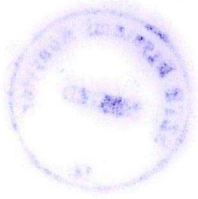
THE STRUCTURE, ORGANIZATION AND EXPRESSION OF PLANT HISTONE H3 GENES

Ph. D. THESIS

By

SHENG-CHENG WU[1,2]

1989



Supervisors:

Prof., Dr. DÉNES DUDITS[1]

Prof., Dr. CHIA-CHEN TAN[2]

Prof., Dr. XUN-MING WANG[2]

1. Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary
2. Institute of Genetics, Fudan University, Shanghai, China

CONTENTS

ACKNOWLEGEMENTS

I express my deep sense of gratitude to my supervisor, Prof., Dr. D.Dudits, for his valuable scientific guidance and generous financial support during the Ph.D. work. I am also very grateful to my Chinese supervisors, Prof. Dr. C.C.Tan and Prof. X.-M.Wang for their scientific supervision, moral encouragement and for providing experimental materials. I thank Dr. G.B.Kiss and Dr. É.Vincze for their courtesy of alfalfa genomic library, Mr. J.Györgyey for his alfalfa cDNA library and screening of H3 cDNA clones, Mr. Z.Végh for his help in DNA sequencing and computer analysis and Mr. L.Bögre for his contribution in studies on somatic embryo culture and ^{3}H-thymidine incorparation experiments. I also wish to thank Miss M. Abraham for her frequent technical assistance, Dr. B.Dusha and Mrs. G.Nöllenburg for their excellent photographic work and other colleagues for technical consultation and essential help in practical work and thoughtful discussions and critical comments on this thesis. Finally, I would extend my great appreciation to the following persons for their generous gifts of hybridization probes:

Prof. R.Wu and Dr. Z.-G.Peng (Cornell University, New York, USA) for rice histone H3 clone, pRH3-2 (Peng and Wu, 1986).

Prof. J.D.Engel (Northwestern University, Illinois, USA) for chicken histone H1, H2a, H2b and H4 clone, pCH1aB1BK1, pCH2eSX30, pCH1aKR-1.3 and pCH3dR8, respectively (Sugarman et al, 1983).

Dr. C.Gigot (Inst. Plant Mol. Biol., Strasbourg, France) for a 196bp clone derived from an Arabidopsis histone H4 gene, H4A748 (Chaboute et al, 1987).

Dr. T.Tabata (Hokkaido University, Sapporo, Japan) for wheat histone H4 clone, pTH011 (Tabata et al, 1983).

ABBREVIATIONS

TCA	trichloroacetic acid
2,4-D	2,4-dichlorophenoxyacetic acid
BAP	benzyl aminopurine
KIN	kinetin
SHM	SH maintenance medium
SHI	SH induction medium
SHR	SH regeneration medium
3'UTR	3' untranslated region
5'UTR	5' untranslated region
RS	repetitive sequence
STR	short tandem repeats
SER	super EcoRI repeats
ISR	interspersed repeats
LTR	long terminal repeats

ABSTRACT

Histone genes are a group of multigene families encoding five small basic polypeptides ubiquitous in eukaryotic organisms. Although a largre numbers of histone genes have been well-characterized in a wide range of organisms, very few of them are from plant species. To extend the knowledge of plant histone genes, two histone H3 genes and one pseudogenes from a monocot plant, Oryza sativa, and one histone H3 genes and four cDNAs from a dicot plant, Medicago sativa, were analyzed. One of the rice H3 genes and the pseudogene are located on a 14.6kb DNA fragment with a 5.8kb spacer separating them. The other rice H3 gene are on a 17.2kb and the alfalfa H3 gene are on a 10.7kb DNA fragment. There are about 50 and 160 copies of H3 sequences in the diploid rice genome and the tetraploid alfalfa genome, respectively. The repetitive copies of H3 sequences are not closely associated in both genomes although histone genes were supposed to be less dispersed in rice genome than in alfalfa or other plant genomes.

In addition to the consensus sequence motifs found in the 5' flanking regions of animal and plant histone genes, multiple copies of the plant-specific motif, CGCGGATC, were detected in the two rice genes and the reverse form of the octamer was found to be adjacent to another specific motif, ACGTCA. The rice H3 pseudogene does not contain any of these specific motifs. Interestingly the pseudogene is associated with a stretch of short tandem repeats in the 5' distal region.

Instead of the animal-specific hairpin structure and its "downstream element", the 3' flanking regions of the rice and alfalfa h3 genes contain unrelated inverted repeats and polyadenylation signal-like sequences. The 3'UTR of alfalfa H3 transcripts were further characterized via cDNA cloning. In contrast to those of animal histone transcripts, the 3'UTR of four alfalfa H3 cDNAs are more than 150bp long and they carry poly(A) tails. The polyadenylation of alfalfa H3 transcripts in plants, calli and somatic embryos was confirmed by Northern blot hybridization.

As other monocot histone genes, the rice H3 genes exclusively prefer G and C to A and T in selection of synonymous codons. In contrast, the dicot alfalfa H3 gene and cDNAs do not follow this coding strategy.

The two rice H3 genes encode an identical protein which varies with that deduced from the alfalfa H3 gene in only one amino acid at position 90, while the alfalfa H3 histone differs from a H3 variant encoded by a minor alfalfa H3 cDNA in four amino acids. Despite these differences, plant H3 histones are highly conserved and they differ from animal H3 histones by four amino acid residues.

The expression of alfalfa H3 genes was studied by Northern blot hybridization in a somatic embryo culture system. The H3 mRNAs are almost undetectable in leaves, more in stem and calli than in roots and highly accumulated in somatic embryos. The significant variation in H3 gene activity probably reflects a complicated regulatory mechanism controlling histone gene expression in higher plants. In partially synchronized callus cells the coincidence of H3 gene expression with DNA synthesis was observed.

Chapter 1.

INTRODUCTION

A Brief Review of Plant Histone Genes

1.1. General Introduction

Histone genes are a group of multigene families encoding five small basic polypeptides ubiquitous in eukaryotic organisms. The histone proteins are the structural elements of chromatin which consists of two components---the nucleosomal particle and the linker region joining the particles. Each of the nucleosomal particle consists of eight core histones, two from each types of the core hiistones (H2a, H2b, H3 and H4), and 146 base pair stretch of DNA wrapped around the core histone octamer (Morse and Simpson, 1988). The histone H1 is associated with the linker region. In the past twenty years, the histone genes have been extensively studied in animals and in lower eukaryotes (review in Kedes, 1979; Hentschel and Birnstiel, 1981; Maxson et al, 1983; Stein et al, 1984). According to the initial studies, the histone genes share a number of structural characteristics. They encode intronless and non-polyadenylated mRNAs; they possess typical RNA polymerase II promoters as well as histone gene-specific consensus sequences, such as the GATCC pentamer, near upstream of the TATA box, and the "cap site", downstream of the TATA box and they have relatively short 5' and 3' untranslated region(UTR). In the 3' flanking regions, most of histone genes carry a highly conserved T-hyphenated dyad symmetric structure and a "downstream element" (Fig.1.1). Now these two elements are known to be essential for the correct and

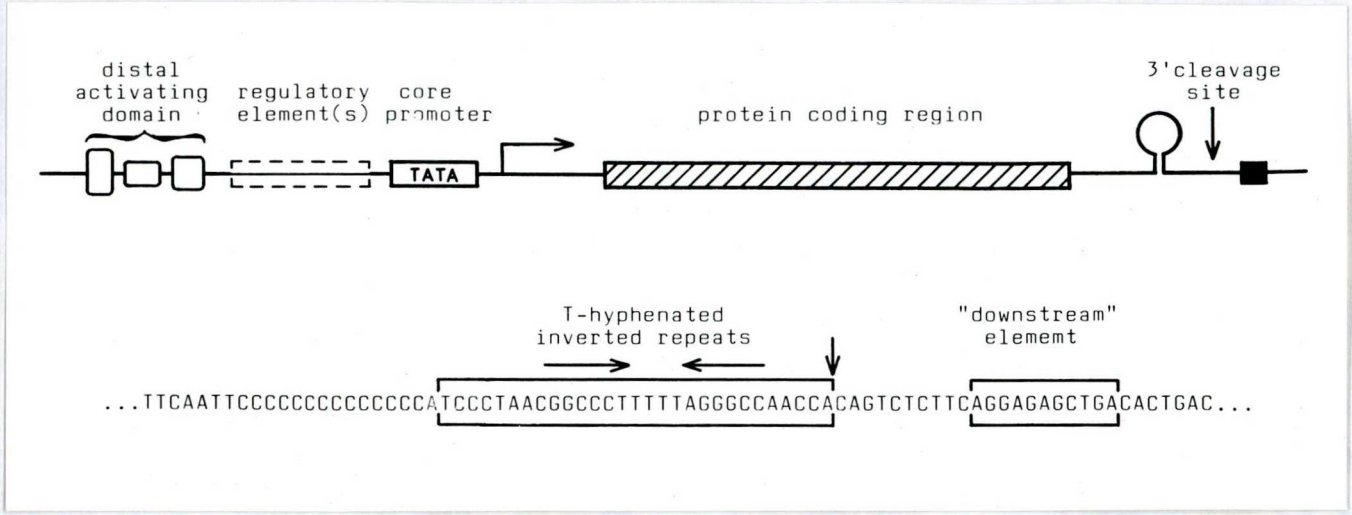


Fig.1.1. Schematic structure of a replication-dependent animal histone gene. This figure is modified from that of Schümperli (1988). In contrast to most other eukaryotic genes, there is no introns. The core promoter consists of the TATA box and the "cap site" where the transcription starts (arrow pointing to the right). The regulatory region contains one or more binding sites for regulated transcription factors. The distal activating domain is composed of multiple constitutive promoter elements, one of which is CAAT box. The signal for 3' processing consists of two conserved sequence elements, the T-hyphenated inverted repeats (hairpin loop) and the "downstream" element (black box). The lower part of this figure is an expanded view of the bipartite RNA 3' processing signal derived from a mouse histone H4 gene (Stauber et al, 1986). Vertical arrow indicates the major 3'cleavage site.

efficient 3'-end processing of histone pre-mRNAs (Birnstiel, 1985; Schaufele et al, 1986; Cotten et al, 1988) and the cell cycle-dependent accumulation of histone mRNA (Schümperli, 1988). Moreover, the regulation of histone genes is typically cell cycle-dependent, with their expression being linked to ongoing DNA replication (Schümperli, 1986; 1988)

Histone genes belong to the class of middle repetitive

sequences with copy number ranging from tens (birds, mammals) to hundreds (sea urchin, frogs, fruit fly, etc). Usually they are clustered and located on certain loci of chromosomes such as in the case of Xenopus borealis (D'Andrea et al, 1985; Grandy and Dodgson, 1987; Turner et al, 1988). However, with the range from highly regular, tandem-repeating quintets, randomly arranged, disperse clusters to non-clustered "orphan" genes, the organization of histone genes does not exhibit a single conserved topology, neither among different genomes nor within a given genome.

Recently, the emphasis in studying the histone genes is laid down not only on the analysis of gene structure and organization, but also on the investigation of molecular mechanism for gene expression and the characterization of variant histone genes. The recent knowledge supplemented with the classical concepts of histone genes can be summarized as the followings:

1. A large number of histone variants and their genes have been detected and studied (Stein et al, 1984; Old and Woodland, 1984). Many of these variant histone genes are more similar to the normal eukaryotic genes than to the histone genes. They are interrupted by introns and transcribed into polyadenylated mRNAs (Engel et al, 1982; Stein et al, 1984; Wu et al, 1986; May et al, 1987; Wells et al, 1987; Wellman et al, 1987; Alonso et al, 1988; Lieber et al, 1988). More intriguingly, histone subtype genes in some species show distinct profiles of expression during development and the course of cell cycle, notably the differential (Ruddell and Jacob-lorena, 1985; Levine et al,

1988; Colin et al, 1988; Carrino et al, 1988), the tissue-specific (Kemler et al, 1986; Trainer et al, 1987; Kim et al, 1987; Brown et al, 1988) and the cell cycle-independent expression (Stein et al, 1984; Wells et al, 1987). The minor histone variants may provide substitution to the predominant histone proteins in chromatin and therefore may play a role(s) in the regulation of chromosome structure and function during different developmental stages or under specific physiological conditions.

2. The cell cycle-dependent regulation of histone genes appears to be controlled by multiple regulatory steps including transcription, pre-mRNA processing, RNA stability and the synthesis of histone proteins (reviews in Schümperli, 1986; 1988; Graves et al, 1987; Marzluff and Pandey, 1988)

3. Another new field in studyiny the histone genes is to determine the regulatory elements governing the histone gene expression. Recently emerged studies suggested that the regulatory elements are located not only in the 5' (Hanly et al, 1985; Oslay et al, 1986; Kroeger et al, 1987; Trainor et al, 1987; Heidl et al, 1988; Dalton and Wells, 1988a; 1988b; Marzluff and Pandey, 1988; Breeden, 1988; Hwang and Chae, 1989), but also in the 3' flanking region (Lüscher et al, 1985; Birnstiel, 1985; Stauber et al, 1986; Levine et al, 1987; Cotten et al, 1988; Stauber and Schümperli, 1988; Schümperli, 1988; Mowry and Steitz, 1988) and probably in the coding region (Grantham et al, 1981; Wells et al, 1986, Thomas et al, 1988; Murray et al, 1989).

In contrast to the enriched information in animals, the study

on plant histone genes had been long ignored until 1982 when an Indian laboratory (Ahmed and Padayatty, 1982; Thomas and Padayatty, 1983a; 1983b; 1984) isolated a 6.64kb DNA fragments carrying histone H2a, H2b and H4 genes from rice embryos. Subsequently, the histone H3 and/or histone H4 genes have been isolated from wheat, corn, Arabidopsis, and rice. Moreover, the cDNAs for histone H3 in barley and H1 in pea have also been isolated and sequenced (Table 1.1). Interestingly, several

Table 1.1. The cloned plant histone genes, pseudogenes and cDNAs

Plant	Clone	Gene	Copy/haploid	Reference
Alfalfa	ALH3-1.1	H3	40	Wu, 1988
	pH3c-1	H3 cDNA		This thesis
	pH3c-3			
	pH3c-7			
	pH3c-10			
	pH3c-11			
	pH3c-12			
	pH3c-17			
Arabidopsis	H3A713	H3	5-7	Chaboute, 1987
	H3A725			
	H4A748	H4	5-7	
	H4A777	H4		
Barley	-	H3 cDNA	10-20	Chojecki, 1986
Corn	H3C2	H3	30-40	Chaubet, 1986
	H3C3			Gigot, 1987
	H3C4			Chaubet, 1986
	H4C7	H4	50-60	Philipps, 1986
	H4C13			Gigot, 1987
	H4C14			Philipps, 1986
Pea	PsH1b	H1 cDNA	-	Gantt, 1986
Rice	pRH3-2	H3	25	Xie, 1987
	pRH3-1	H3 pseud.		
	H3R-11	H3		This thesis
	H3R-12	H3 pseud.		
	H3R-21	H3		
	pIR22	H2A,H2B&H4	-	Thomas, 1983b
Wheat	pTH012	H3	15-18	Tabata, 1984a
	pTH081			Tabata, unpubl.
	pTH011	H4	18-20	Tabata, 1983
	pTH091			Tabata, 1984b

distinct features have been postulated from the limited knowledge of plant histone genes. In comparison with those in animals we present here a brief review on plant histone genes

1.2. The Organization of Plant Histone Genes

The classical picture of histone gene organization was outlined by studies on invertebrate (sea urchin, sea star, fruit fly, etc) where the five main histone genes are usually organized into regular tandem-repeated units on certain loci of chromosomes (Kedes, 1979; Maxson et al, 1983; Stein et al, 1984; Howell et al, 1987). In vertebrates, (birds, mouse, human, etc), however, such regular repeating units are no longer existed, rather, most of the histone genes in these animals are found to be clustered in a random manner. In fact, the organizational topology of histone genes is so diversed that different types of random histone gene clusters can coexist in a genome where the regular tandem-repeating quintets are dominant (Maxson et al, 1983; Stein et al, 1984; Turner et al, 1988). In sea urchin, the histone genes in regular and irregular clusters are transcribed differentially (Maxson et al, 1983; Colin et al, 1988). Furthermore, quite a number of histone "orphans", those who are not closely associated with any other histone genes, increase the complexity of histone gene organization.

In higher plants, the histone genes also belong to the class of intermediate repetitive sequences. The copy number of histone H3 or H4 genes in studied plants is about 20-50 copies per haploid genome with the exception of Arabidopsis (Table 1.1) which possesses the smallest genome size among higher plants

(Meyerowitz, 1989). The Southern hybridization analysis in wheat, rice, corn and Arabidopsis (review in Gigot, 1988; Chaubet et al, 1987) have shown that repetitive copies of histone H3 and H4 genes are located on multiple frements of single and/or double restricted genomic DNA and most of the histone H3 and H4 sequences are not located on the same fragments. These results suggest that both the histone H3 and H4 genes are organized in a very dispersed manner rather than regular repeat units. Nevertheless, as in the case of higher animals, some plant histone genes are found to be clustered in a random manner. The notable plant histone "cluster" is a 6.64kb fragment of rice genomic DNA carrying H2A, H2B and H4 genes with bidirectional transcription (Thomas and Payatthy, 1983). The H2A and H2B genes are closely linked, while the H4 gene is separated by a 3.5kb spacer. A 13kb DNA fragment of wheat genome also carries two histone H3 and H4 genes which are separated by a 1.6kb spacer (Tabata et al, 1983, 1984a). But as the authors proposed, not many of the histone H3 and H4 genes are closely associated in wheat genome (Tabata et al, 1984a). Chaubet et al (1987) recently analyzed the organization of histone H3 and H4 genes in corn and Arabidopsis by using the 5' proximal flanking region as probes. They demonstrated that each probe hybridized to only a few specific bands from all fragments showing hybridization signal with the coding region. These specific bands may represent a certain gene family which is characterized by its specific proximal environments. The low copy number of the histone genes in Arabidopsis also allowed the authors to determine approximately an 11kb spacer between two cloned H3 and

H4 gene, H3A713 and H4A748. Another cloned H4 gene, H4A777 was also shown to be separated from an unidentified H3 gene by a spacer of 4-11kb. Besides, the fact that most of the cloned plant histone genes are "orphans" (Chaubet et al, 1987) also provides an evidence that plant histone genes are not closely clustered. As proposed by Maxson et al (1983), the dispersed organization of histone genes might have resulted from a complex recombination processes. However, one has to consider the role of multiple crosses which is frequently used in plant breeding.

1.3. The Structure of Plant Histone Genes

1.3.1. The coding region

One of the very distinct feature of histone H3 and H4 genes is their high conservation during evolution. Except a rice DNA fragment carrying H2A, H2B and H4 genes, which were identified by hybridization-release translation (Ahmed et al, 1982; Thomas, and padayatty, 1983), all plant histone H3 and H4 genes were isolated by heterologous probes and identified by nucleotide sequence comparison. The homology between plant and animal is about 75-85% for histone H3 or H4 genes and over 95% for their coding products. The plant histone H3 and H4 genes also encode highly conserved small basic proteins with 135 and 102 amino acid residues, respectively, without being interrupted by introns. The plant H4 histones from pea, wheat, corn and Arabidopsis are all identical except that a wheat H4 histone variant deduced from its DNA sequence differs from others by one amino acid change at position 4 (Gigot, 1988; Spiker, 1988). The plant H3 histones are slightly less conserved than H4 histones.

H3 histone variants with up to four amino acid replacements were observed among different or within the same plant species (Spiker, 1988; Gigot, 1988). Interestingly, the amino acid at position 90 seems to be highly variable. It may be a serine, alanine, leucine or arginine (Gigot, 1988).

At the nucleotide level, the homology of histone H3 and H4 genes does not reflect any phylogenetic relationship. The homology can range from 99% to 75% independent of the genome in which the histone genes accomodate (Wells, 1986). This seems also true for plant histone H3 and H4 genes. Because of the high conservation of H3 and H4 histones, the nucleotide divergence is mainly due to the different codon preferences of individual genes. It has been noticed that all the cereal histone genes initially studied prefer exclusively C and G to T and A in their codon choice (Tabata and Iwabuchi, 1983, 1984a, 1984b; Chaubet et al, 1986; Phillips et al, 1986; Xie et al, 1987). The wobble nucleotides used by these genes are composed of over 95% G and C. In some genes, the A-ended codons were never used (Xie et al, 1987). Interestingly, all the four H3 and H4 genes from a dicotyledonous plant, Arabidopsis, do not possess such GC-biased codon usage. It should be worthwhile to survey the coding strategy of histone genes in other dicotyledonous plants. In other eukaryotic genes too the non-random use of synonymous codons has been well documented (Grantham et al, 1981; Murray et al, 1989). Several recent studies have suggested that the coding strategy may reflect gene expressivity (Grantham et al, 1981; Wells et al, 1986; Ernst, 1988; Thomas et al, 1988).

1.3.2. The 5' flanking region

In animals, the flanking regions of histone genes possess several specific motifs. Two of these are in the 5' flanking region, the GATCC pentamer and the "cap site", pyCATTCpu (Hentschel and Birnstiel, 1981). To find out whether plant histone genes exhibit the similar properties, Gigot (1988) has made a comprehensive analysis. He showed that these two specific motifs seemed to exist in plant histone genes with slight modifications. Based on this study, the "cap site" in higher plants is pyCAA/Cpy and the pentamer motif is GCpypyC. But the most intriguing finding is that all the analyzed plant histone H3 and H4 genes carry a highly conserved octamer motif, CGCGGATC, about 200-250bp upstream from the initiation codon. In a few cases, the "CGC box", as designated by Gigot (1988), exists in its reverse form, GATCCGCG. It is unknown whether the orientation of the CGC box has any functional significance. Another plant specific hexamer motif, ACGTCA, originally postulated from wheat histone genes, was also seen in several plant histone genes (Mikami et al, 1987). This hexamer is not found in all the plant histone genes and seems to be less conserved than the CGC octamer. Gel mobility shift assays have showed that this hexamer is the target of a nuclear binding protein(s), HBP-1 (Mikami et al, 1987). The functions of these plant histone gene-specific consensus sequences are unknown so far. However, the recently developed techniques (Schell, 1987, Jefferson, 1987) for reintroducing cloned genes into plant cells have paved the way to the functional analysis of promoter elements involved in regulation. One pioneer experiment carried out by Tabata and Iwabuchi (1987) has demonstrated the faithful

expression of wheat histone H3 and H4 genes in transgenic sunflower seedlings. In yeast and animal, several reports on the characterization of histone promoters involved in cell cycle regulation have recently emerged (review in Breeden, 1988; Schümperli, 1988). These studies revealed that approximately 200 bps of histone 5' flanking region, which includes the regulatory elements and the distal activating domain (Fig.1.1), were sufficient to confer cell cycle regulation on a linked reporter gene. The regulatory elements might exist in a longer 5' flanking region in plant histone genes, since the 5' UTR in plant is generally longer and sometimes the plant specific motifs are found over 250bp away from the initiation codon.

1.3.3. The 3' Flanking Region: Are the Plant Histone mRNAs Polyadenylated?

Evidence documented in the past few years have indicated that, while the 5' flanking region governs gene transcription, the 3' flanking region regulates gene expression at the post-transcriptional level (Birnstiel, 1985; Platt, 1986). Most of eukaryotic genes possess polyadenylated signal and are transcribed into polyadenylated mRNAs. In contrast, most animal histone genes carry two distinct consensus structures and encode non-polyadenylated mRNAs. These two specific structures are the T-hyphenated inverted repeats and its "downstream element" (Fig.1.1). Both of them are involved in the 3' processing, an event mediated by U7 small nuclear RNA (Cotten et al, 1988). The T-hyphenated dyad symmetry element was also shown to be essential for the cell cycle-dependent regulation (review in Schümperli, 1988).

In plant histone genes, dyad symmetry structures were occasionally observed in their 3' flanking regions. But these inverted repeats are rarely similar to the animal ones and they are not followed by the "downstream element". The 3' ends of several plant histone genes from wheat (Tabata and Iwabuchi, 1986), corn (Chaubet et al, 1988), Arabidopsis (Chaboute et al, 1988) and barley (Chojecki, 1986) have been determined by S1 nuclease protection experiments or cDNA cloning. These experiments indicated that the 3'UTR of plant histone genes is more than 150bp, much longer than that of animal histone genes. The latter is about 80bp long. These differences between the 3' flanking regions of plant and animal histone genes suggest that the 3' processing of plant histone pre-mRNAs is different from that of animal histone pre-mRNAs. This suggestion was confirmed at least in some cases when a French group demonstrated that the maize histone H3 and H4 mRNAs are polyadenylated (Chaubet et al, 1988). Subsequently, they again by using the same methods showed the polyadenylation of histone H3 and H4 mRNAs in Arabidopsis and, probably, tobacco and sunflower. The data presented in this thesis also proved the polyadenylation of independent histone H3 transcripts in alfalfa. The cDNAs for a pea H1 and a barley H3 histone were also found to carry a polyA tract. However, an earlier study on rice defined the histone mRNAs in the polyA(-) fraction of total rice RNA (Ahmed et al, 1982). Therefore the literature is insufficient to provide a general picture about the polyadenylation status of plant histone mRNAs.

