1

NEC

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Services des thèses canadiennes

Ottawa, Canada
K1A 0N4

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

DNA Sequence Characterization of Xenopus laevis Histone H5
and Its Evolutionary Implications

by

© Robert Gray Rutledge B.Sc. (Hons.), M.Sc.

A thesis submitted to the Faculty of
Graduate Studies in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Biology Department
Carleton University
Ottawa, Ontario
December, 1986.

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse, ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-35866-1

7
The undersigned hereby recommend to the Faculty of
Graduate Studies acceptance of this thesis, submitted by
Robert Gray Rutledge, B.Sc. (Hons.), M.Sc., in partial fulfillment
of the requirements for the degree of Doctor of Philosophy

Chairman, Department of Biology

Supervisor

Co-supervisor

ABSTRACT

cDNA clones coding for two variants of Xenopus laevis H5 (also called H1^S) were isolated from an M13mp9 erythrocyte cDNA bank by immunological screening and nucleic acid hybridization. Called XLH5A and XLH5B, they are extremely similar in sequence and code for proteins of 195 and 193 amino acids respectively, which may be equivalent to the H1D and H1E variants previously characterized by Risley and Eckhardt (Dev. Biol. 84, 79 1981). X. laevis H5 mRNA is likely polyadenylated and lack the highly conserved, 23 nucleotide "dyad symmetry element" found within the 3' untranslated regions of H1 and core mRNA. Comparison of over 6300 nucleotides from twelve XLH5A cDNA clones revealed six single-nucleotide anomalies that may represent errors introduced by reverse transcriptase, in addition to two XLH5A cDNAs originating from a putative polymorphic mRNA which differed at two nucleotides. Seven XLH5A cDNAs also initiated reverse transcription at an A-rich segment in the middle of the coding region. Partial restriction maps of the XLH5A and XLH5B genes were constructed from Southern blot analysis of X. laevis genomic DNA.

Sequence comparisons show that X. laevis H5 is highly homologous to mammalian H1^O and avian H5, and that all can be divided into 2 variable and 2 conserved protein domains. The latter include a conserved 49-residue carboxyl domain that is absent in H1, and a conserved 80-residue middle domain containing a 21-residue segment whose sequence is distinct from that of H1. Based upon this and other evidence, it is proposed that X. laevis H1^S and mammalian H1^O be included into the H5 subclass.

A detailed comparative DNA sequence analysis of several H5 and H1^O coding regions illustrates that they have a common ancestry, and that some differences in their amino acid sequence can be directly linked to

four mutational processes. These include deletions, single-base mutations; and two forms of DNA sequence modifications called "direct duplication" and "mirror duplication". By resolving past mutational events, the evolution of the amino terminal domains of duck H5, chicken H5 and human H1⁰ is traced back to a hypothetical ancestral H5 gene, which is very similar to X. laevis H5. The establishment of evolutionary lineages from analysis of DNA sequence is also discussed.

To Carol, Lisa and Tara.

ACKNOWLEDGEMENTS

Special thanks to Dr. Vern Seligy whose generous support and long term commitment has been central to the success of this project. For his invaluable advice and support I am again indebted to Dr. Jim Neelin. I would also like to thank Mike Dove for his expert technical support and Vern for providing his artistic expertise during the organization and construction of the figures. I am also indebted for the support and facilities provided by the National Research Council of Canada, Division of Biological Sciences, and the Natural Sciences and Engineering Research Council of Canada, which made this project possible. Special thanks to my wife Carol for the love and support she has given me during these difficult years, and to my friends Shah, Steve and Mike who have shared with me the most exciting, as well as the most difficult phases of this project.

TABLE OF CONTENTS

Title	1
Abstract	iii
Acknowledgements	v
Table of Contents	vi
List of Tables	ix
List of Figures	x
List of Abbreviations	xiii
 I. Introduction	 1
A) Histones and Their Role in Chromatin Structure	1
B) Chromatin Structure and the Regulation of Gene Transcription..	2
C) Vertebrate Linker Histone Diversity	4
D) Characterization of the Vertebrate Linker Histone Subclasses..	7
 II. Materials and Methods	 12
A) Preparation and Characterization of RNA	12
i) RNA Extraction	12
ii) RNA Fractionation	12
iii) <u>In Vitro</u> Translation of mRNA	13
B) cDNA Synthesis	13
C) Construction and Screening of M13mp9 cDNA Banks	15
i) M13mp9 cDNA Bank Construction	15
ii) Immunological Screening of M13mp9 cDNA Banks Using a Polyclonal H5-antibody	16
iii) Rescreening of the S2 Bank Using Probes from XLH5A-1	18
D) DNA Probe Preparation	19
i) Strand-specific Probes for Plaque Hybridizations	19
ii) Single-stranded Probes for DNA and RNA Blot Hybridization..	21

E) DNA Isolation from <u>Xenopus laevis</u> Erythrocytes	21
F) DNA and RNA Blot Analysis	22
G) DNA Sequencing	23
H) Computer Sequence Alignment	27
III. Results	28
A) Cloning of <u>Xenopus laevis</u> H5 cDNA	28
1) RNA Fractionation and M13mp9 cDNA Bank Construction	28
ii) Immunological Screening of the M13mp9 cDNA Banks	32
iii) Rescreening of the S2 cDNA Bank with XLH5A-1	36
B) <u>Xenopus laevis</u> H5 cDNA Sequence Analysis	36
1) Preliminary Characterization of the H5 cDNA Inserts	36
ii) Single-Base Sequence Anomalies	39
iii) Ligation Artifacts	42
iv) Identification of Two Distinct <u>Xenopus laevis</u> H5 cDNA's	44
C) Sequence Analysis of the H5 and H1 ⁰ Histones	54
1) Amino Acid Sequence Comparisons	54
ii) DNA Sequence Comparisons	57
D) Investigation into the DNA Sequence Evolution of H5 and H1 ⁰	62
1) Region 1 and 3 of <u>X. laevis</u> H5 and Human H1 ⁰ : Characterization of Direct Duplication	62
ii) Region 1 of Duck and Chicken H5: Characterization of Mirror Duplication	69
iii) Region 1, Section 1 of Duck and Chicken H5: Derivation of an Ancestral Avian H5	75
iv) Evolution of Ancestral Avian H5 from Ancestral Amphibian H5	80
v) Summary of the DNA Sequence Evolution for Region 1	83
vi) H5 Amino Acid Sequence Evolution for Region 1	90
vii) DNA Sequence Comparisons of Region 3	95

E) Comparison of the 3' and 5' Untranslated Regions of H5 and H1 ^o	95
F) Amino Acid Sequence Comparisons of H1 and H5	104
IV. Discussion	114
A) The Molecular Cloning of cDNA	114
i) A Technique for the Rapid Synthesis and Cloning of dscDNA..	114
ii) The Use of M13 as a Primary Cloning Vector	115
B) The Isolation and Identification of cDNAs Clones Coding for Two Variants of <u>Xenopus laevis</u> H5	116
C) Reclassification of Mammalian H1 ^o and Amphibian H1 ^s into the Same Linker Histone Subclass as Avian H5	119
D) Amino Acid Sequence Characteristics of the H5 Histones	120
E) Investigation of the Mechanism for Amino Acid Sequence Modification	125
F) Summary of the Methodology for Analysis of DNA Sequence Evolution	128
G) The Analysis of DNA Sequence for Establishing and /or Confirming Evolutionary Lineages	131
H) Future Considerations: Elucidating the Evolutionary Origins of H5	133
V. References	135

LIST OF TABLES

Table 1.	Summary of <u>Xenopus laevis</u> H5 cDNA single-base sequence anomalies	41
Table 2.	Comparisons of the cDNA predicted amino acid contents of H5A and H5B with that previously determined from <u>Xenopus laevis</u> H5 protein	49
Table 3.	Summary of Region 2 amino acid sequence homologies between vertebrate and nonvertebrate linker histones	113

LIST OF FIGURES

Figure 1. Schematic of the strategy for nucleotide sequence determination of the M13mp9-H5 cDNA clones	24
Figure 2. <u>In vitro</u> translation and sucrose gradient fractionation of total RNA isolated from immature <u>Xenopus laevis</u> erythrocytes	29
Figure 3. Double-stranded cDNA synthesized from fractionated erythrocyte RNA	31
Figure 4. Schematic of H5 cDNA isolation using immunological screening	33
Figure 5. Characterization of XLH5A-1 and XLH5A-2	35
Figure 6. Schematic of the <u>Xenopus laevis</u> H5 cDNA inserts	37
Figure 7. RNA blot analysis of liver and erythrocyte RNA isolated from anemic <u>Xenopus laevis</u>	40
Figure 8. Nucleotide sequence of non-H5 cDNA segments within XLH5A-36 and XLH5A-16	43
Figure 9. Nucleotide sequence of the <u>Xenopus laevis</u> H5A, and H5B cDNAs	45
Figure 10. Partial restriction mapping of the <u>Xenopus laevis</u> H5 genes	50
Figure 11. Comparison of the cDNA and genomic restriction maps for <u>Xenopus laevis</u> H5A and H5B	52

Figure 12.	H5 and H1 ^o amino acid sequence alignment	55
Figure 13.	DNA sequence alignment of H5 and H1 ^o coding regions	58
Figure 14.	Summary of the major DNA sequence modifications proposed to have occurred within Region 1 of H5 and H1 ^o during the divergence of amphibians, avians and mammals	84
Figure 15.	Amino acid sequences of the predicted evolutionary intermediates for Region 1 of H5 and H1 ^o	91
Figure 16.	Amino acid sequence alignment of the H5 and H1 ^o intermediates for Region 1 that incorporates all proposed DNA mutations	93
Figure 17.	Amino acid sequences of the H5 and H1 ^o intermediates of Region 1 aligned for maximum homology	94
Figure 18.	Preliminary DNA sequence alignment for Region 3 of H5 and H1 ^o coding regions	96
Figure 19.	Alignment of the H5 and H1 ^o 3' untranslated regions	98
Figure 20.	Alignment of the H5 and H1 ^o 5' untranslated regions	101
Figure 21.	Amino acid sequence alignment of Region 2 from several vertebrate H1's	105
Figure 22.	Comparison of the consensus amino acid sequences for Region 2 of vertebrate H1 and H5	106

Figure 23. Amino acid sequence comparisons of Region 2 of trout H1
and mammalian H1t with the consensus sequences of
vertebrate H1 and H5 108

Figure 24. Amino acid sequence comparisons of Region 2 of
several nonvertebrate H1's with the consensus
sequences of vertebrate H1 and H5 110

LIST OF ABBREVIATIONS

bp	base pair
cDNA	DNA made from RNA by reverse transcriptase
Ci	Curie
dNTP	mixture of all four deoxyribonucleotide triphosphates
DTT	dithiothreitol
dscDNA	double-stranded cDNA
IPTG	isopropylthiogalactoside
M	molar
mg	milligram
ml	milliliter
mM	millimolar
mmol	millimole
ng	nanogram
nt	nucleotide
rRNA	ribosomal RNA
SSC	0.15 M NaCl, 0.015 M sodium citrate
SDS	sodium dodecyl sulphate
STE	10 mM Tris pH 8.0, 1 mM EDTA, 100 mM NaCl
TAE	40 mM Tris-Acetate pH 8, 2 mM EDTA
TBS	50 mM Tris pH 8.0, 200 mM NaCl
uCi	microCurie
ug	microgram
ul	microliter
UTR	untranslated region of mRNA

I. INTRODUCTION

A) Histones and Their Role in Chromatin Structure

Histones, small basic proteins found in association with eukaryotic DNA, function to supercoil DNA into a condensed, highly structured form. Consistent with their common role as the major structural proteins of chromatin, most of the histones characterized to date share distinct metabolic and genetic characteristics. For example, the expression of most histone genes is closely coupled to DNA replication (Borun 1975, Pederson 1976, Maxson et al. 1983), so that the majority of histone protein is synthesized only during S-phase. In addition, inhibition of DNA replication with such agents as cytosine arabinoside or hydroxyurea results in the general inhibition of histone synthesis and degradation of histone mRNA. The majority of the histone genes are also organized into clusters containing one gene for each of the five histone classes (Kedes 1979, Hentschel and Birnstiel 1981, Maxson et al. 1983). The vast majority of histone genes characterized to date also lack introns, and histone mRNA generally lacks the 3' polyadenylation characteristic of other eukaryotic mRNAs (Kedes 1979, Hentschel and Birnstiel 1981).

The histones are a relatively simple family of proteins, consisting of only five major classes or types. These can be subdivided into two functional groups, consisting of the "core" histones H2A, H2B, H3 and H4, and the lysine-rich or "linker" histones, which are broadly referred to as the H1 histones (Elgin and Weintraub 1975, Isneberg 1979). Two each of the four core histones assemble in vivo to form octameric protein particles called nucleosome cores, around which DNA is coiled to produce repeating structures resembling "beads on a string" (Kornberg 1977, Felsenfeld 1978, McGhee and Felsenfeld 1980). This represents the lowest

level of chromatin organization, and is generally referred to as the primary chromatin fiber or the 10 nm fiber. The DNA connecting one nucleosome core to the next is called linker DNA and is a site of association for the linker histones.

An important aspect of chromatin function is the ability of the primary chromatin fiber to supercoil, generating a 30 nm diameter fiber, which can be organized into still higher levels of compaction (Thoma et al. 1979, Worcel et al. 1981, Cartwright et al. 1982, Igo-Kemenes et al. 1982). This phenomenon is generally referred to as chromatin fiber condensation. Although very little is known about the mechanism, it has been well established that the linker histones are central to the formation and maintenance of the 30 nm chromatin fiber, and that the amount of linker histone is directly related to the extent of chromatin condensation (Thoma et al. 1979, Thoma and Koller 1981, Allan et al. 1981). It thus appears at present that the primary function of the linker histones is to modulate condensation of the primary chromatin fiber.

B) Chromatin Structure and the Regulation of Gene Transcription

Studies into chromatin function have demonstrated an important link between chromatin fiber condensation and the restriction of gene transcription. This comes in part from the demonstration that inactive genes are contained within highly condensed chromatin structures, and that alteration of higher-order structure occurs within transcriptionally active chromatin (Mathis et al. 1980, Igo-Kemenes 1982, Cartwright 1982, Reeves 1984). This disruption of chromatin condensation can be associated with depletion of linker histone (Mathis et al. 1980, Kimura et al. 1983, Smith et al. 1984b, Karpov et al. 1984), suggesting an inverse relationship between linker histone content and gene

transcription. This is consistent with reports of significant variations in linker histone content within different tissues (Unger-Ullmann and Modak 1979, Bates and Thomas 1981; see also Allan et al. 1981), and by the non-uniform distribution of linker histone within chromatin (Huang and Cole 1984, Smith et al. 1984b; however, see also Komaiko and Felsenfeld 1985, Glotov et al. 1985).

Taken together, the evidence indicates that the linker histones could play a central role in the repression of gene transcription, a proposal supported by several lines of evidence. Hannon et al. (1984) have shown that inhibition of RNA polymerase binding to chromatin in vitro is directly correlated with the linker histone-mediated formation of higher-order structure. Schlissel and Brown (1984) have described H1-mediated repression of transcription of the oocyte 5S RNA gene in somatic tissues. In contrast, the transcriptionally active somatic RNA 5S genes possess "active stable transcriptional complexes" which protect them from H1-mediated repression. This concept of differences in the binding of linker histone within chromatin is further supported by Weintraub's description of a 20-40 Kb supranucleosomal structure within chicken erythrocyte chromatin (Weintraub 1984). Particles which contain inactive genes are "crosslinked" by linker histones, making them resistant to nuclease disruption. In contrast, particles containing previously active genes, are easily disrupted by nuclease digestion, despite the presence of linker histone. This led to the conclusion that linker histone binding differs within active and inactive chromatin. Finally, Karpov et al. (1984), using an elegant technique, illustrated that during its induction and subsequent transcription, the coding region of the hsp 70 gene is progressively depleted first of linker histone and then of core histone. They suggest that selective dissociation of the histones

4

initiates unravelling of the 30 nm transcriptionally inactive chromatin fiber into moderately transcribed 10 nm fiber and actively transcribed linearized DNA.

The evidence thus indicates that the modulation of gene transcription is closely associated with, if not regulated by the modulation of chromatin fiber condensation. Equally important is the central role of the linker histones in chromatin condensation, which indicates that they could be a major component in the regulatory mechanism of gene transcription. Developing a precise understanding of linker histone function could therefore be an important step in elucidating how the control of gene expression is achieved in eukaryotes.

C) Vertebrate Linker Histone Diversity

The linker histones are the most complex and least understood of all the histone classes. Their characterization has revealed extensive structural heterogeneity, related in part to the expression of several forms of linker histones within a single cell type or tissue (Hohmann 1978). Another factor contributing to their complexity is the substantially lower level of sequence conservation of the linker histones, as compared to the core histones. This has led to structural differences between even closely related linker histones.

Despite the difficulties of studying such a large and structurally heterogeneous protein family, a basic understanding of linker histone diversity is now slowly emerging. For example, several types of vertebrate linker histones have been identified to date. These include: the H1 histones, reported to be present even in primitive eukaryotes (Isenberg 1979, Igo-Kemenes 1982, Smith et al. 1984a); mammalian H1t, a testis-specific histone (Seyedin and Kistler 1980, 1981, Cole et al.

1986); mammalian H1^O (Panyim and Chalkley 1971; see below); the supposed erythrocyte-specific avian H5 (Neelin et al. 1964, Neelin 1968, see below) and fish H5 (Miki and Neelin 1975, 1977a, 1977b); amphibian H1^S (Brown et al. 1981, see below) and reptilian H1^S (Rutledge et al. 1981). It is believed that some of these linker histones may represent distinct "subclasses" of linker histones, which differ substantially in their structure and apparent function. Linker histone subclasses could thus represent the first of possibly many levels of linker histone diversity, and will be described in more detail in the next section.

Another level of structural diversity is also evident, and takes the form of linker histone "variants" within the subclasses. The most extensively characterized of these are the primary sequence variants of H1 (Hohmann 1978, 1980). Variants of mammalian H1^O (Medvedev and Medvedeva 1980, Smith and Johns 1980a, Pehrson and Cole 1981, D'Anna et al. 1981) and Xenopus laevis H1^S (Brown et al. 1981, Risley and Eckhardt 1981) have also been characterized.

Although the amino acid sequences for most of these variants have not been determined, gene sequences are available for two variants each of chicken H1 (Sugarman et al. 1983, Coles and Wells 1985), Xenopus laevis H1C (Turner et al. 1983; see also Perry et al. 1985) and Xenopus laevis H1^S (this study). Importantly, their predicted amino acid sequences indicate that each of these pairs of variants possess only small differences in peptide length within otherwise homologous proteins. The limited differences in structure therefore imply that these linker histone variants possess very similar if not identical functions. In addition, the very extensive DNA sequence homologies between these genes, suggest that at least some linker histone variants were formed by gene duplication, followed by independent mutational events which have

produced the structural differences that now distinguish their proteins. Isolation of the genes coding for other linker histone proteins will be necessary however, before this can be established as a general characteristic of the linker histone variants.

Another distinguishing characteristic of the linker histone variants has been their differential expression. Specifically, it has been shown that even when the same H1 variants are present in each tissue within an animal, the relative quantities of each H1 variant differ substantially in a tissue-specific manner (Kinkade 1969, Hohmann 1978, 1980). The relative amounts of some H1 variants have also been found to be modified by developmental and hormonal factors (Hohmann 1978, Ajiro et al. 1981, Hentschel and Birnstiel 1981, Lemox and Cohen 1983), suggesting that the H1 variants could be involved in the modulation of gene expression that is associated with these processes. This has led to speculations that the H1 variants are a major contributor to the extensive functional diversity within chromatin, believed necessary for differential gene expression in eukaryotes (Cole 1977, Hohmann 1978, 1980). Note that the two Xenopus laevis H1^s variants have also been reported to be differentially expressed in a tissue-specific manner, with liver containing about equal amounts of each, whereas the erythrocyte contains almost none of the smaller variant (Brown et al. 1981, Risley and Eckhardt 1981).

Despite the apparent significance of differential expression, the evidence demonstrating only limited structural differences between some of these linker histone variants (see above), brings into doubt the true extent of their functional diversity. However, without the ability to directly test for functional differences between linker histones (e.g. development of a functional assay), the functional significance of the

occurrence and differential expression of linker histone variants will likely remain unclear.

To summarize, vertebrate linker histone diversity appears to consist of at least two major levels of complexity. The most extensive structural differences, which likely initiate substantial functional differences, are characteristic of the linker histone subclasses, represented by such histones as H1^D and H5. Another level of diversity is reflected by linker histone variants, which are characterized by much smaller differences in structure and likely in function and are represented by the primary sequence variants of H1, H1^O and Xenopus laevis H1^S.

D) Characterization of the Vertebrate Linker Histone Subclasses

The largest and best characterized of the linker histone subclasses is H1. As was outlined earlier, the vertebrate H1 histones are similar to the core histones, in the coupling of both H1 and core histone synthesis to DNA replication, the nonpolyadenylated H1 and core mRNA, and the clustering of H1 genes with core histone genes. H1 and core mRNA has also been found to contain 3' untranslated nucleotide sequences that have been conserved over vast evolutionary times (Hentschel and Birnstiel 1981, Coles and Wells 1985).

Avian H5 is another well characterized linker histone and possesses several characteristics that distinguish it from H1. In addition to its supposed erythrocyte-specificity, these include metabolic characteristics unlike those found for the H1 and core histones. Specifically, avian H5 synthesis is not closely coupled to DNA replication, illustrated by the insensitivity of its synthesis to cytosine arabinoside (Seligy et al. 1973) and hydroxyurea (Ruiz-Carrillo et al. 1976), by its synthesis in the

absence of DNA replication, during the final stages of erythrocyte maturation (Appels and Wells 1972, Seligy et al. 1973, Ruiz-Carrillo et al. 1976, 1983, Sung et al. 1977), and by transcription of its gene throughout the cell cycle (Dalton et al. 1986). The amount of H5 also increases dramatically during erythrocyte maturation (Billet and Hindley 1972, Seligy et al. 1973, Ruiz-Carrillo et al. 1974, Weintraub 1978, Urban et al. 1980, Dalton et al. 1985).

These distinctive characteristics have led to the generally held belief that avian H5 has evolved a specialized function distinct from that of H1, possibly playing a role in the repression of the erythrocyte genome (Seligy and Neelin 1970, Billet and Hindley 1972, Medvedev and Kirpicheva 1972, Seligy et al. 1973, Seligy and Miyagi 1974, Tobin and Seligy 1975, Sung et al. 1977, Gasaryan 1978). Other characteristics that distinguish chicken H5 from H1 include polyadenylated mRNA (Molgaard et al. 1980) and the absence of core histone genes linked to the chicken H5 gene (Kreig et al. 1983, Ruiz-Carrillo 1983). Avian H5 mRNA has also been found to lack the highly conserved 3' untranslated nucleotide sequences of the H1 and core mRNAs (Doenecke and Tonjes 1984, Coles and Wells 1985). Despite extensive study however, the precise function of avian H5 has not been elucidated, nor is the significance of its unusual metabolic characteristics clearly understood.

Another linker histone that is distinct from vertebrate H1 is mammalian H1^o, which has been shown to have extensive homologies with avian H5. Specifically, H1^o synthesis is independent of DNA replication (Chabanas et al. 1983, D'Anna et al. 1985) and is insensitive to hydroxyurea treatment (Zlatanova 1980). H1^o has also been reported to increase in amount during reduced proliferative activity in a variety of cultured mammalian cells (Keppel et al. 1977, Keppel et al. 1979, Pehrson

et al. 1980, D'Anna et al. 1980a, D'Anna et al. 1980b, D'Anna et al. 1981, Zlatanova et al. 1980, Gjerst et al. 1981, Pieler et al. 1981) and to accumulate in chromatin during tissue differentiation and maturation (Marsh and Fitzgerald 1973, Garrard and Bonner 1974, Varricchio 1977, Gjerst et al. 1982, Lennox and Cohen 1983, Pina et al. 1984, Gabrielli et al. 1984, Stambolova et al. 1984, D'Anna et al. 1985), similar to the accumulation of H5 observed during avian erythropoiesis. Mammalian H1⁰ also has been found to possess extensive amino acid sequence homology with avian H5 (Smith et al. 1980b, Pehrson and Cole 1981, Doenecke and Tonjes 1986), consistent with their reported homology in tertiary conformation (Cary et al. 1981) and antigenicity (Mura and Stollar 1981, Allan et al. 1982). Like avian H5, human H1⁰ also lacks the conserved 3' untranslated nucleotide sequences present in H1 and core mRNA (Doenecke and Tonjes 1986). The extensive homologies between mammalian H1⁰ and avian H5 has led to the generally held belief that these linker histones have closely related functions.

Despite their similarities, mammalian H1⁰ and avian H5 may still differ significantly in function (Smith et al. 1981, Smith et al. 1984a, Moorman and de Boer 1985). This proposal is based in part upon the supposed erythrocyte-specificity of avian H5, and its putative role in the transcriptional inactivation of the erythrocyte nucleus (see above). H1⁰, with its broad tissue-distribution, could have a more generalized function associated with non-erythroid tissues, and enucleation could represent an "H5-independent" mechanism for nuclear inactivation during mammalian erythropoiesis. In addition to other differences in amino acid sequence, avian H5 also contains larger amounts of arginine as compared to H1⁰, which could be related to differences in their function (Doenecke and Jonjes 1986).

The characterization of an H1⁰/H5-like histone in amphibians (called H1^S by Brown et al. 1981, Rutledge 1982, Rutledge et al. 1984; H1D and H1E by Risley and Eckhardt 1981; H5 by Destree et al. 1979; H1⁰-like by Shimada et al. 1981; H1⁰/H5 by Moorman and de Boer 1985) provided an important opportunity to further investigate the relationship between H1⁰ and avian H5. Specifically, amphibian H1^S is not erythrocyte-specific, but has a broad tissue-distribution similar to that of H1⁰ (Brown et al. 1981, Risley and Eckhardt 1981, Moorman and de Boer 1985). Amphibian H1^S and mammalian H1⁰ are also similar in size and amino acid composition (Brown et al. 1981, Shimada et al. 1981).

In addition to its similarities with H1⁰, amphibian H1^S is also present in erythrocytes. This allowed the study of amphibian H1^S metabolism during erythropoiesis, which demonstrated that Xenopus laevis H1^S possesses the same distinctive metabolic characteristics as avian H5 (Rutledge 1982, Rutledge et al. 1984). Specifically, it was shown that Xenopus laevis H1^S is selectively synthesized in the absence of DNA replication during late stages of erythrocyte maturation, and accumulates three-fold relative to H1. It was concluded from this evidence that the Xenopus laevis H1^S and avian H5 possess closely related if not homologous functions, and the presence of Xenopus laevis H1^S in non-erythroid tissues implies that the role played by Xenopus laevis H1^S or avian H5 is not limited to nucleated erythrocytes. Indeed, because of the extensive homologies between Xenopus laevis H1^S and avian H5, which have been further substantiated by the evidence that will be presented in this thesis, Xenopus laevis H1^S will be subsequently referred to as Xenopus laevis H5.

The metabolic characterization of Xenopus laevis H5 thus provided an important link between H1⁰ and avian H5, and more clearly delineated the

homologies shared by these linker histones by reducing the supposed distinctions between avian H5 and mammalian H1^o. It also supported the proposal that the separate classification of avian H5, mammalian H1^o and amphibian H5 is not justified, and that these proteins belong to a single subclass of linker histone (Pehrson and Cole 1980, Rutledge 1982, Rutledge et al. 1984, Smith et al. 1984a). A more detailed characterization of Xenopus laevis H5 was required however, to further substantiate this proposal.

The major objective of this project was thus to investigate further the homologies between Xenopus laevis H5, avian H5 and mammalian H1^o, and to more precisely characterize the putative functional and evolutionary relationships between these linker histones. This was accomplished by the isolation and sequence characterization of cDNA clones that encode for the two Xenopus laevis H5 variants. A detailed examination of the genetic and sequence homologies between Xenopus laevis H5, avian H5 and mammalian H1^o will subsequently show that these linker histones are truly homologous proteins, and that they are indeed representatives of a distinct subclass of vertebrate linker histone.

II. Materials and Methods

A) Preparation and Characterization of RNA

1) RNA Extraction

RNA was obtained from erythrocytes of female Xenopus laevis (Boreal Laboratories, Mississauga, Ont.) made anemic by two subcutaneous injections of phenylhydrazine (12.5 mg/ml phenylhydrazine HCl, 8.5 mM NaOH) at a dosage of 30 mg/Kg body weight (Thomas and Maclean 1975, Chegini et al. 1979, Rutledge 1982). Eight days after treatment, the animals were anesthetized by a 30 minute immersion in 0.2% MS222 (Sigma Chemical Co., St. Louis, Mo.), and the blood and liver was collected. Blood at this stage of treatment was previously shown to contain actively dividing, immature erythrocytes (late erythroblast/early polychromatic stage), that synthesize large amounts of histone protein (Rutledge 1982, Rutledge et al. 1984). RNA was extracted with guanidinium isothiocyanate (Chirgwin et al. 1979) and further purified by sedimentation through CsCl (Glisin et al. 1974, Chirgwin et al. 1979).

