure 17 poly A^+ RNA from Rat 3A fibroblasts (lane 1), QT6 fibroblasts (lane 2), and HD3 erythroid cells (lane 3) were hybridized to a 27 nucleotide H5 primer. The primer covers the first 9 codons of the H5 coding region. The primer was then extended using reverse transcriptase. The two bands in the HD3 erythroid RNA lane indicates that there are two transcriptional start sites, 3 base pairs apart. Predicted band sizes are 141 bp and 144 bp.

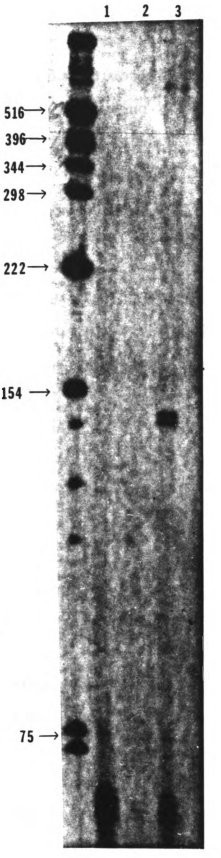


Fig. 17

that the duck and chicken H5 promoter regions do not share extensive areas of homology in the region between the "TTAAAT" box and the octanucleotide around the transcriptional initiation site. Starting at the "TTAAAT" sequence, however, a large stretch of sequence homology extending to -70 is found. Apparently, the noncanonical "TTAAAT" box is under selective pressure.

Immediately upstream of the "TTAAAT" element is a hexanucleotide, which is designated the proximal H1 homology element (3). This element (5′-CGCACC-3′) is located at -30 in the chicken H5 gene and a related sequence (5′-CAC ACC-3′) is located at an identical site in the duck H5 promoter. The 5′-CGCACC-3′ sequence is also present in two chicken H1 genes (2,9) but in the two chicken H1 genes, this element is located at roughly -60 and is separated from the "TATA" element by approximately 10 nucleotides instead of being adjacent as in the H5 genes. The significance of this difference, if any, is unknown.

The proximal H1 homology hexanucleotide sequence almost completely overlaps a second sequence (5′-CACCC-3′). The 5′-CACCC-3′ pentanucleotide has been shown to be required for high levels of rabbit and chicken β-globin gene expression after transfection of these genes into fibroblasts (4, 5). While the location of this sequence varies slightly species to species (4), the pentanucleotide is usually found in the -100 region. In the chicken adult β-globin

gene, this pentanucleotide is present at roughly position -130. In the chicken H5 histone promoter, however, this pentanucleotide is adjacent to the "TATA" element at position -30. Unfortunately, the mutagenesis assay being described here will only determine if this area of the promoter is important for the expression of the H5 histone gene. It will not determine which element (if either) is responsible for any changes detected when this area is mutated.

Further 5′ of these elements is another H1 homology hexanucleotide (5′-CGGGGA-3′) which is designated the distal H1 homology element. This element is located at -60 in both the chicken and duck H5 histone promoter. One chicken H1 histone gene (2) has this element at position -100 while another chicken H1 gene (9) has a related hexanucleotide (5′-CGGGGC-3′) at roughly position -130. The function of this element, if any, is unknown.

After the distal H1 homology element, the spacing of putative promoter elements is different between the duck and the chicken H5 genes. An Sp1 protein binding sequence, 5′-GGGGCGGGG-3′, is located at position -80 in the chicken H5 promoter region. The duck H5 promoter region, however, has this sequence displaced ten nucleotides upstream relative to the chicken H5 gene.

Upstream of the Sp1 binding element in the chicken H5 promoter, between -95 and -115, is a sequence designated

the duck H5 homology element. In the duck H5 promoter, this element consists of two isolated sequences, 14 and 6 nucleotides in length respectively. In the chicken H5 promoter, however, these two sequences have been joined into a 20 base pair element. The function of this element, if any, is unknown.

In summary, six possible promoter elements have been identified to date. They include the octanucleotide near the H5 transcriptional start site, the noncanonical "TATA" element at -20, the proximal H1 homology sequence/"CACCC" element at -30 to -25, the distal H1 homology sequence at -60, the Sp1 binding element at -80, and the duck H5 homology sequences at -115 to -95. Other regulatory sequences may be present in the region of homology (-30 to -70) that exists between the duck and chicken H5 histone promoter regions. There may also exist regulatory elements further upstream (> -150) or downstream (> +20) that have not been detected by homology searches.

The chicken H5 promoter was initially analyzed by making gross deletions in the region between -145 and +20. These deletions were made in the same manner as the linker scanner mutants described in the previous chapter (Figure 15). In this case, however, a single "C" series clone was matched to a series of different "S" series clones. These mutants and the region deleted in each mutant are shown in

Table 1

	duck homol.	Sp1	distal H1	"CACCC"/ prox. H1	"TATA" box	CAP site
-145/-75	-	-	+	+	+	+
-145/-42	-	-	-	+	+	+
-145/-28	-	-	-	-	+/-*	+
-145/+20	-	-	-	-	-	-

*: The -145/-28 mutant deletes the first two thymidines in the "TTAAAT" sequence. It is uncertain if this deletion would effect the function of the sequence.

Figure 18 and Table 1. This procedure yields a series of mutants which have a common boundary on one side (at position -145) but which progressively delete further 3′ towards the H5 histone gene. Also shown in figure 18 is an internal control (+5/+20) which will be described in more detail below.

A number of linker scanner mutants have also been constructed within this region (-145 to +20). A linker scanner mutant requires an exact match between the location

Figure 18 **Size and location of the deletion mutants described in the text. Thin line represents chicken H5 sequence. Putative promoter elements are indicated at the bottem of the page.**