1.4. The Expression of Plant Histone Genes.

It is well known that the expression of histone genes is cell cycle-dependent. The cell cycle dependence of histone gene expression is regulated by multiple mechanisms at different levels, mainly transcription, pre-mRNA processing, mRNA degradation and histone protein synthesis (Stein et al, 1984; Graves et al, 1987; Marzluff and Pandy, 1988; Schumperli, 1986, 1988). The coupling of histone gene expression with the DNA synthesis is the main characteristic of most histone genes in various organisms from yeast to man. Beside that, there are several other regulatory mechanisms related to histone gene expression, such as the temperally uncoupled expression during oogenesis and early embryogenesis in some animal species (Maxson et al, 1983; Stein et al, 1984; Old et al 1985), the cell cycle-independent expression of mammalian basal histone H3.3 gene (Stein et al, 1984; Wells et al, 1987) and the development and tissue-specific expression of minor histone gene variants (Stein et al, 1984; Kemler and Busslinger, 1986; Kim et al, 19887; Lieber et al, 1988; Alonso et al, 1988).

In higher plants, many studies has been carried out on plant histones (Spiker, 1982; 1985; 1988; Waterborg et al, 1987; Mazzolini et al, 1989). In many of these experiments, plant histones from various tissues were mainly characterized, in comparison with animal histones, by gel electrophoresis and immunological assays. Unfortunately, these studies gave very little information about gene regulation. The plant histone mRNAs were first studied in an Indian laboratory by Ahmed and Padayatty (1982). They located the rice histone mRNAs in the polyA(-) fraction of total RNA isolated from 18-day-old

germinated rice embryos by oligo-dT cellulose chromatography and subsequently identified several individual histone mRNAs by cell free translation. The histone mRNAs were found to be very abundant in the 18 days old rice embryos. Recent studies also indicated that the histone mRNAs were highly accumulated in germinated maize embryos (Chaubet et al, 1987, 1988) and alfalfa somatic embryos (Wu et al, 1988). These observations seem to be in agreement with the quick and cell cycle-independent transcription of histone genes during early embryogenesis in animal (Maxson et al, 1983). Consistently, Kato et al (1982) and Zlatanova and Ivanov (1988) have respectively demonstrated that the DNA and histone synthesis are uncoupled in early germination of Vicia faba and maize embryos.

Unfortunately, the expression of plant histone genes in somatic tissues has not yet been studied and it is not known how the plant histone genes are regulated in somatic cells. Since the plant histone genes are structurally very different from animal histone genes, particularly in the 3' flanking region which plays an important role in cell cycle regulation of animal histone genes (Stauber et al, 1986; Stauber and Schümperli, 1988; Schümperli, 1988), one may speculate pecific regulatory mechanisms for plant histone genes.

1.5. Summary

In summary, based on the limited studies discussed above, we have seen that plant histone genes are certainly different from cell cycle-regulated animal histone genes in several aspects (Table 1.2). The main characteristics of plant histone genes are

Table 1.2. Comparison of animal and plant histone genes

	Invertebrates	Vertebrates	Plants
Copy number	100-800	10-60	5-60
Organization	Quintet regular, tandem repeat	Random cluster, no regular repeat	Occasional cluster? no regular repeat
Structure	Short UTR Majority no intron Majority no polyA mRNA		Relative long UTR No intron found All polyA(+) mRNA?
Specific motif	3' T-hyphenated inverted repeats		5' CGCGGATC
Regulation	Cell cycle-dependent in somatic cell Cell cycle-independent during oogenesis and early embryogenesis		? ?

the followings: (1) The multiple copies of plant histone genes usually exist as multigene families and they are not closely associated as animal histone genes are. (2) The presence of plant specific consensus sequences in the 5' flanking region and the absence of animal-specific structural elements in the 3' flanking region. (3) The polyadenylation of histone transcripts in maize and Arabidopsis.

Structure is always related to its function. The differences between plant and animal histone genes suggest that the mechanisms for histone gene evolution, expression and regulation adopted by these two biological categories might be somewhat different. However, these conclusions may still be premature to serve as general rules for histone genes in all plant species, since so far studies on histone H3 and H4 genes are restricted to three monocot cereals and one dicot weed. Furthermore, genes

for the less conserved H1, H2a and H2b histones have not yet been analyzed.

In this thesis, we present data about the organizational and structural analysis of three histone H3 genes, one H3 pseudogenes and four H3 cDNAs from a monocotyledonous plant, Oryza sativa, and a dicotyledonous plant, Medicago sativa. The expression of alfalfa histone H3 genes was studied by using the somatic embryo culture system of alfalfa as a developmental system (Stuart et al, 1985). The presented eight DNA sequences also allowed us to analyze the characteristics in the coding regions of plant histone H3 genes.

Chapter 2.

MATERIALS AND METHODS

2.1. Plant Sources:

2.1.1. Oryza sativa L, subsp. Japanica, cv. early 2439 bred by Y.-X.Cai et al (1985, unpublished) was used for the construction of rice genomic library. Rice seeds were obtained from Jiaxing Institute of Agroscience, Zhejiang Province, China.

2.1.2. Medicago sativa L cv. Nagyszenási commercially available in Hungary was used for the construction of alfalfa genomic library.

2.1.3. Medicago sativa L cv. Regen S. isolate RA3 kindly provided by Stuart (1985) was used for the construction of alfalfa cDNA library, tissue culture and gene expression experiments.

2.2. Chemicals and Enzymes:

Common chemicals were either Hungarian or Chinese products, Fine chemicals were purchased from Sigma or Serva. Most of the enzymes were from Biolab, a few of them were BRL or Boehringer/Mannheim products.

2.3. Plant Tissue Culture:

General procedure for somatic embryo induction was used as described by Stuart et al (1985) with a few modifications. The soft and rootless calli were initiated from roots on agar maintenance medium (SHM), i.e. SH medium (Schenck and Hildebrandt, 1972) supplemented with 1uM 6-benzyl aminopurine

(BAP) and 5uM of 2,4-dichlorophenoxyacetic acid (2,4-D). 10 grams of three week old calli were incubated in 150ml of induction medium (SHI), i.e. SH with 50um of 2,4-D and 5uM kinetin (KIN), for three days with constant shaking (120rpm) at 25°C. The induced calli were then transferred to agar regeneration medium (SHR), i.e. SH with 30mM of proline and 10mM of $(NH_4)_2SO_4$. Somatic embryos became visible after two weeks. For selection of somatic embryos at different stages, 30 day old cultures were used. The whole procedure for somatic embryo induction is illustrated in Figure 3.2.4. The in vitro grown alfalfa plants used for callus initiation and RNA preparation were grown in agar UM medium (Uchimiya and Murashige, 1974) from nodal cutting without adding plant hormones.

2.4. ^{3}H-Thymidine Incorporation:

The procedure was that modified from F.Boldog (Ph.D thesis, Eötvos Lorant University, 1986). The details are described as the followings (steps 4-6 were carried out at 4°C):

1) RA3 calli were collected and suspended in SHI medium (10g/100ml). For control, SHM was used instead.

2) The suspended cell clusters (10ml) were distributed into 25ml flasks and cultured at 25°C with constant shaking (120rpm).

3) Each day callus cells were pulse-labeled with 5.55MBq (150 uCi) of ^{3}H-thymidine per flasks for 12 hours and then collected by filtration. Three labelings were carried out in parallel.

4) The labeled cells were homogenized in a potter homogenizer with 2ml of 7% trichloroacetic acid (TCA). Precipitation of the homogenate was carried out by centrifugation at 20,000g for 10

minutes.

5) The pellets were washed several times with 2ml of 7% TCA until no radioactivity was measured in the supernants and then once with 2ml of ethanol and ether (3:1) by resuspending and centrifugation.

6) The pellets were suspended again with 2ml of ether and then the suspensions were dried by filtration on discs of GF-C glass fiber paper (Whatman).

7) The radio activity on each discs was measured in scintillation cocktail. Final data were the mean cpm from the three sets of labeling.

2.5. Construction and Screening of Genomic Libraries:

2.5.1. Rice: The construction of the rice genomic library has been described previously (Wu et al, 1987). In brief, 10 days old rice seedlings grown in dark at about 28^{o}C were used for total DNA isolation. The method was that described by Murray and Thompson (1980). Purified total DNA was partially digested with MboI and fractioned by density gradient centrifugation in 10-40% sucrose as described by Maniatis et al (1982). The 15-20kb rice DNA fragments were then ligated to BamHI/EcoRI-digested EMBL3 vector DNA (Frischauf et al 1983). The recombinant DNAs were packaged *in vitro* (Maniatis et al, 1982). Finally the rice genomic library was amplified by growing the phages in *E.coli* strain Q359 (Maniatis et al, 1982) and stored at -70^{o}C. To isolate histone genes, about 100,000 phages were grown in *E. coli* strain K802 (Maniatis et al, 1982) and screened by *in situ* plaque hybridization (Benton and David, 1977) with appropriate

DNA probes according to standard procedure (Maniatis et al, 1982).

2.5.2. Alfafla: The alfalfa genomic library was constructed by É.Vincze et al (in preparation). Total DNA was isolated from young alfalfa plants, purified and fractioned as described above. 10-15kb alfalfa DNA fragments were inserted into BamHI-digested pGY97 DNA. pGY97 is a fused vector of lambda EMBL4 (Frischauf et al, 1983) and plasmid pBR322. It can be used either as a plasmid or bacteriophage lambda vector. The recombinant DNAs were packaged in vitro (Maniatis et al, 1982) and the packaged phages were infected into E.coli K-1400 (Cami and Kourilsky, 1978) and grown as bacteria at 28°C in the presence of ampicillin and tetracycline (É.Vincze et al, in preparation). Finally, the bacterial colonies were harvested and stored in 15% glycerol at -70°C in small aliquots. For screening the library, about 100,000 recombinants were induced at 42°C and grown in K802 as phages. The resulting plaques were then in situ hybridized with appropriate DNA probes.

2.6. Southern Blot Hybridization and Estimation of Copy Number:

Genomic DNA or cloned DNAs were digested with appropriate restriction enzymes, separated on agarose gels and blotted according to the standard methods (Maniatis et al, 1982). The copy number of histone H3 genes was estimated according to Gullis (Gullis et al, 1984). Cloned DNA and genomic DNA was dot blotted onto a nitrocelullose filter according to White et al, (1982). The filters were hybridized with appropriate DNA probes. After hybridization, the filters were exposed to X-ray film. For

copy number estimation, the radioactivity of each dot was measured and used for calculation. Hybridization was carried out in a solution containing 50% formamide, 3 x SSC, 0.1% SDS, 0.25% milkpowder (Gloria, Belgium) and $2x10^7$ cpm/ml of ^{32}P-labeled probe DNA (random primer labeling method, Feinberg and Vogelstein, 1983) at 42^oC for 24-30 hours. The filters were washed in 0.1 x SSC and 0.1% SDS at 67^oC for 5 hours.

2.7. RNA Preparation:

Total RNAs were prepared by the method of guanidiniumthiocyanat-LiCl precipitation described by Cathala et al (1983) and quantified by spectrophotometric assay as well as agarose gel electrophoresis. PolyA(+) RNAs and polyA(-) RNAs was prepared from purified total RNAs by oligo-dT cellulose chromotography (Aviv et al, 1973).

2.8. Northern Blot and Dot Blot Hybridization:

Total RNAs, polyA(+) and polyA(-) RNAs were separated on formaldehyde agarose gels and blotted onto nitrocellulose filters (Maniatis et al, 1982). The procedure used for dot blot was that of White et al (1982). Conditions were the same as used for Southern blot hybridization.

2.9. Construction and Screening of Alfalfa cDNA Library:

The cDNNA library was constructed by J.Györgyey et al (unpubl.). RA3 calli were incubated in SHI medium for 60 hours. Total polyA(+) RNA was prepared from the induced calli and used for the construction of cDNA library. The cDNA library was

constructed according to Gubler and Hoffman (1983) in an *in vitro* transcription vector, pGEM-2. Screening of the library was performed according to standard colony *in situ* hybridization procedure (Maniatis et al, 1982).

2.10. DNA Sequencing:

DNA fragments were subcloned into M13 mp18/19 vectors according to Yanisch-Perron et al (1985) and sequenced by the dideoxy chain termination method (Sanger et al, 1977). The sequencing kit (Sequenase) and protocol were provided by United States Biochemical Corporation (USB).

2.11. Computer Analysis: All the DNA sequence data were handled by appropriate programmes from the Cornell DNA Sequence Analysis Package (Fristensky et al, 1982) run on an IBM/DOS computer. For homology searching, Microgenie programme and the DNA data library was that released by GenBank in March, 1988.

Chapter 3.

RESULTS

3.1. THE STRUCTURE AND ORGANIZATION OF RICE HISTONE H3 GENES

3.1.1. Organizations of Isolated Rice Histone H3 Genes

Although two associated histone H3 gene and pseudogene have been identified in a rice genomic clone, λRH3-17, most of the 50

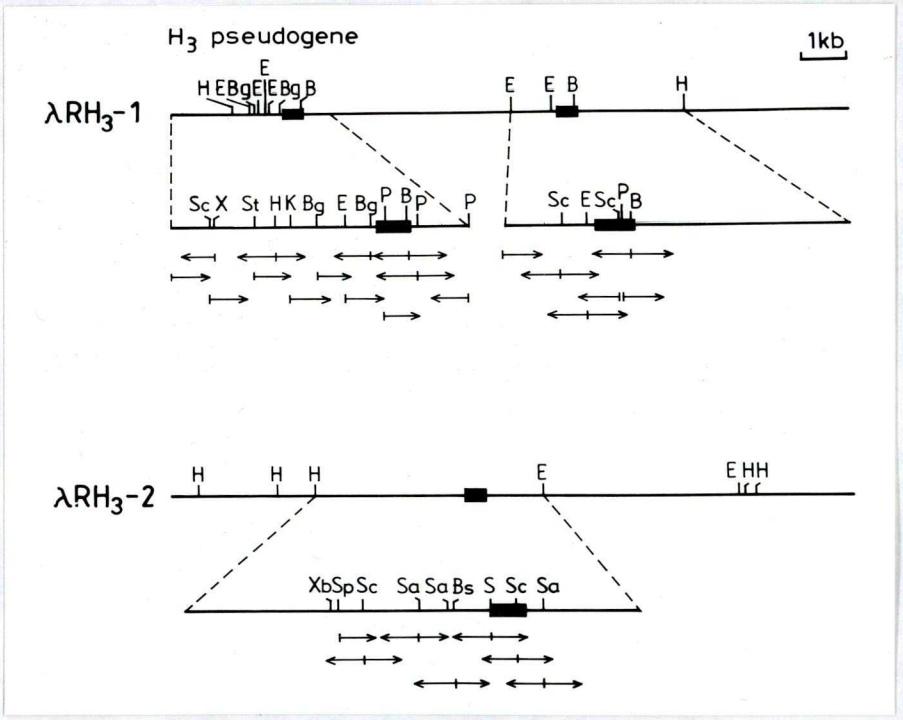


Fig.3.1.1. The restriction maps of λRH3-1 and λRH3-2 and the sequencing strategies (indicated by arrows) of H3 gene subclones, H3R-12 (left) and H3R-11 (right) from λRH3-1 and H3R-21 from λRH3-2. The solid boxes represent coding regions. B=BamHI, Bg=BglII, Bs=BssHII, E=EcoRI, H=HindIII, K=KpnI, P=PstI, S=SalI, Sa=ScaI, Sc=SacI, Sp=SphI, St=StuI, X=XhoI and Xb=XbaI.

copies of H3 sequences remain unknown. In addition the flanking regions of the two cloned H3 gene (pRH3-2) and pseudogene (pRH3-1) were not well characterized (Peng and Wu, 1986; Xie et al, 1987). For further characterization of the structure and organization of rice histone H3 genes, a PstI/PvuII fragment of clone pRH3-2 was used as a probe to screen a rice genomic library (Wu et al, 1987). Two positive clones were isolated from about 100,000 recombinants. Figure 3.1.1. shows the restriction maps of these two rice histone H3 clones. λRH3-1 carries two copies of H3 sequences but shows different organization from that of λRH3-17 (Peng and Wu, 1986)). These two H3 sequences are separated by a 5.8kb spacer and localized on a 0.7 and a 0.5kb of EcoRI/BamHI fragments, respectively. Clone λRH3-2, however, possesses only one copy of H3 gene with distinct restriction pattern as compared to λRH3-1 or λRH3-17.

3.1.2. The Structure of Rice Histone H3 Genes

The nucleotide sequences of the rice histone H3 genes derived from clone λRH3-1 and λRH3-2 were compiled with those from clone λRH3-17 in Figure 3.1.2. The compilation analysis revealed several characteristics of the rice histone H3 genes.

3.2.3.1. The coding region of the rice H3 genes

The two H3 sequences subcloned from λRH3-1 and one H3 sequence subcloned from λRH3-2 were designated as H3R-11, H3R-12 and H3R-21, respectively (Fig.3.1.1). Despite different organization, the coding region of the rice histone genes and pseudogenes are highly conserved. The homoloy among these sequences is over 92% and up to 96.8% between H3R-11 and pRH3-2

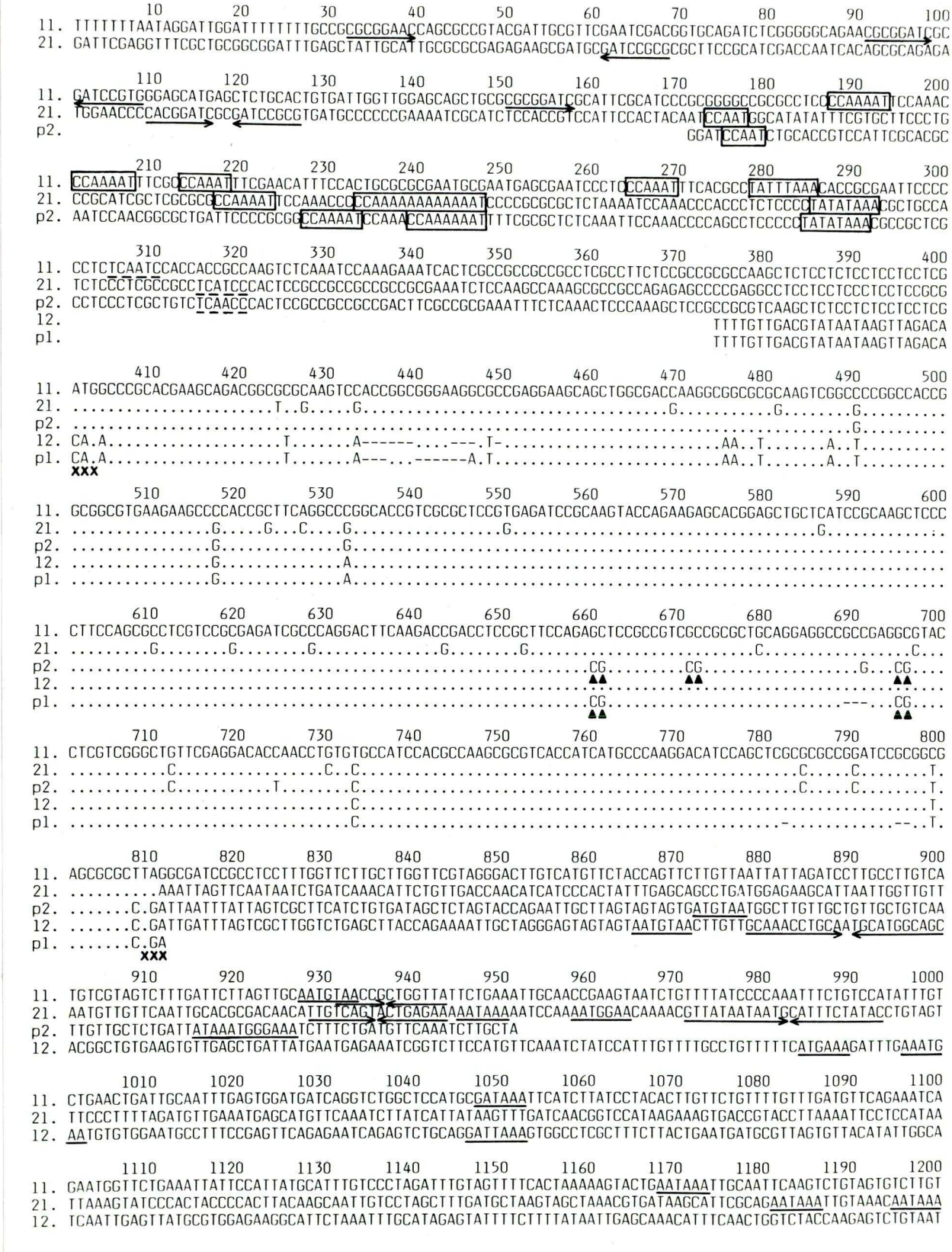


Fig.3.1.2. The nucleotide sequences of rice histone H3 genes and H3 pseudogenes. In the coding regions, the initiation and termination codons are marked with Xs, the putative sequencing errors in pRH3-2 and pRH3-2 are indicated with solid triangles and dashes represent deletions in the pseudogenes. In the 5'-flanking regions, the TATA box and CAAT box are framed, the "cap sites" are shown with dashed lines and the direct or

reverse "CGC boxs" are indicated with arrows in either directions. In the 3'-flanking regions, the palindromic sequences are indicated by inverted arrows and the polyadenylation signal-like sequences are underlined. 11=H3R-11, 21=H3R-21, p2=pRH3-2, 12=H3R-12 and p1=pRH3-1. The sequences of pRH3-1 and pRH3-2 were obtained from Xie et al (1987).

(Table 4.3). As previous studies on pRH3-2 (Xie et al, 1987) and other cereal histone genes (review in Gigot,1988), both H3R-11 and H3R-21 display extremely GC preference at the codon position III (Table 4.4). Among 135 codons excluding the start and stop ones, A-ended codons were never used, while T-ended codons were used only three times in each of the rice genes. 23 of the 25 nucleotide replacements between the coding regions of H3R-11 and H3R-21 were the changes between G and C (Figure 3.1.2). Both H3R-11 and H3R-21 encode an identical protein. This protein is identical to those deduced from corn (Chaubet et al, 1986) and Arabidopsis (Chaboute et al, 1987) H3 genes and it varies with those from pea (Patthy et al, 1973), wheat (Tabata and Iwabuchi, 1984) and alfalfa (Wu et al, 1988) in only one amino acid at positon 90 (Fig.4.5). But surprisingly, the H3 histone encoded by H3R-11 and H3R-21 varies with that (H3r) deduced from the nucleotide sequence of pRH3-2 at three positions despite the

```
          10        20        30        40        50        60        70
1 ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFRPGTVALREIRKYQKSTELLIRKLPFQRL
2 ...........................................................................
3 ...........................................................................
4 ...........................................................................

          80        90       100       110       120       130
1 VREIAQDFKTDLRFQSSAVXALQEAAEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA
2 ..............T...R.......R.....................................
3 ..................A.....................................
4 ..................A.....................................
```

Fig.3.1.3. The amino acid sequences of rice H3 histones derived from their genes. 1, Consensus sequence of plant H3 histones (Wu et al, submitted; also see Fig.4.5) and 2, H3r deduced from pRH3-2 and 3-4, H3 histones deduced from H3R-11 and H3R-21, respectively.

extensive homology (60%) between H3R-11 and pRH3-2 (Fig.3.1.3). The sequence alignment in Figure 3.1.2 indicates that all these amino acid changes in H3r are due to the dinucleotide substitutions of 5'CG to 5'GC. Two identical dinucleotide substitutions are also observed in the previously published pseudogene, pRH3-1. It is possible that these unusual dinucleotide changes in pRH3-2 and pRH3-1 might have resulted from DNA manipulation errors.

In comparison to other rice H3 genes, H3R-12 consists of a total 10bp deletions inside the "coding region" and 11 nucleotide substitutions of G or C to A or T. Some of these substitutions lead to amino acid replacements(not shown). This sequence also lacks an initiation codon and the entire 5' flanking region which characterizes the rice histone H3 genes (see below). In addition, H3R-12 is identical to the previously published rice H3 pseudogene, pRH3-1 (Xie et al, 1987), except that pRH3-1 has five more base pair deletions and contains two dinucleotide substitutions at position 660-661 and 695-696 (Fig.3.1.2). Therefore it is reasonable to conclude that H3R-12 represents a histone H3 pseudogene.

3.1.2.2. <u>The 5' flanking region of rice histone H3 genes</u>

Comparing the 5' flanking sequences of the rice histone genes, the "orphan" gene H3R-21 shows very limited homology to H3R-11 or pRH3-2, while the latter two exhibit relatively high degree of homology (Fig.3.1.4). Despite the homology limitation, however, a number of conservative and distinct sequence motifs were observed among the rice histone genes (Fig.3.1.5). Besides the TATA box and the "cap site", several CnAnTn sequences appear

in each of the rice H3 genes (Fig.3.1.2). These CnAnTn motifs may be the divergent types of CCAAT box as suggested by Anderson et al (1984). Interestingly, the plant histone gene-specific CGC box (Gigot, 1988) exists in the 5' flanking region of H3R-11 and H3R-21 as multiple copies (three copies in H3R-21 and four

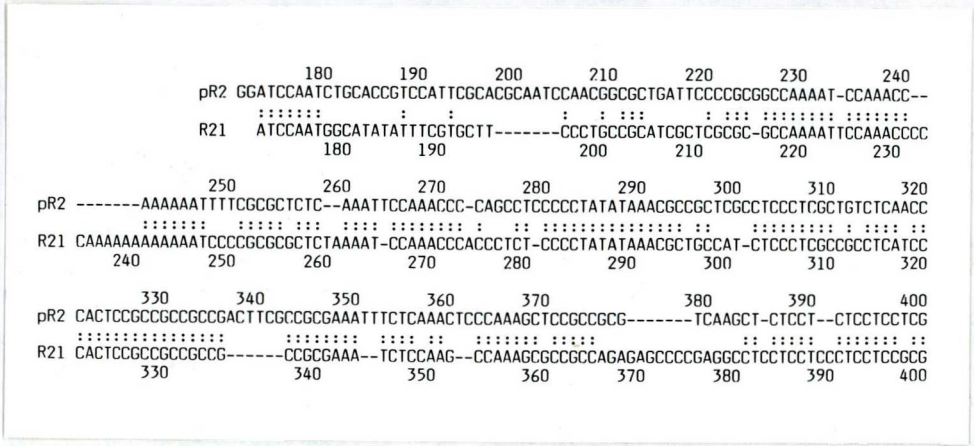


Fig.3.1.4. The homology between the 5' flanking regions of pRH3-2 (pR2) and H3R-21 (R21). sequence numberings correspond to Figure 3.1.2. Dashed are introduced for maximal homology.