11) RNA Fractionation

Enrichment for H5 mRNA (7-17S) was achieved by sedimentation of total RNA through 5-30% isokinetic sucrose gradients in 20 mM Tris pH 8.0, 0.1 M NaCl, 0.25% SDS, 1 mM EDTA. Specifically, 30 ml isokinetic gradients were prepared for a SW27 rotor, using an ISCO gradient former set for 1/3 5% sucrose and 2/3 30% sucrose. RNA samples (3-5 mg) dissolved in buffer were layered onto the top of each gradient and centrifuged at 25,000 RPM, 20°C for 18 hrs. RNA profiles and collection of 1.2 ml fractions were taken simultaneously using an ISCO

gradient fractionator equipped with a UV-5 monitor. The appropriate fractions were pooled, and the RNA was precipitated with 2 vols of ethanol after addition of 1/10 vol 2M potassium acetate, and stored at -20°C . Poly(A)⁺ RNA was isolated from liver RNA by oligo(dT)-cellulose chromatography as described by Maniatis (1982).

iii) In vitro translation of mRNA

Cell free translation of mRNA was carried out using a rabbit reticulocyte cell-free system (lysine and alanine free, BRL) according to the manufacturer's protocol, using 1.0 uCi/ul (³H)-lysine (86 Ci/mmol, Amersham) and 1.0 uCi/ul (³H)-alanine (36 Ci/mmol, Amersham). Reaction products were analyzed using polyacrylamide gel electrophoresis in sodium dodecyl sulfate with a 12.5% acrylamide separating gel and a 7.5% stacking gel. Fluorography was done according to Chamberlain (1979). After fixing in 5% acetic acid, 20% ethanol, the gels were washed in distilled water for 30 min, immersed in 1.0 litre of 1.0 M sodium salicylate (Sigma) pH 6.3 for 30 min and dried onto Whatman 3MM. The dried gel was then exposed to Kodak X-Omat XAR-5 film for 7 days at -80°C .

B) cDNA Synthesis

Double-stranded cDNA was synthesized using a modified Gubler-Hoffman procedure (Gubler and Hoffman 1983), which entailed elimination of a phenol extraction and an ethanol precipitation step before second strand synthesis, producing a continuous two-step reaction. The first step involved a standard reverse transcriptase reaction (Maniatis 1982) in which 10-50 ug of RNA in 5 ul of H₂O and 1.0 ul 100 mM methylmercury hydroxide were mixed together in a 250 ul Eppendorf tube and incubated

at room temperature for 10 minutes. 2.0 ul 700 mM mercaptoethanol was then added and the incubation continued for 5 minutes. 5.0 ul oligo(dT)₁₆₋₁₈ (2 ug/ul), 5.0 ul 5X cDNA Buffer (200 mM Tris pH 7.5, 25 mM MgCl₂, 100 mM KCl, 25 mM dithiothreitol), 2.5 ul 10 mM dNTP (mixture of all four deoxyribonucleotide triphosphates), 1.0 ul ³²P-dATP (5-10 uCi/ul, Amersham), 2.0 ul H₂O, and 1.5 ul (40 units) reverse transcriptase (Life Sciences) were added and the reaction incubated for 30 minutes at 37°C (Berger et al: 1983). One ul of the reaction mixture was then spotted onto a glass filter for later determination of dNTP incorporation (see below).

The second phase of the reaction entailed conversion of the RNA/cDNA hybrids into double-stranded cDNA using RNase H and DNA polymerase (Gubler and Hoffman 1983). The reaction volume was adjusted to 100 ul with 1X cDNA Buffer, and 1 unit RNase H and 20 units E. coli polymerase I were added before incubation at 15°C for 2 hours. After spotting 1.0 ul of the reaction mixture onto a glass filter for later determination of dNTP incorporation, the reaction mixture was extracted 3 times with two volumes phenol/chloroform (1:1), and the mixture was passed through two G-50 "spin columns" in STE (10 mM Tris pH 8.0, 1 mM EDTA, 100 mM NaCl) to remove unincorporated nucleotide triphosphates (Maniatis 1982). The double-stranded, blunt-ended cDNA (dscDNA) was then precipitated with 2 volumes of ethanol and stored at -20°C.

After pelleting, the dscDNA was dissolved in sterile water and used directly for blunt-end ligation to vector DNA (see below). Profiles of the dscDNA were obtained by electrophoresing the ³²P-labelled dscDNA (approx. 3000 CPM per slot) in a 1.0% agarose gel in 40 mM Tris-acetate (pH 8), 2 mM EDTA (TAE) as described by Maniatis (1982). After fixing.

in 5% TCA, the gel was dried onto Whatman 3M without heat, and autoradiographed overnight at -80°C .

Note that the ribonuclease inhibitor, ribonucleoside-vanadyl complex was not used in the preparation of dscDNA, because it was found to be responsible for failed blunt-end ligation of dscDNA into M13mp9, apparently due to its inhibitory effects upon T4 DNA ligase (Puskas et al. 1982).

To determine the amount of cDNA synthesized during first and second strand synthesis, the incorporation of dNTP was determined using a procedure similar to that described by Berger (1984). Specifically, a 1.0 μl sample of reaction mixture was spotted onto a glass filter (Whatman GFC), and dried under a heat lamp. After placing the filter into a scintillation vial, its radioactivity was determined using Cerenkov counting (Kellogg 1983). The filter was then washed with 5% trichloroacetic acid, 20 mM pyrophosphate, followed by ethanol, before drying for 10 minutes under a heat lamp, and the radioactivity determined again using Cerenkov counting. The ratio of TCA precipitable to total CPM determined the fraction of dNTP incorporated into DNA, enabling the amount of cDNA synthesized in each reaction to be estimated. This method was useful for confirming the efficiency of cDNA synthesis, and for quantitating the amount of total dscDNA recovered from the reaction. This assay was also used in the preparation of probes from M13 template DNA (see later section).

C) Construction and Screening of M13mp9 cDNA Banks

1) M13mp9 cDNA Bank Construction

Cloning of the dscDNA was accomplished by blunt-end ligation of 10 ng of dscDNA with 100 ng of Hinc II-cleaved, calf intestinal

phosphatase-treated (Maniatis 1982) M13mp9 replicative form DNA in a 20 μ l reaction volume containing 10 mM Tris pH 7.5, 5mM $MgCl_2$, 5mM DTT, 1 mM ATP, and 3.0 units T4 DNA ligase (BRL). After incubating at room temperature for 6-18 hours, two μ l of the ligation mixture was used to transform 200 μ l of competent E. coli JM103 cells (Hannahan 1983, Messing 1983). Based on a transformation efficiency of 10,000 clones/ng supercoiled, replicative form M13mp9, 500-1000 plaques per ng of ligated double-stranded cDNA was consistently obtained (eight experiments), with 80% of the clones containing cDNA inserts as determined from plaque hybridization to labelled cDNA. After immunological screening (see next section) 5 ml of 10 mM Tris pH 8.0, 1mM $MgCl_2$ was added to each plate, and the top agarose scraped from each plate before centrifugation at 15,000 RPM for 30 minutes. The supernatant which contained the recombinant phage, was then collected and stored at 4°C.

ii) Immunological Screening of M13mp9 cDNA Banks Using a Polyclonal H5-antibody

Transformed cells were plated (Messing 1983) to a density of 1000-2000 plaques per plate (9X9 cm). "Lifts" were taken by laying nitrocellulose (8X8 cm, 0.45 μ m, Schleicher and Schuell) onto each plate, and a replica of the cells taken by lifting the nitrocellulose from one corner. The "replica" cells were then amplified by placing the lifts cell-side up onto a fresh B-plate and incubating for 4-6 hrs at 37°C. To induce expression of the cDNA inserts (Slocombe et al. 1982), the lifts were transferred to B-plates containing 0.5 mM IPTG, followed by incubation overnight at 37°C. The cells were then lysed by exposing the lifts to a chloroform-saturated atmosphere for 15-30 minutes. After drying at room temperature, the lifts were either stored at -80°C or

immediately washed extensively with blocking solution (1% gelatin, 50 mM Tris pH 8.0, 200 mM NaCl), until they were free of cell debris.

A polyclonal Xenopus laevis anti-H5 antibody raised in guinea pig was generously supplied by P. Foster and J. Neelin along with samples of purified Xenopus laevis H1 and H5 protein. The selectivity and sensitivity of the antibody was tested by spotting onto nitrocellulose strips 1, 10, and 100 ng of purified H1 and H5 histone dissolved in water. Using the immunological screening procedure described below, less than 1 ng of H5 but no less than 10 ng of H1 was detected by the H5-antibody.

To reduce cross-reaction to E. coli proteins, 15 ml of H5-antibody diluted in blocking solution, was preabsorbed with 1.0 ml of JM103 lysate, for 60 minutes at room temperature. The lysate was prepared by harvesting E. coli JM103 from a 100 ml overnight culture by centrifuging at 2000 RPM for 5 minutes. After resuspension in 10 ml of 50 mM Tris pH 8.0, 200 mM NaCl (TBS), the cells were boiled for 10 minutes, cell debris removed by centrifugation at 15,000 RPM for 10 minutes, and one ml aliquots of the supernatant stored at -80°C.

The method for immunological screening was based upon the procedure described by Hawkes et al. (1982). Specifically, 5.0 ml of a 1/200 dilution of anti-H5 antibody preadsorbed to JM103 lysate, was added to each washed lift and incubated at room temperature for 2-4 hrs. After extensive washes with TBS and incubation for 15 minutes in blocking solution, 5.0 ml of a 1/500 dilution of a horse radish peroxidase-linked, goat anti-guinea pig immunoglobulin antibody (Kirkegaard and Perry Laboratories Inc.) in blocking solution was added to each lift. Following incubation for 2-4 hrs, the lifts were washed extensively with TBS, and each was immersed in 10 ml of 0.01% H₂O₂, 0.5 mg/ml 4-dichloro-

1-naphthol (BioRad). After 15-30 minutes, the lifts were rinsed in water and dried at room temperature. This produced intense, purple colored spots corresponding to plaques expressing putative H5 peptide.

cDNA banks constructed from total RNA, S1 RNA and S2 RNA (each containing about 10,000 clones) were screened using this procedure, which resulted in identification of two putative H5 cDNA clones from the S2 cDNA bank. These clones were plaque purified using the same immunological screening procedure, and positive identification made by DNA sequencing (see Results section).

111) Rescreening of the S2 Bank Using Probes from XLH5A-1.

An aliquot containing about 45,000 plaque forming units was taken from the S2 bank (collected after immunological screening as described previously), and mixed with 30 ml of 0.6% agarose in B-media and 1.0 ml of early log-phase JM103, before pouring onto a baking dish (16X30 cm) containing bottom agar (300 ml of 1.5% agar in B-media; Maniatis 1982). After incubating overnight at 37°C, a lift (15.5X24.0 cm) was made using nylon membrane (Biodyne A, Pall) and the plaques amplified by incubating the lift for 9 hrs. at 37°C after being placed cell-side up onto a fresh B-plate. The lift was then placed onto Whatman 3MM paper saturated with 0.4 M NaOH, 1.5 M NaCl for 10 minutes, followed by a 10 minute neutralization step in 1.0 M Tris pH 8.0 (Messing 1983). After drying at 80°C for 30 minutes, the cell debris was removed from the lift by extensive washes with 2XSSC (20XSSC = 3.0 M NaCl, 0.3 M Na citrate), 1.0% SDS. Following an overnight hybridization at 65°C with a strand-specific probe prepared from XLH5A-1 (described in a later section), the lift was washed five times at room temperature with 300 ml 2XSSC, 1.0% SDS, before being dried and autoradiographed overnight. Note that

the combination of a strand-specific probe and single-stranded M13 viral DNA allowed selection of H5 clones in one specific orientation. This produced 30 positive cDNA clones which were plaque purified by two rounds of rescreening using the same hybridization procedure.

Before DNA sequencing, clones containing identical cDNA inserts were identified using the "T-tracking" technique described by Messing (1983), which entailed partial DNA sequencing of the 5' terminus of each cDNA insert. This was accomplished without subcloning of insert DNA because each clone had an orientation such that the 5' end of the cDNA insert was adjacent to the universal primer used for DNA sequencing in M13 (Messing 1983). A single "T" sequencing reaction was thus conducted on each of the 30 purified clones, and their "T" patterns compared. Those clones containing identical inserts produced identical patterns, allowing the identification of 11 unique cDNA clones.

D) DNA Probe Preparation

1) Strand-specific Probes for Plaque Hybridization

The ability to use the single-stranded viral DNA of M13mp as a template for DNA polymerase (Messing 1983), allowed the preparation of strand-specific probes needed for selecting M13mp9 cDNA clones with a specific orientation (see previous section). The method involves: 1) synthesis of a DNA strand through the insert DNA using a universal primer, E. coli Pol I Klenow fragment, and ^{32}P -dNTPs; 2) excision of the double-stranded insert DNA from the template DNA with restriction enzymes; 3) purification of insert DNA using preparative electrophoresis. Note that although the insert DNA is double stranded, only the strand complementary to the template DNA is labelled. Specifically, 50-100 μCi each of ^{32}P -dATP, -dCTP and -TTP (800 Ci/mmol,

Amersham or ICN) were placed into a 250 μ l Eppendorf tube and dried in vacuo. Two μ l of XLH5A-1 template DNA (0.5 μ g/ μ l), 2.0 μ l 15mer DNA sequencing primer (BRL) and 1.0 μ l H-buffer (100 mM Tris pH7.5, 50 mM $MgCl_2$, 100mM DTT) were then added to the tube containing the dried isotope, and the mixture was incubated at 50°C for 10 minutes. To this mixture 3.0 μ l of 100 μ M dNTP (equal mixture of dATP, dCTP, dGTP, dTTP), 2.0 μ l H_2O , and 1.0 μ l Klenow fragment (0.5 units/ μ l) were added and the reaction incubated for 1 hr. at 37°C, before supplementing with 1.0 μ l of 10 mM dNTP and incubating for an additional 30 minutes at 37°C.

Excision of insert DNA was carried out by addition of 2.5 μ l 10X restriction buffer (Maniatis 1982), 9.5 μ l H_2O and 1.0 μ l (10 units/ μ l) each of EcoRI and PstI were then added and incubated for an additional 60 minutes. Note that the insert of XLH5A-1 does not contain an EcoRI or a PstI restriction site. The reaction was stopped by addition of 75 μ l STE (10 mM Tris pH 8.0, 1 mM EDTA, 100 mM NaCl) and 1.0 μ l was spotted onto a glass filter for determination of total dNTP incorporated (described previously). The reaction mixture was then twice passed through a G-50 "spin column" (Maniatis 1982) to remove unincorporated dNTP's, before adding electrophoresis loading buffer.

After loading onto a 1.5% gel in TAE (Maniatis 1982), the sample was electrophoresed until the bromophenol blue dye had migrated 4-6 cm. The gel was covered with plastic wrap and a strip of X-ray film placed on top of the gel for 1-5 minutes. After developing the X-ray film, the radioactive band corresponding to the insert fragment was located and the DNA prepared using the DEAE-cellulose method described by Dretzen et al. (1981). A strip of DEAE-cellulose (NA45, 0.45 μ M, Schleicher and Schuell) was placed into a slot cut in front of the band and electrophoresis continued for 5 minutes. The DEAE strip was then placed

into 20 ml of hybridization buffer (1.0 M NaCl, 1.0% SDS, 50 ug/ml salmon sperm DNA, 50 ug/ml polyA, 1 ug/ml M13mp9 viral DNA) and heated in a boiling water bath for 15-30 minutes before hybridization to plaque lifts.

ii) Single-stranded Probes for DNA and RNA Blot Hybridizations

Probes for DNA and RNA blots were prepared from XLH5A-1 template using the same approach described above for plaque hybridization probes. Modifications to the procedure were: 1) reduction of the dNTP concentration in the DNA polymerase reaction to 5-10 uM, to increase the specific-activity of the probe to $>10^9$ DPM/ug, which unfortunately reduced probe stability such that after 48 hrs. the probe becomes fragmented (Burke et al. 1984), dramatically reducing the efficiency of hybridization; 2) eliminating addition of 1mM dNTP; 3) using only PstI to cleave the template DNA; 4) immediately before electrophoresis, addition of 1.0 M NaOH to 200 mM and incubation for 2 minutes at room temperature before neutralization with 1.0 M HCl and addition of electrophoresis buffer.

E) DNA Isolation from Xenopus laevis Erythrocytes

Approximately 1.0 ml of blood was collected from a single animal and the cells were washed once in 0.9% NaCl, followed by resuspension into 10 ml 0.9% NaCl. The cells were lysed by addition of 200 ul 5% saponin and the mixture was stirred at room temperature for 5-10 minutes before pelleting the nuclei by centrifuging at 10,000 RPM for 10 minutes. DNA isolation was according to Maniatis (1982). After resuspending the nuclei in 50 ml 0.9% NaCl, 50 ml of 1.0 M EDTA was added with stirring, followed by the addition of 10 mg Proteinase K (Sigma) in 5.0 ml 0.9%

NaCl and 1.5 ml 30% Sarkosyl (Sigma). The mixture was then incubated at 50°C for 3-24 hours and 20 ml aliquots gently extracted twice with phenol followed by two extractions with phenol/chloroform (1:1). After dialysis overnight against H₂O, the DNA was stored at 4°C and used directly in restriction digests.

F) DNA and RNA Blot Analysis

10-50 ug aliquots of genomic DNA were digested with restriction enzymes according to the manufacturer's protocol (BRL) at an enzyme concentration of 4 units/ug of DNA. 10 ug aliquots of restricted DNA were electrophoresed in 0.8% agarose gel in TAE (Maniatis 1982), 0.5 ug/ml ethidium bromide. After photographing, the DNA was blotted onto a nylon membrane (Gene Screen-Plus, NEN) using 0.4 M NaOH, 1.5 M NaCl (Chomczynski and Qasba 1984; see also Reed and Mann 1985). The blot was prehybridized in 20 ml of 1.0% SDS, 1.0 M NaCl, 50 ug/ml yeast tRNA, at 65°C before a 48 hour hybridization at 65°C in the same buffer containing a high specific-activity (approx. 10⁹ DPM/ug), single-stranded probe, prepared by primer-extension using XLH5A-1 template DNA (described previously). Unhybridized probe was removed by sequential washes in 200 ml 0.1XSSC, 1% SDS at room temperature for 10 minutes, two 200 ml of 2XSSC, 1% SDS at 65°C for 15 minutes, and one 200 ml of 0.1XSSC at room temperature for 30 minutes. The DNA blot was then dried and was autoradiographed for 14 days at -80°C.

10 ug of total erythrocyte RNA, poly(A)⁺ and poly(A)⁻ liver RNA was glyoxalated and electrophoresed as described by Thomas (1983), and blotted onto nylon membrane (Blodyne A, Pall) using 20XSSC. The blot was prehybridized in 20 ml of 1.0% SDS, 1.0 M NaCl, 50 ug/ml yeast tRNA, 50 ug/ml poly(A) before a 24 hour hybridization at 65°C in the same

buffer containing a high-specific activity, single-stranded probe prepared from XLH5A-1 as described previously. Unhybridized probe was removed using the same procedure as described for DNA blots, the blot was autoradiographed at -80°C for 4 days.

G) DNA Sequencing

Nucleotide sequences of the H5 cDNA inserts were determined by the dideoxyribonucleotide method as described by Messing (1983), using ^{35}S -dATP (Amersham) and gradient sequencing gels as described by Biggin et al. (1983). cDNA inserts larger than 400 bp were sequenced using the "deletion banking" method described by Dale et al. (1985). Nucleotide sequence of the opposite strands of XLH5A-1 and XLH5A-2 were determined by subcloning their inserts into the opposite orientation, as described below. The sequencing strategy used for each cDNA clone is summarized in Figure 1.

Subcloning of the cDNA inserts from XLH5A-1 and -2 directly from single-stranded viral DNA was accomplished through primer extension using a standard DNA polymerase reaction. Specifically, 1.0 ug of template was primed with 4 ng of 17mer universal sequencing primer (BRL) and the primer extended with 1.5 units of *E. coli* Pol I Klenow fragment, 1 mM dNTPs, and 1X H-Buffer in a volume of 10 ul for 60 minutes at 37°C . The partially double-stranded template was then digested with the appropriate restriction enzymes, electrophoresed in a 1.0% agarose gel in TAE (Maniatis 1982), and fragments of the insert DNA were prepared using the DEAE-cellulose method (Dretzen et al. 1981). The isolated DNA fragments were then ligated into the appropriate M13 vector, and used to transform competent JM101 or JM103 cells (Hannahan 1983).

Figure 1. Schematic of the strategy for nucleotide sequence determination of the M13mp9-H5 cDNA clones. A total of 13 cDNA clones were isolated from the S2 erythrocyte cDNA-M13mp9 bank using immunological screening and nucleic acid hybridization. DNA sequence was determined using the dideoxyribonucleotide method (Messing 1983) and the "deletion banking" technique of Dale et al. (1985). The scales are in nucleotides starting at the 3' terminus for each cDNA insert, with the arrowed lines above each clone representing the segments sequenced from each deletion clone. A) cDNA clones whose 3' termini correspond to an A-rich region in the middle of the coding region (see Results for details). B) cDNA clones whose 3' termini correspond to the putative polyadenylation site.

A)

B)

H) Computer Sequence Alignment

Computer alignment of DNA sequences was done using the BIOCOMP system provided by Dr. M. Zuker and Ms. L. Brummel of the Biomath Section of the National Research Council of Canada.

III. RESULTS

A) Cloning of Xenopus laevis H5 cDNA

1) RNA Fractionation and M13mp9 cDNA Bank Construction

Immature erythrocytes of anemic Xenopus laevis which had previously been shown to actively synthesize histone protein (Rutledge 1982, Rutledge et al. 1984) provided a rich source of histone mRNA. Although in vitro translation of this RNA produced proteins identified as H1 and core histones (Figure 2B), a product clearly identifiable as H5 was never observed, even after size enrichment. The reason for this was never established, but was assumed to reflect the low abundance of H5 mRNA. On the assumption that it is similar in size to H1 mRNA (about 12S), tentative enrichment of H5 mRNA was conducted using sucrose gradient sedimentation of total RNA, as shown in Figure 2.

Double-stranded cDNA (dscDNA) was synthesized from the fractionated RNA using a modification of the Gubler-Hoffman procedure (Figure 3A; see Methods and Discussion). Oligo(dT)₁₆₋₁₈ was used as the primer for first-strand synthesis on the assumption that like the chicken H5 mRNA, Xenopus laevis H5 mRNA is polyadenylated. The electrophoretic profiles of this dscDNA shown in Figure 3B, demonstrates the effectiveness of the mRNA size fractionation, and illustrates the feasibility of synthesizing oligo(dT)-primed dscDNA from samples that contain large amounts of rRNA. cDNA banks were then constructed by blunt-end ligation of the S1 and S2 dscDNA into the Hinc II site of M13mp9.

Figure 2. In vitro translation and sucrose gradient fractionation of total RNA isolated from immature Xenopus laevis erythrocytes.

A) Initial sedimentation of total erythrocyte RNA through a 5-30% isokinetic sucrose gradient. RNA sedimenting between 5S and 18S (fraction b) was pooled from several gradients and called S1.

Resedimentation of the 18S rRNA was carried out and the leading edge of the 18S peak was pooled (fraction i) and called S2. B) In vitro translation analysis of total RNA (T) and various RNA fraction from the sucrose gradients (a-j in A)). Electrophoretic profiles of ³H-lysine/alanine labelled peptides from rabbit reticulocyte lysate in vitro translation confirmed the presence of histone mRNA. "m" is total histone isolated from X. laevis erythrocytes labelled with ¹⁴C-lysine (Rutledge 1982, Rutledge et al. 1984).

Figure 3. Double-stranded cDNA synthesized from fractionated erythrocyte RNA. A) Preparation of double-stranded, blunt-ended cDNA using a modification of the Gubler-Hoffman procedure (see Methods for details). B) Autoradiographic profiles of ^{32}P -labelled dscDNA electrophoresed in a 1.4% agarose Tris-acetate gel along with end-labelled markers (pBR322 digested with HinfI (a) or TaqI (e)). b: double-stranded dscDNA synthesized from poly(A⁺) RNA isolated from the liver of anemic *Xenopus*. c: and d: dscDNA synthesized from sucrose gradient fractionated RNA S1 and S2 respectively (see Figure 2).

11) Immunological Screening of the M13mp9 cDNA Banks

One of the many advantages of using M13mp as a primary cloning vector (see Discussion) is the ability to exploit it as an expression vector (Slocombe et al. 1982). As described in Methods, lifts were taken from the S1 and S2 M13mp9 cDNA banks, treated with IPTG and the expression of H5 peptide detected with a Xenopus laevis H5 polyclonal antibody probe. This procedure, which is illustrated in Figure 4, led to the identification and purification of two antigenically positive clones, XLH5A-1 and XLH5A-2, from the S2 cDNA bank.

DNA sequencing confirmed that both cDNA inserts encoded for Xenopus laevis H5 (evidence presented in a later section) and that they were in the proper orientation and reading frame for H5 peptide production, (Figure 5A). As outlined in Figure 5B, the 5' ends of both cDNA molecules were truncated 3' to the start codon and thus provided only a partial amino acid sequence for Xenopus laevis H5. The 770 base XLH5A-2 cDNA insert consisted of a 570 nt region coding for 189 amino acids of H5 and a 200 nt 3' untranslated region ending with a 6mer poly(A) tail. XLH5A-1 contained 295 bases coding for 98 amino acids that agreed exactly with the corresponding DNA sequence of XLH5A-2, excluding a 3' 17mer poly(A) track which apparently was formed by initiation of reverse transcription within the middle of the coding region. This was most likely a result of oligo(dT) hybridization to an adenine-rich region (15 of 18 bases adenine), and will be discussed in more detail in a later section.

Figure 4. Schematic of H5 cDNA isolation using immunological screening. dscDNA blunt-end ligated into the HincII site of M13mp9 (A) was used to transform JM103 (B), which were plated at a density of 1000-2000 plaques per plate (9X9 cm) in the absence of IPTG. After incubating overnight at 37°C, nitrocellulose lifts (C) were taken and the cells amplified in the absence of IPTG by placing the lifts cell-side up onto a fresh plate (D) which were incubated for an additional 6 hrs. Expression of insert DNA was then induced by transferring the lifts to plates containing 20 mM IPTG (E) and an overnight incubation at 37°C. After cell lysis with chloroform (F), the lifts were sequentially treated with an H5-antibody (G), horse radish peroxidase linked 2° antibody (H) and a chromogenic-hydrogen peroxide reagent (see Methods for details). Two positive plaques (XLH5A-1 and XLH5A-2) were identified (I and J) and plaque purified by a second round of antibody screening (K) using the same procedure (B-J). Single-stranded viral DNA was then prepared and positive identity of the cDNA inserts obtained by DNA sequencing.

A)

M T M I T P S L A A G P A K P
 ATG ACC ATG ATT ACG CCA AGC TTG GCT GCA GGT C-CT GCC AAA CCC...
 ATG ACC ATG ATT ACG CCA AGC TTG GCT GCA GGT C-CT ACT CCA GCT GCC AAA
 =====> P T P A A K
 M13mp9 Sequences (770) XLH5A-2

B)

Figure 5. Characterization of XLH5A-1 and XLH5A-2. Screening of the S2 bank led to the isolation of two M13mp9 cDNA clones (XLH5A-1 and XLH5A-2) that produced antigenically active H5 peptides upon induction with IPTG. A) Nucleotide sequence of the fusion site between M13mp9 and the two H5 cDNA inserts. DNA sequencing demonstrated that both cDNA inserts began within the coding region of the H5 mRNA, and confirmed that both H5 cDNA inserts were in the proper orientation and reading frame for expression of H5 peptides. Amino acids are shown above the corresponding codons with a dash denoting the site of ligation between M13mp9 and the H5 cDNA inserts. As predicted by the insertion of cDNA into the HincII site of M13mp9, expression would have produced a hybrid protein that contained 11 M13mp9 encoded amino acids fused to the amino termini of the XLH5A-1 and XLH5A-2 encoded H5 peptides. B) Schematic outlining the structure of the XLH5A-1 and XLH5A-2 cDNA inserts. The bottom scale is in nucleotides starting at the putative polyadenylation site, and both cDNA inserts end before the start codon.

iii) Rescreening of the S2 cDNA Bank with XLH5A-1

In an attempt to isolate larger, more complete H5 cDNA clones, the S2 cDNA bank was rescreened using primer-extension probes synthesized from the template DNA of XLH5A-1. This provided several important advantages. First, XLH5A-1 should be highly selective for Xenopus laevis H5 cDNA clones, due to the absence of 5' or 3' untranslated regions which could contain sequences homologous to other cDNA clones. Second, the XLH5A-1 probe would select clones that contain sequences 5' to the adenine-rich region, thus circumventing the isolation of clones containing only the 3' portion of the H5 transcript. Finally, preparation of strand-specific probes by primer-extension (see Methods for details) would permit the selection of cDNA clones with a specific orientation, such that the 5' end of each cDNA insert would be adjacent to the site of the sequencing primer. This greatly facilitated the rapid analysis of the 5' sequences within a large number of M13mp9 cDNA clones, by circumventing the need for subcloning of the cDNA inserts. This strategy led to the isolation of 30 H5 cDNA clones, from which 11 unique clones were identified using "T-tracking" as a screening technique (see Methods for details).