1- Duck H5 homology elements

2- Spl recognition site

3- distal Hl homology sequence

4- proximal Hl homology sequence

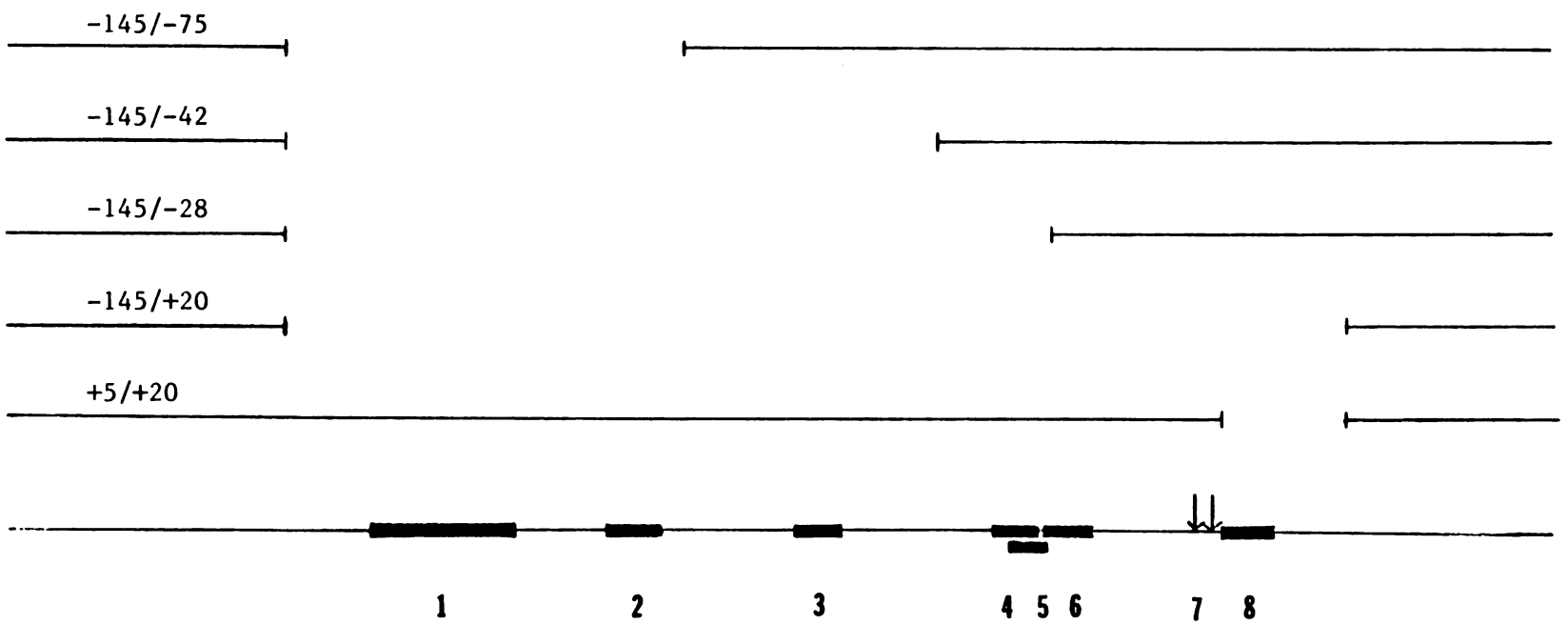
5- globin "CACCC" element

6- "TTAAAT" box

7- transcriptional start sites

8- CAP site octanucleotide

Fig. 18



of the linker in the "S" series clone and in the "C" series clone. When these two clones are digested with HindIII and BamHI and are ligated together (Fig. 15), a mutant is created that maintains the original spacing of the H5 gene promoter, but has replaced 10 base pairs of the original promoter sequence with 10 base pairs of synthetic HindIII linker sequence. Seven such linker scanner mutants have been constructed (Fig. 19). Also shown in this figure are 9 mutants which are the result of ligating clones slightly mismatched as to the location of the HindIII linker between the "S" series clone and the "C" series clone. These mutants change the spacing within the H5 promoter by 1 base pair compared to the normal promoter. Six of these mutants (labeled -1) replace 11 base pairs of the H5 promoter region with only 10 base pairs of HindIII linker, with the result that the spacing between two areas on opposite sides of the linker location has decreased by a single base pair. Likewise, the 3 mutants labeled +1 have increased the promoter length by one base pair.

A final modification to both the gross deletion mutants and the linker scanner mutants is the presence or absence of an enhancer, which is provided by a permuted RSV LTR inserted into the EcoRI recognition site, 1 kb upstream of the transcriptional initiation site (Figure 5). This LTR was obtained from Dr. Maribeth Raines and was constructed

Figure 19 Location of the linker scanner mutants and near linker scanner mutants. Putative promoter elements are shown at the bottom.

1- duck H5 homologies

2- Sp1 binding site

3- distal H1 homology site

4- proximal H1 homology site

5- "CACCC" box

6- "TTAAAT" element

7- transcriptional start sites

8- CAP site octanucleotide

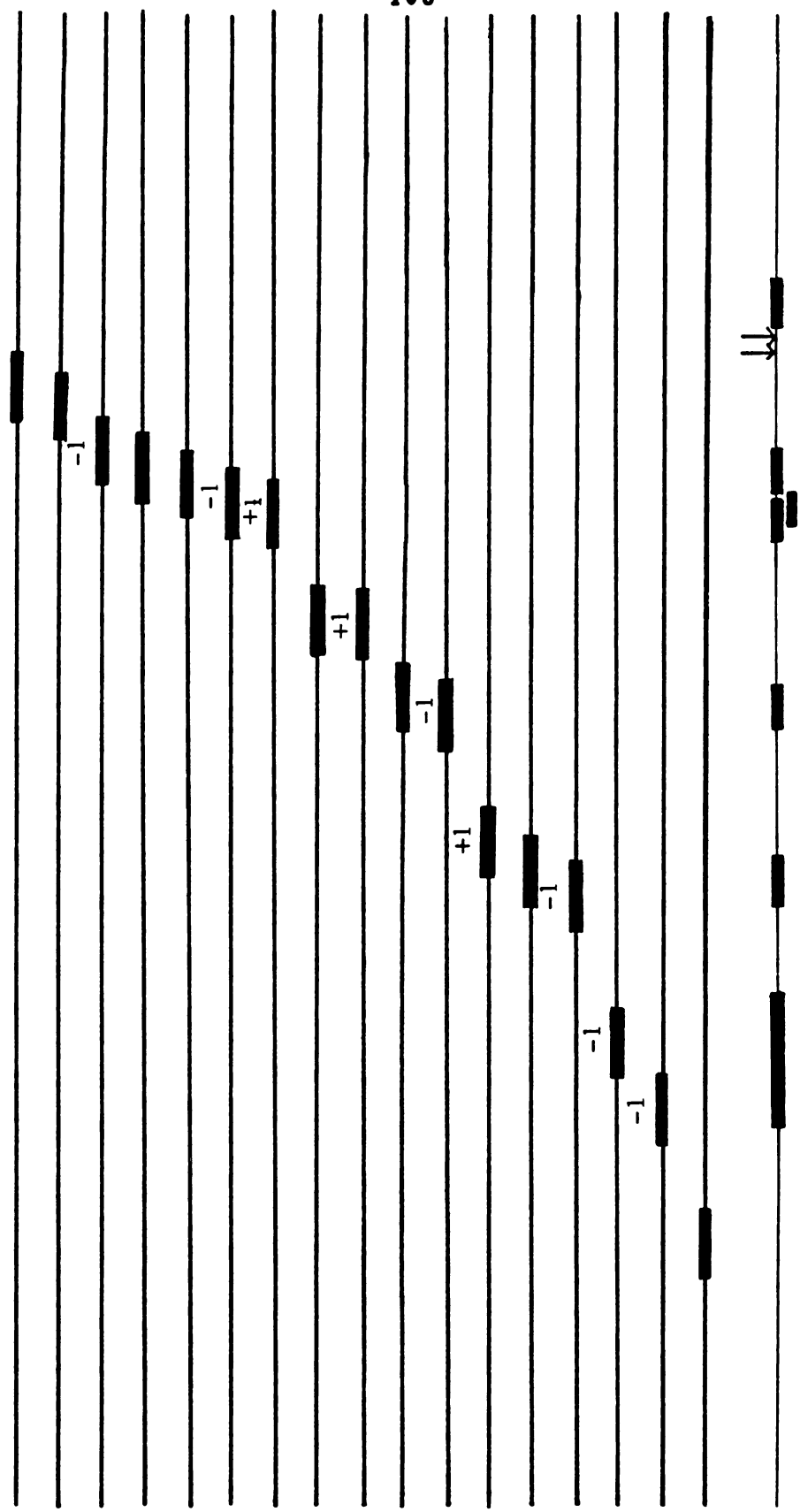


Fig. 19

by digesting two linked RSV LTRs with EcoRI. The EcoRI fragment was then inserted into EcoRI digested pBR322. The structure of the resulting LTR has most of the U3 region removed from the 5′ end of the LTR, and placed at the 3′ end, next to the U5 region. The RSV LTR still retains its enhancer activity, but has decreased promoter activity. The addition of the enhancer activity is meant to boost the level of H5 gene expression to an easily detected level. The RSV LTR has been ligated into the EcoRI site of most mutants constructed to date, with the LTR promoter sequences oriented away from the H5 gene. The LTR is present on the 5′ side of the gene and is anywhere from 500 to 800 base pairs away from the CAP site of the gene, depending on the particular mutant being studied.