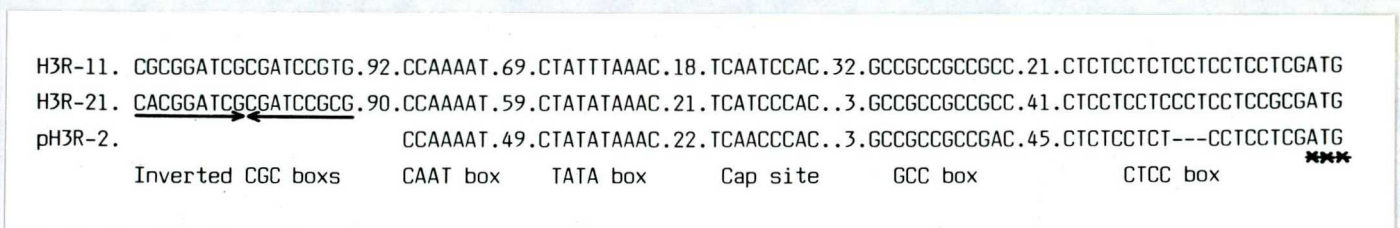


Fig.3.1.5. Conserved sequence motifs in the 5' flanking region of rice histone H3 genes. The asterisks mark the initiation codon and the numbers indicate distances (bp) between motifs.

copies in H3R-11). Two inverted repeats of CGC boxes in both genes are closely linked and they could form a stable hairpin structure (Fig.3.1.2, 3.1.5). The other two CGC boxes in H3R-11 exist as direct form at position 32 and 150, respectively, whereas the third CGC box in H3R-21 exists as reverse form at position 61 (Fig.3.1.2). Additionally, the rice H3 genes also possess other specific consensus sequences. These are a stretch of four GCC repeats designated as "GCC Box" downstream of the "cap site" and a stretch of CT repeats designated as "CTCC Box" adjacent to the initiation codon (Fig.3.1.2 and 3.1.5). The conserved sequence motifs of the rice histone H3 genes are summarized in Figure 3.1.5. In contrast, the 5' "flanking region" of the H3 pseudogene, H3R-12, does not exhibit any of these characteristic motifs (Fig.3.1.7A).

3.1.2.3. The 3' flanking region of rice histone H3 genes

The 3' flanking sequences of the rice histone H3 genes are not as GC-rich as the coding region and the 5' flanking sequence, instead they exhibit slightly AT-rich. Among the five 3' flanking sequences, only a very limited degree of homology

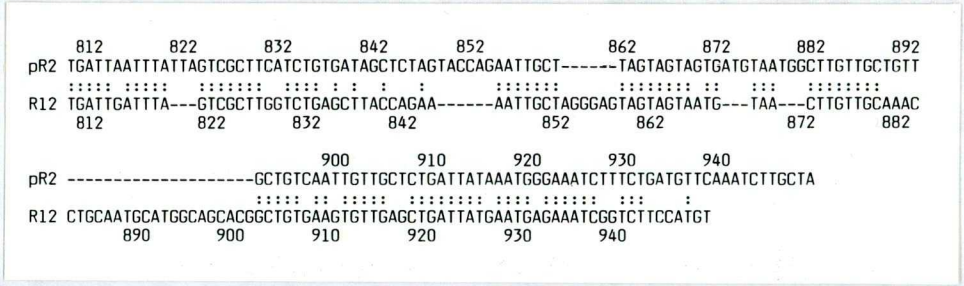


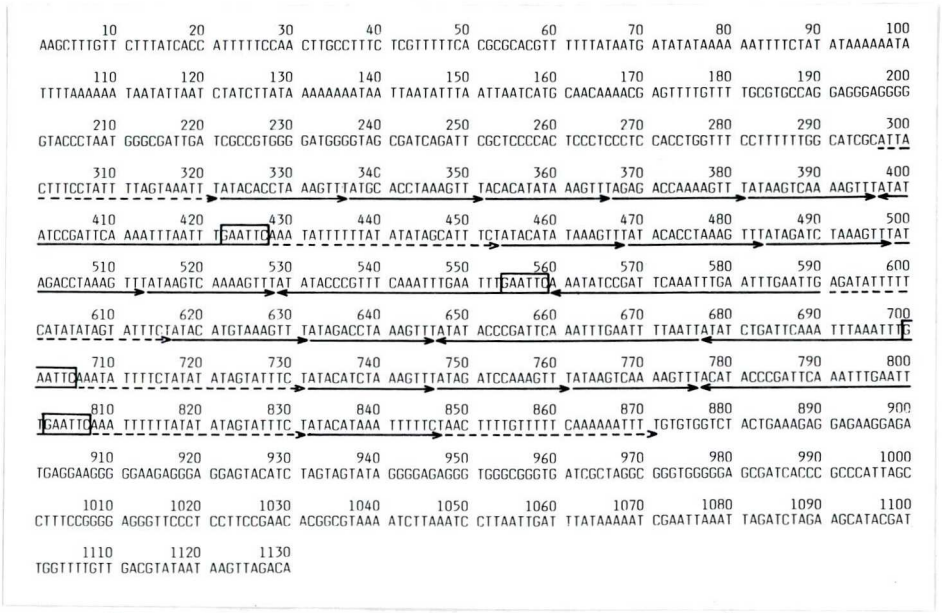
Fig.3.1.6. The homology between 3' flanking regions of pRH3-2 (pR2) and The H3 pseudogene, H3R-12 (R12). sequence numberings correspond to Figure 3.1.2. Dashed are introduced for maximal homology.

between pRH3-2 and H3R-12 was observed (Fig.3.1.6). The T-hyphenated inverted repeats and its "downstream element" of animal histone genes were not detected in the rice H3 genes, rather, several polyadenylation signal-like sequences and short inverted repeats were observed (Fig.3.1.2).

3.1.3. The Rice H3 Pseudogene Is Associated With Short Tandem Repetitive Sequences

To characterize the "5' flanking region" of the H3 pseudogene, H3R-12, the entire "upstream region" of this pseudogene cloned in λRH3-1 (Fig.3.1.1) was sequenced. The complete sequence of this region is shown in the Appendix-2. Figure 3.1.7A includes only the most characteristic 1130 bp sequences adjacent to the "coding region" (Fig.3.1.2). Occasionally interrupted by stretchs of GA-rich sequences, the first 1500 bp "upstream sequence" exhibits extensively AT-rich. More intriguingly, this region includes a stretch of repetitive sequences (RS). Figure 3.1.7A shows the sandwich structure containing the RS and its flanking sequences: 5' AT-rich/G-rich/RS/G-rich/AT-rich 3'. The RS itself is a mosaic composition of three different repeating elements, the 16 bp short tandem repeats (STR), the 26 bp interspersed repeats (ISR) and the 31 bp super-EcoRI repeats (SER). In addition, the AT-rich regions which flanks the G-rich/RS/G-rich structure also display short irregular repetitive sequences which are not specified in Fig.3.1.7.

A



B

```
ATACACCTAAAGTTT
ATGCACCTAAAGTTT
ACACATATAAAGTTT
AGAGACCAAAAGTTT
ATAAGTCAAAAGTTT
ATACATATAAAGTTT
ATACACCTAAAGTTT
ATAGATCTAAAGTTT
ATAGACCTAAAGTTT
ATAAGTCAAAAGTTT
ATACATGTAAAGTTT
ATAGACCTAAAGTTT
ATACATCTAAAGTTT
ATAGATCCAAAGTTT
ATAAGTCAAAAGTTT
ATACAT--AAATTTT
```

D

```
ATTACTTTCCTATTT-TAGTAAATTT
AAATATTTTTTATATATAGCATTTCT
AGATATTTTTCATATATAGTATTTCT
AAATATTTTCTATATATAGTATTTCT
AAA--TTTTTTATATATAGTATTTCT
TAACTTTTGTTTTTCAAAAAATTTTG
```

E

```
ATAT--CCGATTCAAAATTT-AATTTGAATTC
ATATACCCGTTTC-AAATTTGAATTTGAATTC
AAATATCCGATTC-AAATTTGAATTTGAATTG
ATATACCCGATTC-AAATTTGAATTTTAATT-
ATAT--CTGATTC-AAATTTAAATTTGAATTC
ACATACCCGATTC-AAATTTGAATTTGAATTC
```

C

A	T	A	C	A	T	C	T	A	A	A	G	T	T	T
16	14	15	8	13	10	12	10	16	16	16	15	16	16	16
	C	G	G	G	C	A	A				T			
	1	1	5	3	6	2	4				1			
	G		A			G	C							
	1		3			1	1							

Fig.3.1.7. Repetitive sequences (RS) in the "5'-flanking region" of the rice H3 histone pseudogene, H3R-12. (A) The upstream sequence containing the short tandem repeats (STR). Numbering from 5', the STRs, super-EcoRI repeats (SER) and interspersed repeats (ISR) are indicated with direct, reverse and dashed arrows, respectively. The super-EcoRI sites are framed. (B) Sequence comparison of the STRs. (C) Frequency of nucleotide used in the STR. (D) Sequence comparison of the ISRs. (E)Sequence comparison of the SERs. All the repetitive sequence elememts listed from top to bottom in (B), (D) and (E) correspond to those from 5' to 3' in (A).

3.2. CHARACTERIZATION OF HISTONE H3 GENES AND THEIR TRANSCRIPTS IN ALFALFA (Medicago sativa)

3.2.1. Isolation and Structural Analysis of an Alfalfa Histone H3 Gene

To date, about 20 histone genes have been isolated from three monocotyledonous crops, but only four H3 and H4 genes from one dicotyledonous plant (Table 1.1). To study histone genes from a broader taxonomic groups, an alfalfa genomic library was screened for H3 histone clones by in situ plaque hybridization. A BamHI/EcoRI fragment of the rice histone H3 gene, H3R-11 (Fig.3.1.2) was used as a probe. One positive clone out of four, designated as λALH3-1, was further analyzed. The cloned 10.7kb genomic DNA fragment bears only one histone H3 gene (Fig.3.2.1.). Other histone genes (H2a, H2b and H4) have not been detected by DNA hybridization (data not shown). A 1.1kb fragment carrying this H3 gene, designated as ALH3-1.1, was then

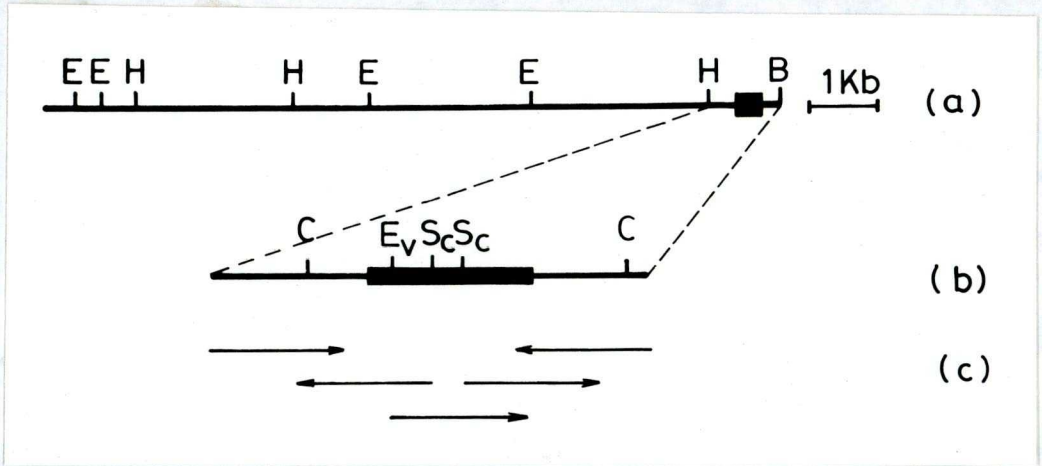


Fig.3.2.1. The restriction map of λALH3-1(a), ALH3-1.1(b) and the sequencing strategy(c) for the alfalfa histone H3 gene. The black box indicates the coding sequence of ALH3-1.1. B=BamHI, C=ClaI, E=EcoRI, Ev=EcoRV, H=HindIII and Sc=SacI.

subcloned and sequenced. The complete nucleotide sequence of ALH3-1.1 is presented in Figure 3.2.2. In comparison with the histone genes from other species, several characteristic features were observed in both the coding region and the flanking sequences of the alfalfa H3 gene.

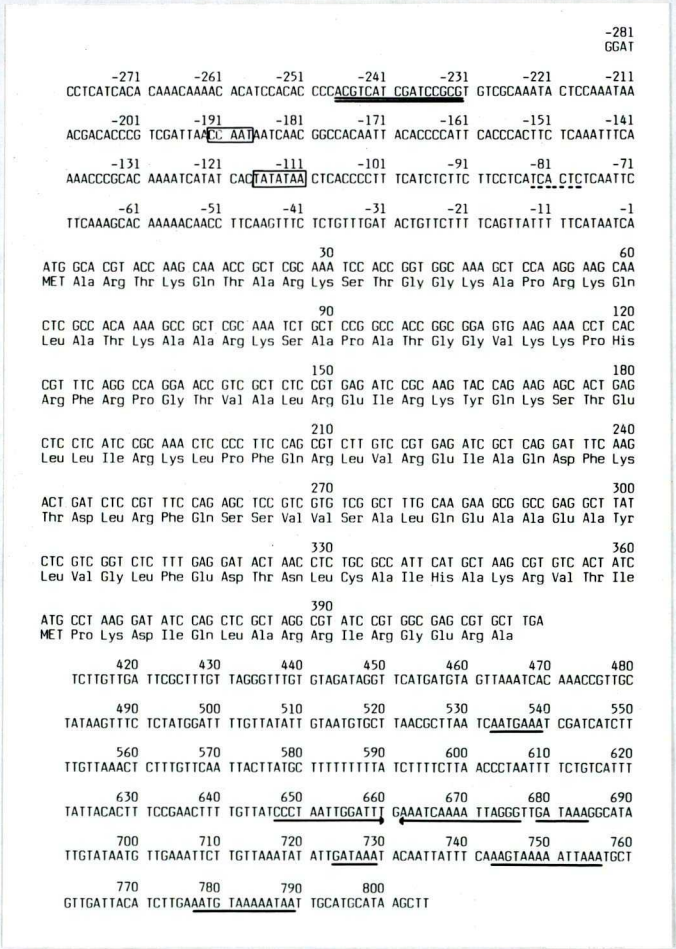


Fig.3.2.2. The nucleotide sequence of an alfalfa histone H3 gene, ALH3-1.1. The coding sequence starting from the initiation codon is arranged in triplets, and deduced amino acids are given below the corresponding codons. In the 5'-flanking region, the TATA box and CAAT box are framed and the "cap site"-like sequence is dotted. The specific block of the ACGTCA and the reverse form of the "CGC box" is double underlined. In the 3'-flanking region, the palindromic sequence is indicated by inverted arrows and the polyadenylation signal-like sequences are underlined.

Table 3.2.1. Percentage homology in the coding region of ALH3-1.1 and deduced H3 protein with those determined in other species.

Species	Clone	Protein	DNA	Reference
Alfalfa	pH3c-11	97.0	78.8	This thesis
Corn	H3C2	99.3	80.8	Chaubet,1986
Rice	pRH3-2	96.3	81.5	Peng,1986
Wheat	TH012	100	81.5	Tabata,1984a
Arabidopsis	H3 A713	99.3	79.8	Chaboute,1987
Pea	-	100	-	Patthy,1973
Human	HuH3-149	94.8	76.4	Wells,1987
Sea urchin	SpL22	95.6	82	Kaumeyer,1986

The amino acid sequence of alfalfa histone protein deduced from the nucleotide sequence of ALH3-1.1 shows very high homology (over 95%) with those determined in other species (Table 3.2.1). This protein is identical to those from wheat and pea, and only one amino acid replacement at position 90 as compared to those from corn and Arabidopsis (Fig.4.5). However, despite the high homology at the protein level, the coding sequence of ALH3-1.1 has only about 80% homology with all the genes compared (Table 3.2.1). Based on these analysis described above, it can be concluded that the isolated alfalfa gene is a histone H3 gene, though its activity has yet to be tested.

The 284 bp of the 5' flanking region possesses most of the typical consensus sequences found in both animal and plant histone genes (Fig.3.2.2). The TATA and CAAT boxes are located at position -111 and -188, respectively. A "cap site"-like motif TCACTC which is probably the transcription initiation point (Hentschel and Birnstiel, 1981) is located about 30 bp downstream from TATA box. Interestingly, a hexamer, ACGTCA,

postulated from wheat and other histone genes (Mikami,et al,1987) was found adjacent to another motif, GATCCGCG, around position -240 of ALH3-1.1 (Fig.3.2.2.). The hexamer is the target of a nuclear binding protein (Mikami et al, 1987) in wheat and the GATCCGCG is the reverse form of a putative plant histone gene-specific motif, CGCGGATC, called "CGC" box (Gigot, 1988).

The 3'-flanking region of the alfalfa histone H3 gene is characterized by an A/T-rich nucleotide composition (72% AT). A long dyad symmetric sequence which might form a stable hairpin structure (Tinoco et al, 1973) is observed 235 bp downstream from the stop codon. Interestingly, several sequences similar to the classical polyadenylation signals, AATAAA, were also observed in the 3' flanking region.

3.2.2. Repetition and Genomic Organization of Histone H3 Genes in Alfalfa

About 160 reiterated copies of histone H3 gene were estimated per tetraploid genome according to the measurement shown in Figure 3.2.3A, assuming that the size of the haploid genome of Medicago sativa is 1.7×10^9bp (Murray et al, 1984). To study the genomic organization of the 160 copies of H3 genes, we analyzed the alfalfa genomic DNA by Southern hybridization. As shown in Figure 3.2.3B, the 160 copies of histone H3 genes are located on more than 20 bands in the double-digested alfalfa genomic DNA. This result suggests that a regular repeat unit of histone genes might be excluded in alfalfa. However, it is possible that some H3 genes are closely linked in certain

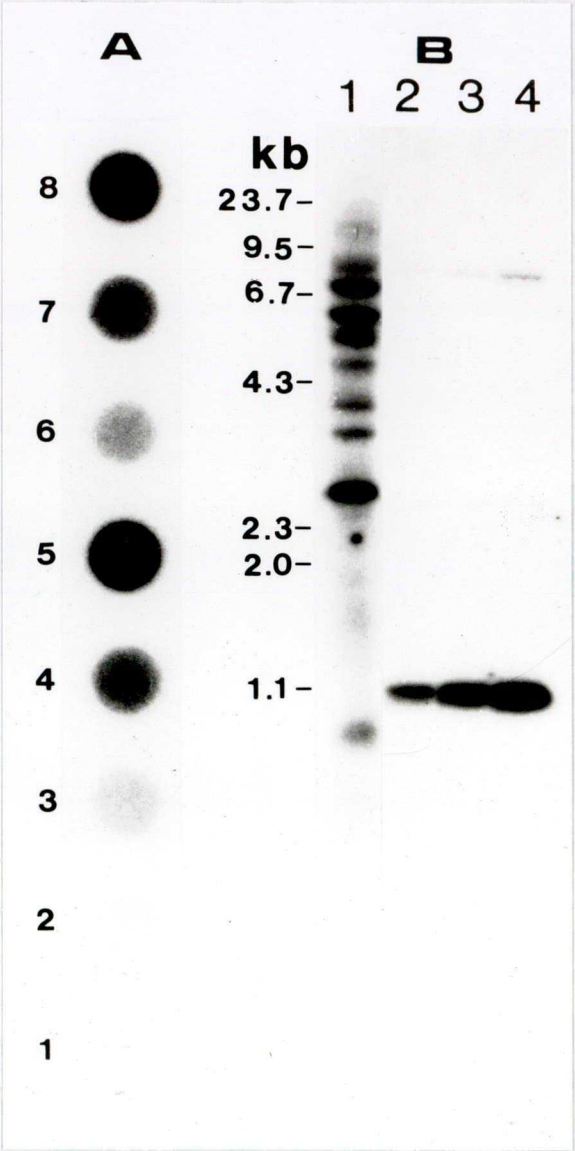


Fig.3.2.3. The copy number estimation(A) and genomic organization(B) of histone H3 genes in alfalfa. (A) 83.25, 166.5, 333.0, 832.5 and 1665 pg of ALH3-1.1 DNA subcloned in M13mp18 vector and 1, 2 and 5 ug of alfalfa Nagyszenási DNA were dot-blotted(1-8), respectively. After hybridization with the ^{32}P-labeled 195bp EcoRV/SacI fragment of ALH3-1.1(Fig.3.2.2), each dot was cut out and the radioactivity was measured. The data were used for calculating the copy number. (B) BamHI and HindIII-digested DNAs were separated on a 1.0% agarose gel, Southern blotted and hybridized with the ^{32}P-labeled 1.1kb BamHI/HindIII fragment of ALH3-1.1. Lane 1, 10ug of Nagyszenasi DNA and lane 2-4, ALH3-1.1 DNA equal to 5, 10 and 20 copies of histone H3 genes corresponding to 10 ug of alfalfa genomic DNA, assuming the genomic size of alfalfa to be 1.7×10^9 bp (Murray et al, 1984).

fragments as is the case for rice (Fig.4.2), since some fragments contain obviously more than one copy of H3 genes (Fig.3.2.3B).

3.2.3. Expression of H3 Genes in Alfalfa

To study the expression of histone H3 genes, we have analyzed the total amount of mRNA homologous to the coding region of ALH3-1.1 during the development of RA3 plants through somatic embryogenesis. Total RNAs were isolated from different parts of *in vitro* grown alfalfa plants, calli and somatic embryos (illustrated in Fig.3.2.4). The RNA samples were analyzed by Northern blot hybridization. As shown in Figure 3.2.5, the quantitative variation of H3 transcripts between tissues is significant. Although H3 transcripts are almost undetectable in leaves and roots, H3 mRNA is accumulated at a relatively high level in stems, similar to that in calli cultured *in vitro*.

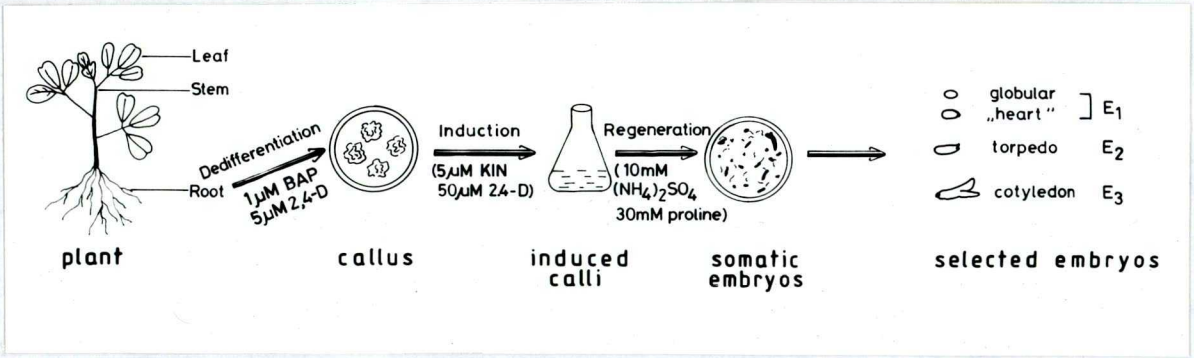


Fig.3.2.4. The procedure for somatic embryo induction from alflafa roots. The SH medium (Schenck and Hoffman,1972) was used throughout the culture with the suplementation of the indicated hormones or chemicals at different stages of tissue culture. Somatic embryos of different stages were selected according to Stuart et al (1985).