B) Xenopus laevis H5 cDNA Sequence Analysis

i) Preliminary Characterization of the H5 cDNA Inserts

The use of M13mp9 as the primary cloning vector, combined with the deletion banking technique recently developed by Dale et al. (1985), allowed the rapid DNA sequencing of all 13 H5 cDNA inserts whose structures are outlined in Figure 6. All but one of the newly isolated cDNA clones were homologous to XLH5A-1 and XLH5A-2. Details about

Figure 6. Schematic of the Xenopus laevis H5 cDNA inserts. Rescreening of the S2 bank with a strand-specific XLH5A-1 probe (corresponding to nucleotides 466-761 and denoted by the heavy arrowed line) led to the isolation of eleven additional H5 cDNA clones, all with an orientation opposite to XLH5A-1 and -2. As expected XLH5A-1 selected only cDNA clones which include sequences 5' to nt 466. The scale is in nucleotides starting at the putative 3' polyadenylation site, with the numbers at the 5' ends of each clone corresponding to the nucleotide position at which each cDNA molecule terminates. The 3' termini of clones shown above the scale all correspond to the adenine-rich region ending at nt 466 (see Figure 9 for details). "A_n" refers to the poly(A) tail found at the 3' terminus of each cDNA clone, and the T₁₆ refers to a 5' 16mer poly(T) tail within XLH5A-16. "N₃₃" at the 3' end of XLH5A-36 denotes the presence of an unidentified 33 nt segment (see text and Figure 8). The arrow heads at nt 1182 of XLH5A-16 is the last identifiable nucleotide of the H5 cDNA and denotes the approximate site of fusion between the H5 cDNA molecule and a second unidentified cDNA molecule (see text and Figure 8). Each "X" denotes the occurrence of a single-base sequence anomaly and "*" denotes a putative single-base polymorphism within XLH5A-9 and XLH5A-2, as listed in Table 1. A) H5A cDNA clones. B) H5B cDNA clone.

XLH5B-1 which codes for a second variant of H5, will be presented in a later section.

Like XLH5A-1, six of the eleven new clones were found to have been formed by initiation of reverse transcription from the adenine-rich region within the center of the coding region (see Figure 9 for details). All of the remaining clones were found to contain a 200 nt 3' untranslated region, ending with a poly(A) tail which varied in length from 4-15 nt. These results indicate that the Xenopus laevis H5 mRNA is polyadenylated, as has been found for chicken H5 mRNA (Molgaard et al. 1980). Further analysis of RNA by hybridization (Figure 7) also supports this conclusion, since most of the XLH5 transcript is detected in the poly(A)⁺ RNA isolated by oligo(dT)-cellulose fractionation.

ii) Single-Base Sequence Anomalies

Figure 6 also outlines the occurrence of several single base anomalies, recognized through the redundancy produced from the sequencing of overlapping cDNA clones. Specifically, DNA sequence comparisons revealed that some of the H5A cDNA clones contained individual nucleotides which did not match the consensus sequence derived from the other H5A cDNA clones. These anomalies occurred at a frequency of about one per 1,400 nt and are summarized in Table 1.

As is discussed below, some of these sequence anomalies are likely legitimate single-base polymorphisms, while others which modify codons such as L->stp within XLH5A-18 and K->M within XLH5A-23 are likely not. The reported in vitro error of about 10^{-3} for reverse transcriptase (Gopinathan et al. 1979, Fields and Winter 1981), as compared to an in vitro error of 10^{-4} for E. coli DNA polymerase (Weymouth and Loeb 1978), indicates that some of these anomalies likely have been produced by errors that

Figure 7. RNA blot analysis of liver and erythrocyte RNA isolated from anemic Xenopus laevis. RNA was blotted onto a nylon membrane, hybridized to XLH5A-1 and autoradiographed as described in Methods. Tracks: a, total immature erythrocyte RNA; b, poly(A)⁺ liver RNA; c, poly(A)⁻ liver RNA. The arrows denote the relative mobility of 28S and 18S rRNA. Based on its mobility relative to 18S rRNA, the H5 mRNA is estimated to be 1300-1500 bases in length.

TABLE 1: Summary of Xenopus laevis H5 cDNA Single-base Sequence Anomalies Outlined in Figure 5

<u>cDNA CLONE</u>	<u>NUCLEOTIDE POSITION*</u>	<u>BASE ANOMALY**</u>	<u>CODON CONVERSION</u>
H5A-4	50	G->C	----
H5A-23	350	A->T	K->M
H5A-4	390	T->G	P->P
H5A-18	736	T->A	L->stp
H5A-14	960	G->A	----
H5A-16	1029	T->C	----
H5B-1 ⁺	281	C->T	R->W
H5A-2+H5A-9 ⁺⁺	128	G->A	----
H5A-2+H5A-9 ⁺⁺	270	T->A	S->T

* refers to the nucleotide position within the H5A consensus cDNA sequence shown in figure 6

** base left of the arrow is the consensus base derived from the H5 cDNA sequence shown in figure 6 and the base right of the arrow presents the nucleotide found within each respective cDNA clone

+ tentative: based upon an anomalous codon (see text)

++ believed to represent mRNA polymorphisms

occur during reverse transcription. Whatever their cause, it is apparent that there could be a significant sequence inaccuracy inherent to cDNA, and thus either the isolation of genomic sequences, the characterization of overlapping cDNA clones or the direct sequencing of mRNA may be necessary for obtaining accurate mRNA sequences.

The lack of a second H5B cDNA clone did not allow a similar analysis for single base anomalies within XLH5B-1. Some of the single nucleotide differences observed between the H5A and H5B cDNA sequences (Figure 9) could therefore be due to single base anomalies within XLH5B-1. As shown in Table 1, the predicted occurrence of a single tryptophan within H5B was thus concluded to be due to a single base sequence anomaly, since the presence of tryptophan in histone protein is unprecedented.

A second form of single-base anomaly involved XLH5A-2 and XLH5A-9, which differ at the same two nucleotides from the consensus sequence derived from XLH5A-2, -4, and -14 (see Table 1). It is highly unlikely that two identical base substitutions could have occurred during cloning of both XLH5A-2 and XLH5A-9. These two single-base differences are therefore believed to represent polymorphisms, isolated as a result of constructing the S2 cDNA bank from mRNA pooled from several individuals.

iii) Ligation Artifacts

Two artifacts were observed during the DNA sequencing of the H5 cDNA clones, which involve segments of what appear to be non-H5 segments ligated to the ends of XLH5A-36 and XLH5A-16 (Figure 6). A 33 nt segment at the 3' end of XLH5A-36, whose DNA sequence is presented in Figure 8A, is absent in all of the other twelve H5 cDNA clones. It was therefore concluded that XLH5A-36 was formed by the ligation of this

33 nt segment to the 3' end of the H5 cDNA molecule, before or during their ligation into M13mp9.

Analysis of XLH5A-16 was somewhat more complex. Initially it appeared that its H5A cDNA insert contained a 1097 nt 5' untranslated region, which would be exceptionally large as compared to the 150 nt 5' untranslated region found in avian H5 mRNA (Doenecke and Tonjes 1984, Ruiz-Carrillo et al. 1983). The presence of a 16mer T-tail at the 5' end of the XLH5A-16 insert suggested however, that a second cDNA molecule was present in an orientation opposite to the H5A cDNA. As outlined in Figure 8B, analysis of the complementary sequence of this 5' segment did reveal an open reading frame encoding an 111 amino acid polypeptide, and a 128 nt 3' untranslated region with a putative polyadenylation signal (AATTAA) and a 16 nt poly(A) tail. Although this provided support for the presence of two cDNA molecules within XLH5A-16, the site of their fusion could not be identified. This was due to the fact that the H5 cDNA molecule within XLH5A-16 extends further than any of the other H5A cDNA's (Figure 6). Based upon the comparison between XLH5B-1 and XLH5A-16, which demonstrated extensive homology up to the last 5' nucleotide of XLH5B-1 (Figure 9), it was possible to tentatively identify at least 1182 nt of H5A sequence. It is likely however, that the 5' end of the H5A cDNA extends even further, although its exact length cannot be determined at this time.

iv) Identification of Two Distinct Xenopus laevis H5 cDNA's

Figure 9 presents the composite nucleotide sequence of the H5A cDNA, along with the nucleotide sequence XLH5B-1. The very extensive DNA sequence homology shared by these two cDNA's, indicates that they were likely produced by a recent duplication of the H5 gene. The

Figure 9. Nucleotide sequence of the Xenopus laevis H5A and H5B cDNAs.

Nucleotide sequences were determined for each cDNA clone using the strategies outlined in Methods, and are numbered from the putative polyadenylation site of H5A. The H5A sequence (upper sequence) is a compilation from all twelve XLH5A cDNA clones (summarized in Figure 6) and includes a 70 nt 5' segment derived exclusively from XLH5A-16. This segment was included because of the extensive sequence homology of this segment with XLH5B-1. Only sequence differences (lower sequence) are given for H5B, whose sequence was derived exclusively from XLH5B-1. "*" denoting sequence gaps introduced for maximum homology. Amino acids encoded by the H5A cDNA is shown above the DNA sequence and amino acid differences within XLH5B-1 are shown below. The adenine-rich region from which many of the H5A cDNA molecules were initiated (see text and Figure 6) is outlined at nt 453-470. An eight amino acid sequence previously determined from the Xenopus laevis H5 protein (Brown 1982) is shown in lower case above the corresponding sequence (nt 700). The two single-base differences outlined above H5A (nt 128 and 270) likely represent allelic polymorphisms within XLH5A-2 and XLH5A-9 (see Table 1). The "(T)" within the tryptophan codon of H5B (nt 276) is a putative single-base sequence anomaly produced during the cDNA synthesis (see text for details).

+400

A K P K K A A K S P A K A K K P K V A E K K V K
 CCAAGCCCCAAAAAGCTGCAAAGTCTCCTGCTAAGGCCAAGAAACCTAAAGTGGCAGAGAAGAAGGTGAA
 -----C-----A-----

+350+300

K V A K K K P A P S P K K A K K T K T V K A K
 GAAAGTAGCTAAGAAAAAGCCTGCACCTTCTCCCAAAAAGCTAAAAAGACAAAACTGTCAAAGCCAAG
 -----C-C-----G-----C-----G-----
 A P R R

T
 --A--+250

P V R A *S K V K K A K P S K P K A K A S P K K
 CCAGTGAGGGCATCCAAGGTAAAGAAGGCCAAGCCCTCCAAACCAAAGGCAAAGCCAGCCCAAAGAAAT
 -----(T)-----C-----A-----
 (W) A

S G R K K +200+150
 CTGGACGGAAGAAG TAAAGCTCCTGCAGTCTCTGGAGAGCTCCTAACAACCTTGTGTACATAGTTTTATA

--A--
 *

+100

ACCTTCACTTTTTTCCAGTGCAACTTTATAATGATGATCAGCTTCCCTGCAATTCCCAAGACCAATTTGT
 -----A-----

+50

GTATTTGACATTGCCTGCACTAGGGGAATTAAGCAAGATTTAATGGCTTCATTTTATTAGTATAACTTTTG
 -----T-----CT-----*-T--T-T-----

+0
 TGTTT A_(n)
 ---G* A₉

proteins encoded by these cDNA molecules differ in only 11 of 193 amino acids, with H5A containing two additional amino acids close to its amino terminus. As discussed in the Introduction, these two proteins would be defined as linker histone variants. The XLH5B-1 cDNA will therefore be subsequently referred to as H5B.

As shown in Figure 9, confirmation that these two variant cDNA's code for Xenopus laevis H5 was provided by the exact agreement between the sequence predicted by H5A and a sequence of nine amino acid residues obtained from X. laevis erythrocyte H5A protein (Brown 1982). As shown in Table 2, the amino acid compositions of the proteins predicted by the two variant H5 cDNA's also agree reasonably well with the previously determined amino acid composition of H5 purified from Xenopus laevis H5 erythrocytes and liver. Based upon this and other evidence summarized in the Discussion, it can be concluded that the H5A and H5B cDNAs likely encode for the two H5 variants previously observed in the tissues of Xenopus laevis (Brown et al. 1981, Risley and Eckhardt 1981).

Figure 10 presents a DNA blot analysis of Xenopus laevis erythrocyte DNA hybridized to a high specific-activity, single-stranded probe synthesized from XLH5A-1. The multiple banding pattern, consisting of strong and weak intensities, was consistently observed. Close examination revealed that the strong and weak bands corresponded to two distinct gene segments, consistent with the presence of two H5 genes. As shown in Figure 11, comparison between the partial restriction maps of these two putative H5 genes and those of the H5A and H5B cDNA's revealed an exact match, with the weak bands corresponding to the H5B gene. This confirmed the presence of two distinct H5 genes within a single individual of Xenopus laevis, and provided a preliminary characterization of the H5A and H5B genes.

TABLE 2: Comparison of the Amino Acid Contents Predicted from the H5A and H5B cDNAs with that Previously Determined from Purified Xenopus laevis H5

AMINO ACID	H5A*	H5B*	ERYTHROCYTE**	LIVER**
Asx	3.5	4.7	3.8	4.7
Thr	3.5	3.6	3.5	3.8
Ser	10.2	11.4	9.6	10.3
Glx	5.6	5.2	6.5	7.3
Pro	8.2	8.8	8.5	8.3
Gly	4.1	3.6	5.1	5.9
Ala	15.4	15.0	17.3	17.1
Val	6.1	5.2	5.9	5.8
Met	0.5	0.5	0.3	0.3
Ile	3.1	3.1	2.3	2.3
Leu	3.1	2.6	4.6	4.4
Tyr	1.5	1.6	1.3	1.3
Phe	0.5	0.5	0.7	0.7
His	1.0	0.5	0.7	0.6
Lys	29.7	28.5	26.9	24.6
Arg	3.6	5.2	2.8	2.9

* from the amino acid sequence predicted from the cDNA sequences of H5A and H5B (Figure 9)

** from P-150 purified erythrocyte and liver H5 (Brown et al. 1981)

Note: that erythrocyte H5 contains about 90% H5A and 10% H5B, and liver H5 contains about 50% of each of H5A and H5B.

Figure 10. Partial restriction mapping of the Xenopus laevis H5A and H5B genes. A) DNA blot analysis of Xenopus laevis erythrocyte DNA isolated from a single individual hybridized to XLH5A-1 (see Methods for details). Lanes: a, molecular weight markers (lambda digested with HindIII and EcoRI: 21.8, 5.24, 4.21, 3.41; 1.98/1.90, 1.71, 1.32, 0.93, 0.84, 0.58 Kb respectively); b, BamHI; c, BamHI/PstI; d, BamHI/EcoRI; e, EcoRI; f, EcoRI/PstI; g, EcoRI/HindIII; h, HindIII (digestion failed); i, HindIII/BamHI; j, HindIII/PstI; k, PstI. The differences in band intensity observed for the multiple bands in some lanes, correspond to differences in the hybridization signal produced by H5A and H5B genomic fragments. Although the precise origins of the high molecular weight bands observed in lanes d, e and g was not established, they were believed to be partially digested restriction fragments. B) Partial restriction map showing the approximate positioning of the H5A and H5B coding regions. B: BamHI, E: EcoRI, H: HindIII; P: PstI. The scale is in kilobase pairs.

A

a b c d e f g h i j k l

B

Figure 11. Comparison of the cDNA and genomic restriction maps for Xenopus laevis H5A and H5B. Confirmation for the presence of two H5 genes within Xenopus laevis was provided by agreement between the restriction sites predicted from the cDNA sequences of H5A and H5B, and the restriction pattern produced from genomic DNA blots (see Figure 10). The scales are in nucleotides starting from the putative polyadenylation sites, with the numbers above each restriction site referring to their nucleotide position. A) Restriction sites predicted from the cDNA sequences of XLH5A and XLH5B (Figure 9). B) Partial genomic restriction maps of the putative H5A and H5B genes determined from DNA blot analysis (Figure 10). Note that only the first restriction site on each side of the XLH5A-1 probe (shown as the double arrow above the scale) could be determined for each of the restriction enzymes. The presence within both H5 genes, of a PstI site in close proximity to the probe provided a reference point which greatly facilitated the partial restriction mapping of these genes. The H5B genomic fragments also produced weaker hybridization signals than did the H5A genomic fragments (see Figure 10 and text).

A)

B)

C) Sequence Analysis of the H5 and H1^o Histones

As was outlined in the Introduction, several lines of evidence have led to the proposal that amphibian H5, avian H5 and mammalian H1^o are functionally homologous. The isolation and sequencing of the Xenopus laevis H5A and H5B cDNA's provides the opportunity to examine in detail the sequence homologies between these three species of linker histone, and to evaluate further their supposed functional homologies.

1) Amino Acid Sequence Comparisons

Figure 12 presents an alignment of all available H5 and H1^o amino acid sequences, from which it is apparent that the sites of length variation are non-randomly distributed, but are clustered within two areas designated Regions 1 and 3. It should be noted that the boundary for Region 1 and the location of the hypervariable site is based in part upon DNA sequence mutations, which will be described in a later section. In addition to the highly conserved Region 2 (also called the central globular region), a second area designated Region 4 also contains highly conserved amino acid sequences.

It is also apparent from Figure 12 that these proteins share very extensive amino acid sequence homologies. This provides strong support for functional homology, with most of the amino acid differences being attributable to either conservative amino acid substitutions or to variations in protein length within specific regions common to all of these linker histones. This comparison also indicates that these proteins likely have evolved from a common ancestral gene. Amino acid sequence comparisons are limited in scope however, providing little direct evidence about the genetic relationships between proteins. The

Figure 12. H5 and H1^o amino acid sequence alignment. All presently available amino acid sequences for H1^o and H5 are aligned for maximum homology, with dots denoting gaps introduced to compensate for differences in length and "X" denotes an unknown residue. The amino acid sequences are divided into four regions, with variations in length restricted to Regions 1 and 3. The "hypervariable region" within Region 1 highlights the segment in which the majority of differences in length are located. BH1^o: bovine H1^o (Pehrson and Cole 1981); HH1^o: human H1^o (Doenecke and Tonjes 1986); XLH5A and XLH5B: Xenopus laevis H5 variants (present study); QH5: quail H5 (Seligy et al. 1976); CH5: chicken H5 (Kreig et al. 1982); PH5: pigeon H5 (Seligy et al. 1976); DH5: duck H5 (Doenecke and Tonjes 1984); GH5: goose H5 (Yaguchi et al. 1979); CONS: consensus sequence for the H5 histones.


```

REGION 1                                REGION 2
(VARIABLE)                             (CONSERVED)
<=====> <=====>
BH1° <=====> hypervariable region MIVAAIQAENRAGTXXQXIQKYVXXXXXKVGENA
HH1° TENST....S.APAACKPKRAKASKSTD HPKYSDMIVAAIQAENRAGSSRQSIQKYIKSHYKVGENA
XLH5A AENS.....AATPAACKPKRSKALKKSTD HPKYSDMILAAVQAEKSRSGSSRQSIQKYIKNHYKVGENA
XLH5B TENS.....APAAKPRRSKASKKSTD HPKYSDMILDAVQAKKSRSGSSRQSIQKYIKNNYTVGENA
QH5 TE.SL..VLSPAP.AKPKRA
CH5 TE.SL..VLSPAP.AKPKRVKASRRSAS HPTYSEMIAAAIRAESRGGSSRQSIQKYIKSHYKVGHNA
PH5 TE.SPIPVPAAPAAKPKPKRVSKRPAS HPPYSDMIAAA
DH5 TD.SPIP..APAPAAKPKRAKAPRKPAS HPSYSEMIVAAIRAESRGGSSRQSIQKYVVKSHYKVGQHA
GH5 TD.SPIP..APAPAAKPKRARAPRKPAS HPTYSEMIAAAIRADKSRGGSSRQSIQKYVVKSHYKVGQHA

CONS TENSPIPVLAAPAAKPKRAKASKKSTD HPKYSDMIVAAIQAENRAGSSRQSIQKYIKSHYKVGENA
AD L PSAT RPSRVPRPAS T E AD VR D N S T V N T QH
T V L S L K G H
K P

REGION 2                                REGION 3
(CONSERVED)                             (VARIABLE)
===== <=====>
BH1° DSEIKLSIKRLVTTGVLKQTKGVGAXGXFRLAK
HH1° DSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEPKK SVAF.KKTKKEIKKVAT.PKKASKPKKAA
XLH5A DSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKADEGKK .PA..KKPKKEIKK.AVSPKKVAKPKKAA
XLH5B DSQIKLSIKRLVTTGTLKQTKGVGASGSFRLAKADEVKK .PA..KKPKKEIKK.AVSPKKAAKPKKAA
CH5 DLQIKLSIRLLAAGVLKQTKGVGASGSFRLAKSDKAKR SP.G..KKKKAARR.STSPKKAAARPKA.
DH5 DLQIKLSIRLLAASVLKQTKGVGASGSFRLAKGDKAKK SPAGRKKKKKAARK.STSPKKAAARPKA.
GH5 DLQIKLAIRLLAASVLKQTKGVGASGSFRLAKGDKAKR SPAGRKKKKKAARK.STSPKKAAARPKA.

CONS DSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEAKK SPAGRKKKKKEIKKVATSPKKAAKPKKAA
LE A R LASSA I A K P R V F P AARR SV VSR R
A T G G T V
V

REGION 3                                REGION 4
(VARIABLE)                             (CONSERVED)
===== <=====>
HH1° .SKAPTCKPKATP..VKK. AKKKLAATPKAKKPKTVKAKPVKASKPKKAKPVKPKAKSSAKRAGKKK
XLH5A KSPAKAKKPKVAEKKVKV AKKKPAPSPKAKKKTIVKAKPVRAKVKKASPPKPKAKASPKKSGRKK
XLH5B KSPAKAKKPKVAEKKVKV AKKKPAPSPKAKKKTIVRAKPVXASKAKKLSKPKAKASPKKSGRKK
CH5 RSPA..KKPKATARK.... ARKKSRSAPKAKKPKTVKAKSRKASKAKKVKRSKPRAKSGARKSPKKK
DH5 RSPA..KKPKAAARK.... ARKKSRSAPKAKKPKTVKAKSLKTSKVKKAKRSKPRAKSGARKSPKKK
GH5 RSPA..KKPKAAARK.... ARKKSRSAPKAKKPKTVKAKSLKTSKPKKARRSKPRAKSGARKSPKKK

CONS KSPAKAKKPKAAAKVKV AKKKLAASPKAKKPKTVKAKPVKASKPKKAKPSKPKAKSSAKKSGKKK
R K PT VTER A PR PRPT R T R SLRTT V VRRP R AGPRRAPR
P S R A SLV

```

next phase of the project was therefore to conduct a comparative DNA sequence analysis on the H5 and H1⁰ proteins ~~for~~ which DNA sequence was available.

ii) DNA Sequence Comparisons

Figure 13 presents an alignment of the coding regions for the human H1⁰, Xenopus laevis H5A, Xenopus laevis H5B, duck H5 and chicken H5; similar to that constructed for the amino acid sequence comparison. Importantly, this comparison allows for a detailed examination of the sources for differences in amino acid sequence between these proteins. From this comparison, common ancestry is most evident within Regions 2 and 4, in which essentially all differences in amino acid sequence can be directly attributed to single base differences. This is directly related to the conservation of length within these regions, which is necessary for the preservation of colinearity between homologous sequences.

Common ancestry is more difficult to establish from examination of Regions 1 and 3, in which the disruption of colinearity caused by variation in length has obscured the sequence homologies. This is due to the fact that the simple introduction of sequence gaps can be rendered ineffective for regions of DNA that have undergone multiple insertions and/or deletions, some of which may overlap. Effective comparative analysis of such regions therefore required a modified approach.

Figure 13. DNA sequence alignment of H5 and H1⁰ coding regions. The coding regions for which DNA sequence is available are aligned for maximum homology, similar to the amino acid sequence alignment shown in Figure 12. The corresponding amino acids are shown above the DNA sequence. Sequence gaps have been introduced to compensate for differences in length, which are clustered within two areas designated Regions 1 and 3. Differences between opposing nucleotides are designated "*" for those that produce differences in amino acid coding and "+" for those that represent codon wobble. HH1⁰: human H1⁰ (Doenecke and Tonjes 1986); XLH5A and XLH5B: Xenopus laevis H5 variants (present study); DH5: duck H5 (Doenecke and Tonjes 1984); CH5: chicken H5 (Kreig et al. 1982).

REGION 1

```

<=====
HH1°  T E N S T S      A P A A K P K R A K A S K K S T
      ACCGAGAATTCCACGTCC      GCCCCTGCGGCCAAGCCCAAGCGGGCCAAGGCCTCCAAGAAGTCCACA
      * + *      ** * +      * + + +      +      ** +      + * +      + + +
      A E N S A A      T P A A K P K R S K A L K K S T
XLH5A GCAGAGAATTACGCGCT      ACTCCAGCTGCCAAACCCAAGCGATCAAAGGCATTAAAGAAATCAACT
      *      +      *      +
      T E N S      A P A A K P R R S K A S K K S T
XLH5B ACAGAGAATTCC      GCTCCAGCTGCCAAACCTAGCGCATCAAAGGCATCAAAGAAATCAACA
      + * * * *      + + +      + + *      ** +      ** + *      ** + *
      T D S P I P A P A P A A K P K R A K A P R K P A
DH5   ACGGACAGCCCCATCCCGCCCCCGCCCCGCGCCCAAGCCCAAGCGGGCGAAGGCTCCTCGGAAGCCGGCG
      *      ***      **      +      +      *      ** +      ****      +
      T E S L V L S P A P      A K P K R V K A S R R S A
CH5   ACGGAGAGCCTGGTCCTATCCCCAGCCCCA      GCCAAGCCCAAGCGGGTGAAGGCATCGCGGCGCTCGGCA

```

REGION 2

```

==> <=====
HH1°  D   H P K Y S D M I V A A I Q A E K N R A G S S
      GAC CACCCCAAGTATTCAGACATGATCGTGGCTGCCATCCAGGCCGAGAAGAACCGCGCTGGCTCCTCG
      +      + +      + *      + * + + +      * +      * + + +
      D   H P K Y S D M I L A A V Q A E K S R S G S S
XLH5A GAC CACCCTAAGTACTCTGACATGATATTGGCTGCGGTGCAAGCAGAGAAGAGTCGCTCTCGTTCATCA
      +      * +      *      +
      D   H P K Y S D M I L D A V Q A K K S R S G S S
XLH5B GAC CATCCTAAGTACTCTGACATGATATTGGATGCTGTGCAAGCAAAGAAGAGTCGCTCCGTTTCATCA
      **      + + **      + *      **      ** + * + *      +      *** + + +
      S   H P S Y S E M I V A A I R A E K S R G G S S
DH5   TCC CACCCAGCTACTCGGAAATGATCGTGGCGGCCATCCGGGCCGAGAAGAGCCGCGGTGGCTCCTCC
      +      *      +      *      + + +      +      +
      S   H P T Y S E M I A A A I R A E K S R G G S S
CH5   TCG CACCCACCTACTCGGAGATGATCGCGCGGCCATCCGTGCGGAAAAGAGCCGCGCGGCTCCTCG

```

REGION 2

```

=====
HH1°  R Q S I Q K Y I K S H Y K V G E N A D S Q I K L
      CGCCAGTCCATTCCAGAAGTATATCAAGAGCCACTACAAGGTGGGTGAGAACCCTGACTCGCAGATCAAGTTG
      + + + +      +      *      + + + +      + *      + **
      R Q S I Q K Y I K N H Y K V G E N A D S Q I K L
XHL5A CGTCAATCAATCCAGAAGTACATCAAAAACCACTATAAAGTTGGAGAGAATGCAGACTCGCAAATCAAACCTG
      +      +      *      *      +      +
      R Q S I Q K Y I K N N Y T V G E N A D S Q I K L
XLH5B CGTCAGTCAATCCAGAAGTACATTAATAAACAACCTATACAGTTGGCGAGAATGCAGATTACAAATCAAACCTG
      +      +      * + + *      + * +      * * +      + * + + + +
      R Q S I Q K Y V K S H Y K V G Q H A D L Q I K L
DH5   CGGCAGTCCATCCAGAAGTACGTGAAGAGCCACTACAAGGTGGGCCAGCAGCCGACCTCCAGATCAAGCTC
      * +      **      + +
      R Q S I Q K Y I K S H Y K V G H N A D L Q I K L
CH5   CGGCAGTCCATCCAGAAGTACATCAAGAGCCACTACAAGGTGGGCCACAACCCGATCTGCAGATCAAGCTC

```

REGION 2

```

=====
S I K R L V T T G V L K Q T K G V G A S G S F R
HH1° TCCATCAAGCGCCTGGTCACCAACGGGTGTCCTCAAGCAGACCAAAGGGGTGGGGGCCTCGGGGTCCTTCCGG
+ + ++ + + * + + + + + + + + +
S I K R L V T S G A L K Q T K G V G A S G S F R
XLH5A TCTATTAAGAACTGGTGACTTCGGGCGCCCTCAAACAGACCAAAGGAGTGGGCGCCTCTGGATCTTTCCGG
+ + + + + + * + + + + + + + + +
S I K R L V T S G T L K Q T K G V G A S G S F R
XLH5B TCTATTAAGAAATTGGTGACTTCTGGCACCCCTCAAACAGACCAAAGGAGTGGAGCCTCTGGATCATTCCGG
+ + + + + + * + + + + + * + + + + + + + + +
S I R R L L A A G V L K Q T K G V G A S G S F R
DH5 TCCATCGGGCGCCTGCTCGCGCGCGGCGTCCTCAAGCAGACCAAAGGGGTGGGGGCCTCCGGCTCCTTCCGG
+ + + + +
S I R R L L A A G V L K Q T K G V G A S G S F R
CH5 TCCATCCGACGTCTCCTGGCTGCCGGCGTCCTCAAGCAGACCAAAGGGGTGGGGGCCTCCGGCTCCTTCCGG

```

REGION 2

```

=====
L A K S D E P K K
HH1° CTAGCCAAGAGCGACGAACCCAAGAAG
+ + + + + + + + + +
L A K A D E G K K
XLH5A TTAGCCAAGGCAGATCAGGGCAAGAAA
+ + + + +
L A K A D E V K K
XLH5B TTAGCCAAGGCAGATGAGGTCAAGAAA
+ + + + +
L A K G D K A K K
DH5 CTGGCCAAGGGGACAAGGCCAAGAAG
+ + + + +
L A K S D K A K R
CH5 TTGGCCAAGAGCGACAAGGCCAAGAGG