The mutants were studied by transfection into QT6 quail fibroblasts. Rather than a stable transformation, where the transfected DNA is integrated into the host cell's chromatin, a transient assay was used. RNA was isolated from the cells 36-48 hours after the transfection. The RNA was afterwards subjected to S1 analysis. The probe in this analysis is a 780 base pair EcoO109-SacI fragment. As shown in Figure 20, a protected fragment 380 base pairs in size is predicted to occur after digestion with nuclease S1. There are a number of areas in this protocol where differences between samples may occur. These variables,

Figure 20 S1 analysis. An EcoO109-SacI fragment was used as the probe. When hybridized to normal or most mutant H5 histone mRNA, a protected fragment 380 bp in size is expected. When this probe is hybridized to the internal control mutant (+5/+20) or to mutant -145/+20, the presence of the HindIII linker 3′ of the CAP sites causes a smaller protected fragment size of 360 bp. The question mark near the 5′ end of the -145/+20 or +5/+20 mRNA indicates that this probe will not indicate whether these mutants are correctly initiated.

HindIII

EcoRI

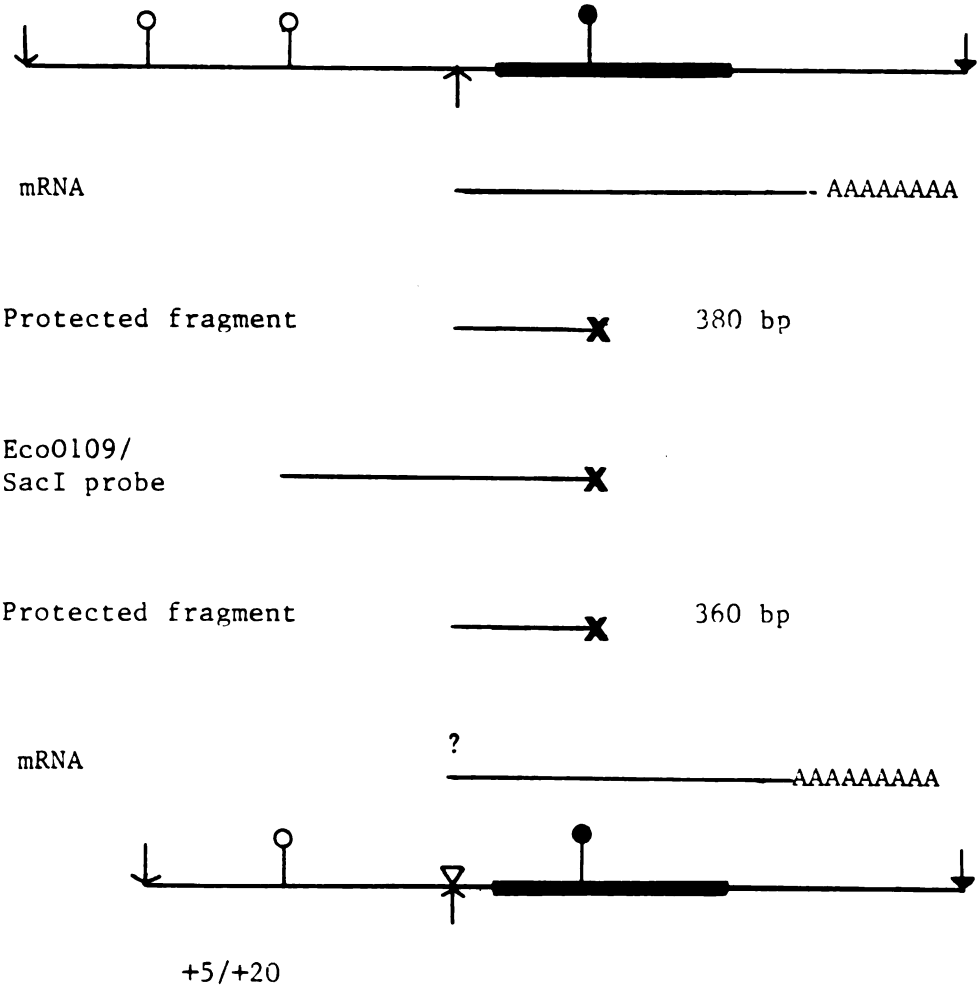
BamHI

SacI

Eco0109

Figure 20

Transcriptional start sites



such as transfection efficiencies, RNA quality, and extent of S1 digestion, could make it difficult to compare the results from different mutants to each other and to the unmutated H5 gene. For this reason, the internal control H5 gene, +5/+20, was constructed. This mutant has had 20 base pairs of DNA immediately downstream of the 5' most CAP site removed and replaced with a 10 base pair HindIII linker. Transcripts from this mutant would also hybridize to the Eco0109-SacI probe, but due to the deletion and linker insertion, the protected fragment after S1 nuclease digestion would be 20 base pairs shorter (360 base pairs) than the normal fragment (Figure 20). The transcriptional activity of this mutant acts as a reference point. The analysis of these mutants, the internal control, and the normal gene, measures the steady state RNA levels. Since the promoter region is the area being mutated and the transcripts themselves have not been altered in structure (except for the mutant delta NarI which will be covered in more detail later), it is reasonably safe to assume that the RNA steady state levels reflect the rate of transcription.

Thirty micrograms of the various mutants described above, were transfected into the QT6 cells along with 30 micrograms of the +5/+20 internal control. All of the clones had the LTR enhancer present in the EcoRI restriction

enzyme recognition site. After the S1 analysis and gel electrophoresis, the samples were exposed to X-Ray film. The results of one film are shown in Figure 21. This film includes a mutant, delta NarI, not yet discussed in detail. This mutant begins at +870, which removes the last 10 codons of the coding region and all of the H5 3' flanking region.

After development of the X-Ray film, the individual lanes were subjected to densitometer scanning. One such tracing (covering only the area of interest) is shown in Figure 22. The area under each curve was determined by cutting out the curve and weighing it on a Mettler balance. The results from an S1 gel are shown in Table 2. The transcriptional level of the normal H5 gene was set at 100%. The area of the internal control peak was then compared to the normal gene's area. It was noted that the internal control steady state H5 transcript levels were roughly 80% the steady state transcript levels for the normal H5 gene. The decrease in transcription level for the internal control gene may be due to the deletion of one of the CAP sites and the removal of the octanucleotide sequence discussed above. For all following samples, the +5/+20 internal control gene was assumed to have a transcription steady state level approximately 80% that of the normal gene. The mutants were compared to the internal control

Figure 21 S1 analysis of various mutants and the internal control gene +5/+20. "a" indicates the size of the normal mRNA protected fragment (380bp) while "b" indicates the size of the internal control gene (+5/+20). The bands which are above and below these two predicted bands appear to represent non-specific initiation sites.