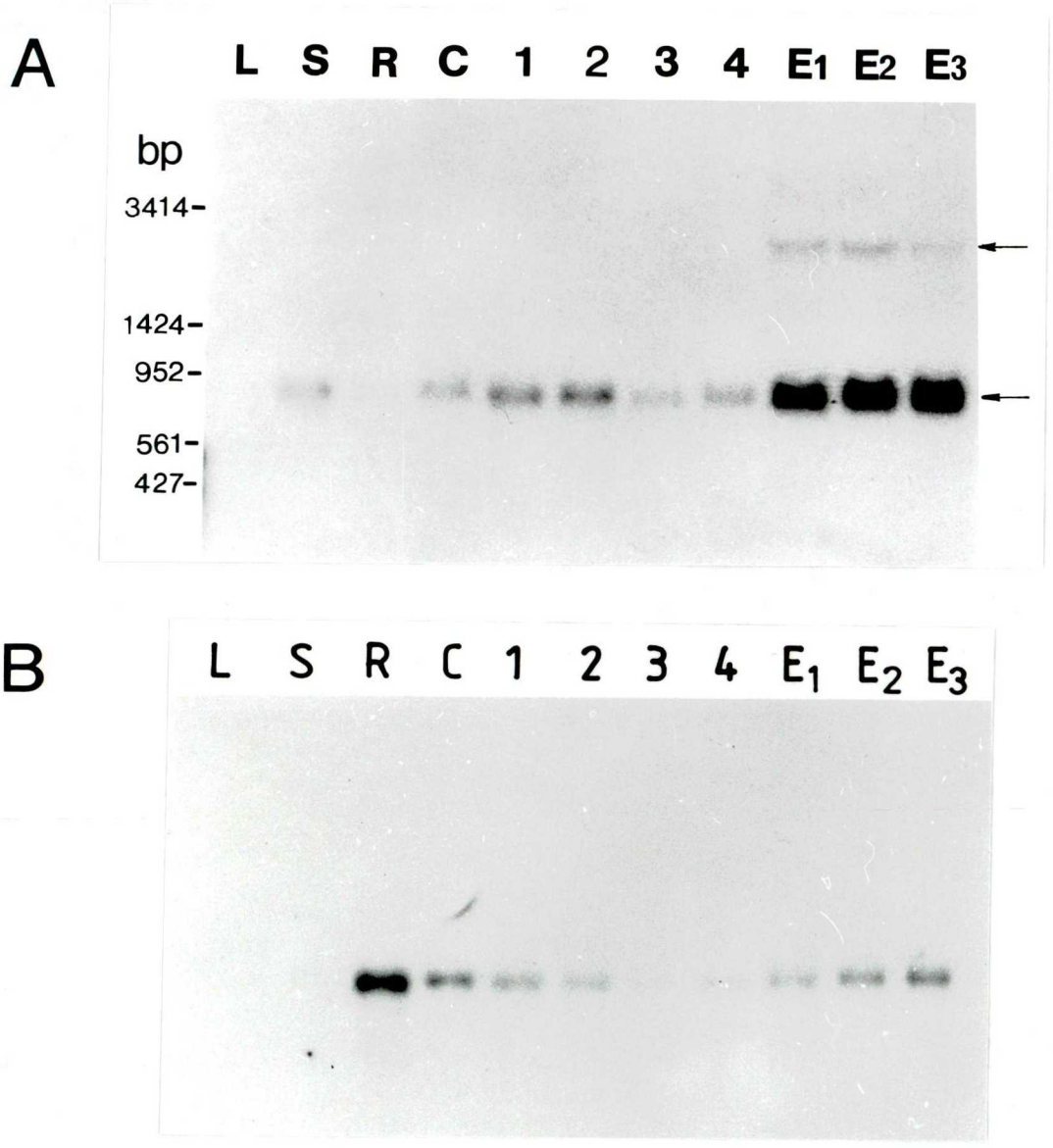


Fig.3.2.5. The accumulation of histone H3 mRNAs in alfalfa tissues. (A) 20ug of total RNA isolated from various sources, was separated on a 1.5% formaldehyde agarose gel and hybridized with the ^{32}P-labeled 1.1kb BamHI/HindIII fragment of ALH3-1.1. (B) a control experiment by using the ^{32}P-labeled 0.7kb PstI fragment of the cDNA clone, pTU4, which encodes a soybean cell wall protein (Hong et al, 1987), as a probe. L=leaf, S=stem, R=root, C=calli and E1-E3=somatic embryos at different stages shown in Figure 3.2.4. Arrows indicate the possible species of H3 transcript detected. Molecular markers were the ClaI and BstNI-digested M13mp18 RF DNA (Yanisch-Perron et al,1985).

However, the level of H3 mRNAs in the plant tissues and calli is considerably lower than those in somatic embryos differentiated from callus cells. Since the development of somatic embryos in RA3 culture system is unsynchronized, the somatic embryos were artificially divided into three stages according to Stuart et al (1985, Fig.3.2.4). As shown in Figure 3.2.5A, the histone H3 transcripts remain at high levels in all somatic embryos at the three stages of embryonic development and little quantitative change can be seen. For comparison, the cDNA pTU04 encoding a soybean cell wall protein (Hong et al, 1987) was used as a probe to hybridize the same RNA preparations and the result is shown in Figure 3.2.5B. In contrast to the H3 mRNAs, the pTU04 mRNA displayed highest level in root. The concentration of pTU04 mRNA reduced during 2,4-D treatment (see below) and increased with the development of the somatic embryos. In addition to the 800 bp main band, a high molecular weight RNA species (about 1900 bp) was also detected by H3 gene probe in the embryos (Fig. 3.2.5A). The level of this unusual RNA species is higher at heart and torpedo stage (E1 and E2 in Fig.3.2.5A) and lower in late embryos (E3). It was speculated that this RNA species might be an embryo-specific H3 transcript.

3.2.4. Accumulation of H3 Transcripts during Induction of Cell Division

As shown in Figure 3.2.5A, the histone H3 genes exhibited an increase of expression during the first two days of 2,4-D treatment. We postulated that this may be related to the increase of cell division frequency. According to our previous

microscopic observations (not shown) during the 2,4-D treatment, the division of callus cells was stimulated and accelerated in the first 48 hours and then decreased afterwards. Meanwhile, the dividing cells become smaller and cytoplasmically condensed. Subsequently many cells follow embryogenic developmental pathway in a basic hormone-free medium (illustrated in Fig.3.2.4). The activation of cell cycle induced by the auxin treatment and/or probably the fresh liquid medium (Sung et al,1984) was analyzed by ^{3}H-thymidine incorporation experiments during the 2,4-D treatment (Fig.3.2.6). As we can see there is a thymidine incorporation peak at the second day of culturing. Parallel with

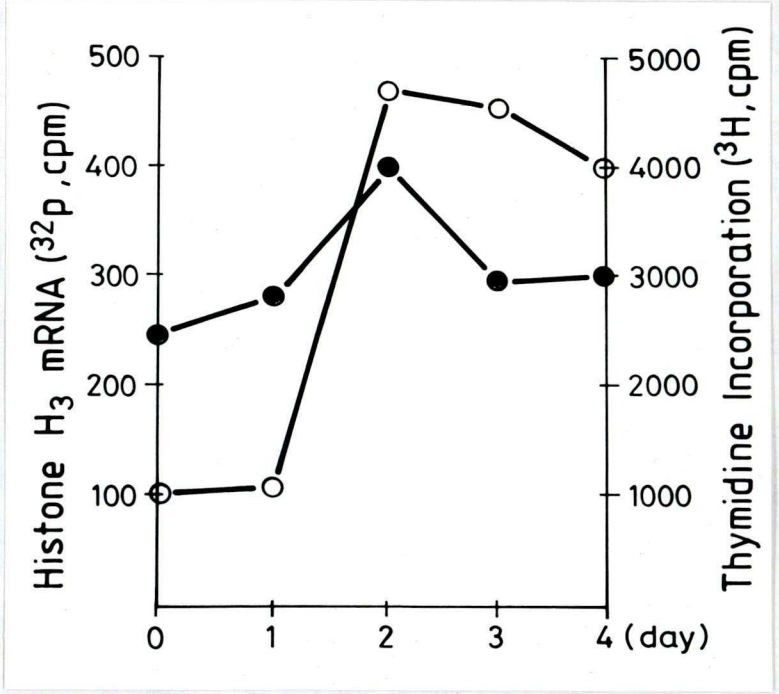


Fig.3.2.6. The timing of H3 mRNA accumulation (-o-) and the incorporation of ^{3}H-thymidine (-.-) in callus cells during the 2,4-D treatment. 10ug of total RNA isolated from SHI-treated calli was dot-blotted and hybridized as described in Figure 3.2.5A. After exposing to X-ray film (not shown), the radioactivity(cpm) on each dots was measured. ^{3}H-thymidine incorporation was performed in parallel as described in Chapter 2.

this change, the expression of H3 histone genes in the callus cells reached the highest level after two days treatment (Fig.3.2.6). These results suggest a certain relation between the timing of H3 gene expression and the activation of DNA synthesis in callus cells. These preliminary data prompted further studies on the use of the H3 gene expression as cell cycle marker.

3.2.5. Alfalfa H3 Histone Transcripts Are Polyadenylated

In addition to the several polyadenylation signal-like sequences located in the 3' flanking region of the alfalfa histone H3 gene, ALH3-1.1, it has been noticed that the alfalfa H3 messenger RNAs possessed unusually high molecular weight at about 800 bp long. Since the coding region of H3 genes is only 405 bp and the 5' UTR normally does not exceed 100bp (Fig.3.2.2), it was predicted that the alfalfa mRNAs carried long 3' UTR with the possible contribution of polyadenylic tail. To test this hypothesis, polyA(+) and polyA(-) mRNAs were isolated by oligo-dT cellulose chromatography (Aviv et al, 1972) from alfalfa plants, calli and somatic embryos. The amount of H3 transcripts in polyA(+) and polyA(-) RNA fractions was analyzed by RNA blot hybridizaton (Fig.3.2.7). In comparison with the non-separated total RNA fractions, the signal detected in the polyA(+) RNA fractions was stronger, whereas no detectable amount of H3 mRNA was observed in the polyA(-) RNA fractions. These results indicated that histone H3 mRNAs are polyadenylated in the studied alfalfa tissues. The results shown in Figure 3.2.7 are also consistent with the findings shown in Figure

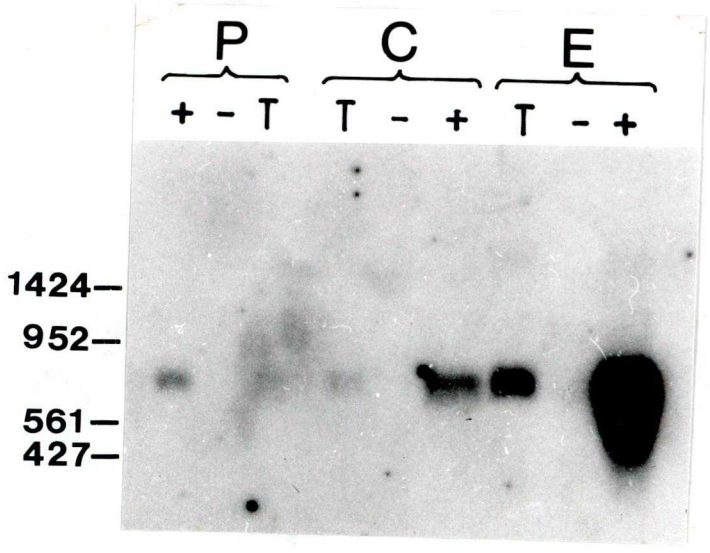


Fig.3.2.7. Northern blot hybridization analysis of total RNA (T), polyA(-) RNA (-) and polyA(+) RNA (+). RNAs were prepared from 16 days old alfalfa plants raised from nobal cutting in vitro (P), three weeks old calli (C) and somatic embryos of mixed stages (E). 25ug of total and polyA(-) RNA and 5ug of polyA(+) RNA were immobilized on nitrocellulose filter and hybridized to the ^{32}P-labeled 320bp EcoRV-StyI fragment of ALH3-1.1. The polyA(+) RNA samples were loaded about 4-fold in excess, assuming an appropriate 5% of polyA(+) RNA in the total RNA. Molecular markers were the same as shown in Fig.3.2.5.

3.2.5 indicating that the histone H3 highly expressed in somatic embryos. To show unequivocally that alfalfa H3 transcripts carry a polyA tail H3 cDNA clones were analyzed. About 40,000 cDNA clones from an alfalfa cDNA library were screened by using the StyI/EcoRV fragment of ALH3-1.1 as a probe. Out of 25 positive clones seven were characterized by restriction enzymes. As shown

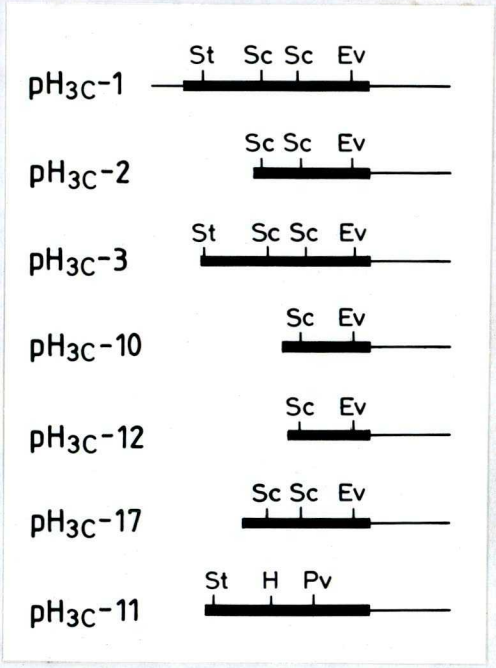


Fig.3.2.8. The restriction maps of cDNA clones for alfalfa H3 histones. Ev= EcoRV, H=HindIII, Pv=PvuII, Sc=SacI, and St=StyI.

in Figure 3.2.8, the seven H3 cDNAs fall into two categories: six of them possess homologous restriction sites with the genomic clone, ALH3-1.1 (Fig.3.2.1), while the other exhibits completely different restriction patterns. The DNA sequencing data of four cDNA clones (Fig.3.2.9) revealed more information about the difference between these two classes of alfalfa H3 cDNAs. Among the isolated cDNAs, only clone pH3c-1 carried a full length H3 transcript, while the others were incomplete copies of H3 mRNAs. Clone pH3c-1, pH3c-3 and pH3c-12 are highly conserved. They vary with ALH3-1.1 only in a few nucleotides in both translated and untranslated region. In contrast, clone pH3c-11 has only 78.8% homology with ALH3-1.1 in the translated region and encodes a longer non-homologous 3' UTR (Fig.3.2.9). The low homology between pH3c-11 and the other alfalfa H3

sequences results in four amino acid replacements (Fig.4.5). These include a change of a hydrophobic Ala to a polar Thr at position 31, a hydrophobic Phe to a polar Tyr at position 41, a polar Ser to a basic His at position 87 and a polar Ser to a hydrophobic Leu at position 90, respectively. As predicted, the 3'UTR of the H3 cDNAs are very long (217 for pH3c-11 and 146 for

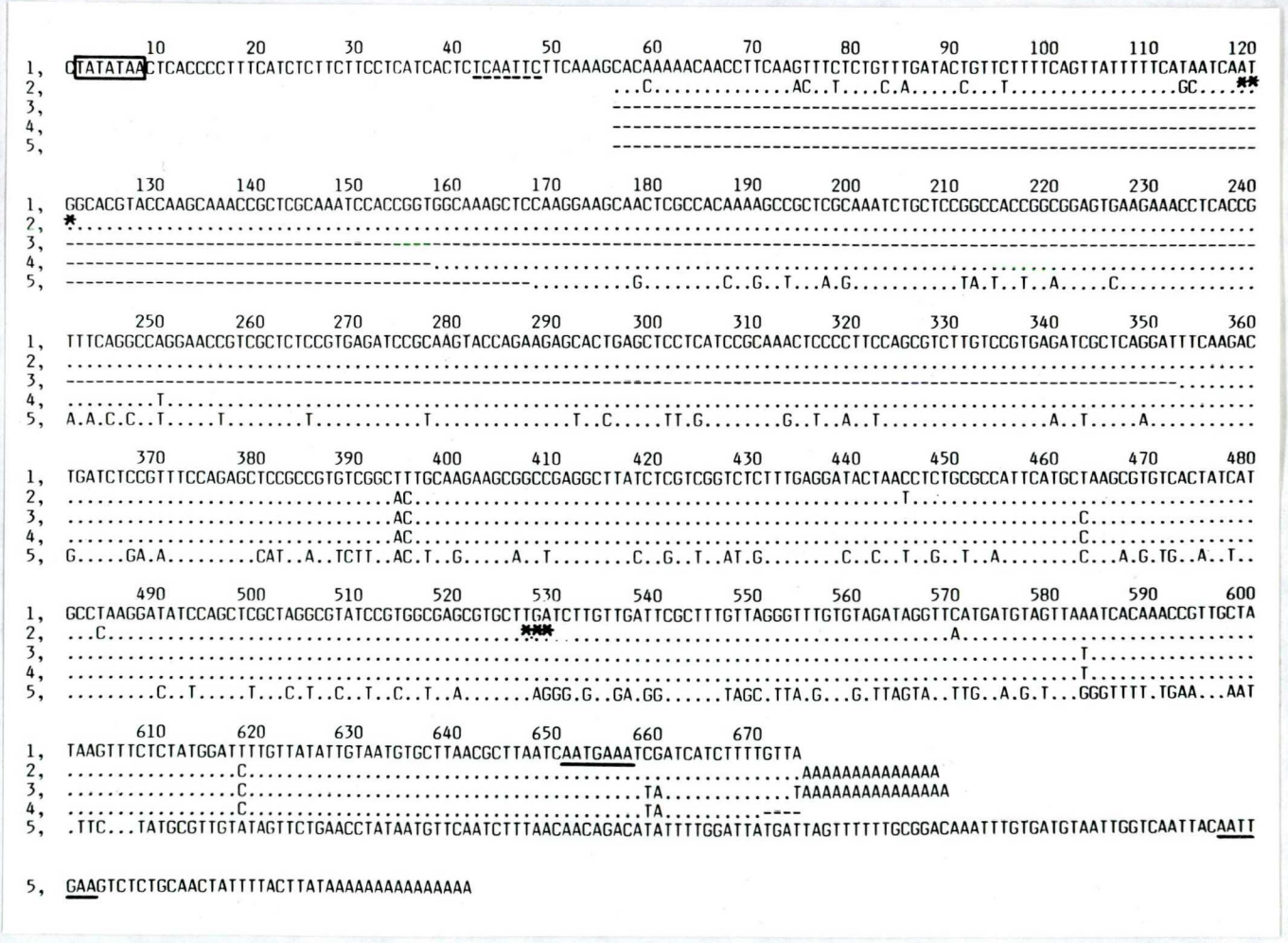


Fig.3.2.9. Comparison of the nucleotide sequence of alfalfa histone H3 cDNAs with that of genomic clone, ALH3-1.1. The TATA box is framed, the "cap site" is indicated by dashed line, the initiation and stop codons are marked with asterisks and the putative polyadenylation signals are underlined. Dots represent base pairs identical to the genomic ones and dashes represent deletions. The polyA tracts in all the cDNAs were represented by 15 A sequences. 1, ALH3-1.1; 2, pH3c-1; 3, pH3c-12; 4, pH3c-3 and 5, pH3c-11.

the others) as compared to those of animal histone mRNAs (Birnstiel, 1985). The detection of polyA tract in three of the sequenced cDNAs is in agreement with the hybridization data shown in Figure 3.2.7. It is conceivable, however, that the H3 cDNA in plasmid pH3c-3 originally possessed a polyA tail which could be lost during cloning (Fig.3.2.9). Based on these results, It can be concluded that independent histone H3 transcripts are polyadenylated in alfalfa at different developmental stages.

About 20bp upstream from the 3' ends of the alfalfa H3 transcripts a highly conserved sequence motif, AAUGAAA, had been identified (Fig.3.2.9 and Table 3.2.2). This motif varies by one

Table 3.2.2. Putative polyadenylation signal sequences of plant histone H3 and H4 genes*

Species	Clone	bp**	Signal	Reference
Alfalfa	ALH3-1.1	122	AAUG-AAA	This thesis
	pH3c-1	122	AAUG-AAA	
	pH3c-3	122	AAUG-AAA	
	pH3c-12	122	AAUG-AAA	
	pH3c-11	188	AAUUG-AA	
Arabidopsis	H4748	149	G-UUGAAA	Chaboute, 1988
	H4777	196	GAUG-AAA	
Corn	H3C2	178	AAUGGAAA	Chaubet,1988
	H3C3	134	AAUGGAAA	
	H3C4	215	AAUGGAAA	
	H4C7	194	GAUG-AAA	
	H4C13	194	GAUG-AAA	
Barley	-	237	GAUG--AA	Chojecki, 1986
pea	PsH1b	65	AAU---AA	Gantt, 1986
	rbcS	101	AAUGGAAA	Coruzzi, 1984
		124	AAUG--AA	
Petunia	Cab91R	83	-AUG-AAA	Dean, 1986
Consensus sequence			A/GAUG(G)AAA	

* Putative polyA signals for a histone H1 cDNA and a rbcS gene from pea and a Cab gene from Petunia are included in this Table. ** Base pairs from the stop codon.

G insertion as compared to the conventional polyadenylation signal, AAUAAA (reviews in Manley, 1988).

3.2.6. Codon Usage of Alfalfa Histone H3 Sequences

Based on the available data on biased codon usage in plant histone genes (Gigot, 1988), the codon usage in the dominant H3 sequences represented by ALH3-1.1 and the minor H3 cDNA, pH3c-11, have been analyzed. By comparing the codons for the abundant amino acids(Ala, Arg, Leu, Lys and thr), three codons were revealed to be preferentially used in ALH3-1.1 (Table 4.5): GCT among four possible codons, encoding Ala, is used 60% of the time; CGT among six possible codons, encoding Arg, is used 60% of the time and CTC among six possible codons, encoding Leu, is used 83% of the time. In pH3c-11, the CGT and CTC are no longer overused, instead, the AAG for Lys and the CTT for Leu appear to be the preferred codons. Besides the codons for abundant amino acids, the difference in codon usage between the dominant and the minor H3 variants is also shown in the codons for the other non-abundant amino acids. A comparison of the nucleotides present at the codon position III revealed that in pH3c-11 the proportion of C is reduced by two times, while the proportion of T is increased to 38.8%, as compared to that in the dominant H3 sequences (Table 4.4).

Subsequently, we compared the nucleotides present at the third position of codons in various histone genes from different organisms. In contrast to the extreme overuse of the G/C-ended codons in all of the monocotyledonous histone genes compared, the histone genes and cDNAs from three dicotyledonous plants do not exhibit such property in codon usage (Table 4.4).

Chapter 4.

DISCUSSION

In this thesis, the isolation and characterization of four genomic clones and seven cDNAs for H3 histones from a cereal, Oryza sativa, and a legume, Medicago sativa have been described and the differential accumulation and polyadenylation of H3 transcripts during the development of alfalfa plants through somatic embryogenesis have also been demonstrated. Although only one type of histone genes was included, several conclusions drawn from this work have widen the horizones of knowledge about plant histone genes. These are discussed as the followings:

4.1. The Organization of Rice Histone Genes

A total of five H3 sequences have been cloned from rice. A previous Southern blot hybridization experiment revealed that five bands in the EcoRI/BamHI digested rice genomic DNA were hybridized to the coding sequence of pRH3-2 (Fig.4.1, reproduced from Xie et al, 1987). Obviously, the 1.3 kb BamHI/BamHI fragment of λRH3-17, the 0.7 kb EcoRI/BamHI fragment of clone λRH3-1 and λRH3-17 and the 0.5 kb EcoRI/BamHI fragment of clone λRH3-1 correspond to band C, D, and E respectively, while the H3 sequence in clone λRH3-2 must correspond to a fragment which has higher molecular weight than band A (Fig.4.1). Since the copy number of H3 genes in the rice diploid genome was estimated to be about 50 (Xie et al, 1987), these five bands may represent different H3 gene families which consists of more than two H3 gene members, while the "orphan" genes such as H3R-21 may be

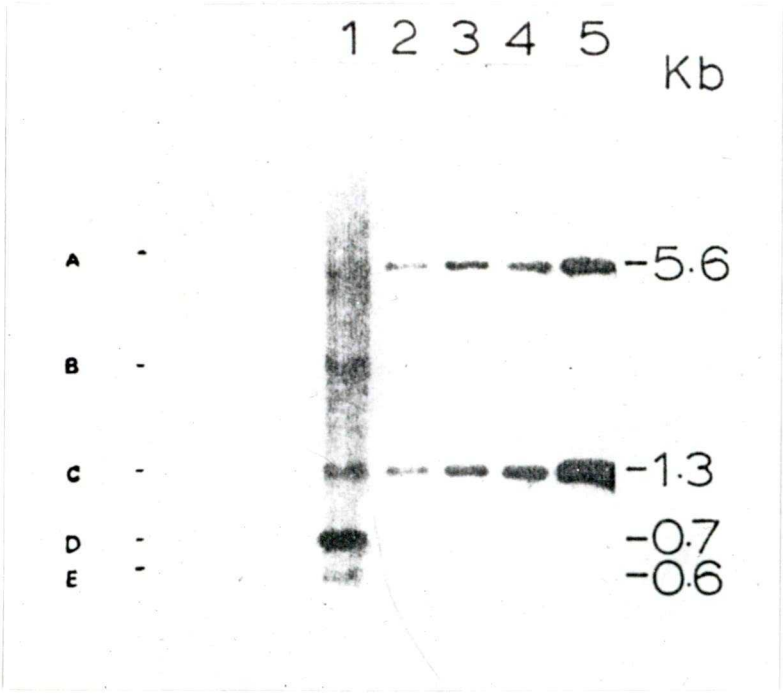


Fig.4.1. Genomic organization of rice histone H3 genes. The data was from Xie et al (1987). BamHI/EcoRI-digested total rice DNA and pRH3-2 plasmid DNA were run on a 0.7% agarose gel in parallel and blotted onto nitrocellulose filter. Hybridization was carried out with the ^{32}P-labeled 220bp PvuII/PstI fragment of pRH3-2 at 42^{o}C for 36 hours. Lane 1, 10ug of rice DNA and lane 2-5, pRH3-2 DNA quantitatively equal to 5, 10, 20 and 40 copies of H3 gene per diploid rice genome, respectively, assuming the genomic size of rice to be 6x10^{8} bp (Iyengar et al, 1979).

involved in those fragments not clearly visualized in this experiment. In addition to the simple picture shown by this Southern hybridization experiment (Fig.4.1), the close linkage of H3 gene and H3 pseudogene in clone λRH3-1 and λRH3-17 and the H2a, H2b and H4 genes in clone pIR22 (Fig.4.2, Thomas and Padayatty, 1983b) also suggest that the histone genes are less dispersed in rice than other higher plants. To see whether clone

λRH3-1 and λRH3-2 carry other histone genes, the coding regions of a wheat histone H4 gene (pTH011, Tabata et al,1983), a chick H2a (pCH2eSX30) and a chick H2b gene (pCH1aKR-1.3, Sugarman et al,1983) were used as probes to hybridize with λRH3-1 and λRH3-2 DNAs, respectively. Unfortunately, these experiments did not show any positive results(not shown). However, the existence of histone H1, H2a and H2b genes in clone λRH3-1 and λRH3-2 may not

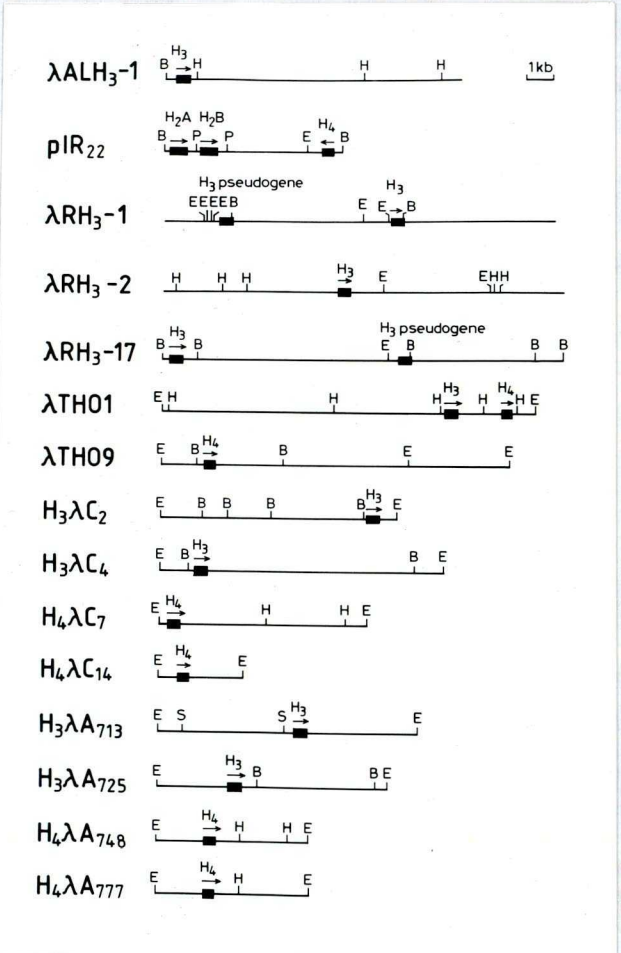


Fig.4.2. The organization of cloned plant histone genes. The black bars represent coding sequences and arrows indicate the transcription orientation of each genes. Abbreviations for restriction enzymes are the same as Figure 3.1.1 and references are as Table 1.1.

be ruled out because of the negative results, since these genes are much less conservative than H3 and H4 genes. In other higher plants, the hybridization patterns of genomic DNA with histone gene probes are not so simple after all. The repetitive copies of histone H3 and H4 genes in alfalfa(Fig.3.2.3), corn and wheat (review in Gigot, 1988) were shown to spread on more than 20 fragments of single or double restricted genomic DNA. In addition all the histone genes cloned so far from these plants were found to be "orphans" with the exception that a wheat genomic clone carried two closely linked H3 and H4 genes (Fig. 4.2). These results suggested that the majority of the H3 and H4 genes exist in a rather dispersed manner in these plant species.