```

REGION 3

```

<=====
S V A F K K T K K E I K
TCAGTGGCCTTC AAGAAGACCAAGAAGGAAATCAAG
+ + + + + + + + + +
P A K K P K K E I K
CCAGCA AAGAAGCCTAAAAAAGAAATTAAA
+ + + + +
P A K K P K K E I K
CCAGCA AAGAAGCCTAAGAAGGAAATTAAA
+ + + + +
S P A G R K K K K K A A R
TCCCCAGCTGGGAGGAAGAAGAAGAAGGCTGCCCCA
+ + + + +
S P G K K K K A V R
TCCCC GGG AAGAAGAAGAAGGCCGTCAAG

```

REGION 3

```

=====
K V A T F K K A S K P K K A A S K A P T K K
HH1° AAGGTAGCCACG CCAAAGAAGGCATCCAAGCCCAAGAAGGCTGCC TCCAAAGCCCCAACCAAGAAA
+ + + + + + + + + + + + + + + +
K A V S P K K V A K P K K A A K S P A K A K K
XLH5A AAA GCAGTTTCACCCAAAAAAGTTGCCAAGCCCAAAAAAGCTGCAAAGTCTCCTGCTAAGGCCAAGAAA
+ + + + +
K A V S P K K A A K P K K A A K S P A K A K K
XLH5B AAA GCAGTTTCACCCAAAAAAGCTGCCAAGCCCAAAAAAGCTGCAAAGTCTCCTGCTAAGGCCAAGAAA
+ + + + +
R S T S P R K A A R P R K A R S P A K K
DH5 AGG TCCACGTACCCAGGAAGGCGGCGAGGCCAGGAAA GCCAGGTCCCGGGC AAGAAG
+ + + + +
R S T S P K K A A R P R K A R S P A K K
CH5 AGG TCCACGTCTCCAAGAAGGCGGCGAGGCCAGGAAAG GCCAGGTACCGGGC AAGAAG

```

REGION 3

```

=====>
HH1°  P K A A P      V K K
      C C C A A G C C G C C C C G      G T C A A G A A G
      +   **   ***      +   +
      P K V A E K K V K K V
XLH5A C C T A A A G T G C C A G A G A A G G T G A A G A A G T A
      +   +
      P K V A E K K V K K A
XLH5B C C T A A A G T G C C A G A G A A G A A G T G A A G A A G C A
      +   **   +   **   +
      P K A A A R K
DH5   C C C A A G C C G C C G C C A G G A A A
      *
      P K A T A R K
CH5   C C C A A G C C A C C G C C A G G A A G

```

REGION 4

```

<=====
      A K K K L A A T P K K A
G C C A A G A A G A A G C T G G C T G C C A C G C C C A A G A A G C C
      +   +   **   **   **   +
      A K K K P A P S P K K A
G C T A A G A A A A G C C T G C A C C T T C T C C C A A A A A G C T
      *
      P K K K P A P S P R K A
C C T A A G A A A A G C C T G C A C C T T C T C C C A G A A A G C T
      *   *   +   *   *   *   *   *   *   *
      A R K K S R A S P K K A
G C C A G G A A G A A G T C G A G G G C C A G C C C C A A G A A G C C
      +
      A R K K S R A S P K K A
G C C A G G A A G A A G T C G C G G G C A A G C C C C A A G A A G G C C

```

REGION 4

```

=====
HH1°  K K P K T V K A K P V K A S K P K K A K P V K P
      A A A A A C C C A A G A C T G T C A A A G C C A A G C C G T C A A G G C A T C C A A G C C C A A A A G G C C A A C C A G T G A A A C C C
      *   +   +   +   +   +   +   +   +   +   +   +   +
      K K T K T V K A K P V R A S K V K K A K P S K P
XLH5A A A A A A G A C A A A A C T G T C A A A G C C A A G C C A G T G A G G G C A T C C A A G G T A A A G A A G G C C A A G C C C T C C A A A C C A
      +   *
      K K T K T V R A K P V X A S K A K K A K L S K P
XLH5B A A A A A G A C A A A A C C G T C A G A G C C A A G C C A G T G X G G G C A T C C A A G G C A A A G A A G G C C A A G C T T T C C A A A C C A
      +   *   +   +   +   +   +   +   +   +   +   +   +
      K K P K T V K A K S L K T S K V K K A K R S K P
DH5   A A G A A G C C A A A G A C T G T T A A G G C C A A G T C G C T G A A G A C A T C C A A G G T G A A G A A G C C A A A G C G G T C A A A A C C C
      *   *   +   *   +   *   +
      K K P K T V K A K S R K A S K A K K V K R S K P
CH5   A A G A A G C C A A A G A C T G T T A A G G C C A A G T C G C G G A A G G C C T C C A A G G C C A A A A G G T G A A G C G G T C G A A A C C C

```

REGION 4

```

=====>
HH1°  K A K S S A K R A G K K K
      A A A G C C A A A G T C C A G T G C C A A G A G G G C C G G C A A G A A G T G A
      +   +   +   +   +   +   +
      K A K A S P K K S G R K K
XLH5A A A G G C C A A A G C C A G C C C A A A G A A A T C T G G A C G G A A G A A G T A A
      +
      K A K A S P K K S G R K K
XLH5B A A G G C C A A A G C C A A G C C C A A A G A A A T C T G G A C G G A A G A A G T A A
      **   +   *   **   *   ****   +   +   +   +
      R A K S G A R K S P K K K
DH5   A G A G C C A A A T C C G G C G C C C G A A A T C G C C C A A A A A G A A G T G A
      +   +   +   +
      R A K S G A R K S P K K K
CH5   A G A G C C A A G T C T G G C G C C C G A A A T C G C C C A A G A A G A A G T G A

```

D) Investigation into the Sequence Evolution of H5 and H1⁰

1) Region 1 and 3 of Xenopus laevis H5 and Human H1⁰: Characterization of Direct Duplication

The analytical approach used in the following sections was based on the premise that once the mutational events responsible for disruption of colinearity were recognized, effective sequence comparisons could be constructed.

It is evident from Figures 9 and 13 that XLH5A and XLH5B are most likely products of a recent gene duplication. The 6 bp gap within Region 1 (shown below) is the only observed disruption of colinearity within the coding regions of these genes, with all other sequence differences likely to be products of single base mutations. Note that in these comparisons, '*' denotes nucleotide mismatches that produce differences in amino acid coding and that '+' denotes nucleotide mismatches which have no effect on amino acid coding due to codon redundancy.

	A	E	N	S	A	A	T	P	A	A	K	P	K	R	S	K	A	L	K	K	S	T	D	H		
XLH5A	G	C	A	G	A	A	T	T	G	A	G	C	C	G	C	T	A	C	C	A	A	C	C	A	A	
	*																									
XLH5B	A	C	A	G	A	A	T	T	C	C	G		C	T	C	A	G	C	T	G	C	C	A	A	C	
	T	E	N	S					A	P	A	A	K	P	R	R	S	K	A	S	K	K	S	T	D	H

It is clear that this gap has been produced by either a 6 bp insertion into XLH5A or a 6 bp deletion within XLH5B. Additional information is needed however to precisely establish its origin.

When Region 1 of human H1^o is aligned with XLH5A, no difference in length is revealed:

```

      <=====>
      A E N S A A T P A A K P K R S K A L K K S T D H
XLH5A GCAGAGAATTCAGCCGCTACTCCAGCTGCCAAACCCAAGCGATCAAAGGCATTAAAGAAATCAACTGACCAC
      * +      ** ** + + + + +      +      ** +      + **      + + +
HH1o ACCGAGAATTCACGTCCGCCCTGCGGCCAAGCCCAAGCGGGCCAAGGCCTCCAAGAAGTCCACAGACCAC
      T E N S T S A P A A K P K R A K A S K K S T D H
      <=====>

```

It is difficult however, to conclude that the outlined cluster of nucleotide mismatches are due solely to single base mutations, suggesting instead the occurrence of additional mutational events. Close examination of the human H1^o sequence does reveal a possible sequence duplication that is absent within XLH5A. This was initially recognized through the presence of what I have called "direct duplicates", as described below:

```

      A E N S A A T P A A K P K R S K A L K K S T D H
      ----->----->
XLH5A GCAGAGAATTCAGCCGCTACTCCAGCTGCCAAACCCAAGCGATCAAAGGCATTAAAGAAATCAACTGACCAC
      * +      ** ** + + + + +      +      ** +      + **      + + +
HH1o ACCGAGAATTCACGTCCGCCCTGCGGCCAAGCCCAAGCGGGCCAAGGCCTCCAAGAAGTCCACAGACCAC
      ----->----->
      T E N S T S A P A A K P K R A K A S K K S T D H

      DIRECT          NO
      DUPLICATES      DUPLICATES
      S T              S A
      ----->        ----->
      TCCACG          TCAGCC
HH1o  * +          XLH5A * + * +
      TCCGCC          GCTACT
      ----->        ----->
      S A              A T

```

I propose that these direct duplicates are produced by a DNA sequence modification which I will call "direct duplication"; by "duplication" I am not implying a mechanism, but rather the outcome of an event or series of events, which ultimately results in the following form:

DIRECT DUPLICATION

```

----->
XXX12345XXX =====> XXX1234512345XXX
XXX67890XXX           XXX6789067890XXX

```

Many other direct duplicates have been found within these genes, suggesting that direct duplication could be a frequent form of DNA mutation. The largest observed direct duplicate is found within Region 3 of XLH5A:

```

      (A)                      (B)
      ----->
XLH5A TCACCCAAAAAAGTTGCCAAGCCCAAAAAAGCTGCAAAGTCT
      S P K K V A K P K K A A K S

```

```

      P K K V A K
(A)----->
      CCCAAAAAAGTTGCCAAG
      * +
      CCCAAAAAAGCTGCAAAG
(B)----->
      P K K A A K

```

18 bp

Examination of the other H5 genes reveals the presence of homologous duplicates as is shown below:

REGION 3

```

      K V A T      P K K A S K P K K A A      S K A P T
      ----->
HH1o AAGGTAGCCACG  CCAAAGAAGGCATCCAAGCCCAAGAAGGCTGCC  TCCAAAGCCCCAACC
      +   + + + +   + + + + +   + +   +   + + + + +
      K   A V S P K K V A K P K K A A K S P A K A
      ----->
XLH5A AAA  GCAGTTTCACCCAAAAAAGTTGCCAAGCCCAAAAAAGCTGCAAAGTCTCCTGCTAAGGCC
      *
      K   A V S P K K A A K P K K A A K S P A K A
      ----->
XLH5B AAA  GCAGTTTCACCCAAAAAAGCTGCCAAGCCCAAAAAAGCTGCAAAGTCTCCTGCCAAGGCC
      **   * + + + +   ** + + + +   **   + * + + +
      R   S T S P R K A A R P R K      A R S P A
      ----->
DH5  AGG  TCCACGTACCCAGGAAGGCGGCGAGGCCAGGAAA  GCCAGGTCGCCGCGC
      +   *   +
      R   S T S P K K A A R P R K      A R S P A
      ----->
CH5  AGG  TCCACGTCTCCCAAGAAGGCGGCGAGGCCAGGAAG  GCCAGGTCACCGGCC

```

These direct duplicates are compared below:

	(A)	(B)
	=====>	=====>
HH1 ^o	CCAAAGAAGGCATCCAAG	CCCAAGAAGGCTGCC...
XLH5A	CCCAAAAAAGTTGCCAAG	CCCAAAAAAGCTGCAAAG
XLH5B	CCCAAAAAAGCTGCCAAG	CCCAAAAAAGCTGCAAAG
DH5	CCCAGGAAGGCGCGGAGG	CCCAGGAAA...GCCAGG
CH5	CCCAAGAAGGCGCGGAGG	CCCAGGAAG...GCCAGG
CONSENSUS	CCCAAGAAGGCXGCCAAG	CCCAAGAAGGCTGCCAAG

		A
	P K K A A K	
(A)	=====>	
	CCCAAGAAGGCXGCCAAG	
CONSENSUS	?	
	CCCAAGAAGGCTGCCAAG	
(B)	=====>	
	P K K A A K	

This indicates that these direct duplicates could have been formed by direct duplication of the 18 bp consensus sequence before the divergence of these genes. If this proposal is correct, it can be further concluded that after the duplication event, but before the divergence of duck and chicken, the precursor avian H5 underwent a deletion of an alanine codon. The HH1^o gene has also likely undergone deletion of a lysine codon. Interestingly, both of these supposed deletions occurred within the same duplicate although the significance of this is unclear.

Other direct duplicates are shown below:

	REGION 3
	----->
XLH5A	AGAGAAGAAGGTGAAGAAAGTA
	+ * + * * + * + * +
	----->
CH5	CGCCAGGAAGGCCAGGAAGAAG
	+
	----->
DH5	CGCCAGGAAAGCCAGGAAGAAG

K K V	A R K	A R K	
----->	----->	----->	
XLH5A GAAGAAGGT *	CH5 GCCAGGAAG	DH5 GCCAGGAAA †	9 bp
GAAGAAAGT	GCCAGGAAG	GCCAGGAAAG	
----->	----->	----->	
K K V	A R K	A R K	

REGION 3

XLH5A GCCTAAAAAAGAAATTAAAAAAGCAGTTTC

K K E I	
----->	
TAAAAAAGAAAT	
* *	12 bp
TAAAAAAGCAGT	
----->	
K K A V	

It should be noted that the occurrence of a direct duplication event cannot be conclusively established solely by the presence of direct duplicates. It is possible, indeed likely that some of the direct duplicates shown above are due to the fortuitous occurrence of repetitive amino acid sequences, which have evolved independently of any duplication event. Further examination of Region 1 (to be presented later) will however provide strong evidence that sequence duplication is in fact a major form of DNA mutation.

Several general characteristics of direct duplication can be predicted. For example, most (with exceptions to be presented later) were found to produce inserts that preserve reading frame, reflecting a strong selection for the preservation of protein function. Preservation of protein function may also be a central factor in establishing a potential tolerance to the amino acid insertions produced by direct duplication. Specifically, in contrast to insertion of exogenous DNA which can introduce "foreign" amino acids that would have a high

potential for disrupting protein function, direct duplication can introduce insertions that are identical to adjacent amino acid sequences. This can reduce some of the potential effects produced by direct duplication, and allow preservation of protein function when changes in peptide length can be tolerated. The extensive variations in peptide length observed within Regions 1 and 3 are likely related to such a functional tolerance within H5, and is consistent with the frequent occurrence of sequence duplicates observed within these regions of H5.

An exception to this generalization involves the codon in which the duplicated sequence is inserted. Although all examples presented above produce perfect in-frame insertions such that all codons are preserved, the codons formed at the end of each duplicate is dependent upon the site of insertion, thus making the identity of these codons unpredictable. Further discussion of this phenomenon will be presented, with examples, during a more detailed examination of Region 1.

Importantly, direct duplication represents the generation of new DNA sequence, which must be removed in order to restore the colinearity between related genes. For example, based upon the proposed direct duplication within HH1⁰, the alignment with XLH5 can be modified as shown below:

```

XLH5B ACAGAGAAT      TCCG.....CTCCAGCTGCCAAACCTAGGCGATCAAAGGCATCAAAGAAATCAACA
      *              +              *              *              +
XLH5A GCAGAGAAT      TCAGCCGCTACTCCAGCTGCCAAACCCAAGCGATCAAAGGCATTAAAGAAATCAACT
      * +            + *            + *            + *            + *            + *
HH10 ACCGAGAATTCCACGTCCGCCCT.....GCCGCCAAGCCCAAGCGGGCCAGGCCCTCCAAGAAGTCCACA
      ----->----->
XLH5B  T E N . . S . . A P A A K P R R S K A S V K S T
XLH5A  A E N . . S A A T P A A K P K R S K A L K K S T
HH10   T E N S T S A P . . A A K P K R A K A S K K S T

```

This alignment predicts the occurrence of a 6 bp deletion within HH1° (denoted by dots) and suggests that the gap in XLH5B was produced by a 6 bp deletion. It can therefore be proposed that HH1° has undergone a 6 bp deletion and a 6 bp direct duplication (the order of which cannot be determined), that not only preserved reading frame but also peptide length. In addition, the lack of an homologous duplicate within XLH5A indicates that this direct duplication event most likely occurred after divergence of the mammals and amphibians. It can be further deduced that this direct duplication within HH1° produced a perfect insertion of the amino acids "SA", which were later modified to "ST". Furthermore, although the 6 bp deletions within XLH5B and HH1° overlap by only one base, both have produced deletions of "TP". The most important conclusion however, is the confirmation that HH1° and XLH5 have likely been derived from a common ancestral H5 gene.

The next step was to construct a sequence representing the hypothetical ancestral sequence from which XLH5 and HH1° were derived. This entailed removal of duplicated sequences, restoration of deleted sequences, and elimination of supposed single base mutations as predicted from a consensus of these sequences:

```

HH1° ACCGAGAATTCCGCCCT.....GCGGCCAAGCCCAAGCGGGCCAAGGCCTCCAAGAAGTCCACAGACCAC
HH1° duplc   TCCACG
XH5A GCAGAGAATTCCAGCGCTACTCCAGCTGCCAAACCCAAGCGATCAAAGGCATTAAAGAAATCAACTGACCAC
XH5B ACAGAGAATTCCG.....CTCCAGCTGCCAAACCTAGCGATCAAAGGCATCAAAGAAATCAACAGACCAT

aAMP ACWGAGAATTCWGGCCCTACTCCAGCWGCCAAWCCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGACCAC
      T E N S A P T P A A K P K R A K A S K K S T D H

```

Note that "W" denotes codon redundancy. It is not proposed that this is an exact replica of the H5 ancestral sequence, but rather it is a close representative which will be useful in further analysis. I have called this sequence "ancestral amphibian H5" or "aAMP", although the

species from which this ancestral sequence might have been derived cannot be established at present.

ii) Region 1 of Duck and Chicken H5: Characterization of Mirror Duplication

The next phase of the analysis was to examine Region 1 of the chicken and duck H5 genes, which are compared below with the aAMP H5 sequence:

```

      T E S L V L S P A P   A K
CH5  ACGGAGAGCCTGGTCCTATCCCCAGCCCCA   GCCAAG
      **** * **          * +
      T E N S A P           T P A A K
aAMP ACWGAGAATTCWGGCCCT      ACTCCAGCWGCCAAW
      * *** ** +          * + +
      T D S P I P A P A P A A K
DH5  ACGGACAGCCCCATCCCGGCCCGGCCCGGCCCAAG
      <=====>

```

Section 1

```

      P K R V K A S R R S A S H
CH5  CCCAAGCGGGTGAAGGCATCGCGGCGCTCGGCATCGCAC
      *          ** *** * ***
      P K R A K A S K K S T D H
aAMP CCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGACCAC
      * ** ++ * ***
      P K R A K A P R K P A S H
DH5  CCCAAGCGGGCGAAGGCTCCTCGGAAGCCGCGTCCAC
      <=====>

```

Section 2

The two major clusters of mismatches (designated as Section 1 and Section 2), in combination with the sequence gaps, suggest the occurrence of other sequence modifications in addition to single bp mutations. Comparison of the very closely related CH5 and DH5 further indicates that at least some of these sequence modifications have occurred after the divergence of duck and chicken:

T E S L V L S P A P A K
 CH5 ACGGAGAGCCTGGTCCTATCCCGAGCCCCA GCCAAG
 * ** * * +
 DH5 ACGGACAGCCCCATCCCGGCCCCGGCCCCGGCCGCAAG
 T D S P I P A P A P A A K
 <=====>

Section 1

P K R V K A S R R S A S H
 CH5 CCCAAGCGGGTGAAGGCATCGCGCGCTCGGCATCGCAC
 * ** + *** + +
 DH5 CCCAAGCGGGCGAAGGCTCCTCGGAAGCGCGCTCCAC
 P K R A K A P R K P A S H
 <=====>

Section 2

It should be noted that these are the most dissimilar areas within the entire coding regions of the duck and chicken H5 genes. Examination of the DH5 Section 2 reveals the occurrence of a second type of sequence duplication that I have called "mirror duplication"; again I use the term "duplication" without implying a mechanism but rather an outcome which takes the following form:

MIRROR DUPLICATION

-----> ----->-----
 XXX12345XXX =====> XXX1234554321XXX
 XXX67890XXX XXX6789009876XXX

(DUPLICATES OF
 OPPOSITE POLARITY)

This produces an insert that is a mirror image of preceding sequences and is not an inverted repeat which takes the following form:

INVERTED REPEATS

----->
 XXX1234509876XXX
 XXX6789054321XXX
 <-----

Mirror duplication produces an interesting situation in which the polarity of the duplicated DNA is reversed, preventing direct duplication of amino acid sequence as can occur with direct duplication. Several mirror duplicates have been found during examinations of these

H5 gene sequences, suggesting that this is a frequently occurring form of DNA mutation.

A 15 bp mirror duplicate within Section 2 of DH5 is outlined below:

```

      P K R V K A S R R S A S H
CH5  GCCCAAGCGGGTGAAGGCATCGCGGCGCTCGGCATCGCAC
      *      ** +      ***      + +
DH5  GCCCAAGCGGGCGAAGGCTCCTCGGAAGCGGCGTCCCAC
      (A)-----><----- (B)
      P K R A K A P R K P A S H

```

```

READING FRAME =====>
      K R A K A
(A)----->
      AGCGGGCGAAGGCTC
      *      *
      TCGGGCGGAAGGCTC
(B)----->
      A P K R P
<===== READING FRAME

```

15-bp

It should be noted that the first mismatching A-T pair was initially included in this mirror duplicate in order to allow preservation of reading frame, and illustrates the inherent difficulty of establishing the ends of mirror duplicates. Evidence to be presented later, did subsequently confirm that this mirror duplicate was formed by duplication of 15 bp. Note also, that in addition to a reversed orientation, the newly formed sequence uses a different reading frame from that of the copied sequence. As would be expected, this produces a completely different amino acid sequence, although the new amino acids are not unusual for those found in the amino terminus of these linker histones. This large mirror duplication is the only apparent mutation within Section 2 of DH5.

Mutations within CH5 are somewhat more complex involving both mirror and direct duplications. A 6 bp mirror duplicate is outlined below:

SECTION 2

P K R V K A S R R S A S H
 CH5 GCCCAAGCGGGTGAAGGCATCGCGCGCTCGGCATCGCAC
 (A)-----><-----<----->(B)

S R
 (A)----->
 ATCGCG
 * 6 bp
 CTCGCG
 (B)----->
 S R

Note that the new amino acids are similar to the original duplicate sequence despite the reversal of sequence polarity and change in reading frame. Again the first mismatched A-C has been included in these mirror duplicates. This is not however based solely on preservation of reading frame, in that the duplicated sequence is nested within a 9 bp direct duplicate:

P K R V K A S R R S A S H
 -----><-----
 CH5 GCCCAAGCGGGTGAAGGCATCGCGCGCTCGGCATCGCAC
 ----->*****<----->
 (A) (B)
 A S R
 (A)----->
 GGCATCGCG
 * 9 bp
 GGCATCGCA
 (B)----->
 A S H

This fortuitously allows the ends of this mirror duplication to be established. It can also be concluded that the A-C mismatch within the mirror duplicates is due to single base mutation of A to C in the second mirror duplicate, since A occurs at both homologous positions within the direct duplicates (denoted by lower case). This presents a perplexing situation in that the mirror duplication would have produced an amber mutation (TAG) at its inception, which upon mutation of A to C

would produce the present (TCG) serine codon (see below). Note that the CH5 duplications produce 6 and 9 bp inserts which preserve reading frame. This results in a net increase of 15 bp, as did the mirror duplication within DH5.

From this analysis, the following can be proposed for the sequential evolution of Section 2 of CH5:

<====> HIGHLY CONSERVED AMINO ACIDS
K A S H P
PRECURSOR SEQUENCE AAGGCATCGCACCCC
----->

DIRECT DUPLICATION (9 bp)

K A S Q A S H P
----->----->
AAGGCATCGCAGGCATCGCACCCC
*
G

Note that formation of "Q" is a result of the site of insertion. This could have disrupted the proper functioning of H5 and led to selection of a single base mutation which produced conversion of Q->R. Note also the preservation of the highly conserved "H".

A->G MUTATION CONVERTING Q->R

K A S R A S H P
----->----->
AAGGCATCGCGGCGCATCGCACCCC
----->

MIRROR DUPLICATION (6 bp):
PRODUCTION OF AN AMBER MUTATION

----->*****----->
-----><-----
K A S R R stp A S H P
AAGGCATCGCGCGCTAGGCATCGCACCCCT
*
C

A->C MUTATION CONVERTING stp->S

```

----->*****----->
-----><-----
A K P K R V K A S R R S A S H P T Y
CH5 GCCAAGCCCAAGCGGGTGAAGGCATCGCGgcgctcggcatcgcaCCCCACCTAC
      *      ++ +      ****      + +      *
DH5 GCCAAGCCCAAGCGGGCGGAAGGCTCctcggaagcggcgtCCCACCCCAGCTAC
A K P K R A K A P R K P A S H P S Y
-----><-----

```

(LOWER CASE DENOTES INSERTED BASES)

Removal of the duplicated sequences produces representations of the duck and chicken H5 precursor sequences (pDH5 and pCH5) prior to the duplication events, as shown below:

SECTION 2

```

A K P K R V K A S H P T Y
pCH5 GCCAAGCCCAAGCGGGTGAAGGCATCGCACCCCACCTAC
      *      ++ +      *
pDH5 GCCAAGCCCAAGCGGGCGAAGGCTCCCCACCCCAGCTAC
A K P K R A K A P H P S Y

```

This reveals that, excluding single base mutations, the precursor sequences are identical and thus likely represent the ancestral sequence for Section 2 of H5 before divergence of chicken and duck.