1- +5/+20 + normal H5 gene (+LTR)

2- +5/+20 + -145/+75 (+LTR)

3- +5/+20 + -145/-42 (+LTR)

4- +5/+20 + -145/-28 (+LTR)

5- normal H5 gene + -145/+20 (+LTR)

6- +5/+20 + delta NarI (+LTR)

7- +5/+20 + normal H5 gene (-LTR)

8- +5/+20 (+LTR)

9- QT6

10- HD3

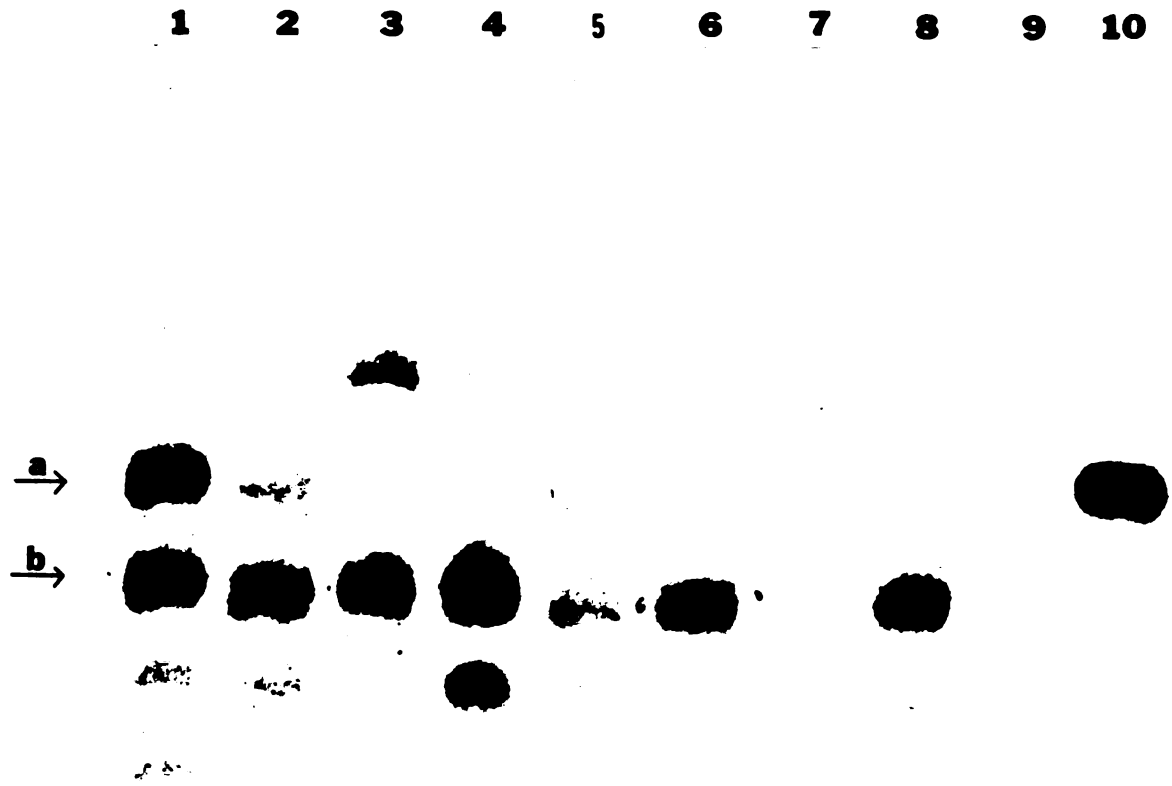


Fig. 21

Table 2

with enhancer

normal H5	0.25 grams ± 0.01 grams	(100%)
+5/+20	0.20 grams ± 0.01 grams	(~80%)
-145/-75	0.07 grams ± 0.01 grams	(~40%)
+5/+20	0.18 grams ± 0.02 grams	(~80%)
-145/-42	0.07 grams ± 0.00 grams	(~20%)
+5/+20	0.40 grams ± 0.03 grams	(~80%)
-145/-28	0.05 grams ± 0.00 grams	(~10%)
+5/+20	0.39 grams ± 0.02 grams	(~80%)
normal H5	0.06 grams ± 0.00 grams	(100%)
-145/+20	0.09 grams ± 0.01 grams	(~150%)
delta NarI	0.02 grams ± 0.00 grams	(~10%)
+5/+20	0.28 grams ± 0.02 grams	(~80%)

no enhancer

normal H5	0.06 grams ± 0.00 grams	(100%)
+5/+20	0.05 grams ± 0.00 grams	(~80%)

The S1 gel in Figure 21 was subjected to densitometer scanning; each lane was scanned three separate times. The peaks from each band was cut out and weighed on a Mettler balance. The three measurements were averaged together to yield the final result. The "±" number indictes the variation between seperate measurements. Variation below the threshold of significant figures are indicated as ± 0.00 %. The transcriptional level of the normal H5 gene is considered to be 100%. All other percentages are comparisons to the normal H5 gene. The internal control, +5/+20, is considered to be expressed at roughly 80% for all samples.

Figure 22 Sample densitometer tracing of one S1 lane. The lane shown here is for lane 1, normal H5 + +5/+20. "a" is the protected fragment from the normal H5 gene while "b" represents the protected fragment from the internal control gene, +5/+20. The small peaks below bands "a" and "b" may represent S1 degradation products, or more likely, non-specific initiation sites

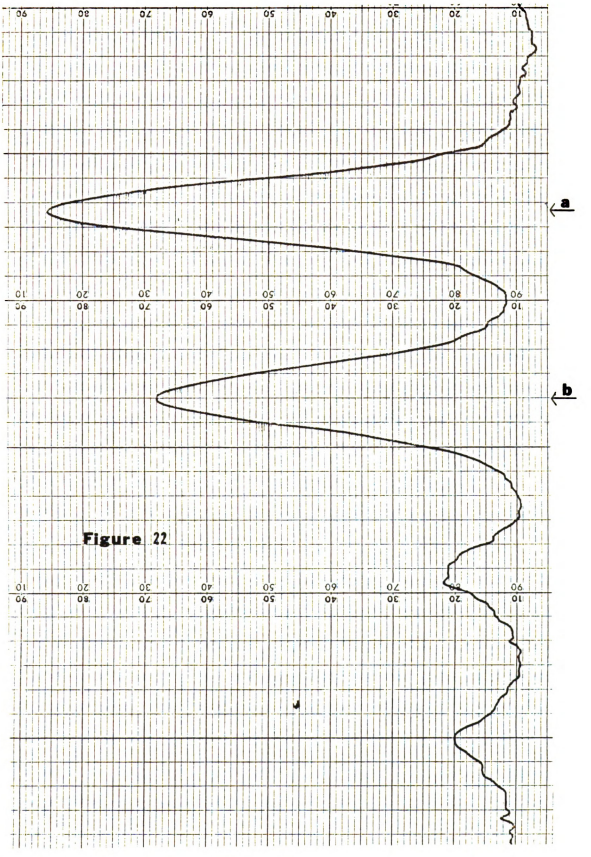


Figure 22

to obtain a level of transcription relative to the normal H5 gene. For example, the internal control in one sample was found to have an area of 0.18, while the mutant -145/-75 had an area of 0.07. Since +5/+20 was considered to have only 80% of the activity of the normal H5 gene, the normal H5 gene would have had an area of roughly 0.22. Compared to this number then, the mutant -145/-75 has a transcriptional activity roughly 40% that of the normal H5 gene.

As predicted, as more of the putative H5 promoter is deleted, the transcriptional activity of the H5 gene decreases. Removal of the duck H5 homologies and the Sp1 binding site (mutant -145/-75) decreases the activity of the gene to a level roughly 40% that of the normal H5 gene. It is still uncertain if the duck H5 homologies have any function. All of this decrease could possibly be due to the loss of the Sp1 binding site.