4.2. H3R-12 Represents A Rice Histone H3 Pseudogene

In comparison with the other rice H3 genes, H3R-12 obviously represents a H3 histone pseudogene for several reasons. In addition to a total 10 bp deletion in the "coding region", H3R-12 lacks the initiation codon and the entire 5' flanking sequence which is characteristic for the rice histone H3 genes (Fig.3.1.2, Fig.3.1.5). Moreover, there are 11 base substitutions of G or C to A or T in the first 150 bp of the "coding region" and some of these base substututions resulted in amino acid replacements (Fig.3.1.2). These base substitutions are consistent with the assumption that most base changes in pseudogenes no longer under selective pressure are due to losses of C's and G's, with the preferred changes from C to T and G to A (Gojobori et al, 1982). Interestingly, this H3 pseudogene is identical to a previously published rice H3 pseudogene, pRH3-1

(Xie et al, 1987), except that pRH3-1 constitutes more deletions and two dinucleotide substitutions at position 660-661 and 695-696. The dinucleotide substitutions in pRH3-1 were supposed to be due to sequencing errors (marked with solid triangles in Fig.3.1.2). The high homology between these two pseudogenes may suggest that they arose in the same period, probably due to a duplication. Since the nucleotide divergence between the rice H3 pseudogenes and their "parent" (i.e. the associated intact H3 gene) is about 7%, the rice H3 pseudogenes might have arisen about 5 million years ago, if the rate of accumulation of neutral point mutations in pseudogenes is about 1.3% per million years during evolution (Miyata and Yasunaya, 1981; Wells et al, 1987).

Pseudogenes have been frequently found among eukaryotic genes, especially among the best-characterized "house-keeping genes (review in Vanin, 1983 and 1985). These pseudogenes generally fall into two categories: the replicative pseudogene, which retains introns and is in tandem with a functional parent (Vanin, 1983), and the processed pseudogene, which is free of intron and dispersed throughout the genome and therefore it was supposed to generate from processed mRNA intermediate (Vanin, 1985; Weiner, 1986). Since both the two rice pseudogenes are associated with an intact H3 gene, they may be classified into the replicative pseudogene category. The replicative pseudogenes might resulted from unequal crossover, gene conversion, insertion of mobile elements, etc (Maxson et al, 1983; Liu et al, 1987).

Considering the RS region in clone H3R-12 might belong to

part of some specific eukaryotic sequences, such as highly repetitive sequences, rDNAs and tDNAs (Wu and Wu, 1988) and the terminal inverted repeats of transposable elements (Ikenaya and Saigo, 1982; Liebermann et al, 1983), we searched for the possible sequences in GenBank homologous to the rice RS and its flanking region. Only very limited and short sequences presented in the GenBank were found homologous to the AT-rich sequences flanking the rice RS and little to the RS itself (searching results not shown). Most of the detected homologous sequences are within the AT-rich region of random sequence samples and they seem not to make much sense. Nevertheless, the distinct organization of the rice RS region is similar to the outer domain (OD) of the long terminal repeats (LTR) of several transposable elements, such as the TU transposons of sea urchin (Liebermann et al, 1983; 1986) and the foldback transposable elements of Drosophila (FB4, Potter et al, 1982). The ODs of these transposable elements mainly consist of short direct tandem repeats of 10-31 bp which may construct larger repeating units as the basic repeating units of the rice RS do (Fig.3.2.7A). To see whether the RS is one part of the LTR of a possible rice transposable element, further studies are required.

4.3. The Consensus Sequence Motifs In The 5' Flanking Region of Plant Histone Genes

4.3.1. The GATCC pentamer and the cap site

In animal, two specific motifs were found in the 5' flanking region of most histone genes: a "cap site", pyCATTCpu where the

transcription starts, located about 20 bp downstream from the TATA box, and a pentamer, GATCC, located about 10 bp upstream from the TATA box (Hentschel and Birnstiel,1981). These two consensus motifs seem to exist as well, with slight modification, in plant histone genes. As shown in Table 4.1, the pentamer motif for plant histone H3 or H4 genes is GCTCC and the "cap site" is pyCAA/CpyC. The unusual location of the two GATCC motifs which was noticed previously (Wu et al, 1988) in the alfalfa H3 gene, ALH3-1.1, may reflect a casual event and has not any functional implification. The real pentamer, if there is any, for this gene may be probably the GCACA 15 bp upstream from the TATA box (Table 4.1).

In addition to the modification of the pentamer and the "cap site", the TATA and CAAT boxes of plant histone genes also exhibit some differences from those of animal gene's. It is interesting that, whereas flanked mostly by G's in animal histone genes (Hentschel and Birnstiel, 1981), most of the TATA boxes of plant histone genes are flanked by C's (Table 4.1). As in other plant genes (Anderson et al, 1984; Heidecker, 1986), the CAAT boxes postulated from plant histone genes are very variable, although the CnAnTn boxes appear to be dominant in monocots (Table 4.1).

4.3.2. The plant specific motifs

Recent studies on wheat, maize and Arabidopsis histone H3 and H4 genes have revealed a highly conserved octamer, CGCGGATC, existing in the 5' flanking region of all these genes. These octamer, so called "CGC box", is located about 100 to 150bp upstream from the TATA box. In some cases, this CGC box exists

Table 4.1 Consensus motifs in the 5'-flanking region of plant histone genes*

Gene	CGC Box	CAAT Box	Pentamer	TATA Box	CAP site
ALH3-1.1	-122 GATCCGCG	-74 CCAAT	-17 GCACA	CTATATAAAC	+33 TCACTC
H3A713	-185 CGCGGATC	-70 CAAT...CACA	-7 GCTTC	CTATAAAAC	+30 TCAACA
	-91 CGCGGATC	-32 CAnT			
H3A725	-123 CGCGGATC	-81 CAAAT..GAAC	-	ATAAATATC	+40 TCAACT
H4A748	-140 CGCGGATC	-92 ACAT...CATT	-17 GACTC	CTATAAAAC	+24 TCACCC
		-26 CAnT			
H4A777	-141 CGCGGATC	-93 ACAT...CATT	-17 GACTC	CTATAAAAC	+24 TCACCC
		-27 CAnT			
H3C2	-116 CGCGGATC	-96 CCACC.TCATC	-15 GCTCC	TTACAAAAC	+28 CCGAAC
		-68 GCAATGCCACT	-8 GCTCC		
H3C3	-130 GATCCGCG	-101CCATTCCCACG	-6 GCTCC	CTATTAATC	+31 CCAACG
		-49 CAnT			
		-23 CAnT			
H3C4	-125 GATCCGTG	-93 CCACC.TCATC	-13 GCTCC	CTACAAATAC	+32 TCACTG
	-113 CGCGGATC	-75 CCATCTCCATC			
H4C7	-100 GATCCGCG	-92 GCATC..TACC	-25 GCTCC	TTAAATAAC	+30 CCACTC
	-96 CGCGGCATC	-53 CCATC..GAAC			
H4C13	-126 GATCCGTG	-78 TCATTCTCACC	-25 GCTCC	TTAAATAAC	+28 CCACTC
	-101 GATCCGCG	-53 CCATC..GAAC			
	-97 CGCGGCATC				
H4C14	-45 CGCGGATC	-80 CCATC..CAAC	-24 GCCTC	CTATATTAC	+35 CCAATC
		-68 TCATC.CCACC	-12 GCCTC		
		-32 CCATC..CAAC			
pRH3-2	-	-109CCAAT	-30 GCTCT	CTATATAAAC	+31 TCAACC
		-57 CCAnT			
		-45 CCAnTn			
H3R-11	-247 CGCGGAAC	-93 CnAnTT	-23 GAATC	CTATTTAAAC	+27 TCAATC
	-187 CGCGGATC	-78 CnAnTn			
	-177 GATCCGTG	-66 CnAnTn			
	-128 CGCGGATC	-17 CnAnTn			
H3R-21	-224 GATCCGCG	-112CCAAT	-29 GCTCT	CTATATAAAC	+30 TCATCC
	-176 CACGGATC	-68 CCAnTT			
	-166 GATCCGCG				
TH011	-94 GATCCGCG	-54 CCACT.CCATC	-23 GCCTC	CTTTAAGAC	+30 TCACAG
	-90 CGCGGCATC		-6 GACCC		
TH012	-131 AATCCGCG	-92 CCACT.CCACT	-19 GCTCT	CTATTTAAC	+32 TCACCC
	-127 CGCGGCAT	-81 GCATC.CCAAC	-8 GCTCC		
		-38 CAnT			
TH081	-120 GATCCGCG	-107CCAC...CATC	-24 GCCCG	CTAAAACAC	N.D.
		-104CCATC..CACG			
		-41 CAnT			
TH091	-75 CGCGGATC	-33 CAnT	-22 GTTCC	CATATAAAC	+29 TAATCC
CONS					
Plant	CGCGGATC	CCAnTn	GCpypyC	CTATAT/AAAC	pyCAA/CpyC
	or GATCCGCG	or CCATC..CAXT			
Animal		CCAAT	GATCC	GTATAAATAG	pyCATTCpu

* Modified from Gigot (1988). For individual references see Table 1.1.

Table 4.2. The putative ACGTCA-GATCCGCG block in the 5' flanking region of plant histone H3 and H4 genes

Species	clone		AC-GA block	
Alfalfa	ALH3-1.1	-130	ACG TCAtcGATCCGCG	-115
Corn	H3C3	-139	ACGgaTCAc GATCCGCG	-123
	H3C7	-109	ACG TCAgcGATCCGCG	-94
	H3C13	-109	ACG TCAgcGATCCGCG	-94
Wheat	H3TH012	-139	ACG TCAccAATCCGCG	-124
	H3TH081	-126	ACG CAg GATCCGCG	-113
	H3TH011	-102	ACG TCAccGATCCGCG	-87
Rice	H3R-21	-175	ACGgaTC gcGATCCGCG	-159
Consensus			ACGTCA(g)cGATCCGCG	

Numbering and References are the same as table 4.1.

in its reverse form, 5'GATCCGCG3' (review in Gigot, 1988). As expected, both the rice and alfalfa H3 genes possessed this specific motif. Intriguingly, however, we have observed the multiple copies of CGC box, either direct or reverse form, in the rice histone H3 genes (Fig.3.1.2) and the association of reverse CGC box with a hexamer motif, ACGTCA, in the alfalfa H3 gene (Fig.3.2.2). Further investigation revealed that the observations were not restricted to rice and alfalfa (Table 4.1 and 4.2). Multiple copies of CGC boxes were also found in an Arabidopsis H3 gene, two wheat H3 and H4 genes and two highly conserved corn H4 genes. In the case of wheat and corn genes, two of the octamers are inverted and 4bp-overlapped (Table 4.1).

The association of ACGTCA with GATCCGCG is also found in three H3 genes of corn, one H3 gene of rice and two H3 and one H4 genes of wheat (Table 4.2). It seems that the hexamer appears only in those genes which possess a reverse form of CGC box.

Here we designate the ACGTCA(g)cGATCCGCG as AC-GA block. The ACGTCA hexamer was originally reported in wheat histone H3 and H4 genes (Mikami, 1987). Gel mobility shift assays have demonstrated that the ACGTCA hexamer is the binding target of a partially purified nuclear protein(s), HBP-1 (Mikami et al, 1987). The function of this hexamer, as suggested by the authors, is probably related to cell-cycle regulation of wheat histone genes or cAMP-mediated gene regulation. The attachment of ACGTCA to GATCCGCG may imply functional interaction between this two motifs. It is also possible that the AC-GA block might exist as an intact functional cis unit.

4.4. The Polyadenylation of Histone Transcripts

It was an early generally accepted dogma that histone transcripts are non-polyadenylated. This concept has been under revision in the last few years. In animal, where the non-polyadenylated histone mRNAs are quantitatively dominant, a number of polyadenylated histone mRNAs have been discovered recently. These polyadenylated mRNAs generally encode minor histone variants and their accumulation is not cell cycle-dependent (reviews in Stein et al, 1984; also see Wells et al, 1987; Lieber et al, 1988; Alonso et al, 1988). Interestingly, there are two exceptions. In pre-matured frog oocytes (review in Maxson et al, 1983) and in avian haploid round spermatids (Challoner et al, 1989), a significant proportion of histone mRNAs are polyadenylated, despite the polyA(+) mRNAs are identical to, except the polyA tail, the non-polyadenylated histone mRNAs in proliferating somatic cells.

These two exceptional cases have resulted in the hypothesis that essential structural elements for polyadenylation coexist with those for the U7 snRNA-mediated 3' end formation in replication-dependent histone genes (Challoner et al, 1989). However, the observation of the polyadenylation in H3 mRNAs of alfalfa cannot be considered as a minor case or a transitory event, since most of the H3 mRNAs (if not all) are polyadenylated in tissues of various developmental stages (Fig.3.2.7). Consistently, the histone mRNAs are polyadenylated throughout the life cycle in yeast (Fahrner, 1980) and a siliated protozoan, Tetrahymena. Recently, the polyadenylation of histone H3 and H4 mRNAs was also demonstrated in germinating embryos and seedlings of maize (Chaubet et al, 1988), cell suspensions and seedlings of Arabidopsis and probably seedlings of sunflower or cell suspension of tobacco (Chaboute et al, 1988). However, an early study on rice, which defined the histone mRNAs in the non-polyadenylated fractions (Ahmed and Padayatty, 1982), suggests that a more comprehensive investigation is necessary for a general conclusion that the histone transcripts are polyadenylated in higher plants.

In addition to the different regulatory mechanisms for the 3' end formation (compare reviews in Mowry and Steitz, 1988 and Humphrey and Proudfoot, 1988) and RNA stability, the polyadenylation of histone mRNA may also reflect an important event in gene evolution. If the non-poly(A) histone genes are modern products as suggested by Wells et al (1986), the poly(A) histone genes in at least some species of higher plants, fungi and protozoa might have failed to catch up with the evolutionary

progress that the histone genes in higher animals have achieved, or they would have adopted a different evolutionary pathway.

4.5 The Polyadenylation Signals In Alfalfa H3 Histone Genes

In animal, the histone genes which are transcribed into non-polyadenylated histone mRNAs possess a highly conserved T-hyphenated hairpin structure and a consensus sequence just downstream from the 3' end of their transcripts (Fig.1.1). These two structural elements are essential for efficient and correct 3' cleavage of histone pre-mRNAs, an U7 snRNA-mediated event which has been demonstrated recently (Cotten et al, 1988). The histone genes which encode polyadenylated mRNAs lack the T-hyphenated inverted repeats and its "downstream element", thus the 3' cleavage of their pre-mRNA is supposed to be regulated by a different mechanism. For polyadenylated mRNAs, at least a highly conserved hexamer, AAUAAA, located 10-30 bp upstream from the 3'-end, and a stretch of less conserved U or GU repeats downstream of the 3'-end are the essential cis-regulatory elements for the 3' cleavage and polyadenylation (review in Birnstiel, 1985; Manley, 1988 and Humphrey and Proudfoot, 1988). Many polyadenylated histone mRNAs including those from animal and protista lack the typical polyadenylation signal, AAUAAA. The putative polyadenylation signal, AAUGAAA, for alfalfa H3 mRNAs also varies with the AAUAAA by one G insertion. This structural feature may not necessarily resulted in a different mechanism for the cleavage-polyadenylation reaction, since many divergent types of AAUAAA have been reported (Manley, 1988) and pre-messenger RNAs are occasionally found to be cleaved and

polyadenylated at multiple sites *in vivo* (for animal see ref. in Humphrey and Proudfoot,1988; for plant see Dean et al,1985). Similar forms of this putative polyadenylation signal also exist in histone H3 and H4 mRNAs of maize and *Arabidopsis* (Table 3.2.2), which have recently been shown to be polyadenylated (Chaubet et al, 1988; Chaboute et al, 1988). Coincidently, AAUGAA or AAUGGAAA sequences were also suggested to be the polyadenylation signal in a Cab gene of *Petunia* (Dean et al, 1985), in a rbcS gene of pea (Coruzzi et al, 1984) and a octopine synthase gene of Ti plasmid (Dhaese et al, 1983). In addition the putative polyadenylation signals in the studied alfalfa H3 sequences (Fig.3.2.9) and a maize H3 gene, H3C3 (Chaboute et al, 1988), are all preceded by a C. The resulting CAAUG pentamer overlapped with the AAUGAAA heptamer is similar to the CApyUG motif postulated from many polyadenylated mRNAs (Manley, 1988). The latter usually lies just upstream or just downstream of the AAUAAA. In two *Arabidopsis* H4 genes the CAATG motif is found 10bp downstream from the putative poly(A) addition site (Chaboute et al, 1988). Moreover, there is a stretch of T-rich sequence about 20bp downstream from the 3' cleavage site of ALH3-1.1 (Fig.3.2.2 and 4.3). Thus, it is reasonable to propose that the cleavage-polyadenylation of alfalfa histone H3 pre-mRNAs occurs by a mechanism similar to that described for animal pre-mRNAs (Humphrey and Proudfoot, 1988), although modifications in plant cells may not be omitted (Hunt et al, 1987).

In the RESULTS we pointed out the existence of a palindromic structure followed by several polyadenylation signal-like

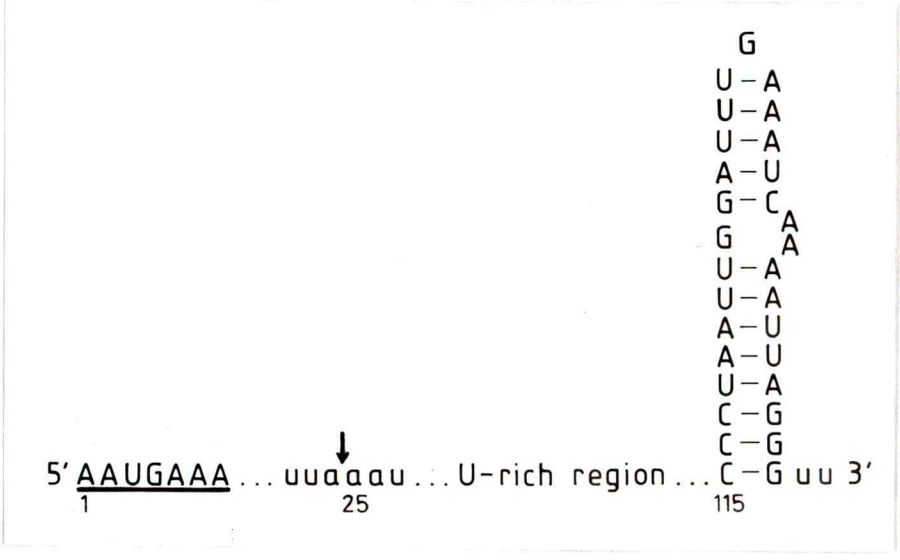


Fig.4.3. The putative hairpin structure in the 3' UTR of the major alfalfa H3 pre-transcripts derived from the alfalfa genomic clone, ALH3-1.1. The nucleotides are numbered from the underlined polyadenylation signal. The arrow points to the 3' cleavage site. The cryptic polyadenylation signal-like sequences (Fig.3.2.2) following the hairpin structure are not shown.

sequences in the alfalfa histone H3 gene, ALH3-1.1 (Fig.3.2.2). These polyadenylation signal-like sequences may be recognized in vivo, similarly to those in other plant genes (Dean et al, 1985). However, the 3' end cleavage and poly(A) addition at alternative sites may not be seen since this hairpin structure may play a critical role in preventing 3' end formation at sites downstream of it (Fig.4.3). Hunt (1988) has recently demonstrated that cryptic polyadenylation sites in a pea rbcS gene can only be seen to be functional if the principal polyadenylation sites were deleted. Interestingly, a hairpin

structure, GTCAAATCGTGGCCTCTAATGAC, was found in this rbcS gene 248bp downstream from the stop codon (Coruzzi et al, 1984). The position of this hairpin structure is just between the principal and cryptic polyadenylation sites defined by Hunt (1988). It is likely that this palindromic structure in the rbcS gene may also have the same function as suggested by us for that in ALH3-1.1. Coincidently, in other plant genes which encode transcripts with alternative 3' ends in vivo (Dean et al, 1985), we failed to find sequences which may form stable hairpin structure between any two 3' ends. Figure 4.3 shows the putative hairpin structure in the 3' UTR of the alfalfa H3 pre-mRNA derived from the genomic clone, ALH3-1.1. The hairpin structure may participate in the recognition by cleavage/polyadenylation factors (Christofori et al, 1988) or, alternatively, act as a structural factor for transcription termination (Platt, 1986). It should be of interest to prove this speculation experimentally.

4.6. The Expression of Histone H3 Genes In Alfalfa

In RNA blot hybridizations, we have found a significant variation in the accumulation of H3 transcripts among various tissues of different developmental stages in alfalfa (Fig.3.2.5). Comparable results were also observed in maize by Gigot et al (Abstracts of the 2nd Congress of ISPMB, Jerusalem, 1988). These differences may not be simply explained by using an analogy with animal somatic cells, where the majority of histone RNA synthesis is cell cycle-dependent (Schümperli, 1986). Nevertheless, the high amount of H3 mRNAs in somatic embryos may suggest a similarity with the DNA replication-uncoupled histone

synthesis which is seen in animal oocytes and early embryos (Maxson et al, 1983 and Old et al, 1985). However, differences have to be considered as for example the source of histone mRNA from which the uncoupled histones are translated. In some animals, the high demand for histone mRNAs at the early stage of embryogenesis is provided by a RNA pool generated during oogenesis, while the expression of histone genes remains at a very low level (Maxson et al, 1983). Probably, in contrast, the histone genes are highly active throughout embryogenesis in plants because of the lack of animal-like oogenesis which provides the histone mRNA pool. Consistent with this hypothesis, V.Raghavan (see Abstracts of XIV Botanical Congress, 1987, Berlin) found that the H4 gene is highly expressed in newly formed cells in rice anther cultures during embryogenic division. A study on Vicia faba also demonstrated that the synthesis of histone proteins in early germination depends on newly transcribed messengers (Kato et al, 1982).

4.7. Is the Expression of Plant Histone Genes Regulated Replication-Dependently?

The cell cycle-dependence of histone gene expression has been well characterized in animal somatic cells (Maxson et al, 1983; Graves et al, 1987; Marzluff and Pandey, 1988; Schümperli, 1986; 1988). In alfalfa, we showed that the maximal amount of total H3 transcripts coincided with ^{3}H-thymidine incorporation peak in partially synchronized callus cells (Fig.3.2.6). Recently Mikami et al (1987) also found that the binding of a nuclear protein(s), HBP-1, to a specific 5' hexamer, ACGTCA, in histone

H3 and H4 genes showed highest affinity in S phase-abundant cell population of partially synchronized wheat seedlings. These experiments primarily suggested that the expression of histone genes is cell cycle-dependent in plants, as the case in animals. However, the cell culture systems used in both alfalfa and wheat were not well defined and the experiments lacked comprehensive control. Therefore, further studies with highly synchronized cell populations are required to demonstrate a link between histone gene expression and DNA replication.