Comparison of these precursor sequences with the human H1⁰ and Xenopus laevis H5 sequences is shown below:

SECTION 2

```

A K P K R A K A S K K S T D H P K Y
HH10 GCCAAGCCCAAGCGGGCCAAGGCCTCCAAGAAGTCCACAGACCACCCCAAGTAT
      + + *      ++ +      + +      + +      + +
A K P R R S K A S K K S T D H P K Y
XLH5B GCCAAACCTAGCGGATCAAAGGCATCAAAGAAATCAACAGACCATCCTAAGTAC
      + *      *      + +
A K P K R S K A L K K S T D H P K Y
XLH5A GCCAAACCCAAGCGATCAAAGGCATTAAGAAATCAACTGACCACCCTAAGTAC
      +      ++++      *      +      ++ **
A K P K R V K A S H P T Y
pCH5 GCCAAGCCCAAGCGGGTGAAGGCATC.....GCACCCCACCTAC
      +      *      ++      +      *
A K P K R A K A P H P S Y
pDH5 GCCAAGCCCAAGCGGGCGAAGGCTCC.....CCACCCCAGCTAC

```

<----->

15 bp

This clearly illustrates that after divergence from amphibian H5 but before divergence of the duck and chicken H5, the precursor avian H5 gene underwent a 15 bp deletion. Both the chicken and duck H5 genes then underwent independent mutational events which acted not only to replace the deleted 15 bp segment, but also produced amino acid insertions that closely matched the lost amino acid sequence, despite the fact that two mirror duplications were involved. This is illustrated below with the newly formed amino acids denoted by lower case:

II

CH5	AKPKRVKASrrsashPTY
DH5	AKPKRAKAprkpashPSY
XLH5B	AKPRRSKASKKSTDHPKY
XLH5A	AKPKRSKALKKSTDHPKY
HH1°	AKPKRAKASKKSTDHPKY

Analysis of Section 2 has thus provided strong evidence for the frequent occurrence of direct and mirror duplication, and has illustrated the potential of this analytical approach for resolving the DNA sequence evolution of related genes.

iii) Region 1, Section 1 of Duck and Chicken H5: Derivation of an Ancestral Avian H5

The next step in the analysis was to identify the mutational events that have led to other sequence differences within Section 1 of chicken and duck H5. This revealed the occurrence of a 9 bp mirror duplication within DH5:

SECTION 1

T E S L V L S P A P A K
CH5 ACGGAGAGCCTGGTCTATCCCCAGCCCA...GCCAAG
 * ** *** + +
DH5 ACGGACAGCCCATCCGGGCCCGGCGCCGGCCAAG
 (A)-----><----- (B)
T E S P I P A P A P A A K

(A) $\begin{array}{c} \text{P} \quad \text{A} \quad \text{P} \\ \text{-----} \\ \text{CCGGCCCCG} \end{array}$ 9 bp

(B) $\begin{array}{c} \text{CCGGCCCCG} \\ \text{-----} \\ \text{A} \quad \text{P} \quad \text{A} \end{array}$

This is another example of a mirror duplication that has produced similar amino acids. Because no single base mutations have occurred within the mirror duplicates and there is no evidence for an homologous duplication within CH5, this has likely been a recent duplication event. It should be noted that these duplicates are likely present in goose and pigeon H5's, in that their amino acid sequences within this area are identical to that of DH5 (Figure 12). This mirror duplication event thus likely occurred after the divergence of chicken and duck but before the divergence of goose, pigeon and duck. Based upon this mirror duplication the follow modification to the sequence alignment can be made:

T E S L V L S P A P . A K
CH5 ACGGAGAGCCTGGTCCTATCCCCAGCCCA.....GCCAAG
+ ** + +
DH5 ACGGACAGCCCCATC.....CCGGCCCGGCCCGGCCAAG
-----><-----
T E S P I P A P A P A A K

Further analysis reveals that these mirror duplicates are nested within a direct duplicate, similar to that previously described in Section 2 of CH5:

```

      (A)          (B)
      ----->----->----->
DH5  ACGGACAGCCCCATCCCGGCCCCGGCCCCGGCCGCAAG
      -----><-----
          S P I
      (A)----->
          CAGCCCCATCC
            *  *  *  11 BP
          CGGCCCCGGCC
      (B)----->
          P A P A

```

This an example of a direct duplication producing completely different amino acids due to a shift in reading frame within the duplicated sequence. It is also an example of an insert (11 bp) which would have disrupted reading frame at its inception. Examination of the CH5 sequence reveals that it contains a sequence similar to the second DH5 direct duplicate:

```

      T E S L V L S P A P          A K
      ----->----->
CH5  ACGGAGAGCCTGGTCCTATCCCCAGCCCCA.....GCCAAG
      +   *+*          +   +
DH5  ACGGACAGCCCCATC.....CGGCCCCGGCCCCGGCCGCAAG
      -----><-----
      ----->----->----->
      T E S P I          P A P A P A A K
          P A P A
CH5  CAGCCCCAGCC
          *
          S P I
DH5  CAGCCCCATCC  11 bp
          *  *
          P A P A
DH5  CGGCCCCGGCC

```

This suggests that identical direct duplicates are presence in CH5. Examination of CH5 sequence opposite the first DH5 direct duplicate did reveal an homologous sequence but one which has suffered several inserts:

T E S L V L S P A P A K
 CH5 ACGGAGAGCCTGGTCCTATCCCCAGCCCCAGCCAAG
 -----****--*--->*----->

-----****--*--->*
 TGGT T C
 GAGCC CC ATCC
 * *
 CAGCC....CC.AGCC
 ----->

The following modified sequence alignment can thus be constructed,
 after incorporating these duplications:

-----****--*--->*----->
 T E S L V L S P A P A K
 CH5 ACGGAGAGCCTGGTCCTATCCCCAGCCCCA.....GCCAAG
 * + +
 DH5 ACGGACAGCC....CC.ATCC.CGGCCCCCGGCCCGGCCGCAAG
 T D S P I P A P A P A A K
 -----><-----
 ----->-----*****-->

The following summary of events can thus be proposed for the DNA
 sequence evolution of Region 1 of chicken H5:

----->
 AVIAN PRECURSOR ACGGAG CAGCCCCAGCC AAG

DIRECT DUPLICATION (11 bp)

----->----->
 T D S P I P A P A K
 ACGGACAGCCCCATCCCAGCCCCAGCCAAG
 * *+
 G TG

THREE SINGLE BASE MUTATIONS: conversion of D->E and P->L

----->----->
 T E S L I P A P A K
 ACGGAGAGC CTG ATCCCAGCCCCAGCCAAG
 -->

MIRROR DUPLICATION OF CTG: insertion of V

T E S L V I P A P A K
 --><--
 ACGGAGAGCCTGGTCATCCCAGCCCCAGCCAAG
 ->

MIRROR DUPLICATION OF TC: frame shift mutation

T E S L V L S Q P Q P S
 --><--
 ACGGAGAGCCTGGTCCTATCCCAGCCCCAGCCAAG
 <=>

DUPLICATION OF ONE C: restoration of reading frame

T E S L V L S P A P A K
 <==>
 CH5: ACGGAGAGCCTGGTCCTATCCCAGCCCCAGCCAAGC

This leads to the following modified alignment in which lower case denotes inserted bases:

```

-----*****-->*-----*****-->
--><--
      T E S L V L S P A P           A K
CH5:  ACGGAGAGCCTGgtcctATCcCCAGCCCCA.....GCCAAGC
      _ *  *+ --><-  _ +      +
DH5:  ACGGACAGCCCC.....ATC.CCGGCCCCGgccccgcccGCCAAGC
      T D S P           I P A P A P A A K
                        -----><-----
----->-----*****-->

```

It should be noted that after removing the inserts to produced representations of the precursor sequences (pCH5 and pDH5), identical amino acids (excluding the conversion of E to D and L to P) are encoded within these regions of the DH5 and CH5 genes:

```

          C1          C2
          ----->----->
      T E S L I P A P A K
pCH5:  ACGGAGAGCCTGATCCCAGCCCCAGCCAAGC
          *  *+      +      +
pDH5:  ACGGACAGCCCCATCCGGCCCCCGGCCAAGC
      T D S P I P A P A K
          ----->----->
          D1          D2

```

These sequences should thus closely represent the avian H5 precursor.

iv) Evolution of Ancestral Avian H5 from Ancestral Amphibian H5

The next phase was to compare the avian precursor H5 sequence with aAMP. The first step taken was to establish a composite or consensus of the duplicate sequence:

C1 gAGCctgATCC
D1 CAGCCCCATCC
C2 CAGCCCCAGCC
D2 CgGCCCCgGCC

CONSENSUS CAGCCCCAGCC

T

Derivation of a consensus sequence allows elimination of single base changes that are supposed to have occurred after the direct duplication (denoted as lower case). Note that it is not possible at this point to distinguish between G and T in the ninth position within this consensus sequence. The fact that the G/T mismatch is identical within both CH5 and DH5, suggests that this single base mutation likely occurred before divergence of duck and chicken, where as all other single base mutations likely occurred after their divergence.

Initial examination of the aAMP H5 sequence revealed the absence of an homologous direct duplicate, indicating that this duplication event occurred after divergence of the ancestral avian H5 from amphibian H5. Removal of the duplicated sequence produces the sequence shown below:

pCH5: ACGGA AAGCCCAAGCGG
CONSENSUS CAGCCCCAGCC
pDH5: ACGGA T AAGCCCAAGCGG

Note that removal of the 11 base duplicate disrupts reading frame, and that these identical sequences will be referred as the ancestral avian H5 or aAVN. DNA sequence comparison of aAVN indeed reveals the

presence of the consensus sequence within aAMP, although additional sequences are also present:

	T E N	S A P T P A A	K P K R
aAMP	ACWGAGAATT	CWGGCCCTACTCCAGCGWGCC	AAWCCWAAGCGW
aAVN	ACGGA....	CAGCCCC.....AGCC...	AAGCCCAAGCGG

(T)

DUPLICATE, CONSENSUS: -----*****-->

From this it can be concluded that the unresolved G/T within the consensus sequence originated as G. It is also apparent that in addition to a direct duplication, three deletion events of 5, 6 and 3 bp have occurred during the evolution of the avian H5 gene. From this the following order of events can be proposed:

	T E N S A P T P A A K P K R
	-----*****-->
aAMP H5	ACWGAGAATTCAGCCCCCTACTCCAGCGWGCCAAWCCWAAGCGW
	<====>

DELETION OF 6 (OR 3) BP: net deletion of "TP"

	T E N S A P A A K P K R
	----->
	ACWGAGAATTCAGCCCCAGCGWGCCAAWCCWAAGCGW
	<=>

It is apparent that the 6 bp deletion occurred before the direct duplication of the arrowed sequence. However, the 3 bp deletion could have occurred either before or after the direct duplication, and could involve either alanine codon. Note also that the 6 bp deletion of "TP" is the same as the deletions within XLH5B and HH1⁰ sequences described earlier.

2 of/de 2

THE QUALITY OF THE IMAGE IS
MICRO
THE QUALITY OF THE IMAGE IS

DELETION OF 3 (OR 6) BP

T E N S A P A K P K R
ACWGAGATTTCAGCCCCAGCCCAAWCCWAAGCGW

DIRECT DUPLICATION OF 11 BP (OR DELETION OF 5 BP) CAUSING DISRUPTION OF READING FRAME

T E N S A P A N P N P S ? S
ACWGAGAATTTCAGCCCCAGCCCAGCCCCAGCCCAAWCCWAAGCGW
<===>

Disruption of reading frame is another example indicating that the avian H5 gene has passed through an inactive intermediate.

DELETION OF 5 BP (OR 11 BASE DUPLICATION) WHICH RESTORES READING FRAME

(NS)
OLD READING FRAME: T E A P A ***** K P K R
T E S P S P A P A K P K R
AVIAN PRECURSOR ACWGACAGCCCCAGCCCAGCCCCAGCCCAAWCCWAAGCGW

Note that the change in reading frame produced by combined action of the deletion and duplication, and that again the order of these two mutational events can not be resolved. Comparison of this resulting sequence with the precursor sequences of the chicken and duck H5 genes (presented earlier), is shown below:

PRECURSOR CH5: T E S L I P A P A K P K R
ACGGAGAGCCTGATCCAGCCCCAGCCCAAGCCCAAGCGG
* * *
AVIAN PRECURSOR: T D S P S P A P A K P K R
ACWGACAGCCCCAGCCCAGCCCCAGCCCAAWCCWAAGCGW
* * *
PRECURSOR DH5: T D S P I P A P A K P K R
ACGGACAGCCCCATCCCGGCCCGCCCAAGCCCAAGCGG

Based upon this comparison, it can be proposed that the S→I conversion within pCH5 and pDH5 likely occurred before the divergence of chicken and duck. It is possible but less likely that this was produced by independent mutations that have produced identical single base changes. From this, it can be proposed that the following single base changes occurred within pAVN H5 before divergence of the chicken and duck:

REMOVAL OF IDENTICAL SINGLE-BP CHANGES:

```

      T D S P S P A P A K P K R
pAVN: ACWGACAGCCCCAGCCCAGCCCCAGCCAAWCCWAAGCGW
      + + + + +
aAVN H5: ACGGACAGCCCCATCCCAGCCGACCAAGCCCAAGCGG
      T D S P I P A P A K P K R

```

Thus, aAVN H5 represents the avian H5 sequence directly preceding the divergence of the duck and chicken. The following single base changes thus likely occurred after their divergence:

```

      T E S L I P A P A K P K R
pCH5: ACGGAGAGCCTGATCCCAGCCCCAGCCAAGCCCAAGCGG
      * * +
aAVN H5: ACGGACAGCCCCATCCGAGCCCAAGCCCAAGCGG
      + + +
pDH5: ACGGACAGCCCCATCCCGGCCCGGCCAAGCCCAAGCGG
      T D S P I P A P A K P K R

```

v) Summary of the Proposed DNA Sequence Evolution for Region 1

The preceding section presented an analysis of some of the possible DNA sequence modifications that have occurred within Region 1 of human H1⁰, chicken H5, duck H5, Xenopus laevis h5A and Xenopus laevis H5B. From this analysis it was possible to resolve details about the DNA sequence evolution of Region 1, and to allow the origins of these genes to be traced back to a common ancestral H5 gene. This is outlined in Figure 14, which summarizes the major mutational events that occurred after the divergence of these genes.

Figure 14. Summary of the major DNA sequence modifications proposed to have occurred within Region 1 of H5 and H10 during the divergence of amphibians, avians, and mammals. Examination of the DNA sequence of Region 1 from Xenopus laevis H5, human H10, duck H5 and chicken H5 revealed the probable occurrence of many major mutational events. Based primarily upon DNA sequence comparisons, a detailed analysis led to the derivation of an ancestral amphibian H5 sequence from which Region 1 of these H5 genes likely evolved (see text for details). The specific DNA mutations and their order of occurrence, as outlined, includes ancestral avian H5, which represents the gene from which the duck and chicken H5 genes were derived. Note that all sequence modifications within Region 1 of these genes can be directly attributed to either deletions, sequence duplications or single base mutations. MD: mirror duplication; DD: direct duplication; bp: base pair. Single base mutations are not indicated in the scheme.

The following section more clearly documents the details of this analysis, by summarizing the evolution of Region 1 from the ancestral amphibian H5, for each of the five H5 histones. These scenarios are based upon the most likely order of occurrence for each of the mutational events, as can be best determined from the information available at present.

Derivation of the ancestral amphibian H5 (aAMP) is summarized below, in which dots denote deletions and the direct duplicate sequence (duplic) within human H1⁰ is shown below HH1⁰:

```

pAVN ACGGA.....CAGCCCC.....AGCC...AAGCCCAAGCGGCTGAAGGCATC.....GCAC
HH10 ACCGAGAATTCCGCCCT.....GCGGCGAAGCGCAAGCGGCCAAGGCCTCCAAGAAGTCCACAGACCAC
HH10 duplic: TCCACG
XH5A GCAGAGAATTCCAGCCGCTACTCCAGCTGCCAAACCCAAGCGATCAAAGGCATTAAAGAAATCAACTGACCAC
XH5B ACAGAGAATTCCG.....CTQCAGCTGCCAAACCTAAGCGATCAAAGGCATCAAAGAAATCAACAGACCAT
aAMP ACWGAGAATTWGGCCCTACTCCAGCWGCCAAWCCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGACCAC
      T E N S A P T P A A K P K R A K A S K K S T D H

```

1) EVOLUTION OF XENOPUS LAEVIS H5A AND H5B FROM ANCESTRAL AMPHIBIAN-H5

```

      T E N S A P T P A A K P K R A K A S K K S T D H
aAMP ACWGAGAATTWGGCCCTACTCCAGCWGCCAAWCCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGACCAC
pXLH5 ACWGAGAATTWGGCCCTACTCCAGCWGCCAAWCCWAAGCGWTCWAAGGCWTCWAAGAAATCWACAGACCAC
      T E N S A P T P A A K P K R S K A S K K S T D H

```

1 SINGLE BP MUTATION: conversion of A→S

THE PRECURSOR XENOPUS H5 GENE BECAME DUPLICATED, FORMING XLH5A AND XLH5B

XLH5A

3 SINGLE BP MUTATIONS: modification of 3 amino acids

```

      T E N S A P T P A A K P K R S K A S K K S T D H
pXLH5 ACWGAGAATTWGGCCCTACTCCAGCWGCCAAWCCWAAGCGWTCWAAGGCWTCWAAGAAATCWACAGACCAC
      *
      *
      *
XLH5A GCAGAGAATTCCAGCCGCTACTCCAGCTGCCAAACCCAAGCGATCAAAGGCATTAAAGAAATCAACTGACCAC
      A E N S A A T P A A K P K R S K A L K K S T D H

```

XLH5B

T E N S A P T P A A K P K R S K A S K K S T D H
 pXLH5 ACWGAGAATTCWGCCCCCTACTCCAGCWGCCAAWCCWAAGCGWTCWAAGGCWTCWAAGAAATCWACAGACCAC
 <====>

6 BP DELETION: perfect deletion which preserves the alanine;
 net deletion of "PT"

T E N S A P A A K - P K R S K A S K K S T D H
 pXLH5B ACWGAGAATTCWGCTCCAGCWGCCAAWCCWAAGCGWTCWAAGGCWTCWAAGAAATCWACAGACCAC
 *
 XLH5B ACAGAGAATTCGCTCCAGCTGCCAAACCTAGGCGATCAAAGGCATCAAAGAAATCAACAGACCAT
 T E N S A P A A K P R R S K A S K K S T D H

1 SINGLE BP MUTATION: conversion of K->R

2) EVOLUTION OF HUMAN H1⁰ FROM ANCESTRAL AMPHIBIAN H5

T E N S A P T P A A K P K R A K A S K K S T D H
 aAMP ACWGAGAATTCWGCCCCCTACTCCAGCWGCCAAWCCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGACCAC
 <====>

6 BP DELETION: no disruption of amino acid coding, perfect deletion of "TP"

T E N S A P A A K P K R A K A S K K S T D H
 pHH1⁰-1 ACWGAGAAT TCWGCC CCTGCWGCCA WCCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGACCAC
 1----->

6 BP DIRECT DUPLICATION: no disruption of amino acid coding, perfect
 duplication of "SA"

pHH1⁰-2
 T E N S A S A P A A K P K R A K A S K K S T D H
 ----->----->
 ACWGAGAATTCWGCTCWGCCCTGCWGCCA WCCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGACCAC
 * +
 ACCGAGAATTCACGTCGCCCCCTGCGGCCAAGCCCAAGCGGCCAAGGCCTCCAAGAACTCCACAGACCAC
 T E N S T S A P A A K P K R A K A S K K S T D H
 HH1⁰

3 SINGLE BASE MUTATIONS: conversion of A->T

Note that the deletion of "TP" produces an amino acid sequence
 within pHH1⁰-2 that is identical to that produced by the "PT" deletion
 within XLH5B.

3) EVOLUTION OF ANCESTRAL AVIAN H5 FROM ANCESTRAL AMPHIBIAN H5

T E N S A P T P A A K P K R A K A S K K S T D H
 aAMP ACWGAGAATTCWGCCCCCTACTCCAGCGWCCAAWCCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGACCAC
 ~<====>

6 BP DELETION: net deletion of "TP"

T E N S A P A A K P K R A K A S K K S T D H
 pAVN1 ACWGAGAATTCWGCCCCA GCW GCC AAWCCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGACCAC
 <=>or<=>

3 BP DELETION: perfect deletion of either "A"; note that this deletion could have occurred after the 11 bp direct duplication shown below

T E N S A P A K P K R A K A S K K S T D H
 pAVN2 ACWGAGAATT CWGCCCCAGCC AAWCCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGACCAC
 ----->

11 BP DIRECT DUPLICATION: duplication disrupts reading frame

T E N S A P A N P N P ? ? S ? ? R ? ? R N ? N T
 ----->----->
 pAVN3 ACWGAGAATTCWGCCCCAGCCCAGCCCCAGCCAAWCCWAAGCGWGCWAAGGCWTCWAAGAAATCWACAGAC
 <==>

5 BP DELETION: deletion of "NS"; reading frame restoration; conversion of E to D; change in reading within first half of the duplicate producing "APA" conversion to "SPS"; net insertion of "PAPA"

NS
 OLD R F T E A P A * * * * K P K R.....
 T D S P S P A P A K P K R A K A S K K S T D H
 ----->----->
 pAVN4 ACWGACAGCCCCAGCCCCAGCCCCAGCCAAWCCWAAGCGWGCWAAGGCWTC WAAGAAATCWACAGA CCAC
 <=====>

15 BP DELETION: net deletion of "KKSTD" (or 3 bp deletion of "A"; see above)

T D S P S P A P A K P K R A K A S H
 pAVN5 ACWGACAGCCCCAGCCCCAGCCCCAGCCAAWCCWAAGCGWGCWAAGGCWTCACAC
 *
 aAVN ACGGACAGCCCCATCCCAGCCCCAGCCAAGCCCAAGCGGGCGAAGGCATCCCAC
 T D S P I P A P A K P K R A K A S H

1 SINGLE BP MUTATION: net conversion of S->I

Note that the initial deletion of "TP" to form pAVN-1 is identical to that found for pXLH5B and pHH1⁰, although in all three cases, different bases are involved. This indicates that this has been a

preferential mutation during the evolution of H5 subsequent to divergence from aAMP H5. This is summarized below with lower case denoting deleted bases:

NET DELETION OF "TP" FOR ALL THREE DELETIONS

	T E N S A P T P A A K		T E N S A P A A K
pXLH5B	ACWGAGAATTCWGccccctACTCCAGCWGCCAAW	====>	ACWGAGAATTCWGCTCCAGCWGCCAAW
pHH1 ⁰ -1	ACWGAGAATTCWGCCCCCTactccaGCWGCCA AW	====>	ACWGAGAATTCWGCCCCCTGCWGCCA AW
pAVN1	ACWGAGAATTCWGCCCCtactccAGCWGCCAAW	====>	ACWGAGAATTCWGCCCCAGCWGCCAAW

DIVERGENCE OF THE CHICKEN AND DUCK H5 GENES FROM ANCESTRAL AVIAN H5

4) EVOLUTION OF DUCK H5 FROM aAVN H5

	T D S P I P A P A K P K R A K A S H
aAVN	ACGGACAGCCCCATCCCAGCCCCAGCCAAGCCCAAGCGGGCGAAGGCATCCCAC
	+ + +
pDH5-1	ACGGACAGCCCCATCCCGGCCCCGGGCAAGCCCAAGCGGGCGAAGGCTCCCCAC
	T D S P I P A P A K P K R A K A P H

5 SINGLE BP MUTATIONS: net conversion of S->P

	T D S P I P A P A K P K R A K A P H
pDH5-1	ACGGACAGCCCCATCCCGGCCCCGGGCAAGCCC AAGCGGGCGAAGGCTC CCCAC
	----->

15 BP MIRROR DUPLICATION: net insertion of "PRKRA" and conversion of P->T

	T D S P I P A P A K P K R A K A P R K R A T H
	-----><-----
pDH5-2	ACGGACAGCCCCATC CCGGCCCGG GCCAAGCCCAAGCGGGCGAAGGCTCCTCGGAAGCGGGCGACCCAC
	----->

9 BP MIRROR DUPLICATION: perfect duplication; net insert of "APA"

	T D S P I P A P A P A A K P K R A K A P R K R A T H
	-----><-----
pDH5-3	ACGGACAGCCCCATCCCGGCCCGGCCCGGCCCAAGCCCAAGCGGGCGAAGGCTCCTCGGAAGCGGGCGACCCAC
	-----><-----
	ACGGACAGCCCCATCCCGGCCCGGCCCGGCCCAAGCCCAAGCGGGCGAAGGCTCCTCGGAAGCGGGCGTCCCAC
	T D S P I P A P A P A A K P K R A K A P R K P A S H
	DH5

TWO SINGLE BP MUTATIONS: conversion of R->P and T->S

5) EVOLUTION OF CHICKEN H5 FROM aAVN H5

aAVN T D S P I P A P A K P K R A K A S H
ACGGACAGCCCCATCCCAGCCCCAGCCAAGCCCAAGCGGGCGAA GGCATCGCA C
----->

9 BP DIRECT DUPLICATION: net insertion of "QAS"; note the preservation of H

T D S P I P A P A K P K R A K A S Q A S H
----->----->
pCH5-1 ACGGACAGCCCCATCCCAGCCCCAGCCAAGCCCAAGCGGGCGAAGGCATCGCAGGCATCGCAC
*
G

1 SINGLE BP MUTATION: conversion of Q->R

T D S P I P A P A K P K R A K A S R A S H
pCH5-2 ACGGACAGCCCCATCCCAGCCCCAGCCAAGCCCAAGCGGGCGAAGGC ATCGCG GGCATCGCAC
----->

6 BP MIRROR DUPLICATION: imperfect insertion, preserves the "R"; net insertion of "R" and a stop codon causing gene inactivation

T D S P I P A P A K P K R A K A S R R stp.....
-----><-----
pCH5-3 ACGGACAGCCCCATCCCAGCCCCAGCCAAGCCCAAGCGGGCGAAGGCATCGCGGCGCTAGGCATCGCAC
*
C

SINGLE BP MUTATION: A->C converts stop into "S"

T D S P I P A P A K P K R A K A S R R S A S H
pCH5-4 ACGGACAGCCCCATCCCAGCCCCAGCCAAGCCCAAGCGGGCGAAGGCATCGCGGCGCTCGGCATCGCAC
*+
TG

2 SINGLE BP MUTATIONS: conversion of P->L

T D S L I P A P A K P K R A K A S R R S A S H
pCH5-5 ACGGACAGC CTG ATCCCAGCCCCAGCCAAGCCCAAGCGGGCGAAGGCATCGCGGCGCTCGGCATCGCAC
-->

3 BP MIRROR DUPLICATION: perfect insertion of V

pCH5-6
T D S L V I P A P A K P K R A K A S R R S A S H
--><-----
ACGGACAGCCTGG TC ATCCCAGCCCCAGCCAAGCCCAAGCGGGCGAAGGCATCGCGGCGCTCGGCATCGCAC
-->

2 BP MIRROR DUPLICATION: disruption of reading frame:

pCH5-7

T D S L V L S Q P Q P S P S G R R H R G A R H R

-XX-

ACGGACAGCCTGCTCCTATC*CCAGCCCCAGCCAAGCCCAAGCGGGCGAAGGCATCGCGGCGCTCGGCATCGCAC

<==>

1 BP DUPLICATION: insertion of C; restoration of reading frame; net
insertion of "L"; conversion of I->S

pCH5-8

T D S L V L S P A P A K P K R A K A S R R S A S H

ACGGACAACCTGCTCCTATCCCCAGCCCCAGCCAAGCCCAAGCGGGCGAAGGCATCGCGGCGCTCGGCATCGCAC

ACGGAGAGCCTGCTCCTATCCCCAGCCCCAGCCAAGCCCAAGCGGGTGAAGGCATCGCGGCGCTCGGCATCGCAC

T E S L V L S P A P A K P K R V K A S R R S A S H

CH5

2 SINGLE BP MUTATIONS: conversion of D->E and A->V

vi) H5 Amino Acid Sequence Evolution for Region 1

One of the unexpected results produced by the DNA sequence analysis of Region 1 was the derivation of evolutionary intermediates which must closely resemble pre-existing forms of the H5 protein. These are listed in Figure 15.

These predicted intermediates can provide information about the protein evolution of H5 which would be difficult, if not impossible to obtain using other methodologies. Many, for example, likely represent extinct forms of H5. Some however may closely resemble present day forms of H5, whose characterization would not only provide the opportunity to verify occurrence of the proposed DNA sequence modifications, but would also provide the information necessary for more precisely defining the origins and order of their occurrence. One of the best candidates for characterization is turtle H5, in that the reptilian H5 might contain sequences similar to those predicted for the ancestral avian H5.

		# OF AA's
aAMP	TENSAPTPAAKPKRAKASKKSTDH	24
pHH1 ⁰ -1	TENSAPAAKPKRAKASKKSTDH	22
pHH1 ⁰ -2	TENSASAPAAKPKRAKASKKSTDH	24
HH1 ⁰	TENSTSAPAAKPKRAKASKKSTDH	24
pXLH5	TENSAPTPAAKPKRSKASKKSTDH	24
XLH5A	AENSAATPAAKPKRSKALKKSTDH	24
pXLH5B	TENSAPAAKPKRAKASKKSTDH	22
XLH5B	TENSAPAAKPRRSKASKKSTDH	22
PAVN-1	TENSAPAAKPKRAKASKKSTDH	22
PAVN-2	TENSAPAKPKRAKASKKSTDH	21
PAVN-3	INACTIVE	
PAVN-4	TDSPSPAPAKPKRAKASKKSTDH	23
PAVN-5	TDSPSPAPAKPKRAKASH	18
aAVN	TDSPIPAPAKPKRVKASH	18
pDH5-1	TDSPIPAPAKPKRAKAPH	18
pDH5-2	TDSPIPAPAKPKRAKAPRKRATH	23
pDH5-3	TDSPIPAPAPAAKPKRAKAPRKRATH	26
DH5	TDSPIPAPAPAAKPKRAKAPRKPASH	26
pCH5-1	TDSPIPAPAKPKRAKASQASH	21
pCH5-2	TDSPIPAPAKPKRAKASRASH	21
pCH5-3	INACTIVE	
pCH5-4	TDSPIPAPAKPKRAKASRRSASH	23
pCH5-5	TDSPIPAPAKPKRAKASRRSASH	23
pCH5-6	TDSPVIPAPAKPKRAKASRRSASH	24
pCH5-7	INACTIVE	
pCH5-8	TDSPVLSPAPAKPKRAKASRRSASH	25
CH5	TESLVLSPAPAKPKRVKASRRSASH	25

Figure 15. Amino acid sequences of the predicted evolutionary intermediates for Region 1 of H5 and H1⁰. Sequence analysis of Region 1 of Xenopus laevis H5, human H1⁰, duck H5 and chicken H5 led to a detailed proposal for the derivation of Region 1 for these genes from an ancestral amphibian H5 gene. This led to the reconstruction of 21 evolutionary intermediates for H5 (see text for details) which are listed here. The amino acid length of each intermediate is indicated to the right. The three inactive intermediates are products of frame shift mutations within pAVN-3 and pCH5-7, and an amber mutation within pCH5-3.

The alignment shown in Figure 16 attempts to incorporate all predicted mutational events. Unfortunately this form of alignment obscures any conservation of amino acid sequence. To examine amino acid sequence conservation, a second alignment is presented in Figure 17, in which the sequences have been aligned for maximum homology. Note that this type of alignment obscures the underlying DNA sequence modifications which have led to changes in amino acid sequences, illustrating the inability of amino acid sequence to provide reliable information about changes in DNA sequence. Note for example that the conservation of "S" at the fourth position does not represent a conservation of this codon within duck and chicken H5, but is predicted to have been formed within pAVN4, independently from the original amphibian serine codon. A similar situation occurred for the conservation of the histidine within duck H5.

It is also apparent from the alignment shown in Figure 17 that there is an underlying consistency in length variation, with the majority of amino acid deletions and insertions clustered within a hypervariable region located near to the amino terminus. Preferred amino acid sequences are also apparent. Note for example the restoration of "APAAK" within DH5, as well as the independent restoration within duck and chicken H5's of the five amino acids within Section 2, lost during evolution of the aAVN H5. Another important observation is the lack of any unusual amino acids (with the possible exception of the "Q" within pCH5-4), despite the large number of DNA sequence modifications that are predicted to have occurred within this region of H5. This provides indirect support for the validity of the DNA sequence analysis, with no predicted existence of H5 intermediates that would have suffered obvious and dramatic functional modifications.