Further deletion past the distal H1 homology site (-145/-42) causes a further decrease to ~20% the transcriptional activity of the normal H5 gene. It should be noted that this decrease may not be due only to the loss of the distal H1 homology. While this sequence is the most likely reason for the decrease, nearly 100 base pairs of sequence have also been deleted. There may be other elements involved that have not been identified to date.

The mutant -145/-28, which deletes past the proximal H1 homology/"CACCC" site, shows a further decrease in transcription to ~10%. Interestingly, the next mutant in the series (-145/+20) which deletes past the CAP sites, shows a substantial increase in transcription activity relative to the normal H5 gene. The reasons for this increase are not clear. A probable explanation is that this band includes transcripts initiated from cryptic start sites in the 5′ flanking region. The mutant is similar to the internal control (see Figure 20) in that the Eco0109-SacI S1 probe is not homologous to the -145/+20 transcripts when the HindIII linker at +20 is reached. Therefore, the S1 probe in this case will indicate all transcripts that initiated before the HindIII discontinuity. The probe will not show whether the transcripts were initiated at the correct location.

The mutant delta NarI, which removes the 3′ flanking region, has a transcriptional level roughly 10% that of the normal H5 gene. While this mutant deletes the H5 enhancer sequence, this is unlikely to be the reason for the decrease since the enhancer has been shown to be erythroid-specific. Deletion of this sequence should not effect the level of transcription in non-erythroid quail fibroblasts. It is possible that other regulatory elements exist in the 3′ flanking region of the H5 histone gene.

Another possible explanation is that H5 transcript processing has been affected. As described in the introduction, the H5 transcript is polyadenylated but the 3' flanking region does not contain the usual "ATAAAA" polyadenylation signal. The flanking region does however, have a number of sequences that could form stem-loop structures which could aid in the terminal processing of the H5 transcript (6). Deletion of these sequences could greatly decrease the level of correctly processed H5 transcripts. These incorrectly processed messages may then be rapidly degraded.

The last sample which is shown on Table 2, is the normal H5 gene and the internal control H5 gene without the LTR enhancer sequence attached. As expected, the transcriptional level of the two genes without the attached LTR was substantially lower than the same genes with the LTR. The amount of the difference is uncertain. It is of interest, however, that the internal control gene (+5/+20) still shows roughly 80% transcriptional activity compared to the normal H5 gene. This is the same as the ratio seen for the two genes with the LTR attached. At least in this case, the LTR is only increasing the levels of transcription. It is not providing substitute sequences for elements missing in the +5/+20 mutant.

Discussion

The putative promoter region of the chicken H5 histone gene has been partially analyzed by *in vitro* mutagenesis. The mutants were studied by transfection into QT6 quail fibroblasts in a transient assay. In a transient assay, RNA is isolated from the cells 36-48 hours after the plasmid(s) was transfected into the host cell. One advantage of transient assays is that the transfected plasmid DNA is not integrated into the host cell's chromatin (10,11). In a stable transfection, the level of expression of the transfected gene often depends upon where the gene integrated into the chromatin. The sequences and chromatin configuration flanking the newly integrated gene will create differences in the gene's transcription levels. In a transient assay, however, RNA is isolated and analyzed before the plasmid has been able to integrated. Since the gene's flanking sequences is equivalent plasmid to plasmid, this variation is eliminated.

The putative chicken H5 gene promoter was analyzed by using a series of deletions which have a common boundary at -145, but delete various amounts of the H5 promoter region. All mutants tested to date include an LTR enhancer. Deletion of the six possible promoter sequences discussed in the beginning of this chapter leads to decreased levels of transcription. Transcription levels dropped sharply (to

roughly 40% that of the normal H5 gene) when the duck H5 homologies and the Sp1 recognition site were deleted. The Sp1 recognition site is possibly responsible for much of this decrease. It is still not certain, however, if the duck H5 homologies have a role in regulating the transcriptional activity of the H5 gene. Other mutants have been constructed which delete only the duck H5 homology sequences and will be tested to determine if the loss of the duck H5 homology elements decreases the level of H5 transcription.

The transcriptional activity of the H5 gene continued to decrease as first the distal H1 homology sequence (~20% the normal gene's transcription level), then the proximal H1 homology sequence/"CACCC" box (~10% the transcription level of the normal H5 gene) were removed. Since other DNA sequences besides the target elements were removed in these deletions, it is not possible to state that the reduction in the transcriptional activity of the H5 gene was due only to the loss of the distal or proximal H1 homology sequences. It is apparent, however, that elements exist within the deleted areas which contribute to the expression of the H5 gene.

The -145/+20 mutant, which deletes past the transcriptional start sites, actually showed a higher level of transcription than the normal H5 gene. As described above, this

is probably due to cryptic start sites present in the 5′ flanking region. The Eco0109 probe used for the S1 analysis is isolated from the normal H5 gene and would not identify the transcriptional initiation sites for either the +5/+20 internal control clone or the -145/+20 mutant (Fig. 20). All transcripts, whether correctly initiated or not, would be present within the one protected fragment. It would be of interest to test the -145/+20 mutant with a Eco0109 probe taken from the -145/+20 clone itself. This would identify the transcriptional start sites for this clone and may allow the "correct" -145/+20 transcript to be identified from the transcripts started from the cryptic sites.

A large deletion mutant (delta NarI) which removes the 3′ flanking region was also tested. The level of transcription for this mutant is decreased to below 10%, compared to the normal gene. Unlike the mutants described above, delta NarI has not changed the promoter region in any fashion. Rather than affecting the rate at which the gene is transcribed, this mutant appears to alter the way the mRNA is processed after transcription. While the H5 transcript is polyadenylated, the H5 gene does not use the standard sequence, "AAUAAA", for this modification. The H5 message may use a series of stem-loop motifs as the means of processing its 3′ terminus (6). Removal of these

stem-loop elements would then seriously effect H5 transcript processing. Smaller deletions in this region may clarify what sequences are responsible for the decrease in the levels of histone H5 messages.

Two potential problem areas which have not yet been discussed are the effects of the LTR enhancer on the mutants and the possibility of no effect when a potential promoter element is deleted. The last problem has not yet been encountered, since all deletions tested so far have had some effect on the level of H5 histone transcripts. As the promoter region is analyzed in more detail, however, it is possible that the deletion of a suspected promoter element will not have an detectable effect. A negative result would be difficult to interpret. There would be uncertainty as to whether this sequence is not a promoter element or whether the element is a promoter element, but it is not used correctly in a non-erythroid, quail fibroblast cell line. An example of this ambiguity that may occur for the H5 promoter, is the "H1 box" which was studied in one of the chicken H1 genes. Since this sequence has been found in all H1 histone promoters analyzed to date, it is likely that the "H1 box" is a promoter element. Deletion of this sequence in a chicken H1 histone promoter, however, showed no effects when the H1 histone gene was tested in the heterologous Xenopus oocyte or HeLa cell systems (12). Such

ambiguity may arise in deletions of the H5 promoter elements.