4.8. The Diversity of H3 Sequences In Alfalfa

The nucleotide sequence compilation of the H3 cDNAs with ALH3-1.1 confirms the restriction enzmye analysis (Fig.3.2.8) showing that the cDNAs belong to two classes: a dominant class and a minor class. The dominant class is represented by clone pH3c-1, pH3c-3 and pH3c-12. These sequences vary only in very few nucleotides in both translated and untranslated region with the genomic clone, ALH3-1.1 (Fig.3.2.9). Clone pH3c-7, pH3c-10 and pH3c-17 are believed to fall into this class because of their identical restriction patterns with the dominant class (Fig.3.2.8). In contrast, clone pH3c-11 is considered to represent a minor gene which has only 78.8% homology with the major class in the translated region and encodes a longer non-homologous 3' UTR (Fig.3.2.9). The low homology between these two classes of alfalfa H3 genes results in four unique amino acid replacements (Fig.4.5). Surprisingly, the amino acid sequence of this minor alfalfa H3 variant is identical to that of a barley H3 histone deduced from a partial H3 cDNA sequence

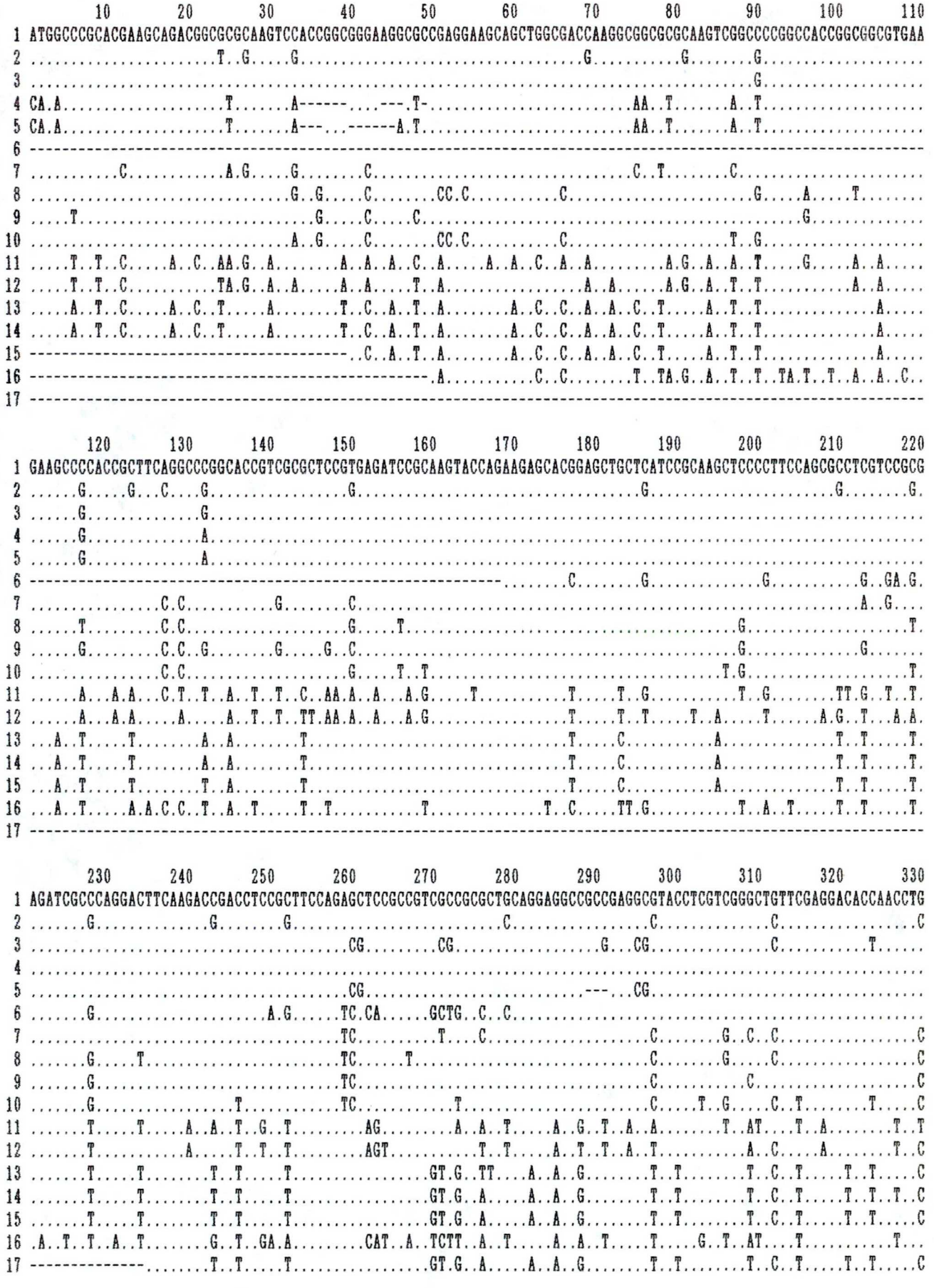
(Chojecki, 1986), although barley and alfalfa belong to two taxonomic groups which adopt completely different coding strategies(see below). Despite the four amino acid changes, the 21.2% nucleotide divergence between these two classes of alfalfa H3 genes is mostly due to the different preference of synonymous codons (Table 4.4 and Table 4.5). If assuming that the rate of accumulation of neutral point mutation among histone H3 genes during evolution is about 0.45% per million year (Miyata and Yasunaga, 1980), these two classes of alfalfa H3 genes may have coexisted for more than 45 millon years. In attempt to study the expression of this minor H3 gene, we used the 3'UTR of pH3c-11 as a specific probe to hybridized the same RNA samples as those in Figure 3.2.5. Under the same hybridization conditions, we could not detect the presumed longer mRNA band possibly encoded by the minor H3 gene. Thus we propose that the expression level of this gene might be very low in the studied tissues.

We have estimated about 160 copies of histone H3 sequences existing in the tetraploid alfalfa genome (Fig.3.2.3). This high copy number may reflect a composition of major and minor gene families and probably pseudogenes. A recent study on alfalfa histone proteins also revealed the existence of different subtypes of H3 histones (Waterborg et al, 1987). The diversity of alfalfa cDNAs is consistent with these observations.

4.9. The Codon Usage of Plant Histone Genes

We have analyzed the codon usage of alfalfa histone H3 genes and pointed out that the 21.2% nucleotide divergence between the dominant and minor histone H3 genes is mostly resulted from

their choice of different coding strategies. We also noticed that the alfalfa histone H3 sequences display no exclusive GC preference in their codon usage. In attempt to obtain more generalized information, we analyzed the coding region of all plant histone H3 genes by compiling them with that of the rice H3 gene, H3R-11. As shown in Figure 4.4, most of the base substitutions between monocot and dicot H3 genes are being C's and G's to T's and A's. Such dramatical difference leads to a 20% or more nucleotide divergence between monocot and dicot H3 genes (Table 4.3). Because of the high conservation of histone H3 proteins (Fig.4.5), the nucleotide divergence is mostly contributed by the unique codon usage of individual genes. Table 4.4 shows the nucleotide composition of the codon position III used by plant histone genes. Among all monocot histone H3 and H4 genes, the A and T-ended codons, particularly the A-ended codons, are almost completely avoided. They compose less than 10% of all the possible codons. In contrast to the exclusive preference of G and C, the dicot histone genes select a totally different coding strategy. Here the A and T-ended codons are no longer depressed. They consist over 40% of the possible codons. Such more or less-balanced choice of G, C, A, and T in codon position III is also adopted by mammal histone H3.3 genes (represented by human H3.3 gene in Table 4.4), a gene family which expresses at a low level throughout the cell cycle (review in Stein et al, 1984, also see Wells et al, 1987). Interestingly, the proportion of A and T-ended codons used by cell-cycle dependent animal H3 genes (represented by a sea urchin late H3 gene) is just between those of monocot and dicot



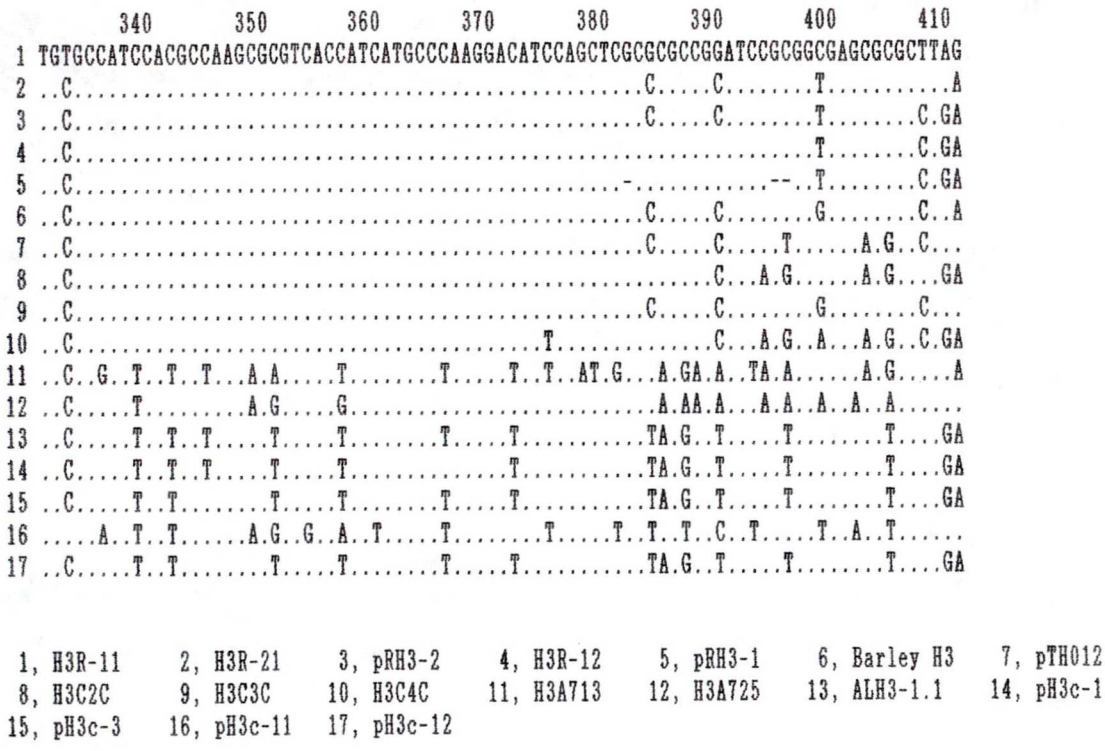


Fig.4.4. Compilation analysis of plant histone H3 genes. The coding regions or assumed coding regions of plant histone H3 genes, pseudogenes and cDNAs were compiled with that of the rice H3 gene, H3R-11. Deletions are marked by dashes and the nucleotides identical to those in H3R-11 are dotted. For comparison, the presumed sequencing errors in sequence 3 and 5 remain unchanged (Xie et al, 1987).

Table 4.3. Homology of plant histone H3 genes with the coding region of H3R-11 (%)

Species	clone	Homology	Reference
Monocot			
Rice	H3R-21	93.7	This thesis
	pRH3-2	96.8*	Peng,1986
	H3R12	93.2	This thesis
	pRH3-1	92.0*	Xie,1987
Barley	H3	89.3	Chojecki,1986
Wheat	pTH012	92.7	Tabata,1984a
Corn	H3C2	91.7	Chaubet,1986
	H3C3	93.9	Gigot,1987
	H3C4	90.3	Chaubet,1986
Dicot			
Arab.	H3A713	75.2	Chaboute,1987
	H3A725	80.3	
Alfa.	ALH3-1.1	82.5	Wu,1988
	pH3c-11	75.7	This thesis

*The supposed sequencing errors (Fig.3.1.2) were corrected before comparison. Arab.=Arabidopsis and Alfa.=Alfalfa.

Table 4.4. Comparison of the nucleotide composition presents at the third position of the codons(%)

Species	Clone	G	C	A	T	Reference
Monocot						
Barley	H3	40.0	60.0	0	0	Chojecki,1986
Corn	H3C2	41.5	52.6	0.7	5.2	Chaubet,1986
	H3C3	43.0	56.3	0	0.7	Gigot,1987
	H3C4	40.0	51.1	1.5	7.4	Chaubet,1986
	H4C7	37.3	52.9	2.9	6.9	Philipps,1986
	H4C13	37.3	52.9	3.9	5.9	Gigot,1987
	H4C14	36.3	57.8	1.0	4.9	Philipps,1986
Rice	H3R-11	41.5	56.3	0	2.2	This thesis
	H3R-21	47.4	50.4	0	2.2	
	pRH3-2	42.2	55.6	0	2.2	Peng,1986
Wheat	pTH012	36.3	61.5	0.7	1.5	Tabata,1984a
	pTH011	31.4	64.7	2.0	2.0	Tabata,1983
	pTH091	29.4	67.6	2.0	1.0	Tabata,1984b
Dicot						
Alfa.	ALH3-1.1	21.5	37.8	11.9	28.9	Wu,1988
	pH3c-11	25.0	20.0	15.8	39.2	This thesis
Arab.	H3A713	28.1	22.2	25.2	24.4	Chauboute,1987
	H3A725	25.9	31.1	22.2	20.7	
	H4A748	25.5	28.4	15.6	30.4	
	H4A777	24.5	30.4	17.6	27.5	
Pea*	PsH1b	32.5	18.1	24.9	24.5	Gantt,1986
Animal						
Human**	HuH3-149	22.2	21.5	22.2	34.1	Wells,1987
S. ur.	SpL22	27.4	46.7	5.9	20.0	Kaumeyer,1986

* Full length histone H1 cDNA. ** Human histone H3.3 gene. Alfa.=Alfalfa, Arab.=Arabidopsis and S.ur.=Sea urchin

histone genes (Table 4.4).

Comparison of the individual codons used by plant histone H3 genes provides more detailed information (Table 4.5). Generally, the monocot H3 genes use only the G and C-ended codons with slight preference of C-ended codons. This narrow selection of codons results in high conservation among monocot H3 genes (over 90%, Table 4.3). In contrast, the dicot H3 genes select codons

Table 4.5. The Number of individual codons used in histone H3 genes from different species*

A.A	Codon	Alfalfa A-1.1	Alfalfa pc11	Arabidopsis A713	Arabidopsis A725	Rice R-11	Rice R-21	Corn H3C2	Wheat TH012	Human Hu149	S.ur. SpL22
Phe	TTT	1	2	1	0	0	0	0	0	2	0
	TTC	4	2	4	5	5	5	5	5	2	4
Leu	TTA	0	0	0	1	0	0	0	0	0	0
	TTG	1	2	3	0	0	0	0	0	1	1
	CTT	1	7	4	5	0	0	0	0	4	3
	CTC	10	1	2	5	7	9	8	8	1	4
	CTA	0	0	1	0	0	0	0	1	1	0
	CTG	0	3	2	1	5	3	4	3	5	4
Ile	ATT	1	5	3	1	0	0	1	0	4	0
	ATC	6	2	4	6	7	7	6	7	3	7
	ATA	0	0	0	0	0	0	0	0	1	0
Met	ATG	1	1	1	1	1	1	1	1	2	3
Val	GTT	0	2	3	1	0	0	0	0	1	1
	GTC	4	3	2	4	5	5	4	2	0	5
	GTA	0	0	0	0	0	0	0	0	1	0
	GTG	2	1	1	1	1	1	2	4	3	0
Ser	TCT	1	1	0	1	0	0	0	0	1	1
	TCC	2	0	1	0	2	1	2	4	1	2
	TCA	0	0	1	1	0	0	0	0	0	0
	TCG	2	0	0	0	1	2	2	1	1	0
	AGT	0	1	0	1	0	0	0	0	2	1
	AGC	2	1	3	2	2	2	1	1	1	2
Pro	CCT	2	4	2	1	0	0	1	0	2	1
	CCC	1	0	0	2	4	2	4	4	3	3
	CCA	2	1	2	2	0	0	0	0	1	2
	CCG	1	0	2	1	2	4	1	2	0	0
Thr	ACT	4	3	3	2	0	0	0	0	4	1
	ACC	5	3	5	5	7	5	6	8	2	9
	ACA	1	1	2	1	0	0	0	0	4	0
	ACG	0	1	0	2	3	5	4	2	0	0
Ala	GCT	11	9	6	11	1	2	2	1	8	7
	GCC	6	2	3	5	10	10	8	14	4	10
	GCA	1	4	5	0	0	0	1	0	4	1
	GCG	1	0	6	4	9	8	9	4	2	0
Tyr	TAT	1	0	1	0	0	0	0	0	2	0
	TAC	1	3	1	2	2	2	2	2	1	3

His	CAT	1	2	1	0	0	0	0	0	2	1
	CAC	1	1	1	2	2	2	2	2	0	1
Gln	CAA	3	1	3	0	0	0	0	0	1	0
	CAG	5	6	5	8	8	8	8	8	7	8
Asn	ATT	0	1	1	1	0	0	0	0	0	0
	ACC	1	0	0	0	1	1	1	1	1	1
Lys	AAA	6	2	5	5	0	0	0	0	8	0
	AAG	8	9	9	9	14	14	14	14	5	13
Asp	GAT	4	2	3	1	0	0	1	0	1	0
	GAC	0	2	1	3	4	4	3	4	3	3
Glu	GAA	1	3	4	5	0	0	0	0	5	1
	GAG	6	4	3	2	7	7	7	7	2	7
Cys	TGT	0	1	0	0	1	0	0	0	1	0
	TGC	1	0	1	1	0	1	1	1	0	1
Arg	CGT	10	6	5	3	1	0	1	1	8	10
	CGC	4	4	1	0	13	8	13	13	6	6
	CGA	0	1	0	1	0	0	0	0	1	1
	CGG	0	0	0	0	2	8	1	0	0	1
	AGA	0	1	5	7	0	0	0	0	1	0
	AGG	3	3	6	6	1	1	2	3	2	1
Gly	GGT	2	1	0	0	0	1	1	0	4	1
	GGC	3	0	1	0	5	4	5	7	1	2
	GGA	2	4	6	7	0	0	0	0	2	3
	GGG	0	1	1	0	2	2	1	0	1	0

*The Genes compared were from the same sources as Table 4.**4**. Abbreviations used: S.ur.=Sea urchin; A-1.1=ALH3-1.1; pc-11=pH3c-11; A713=H3A713; A725=H3A725; R-11=H3R-11; R-21=H3R-21; TH012=pTH012 and Hu149=HuH3-149.

with much wider range which leads to a higher divergence (about 20%, Table 3.2.1) among these genes. The codon preference in dicot histone genes does not follow a general rule. For example, ALH3-1.1 of alfalfa favors CTC, ATC, GCT, GAT, GAG and CGT, alfalfa minor H3 gene prefers CTT, CCT, GCT and AAG, H3A725 of Arabidopsis overuses TTC, ATC, GCT, CAG and GGA, while A713 does

not obviously favor any individual codons except GGA (Table 4.5). Despite the divergence, however, some codons are commonly favored among H3 genes, such as CAG and, except those of human H3.3 and/or alfalafa minor H3 gene, AAG, TTC, ATC, ACC. The codon GCT is frequently used among dicot H3 and human H3.3 genes. As most of the eukaryotic genes (Murray et al, 1989) the codon XTA (X=A,T,C,G) is rarely used in all the compared H3 genes.

Very recently Murray et al (1989) have made a comprehensive investigation on codon usage in plant genes. Generally, the "codon dialect" of plant histone genes analyzed here is in agreement with that of other plant genes. As in the case of histone genes, the monocot genes highly active in leaves such as CAB or RuBPC SSU genes almost completely avoid the use of A and T at codon position III. But this codon bias seems not as pronounced in non-leaf genes such as ADH, zein 22 kDa subunit, sucrose synthetase and ATP/ADP translocator genes. Several studies have suggested that the difference in codon usage among genes reflects the modulation of gene expression rather than phylogenetic relationship (Grantham et al, 1981; Wells et al, 1986; Ernst, 1988). Possibly the wobble codons differ in the affinity of codons and their anticodons (Thomas et al, 1988). Accordingly, the activity of histone genes may be somewhat different between monocot and dicot plants. The very different coding strategies between monocot and dicot histone genes may serve as a model system to investigate the possible regulatory function of genetic codes.

4.10. The H3 Histones in Plants Are Different from Those in Animals

In alfalfa, we observed two different H3 histone variants represented by dominant and minor cDNA species. As in animals, in fact, many plant species were found to contain more than one histone variants for each type of histone (Spiker, 1985; 1988; Waterborg et al, 1987). The amino acid sequences of some of these histone variants have been either directly determined or deduced from their genes (Spiker, 1988). However, despite the divergence, plant H3 histones are highly conserved. Interestingly, at four positions the amino acids of plant H3 histones are different from those of animals' (Fig.4.5). The substitutions at position 41 and 53 are both "conservative" (Von Holt et al, 1979), a term used for those changes between the amino acids with similar chemical properties. The most intriguing difference between the animal and plant H3 histones is located at position 90. At this position, the amino acids can be either a hydrophobic Ala and Leu or a polar Ser (Fig.4.5) in plant H3 histone, while in animal, at the same position, the amino acid is exclusively either a Met in H3.1 and H3.2 histones or a Gly in H3.3 histone which was considered to be an ancient protein (Wells et al, 1986). It should be of particular interest to study whether these amino acid replacements in plant H3 histones have any structural effect on plant chromatin. In comparison with the plant H3 consensus sequence, the amino acid at position 41 of the alfalfa minor H3 variant is identical to that in animal H3 consensus (Fig.4.5). Interestingly, two H3 histones (Hayashi et al, 1984) from a ciliate protozoa,

```
          10        20        30        40        50        60        70
1  ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRL
2  ....................................F...........K.................
3  ....................................F...........K.................
4  ....................................F...........K.................
5  ....................................F...........K.................
6  ....................................F...........K.................
7  ....................................F.............................
8  ....................................F...........K.................
9                                                    .................
10 ....................................F...........K.................
11                 ..............T.....................K.................
12 ....................................F...........K.................
13 ............A.........S.............I.....F...........K.....D...........
14 ............V..................VS.......KF...........K...T.D...........
15 ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFRPGTVALREIRKYQKSTELLIRKLPFQRL

          80        90       100       110       120       130
1  VREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA
2  ....................S.....A.....................................
3  ....................S.....S.....................................
4  ....................S.....A.....................................
5  ....................A.....A.....................................
6  ....................A.....A.....................................
7  ....................A.....A.....................................
8  ....................A.....S..........S..............V...........
9  ..................H..L.....A.....................................
10 ....................S.....A.....................................
11 ..................H..L.....A.....................................
12 ....................A.....A.....................................
13 ..D..HE..AE........L.....A...................R.....T..M..........F
14 ..D..MEM.N.I....Q.IL.....A...................R.....T..M..........F
15 VREIAQDFKTDLRFQSSAVXALQEAAEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA
```

1, Animal consensus (Wells,1986)
2, Pea a (Patthy,1973)
3, Pea b (Patthy,1973)
4, Wheat (Tabata,1984)
5, Arabidopsis (Chaboute,1987)
6, Corn (Chaubet,1986)
7, Cycad a (Brandt,1986)
8, Cycad b (Brandt,1986)
9, Barley (Chojecki,1986)
10, Alfalfa ALH3-1.1
11, Alfalfa pH3c-11
12, Rice
13, Tetrahymena a (Hayashi,1984)
14, Tetrahymena b (Hayashi,1984)
15, Plant H3 Consensus

Fig.4.5. Compilation analysis of plant H3 histones. All plant and two protozoan H3 histones are compiled with the consensus amino acid sequence of animal H3 histones (Wells, 1986). The letter X in the consensus amino acid sequence of plant H3 histone (No.15) represents the variable amino acids at position 90. The H3r deduced from pRH3-2 (Peng and Wu, 1986) is excluded in this figure because of the putative errors described in the text.

Tetrahymena, where the H3 mRNAs are polyadenylated (Baum et al, 1983), possess the same amino acids at all the four positions as the plant H3 histone consensus, whereas the H3 histones from fungi (Wells, 1986), where the H3 transcripts are also polyadenylated (Fahrner et al, 1982), are closer to the animal H3.3 than the plant H3 histone (Wells, 1986). The studies on H4 histones also revealed two "conservative" amino acid replacements between plant and animal at position 60 and 77 (Spiker, 1988). The conservation and variation of different classes of H3 histone among eukaryotes may be of evolutionary importance. This remains to be studied.

Surprisingly, The H3 histone (H3r) deduced from the nucleotide sequence of pRH3-2 (Peng and Wu, 1986) varies with that encoded by H3R-11 and H3R-21 in three positions (Fig.3.1.3) although the homology between H3R-11 and pRH3-2 is as high as 96% (Table 4.3). Two of these replacements happened at position 86 and 98 where the amino acids are so conservative that they were never changed during the course of evolution (Wells, 1986). Although the third replacement occured at a variable position (Fig.4.5), the change from a neutral Ala to a basic Arg may dramatically affect the biochemical property of H3 protein, particularly taking the identical change at position 98 into account (Fig.3.1.3). Moreover, the nucleotide compilation of rice H3 sequences revealed that all these three amino acid changes in H3r were resulted from dinucleotide substitutions, as compared to H3R-11 or H3R-21, from 5'GC3' to 5'CG3' (marked with solid triangles in Fig.3.1.2). Similar dinucleotide substitutions were also observed in two positions of the

pseudogene, pRH3-1 (Fig.3.1.2). All these facts support the explanation that the distinct amino acid changes in H3r might be due to DNA sequencing errors.