<=====> hypervariable region

aAMP	TENSA	PTPA	A	KPKRAKASKKSTD	H
pH1 ^o -1	TENSASAPTPA	A		KPKRAKASKKSTD	H
pH1 ^o -2	TENSASAP..A	A		KPKRAKASKKSTD	H
H1 ^o	TENSTSAP..A	A		KPKRAKASKKSTD	H
XLH5A	AENSA	ATPA	A	KPKRSKALKKSTD	H
pXLH5B	TENSA	..PA	A	KPKRAKASKKSTD	H
XLH5B	TENSA	..PA	A	KPRRSKASKKSTD	H
PAVN-1	TENSA	P..A	A	KPKRAKASKKSTD	H
PAVN-2	TENSA	P...	A	KPKRAKASKKSTD	H
PAVN-4	TD..S	P...	SPAP	AKPKRAKASKKSTD	H
PAVN-5	TD..S	P...	SPAP	AKPKRAKAS.....	H
aAVN	TD..S	P...	IPAP	AKPKRVKAS.....	H
PDH5-1	TD..S	P...	IPAP	AKPKRAKAP.....	H
PDH5-2	TD..S	P...	IPAP	AKPKRAKAP.....RKRATH	
PDH5-3	TD..S	P...	IPAPAPA	AKPKRAKAP.....RKRATH	
DH5	TD..S	P...	IPAPAPA	AKPKRAKAP.....RKPASH	
pCH5-1	TD..S	P...	IPAP	AKPKRAKAS.....	QASH
pCH5-2	TD..S	P...	IPAP	AKPKRAKAS.....	RASH
pCH5-4	TD..S	P...	IPAP	AKPKRAKAS.....	RRSASH
pCH5-5	TD..S	L...	IPAP	AKPKRAKAS.....	RRSASH
pCH5-6	TD..S	L...V	IPAP	AKPKRAKAS.....	RRSASH
pCH5-8	TD..S	L...VL	SPAP	AKPKRAKAS.....	RRSASH
CH5	TE..S	L...VL	SPAP	AKPKRVKAS.....	RRSASH

Figure 16. Amino acid sequence alignment of the H5 and H1^o intermediates for Region 1 incorporating all proposed DNA mutations. The amino acid sequences of the intermediates listed in Figure 15 are aligned, with gaps denoting insertions and dots denoting deletions. This comparison demonstrates the complex nature of the DNA sequence modifications that have likely occurred within Region 1 of H5 during the evolutionary history of these species. It also illustrates the difficulty of resolving DNA sequence modifications from amino acid sequence alone.

aAMP	TENS	APTPAAKPKRAKASKKSTDH
pH1 ⁰ -1	TENS	APAAPKPKRAKASKKSTDH
pH1 ⁰ -2	TENS	AS APAAPKPKRAKASKKSTDH
H1 ⁰	TENS	TS APAAPKPKRAKASKKSTDH
XLH5A	AENS	AATPAAKPKRSKALKKSTDH
pXLH5B	TENS	APAAPKPKRAKASKKSTDH
XLH5B	TENS	APAAPKRRSKASKKSTDH
pAVN-1	TENS	APAAPKPKRAKASKKSTDH
pAVN-2	TENS	AP.AKPKRAKASKKSTDH
pAVN-4	TD.SP	SPAP.AKPKRAKASKKSTDH
pAVN-5	TD.SP	SPAP.AKPKRAKAS.....H
aAVN	TD.SP	IPAP.AKPKRVKAS.....H
pDH5-1	TD.SP	IPAP.AKPKRAKAP.....H
pDH5-2	TD.SP	IPAP.AKPKRAKAPRKRATH
pDH5-3	TD.SPI	PIPAPAPAAKPKRAKAPRKRATH
DH5	TD.SPI	PIPAPAPAAKPKRAKAPRKPASH
pCH5-1	TD.SP	IPAP.AKPKRAKAS..QASH
pCH5-2	TD.SP	IPAP.AKPKRAKAS..RASH
pCH5-4	TD.SP	IPAP.AKPKRAKASRRSASH
pCH5-5	TD.SL	IPAP.AKPKRAKASRRSASH
pCH5-6	TD.SLV	IPAP.AKPKRAKASRRSASH
pCH5-8	TD.SLVL	SPAP.AKPKRAKASRRSASH
CH5	TE.SLVL	SPAP.AKPKRVKASRRSASH
		<===> hypervariable region
CONS	TENSPV	SPAPAAKPKRAKASKKSTDH
	AD	LILAAT R V LRRRAS
		AI S P P T
		T Q

Figure 17. Amino acid sequences of the H5 and H1⁰ intermediates for Region 1 aligned for maximum homology. The intermediate sequences listed in Figure 15 are aligned with gaps denoting hypervariable amino acid positions and dots denoting deletion of conserved amino acids. CONS: consensus sequence for Region 1 of H5. This comparison outlines some of the variation and conservation of amino acids proposed to have occurred during the divergence of these linker histones from a common ancestral H5 gene. It is also apparent from this alignment that the majority of length variations occur within a selective site denoted as the hypervariable region.

This excludes the three inactive intermediates, which likely represent a special type of functional modification, as will be described in the Discussion.

vii) DNA Sequence Comparisons of Region 3

Although the success of resolving the mutational events within Region 1 implies that a similar approach would be successful for Region 3, initial examination revealed it to be even more complex. The occurrence of major mutational events within all of the genes was one of the major contributing factors obscuring the sequence analysis, and prevented an extensive analysis of Region 3. A partial analysis was done however, producing the alignment shown in Figure 18. Unfortunately without the additional H5 sequences needed to more clearly resolve the mutational events within Region 3, little can be concluded from this analysis.

E) Comparisons of the 3' and 5' Untranslated Regions of H5 and H1⁰

One of the interesting observations to come from the initial comparison of the Xenopus laevis H5A and H5B cDNA sequences, was the great contrast between the extent of sequence conservation of the 3' and 5' untranslated regions (see Figure 9). The high level of conservation implies that much of the 3' untranslated region of the Xenopus laevis H5 mRNA contain sequences that are important to its function. This proposal is supported by the sequence alignment shown in Figure 19, which outlines several relatively large blocks of conserved sequence within the human H1⁰, duck H5, chicken H5 and Xenopus laevis H5 3' UTR's.

Figure 18. Preliminary DNA sequence alignment for Region 3 of the H5 and H1⁰ coding regions. Preliminary DNA sequence analysis for Region 3 of these five H5 genes produced this alignment, in which gaps have been introduced to compensate for suspected sequence duplications and dots represent potential deleted nucleotides. Mirror and direct duplicates are outlined with arrows. "*" denote nucleotide mismatches which modify amino acid coding and "+" denotes nucleotide mismatches which have no effect due to codon wobble. Extensive DNA sequence modifications, many of them deletions within Xenopus laevis H5, contributed greatly to the complexity of this analysis, which ultimately prevented its completion. HH1⁰: human H1⁰; XLH5A: Xenopus laevis H5A; DH5: duck H5; CH5: chicken H5.

	S V A F	K K	T K K E I K	K V
MH1 ^o	GTCAGTGGCCTTC.....AAGAAG		ACCAAGAAGGAAATCAA.....GAAGGTA	
	+ *		* + + + +	
	---->---> -><-			
	P A K K		P K K E I K K A V	
XLH5A	A... GCA GCA...AAGAAG		CCTAAAAAAGAAATTAAAAAAGCAGTT.....	
		+ *	----->----->	
			+ + + **	
	S P A G R K K K		K K A A R R	
DH5	GTCC CCA GCTGGGAGGAAGAAGAAG..		.AAGAAGGCTGCCCGAAGGT.	
		----->----->		
	*	*		
	S P G K K		K K A V R R	
CH5	GTCC CCC ...GGGAAGAAG ..		.AAGAAGGCCCTCAGGAGGT.	
			+ * + +	

HH1° A T P K K A S K P K K A A S K A P T
 ECACG...CCAAAGAGGCATCCAAGCCCAAGAAGGCTGC.....CTCC....AAAGCCCAACC
 ----->----->
 + + + *** + + * + -><-
 S P K K V A K P K K A A K S P A K A
 XLH5A TCACCCAAAAAGTTGCCAAGCCCAAAAAAGCTGCAAAGTCTCCTGCTAAGG... ..CC
 ----->----->
 + ** + * + * ** + * + + +
 S T S P R K A A R P R K A R S P A K
 DH5 .CCACGTCACCGAGGAAGGCGGCGAGGCCAGGAAGCC...AGGTCGCCGCGGAAG... ..
 ----->----->
 + * + + +
 S T S P K K A A R P R K A R S P A K
 CH5 .CCACGTCCTCCAAGAAGGCGGCGAGGCCAGGAAGCC...AGGTCACCGGCGGAAG... ..
 ----->----->

	K K P K A A P V	K K	A K
HH1°	AAGAAACCCAAAGCCGCCCGGTC	AAGAAG	GCCAAG
	---><---		
	+ ** + **	+ ** +	
	K K P K V A	E K K V K K V A K	
XLH5A	AAGAAACCTAAAGTGGCA	GAGAAGAAGGTGAAGAAAGTAGCTAAG	
	-----***-----><-----	----->-----	
	+ + ** + **		+ *
	K P K A A	A	R K A R
DH5	AAGCCCAAAGCCGCC	GCC	AGGAAA GCCAGG
		-----*****----->	-----
	* *		+
	K P K A T	A	R K A R
CH5	AAGCCCAAAGCCACC	GCC	AGGAAG GCCAGG
		-----*****----->	

Figure 19. Alignment of the H5 and H1⁰ 3' untranslated regions. To examine for possible regulatory sequences that have been conserved during amphibian, avian and mammalian evolution, the 3' untranslated regions of the H5 mRNA's were aligned, with gaps introduced to maximize homology. This reveals several blocks of conserved sequence, the consensus sequences of which are shown below the alignment. There is no evidence for a conserved sequence analogous to a polyadenylation signal within these sequences. Although the site of polyadenylation (denoted by "A_n") of the chicken and *Xenopus* mRNA's have been established from examination of their cDNA sequences, the 3' termini of the duck H5 and human H1⁰ mRNA's are unknown (denoted by "..."). Within the consensus sequences, "X" denotes any base, "." denotes variable base and lower case denotes absence of a base within one or two of the gene sequences.

H: human H1⁰ (Doenecke and Tonjes 1986); A and B: *Xenopus laevis* H5A and H5B respectively (this study); D: duck H5 (Doenecke and Tonjes³ 1984); C: chicken H5 (Kreig et al. 1982). Note that although a putative polyadenylation signal (TATAA) is evident near the 3' end of the chicken H5 and *Xenopus* H5A mRNA's, this putative signal was not identifiable in the other H5 sequences.

A)

H TGA CAA TGAAGTCTTTTCTTGCGGACA CTGCTCCTGTCTCCTA TTTTCTGTAAATAA TTTF
 A TAA AGCTCCTGCAGTCTCT GGA CAG CTCCTAACAAC TTGTGTACATAG TTTF
 B TAA AGCTCCTGCAGTCTCT GGA CAG CTCCTAACAAC TTGTGTACATAG TTTF
 D TGA GAAGTCT GGAGGCCTTTGTCCAGACTCCCCC GTTGGTTTCTGTAAATAGCTTTT
 C TGA GCAGCCT GGGGG CTTTGTCCAGGCTCTCCCCA TTGGTTTCTGTAAATAGCTTTT

GAAGTCT GGA cCAG.CTCCTA tTTCTGTAAATAG.TTTT
 C C G T TCC G C A

H CTCC TTTTTTCTCTCTTGATGCTCACCACCACCT TTTGCCCCCTTCTGTTCTGACTTTTATAA GA
 A ATAACCTTCACTTTTTT TCCAG TGCA ACTTTATAATGA
 B ATAACCTTCACTTTTTT TCCAG TGCA ACTTTATAATGA
 D GCCTTTA TTTT ACCTCTTAC TA.TTGCC TTTTATTTTTTT
 C GCCTTTA TTTT ACCTCTTCTCTATTTGCAA ATTTTATAA G

cCTTCa.TTTTT ACCAC tTGCA aCTTTATAAtG
 CT T TG C T TT T

H GACAGGATTTGGATT CT TCAGAAATT ACAG AATAATTC ATTTTTCCTTAACCAGTTGTG
 A TGATCAGC TTCCCT GCAATTCCC AAGACCAATTTGTGTATTTGACATTGCC TGCAC
 B TGATCAGC TTCCCT GCAATTCCC AAGACCAATTA TGTATTTGACTTTGCC TGCAC
 D GTACGAATCGATCT ATTT CTGTCTGG AAACAGCTAAAA
 C TTGATCT ATT CCT AAGAGCTAAAA

IGATCT...TT..CT aAGAGCaATAA
 TA C C T ATT

H CAAGG ACAGCAACAACCAATGTAATG ATGAGAATGTACTT AT
 A TAGGGGAATTAAGCAAG ATTT AATGG CTTCATTTT ATTAGT ATAA
 B TAGGGGAATTAAGCAAG ATTT AACTG CTTCATTTT AATTA TTAT.
 D TAAGGCAATTAATGAAAGAAAAAGG AACAT CTTCAGCATG AAG AGTTCCC
 C CAAGGCAACGAATGAAAGAAAAA AAGAAACAAAAATGG AACTT CTTCAT ATGGAAG AGTTCCC

TAAGGcaatXAAGCAAX ATGX.AACXT.CTTcAT aag.agtt
 C G . g c C TG AT T G G t

H ATTTTGT TTTGCTATAACCT...
 A CTTTTGTGTTT.(A)_n
 B TTTTGTGT G.(A)_n
 D CTTTTTACGA GCTGCTGGTC...
 C GTTTATAAAA GCT.(A)_n

TTTTGT gct
 ATA

This comparison revealed no evidence of sequences related to the highly conserved 23 base sequence found in H1t, H1 and core mRNAs (Birnstiel et al. 1985, Coles and Wells 1985, Cole et al. 1986). In addition, the human H1^o and Xenopus laevis H5 3' UTR's do not contain sequences related to either the remnants of the 3' dyad symmetry element described in chicken H5 (Coles and Wells 1985) or of the two 3' dyad symmetry sequences described in chicken and duck H5 mRNA's (Kreig et al. 1982, Doenecke and Tonjes 1984, 1986).

Examination of the 5' UTR's of these histones revealed much less sequence conservation, as was suggested by the sequence gaps observed within the 5' UTR's of the closely related Xenopus laevis H5A and H5B mRNA's (see Figure 9). This lack of sequence conservation made an effective sequence alignment of the 5' UTR's very difficult, although as shown in Figure 20, some conserved sequences were observed.

Homology between human H1^o, duck H5, chicken H5 and Xenopus laevis H5 is thus supported further by the sequence homologies found within their 3' and 5' UTR's. Although the function of these conserved sequences are unknown at present, it is possible that they maybe involved in determining some of the distinctive metabolic characteristics shared by these linker histones.

Figure 20. Alignment of the H5 and H1⁰ 5' untranslated regions. Similar to the alignment of the 3' UTRs (Figure 19), alignment of the 5' UTRs revealed several blocks of conserved sequence. The consensus for each block is shown below the alignment. H: human H1⁰; A and B: Xenopus laevis H5A and H5B, respectively; D: duck H5; C: chicken H5.

H CCGCGGCTGTCTCT CA
 A GAGAGAGTGGGCTGTCTACTGAGCAGGAGCTCCACACAAACAAACATTGCGCTCCACTCGACGGGA TCCCTCA
 B TAGAGAGTGTGGCTGTCTACCGAGCGATACCACAAAGCAAACAAACAGTGGCTGCGAGTTGACGGGA TCCTTCA
 D TCGGGGGCGGGCGGGACGGGACGGGGGGGGGACACGGGG GACAGGCAGTCCT CC
 C TCGGGGGCGGGG AGAGGGGGGACACGGG GACAGGCAGTCCT CC

CONSENSUS:

GACGGG..TCCT.CA
 G A CA C

H CCGCGGTCCGCGCGCGCGG TAAATACCGGATGGCGCG CCAAGCGCGG
 A AAGAAGCAAAGTGAAGCTGATCTGCTCACAATTGTACAGACA.CGAGAGTGTCCAGGAAGCCATTG
 B AAGAAGGAAAGGGAAGCCTGCTCTGCTCACAATTCTACACACA.....AAGAAGCCATTG
 D CCGCGGTCCG CACCACACCCTTAAATTC
 C CCGCGGTCCG TGCGGCAC CTAAAT

CONSENSUS:

TAAATTC
 C AG

H ACGCGGAGCTGGGAAAAGGGAGGCAGAGGAGGCGGAGGCAGAGGCAGAGGCCGCTGCGGAGACCA
 A CGAGTCAGGGCTTGAC.ATTTATCCTACAATTTATGCGCGACTGACAAAGAAACAGGTGCAAATTCCAAATA
 B CGAGTCAGGGCTTGAC....TATTCTAGAACATATCCTCAACTGACGAGGAAACAGGTGCAAATTCCAAACA
 D
 C

H CGACAGACCGGCGGGGCTGGGCCTCGAAAGCCGGCTCGGCGAGCTCTCCCGACACCCGAGCCGGGGAGGAAA
 A CTTGGG..... AGTGCAAGAATAACTTCACTAT.....TTCTCAGGACTGACTGT
 B CTCAGG.CACAAAAGCC AAATATAGAATAACTTCACTAT.TTGGAGAACTGCAGTTGAGAGAACTGACTGA
 D GGGGCAG
 C GCGTCTGCTGGTGGCGA

H AGCAGCGACTCCTCGCTCGCATCCCCGGGAGCCGCACTCCAGACTGGCCCGGTAGTCAGGGGCTCAGGAGCAGAT
 A ATTTGGAAAGGTCACTCCAGCCATCTCTGAGAACTTTGTTACCAAGCTCATATCTGCC
 B ATTTGGAAAGGTCACTCTAG.CCATCTCTGAGAACTTTGTTACCGGGCTCCTATTGTCCATTGAATTCCAAC
 D
 C

H CCGAGGCAGG CTTTC TC
 A ACAGG TTGC TGGG
 B CTCAGTAACCAGCCTCTTTGCATCTTGCTGACAGG TTGC TGGG
 D CGCCGGGTGCGCCGAGAGGCGGCGGAGAGGCAGCAGAGGTGGCGGCAACGGTGCCCCGTGTTGCCTG
 C CGCGCGGCCGAGACGCAGCGGCGGCGGCAGCAGGAGCA GCCCCACATCCGGTTGT TGCTGGC

CONSENSUS:

GCAGG
 a c

TTGC.TG
 t c

H AGCCTCCGACGAGGGCTGGCCCTTGGAAGGCG
 A TCTAACCTCTTGCTTCACCGAAA
 B AAAGCCTTGCCTTGGAATAAGCTCCTGCTGCACTGCAA
 D TAGCTCCTTCTCGTAGCCTGCCCGGGTGGCCCCGGTGGCAGCTGGCCGACCCGG.
 C GGCTCCTTT

CONSENSUS: AGCTCCTG
 GC CTTCT

H	CCTTCAA	CAGCCGG	ACCAGACA	GGCCACCATG
A	CTTTCATTCTTCAGACTACACCTGAA			GTGAAGGTTA TAA AATG
B	GTTTCACTCTTGAGACTACACCTGAACC			GAAG TAACAATG
D	TCCCGGTTTCCACCG CAG CAGCACCGGGA			GAAGACTGGC CATG
C	TTTAA GCTCCCTA ACCCCAGTGCCCTGCCGTGGGCTGAAGC			GGCGG CATG

CONS:	TTTCA	ACACXGA	GAAG	CATG
	C AC	G CG		A

F) Amino Acid Sequence Comparisons of H1 and H5

Although it has been well established that H1 and H5 are closely related proteins, much evidence has been reported indicating that the H5 histones are functionally distinct from H1 (see Introduction). To further investigate the functional relationship between H1 and H5 linker histones, an amino acid sequence comparison was conducted. It was necessary to restrict this analysis to the highly conserved central region (equivalent to Region 2 from Figure 12) because of considerable variation in length and sequence within the amino and carboxyl regions of H1, as noted previously for Regions 1 and 3 of the H5 histones (see Figure 12). It should be noted that examination of the amino amino acids within the carboxyl termini of mammalian, avian and Xenopus laevis H1's failed to reveal a conserved region comparable to Region 4 of H5, suggesting that this conserved region may be characteristic of the H5 histones.

Figure 21 presents a compilation of mammalian, avian and Xenopus laevis H1 amino acid sequences, from which a consensus sequence for Region 2 was derived. The consensus sequences for Region 2 of H1 and H5 are compared in Figure 22, revealing an overall homology of 51%. This comparison also reveals however, that the sequence homologies are not randomly distributed, since nearly half of the mismatches are clustered within a central 21 amino acid segment. According to the asymmetrical distribution of amino acid homologies, Region 2 of H1 and H5 was subdivided into three sections.

As outlined in Figure 22, Sections 1 and 3 possess homologies of 64% and 74% respectively, as compared to only 14% within Section 2. It is important to note that although there is little sequence homology between Section 2 of H1 and H5, this segment does in fact contain amino

A) GPPVSELITKAVPASKERN
 B) GPPVSELITKAVAASKERSGVSLAALKKALAAAGYDVEKNNSRIKGLKSLVSKGTLVQTKGTGASGSF
 C) GPPVSELITKAVAASKERSGVSLAALKKALAAAGYDVEKNNSR
 D) GPPVSELITKAVAASKERNGLSLAALKKALAAAGYDVEKNNSRIKGLKSLVSKGTLVETKGTGASGSFKLDKK
 E) GPPVSELITKAVAASKERSGVSLAALKKALAAAGY
 F) GP.VSELITKAVSASKXXXXXXXXXXAAGGYDVEKGNSR
 G) GPSVTELITKAVSASKERKGLSLAALKKALAAAGYDVEKNNSRIKGLKSLVSKGTLVQTKGTGASGSFRLNKK
 H) GPSVTELITKAVSASKERKGLSLAALKKALAAAGYDVEKNNSRIKGLKSLVSKGTLVQTKGTGASGSFKLNKK
 I) GPSASELIVKSVSASKERGGVSLAALKKALAAAGYDVERNNNSRLKALKALVTGTLTQVKGSGASGSFKLNKK
 J) GPSASELIVKSVSASKERGGVSLAALKKALAAAGYDVERNNNSRLKALKALVTGTLTQVKGSGASGSFKLNKK
 K) GPSVSELITKAVSASKERSGVSLAALKKALAAAGYDVEKNNSRIKGLKSLVSKGTLVQTKGTGASGSFKLNKK
 PAT V S A G L A N R G L A A T T E V S R D
 P N
 K

Figure 21. Amino acid sequence alignment of Region 2 from several vertebrate H1's. As an initial step for a comparative amino acid sequence analysis of vertebrate H1 and H5, a consensus sequence was derived for Region 2 of H1 by alignment of several vertebrate H1 histones. A) human H1 (Wells 1986); B) bovine H1 (Wells 1986); C), D), and E) rabbit H1-2, H1-3, H1-4 respectively (von Holt et al. 1979); F) chicken H1 CEL-5 (von Holt et al. 1979); G) and H), chicken H1-O1 and H1A respectively (Wells 1986); J) and K) Xenopus laevis H1C1 and H1C2 respectively (Wells 1986).

acids that are highly conserved within each of these subclasses. The vertebrate H1 and H5 histones thus contain a highly conserved 21 amino acid sequence which is distinctive to each of these linker histone subclasses.

The clear distinction demonstrated between H1 and H5 by the amino acid sequence comparison shown in Figure 22, suggests that homologies with other species of linker histones could be effectively examined through comparisons of the conserved Region 2. To evaluate this approach, trout H1 and mammalian H1t amino acid sequences were first compared to the vertebrate H1 and H5 consensus sequences. This analysis which is presented in Figure 23, demonstrates that the trout H1's contain extensive homology with the vertebrate H1 consensus sequence. As summarized in Table 3, the trout H1's could therefore be clearly classified as a typical vertebrate H1.

H1t was also found to have extensive homologies with the vertebrate H1 consensus sequence, showing 59%, 100% and 85% homology respectively within Sections 1, 2 and 3 of Region 2. Section 1 does however, contain several contrasting amino acid differences, including two acidic to basic, and one basic to acidic amino acid conversions. Section 2 of Region 2 apparently does not contain amino acid sequences that are unique to vertebrate H1, as was suggested by the comparison between H5 and vertebrate H1. It also suggests that Section 1 may contain amino acids that may be functionally distinctive.

Several nonvertebrate linker histone comparisons are presented in Figure 24. It is evident from these comparisons that no nonvertebrate linker histones examined possesses any clear relationship to either vertebrate H1 or H5, showing only limited sequence homology as is summarized in Table 3.

Figure 23. Amino acid sequence comparisons of the consensus sequences of Region 2 from vertebrate H1 and H5 with trout testis H1 and mammalian H1t. To evaluate the amino acid sequence homologies the consensus sequence of vertebrate H1 and H5 (Figure 22) are compared to Region 2 from A) trout testis H1 (Macleod et al. 1977), B) trout H1 (Mezquita et al. 1985) C) boar H1t (Cole et al. 1984) and D) rat H1t (Cole et al. 1986). In these comparisons "+" denotes amino acids that are common to both vertebrate H1 and H5, and "*" denotes amino acids unique to either vertebrate H1 or H5. The percent homologies for each section are summarized in Table 3.

Figure 24. Amino acid sequence comparisons of Region 2 of several nonvertebrate H1's with the consensus sequences of H5 and vertebrate H1. To evaluate the amino acid sequence homologies between the vertebrate and nonvertebrate linker histones, the consensus sequences of vertebrate H1 and H5 from Figure 22 are compared to Region 2 from several nonvertebrate linker histones. In these comparisons dots denote gaps introduced to maximize sequence homology, "+" denotes an amino acid that is common to vertebrate H1 and H5, and "*" denotes an amino acid unique to either vertebrate H1 or H5. The percent homologies for each section of H1 and H5 are summarized in Table 3. A) Echinolampus H1 (Wells 1986), B) Sea urchin H1-Sp (Wells 1986), C) Parachinus H1 (Wells 1986), D) Drosophila H1 (Levy et al. 1982, Barbero et al. 1982, Wells 1986), E) Maize H1 (Hurley and Stout 1980), F) Platyneria H1A (Kmieciak et al. 1985), G) Platyneria H1B (Kmieciak et al. 1985).

E) Maize H1

H5

```

P
S   L   K   A           H           A T           G
T E AD VR D N S T     V N T QH LE     A R LASSA I     A
HPKYSDMIVAAIQAESRGGSS RQSIQKYIKSHYKVGENADSQ IKLSIKRLVTTGVLKQTKGVGASGSFRLAKS
* + + + * + + + + * + + + + * + + + + * + + + +
YAEMISEAITSKERTGSS ZYAIKAYVEDKHDKLPXDFX XXXXXXXXXXXXXXXXXXXXXXXXFKLPXX
* + + * + + + + * + + * + + + + * + + + +
GPSVSELITKAVSASKERSGVs LAALKKALAAGGYDVEKNNSR IKLALKSLVTGKTLVQTKGTGASGSFKLNKK
PAT V S A G L A N RG L G A S TEV S R D
P N
K

```

H1

F) Platyneria H1A

H5

```

P
S   L   K   A           H           A T           G
T E AD VR D N S T     V N T QH LE     A R LASSA I     A
HPKYSDMIVAAIQAESRGGSS RQSIQKYIKSHYKVGENADSQ. IKLSIKRLVTTGVLKQTKGVGASGSFRLAKS
* + + * + + + * + + + * + + + * + + + + * + + + +
HPPVATMVVTTAILGLKDRKGSS MVAIKKYIAANYRVDVARLAPF IRKFIRKAV.....KQTKGTGASGSFVRNKT
* + + * + + * + + + * + + * + + * + + + + * + + + +
GPSVSELITKAVSASKERSGVs LAALKKALAAGGYDVEKNNSR. IKLALKSLVTGKTLVQTKGTGASGSFKLNKK
PAT V S A G L A N RG L G A S TEV S R D
P N
K

```

H1

G) Platyneria H1B

H5

```

P
S   L   K   A           H           A T           G
T E AD VR D N S T     V N T QH LE     A R LASSA I     A
HPKYSDMIVAAIQAESRGGSS RQSIQKYIKSHYKVGENADSQ. IKLSIKRLVTTGVLKQTKGVGASGSFRLAKS
* + + * + + + * + + + * + + + * + + + + * + + + +
HPPVATAVVAAILGLKDRKGSS MVAIKKYLAANYRVDVARLGPF IRRFVRKAVAEGLVQNKG.....SFRVNKT
* + + * + + * + + + * + + * + + * + + + + * + + + +
GPSVSELITKAVSASKERSGVs LAALKKALAAGGYDVEKNNSR. IKLALKSLVTGKTLVQTKGTGASGSFKLNKK
PAT V S A G L A N RG L G A S TEV S R D
P N
K

```

H1

TABLE 3: Summary of Region 2 Amino Acid Sequence Homologies Between
Vertebrate and Nonvertebrate Linker Histones

SECTION: SPECIES*	#1		#2		#3	
	H5	H1	H5	H1	H5	H1
H1 vs H5**	64%		14%		74%	
TROUT TESTIS H1	45%	82%	14%	95%	64%	87%
TROUT H1	48%	96%	14%	95%	64%	90%
BOAR H1t	36%	59%	14%	100%	61%	87%
RAT H1t	36%	59%	14%	100%	74%	84%
ECHINOLAMPUS	68%	59%	28%	14%	51%	48%
SEA URCHIN-SP	59%	45%	33%	24%	51%	75%
PARACHINIUS	64%	50%	24%	19%	55%	45%
DROSOPHILA	59%	45%	25%	28%	48%	58%
MAIZE	58%	50%	30%	15%	---	---
PLATYNERIA-A	54%	45%	28%	24%	61%	52%
PLATYERNIA-B	54%	45%	24%	28%	45%	42%

* from Figure 23 and 24

** from Figure 22

IV. DISCUSSION

A) The Molecular Cloning of cDNA

r) A Technique for the Rapid Synthesis and Cloning of dscDNA.

One of the most successful techniques employed during this project was a modification to the Gubler-Hoffman procedure, which reduced the preparation time of blunt-ended, double-stranded cDNA to less than three hours. This modification entailed elimination of a phenol extraction and ethanol precipitation before second-strand synthesis, producing a single, two-step reaction. Another important variation was the use of RNA preparations containing large amounts of rRNA, which circumvented poly(A)⁺ RNA isolation and allowed efficient synthesis and recovery of even small amounts of dscDNA. Although not employed in this project, replacement of oligo(dT) by random primers could also allow the preparation of dscDNA from poly(A)⁻ RNA, in addition to promoting synthesis of very large dscDNA molecules (R.G. Rutledge, V.L. Seligy, M.-J. Cote, K. Dimock, L.L. Lewin and M.P. Tenniswood manuscript in preparation). Finally, because of the speed and ease of this method, mRNA preparations can be quickly examined by preparing electrophoretic profiles of their dscDNA.