The second possible problem area is in the use of the LTR enhancer. While the LTR may increase H5 transcription to a more easily detectable level, it may also interfere with the promoter element analysis. Comparison of the normal H5 gene to the internal control gene yields roughly the same ratio (~80%) with or without the LTR enhancer (Table 2). It cannot be stated with certainty, however, that the LTR will not effect the other mutants. The loss of, for example, the Sp1 recognition element may be counteracted in some way by the LTR enhancer so that the apparent decrease in the levels of the H5 histone transcript is not as great as in the absence of the LTR. The dilemma is that without the LTR enhancer, the levels of the H5 transcript is extremely low. Until the series of mutants have been tested without the LTR enhancer, it will be assumed that the LTR only increases the level of transcription.

References

1) Coles, L., Wells, J.R.E. NAR 13 pp. 585-594 (1985)

2) Sugerman, B., Dodgson, J., Engel, J. J. Biol. Chem. 258 pp. 9005-9016 (1983)

3) Ruiz-Carrillo, A., Affloter, M., Renaud, J. J. Mol. Biol. 170 pp. 843-859 (1983)

4) Dierks, P., VanOoyen, A., Cochran, M., Dobkin, C., Reiser, J., Weissman, C. Cell 32 pp. 695-706 (1983)

5) Myers, R., Tilly, K., Maniatis, T. Science 232 pp. 613-618 (1986)

6) Doenecke, D., Tönjes, R. J. Mol. Biol. 178 pp. 121-135 (1984)

7) Krieg, P., Robins, A., D'Andrea, R., Wells, J.R.E. NAR 11 pp. 619-627 (1983)

8) Ruiz-Carrillo, A., Affloter, M., Renaud, J. J. Mol. Biol. 170 pp. 843-859 (1983)

9) Coles, L., Wells, J.R.E. NAR 13 pp. 585-594 (1985)

10) Mellon, P., Parker, V., Gluzman, Y., Maniatis, T. Cell 27 pp. 279-288 (1981)

11) Treisman, R., Green, M., Maniatis, T. PNAS 80 pp. 7428-7432 (1983)

12) Younghusband, H., Sturm, R., Wells, J.R.E. NAR
14 pp. 635-644 (1986)

Chapter 5

Methylation of the H5 gene

A number of studies (1,2,3,4), have shown that the level of DNA methylation for a given gene is related to the transcriptional activity of the gene. Transcriptionally active genes tend to be hypomethylated (undermethylated) while inactive genes tend to be hypermethylated. While hypomethylation is not completely responsible for rendering a gene transcriptionally active, it is nevertheless a useful marker for determining if a gene may be active in a given cell type. The level of DNA methylation within a gene is usually determined by digestion of the DNA with restriction enzymes. The three most commonly used restriction enzymes are CfoI, MspI, and HpaII. It should be noted that most cytosine methylation (in eukaryotes) occurs on a cytosine followed by a guanine residue ($C^{me}G$). The recognition sequence for all three restriction enzymes includes this pair of nucleotides. As shown in Table 3, CfoI can only digest the DNA if its recognition site (GCGC) is not methylated. The other two enzymes, HpaII and MspI, are a set of enzymes that have the same recognition site (CCGG). HpaII however cannot digest the DNA if either cytosine residue within this recognition sequence is methylated. MspI is able to digest the DNA if the central cytosine nucleotide is methylated but cannot cleave the DNA if the 5′

Table 3

	CCGG	$CC^{me}GG$	$C^{me}CGG$	GCGC	$GC^{me}GC$
CfoI	----	----	----	+	-
HpaII	+	-	-	----	----
MspI	+	+	-	----	----

cytosine is modified. The ability or inability of these enzymes to digest the DNA reflects the degree of DNA methylation within the gene.

The methylation status of the H5 gene was determined for several cell types. Reticulocytes and erythroblasts (which were slightly contaminated with white blood cells) from both normal and RAV-1 infected chickens were obtained from Dr. MariBeth Raines. The RAV-1 infected chickens show the symptoms of erythroblastosis, due to the activation of the c-erb gene locus. As a control, chicken sperm DNA was also digested with the same three enzymes. Sperm DNA is known to be extremely hypermethylated and most, if not all, detectable methylation sites should be modified. The DNA from all five cell types were digested with the various restriction enzymes and transferred to nitrocellulose paper

Figure 23 Southern blot of DNA from various cell types digested with CfoI, HpaII, or MspI. The two unusual bands are indicated by "X's".

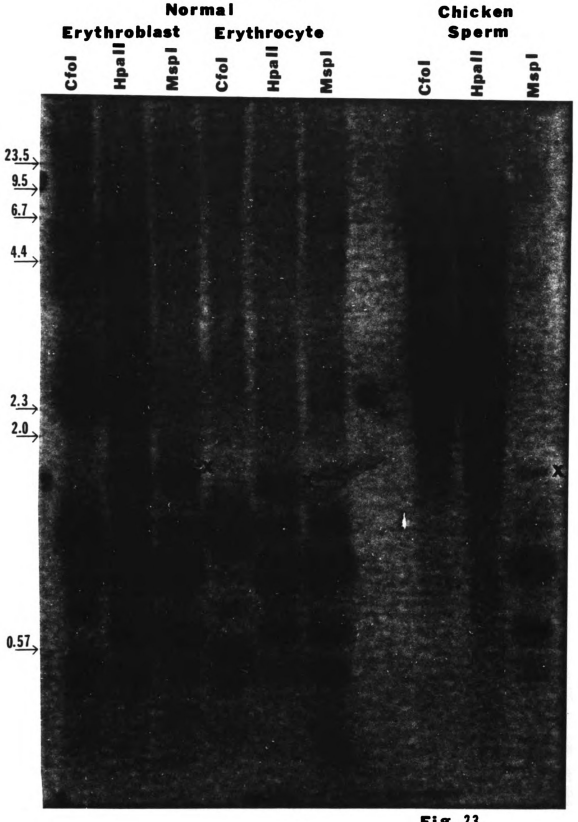

Normal
Erythroblast
Erythrocyte
Chicken
Sperm
Cfol
HpaII
MspI
Cfol
HpaII
MspI
Cfol
HpaII
MspI
23.5
9.5
6.7
4.4
2.3
2.0
0.57

Fig. 23

Figure 24 DNA from RAV-1 infected erythroblasts and reticulocytes digested with CfoI, HpaII, or MspI. Unusual bands are indicated by "X's".

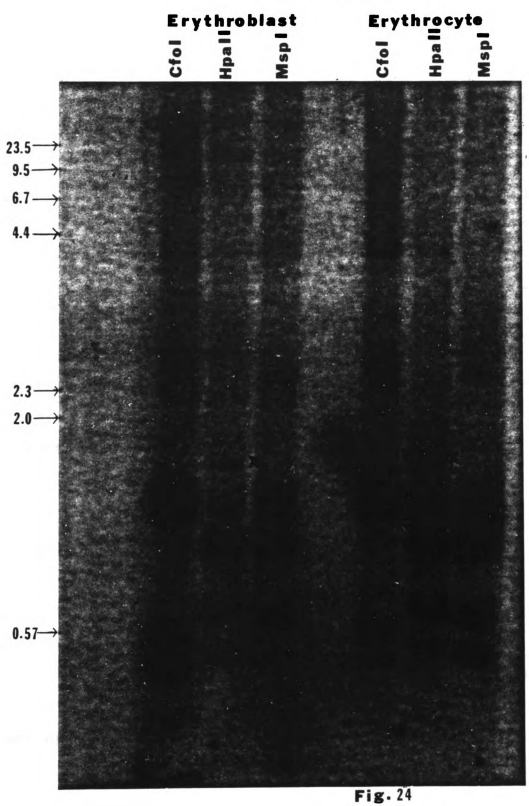

RAV−1 infected
Erythroblast
Erythrocyte
CfoI
HpaII
MspI
CfoI
HpaII
MspI
23.5
9.5
6.7
4.4
2.3
2.0
0.57

Fig. 24

by Southern blotting. The resulting blots were probed with EcoRI-BamHI insert from pH5-BR 2.4, with the results shown in Fig. 23 and 24.