4.11. The General Structure of Plant Histone H3 Genes

In summary, the plant histone H3 genes not only possess the essential characteristics of animal H3 genes, but also exhibit their own specificities.

a) The size of the coding region of plant histone H3 genes is 405 bp which encodes a protein of 135 amino acid residues. This protein is identical to its animal counterpart except four amino acid differences at position 41, 53, 90 and 96, respectively.

b) The codon usage of plant histone H3 genes is highly unique. The monocotyledonous genes exhibit exclusive GC preference in their codon choice. The dicotyledonous genes do not follow the same coding strategy, rather, they differ individually.

c) In addition to, with slightly modifications, all histone gene-specific consensus sequences found in animals, the 5' flanking region of plant H3 genes possesses plant-specific motifs, either CGC box or AC-GA block.

d) In comparison with animals, the 3' untranslated region of plant histone H3 transcripts is much longer and does not contain the highly conserved T-hyphenated hairpin structure and the "downstream element" (Fig.1.1).

e) Most of the H3 transcripts (if not all) in alfalfa, Arabidopsis and corn are polyadenylated.

It has been known that the unique codon usage, 5' specific

consensus motifs and the polyadenylation of transcripts are all the common features of plant histone H4 genes. However, the characteristics of other plant histone genes have yet to be studied.

REFERENCES

Ahmed,C.M.I. and Padayatty,I.D., 1982, Identification of individual histone mRNAs from rice embryos, Indian J. Biochem. Biophys. 19: 160-166.

Alonso,A., Breuer,B., Bouterrfa,H. and Doenecke,D., 1988, Early increase in histone $H1^{0}$ mRNA during differentiation of F9 cells to parietal endoderm, EMBO J. 7: 3003-3008.

Anderson,O.D., Litts,J.C., Gautier,M.-F. and Greene,F.C., 1984, Nucleic acid sequences and chromosome assignment of a wheat storage protein genes, Nucl. Acids Res. 12: 8129-8144.

Aviv,H. and Leder,P., 1972, Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose, Proc. Natl. Acad. Sci. USA 69: 1408-1412.

Baum,G.A., Calzone,F., Bowen,J.K., Allis,C.D. and Gorovsky,M.A., 1983, Multiple independently regulated messages for histone H3 and H4 in Tetrahymena, Nucl. Acids Res. 11: 3903-3917.

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

Brandt,W.F. and Von Holt,C., 1986, The primary structure of histone H3 from cycad pollen, FEBS Lett. 194: 278-281.

Breeden,L., 1988, Cell cycle-regulated promoters in budding yeast, Trends in Genet. 4: 249-253.

Brown,D.T., Yang,Y.-S. and Sittman,D.B., 1988, Histone gene switching in murine erythroleukemia cells is differentiation-specific and occurs without loss of cell cycle regulation, Mol. Cell. Biol. 8: 4406-4415.

Cami,B. and Kourilsky,P., 1978, Screening of cloned recombinant DNA in bacteria by in situ colony hybridization, Nucl. Acids Res. 5: 2381-2390.

Cathala,G., Savouret,J.-F., Mendez,B., West,B.L., Karin,M., Martial,J.A. and Baxter,J.D., 1983, A method for isolation of intact, translationally active ribonucleic acid, DNA 2: 329-335.

Chaboute,M.-E., Chaubet,N., Ehling,M. and Gigot,C., 1987, Genomic organization and nucleotide sequences of two histone H3 and two histone H4 genes of Arabidopsis thaliana, Plant Mol. Biol. 8: 179-191.

Chaboute,M.-E., Chaubet,N., Clement,B., Gigot,C. and Philipps,G., 1988, Polyadenylation of histone H3 and H4 mRNAs in dicotyledonous plants, Gene 71: 217-223.

Challoner,P.B. Moss,S.B. and Groudine,M., 1989, Expression of replication-

dependent histone genes in avian spermatids involves an alternate pathway of mRNA 3'-end formation, Mol. Cell. Biol. 9: 902-913.

Chaubet,N., Philipps,G., Chaboute,M.-E., Ehling,M. and Gigot,C., 1986, Nucleotide sequences of two corn histone H3 genes. The organizations of the corn histone H3 and H4 genes, Plant Mol. Biol. 6: 253-263.

Chaubet,N., Chaboute,M.-E., Philipps,G. and Gigot,C., 1987, Histone genes in higher plants: organization and expression, Dev. Genet. 8: 461-473.

Chaubet,N., Chaboute,M.-E., Clement,B., Ehling,M., Philipps,G. and Gigot,C., 1988, The histone H3 and H4 mRNA are polyadenylated in maize. Nucl. Acids Res. 16: 1295-1304.

Chojecki,J., 1986, Identification and characterization of a cDNA clone for histone H3 in barley, Carlsberg Res. Commun. 51: 211-217.

Christofori,G. and Keller,W., 1988, 3' cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor and a snRNP, Cell 54: 875-889.

Colin,A.M., Catlin,T.L., Kidson,S.H. and Maxson,R., 1988, Closely linked early and late histone H2B genes are differentially expressed after microinjection into sea urchin zygotes, Proc. Natl. Acad. Sci. USA 85: 507-510.

Coruzzi,G., Broglie,R., Edwards,C. and Chua,N.-H., 1984, Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase, EMBL J. 3: 1671-1679.

Cotten,M., Gick,O., Vasserot,A., Schaffner,G. and Birnstiel,M.L., 1988, Specific contacts between mammalian U7 snRNA and histone precursor RNA are indispensable for the in vitro 3' RNA processing reaction, EMBO J. 7: 801-808.

Dalton,S. and Wells,J.R.E., 1988a, Maximal binding levels of an H1 histone gene-specific factor in S-phase correlate with maximal H1 gene transcription, Mol. Cell. Biol. 8: 4576-4578.

Dalton,S. and Wells,J.R.E., 1988b, A gene-specific promoter element is required for optimal expression of the histone H1 genes in S-phase, EMBO J. 7: 49-56.

D'Andrea,R.J., Coles,L.S., Lesnikowski,C., Tabe,L. and Wells,J.R.E., 1985, Chromosomal organization of chichen histone genes: preferred association and inverted duplications, Mol. Cell. Biol. 5: 3108-3115.

Dean,C., Tamaki,S., Dunsmuir,P., Favreau,M., Katayama,C., Dooner,H. and Bedbrock,J., 1986, mRNA transcripts of several plant genes are

polyadenylated at multiple sites in vivo, Nucl. Acids Res. 14: 2229-2240.

Dhaese,P., Greve,H.D., Giden,J., Seurinck,J., Van Montagu,M., Schell,J., 1983, Identification of sequences involved in the polyadenylation of higher plant nuclear transcripts using Agrobacterium T-DNA genes as models, EMBO J. 2: 419-426.

Engel,J.D., Sugarman,B.J. and Dodgson,J.B., 1982, A chicken histone H3 gene contains intervening sequences, Nature, 297: 434-436.

Ernst,J.F., 1988, Codon usage and gene expresssion, Trends in Biotec. 6: 19-199.

Fahrner,K., Yarger,J. and Hereford,L., 1980, Yeast histone mRNA is polyadenylated, Nucl. Acids Res. 8: 5725-5737.

Feinberg,A.P. and Vogelstein,B., 1983, A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity, Anal. Biochem. 132: 6-13.

Frischauf,A.M., Lehrach,H. Poustka,A. and Murray,N., 1983, Lambda replacement vectors carrying polylinker sequences, J. Mol. Biol. 170: 827-842.

Fristensky,B., Lis,J. and Wu,R., 1982, Portable microcomputer software for nucleotide sequence analysis, Nucl. Acids Res. 10: 6451-6463.

Gantt,J.S. and Key,J.L., 1987, Molecular cloning of a pea H1 histone cDNA. Eur. J. Biochem. 66: 119-125.

Gigot,C., Chaubet,N., Chaboute,M.-E., Ehling,M., Philipps,G., 1987, Nucleotide sequences of two histone H3 and H4 genes of corn. Further insight into the structure and organization of the histone genes in higher plants, Plant Physiol. Biochem., 25: 235-247.

Gigot,C., 1988, Histone genes in higher plants, In Architecture of Eukaryotic Genes, Kahl,G.(ed), pp.229-240, Verlags-gesellschaft, Frankfurt.

Gojobori,T., Li,W.-H. and Graur,D., 1982, Patterns of nucleotide substitution in pseudogenes and functional genes, J. Mol. Evol. 18: 360-369.

Grandy,D.K. and Dodgson,J.B., 1987, The structure and organization of the chicken H2B histone gene family, Nucl. Acids Res. 15: 1063-1080.

Grantham,R., Gauthier,C., Gouy,M., Jacobzone,M. and Mercier,R., 1981, Codon catalog usage is a genome strategy modulated for gene expressivity. Nucl. Acids Res. 9: r43-r74.

Graves,R.A., Pandey,N.B., Chodchoy,N. and Marzluff,W.F., 1987, Translation is required for regulation of histone mRNA degradation, Cell 48: 615-626.

Gubler,U. and Hoffman,B.J., 1983, A simple and very efficient method for generating cDNA libraries, Gene 25: 263-269.

Gullis,C.A., Rivin,C.J. and Walbot,V., 1984, A rapid procedure for the determination of the copy number of repetitive sequences in eukaryotic genomes, Plant Mol. Biol. Rep. 2: 24-31.

Hanly,S.H., Bleecker,G.C. and Heintz,N., 1985, Identification of promoter elements necessary for transcriptional regulation of a human histone H4 gene in vitro, Mol. Cell. Biol. 5: 380-389.

Hayashi,T., Hayashi,H., Fusauchi,Y. and Iwai,K., 1984, Tetrahymena histone H3. purification and two variant sequences, J. Biochem. 95: 1741-1749.

Heidecker,G., 1986, Structural analysis of plant genes, Ann. Rev. Plant Physiol. 37: 439-466.

Heidl,L.M., Veil,T.S. and Perry,M., 1988, Promoter sequences required for transcription of Xenopus laevis histone genes in injected frog oocyte nuclei, Mol. Cell. Biol. 8: 3676-3682.

Hentschel,C.C., Birnstiel,M.L., 1981, The organization and expression of histone gene families, Cell 25: 301-313.

Hong,J.C., Nagao,R.T. and Key,J.L., 1987, Characterization and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean, J. Biol. Chem. 262: 8367-8376.

Howell,A.M., Cool,D., Hewitt,J., Ydenberg,B., Smith,M.J. and Honda,B.M., 1987, Organization and unusual expression of histone genes in the sea star Pisater ochraceus, J. Mol. Evol. 25: 29-36.

Humphrey,T. and Proudfoot,N.T., 1988, A beginning to the biochemistry of polyadenylation, Trends in Genet. 4: 243-245.

Hunt,A.G., Chu,N.M., Odell,J.T., Nagy,F., Chua,N.-H., 1987, Plant cells do not properly recognize animal polyadenylation signals, Plant Mol. Biol., 8: 23-35.

Hunt,A.G., 1988, Identification and characterization of cryptic polyadenylation sites in the 3' region of a pea Rubisco small subunit gene, DNA 7: 329-336.

Hwang,I. and Chae,C.-B., 1989, S-phase-specific transcription regulatory elements are present in a replication-independent testis-specific H2B histone gene, Mol. Cell. Biol. 9: 1005-1013.

Ikenaya,H. and Saigo,K., 1982, Insertion of a movable genetic element, 297, into the TATA box for the H3 histone gene in Drosophila melanogaster, Proc. Natl. Acad. Sci. USA 79: 4143-4147.

Iyengar,G.A.S., Gaddipati,J.P. and Sen,S.K., 1979, Characteristics of nuclear DNA in the genus Oryza, Thoer. Appl. Genet. 54: 219-224.

Jefferson,R.A., 1987, Assaying chimeric genes in plants: the GUS gene fusion system, Plant Mol. Biol. Rep. 5: 387-405.

Kato,A., Fukuei,K. and Tanifuji,S., 1982, Histone synthesis during the early stages of germination in Vicia faba embryonic axes, Plant Cell Physiol. 23: 967-976.

Kaumeyer,J.F. and Weinberg,E.S., 1986, Sequence, organization and expression of late embryonic H3 and H4 genes from the sea urchin, Nucl. Acids Res. 14: 4557-4576.

Kedes,L.H., 1979, Histone genes and histone messengers, Ann. Rev. Biochem. 48: 837-870.

Kemler,I. and Busslinger,M., 1986, Characterization of two nonallelic pairs of late histone H2A and H2B genes of sea urchin: differential regulation in the embryo and tissue-specific expression in the adult, Mol. Cell. Biol. 6: 3746-3754.

Kim,Y.-Z., Hwang,I., Tres,L., Kievszenbaum,A. and Chae,C.-B., 1987, Molecular cloning and differential expression of somatic and testis-specific H2B histone gene during rat spermatogenesis, Dev. Biol. 124:23-34.

Kroeger,P., Stewart,C., Schaap,T., Van Wijnen,A., Hirshman,J., Helms,S., Stain,G. and Stein,J., 1987, Proximal and distal regulatory elements that influence in vivo expression of a cell cycle-dependent human H4 histone gene, Proc. Natl. Acad. Sci. USA 84: 3982-3986.

Levine,B.J., Chodchoy,N., Marzluff,W.F. and Skoultchi,A.L., 1987, Coupling of replication type histone mRNA levels to DNA synthesis requires the stem-loop sequence at the 3' end of the mRNA, Proc. Natl. Acad. Sci. USA 84: 6189-6193.

Lieber,T., Angerer,L.M., Angerer,R.C. and Childs,G., 1988, A histone H1 protein in sea urchin is encoded by a polyA+ mRNA, Proc. Natl. Acad. Sci. USA 85: 4123-4127.

Liebermann,D., Hoffman-Liebermann,B., Weinthal,J., Childs,G., Maxson,R., Mauron,A., Cohen,S.N. and Kedes,L., 1983, An unusual transposon with long terminal inverted repeats in the sea urchin S. purpuratus, Nature

306: 342-347.

Liebermann,D., Hoffman-Liebermann,B., Troutt,A.B., Kedes,L. and Cohen,S.N., 1986, Sequeces from sea urchin TU transposons are conserved among multiple eukaryotic species, including human, Mol. Cell. Biol., 6: 218-226.

Liu,T.-J., Liu,L. and Marzluff,W.F., 1987, Mouse histone H2a and H2b genes: four functional genes and a pseudogenes undergoing gene conversion with a closely linked functional gene, Nucl. Acids Res. 15: 3023-3040.

Lüscher,B., Stauber,C., Schindler,R. and Schumperli,D., 1985, Faithful cell-cycle regulation of a recombinant mouse histone H4 gene is controlled by sequences in the 3'-terminal part of the gene, Proc. Natl. Acad. Sci. USA 82: 4389-4393.

Maniatis,T., Fritsch,E.F. and Sambrook,J., 1982, Molecular Cloning, A laboratory manual, CSH, New York.

Manley,J.L., 1988, Polyadenylation of mRNA precursors, Biochem. Biophys. Acta 950, 1-12.

Maxson,R., Cohn,R. and Kedes,L., 1983, Exppression and organization of histone genes, Ann. Rev. Genet. 17: 239-277.

May,G.S. and Morris,N.R., 1987, The unique histone H2A gene of Aspergillus nidulans contains three introns, Gene 58: 59-66.

Marzluff,W.F., Pandey,N.B., 1988, Multiple regulatory steps control histone mRNA concentrations, Trends in Biochem. Sci. 13: 49-52.

Mazzolini,L., Vaeck,M., and Van Montagu,M., 1989, Conserved epitopes on plant H1 histones recognized by monoclonal antibodies, Eur. J. Biochem. 178: 779-787.

Meyerowitz,E.M., 1989, Arabidopsis, a useful weed, Cell 56: 263-269.

Mikami,K., Tabata,T., Kawata,T., Nakayama,T. and Iwabuchi,M., 1987, Nuclear protein(s) binding to the conserved DNA hexameric sequence postulated to regulate transcription of wheat histone genes. FEBS Lett. 223: 273-278.

Miyata,T. and Yasunaya,T., 1980, Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application, J. Mol. Evol. 16: 23-26.

Miyata,T. and Yasunaya,T., 1981, Rapidly involving mouse a-globin-related pseudogene and its evolutionary history, Proc. Natl. Acad. Sci. USA 78: 450-453.

Morse,R.H. and Simpson,R.T., 1988, DNA in the nucleosome, Cell 54: 285-287.

Mowry,K.L. and Steitz,J.A., 1988, snRNP mediators of 3' end processing: functional fossils? Trends in Biochem. Sci. 13: 447-451.

Murray,M.G., Thompson,W.F., 1980, Rapid isolation of high molecular weight plant DNA, Nucl. Acids Res. 8: 4321-4327.

Murray,M.G., Kennard,W.C., Drong,R.F. and Slighton, J.C., 1984, Use of a recombination-deficient phage lambda system to construct wheat genomic libraries, Gene 30:237-240.

Murray,E.E., Lotzer,J. and Eberle,M., 1989, Codon usage in plant genes, Nucl. Acids Res. 17: 477-498.

Old,R.W., Sheikh,S.A., Chambers,A., Newton,C.A., Mohammed,A. and Aldridge,T.C., 1985, Individual Xenopus histone genes are replication-indipendent in oocytes and replication-dependent in Xenopus or mouse somatic cells, Nucl. Acids Res. 13: 7341-7358.

Old,R.W. and Woodland,H.R., 1984, Histone genes: not so simple after all, Cell 38: 624-626.

Oslay,M.A., Gould,J., Kim,S.-Y., Kane,M. and Hereford,L., 1986, Identification of sequences in a yeast histone promoter involved in periodic transcription, Cell 45: 537-544.

Patthy,L., Smith,E.L. and Johnson,J., 1973, Histone III: V, The amino acid sequence of pea embryo histone III, J. Biol. Chem. 248: 6834-6840.

Peng,Z.-G. and Wu,R., 1986, Rapid nucleotide sequencing strategy and its application in analysing a rice histone H3 gene, Gene 45: 247-252.

Philipps,G., Chaubet,N., Chabout,M.-E., Ehling,M. and Gigot, C., 1986, Genomic organization and nucleotide sequences of two corn histone H4 genes, Gene 42: 225-229.

Platt,T., 1986, Transcription termination and the regulation of gene expression, Ann. Rev. Biochem. 55: 339-372.

Potter,S.S., 1982, DNA sequence of a foldback transposable element in Drosophila, Nature 297: 201-204.

Ruddell,A., Jacobs-Lorena,M., 1985, Biphasic pattern of histone gene expression during Drosophila oogenesis, Proc. Natl. Acad. Sci. USA 82: 3316-3319.

Sanger,F., Nicklen,S. and Coulson,A.R., 1977, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Schaufele,F., Gilmartin,G.M., Bannwarth,W. and Birnstiel,M.L., 1986, Compensatory mutations suggest that base-paring with a small nuclear RNA is required to form the 3' end of H3 messenger RNA, Nature 323: 771-781.

Schell,J., 1987, Transgenic plants as tools to study the molecular organization of plant genes, Science 237: 1176-1182.

Schenck,R.U. and Hildebrandt,A.C., 1972, Medium and techniques for induction and grouth of monocotyledonous and dicotyledonous plant cell culture, Can. J. Bot. 50: 199-204.

Schümperli,D., 1986, Cell cycle regulation of histone gene expression. Cell 45: 471-472.

Schümperli,D., 1988, Multiple regulation of replication-dependent histone genes, Trends in Genet. 4: 187-191.

Song,Z.R., Fienberg,A., Chorneau,R., Borkird,C., Furner,I. and Smith,J., 1984, Developmental biology of embryogenesis from carrot culture, Plant Mol. Biol. Rep. 2: 3-14.

Spiker,S., 1982, Histone variants in plants, J. Biol. Chem. 257: 14250-14255.

Spiker,S., 1985, Plant chromatin structure, Ann. Rev. Plant Physiol. 36: 235-253.

Spiker,S., 1988, Histones and HMG proteins of higher plants, In Architecture of Eukaryotic Genes, Kahl,G.(ed), pp.143-162, Verlags-gesellschaft, Frankfurt.

Stauber,C., Luscher,B., Eckner,R. and Schümperli,D., 1986, A signal regulating mouse histone H4 mRNA levels in a mammalian cell cycle mutant and sequences controlling RNA 3' processing are both contained within the same 80bp fragment, EMBO J. 5: 3297-3303.

Stauber,C., Schümperli,D., 1988, 3' processing of pre-mRNA plays a major role in proliferation-dependent regulation of histone gene expression, Nucl. Acids Res. 16: 9399-9414.

Stein,G.S., Stein,J.L. and Marzluff,W.F., 1984, Histone genes, structure, organization and regulation, John Wiley & Sons, New York.

Stuart,D.A., Nelson,J., Strickland,S.G. and Nichol,J.W., 1985, Factors affecting developmental processes in alflfa cell cultures, In Tissue Culture in Forestry and Agriculture, Henke,R.R., Hughes,W.K., Constantin,M.P. and Hollaeder,A.(eds) pp.59-73, Plenum Press, New York.

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

Tabata,T., Sakaki,K and Iwabuchi,M., 1983, The structural organization and DNA sequence of a wheat histone H4 gene. Nucl. Acids Res. 11: 5865-5875.

Tabata,T., Fukasava,M., Iwabuchi,M., 1984a, Nucleotide sequence and genomic organization of a histone H3 gene. Mol. Gen. Genet. 196: 397-400.

Tabata,T. and Iwabuchi,M., 1984b, Molecular cloning and nucleotide sequence of a variant wheat histone H4 gene, Gene 31: 285-289.

Tabata,T. and Iwabuchi,M., 1986, Characteristic sequences the in 3'-flanking region of wheat histone genes, Plant Cell Physiol. 27: 929-933.

Tabata,T., Terayama,C., Mikami,K, Uchimiya,H. and Iwabuchi,M., 1987, An accurate transcription of wheat histone genes in sunflower cells, Plant Cell Physiol. 28: 73-82.

Thomas,G. Vasavada,H.A., Zachariah,E. and Padayatty,J.D., 1983a, Cloning of rice DNA and identification of the cloned H2a, H2b and H4 histone genes, Ind. J. Biochem. Biophy. 20: 8-15.

Thomas,G. and Padayatty,J.D., 1983b, Organization and bidirectional transcription of H2A, H2B and H4 histone genes in rice embryos, Nature 306: 82-84.

Thomas,G. and Padayatty,J.D., 1984, Restriction map and partial sequence of a rice DNA fragment carrying histone H2a, H2b and H4 genes, Ind. J. Biochem. Biophy. 21: 1-6.

Thomas,L.K., Dix,D.B. and Thompson,R.C., 1988, Codon choice and gene expression: synonymous codons differ in their ability to direct aminoacrylated-transfer RNA binding to ribosomes in vitro, Proc. Natl. Acad. Sci. USA, 85: 4242-4246.

Tinico,I., Borer,P., Dengler,B., Levine,M., Uhlenbeck,O., Crothers,D. and Gralla,J., 1973, Improved estimation of secondary structure in ribonucleic acids, Nature 246: 40-41.

Trainor,C.D., Stamler,S.J. and Engel,J.D, 1987, Erythroid-specific transcription of the chicken histone H5 gene is directed by a 3' enhancer, Nature 328: 827-930.

Turner,P.C., Bagenal,E.B.D., Vlad,M.T. and Woodland,H.R., 1988, The organization and expression of histone genes from Xenopus borealis, Nucl. Acids Res. 16: 3471-3485.

Uchimiya,T., Murashige,M., 1974, Evolution of parameters in the isolation of viable protoplasts from cultured tobacco cells, Plant Physiol. 54: 936-944.

Vanin,E.F., 1983, Globin pseudogenes, In Regulation of Hemoglobin biosythesis, Goldwasser,E.(ed), pp.69-88, Elsevier Biomedical, New York.

Vanin,E.F., 1985, Processed pseudogenes: Characteristics and evolution,

Ann. Rev. Genet. 19: 253-272.

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

Waterborg,J.H., Winicov,I. and Harington,R.E., 1987, The histone variants and acetylated species from the alfalfa plant Medicago sativa, Arch. Bioch. Biophys. 256: 167-178.

Weiner,A.M., 1986, Non-viral retroposons: genes, pseudogenes and transposable elements generated by the reverse flow of genetic information, Ann. Rev. Biochem. 55: 631-661.

Wellman,S.C., Casano,P.J., Pilch,D.R., Marzluff,W.F. and Sittman,D.B., 1987, Characterization of mouse H3.3-like histone genes, Gene 59: 29-39.

Wells,D., Bains,W. and Kedes,L., 1986, Codon usage in histone gene families of higher eukaryotes reflects functional rather than phylogenetic relationships, J. Mol. Evol. 23: 224-241.

Wells,D.E., 1986, Compilation analysis of histones and histone genes, Nucl. Acids Res. 14: r119-149.

Wells,D.E., Hoffman,D. and Kedes,L., 1987, Unusual structure, evolutionary conservation of non-coding sequences and numerous pseudogenes characterize the human H3.3 histone multigene family, Nucl. Acids Res. 15: 2871-2880.

White,B.A. and Bancroft,F.C., 1982, Cytoplasmic dot hybridization, J. Biol. Chem. 257: 8569-8572.

Wu,S.-C., Cai,Y.-X. and Wang,X.-M., 1987, Isolation and organizational analysis of histone H3 genes from rice, Acta Genetica Sinica 9: 89-98.

Wu,S.-C., Bögre,L., Vincze,É., Kiss,G.B. and Dudits,D., 1988, Isolation of an alfalfa histone H3 gene: structure and expression, Plant Mol. Biol. 11: 641-649.

Wu,T, and Wu,R., 1988, A new rice repetitive DNA shows sequence homology to both 5S RNA and tRNA, Nucl. Acids Res. 15: 5913-5923.

Xie,Y., Peng,Z.-G., Cai,Y.-X. and Wu,R., 1987, Cloning and analysis of histone H3 gene and Rubisco small subunit gene of rice, Scientia Sinica 30: 706-719.

Yanisch-Perron,C., Vieira,J. and Messing,J., 1985, Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors, Gene 33: 103-119.