Another valuable technique employed in this project was the use of blunt-end ligation for the construction of cDNA banks. This approach avoided difficult, sometimes ineffective manipulations, such as linker addition or homopolymer tailing, and dramatically reduced the preparation time of the cDNA banks. When combined with the modified Gubler-Hoffman procedure, this allowed cDNA banks to be constructed from RNA in less than 24 hours. One of the potential disadvantages of this approach is the inefficiency of blunt-end ligation as compared to

cohesive-end ligation (Maniatis 1982). This did not prove to be a problem, however, in that the blunt-end ligations here consistently produced 500-1000 clones per ng dsDNA, an efficiency that is more than sufficient for preparing even very large cDNA banks.

ii) The Use of M13 as a Primary Cloning Vector

The choice of M13mp9 as the primary cloning vector for Xenopus laevis erythrocyte cDNA was based in part upon the versatility of M13 (Messing 1983). In addition to offering the advantages of a viral cloning vector (e.g. circumvention of antibiotic selection and ease of clone storage), M13 also provides, through the ability to prepare double-stranded "replicative form" DNA, all of the advantages of a plasmid cloning vector. Most important however, cloning with M13 provides several exceptional advantages related to the production of single-stranded viral DNA in the absence of host lysis. For example, the production of single-stranded DNA in addition to producing efficient plaque hybridizations, provides the ability to select clone inserts with a specific orientation. This can be of particular importance in the isolation and characterization of cDNA clones, as was illustrated in the Results section. Host viability facilitates the preparation of large numbers of high-titer culture supernatants, which allows the rapid electrophoretic characterization and purification of single-stranded viral DNA from a large number of M13 clones.

One of the greatest advantages of M13 is the ability to use its single-stranded viral DNA in a variety of DNA polymerase reactions. The most prominent application has been DNA sequencing using the dideoxynucleotide method, for which the M13mp cloning vectors were originally developed (Messing 1983). With the deletion technique of Dale

et al. (1985), the DNA sequence of a large number of clones can be rapidly determined, as described for the H5 cDNA clones (see Methods and Results). M13mp9 viral DNA was also used in this project to prepare single-stranded probes of high-specific activity, needed for genomic DNA blot analysis and M13 plaque hybridizations. Using a similar method, Burke et al. (1984) have also used M13mp viral DNA for preparing probes for high-sensitivity S1 mapping.

Although not employed during this project, M13 viral DNA has been effectively used in hybridization-arrest translation and hybridization-selection analysis of insert DNA (Chandler 1984), and for the in vitro mutagenesis of DNA (Messing 1983, Kramer et al. 1984, Zoller and Smith 1984, Nisbet and Beilharz 1985). M13mp can also serve as efficient expression vectors (Slocombe et al. 1982, Georges et al. 1984, Lathe et al. 1984). Thus, in addition to being a versatile cloning vector, M13 can also be exploited in a variety of ways to provide often superior methodologies for the molecular cloning and analysis of DNA.

B) Isolation and Identification of cDNA Clones Coding for Two Variants of Xenopus laevis H5

The isolation of XLH5A-1 and XLH5A-2 described in Results, is the first reported demonstration of immunological screening of M13mp9 cDNA clones. Their successful identification depended in part upon the antigenicity of the central globular region (Region 2), which is known to be relatively resistant to protease degradation (Cary et al. 1981), and could therefore have contributed to the stability of the H5 peptide within E. coli. (Positive identification of XLH5A and XLH5B was provided by amino acid sequence comparisons which indicated that they encode for proteins highly homologous to both mammalian H1⁰ and avian H5.

(Figure 12). Combined with the extensive homology within their 5' and 3' UTRs (Figures 19 and 20), the evidence clearly established that XLH5A and XLH5B are homologous to avian H5 and mammalian H1⁰.

Several lines of evidence further indicate that XLH5A and XLH5B code for the two H5 proteins variants previously identified in the tissues of Xenopus laevis (called H1^S by Brown et al. 1981, H1D and H1E by Risley and Eckhardt 1981; see Introduction). First, the initial identification of XLH5A-1 and -2 by use of a polyclonal antibody selective for H5 (see Methods) indirectly links the peptides encoded by these cDNAs to H5 protein. Second is the prediction within the XLH5A cDNA of a nine-residue sequence which is identical to that previously determined from Xenopus laevis erythrocyte H5 (Brown 1982; Figure 9), but which differs in detail in the four other H5 sequences (Figure 12). Third, as summarized in Table 2, the amino acid composition of the predicted proteins from both XLH5A and XLH5B agree with the previously determined amino acid composition of H5 purified Xenopus laevis erythrocytes and liver (Brown et al. 1981), although the differences in the reported compositions of H5 from liver and erythrocytes do not precisely reflect the putative differences in the proportions of H5A and H5B within these tissues. Fourth, the two additional amino acids within the amino terminus XLH5A agrees with the slightly lower electrophoretic mobility in SDS of H5A as compared to H5B (Brown et al. 1981, Risley and Eckhardt 1981), although the molecular weight differences calculated by Risley and Eckhardt (1981) would predict a difference of more than two amino acids. Fifth, Brown (1982) and Smith et al. (1984a) both reported that the electrophoretic differences between the two H5 variants were due to a small amino terminal segment that was lost after cyanogen bromide cleavage of methionine. The present results

confirm that the XLH5A and XLH5B proteins would in each case, produce a single large peptide fragment of identical length after cleavage of methionine. Sixth, Risley and Eckhardt (1981) concluded from V8 protease cleavage mapping that H1D and H1E have very similar amino acid sequences, consistent with the extensive similarity in the amino acid sequences predicted by XLH5A and XLH5B. Seventh, the identification of only two H5 genes by genomic DNA blot analysis (Figure 10) is consistent with the characterization of only two protein variants of H5 within Xenopus laevis (Brown et al. 1981, Risley Eckhardt 1981, Smith et al. 1984a). Eighth the recovery of a single XLH5B cDNA clone from the S2 bank relative to twelve XLH5A cDNA clones, is consistent with reports that the amount of H5B protein in erythrocytes is about ten times lower than H5A (Brown et al. 1981, Risley and Eckhardt 1981).

In addition to providing a preliminary characterization of the H5A and H5B genes, DNA blot analysis gave no obvious indication that these genes contain sizeable introns. The analysis also revealed differences in the hybridization signals produced by each gene (see Figure 10). This could be due to reduced hybridization efficiency produced by the 8% mismatch between the XLH5A-1 probe and the H5B gene. It is also possible that H5A has a higher gene copy number than H5B. The very low degree of sequence homology between X. laevis H5 genes and those of X. laevis H1 (Turner et al. 1983, Perry et al. 1985), along with the differences in the respective endonuclease restriction site patterns of these genes (Turner et al. 1983, Perry et al. 1985), also indicates that the genomic sequences detected by the XLH5A-1 probe are those of H5A and H5B.

C) Reclassification of Mammalian H1⁰ and Amphibian H1^S into the Same Linker Histone Subclass as Avian H5

As was outlined in the Introduction, studies on the metabolic and structural characteristics of amphibian H1^S, avian H5 and mammalian H1⁰ have revealed extensive homologies. The sequence characterization of the Xenopus laevis H5A and H5B cDNA's has extended these homologies to the nucleic acid sequence level, and has provided substantial evidence for including mammalian H1⁰, amphibian H1^S and avian H5 into the same linker histone subclass. This evidence, which also clearly distinguishes these linker histones from H1, encompasses three major aspects of mRNA sequence.

1) The presence of several blocks of homology within the 3' and 5' untranslated regions of the Xenopus laevis H5, avian H5 and mammalian H1⁰ mRNA's (Figure 19 and 20). Although the function of these conserved nucleotide sequences has not been established, they may be involved in dictating some of the metabolic characteristics that distinguish H5 from vertebrate H1. This is consistent with the absence within these mRNA's of the conserved 23 nucleotide dyad symmetry element found within the 3' UTR of H1 and core mRNAs (Birnstiel et al. 1985, Coles and Wells 1985). The polyadenylation of Xenopus laevis H5 mRNA (see Results) demonstrates another common feature with chicken H5 (Molgaard et al. 1980) and human H1⁰ (Kress et al. 1986), which also distinguishes these linker histones from H1.

2) The comparative DNA sequence analysis (to be discussed in detail in a later section) provided evidence that amphibian H5, mammalian H1⁰ and avian H5 likely evolved from a common ancestral H5 gene.

3) The demonstration of extensive amino acid sequence homologies between mammalian H1⁰, amphibian H5 and avian H5 (Figure 12). This

included a highly conserved 21 amino acid segment within Region 2 of these proteins (Section 2 of Figure 22), which differs greatly from its analogous segment within the vertebrate H1 histones. A conserved carboxyl segment (Region 4 in Figure 12), that is absent within the vertebrate H1's, also implies that these linker histones differ significantly in function from vertebrate H1.

The extensive and substantial homology in the regulation of expression, amino acid and mRNA sequence, and common genetic ancestry, argues that avian H5, mammalian H1^O and amphibian H1^S should be placed into the same linker histone subclass. In view of the extensive differences between these linker histones and H1, it is apparent that the initial distinction given to avian H5 by its placement into a separate histone group was particularly justified, and that this distinction should be extended to mammalian H1^O and amphibian H1^S. In addition to more clearly delineating their homologies, this reclassification would also more properly distinguish these linker histones from H1. Specifically, the names "H1^O" and "H1^S" over-emphasize their homology with H1. The reclassification of H1^O and H1^S into the H5 subclass would thus not only correct this inaccuracy but would also reserve this form of nomenclature for such linker histones as H1t, which have been shown to be closely related to H1.

D) Amino Acid Sequence Comparisons of the H5 Histone Proteins

Much can be learned about the functional and structural characteristics of a protein, by comparing the amino acid sequence of its homologues from several different animal species. By determining the conservation or modification of specific amino acids and the sites of peptide length variation, the evolutionary constraints that have

acted upon the amino acids of a protein can be inferred, and the potential significance of specific changes in amino acid sequence can be evaluated. This approach, illustrated by the amino acid sequence comparisons of H5 outlined in Figure 12, established that variations in peptide length within the H5 histones are restricted to two specific areas. These are adjacent to segments containing conserved amino acid sequences, whose lengths have likely been invariant during the evolution of mammals and birds. The nonrandom distribution of variations in amino acid sequence and peptide length, indicates that the function of H5 has placed specific regional demands upon changes to its protein structure. H5 can thus be subdivided into four regions (outlined in Figure 12), each possessing distinctive characteristics which likely contribute to different aspects of H5 function.

The absence within vertebrate H1 of a carboxyl segment equivalent to Region 4, indicates that this segment may be involved in determining functional differences between vertebrate H1 and H5. However, it is not clear whether Region 4 is exclusive to H5, because no detailed examination of linker histones other than vertebrate H1 was conducted. This is in contrast to Region 3 of H5, whose variability of length and general amino acid composition is very similar to the carboxyl region of the vertebrate H1's, and suggests that a domain analogous to Region 3 of H5 is an innate characteristic of other linker histones.

It can be further concluded that H5 function may, to some extent, be insensitive to length variations within Region 3. This proposal is based upon the length differences within Region 3, observed between very closely related H5 histones (e.g. duck and chicken H5, see Figure 12), and between the carboxyl regions of recently duplicated linker histone variants (e.g. Xenopus laevis H1C1 and H1C2, Turner et al. 1983). On

the assumption that significant changes in function could not have evolved between such closely related proteins (see later discussion), it appears that variations in peptide length within Region 3 (or its equivalent) likely do not initiate large changes in linker histone function. Unfortunately, the great variability in the sequence and length of Region 3 makes it difficult to conduct an effective amino acid sequence comparison and to resolve other functionally significant aspects within this segment of H5.

This variability is in contrast to Region 2 whose universal presence, relatively invariant length and high level of amino acid sequence conservation, allows effective amino acid comparisons between even greatly divergent linker histones. Most striking is the extensive amino acid sequence conservation, which includes amino acids that have been preserved over extended evolutionary periods, as is seen for the hexapeptide ending in phenylalanine near to the carboxyl end of Region 2. This implies that the function of Region 2 is both critical and in some aspects universal to the linker histones. Importantly, amino acid sequence comparisons have also revealed significant amino acid changes within Region 2 (see Figures 23 and 24), indicating that the function of Region 2 has likely not been entirely invariant during eukaryotic evolution. The difficulty thus lies in distinguishing those changes that signify distinct differences in function from those that represent simple amino acid drift.

This issue has been addressed to some extent by comparing Region 2 from vertebrate H1 and H5, which demonstrated that Region 2 can be subdivided into three sections (see Figure 22). Specifically, vertebrate H1 and H5 were found to share a 64% and 71% homology within Sections 1 and 3 respectively, in contrast to only 14% homology within

Section 2. However, despite their lack of homology, the amino acids of Section 2 have been highly conserved within each subclass. This indicates that these amino acids must have a critical function which is most certainly different within vertebrate H1 and H5. This analysis also demonstrates that a clear distinction between vertebrate H1 and H5 can be made even from such a limited amino acid comparison, and suggests that comparisons of Region 2 could be used to distinguish other functional forms of linker histone.

Further comparisons shown in Figure 23 confirmed that trout H1 could in fact be clearly recognized as a typical vertebrate H1. In contrast, it was expected that Region 2 of H1t would be distinctly different, based upon the reports that H1t is antigenically distinct from other vertebrate H1's, avian H5 and mammalian H1⁰ and is present only in mammalian testis (Markose et al. 1985). Comparisons (Figure 23) revealed however, that Region 2 of H1t has extensive homologies with the vertebrate H1 consensus sequence, sharing 59%, 100% and 85% of the residues within Sections 1, 2 and 3 respectively. Although some important differences are observed within Section 1, the evidence clearly indicates that mammalian H1t and H1 are more closely related than either is with H5. This is consistent with the presence in the 3' UTR of the H1t mRNA, of the 23 bp dyad symmetry element found within H1 mRNA (Cole et al. 1986). Therefore, it could be concluded from this evidence that H1t is a functionally divergent form of H1.

The comparison was also expanded to include Region 2 of several nonvertebrate H1's (Figure 24). No dramatic distinctions were observed however, providing no evidence for direct homology between vertebrate H1 or H5 and any of the nonvertebrate linker histones examined. Thus although the examination of Region 2 has been successful for analysis of

the vertebrate linker histones, it is apparent that further work is necessary before the significance of amino acid differences within Region 2 of nonvertebrates can be more precisely resolved. Most valuable would be the development of an assay for directly comparing linker histone function, that would allow the functional significance of specific differences in amino acid sequence to be determined.

The comparative DNA sequence analysis conducted on the portion of the H5 gene coding for Region 1 (see Results), provided additional information about the possible evolution of H5. From this analysis the DNA sequence evolution of Region 1 could be traced back to a common ancestral H5 gene, and led to the derivation of 18 evolutionary intermediates of Region 1 which could represent pre-existing forms of H5. Examination of these intermediate forms revealed specific changes in the sequence of Region 1, that have likely occurred during the evolution of avian H5, mammalian H1⁰ and Xenopus laevis H5 (see Figure 17). These predicted changes included extensive variation in the length of Region 1, which is consistent with the reported differences in the amino termini of H5 within several bird species (Seligy et al. 1976) and with the differences observed between Region 1 of human H1⁰, Xenopus laevis H5, duck H5 and chicken H5 (Figure 12). It was also found that the majority of these variations in length were restricted to a single site located close to the amino terminus of H5. Importantly, it could be concluded from these results that the function of H5 tolerates at least small changes in length when they occur at this site within Region 1, similar to the functional tolerance of Region 3 discussed earlier.

One unusual aspect in the evolution of Region 1 of avian H5, is the predicted occurrence of a five amino acid deletion within a segment that has been absolutely preserved between amphibian H5 and mammalian H1⁰ (refer to pAVN-5 of Figure 17). Conservation of these amino acids within human H1⁰ and X. laevis H5 suggests that they have a function which is under strong evolutionary selection, and implies that their deletion would have affected the proper function of the precursor avian H5. Consistent with this is the independent selection within duck and chicken H5 of mutations that have produced insertions that are strikingly similar to the deleted amino acids. This suggests that any loss in function produced by this deletion has at least in part been restored within duck and chicken H5. Unfortunately, without the ability to effectively determine the precise functional effects of changes in amino acid sequence, the true significance of these mutational events cannot be established at present.

E) Investigation of the Mechanism for Acid Sequence Modification

In an attempt to resolve some of the mutational events that have led to changes in amino acid sequence of the H5 histones, a comparative DNA sequence analysis was conducted on the coding regions of several H5 genes. As described in Results this led to the identification of two unusual types of DNA mutations, which take the form of sequence duplications. These were called mirror and direct duplications to signify their structural forms (for details see Results). Although the mechanism for these duplication events was not established, the DNA sequence comparisons of the H5 histones did provide evidence for the frequent occurrence of these mutational processes. Importantly, no examples for the insert of foreign DNA were identified, implying that most, if not all amino acid insertions that have

occurred within these linker histones were products of DNA sequence duplications. Evidence for "tandem repeats" has also been reported by several other groups (Turner et al. 1983, Foran et al. 1985, Boshardt and zur Hausen 1986) as well as evidence for "intra-genic" duplications from examination of protein sequences (Barker et al. 1978).

Examination of the H5 coding region thus revealed only four major mutational processes that have led to amino acid sequence modifications: 1) single base mutations, 2) deletions, 3) direct duplication and 4) mirror duplication. Once these mutational processes were identified, the occurrence of specific DNA sequence modifications could be demonstrated by comparing the gene sequences of the most closely related H5's. This in turn led to the development of a methodology (to be summarized in the next section) that produced detailed scenarios reconstructing the DNA sequence evolution of Region 1 for each of the H5 histones examined.

One of the most difficult aspects of this analysis was in resolving the order of the mutational events, which in many cases could not be firmly established. Importantly, the 18 predicted intermediate forms of Region 1, whose amino acid sequences likely represent preexisting forms of H5, provide testable products for confirming the accuracy of the analysis. The most direct approach would be the sequence characterization of other present day forms of H5, and to compare their structures with the predicted evolutionary intermediates of Region 1. This could not only provide confirmation for the occurrence of specific mutational events, but could as well establish more accurately the order of their occurrence.

A less direct approach is to examine these intermediates for any unusual or inconsistent amino acid sequences, which could signify dramatic changes in function. This is based upon the assumption that the preservation of protein function is a major factor for selecting changes in

amino acid sequence. Specifically, it would be highly unlikely that functionally impaired forms of H5 would survive and become fixed within the parental population. Even when the modified gene is heterozygous with an intact H5 gene, the expression and incorporation into chromatin of a functionally impaired H5 protein, would most certainly disrupt chromatin function. Functional impairment would thus most likely constitute a constitutive lethal, which would immediately lead to extinction of the modified gene.

Although the effects upon protein function of some mutations are apparent, such as for frame shift mutations, the effects of others, such as single amino acid changes or modification of peptide length, can be more difficult to evaluate. However, comparison of these intermediate sequences did reveal a high level of consistency in the form and site of the amino acid modifications produced by the predicted DNA mutations (Figure 17). No highly unusual intermediates for Region 1 of H5 were predicted, although some potential modification of function was evident. As was discussed previously, the 15 bp deletion within Section 2 of pAVN-4 may have affected H5 protein function. It is certainly significant however, that none of the predicted intermediates contain highly unusual amino acids such as phenylalanine, methionine or tryptophan, which would be expected to drastically disrupt the proper function of H5.

The prediction of three inactive H5 intermediates requires special consideration, in that they represent a unique form of functional disruption. Specifically, the occurrence of frame shift and nonsense mutations at a site close to the amino terminus would produce complete loss of protein activity. Importantly, this would prevent the direct selection against these H5 genes, due to the inability to express an H5 protein. Unlike the production of a functionally impaired H5 protein, these modified

genes could thus survive when they are heterozygous with an intact H5 gene, and would constitute recessive lethals. Such "inactive" genes could therefore potentially survive for a period of time sufficient for the occurrence of a second DNA mutation, which could restore protein activity.

F) Summary of the Methodology for Analysis of DNA Sequence Evolution

As has been discussed in detail previously; the analysis described in Results was successful in resolving some of the amino acid sequence modifications that have occurred within Region 1 of H5 during the evolutionary histories of Xenopus laevis, human, duck and chicken. This could represent an effective approach for the study of protein sequence evolution via DNA sequence analysis, in addition to providing information for resolving evolutionary lineages, as will be discussed in the next section. A similar approach should be successful in recognizing the DNA mutations and their associated amino acid sequence modifications, that have occurred within other protein families. To more clearly describe this approach a discussion of my methodology is presented. Similar approaches have been described by Dayhoff (1969) and Fitch (1976) for examining the evolution of cytochrome c using computer analysis of protein and DNA sequences. Computer analysis of tRNA (Cedergren et al. 1981) and ribosomal RNA (Olsen et al. 1986) sequences have also been employed for establishing evolutionary lineages. However, the major difference between these methodologies and that described here, is that computer analysis has been restricted to statistical examination of DNA sequence homologies, primarily involving the resolution of single base pair mutations (Felsenstein 1982). The ability to recognize specific deletions and insertions as described here, greatly increases the accuracy of DNA sequence analysis for investigating protein evolution and for the establishment of evolutionary lineages.

The steps taken during the analysis of the H5 genes can be summarized as follows:

- 1) computer DNA sequence alignment between each of the genes
- 2) identification of modified regions: location of sequence breaks and areas of extensive nucleotide differences
- 3) identification of duplicated sequences
- 4) realignment of the genes after removal of duplicated sequences
- 5) identification of deletions and assignment of single bp changes
- 6) derivation of a hypothetical ancestral sequence for each gene
- 7) development of hypothetical evolutionary intermediates: resolution of further mutational events
- 8) construction of evolutionary histories linking each of the genes

The initial step was to generate computer alignments between all combinations of the five H5 coding sequences. Although such alignments were generally found to be inaccurate for highly modified regions, it was effective in establishing highly conserved areas and for identifying changes in length.

The major problem associated with alignment of homologous bases within two related DNA sequences, is the disruption of colinearity caused by deletions and insertions. The next step was therefore to identify the regions in which such mutations may have occurred. This was done primarily by locating regions which contained computer generated sequence breaks and/or contained extensive nucleotide mismatches.

Insertions proved to be relatively easy to resolve, with all observed insertions produced by the duplication of adjacent sequences. In attempting to resolve sequence duplications, the occurrence of single nucleotide mutations subsequent to the duplication event must be taken

into account, although such mutations did not for the most part greatly effect the analysis. Importantly, duplicated sequences represent the generation of new DNA sequence and thus must be removed before an effective sequence alignment can be achieved.

Deletions are the most difficult form of mutation to resolve, since derivation of deleted sequences can only be accomplished by reference to other related genes which contain the deleted sequence. It was therefore fortunate that the *Xenopus* H5A gene has not undergone any major mutations within Region 1, in that it provided an important reference sequence from which deletions within mammalian H1⁰ and avian H5 were resolved. This is in contrast to Region 3, in which all of the H5 genes examined have likely undergone extensive mutations, making the DNA sequence analysis of this region very difficult (see Results).

Once deletion and duplication events have been resolved, single bp changes can be assigned, followed by the derivation of a representative ancestral sequence for each of the most closely related genes. The next step is the resolution of additional mutational events that have produced differences between the most divergent genes, which may consist of multiple, overlapping events that obscure the alignment of the DNA sequences. This step is followed by development of evolutionary scenarios linking each of the genes, in which evolutionary intermediates are derived, based upon the most likely order of the mutational events.

There were limitations inherent to this analysis of the H5 genes. First, was the restriction of this analysis to the coding regions of the H5 genes. The specialized constraints imposed by protein coding must obviously be an important factor in the fixation of modified genes, and therefore the frequency and form of DNA sequence modifications are

likely different outside of the coding region. Second, the specialized structural role played by the H5 proteins, which has led to a high degree of sequence conservation (characteristic of all the histone classes), has likely been central to the success of this analysis. The evolutionary analysis of less conserved proteins would therefore be more difficult, as was found in the analysis of Region 3 of H5. Third, the number of H5 genes used in the analysis was limiting. This led to some difficulties in resolving the evolutionary intermediates for the avian H5 genes.

G) The Analysis of DNA Sequence for Establishing and/or Confirming Evolutionary Lineages.

One of the most exciting aspects of the comparative DNA sequence analysis is the ability to establish evolutionary relationships between diverse species. The identification of specific DNA sequence mutations within H5 provides the opportunity to clearly resolve the ancestry of individual animal species through DNA sequence characterization of their H5 genes. This is well illustrated by the 15 bp deletion within Section 2 of Region 1 that is predicted to have occurred within the ancestral avian H5 gene. It is clear from the analysis that after the divergence of duck and chicken, independent mutational events occurred close to the site of this deletion. This not only provides a clear distinction between duck and chicken, but it can also be used to establish the ancestry of other bird species, based upon the DNA sequence of this segment of their H5 gene. It is also apparent that this 15 bp deletion has not occurred with the human H1⁰ and thus provides a clear marker for separating the mammalian and avian lineages. Another illustration comes from examining the amino acid sequences for

the amino termini of quail, pigeon and goose H5, which can provide information similar to DNA sequence. These were presented in Figure 12 and are reproduced below:

QH5	TE.SL..VLSPAP.AKPKRA
CH5	TE.SL..VLSPAP.AKPKRVKASRRSASH
PH5	TE.SPIPVPAAPAAKPKPKRVSKRPASH
DH5	TD.SPIP..APAPAAKPKRAKAPRKPASH
GH5	TD.SPIP..APAPAAKPKRARAPRKPASH

These amino acid sequences indicate that chicken and quail have likely evolved from the same avian lineage. Pigeon, duck, and goose also appear to have evolved from a common lineage, but one which has diverged from the chicken and quail lineage. Note however, that examination of DNA sequence is necessary to confirm these relationships, because of the potential inaccuracies produced by the redundancy of amino acid coding, and the complexity of the DNA mutations that produce changes in amino acid sequence (see Figure 16). For example, the DNA sequence analysis has indicated that the absence of alanine within quail and chicken H5 (AP.AKP) is not due to a deletion, as could be concluded by the presence of an homologous alanine within the other H5's (APAAKP). Rather, it was established from the DNA sequence analysis that both the duck and chicken H5's initially lacked this alanine, and that an insertion of "APA" during the evolution of the duck H5 produced the present day alanine (i.e. AP.AKP → APAPAAKP; refer to pDH5-2). It is thus apparent that the DNA sequence analysis provides much greater information than amino acid sequence alone, and can allow evolutionary relationships between species to be more clearly resolved.

Although sequence characterization of any gene family could provide information similar to that obtained from H5, it is apparent that to be successful, there has to be a balance between the degree of sequence

variation and conservation. In this respect, H5 has proven to be an effective candidate, consisting of hypervariable, variable and conserved segments. This combination provides the information necessary for developing evolutionary lineages between both greatly divergent and closely related species.

H) Future Considerations: Elucidating the Evolutionary Origins of H5

Although this project has provided insights into some important aspects of H5, many questions must still be answered before a clear understanding of this linker histone subclass is achieved. These include: what are the precise evolutionary origins of H5 and what are the evolutionary relationships between vertebrate H1 and H5? Specifically, was the ancestral H5 gene derived from an ancestral vertebrate H1 or did the evolution of vertebrates entail the derivation of these two distinct linker histone subclasses from a third form of linker histone? What are the functional relationships between H5 and H1, and how has chromatin function been modified during the evolution of H5 and vertebrate H1?

The sequence characterization of Xenopus laevis H5 has definitively established that H5 is present within lower vertebrates. This is consistent with reports of H5-like proteins in several species of amphibians (Panyim et al. 1971, Risley and Eckhardt 1981, Brown et al. 1981, Shimada et al. 1981, Smith et al. 1984, Moorman and de Boer 1985) and reptiles (Panyim et al. 1971, Rutledge et al. 1981, Svinarchuk et al. 1982). The amino acid sequence comparisons of Region 2 presented in Results also revealed little homology of vertebrate H1 and H5 with several non-vertebrate H1 proteins. This is consistent with the survey of Smith et al. (1984a) which revealed no evidence of H5-like histones

in a crustacean, two fungi and a plant, and led to their suggestion that this type of linker histone is restricted to the vertebrates.

The evidence thus suggests that H5 evolved before or early in vertebrate evolution. The reports of an H5-like histone in some fish (Miki and Neelin 1975, 1977a, 1977b, Zalenskii 1981, Svinarchuk 1985) is thus particularly significant, suggesting that the characterization of fish H5 could provide the information necessary for elucidating the evolutionary origins of H5. Importantly, the identification of primitive forms of H5 which may be present in fish, could also provide the opportunity to investigate changes in chromatin function that may have been associated with the evolution of H5.

Finally, the association of H5 accumulation with the restriction of RNA transcription during reduced cellular activity (see Introduction) illustrates the importance for better understanding the mechanism controlling the expression and incorporation of H5 into chromatin. A detailed examination of the structure and regulation of the H5 gene could therefore provide valuable insights into how H5 function is related to the regulation of gene expression.

V. REFERENCES

Ajiro, K., T.W. Borun and D. Solter. 1981. Quantitative changes in the expression of histone H1 and H2B subtypes and their relationship to the differentiation of mouse embryonal carcinoma cells. *Developmental Biology* 86, 206-211.

Allan, J., G.J. Cowling, N. Harborne, P. Cattini, R. Craige and H. Gould. 1981. Regulation of the higher-order structure of chromatin by histones H1 and H5. *J. Cell Biol.* 90, 279-288.

Allan, J., B.J. Smith, B. Dunn and M. Bustin. 1982. Antibodies against the folding domain of histone H5 cross-react with H1⁰ but not with H1. *J. Biol. Chem.* 257, 10533-10535.

Appels, R., J.R.E. Wells and A.F. Williams. 1972. Characterization of DNA-bound histone in the cells of the avian erythropoietic series. *J. Mol. Biol.* 70, 425-434.