Since the sequence of the H5 gene is known, the sizes of the bands that should result from the enzymatic digests (if the recognition sites are not methylated) may be predicted. The sizes of the predicted bands are listed in Table 4, and are seen in Figures 23 and 24. The normal reticulocyte and erythroblast DNA is shown in Figure 23, on the left. Both CfoI and HpaII were able to digest the DNA to give the predicted banding pattern, indicating that the H5 gene for both reticulocytes and erythroblasts is hypomethylated. As mentioned in the Literature Review, the H5 gene is known to be transcriptionally active in both erythroblasts and reticulocytes. A number of high molecular weight, H5 hybridizing bands are present in both the CfoI and HpaII lanes. These bands most likely came from the DNA of the contaminating white blood cells. The reticulocytes and erythroblasts were isolated by banding on a Percoll gradient. This procedure however, would not purify the erythroblasts from the white blood cells. Since the H5 gene is not transcriptionally active in white blood cells, it is likely that the gene would be hypermethylated, rendering CfoI and HpaII incapable of digesting the DNA.

Table 4

HpaII/MspI	CfoI
1138	1180
879	1111
560	482
371	377
158	180
74	49
70	
45	
36	
30	

Table 4 Chicken DNA from various sources was digested with either CfoI, HpaII, or MspI, and probed with the 2.4 kilo base pair insert from the plasmid pH5-BR 2.4. From the known sequence of the H5 gene, the sizes of the restriction enzyme fragments that will hybridize to the H5 probe may be predicted. The fragments listed below the gap in the table were not detected on these blots.

The erythroblast and reticulocyte DNA from RAV-1 infected chickens are shown in figure 24. The chickens infected with RAV-1 exhibit symptoms of erythroblastosis, and the levels of erythroblast cells are higher in these chickens than in normal chickens. Since so many erythroblasts are present, problems with contaminating white blood cells are minimized. RAV-1 erythroblast DNA digested with either CfoI or HpaII shows none of the higher molecular weight bands that were detected with normal erythroblast DNA. The chicken sperm DNA in the center is very highly methylated and neither CfoI or HpaII was able to digest the DNA around the H5 locus. MspI was, as expected, able to digest the methylated DNA to yield a banding pattern similar to the MspI banding pattern of erythroid DNA.

There are two bands of interest that were not predicted to have occured (indicated by arrows). One of these bands is present in both the MspI digestion of normal erythroblast DNA and in the MspI digestion of the chicken sperm DNA. This band is not present in the MspI digest of either reticulocyte DNA or the RAV-1 erythroblasts and may indicate a modified MspI recognition site that MspI cannot digest. For the normal erythroblast DNA lane, this modified MspI site is probably from the contaminating white blood cell DNA rather than from the erythroblast DNA. While this site may represent a $C^{me}CGG$ modification (Table 3), this

type of methylation is not common. Two groups (5,6) have found that two sequences, GGCC^{me}GG and CC^{me}GGCC, are highly resistant to digestion with MspI. Two such sequences occur within the H5 gene region being studied. One of these sequences (GGCCGG) separates a 1138 base pair fragment and a 371 base pair fragment. The combined fragment size of 1509 base pairs is roughly the same size as the unpredicted MspI band. Therefore, it is most likely that this MspI restriction enzyme fragment is the result of the inability of MspI to fully digest a GGCC^{me}GG sequence within the methylated H5 DNA. A second such sequence within the H5 gene (CCGGCC), combines a 158 base pair fragment and a 74 base pair fragment to yield a fragment 232 base pairs in length. A band of this size however, was not detected on these blots.

The other band of interest is present in both normal and RAV-1 infected reticulocyte and erythroblast DNA digested with HpaII. This band apparently is the result of a HpaII/MspI recognition site that remains methylated even though the H5 gene is transcriptionally active. The most likely candidate is the recognition site at -528. This site falls at the junction of a 560 bp and a 879 bp fragment to yield a total fragment size of roughly 1440 bp. This size is in agreement with the unusual band's size. The significance of this methylation is unclear. This site may lie outside the H5 gene domain and loss of the methylation may not be

neccessary. This is not likely, since there is another restriction enzyme site 5′ of this site which shows the predicted methylation pattern.

Summary

The H5 gene is hypomethylated in both the reticulocyte and the erythroblast stages of erythropoeisis, but is hypermethylated in both the contaminating white blood cell DNA and in the chicken sperm DNA. The amount of methylation correlates with the known transcriptional activity of the H5 gene. The H5 gene is known to be transcribed in both the reticulocyte and erythroblast but is not transcribed in non-erythroid cells. It has also been found that one of the HpaII/MspI recognition sites remains methylated, even when the H5 gene is being actively transcribed.

H5 Gene Restriction Enzyme Site Polymorphisms

DNA sequences may be altered by naturally occuring mutations and passed to the organism's offspring. Certain mutations may either remove a normally present restriction enzyme recognition site or create a new restriction enzyme recognition sequence within the organism's DNA. Different organisms within the same species therefore, may show

different fragment patterns (polymorphisms) around a given gene locus when digested with a given restriction enzyme.

The region surrounding the H5 gene locus was analyzed for SacI restriction site polymorphisms. DNA from a number of inbred chicken lines, as well as DNA from the related Galliforms Japanese quail (Coturnix coturnix), the turkey (Meleagris gallopavo), and the ring-necked pheasant (Phasianus colchicus) was obtained from Dr. Gene Smith (USDA Regional Poultry Laboratories) and digested with the restriction enzyme SacI. The DNA fragments were transferred to nitrocellulose filters by Southern blotting and probed with the EcoRI-BamHI insert of pH5-BR 2.4. From the restriction enzyme map of the lambda phage clones (Figure 5), three SacI fragments should hybridize to the H5 probe if no restriction enzyme site polymorphisms have occurred. The sizes of the three predicted fragments are 350 bp, 2.3 kb, and 6.5 kb in length. As shown in Fig. 25, all of the chicken line DNA's had the predicted fragments. Even though none of the SacI recognition sites are within the H5 gene, and are not under any known selective pressure, none of the sites have been mutated. A few of the chicken DNAs have fainter high molecular weight bands between the predicted 2.3 kb and 6.5 kb bands. These are most likely the result of incomplete SacI digestion rather than polymorphisms.

Turkey DNA has a SacI pattern similar to the chicken,

but the upper band appears to be slightly larger than the 6.5 kb band of the chicken. The exact amount of the increase cannot be determined from this gel but may be up to 1 kb. Both the quail and the pheasant DNAs show large differences from the chicken DNAs. Both of these H5 histone genes have lost the smaller 350 base pair SacI fragment. In the pheasant DNA, the 5' SacI site flanking the 350 bp fragment has been lost, and the fragment has been combined with the the 2.3 kb fragment to yield a new fragment 2.6 kb in length. In the quail genome, both the 350 bp and 2.3 kb fragments are missing, and replaced with a fragment 3.5-4.0 kb in length.