Barbero, J.L., G. Gonzalez de Buitrago and R. Rodriguez. 1982. Structural homology between the central domain of the histones from sea urchin and the fruit fly Ceratitis capitata. *Comp. Biochem. Physiol.* 73B, 475-479.

Barker, W.C., L.K. Ketcham and M.O. Dayhoff. 1978. A comprehensive examination of protein sequences for evidence of internal gene duplication. *J. Mol. Evol.* 10, 265-281.

Bates, D.L. and J.O. Thomas. 1981. Histone H1 and H5: one or two molecules per nucleosome? *Nuc. Acid Res.* 9, 5883-5894.

Berger, S.L., D.M. Wallace, R.S. Puskas and W.H. Eschenfeldt. 1983. Reverse transcriptase and its associated ribonuclease H: Interplay of two enzyme activities controls the yield of single-stranded complementary deoxyribonucleic acid. *Biochemistry* 22, 2365-2372.

Berger, S.L. 1984. The use of Cerenkov radiation for monitoring reactions performed in minute volumes: examples from recombinant DNA technology. *Anal. Biochem.* 136, 515-519.

Biggin, M.D., T.J. Gibson and G.F. Hong. 1983. Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* 80, 3963-3965.

Billet, M.A. and J. Hindley. 1972. A study of the quantitative variation of histones, and their relationship to RNA synthesis, during erythropoiesis in adult chicken. *Eur. J. Biochem.* 28, 451-462.

Birnstiel, M.L., M. Busslinger and K. Strub. 1985. Transcription termination and 3' processing: the end is in site! *Cell* 41, 349-359.

Borun, T.W. 1975. Histones, differentiation, and the cell cycle. In "Results and Problems in Cell Differentiation", Vol 7, Cell Cycle and Cell Differentiation (eds. J. Reinert and H. Holtzer) N.Y.: Springer-Verlog. pp 249 -290.

Boshardt, M. and H. zur Hausen. 1986. Human papillomaviruses in Buschke-Lowenstien tumors: physical state of the DNA and identification of a tandem duplication in the noncoding region of a human papillomavirus-6 subtype. J. Virology 58, 963-966.

Brown, G.L. 1982. Lysine-rich histones of the amphibians Xenopus and Ambystoma. Ph.D. Thesis, Carleton University.

Brown, G.L., R.G. Rutledge and J.M. Neelin. 1981. Anuran erythrocytes and liver both contain satellite histone H1^S. Life Sciences 28, 2993-2999.

Burke, J.F. 1984. High-sensitivity S1 mapping with single-stranded [³²P]DNA probes synthesized from bacteriophage M13mp templates. Gene 30, 63-68.

Cartwright, I.L., M.A. Keene, G.C. Howard, S.M. Abmayr, G. Fleischmann, K. Lowenhaupt and S.C.R. Elgin. 1982. Chromatin structure and gene activity: the role of nonhistones chromosomal proteins. CRC Crit. Rev. Biochem. 13, 1-86.

Cary, P.D., M.L. Hines, E.M. Bradbury, B.J. Smith and E.W. Johns. 1981. Conformation studies of histone H1⁰ in comparison with histones H1 and H5. Eur. J. Biochem. 120, 371-377.

Cedergren, R.J., D. Sankoff, B. LaRue and H. Grosjean. 1981. The evolving tRNA molecule. CRC Crit. Rev. Biochem. 11, 35-104.

Chabanas, A., J.J. Lawrence, J. Humbert and H. Eisen. 1983. Cell cycle regulation of histone H1⁰ in CHO cells: a flow cytofluorimetric study after double staining of the cells. EMBO J. 2, 833-837.

Chamberlian, J.P. 1979. Fluorographic detection of radioactivity in polyacrylamide gels with the water-soluble fluor, sodium salicylate. Anal. Biochem. 98, 132-135.

Chandler, P.M. 1982. The use of single-stranded phage DNAs in hybrid arrest and release translation. Anal. Biochem. 127, 9-16.

Chegini, N., V. Aleporou, G. Bell, V.A. Hilder, and N. Maclean. 1979. Production and fate of erythroid cells in anemic Xenopus laevis. J. Cell Sci. 35, 403-415.

Chirgwin, J.M., A.E. Przybyla, R.J. MacDonald and W.J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294-5299.

Chomczynski, P. and P.K. Qasba. 1984. Alkaline transfer of DNA to plastic membrane. Biochem. Biophys. Res. Commun. 122, 340-344.

- Cole, R.D. 1977. Special features of the structures of H1 histones. In "The Molecular Biology of the Mammalian Genetic Apparatus", T'so, P., Ed. Elsevier North-Holland Biomedical Press. pp. 93-104.
- Cole, K.D., R.G. York and W.S. Kistler. 1984. The amino acid sequence of boar H1t, a testis-specific H1 histone. J. Biol. Chem. 259, 13695-13702.
- Cole, K.D., J.C. Kandala and W.S. Kistler. 1986. Isolation of the gene for the testis-specific H1 histone variant H1t. J. Biol. Chem. 261, 7178-7183.
- Coles, L.S. and J.R.E. Wells. 1985. An H1 histone gene-specific 5' element and evolution of H1 and H5 genes. Nuc. Acid Res. 13, 585-594.
- Dale, R.M.K., B.A. McClure and J.P. Houchins. 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: Application to sequencing the corn mitochondrial 18 S rDNA. Plasmid 13, 31-40.
- Dalton S., J.R. Coleman and J.R. Wells. 1986. Transcription of the histone H5 gene is not S-phase regulated. Mol. Cell. Biol. 6, 601-606.
- D'Anna, J.A., R.A. Tobey and L.R. Gurley. 1980a. Concentration dependent effects of sodium butyrate in Chinese hamster cells: Cell cycle progression, inner-histone acetylation, and induction of an H1-like protein. Biochemistry 19, 2656-2671.
- D'Anna, J.A., L.R. Gurley, R.R. Becker, S.S. Barkam, R.A. Tobey and R.A. Walters. 1980b. Amino acid analysis and cell cycle dependent phosphorylation of an H1-like, butyrate-enhanced protein (BEP; H1^o; IP125) from chinese hamster cells. Biochemistry 19, 4331-4341.
- D'Anna, J.A., L.R. Gurley and R.R. Becker. 1981. Histones H1^oa and H1^ob are the same as CHO histones H1(III) and H1(IV): New features of H1^o phosphorylation during the cell cycle. Biochemistry 20, 4501-4505.
- D'Anna, J.A., M.M. Thayer, R.A. Tobey and L.R. Gurley. 1985. G1- and S-phase synthesis of histones H1 and H1^o in mitotically selected CHO cells: Utilization of high-performance liquid chromatography. Biochem. 24, 2005-2010.
- Dayhoff, M.O. 1969. Computer analysis of protein evolution. Sci. Am. 221, 86-95.
- Destree, O.H.J., H.J. Hoenders, A.F.M. Moorman and R. Charles. 1979. Histones of Xenopus laevis erythrocytes: Purification and characterization of the lysine-rich fraction. Biochim. Biophys. Acta 577, 61-70.
- Doenecke, D. and R. Tonjes. 1984. Conserved dyad symmetry structures at the 3' end of H5 histone genes. Analysis of the duck H5 gene. J. Mol. Biol. 178, 121-135.

- Doenecke, G. and R. Tonjes. 1986. Differential distribution of lysine and arginine residues in the closely related histones H1^o and H5: Analysis of a human H1^o gene. *J. Mol. Biol.* 187, 461-464.
- Dretzen, G., M. Bellard, P. Sassone-Corsi and P. Chambon. 1981. A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *Anal. Biochem.* 112, 295-298.
- Elgin, S.C.R. and H. Weintraub. 1975. Chromosomal proteins and chromatin structure. *Ann. Rev. Biochem.* 44, 725-774.
- Felsenstein, J. 1982. Numerical methods for inferring evolutionary trees. *Quart. Rev. Biol.* 57, 379-404.
- Fields, S. and G. Winter. 1981. Nucleotide-sequence heterogeneity and sequence rearrangements in influenza virus cDNA. *Gene* 15, 207-214.
- Fitch, W.M. 1976. The molecular evolution of cytochrome c in eukaryotes. *J. Mol. Evol.* 8, 13-40.
- Foran, D.R., P.J. Johnson and G.P. Moore. 1985. Evolution of two actin genes in the sea urchin Strongylocentrotus franciscanus. *J. Mol. Evol.* 22, 108-116.
- Gabrielli, F., D.P. Aden, S.C. Carrel, C. von Bahr, A. Rane, C.A. Angeletti and R. Hancock. 1984. Histone complements of human tissues, carcinomas, and carcinoma-derived cell lines. *Mol. Cell. Biochem.* 65, 57-66.
- Garrard, W.T. and J. Bonner. 1974. Changes in chromatin proteins during liver regeneration. *J. Biol. Chem.* 249, 5570-5579.
- Gasaryan, K.G., N.B. Andreeva and T.Y. Vishnevskaya. 1978. Existence of differences in repressor properties between serine-rich histone (H5, F2c) from immature and mature pigeon erythroid cells. *Differentiation* 10, 123.
- Georges, F., R. Brousseau, J. Michniewicz, G. Prefontaine, J. Stawinski, W. Sung, R. Wu and S.A. Narang. 1984. Synthesis of a human insulin gene VII. Synthesis of preproinsulin-like human DNA, its cloning and expression in M13 bacteriophage. *Gene* 27, 201-211.
- Gjerst, R., F. Ibarrando, S. Saragosti, and H. Eisen. 1981. Distribution of IP125 in chromatin and its possible involvement in chromatin condensation. *Biochem. Biophys. Res. Commun.* 99, 349-357.
- Gjerst, R., C. Gorka, S. Hasthorpe, J.J. Lawrence and H. Eisen. 1982. Developmental and hormonal regulation of protein H1^o in rodents. *Proc. Natl. Acad. Sci.* 79, 233-2337.
- Glisin, V., R. Crkvenjakov and C. Byus. 1974. Ribonucleic acid isolated by cesium chloride centrifugation. *Biochemistry* 13, 2633-2637.

- Glotov, B.O., L.G. Nikolaev, V.K. Dashkevich and S.F. Barbashov. 1985. Histone crosslinking patterns indicate dynamic binding of histone H1 in chromatin. *Biochim. Biophys. Acta.* 824, 185-193.
- Gopinathan, K.P., L.A. Weymouth, T.A. Kunkel and L.A. Leob. 1979. Mutagenesis in vitro by DNA polymerase from an RNA virus. *Nature* 178, 857-859.
- Gubler, U. and B.J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 25, 263-269.
- Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* 166, 557-580.
- Hannon, R., E. Bateman, J. Allan, N. Harborne and H. Gould. 1984. Control of RNA polymerase binding to chromatin by variation in linker histone composition. *J. Mol. Biol.* 180, 131-149.
- Hawkes, R., E. Niday and J. Gordon. 1982. A dot-immunobinding assay for monoclonal and other antibodies. *Anal. Biochem.* 119, 142-147.
- Hentschel, C.C. and M.L. Birnstiel. 1981. The organization and expression of histone gene families. *Cell* 25, 301-313.
- Hohmann, P. 1978. The H1 class of histone and diversity in chromosomal structures. *Subcell. Biochem.* 5, 87-127.
- Hohmann, P. 1980. Species and cell specific expression of H1 histones in tissue culture cells. *Arch. Biochem. Biophys.* 205, 198-209.
- Huang, H., R.D. Cole. 1984. The distribution of H1 histones is nonuniform in chromatin and correlates with different degrees of condensation. *J. Biol. Chem.* 259, 14237-14242.
- Hurley, C.K. and J.T. Stout. 1980. Maize histone H1: A partial structural characterization. *Biochemistry* 19, 410-416.
- Igo-Kemenes, T., W. Horz and H.G. Zachau. 1982. Chromatin. *Annu. Rev. Biochem.* 51, 89-121.
- Isenberg, I. 1979. Histones. *Ann. Rev. Biochem.* 48, 159-191.
- Karpov, V.L., O.V. Preobrazhenskaya and A.D. Mirzabekov. 1984. Chromatin structure of hsp 70 genes, activated by heat shock: Selective removal of histones from the coding region and their absence from the 5' region. *Cell* 36, 423-431.
- Kedes, L.H. 1979. Histone genes and histone messengers. *Ann. Rev. Biochem.* 48, 837-870.
- Kellogg, T.S. 1983. The effect of sample composition and vial type on Cerenkov counting in a liquid scintillation counter. *Anal. Biochem.* 134, 137-143.

- Keppel, F., B. Allet and H. Eisen. 1977. Appearance of a chromatin protein during the erythroid differentiation of Friend virus-transformed cells. *Proc. Natl. Acad. Sci. USA* 74, 653-656.
- Keppel, F. B. Allet and H. Eisen. 1979. Biochemical properties and localization of the chromosomal protein IP125. *Eur. J. Biochem.* 96, 477-482.
- Kimura, T., F.C. Mills, J. Allan and H. Gould. 1983. Selective unfolding of erythroid chromatin in the region of the active B-globin gene. *Nature* 306, 709-712.
- Kinkade, J.M. 1969. Qualitative species differences and quantitative tissue differences in the distribution of lysine-rich histones. *J. Biol. Chem.* 244, 3375-3386.
- Kmiecik, D., D. Sello, D. Belaiche and P. Sautiere. 1985. Primary structure of the two variants of a sperm-specific histone H1 from the annelid Platynereis dumerilii. *Eur. J. Biochem.* 150, 359-370.
- Komaiko, W. and G. Felsenfeld. 1985. Solubility and structure of domains of chicken erythrocyte chromatin containing transcriptionally competent and inactive genes. *Biochem.* 24, 1186-1193.
- Kornberg, R.D. 1977. Structure of chromatin. *Ann. Rev. Biochem.* 46, 931-954.
- Kramer, W.V., Drutsa, H. Jansen, B. Kramer, M. Pflugfelder and H. Fritz. 1984. The gapped duplex DNA approach to oligonucleotide-directed mutation construction. *Nuc. Acid Res.* 12, 9441-9456.
- Kreig, P.A., A.J. Robins, M.J. Gait, R.C. Titmas and J.R.E. Wells. 1982. Chicken histone H5: selection of a cDNA recombinant using an extended synthetic primer. *Nuc. Acid Res.* 10, 1495-1502.
- Kreig, P.A., A.J. Robins, A. Colman and J.R.E. Wells. 1982. Chicken Histone H5 mRNA: the polyadenylated RNA lacks the the conserved histone 3' terminator sequence. *Nuc. Acid Res.* 10, 6777-6785.
- Kreig, P.A., A.J. Robins, R. D'Andrea and J.R.E. Wells. 1983. The chicken H5 gene is unlinked to core and H1 histone genes. *Nuc. Acid Res.* 11, 619-627.
- Kress, H., R. Tonjes and D. Deonecke. 1986. Butyrate induced accumulation of a 2.3 kb polyadenylated H1⁰ histone mRNA in HeLa cells. *Nuc. Acid Res.* 14, 7189-7197.
- Lathe, R.F., M.P. Kieny, D. Schmitt, P. Curtis and J.P. Lecocq. 1984. M13 bacteriophage vectors for the expression of foreign proteins in Escherichia coli: the rabies glycoprotein. *J. Mol. App. Gen.* 2, 331-342.
- Lennox, R.W. and L.H. Cohen. 1983. The histone H1 complements of dividing and nondividing cells of the mouse. *J. Biol. Chem.* 262-268.

- Levy, S., I. Sures and L. Kedes. 1982. The nucleotide and amino acid coding sequence of a gene for H1 histone that interacts with euchromatin: The early embryonic H1 gene of the sea urchin Strongylocentrotus purpuratus. J. Biol. Chem. 257, 9438-9443.
- Macleod, A.R., N.C.W. Wong and G.H. Dixon. 1977. The amino acid sequence of trout testis histone H1. Eur. J. Biochem. 78, 281-291.
- Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. Molecular Cloning. Cold Spring Harbor Laboratories.
- Marsh, W.H. and P.J. Fitzgerald. 1973. Pancreas acinar cell regeneration. XIII. Histone synthesis and modification. Fed. Proc. 32, 2119-2125.
- Markose, E.R. and M.R. Satyanarayana Rao. 1985. Testis-specific histone H1t is antigenically distinct among H1 subtypes. J. Biol. Chem. 260, 16263-16268.
- Mathis, D., P. Oudet and P. Chambon. 1980. Structures of transcribing chromatin. Prog. Nuc. Acid. Res. Mol. Biol. 24, 1-55.
- Maxson, R., R. Cohn and L. Kedes. 1983. Expression and organization of histone genes. Ann. Rev. Genet. 17, 239-277.
- McGhee, J.D. and G. Felsenfeld. 1980. Nucleosome structure. Ann. Rev. Biochem. 49, 1115-1156.
- Medvedev, Zh.A. and N.D. Kirpicheva. 1972. Time of the erythrocyte-specific histone fraction F2c synthesis during erythropoiesis in birds. Molek. Biol. 6, 607.
- Medvedev, Zh.A. and M.N. Medvedeva. 1980. High H1⁰/H1 histone ratio in spontaneous hepatomas in aging mice. IRCS Med. Sci. 8, 431.
- Messing, J. 1983. New M13 vectors for cloning. Methods Enzy. 101, 20-78.
- Mezquita, J., W. Connor, R.J. Winkfein and G.H. Dixon. 1985. An H1 histone gene from rainbow trout (Salmo gairdnerii). J. Mol. Evol. 21, 209-219.
- Miki, B.L.A. and J.M. Neelin. 1975. The histones of rainbow trout erythrocytes include an erythrocyte-specific histone. Can. J. Biochem. 53, 1158-1169.
- Miki, B.L.A. and J.M. Neelin. 1977a. Histone H5 from carp erythrocytes. Can. J. Biochem. 55, 384-389.
- Miki, B.L.A. and J.M. Neelin. 1977b. Comparison of the histones from fish erythrocytes. Can. J. Biochem. 55, 1220-1227.
- Molgaard, H.V., M. Perucho and A. Ruiz-Carrillo. 1980. Histone H5 messenger RNA is polyadenylated. Nature 283, 502-504.

- Moorman, A.F.M. and P.A.J. de Boer. 1985. Immunohistochemical distribution of the histone H1⁰/H5 variant in various tissues of adult Xenopus laevis. *Cell Differ.* 16, 109-117.
- Mura, C.V. and B.D. Stollar. 1981. Serological detection of homologies of H1⁰ and H5 with H1 histones. *J. Biol. Chem.* 256, 9767-9769.
- Neelin, J.M., ~~P.X. Gallahan~~, D.C. Lamb and K. Murray. 1964. The histones of chicken erythrocyte nuclei. *Can. J. Biochem. Physiol.* 42, 1743-1752.
- Neelin, J.M. 1968. Variability in cell-specific and common histones of avian erythrocytes. *Can. J. Biochem.* 46, 241-247.
- Nisbet, I.T. and M.W. Beilharz. 1985. Simplified DNA manipulations based on in vitro mutagenesis. *Gene Techn.* 2, 23-29.
- Olsen, G.J., D.J. Lane, S.J. Giovannoni and N.R. Pace. 1986. Microbial ecology and evolution: a ribosomal RNA approach. *Ann. Rev. Microbiol.* 40, 337-365.
- Panyim, S., D. Bilek and R. Chalkley. 1971. An electrophoretic comparison of vertebrate histones. *J. Biol. Chem.* 246, 4206-4251.
- Pederson, T. 1976. Cellular aspects of histone synthesis. In "Protein Synthesis", Vol 2, E.H. McConkey, ed. (N.Y.: Marcel Dekker), pp. 69-123.
- Pehrson, J. and R.D. Cole. 1980. Histone H1⁰ accumulates in growth-inhibited cultured cells. *Nature* 285, 43-44.
- Pehrson, J.R. and R.D. Cole. 1981. Bovine H1⁰ histone subfractions contain an invariant sequence which matches histone H5 rather than H1. *Biochem.* 20, 2298-2301.
- Perry, M., G.H. Thomsen and R.G. Roeder. 1985. Genomic organization and nucleotide sequence of two distinct histone gene clusters from Xenopus laevis: Identification of novel conserved upstream sequence elements. *J. Mol. Biol.* 185, 479-499.
- Pieler, C., G.R. Adolf and P. Sevetly. 1981. Accumulation of histone H1⁰ during chemically induced differentiation of murine neuroblastoma cells. *Eur. J. Biochem.* 115, 329-333.
- Pina, B., P. Martinez, L. Simon and P. Suau. 1984. Differential kinetics of histone H1⁰ accumulation in neuronal and glial cells from at cerebral cortex during postnatal development. *Biochem. Biophys. Res. Commun.* 123, 697-702.
- Puskas, R.S., N.R. Manley, D.M. Wallace and S.L. Berger. 1982. Effect of ribonucleoside-vanadyl complexes on enzyme-catalyzed reactions central to recombinant deoxyribonucleic acid technology. *Biochemistry* 21, 4602-4608.
- Reed, R.C. and D.A. Mann. 1985. Rapid transfer of DNA from agarose gels to nylon membrane. *Nuc. Acid Res.* 13, 7207-7221.

Reeves, R. 1984. Transcriptionally active chromatin. *Bioch. Biophys. Acta* 782, 343-393.

Risley, M.S. and R. A. Eckhardt. 1981. H1 histone variants in Xenopus laevis. *Devel. Biol.* 84, 79-87.

Ruiz-Carrillo, A., L.J. Wangh, V.C. Littau and V.G. Allfrey. 1974. Changes in histone acetyl content and in nuclear non-histones protein composition of avian erythroid cells at different stages of maturation. *J. Biol. Chem.* 249, 7358-7368.

Ruiz-Carrillo, A., L.J. Wangh and V.G. Allfrey. 1976. Selective synthesis and modification of nuclear proteins during maturation of avian erythroid cells. *Arch. Biochem. Biophys.* 174, 273-290.

Ruiz-Carrillo, A., M. Affolter and J. Renaud. 1983. Genomic organization of the genes coding for the six main histones of the chicken: complete sequence of the H5 gene. *J. Mol. Biol.* 170, 843-859.

Rutledge, R.G., C.E. Shay, G.L. Brown and J.M. Neelin. 1981. The similarity of histones from turtle erythrocytes and liver. *Can. J. Biochem.* 59, 273-279.

Rutledge, R.G. 1982. Histone synthesis during erythropoiesis in Xenopus laevis, M. Sc. Thesis, Carleton University, Ottawa, Canada.

Rutledge, R.G., J.M. Neelin and V.L. Seligy. 1984. Uncoupled synthesis of H1^O-like histone H1^S during late erythropoiesis in Xenopus laevis. *Eur. J. Biochem.* 144, 191-198.

Schlissel, M.S. and D.D. Brown. 1984. The transcriptional regulation of Xenopus 5S RNA genes in chromatin: The roles of active stable transcription complexes and histone H1. *Cell* 37, 903-913.

Seligy, V.L. and J.M. Neelin. 1970. Transcription properties of stepwise acid-extraction of chicken erythrocyte chromatin. *Biochem. Biophys. Acta* 213, 380-390.

Seligy, V.L., G.H.M. Adams and J.M. Neelin. 1973. The role of histones in avian erythropoiesis. In *Proc. Int'l. Cong. on the "Biochemistry of Gene Expression in Higher Organisms"*. eds. J. Pollak and J.W. Lee, Australia and New Zealand Book Co. TTY, Ltd. pp 177-190.

Seligy, V.L. and M. Miyagi. 1974. Comparison of template property changes after salt extraction of avian erythrocyte and liver chromatin. *Eur. J. Biochem.* 46, 259-269.

Seligy, V., C. Roy, M. Dove and M. Yaguchi. 1976. Species variability of N-terminal sequence of avian erythrocyte-specific histone H5. *Biochem. Biophys. Res. Commun.* 71, 196-202.

Seyedin, S.M. and W.S. Kistler. 1980. Isolation and characterization of rat testis H1t: An H1 histone variant associated with spermatogenesis. *J. Biol. Chem.* 255, 5949-5954.

Seyedin, S.M., R.D. Cole and W.S. Kistler. 1981. H1 histones from mammalian testis. The widespread occurrence of H1t. *Exptl. Cell Res.* 136, 399-405.

Shimada, T., Y. Okihama, C. Murata and R. Shukuya. 1981. Occurrence of H1^o-like protein and protein A24 in the chromatin of bullfrog erythrocytes lacking histone 5. *J. Biol. Chem.* 256, 10577-10582.

Slocombe, P., A. Easton, P. Boseley and D.C. Burke. 1982. High-level expression of an interferon $\alpha 2$ gene cloned in phage M13mp7 and subsequent purification with a monoclonal antibody. *Proc. Natl. Acad. Sci. USA* 79, 5455-5459.

Smith, B.J. and E.W. Johns. 1980a. Isolation and characterization of subfractions of nuclear protein H1^o. *FEBS Lett.* 110, 25-29.

Smith, B.J., J.M. Walker and E.W. Johns. 1980b. Structural homology between a mammalian H1^o subfraction and avian erythrocyte-specific histone H5. *FEBS Lett.* 112, 42-44.

Smith, B.J., Y. Cook, E.W. Johns and R.A. Weiss. 1981. Absence of H1^o from quiescent chicken cells. *FEBS Lett.* 135, 77-80.

Smith, B.J., M.R. Harris, C.M. Sigournay, E.L.V. Mayes and M. Bustin. 1984a. A survey of H1^o- and H5-like protein structure and distribution in higher and lower eukaryotes. *Eur. J. Biochem.* 138, 309-317.

Smith, R.D., J. Yu, A. Annunziato and R.L. Seale. 1984b. B-globin gene family in murine erythroleukemia cells resides within two chromatin domains differing in higher order structure. *Biochem.* 23, 2970-2976.

Stambolova, M., V. Simeonova, L. Srebrevva, J. Zlatanova and R. Tsanev. 1984. Histone H1^o in developing rat brain cells. *Differentiation* 28, 191-194.

Sugarman, B.J., J.B. Dodgson and J.D. Engel. 1983. Genomic organization, DNA sequence, and expression of chicken embryonic histone genes. *J. Biol. Chem.* 258, 9005-9016.

Sung, M.T., J. Harford, M. Bundmann and G. Vidalakas. 1977. Metabolism of histones in avian erythroid cells. *Biochemistry* 16, 279-285.

Svinarchuk, F.P., S.M. Rozov and V.A. Berdnikov. 1982. Histone H1 of reptiles homologous to H5 of birds. *Mol. Biol.* 16, 551-559.

Thoma, F., Th. Koller and A. Klug. 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell. Biol.* 83, 403-427.

Thoma, F., and Th. Koller. 1981. Unravalled nucleosomes, nucleosome beads and higher-order structures of chromatin: Influence of non-histone components and histone H1. *J. Mol. Biol.* 149, 709-733.

Thoma, F., R. Losa and T. Koller. 1983. Involvement of histone H1 and H5 in the structural organization of soluble chromatin. *J. Mol. Biol.* 167, 619-640.

Thomas, N. and N. Maclean. 1975. The erythroid cells of anemic Xenopus laevis. I. Studies on the cellular morphology and protein and nucleic acid synthesis during differentiation. J. Cell Sci. 19, 509-520.

Thomas, P.S. 1983. Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods Enzymology 100, 255-266.

Tobin, R.S., and V.L. Seligy. 1975. Characterization of chromatin-bound erythrocyte histone V (f2c): Synthesis, acetylation and phosphorylation. J. Biol. Chem. 250, 358-364.

Turner, P.C., T.C. Aldridge, H.R. Woodland and R.W. Old. 1983. Nucleotide sequences of H1 histone genes from Xenopus laevis. A recently diverged pair of H1 genes and an unusual H1 pseudogene. Nuc. Acid Res. 11, 4093-4107.

Unger-Ullmann, C. and S.P. Modak. 1979. Gel electrophoresis analysis of histones in lens epithelium, lens fibre, liver, brain, and erythrocytes of late chick embryos. Differentiation 12, 135-144.

Urban, M.K., J.M. Neelin and T.W. Betz. 1980. Correlation of chromatin composition with metabolic changes in nuclei of primitive erythroid cells of chicken embryos. Can. J. Biochem. 9, 726-731.

Varricchio, F. 1977. Postnatal increase in histone H1a in the rat pancreas. Arch. Biochem. Biophys. 179, 715-717.

Von Holt, C., W.N. Strickland, W.F. Brandt and M.S. Stickland. 1979. More histone structures. FEBS Lett. 100, 201-218.

Weintraub, H. 1978. The nucleosome repeat length increases during erythropoiesis in the chick. Nuc. Acid Res. 5, 1179-1188.

Weintraub, H. 1984. Histone-H1-dependent chromatin superstructures and the suppression of gene activity. Cell 38, 17-27.

Wells, D.E. 1986. Compilation analysis of histones and histone genes. Nuc. Acid Res. 14, r119-r149.

Weymouth, L.A. and L.A. Leob. 1978. Mutagenesis during in vitro DNA synthesis. Proc. Natl. Acad. Sci. USA 75, 1924-1928.

Worcel, A., S. Strogatz and D. Riley. 1981. Structure of chromatin and the linking number of DNA. Proc. Natl. Acad. Sci. USA 78, 1461-1465.

Yaguchi, M., C. Roy and V.L. Seligy. 1979. Complete amino acid sequence of goose erythrocyte H5 histone and the homology between H1 and H5 histones. Biochem. Biophys. Res. Commu. 90, 1400-1406.

Zalenskii, A.O., L.V. Brykova, F.A. Gorel', V.A. Berndnikov, and I.A.

Zalenskaya. 1981. Biochemistry (USSR) 46, 397-401.

Zlatanova, J.S. 1980. Synthesis of histone H1⁰ is not inhibited in hydroxyurea treated Friend cells. FEBS Lett. 112, 199-202.

Zoller, M.J. and M. Smith. 1984. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. DNA 3, 479-488.

END

09|03|88

FIN