The chicken lines tested in this experiment do not exhibit any SacI restriction site polymorphisms. While none of the SacI restriction enzyme sites are within the H5 gene itself, none have been mutated and lost. This experiment, however, does not rule out that the various chicken H5 genes may have other restriction enzyme site polymorphisms. The three other avian H5 genes tested showed similarities to the chicken H5 gene, but showed the loss of one or more of the SacI recognition sites that are present in the chicken.

The SacI DNA blot was also probed with the chicken α^A-globin gene. As shown in Fig. 26, the chicken lines again did not exhibit any SacI polymorphisms. The turkey α^A-

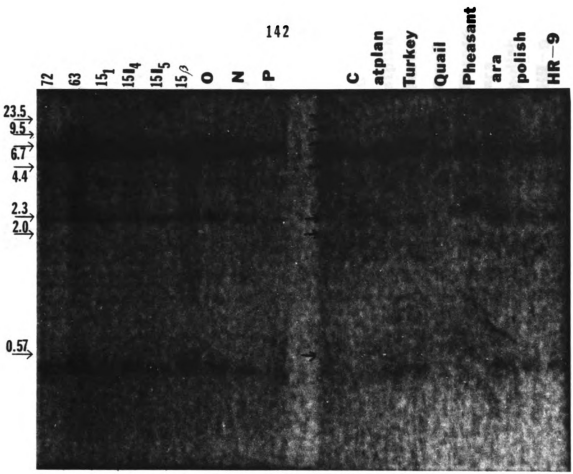


Figure **25** DNA from a number of inbred chicken lines as well as DNA from the domestic turkey, ring-necked pheasant and the Japanese quail were digested with SacI and transferred to nitrocellulose by Southern blotting. The filter was probed with the insert from pH5-BR 2.4. From the restriction enzyme map of the phage clone cH5-3 (Fig. 5), three fragments, 6.5 kb, 2.3 kb, and 0.4 kb in size, should be detected if no SacI restriction enzyme site polymorphisms have occured.

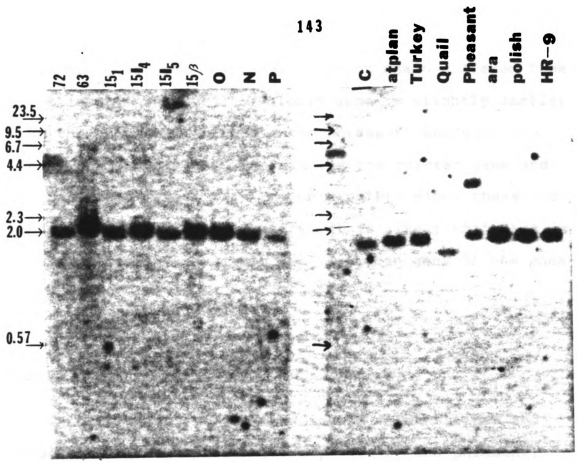


Figure **26** Nitrocellulose filter described in Figure 23 probed with the insert from the clone pBRα7-1.7 (11), which contains an EcoRI-BamHI fragment covering the α^A globin gene of the chicken.

globin gene appeared to be the same size as the chicken α^A-globin gene, but quail and pheasant again differed from the chicken. The quail α^A-globin gene is slightly smaller than the chicken gene, while the pheasant gene gave two bands, one roughly the same size as the chicken gene and, one band much larger than the chicken's. Since these two bands are of the same intensity, it is likely that this is an allelic variation within the α^A-globin gene of the pheasant.

Discussion

In conclusion, 14 separate inbred chicken lines and three other avian species (turkey, quail, and pheasant) were digested with the restriction enzyme SacI and analyzed at two different gene loci. None of the chicken lines showed evidence of SacI restriction site polymorphisms either around the H5 gene or the α^A-globin genes. It is apparent that none of the chicken lines have evolved separately from one another for a long enough time span to display any differences in their SacI fragment sizes for the two genes studied here.

The three other avian species analyzed here are related to the chicken but have evolved separately for various lengths of time. On non-molecular evidence, the ring-necked

Table 5

Lysozyme immunological index of dissimilarity

	Anti-chicken antisera A24	Anti-chicken antisera A54
Chicken	1.00	1.00
Turkey	1.66	1.16
Japanese quail	1.86	1.81
Ring-necked pheasant	2.20	1.70

Table from Arnheim and Wilson (8)

Table 6

Immunological distance of the albumin and transferrin proteins

	Albumin	Transferrin
chicken	0.0	0.0
turkey	18.0	32.0
quail	23.0	35.0
ring-necked pheasant	29.0	35.0

Table from Prager and Wilson (9)

pheasant was originally considered to be closer to the chicken evolutionarily than either the Japanese quail or the turkey (7). The study of the avian lysozyme, albumin, and transferrin proteins has altered that view. Immunologically, the lysozyme, albumin, and transferrin proteins of the chicken are closer to that of the turkey while the ring-necked pheasant and quail are less similar (8,9,10).

The SacI digestions appear to support this arrangment of the avian species. From the SacI fragment sizes of the H5 and α-globin genes, the turkey would appear be the closest relative to the chicken of the three species tested. The turkey has an H5 gene SacI fragment pattern nearly identical to that of the chicken, while the α^A-globin gene SacI band is the same size as the chicken's.

The pheasant would appear to be the next closest relative. The pheasant H5 gene SacI banding pattern is similar to that of the chicken, and the loss of the smaller SacI fragment appears to be the result of the loss of a single SacI recognition site compared to the chicken gene. One of the α^A-globin SacI bands of the pheasant is the same size as that of the chicken. The pheasant however, has another α^A-globin band that apparantly is the result of allelic variation.

Of the three avian species tested here, the quail appears to the least related to the chicken. The quail H5 gene

SacI banding pattern indicates several changes from the chicken. To obtain the quail SacI pattern, it appears as though several SacI sites have been altered. The α^A-globin genes of the quail and chicken are also different. The quail α^A-globin band is roughly 300 base pairs smaller than the chicken α^A-globin band.

Determining how close the four avian species are related to each other and how different the various chicken lines are from each other is based only on studying two genes and digestion with one restriction enzyme, Sac I. While the SacI digestions agree with the earlier protein studies in how the three galliform species are related to the chicken, a more complete study would require several different restriction enzyme digestions and more genes being analyzed.

References

1) Jones, P., Taylor, S. Cell 20 pp. 85-93 (1980)

2) Christy, B., Scangos, G. PNAS 79 pp. 6299-6303 (1982)

3) Cedar, H. PNAS 78 pp. 4246-4250 (1981)

4) Fradin, A., Manley, J., Prives, C. PNAS 79 pp. 5142-5146 (1982)

5) Busslinger, M., deBoer, E., Wright, E., Grosveld, F., Flavell, R. NAR 11 pp. 3559-3570 (1983)

6) Keshet, E., Cedar, H. NAR 11 pp. 3571-3580 (1983)

7) Arnheim, N., Prager, E., Wilson, A. J. Biol. Chem. 244 pp. 2085-2094 (1969)

8) Arnheim, N., Wilson, A. J. Biol. Chem. 242 pp. 3951-3956 (1967)

9) Prager, E., Wilson, A. PNAS 72 pp. 200-204 (1975)

10) Jolles, J., Schoentgen, F., Jolles, P., Prager, E., Wilson, A. J. Mol. Evol. 8 pp. 59-78 (1976)

11) Engel, J., Rusling, D., McCune, K., Dodgson, J. PNAS 80 pp. 1392-1396 (1983)