Zlatanova,J. and Ivanov,P., 1988, DNA and histone synthesis are uncoupled during germination of maize embryos, Plant Sci. 58: 71-76.

Appendix 1.

Papers and Publications

1. Wu,S.-C., Cai,Y.-X. and Wang,X.-M., 1987, Isolation and organizational analysis of histone H3 genes from rice, Acta Genetica Sinica 9: 89-98.

2. Wu,S.-C., Bögre,L., Vincze,É., Kiss,G.B. and Dudits,D., 1988, Isolation of an alfalfa histone H3 gene: structure and expression, Plant Mol. Biol. 11: 641-649.

3. Wu,S.-C., Györgyey,J. and Dudits,D., 1989, Polyadenylated H3 histone transcripts and H3 histone variants in alfalfa, Nucl. Acids Res. submitted.

4. Wu,S.-C., Végh,Z., Wang,X.-M., Tan,C.C. and Dudits,D., 1989, The nucleotide sequences of two rice histone H3 genes, Nucl. Acids Res. submitted.

Appendix 2.

The Determined Nucleotide Sequences of the H3 Histone Genes, Pseudogene and cDNAs Isolated from Rice and Alfalfa.

All numberings start from 5' of the sense strand. The initiation and stop codons are indicated with Xs. All the sequences have been handled by an IBM MS-DOS computer and the DNA sequences analyzed in the TEXT have been deposited in the Sequence Data Library of the European Molecular Biology laboratory. The file names stored in the computer and the accession numbers assigned in the EMBL Data Library are listed as the followings:

Computer: IBM, VARYTER XT. Operation: MS-DOS. Editor: Pathminder

Sequences	Genes	Acc.No.	File Names
ALH3-1.1	Alfalfa H3 gene	X13673	ALH3-11.NUM
pH3c-1	Alfalfa H3 cDNA	X13674	PH3C1.NUM
pH3c-3	Alfalfa H3 cDNA	X13675	PH3C3.NUM
pH3c-11	Alfalfa H3 minor cDNA	X13676	PH3C11.NUM
pH3c-12	Alfalfa H3 cDNA	X13677	PH3C12.NUM
H3R-11	Rice H3 gene	X13678	RH3-05FI.NUM
H3R-12	Rice H3 pseudogene	X13679	RH3-08FI.NUM
H3R-21	Rice H3 gene	X13680	52RH3FI.NUM

FILE RH3-05FI.NUM, RICE HISTONE H3 GENE, H3R-11, 14.12.1988, BY S.-C.WU

```
       10         20         30         40         50         60         70         80         90        100        110
GAATTCTGCA AGTTAATATG TGTGCTTTCT CCATAACCAT GTACTAGTTG CTATTTCCAC TACATGAAAA AATGTTACAA ATTCATGGCG AGATTTCAAA AAGGAACCTA

      120        130        140        150        160        170        180        190        200        210        220
GTATTTCTTT TTATAAAAAA AATCCATTCC AAATTCATGG CGAGATTTCA AAAGGAAAGT TATTTGTGCA TTTCGGCACA CCCAAATATG TTTGTAATAA CTGAAAAATA

      230        240        250        260        270        280        290        300        310        320        330
TTGTGAAAAT AATACTCGGT GTTATCATTA ACCAAGTACA GTATGACCAA TTTTCAGGAA ATTTAACGGC GGTTTTATTT CCATACGGCC TTAAGCGATG AGTGGCAGCG

      340        350        360        370        380        390        400        410        420        430        440
TACGGTATGA ATTGAGAAAA ATGGCCGGGC CAGCAATACT TTCCCATCCA GGCGTCTGGA AAAGCCCAAG CTCCCGGGGC TTCACATGGA TGGGCCTGAT AATTAGTTAA

      450        460        470        480        490        500        510        520        530        540        550
ATTTTCCGGC CCATACATAG GATAGGCCCA TGCGAGGAAG ATGAGATGAT GCATCATGAT CTGAGGAGAG ATGGGCCTAT GGGTTGAGGG CCTGGGCCTT CTTGCGAAGC

      560        570        580        590        600        610        620        630        640        650        660
AGGCCAGTCT TCAGATGCTT AGGTTTTGCA TACAGCCCTA CAGTGTTTCA ACTTTGGCGC TGTTTCAAAT GCAGAAATTT TAAACGAAAT TTAAAATGCA ATTTTTTTAA

      670        680        690        700        710        720        730        740        750        760        770
TAGGATTGGA TTTTTTTTGC CGCGCGGAAC CAGCGCCGTA CGATTGCGTT CGAATCGACG GTGCAGATCT CGGGGGCAGA ACGCGGATCG CGATCCGTGG GAGCATGAGC

      780        790        800        810        820        830        840        850        860        870        880
TCTGCACTGT GATTGGTTGG AGCAGCTGCG CGCGGATCGC ATTCGCATCC CGCGGGGCCG CGCCTCCCCA AAATTCCAAA CCCAAAATTT CGCCCAAATT TCGAACATTT

      890        900        910        920        930        940        950        960        970        980        990
CCACTGCGCG CGAATGCGAA TGAGCGAATC CCTCCCAAAT TTCACGCCTA TTTAAACACC GCGAATTCCC CCCTCTCAAT CCACCACCGC CAAGTCTCAA ATCCAAAGAA

     1000       1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
ATCACTCGCC GCCGCCGCCT CGCCTTCTCC GCCGCGCCAA GCTCTCCTCT CCTCCTCCTC GATGGCCCGC ACGAAGCAGA CGGCGCGCAA GTCCACCGGC GGGAAGGCGC
                                                                  XXX
     1110       1120       1130       1140       1150       1160       1170       1180       1190       1200       1210
CGAGGAAGCA GCTGGCGACC AAGGCGGCGC GCAAGTCGGC CCCGGCCACC GGCGGCGTGA AGAAGCCCCA CCGCTTCAGG CCCGGCACCG TCGCGCTCCG TGAGATCCGC

     1220       1230       1240       1250       1260       1270       1280       1290       1300       1310       1320
AAGTACCAGA AGAGCACGGA GCTGCTCATC CGCAAGCTCC CCTTCCAGCG CCTCGTCCGC GAGATCGCCC AGGACTTCAA GACCGACCTC CGCTTCCAGA GCTCCGCCGT

     1330       1340       1350       1360       1370       1380       1390       1400       1410       1420       1430
CGCCGCGCTG CAGGAGGCCG CCGAGGCGTA CCTCGTCGGG CTGTTCGAGG ACACCAACCT GTGTGCCATC CACGCCAAGC GCGTCACCAT CATGCCCAAG GACATCCAGC

     1440       1450       1460       1470       1480       1490       1500       1510       1520       1530       1540
TCGCGCGCCG GATCCGCGGC GAGCGCGCTT AGGCGATCCG CCTCCTTTGG TTCTTGCTTG GTTCGTAGGG ACTTGTCATG TTCTACCAGT TCTTGTTAAT TATTAGATCC
                          X XX
     1550       1560       1570       1580       1590       1600       1610       1620       1630       1640       1650
TTGCCTTGTC ATGTCGTAGT CTTTGATTCT TAGTTGCAAT GTAACCGCTG GTTATTCTGA AATTGCAACC GAAGTAATCT GTTTTATCCC CAAATTTCTG TCCATATTTG

     1660       1670       1680       1690       1700       1710       1720       1730       1740       1750       1760
TCTGAACTGA TTGCAATTTG AGTGGATGAT CAGGTCTGGC TCCATGCGAT AAATTCATCT TATCCTACAC TTGTTCTGTT TTGTTTGATG TTCAGAAATC AGAATGGTTC

     1770       1780       1790       1800       1810       1820       1830       1840       1850       1860       1870
TGAAATTATT CCATTATGCA TTTGTCCCTA GATTTGTAGT TTTCACTAAA AAGTACTGAA TAAATTGCAA TTCAAGTCTG TAGTGTCTTG TATCTGTACC ATAGTGCTAA

     1880       1890       1900
AAAAAAATAC TGGAGCATCC ATGCTATTTT CATCA
```

FILE RH3-08FI.NUM, RICE HISTONE H3 PSEUDOGENE, H3R-12, 14.12.1988, BY S.-C.WU

```
      10         20         30         40         50         60         70         80         90        100        110
GATCGATCAG CAATGAGCAG CAAAGTGTGT ATGAGGTTGG CTTTGCAGAG GCTGTCACGG TTTAAGGATG CAGGGTGCAT CCTGTCACAA TGATCATAAG ACCTCAGTCA

     120        130        140        150        160        170        180        190        200        210        220
GCTATAGGTG GCATCGGCGT CTGGGACTGC ACCACCGGCG AGCAGGCCGA CTTCTTCTAC GAGCCGCCGG GCTGTGCACT CGGCGACGCC GACAAGCTCC AGTGGCTCGA

     230        240        250        260        270        280        290        300        310        320        330
CGGCACGAGC ACGCTCATGG CAGCGACCAT GTTCCTACGG ACAGACACCT CCTTCATCAT CCTGCTGGAC TTCCGGGACA AGAAGAACGT GGCCTGGTTG TGGTCCGACG

     340        350        360        370        380        390        400        410        420        430        440
TCGGCACGCC GGCGTCGCTG GAGGACAAGA ACGTGCTACA CGCCATCGCT ATGGAGGACG GGCGGTGCGC GCGCCCCCTC TCCCCCGCTG CCGTGCCGCG CGCCGCTACT

     450        460        470        480        490        500        510        520        530        540        550
CTCCGCCGCC GAGCTCGCCA TCGCGTCGGC CGTCTCTCTC CCTCTCTCCA GCGCCGCGCC CCCTGCTCCG CCGCCGGCCG CCTCACGCAA GTCTGCCACC ACGCCTGCTC

     560        570        580        590        600        610        620        630        640        650        660
GAGAGAGGAG AGAAAGAAGA GAAAAAGAAA GGATGTGGGG CCTACACCGC CCCGCCGCCG CCTCGCCGTG GCTGAGCCGT CCAAGCTCGG GCCAGCCGCC GTCGCACACT

     670        680        690        700        710        720        730        740        750        760        770
CGCACCTCTG CTCTGCTGCA AGGAAGAAGA AGGGGAAAAA AAGATAGAGA GAGGAAGAAC AGTGTCCAGG GGCATTTTGG TACTTACACA ACTCTTTCTC TCTCCATCTT

     780        790        800        810        820        830        840        850        860        870        880
GACCGGAAAT AATAAAATAA TGGGAAGAAT GTCCAGCAGC TAATTACCAC TTTTTTACAG TGTTCACGAG CAAATACACG ATCTTCGGGT GTCTCACAGC AAAGACCACA

     890        900        910        920        930        940        950        960        970        980        990
ATCTTTGAGT GTCCTGTAGC AAATTTACCC TTCACATTAA GGCCTTGTTT GGCTAATGGG TTATGGAAAT GATTTCCCTC CCACCAAAAG CCATCTTTAA ATCTTAACCA

    1000       1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
TTCATTCTCC CTCTTCCATT TAATCCTAAC TCTTTATTTA TATTTCAATC CCAATCCCAC CTCTCATTTC CCATTATCCA AATACACCAT AAGATCCTGT TAGGTCTGTA

    1110       1120       1130       1140       1150       1160       1170       1180       1190       1200       1210
ATCAGCCCAA TACAGTTGTA AATTACTGGC CCGTATATTT TACGTCGAGT TTATTCATCA TTAAAGCTTT GTTCTTTATC ACCATTTTTC CAACTTGCCT TTCTCGTTTT

    1220       1230       1240       1250       1260       1270       1280       1290       1300       1310       1320
TCACGCGCAC GTTTTTTATA ATGATATATA AAAAATTTTC TATATAAAAA ATATTTTAAA AAATAATATT AATCTATCTT ATAAAAAAAA TAATTAATAT TTAATTAATC

    1330       1340       1350       1360       1370       1380       1390       1400       1410       1420       1430
ATGCAACAAA ACGAGTTTTG TTTTGCGTGC CAGGAGGGAG GGGGTACCCT AATGGGCGAT TGATCGCCGT GGGGATGGGG TAGCGATCAG ATTCGCTCCC CACTCCCTCC

    1440       1450       1460       1470       1480       1490       1500       1510       1520       1530       1540
CTCCACCTGG TTTCCTTTTT TGGCATCGCA TTACTTTCCT ATTTTAGTAA ATTTATACAC CTAAAGTTTA TGCACCTAAA GTTTACACAT ATAAAGTTTA GAGACCAAAA

    1550       1560       1570       1580       1590       1600       1610       1620       1630       1640       1650
GTTTATAAGT CAAAAGTTTA TATATCCGAT TCAAAATTTA ATTTGAATTC AAATATTTTT TATATATAGC ATTTCTATAC ATATAAAGTT TATACACCTA AAGTTTATAG

    1660       1670       1680       1690       1700       1710       1720       1730       1740       1750       1760
ATCTAAAGTT TATAGACCTA AAGTTTATAA GTCAAAAGTT TATATACCCG TTTCAAATTT GAATTTGAAT TCAAATATCC GATTCAAATT TGAATTTGAA TTGAGATATT

    1770       1780       1790       1800       1810       1820       1830       1840       1850       1860       1870
TTTCATATAT AGTATTTCTA TACATGTAAA GTTTATAGAC CTAAAGTTTA TATACCCGAT TCAAATTTGA ATTTTAATTA TATCTGATTC AAATTTAAAT TTGAATTCAA

    1880       1890       1900       1910       1920       1930       1940       1950       1960       1970       1980
ATATTTTCTA TATATAGTAT TTCTATACAT CTAAAGTTTA TAGATCCAAA GTTTATAAGT CAAAAGTTTA CATACCCGAT TCAAATTTGA ATTTGAATTC AAATTTTTTA
```

```
     1990       2000       2010       2020       2030       2040       2050       2060       2070       2080       2090
TATATAGTAT TTCTATACAT AAATTTTTCT AACTTTTGTT TTTCAAAAAA TTTTGTGTGG TCTACTGAAA GAGGAGAAGG AGATGAGGAA GGGGGAAGAG GGAGGAGTAC

     2100       2110       2120       2130       2140       2150       2160       2170       2180       2190       2200
ATCTAGTAGT ATAGGGGAGA GGGTGGGCGG GTGATCGCTA GGCGGGTGGG GGAGCGATCA CCCGCCCATT AGCCTTTCCG GGGAGGGTTC CCTCCTTCCG AACACGGCGT

     2210       2220       2230       2240       2250       2260       2270       2280       2290       2300       2310
AAAATCTTAA ATCCTTAATT GATTTATAAA AATCGAATTA AATTAGATCT AGAAGCATAC GATTGGTTTT GTTGACGTAT AATAAGTTAG ACACAGACCC GCACGAAGCA

     2320       2330       2340       2350       2360       2370       2380       2390       2400       2410       2420
GACGGCGTGC AAGTCAGGGA CTCGAGGAAG CAGCTGGCGA CCAAGGCAAC GTGCAAGTCA GCTCCGGCCA CCGGCGGCGT GAAGAAGCCG CACCGCTTCA GGCCAGGCAC

     2430       2440       2450       2460       2470       2480       2490       2500       2510       2520       2530
CGTCGCGCTC CGTGAGATCC GCAAGTACCA GAAGAGCACG GAGCTGCTCA TCCGCAAGCT CCCCTTCCAG CGCCTCGTCC GCGAGATCGC CCAGGACTTC AAGACCGACC

     2540       2550       2560       2570       2580       2590       2600       2610       2620       2630       2640
TCCGCTTCCA GAGCTCCGCC GTCGCCGCGC TGCAGGAGGC CGCCGAGGCG TACCTCGTCG GGCTGTTCGA GGACACCAAC CTGTGCGCCA TCCACGCCAA GCGCGTCACC

     2650       2660       2670       2680       2690       2700       2710       2720       2730       2740       2750
ATCATGCCCA AGGACATCCA GCTCGCGCGC CGGATCCGCG GTGAGCGCGC CTGATTGATT TAGTCGCTTG GTCTGAGCTT ACCAGAAAAT TGCTAGGGAG TAGTAGTAAT
                                                      XXX

     2760       2770       2780       2790       2800       2810       2820       2830       2840       2850       2860
GTAACTTGTT GCAAACCTGC AATGCATGGC AGCACGGCTG TGAAGTGTTG AGCTGATTAT GAATGAGAAA TCGGTCTTCC ATGTTCAAAT CTATCCATTT GTTTTGCCTG

     2870       2880       2890       2900       2910       2920       2930       2940       2950       2960       2970
TTTTTCATGA AAGATTTGAA ATGAATGTGT GGAATGCCTT TCCGAGTTCA GAGAATCAGA GTCTGCAGGA TTAAAGTGGC CTCGCTTTCT TACTGAATGA TGCGTTAGTG

     2980       2990       3000       3010       3020       3030       3040       3050       3060       3070       3080
TTACATATTG GCATCAATTG AGTTATGCGT GGAGAAGGCA TTCTAAATTT GCATAGAGTA TTTTCTTTTA TAATTGAGCA AACATTTCAA CTGGTCTACC AAGAGTCTGT

     3090       3100       3110       3120       3130       3140       3150       3160       3170       3180       3190
AATAAATTAT TACAAAATGG TCATAGTTTA ATTCACACAA ACTCAGTTTA CAAATCCGAA TCTGCATATT GAAAAGCTAT ATCATGGTTC GTTATTAAAG CATAGTTGTT

     3200       3210       3220       3230       3240       3250       3260       3270       3280       3290       3300
CAGTTCGTTA CTTAACACAA AAAACAGTTT TGGTGATGCT TGTAGCAGAG GTTAAAAGTA GCATCATATT GTGTTATTTG CGACACTCCT CAATAGCAGA ACGCAGCGCC

     3310       3320       3330       3340       3350       3360       3370       3380       3390       3400       3410
CGATTTGGAA CGAGTTTCTT GTGTTCAAGT TCAACATTGG TAATTGGTGA TGTGTTATGG CCTCTGCGGA ACCAATCCCT GATAAACTTA GCTTCGTAGG TGTGTCCATC

     3420       3430       3440       3450       3460       3470       3480       3490       3500       3510       3520
AGCTGCCATA AGAGGATCAT TCATGACTTC CTGAAATTTT TTAACCACAT AAATTATTGT GAGTCATCAG TCACAAAATG GGTGGGGATT CTCTGCAATA TTACACCTGC


AG
```

FILE 52RH3FI.NUM, RICE HISTONE H3 GENE, H3R-21, 14.12.1988, BY S.-C.WU

```
      10         20         30         40         50         60         70         80         90        100        110
TCTAGAGTGG TGACTGACCA CTGACCAGAG TGACCTGCAA CCTCGTCGAA GCAAGGCCAT CAATTCCCCT CCACCCACAT CCCCGCACTT GGCATGCATG CAATTCAGGG

     120        130        140        150        160        170        180        190        200        210        220
ATGACAGAAC GCAAGGCACG AAAATTGATA CCTCTCGGTG GGGGCATTCC GCCGAACACC ATCGGCGCAT TCCCTCTCTC TTTCCTCTTC CTCCTCTCTC GCCTCACAGC

     230        240        250        260        270        280        290        300        310        320        330
AGCAGGCGGG AAGGATCTTG CCGCCACGGC GCGCAAGGAG AGGGGTCGCG GTCGGTCACG CGTGCTAGCA ATCAGCCGCG GCGCGCACGC GGAGAGGAAT CGCTAGCCAC

     340        350        360        370        380        390        400        410        420        430        440
GGTGCGCGCC GAAGGAGGGG TCACCGGTTG CGGCACATAT GCGGGAAGGA GTTCGCTGGT AGCGTGCGTG CGTGCGTGAG GAGGAACTCG TCGTCTGGAC ATGCATGGGA

     450        460        470        480        490        500        510        520        530        540        550
AGGAGCTCGT CGTGTCCTCC CCCACCTTCT CCTGCCTTCC TCCCTTACCT TCCACGCCGA GCCCTAGTCA CCATCGGGCC TCGCCTCCTC GTGTGAGTCA GCTCACCGGG

     560        570        580        590        600        610        620        630        640        650        660
ACGACCATCG CCGCATCTGT GTGAACGACA CGAGCGTAGG GTGGAGAGGC GTGCGAACTC ACATGTGAAC TTTTTTTTTT CCATGTCTCA CGGACATGTG GGTCCACCGT

     670        680        690        700        710        720        730        740        750        760        770
ATATGGTTTT CCATATCACA CTAGCATGTG GGCCCCACCT GCAAAAAAAC CACTTTTCAA ACCACCAAAG GAGTCTATTT GTATGGGTTT TCGAAGATAG AGGATGCATT

     780        790        800        810        820        830        840        850        860        870        880
ATACTTGGTT TTGTAATTGA GGAATGTGAT TTAATCAAGA GCAAGGGATG AGGGAGGCCA GATATACTTA ATCCTCTAAA AAATGGCTGC CATCTTCCGA CTTATGGGCC

     890        900        910        920        930        940        950        960        970        980        990
GAGTAAATGG TCTGGGCTGG GCTGGGCTGG CAAGGGTGTA TGACTTTGAT GCTGAACTTC CAGTTACATA CTTCATTTAG CATTGAACTT AGTACTGTGG TTAATTACTT

    1000       1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
ATACCTAGAC CTAAGGTGAC ATCATATAGA GTTTTCTCTA ATTATCACGT AGGATTGGTG TAGTAACAAG ATAGTACATC AGTACATTCA TTTTATATGT ATACTGGTAT

    1110       1120       1130       1140       1150       1160       1170       1180       1190       1200       1210
ATATCTACGT GTTTGAAGTA CAACACATTA ACACCCATGT TTCCTATTAG CTTAAAAAAG AATATAAACT TCTGAACTGA CACCGATAGC AGTATTGGAA GATAAACACG

    1220       1230       1240       1250       1260       1270       1280       1290       1300       1310       1320
GAAGAAACCA TATAAAAGGT TCAAACATAA TGATGATACA AATATTTGTT AATATTGTTC GACCGAAGTC TATATTTACG AGAGTACTAA CCTTTTCCCA AACCGTAGCA

    1330       1340       1350       1360       1370       1380       1390       1400       1410       1420       1430
TAATTACAAC TGACGCGTGA GTTGTCTGGA TTCGAGGTTT CGCTGCGGCG GATTTGAGCT ATTGCATTGC GCGCGAGAGA AGCGATGCGA TCCGCGCGCT TCCGCATCGA

    1440       1450       1460       1470       1480       1490       1500       1510       1520       1530       1540
CCAATCACAG CGCAGAGATG GAACCCCACG GATCGCGATC CGCGTGATGC CCCCCGAAAA TCGCATCTCC ACCGTCCATT CCACTACAAT CCAATGGCAT ATATTTCGTG

    1550       1560       1570       1580       1590       1600       1610       1620       1630       1640       1650
CTTCCCTGCC GCATCGCTCG CGCGCCAAAA TTCCAAACCC CCAAAAAAAA AAAATCCCCG CGCGCTCTAA AATCCAAACC CACCCTCTCC CCTATATAAA CGCTGCCATC

    1660       1670       1680       1690       1700       1710       1720       1730       1740       1750       1760
TCCCTCGCCG CCTCATCCCA CTCCGCCGCC GCCGCCGCGA AATCTCCAAG CCAAAGCGCC GCCAGAGAGC CCCGAGGCCT CCTCCTCCCT CCTCCGCGAT GGCCCGCACG
                                                                                                    XX X
    1770       1780       1790       1800       1810       1820       1830       1840       1850       1860       1870
AAGCAGACGG CTCGGAAGTC GACCGGCGGG AAGGCGCCGA GGAAGCAGCT GGCGACGAAG GCGGCGCGGA AGTCGGCGCC GGCCACCGGC GGCGTGAAGA AGCCGCACCG

    1880       1890       1900       1910       1920       1930       1940       1950       1960       1970       1980
GTTCCGGCCG GGCACCGTCG CGCTCCGGGA GATCCGCAAG TACCAGAAGA GCACGGAGCT GCTGATCCGC AAGCTCCCCT TCCAGCGGCT CGTCCGGGAG ATCGCGCAGG
```

```
      1990       2000       2010       2020       2030       2040       2050       2060       2070       2080       2090
ACTTCAAGAC GGACCTCCGG TTCCAGAGCT CCGCCGTCGC CGCGCTCCAG GAGGCCGCCG AGGCCTACCT CGTCGGGCTC TTCGAGGACA CCAACCTCTG CGCCATCCAC

      2100       2110       2120       2130       2140       2150       2160       2170       2180       2190       2200
GCCAAGCGCG TCACCATCAT GCCCAAGGAC ATCCAGCTCG CCCGCCGCAT CCGCGGTGAG CGCGCTTAAA ATTAGTTCAA TAATCTGATC AAACATTCTG TTGACCAACA
                                                                    XXX
      2210       2220       2230       2240       2250       2260       2270       2280       2290       2300       2310
TCATCCCACT ATTTGAGCAG CCTGATGGAG AAGCATTAAT TGGTTGTTAA TGTTGTTCAA TTGCACGCGA CAACATTGTC AGTACTGAGA AAAATAAAAA TCCAAAATGG

      2320       2330       2340       2350       2360       2370       2380       2390       2400       2410       2420
AACAAAACGT TATAATAATG CATTTCTATA CCTGTAGTTT CCCTTTTAGA TGTTGAAATG AGCATGTTCA AATCTTATCA TTATAAGTTT GATCAACGGT CCATAAGAAA

      2430       2440       2450       2460       2470       2480       2490       2500       2510       2520       2530
GTGACCGTAC CTTAAAATTC CTCCATAATT AAAGTATCCC ACTACCCCAC TTACAAGCAA TTGTCCTAGC TTTGATGCTA AGTAGCTAAA CGTGATAAGC ATTCGCAGAA

      2540       2550       2560       2570       2580       2590       2600       2610       2620       2630       2640
TAAATTGTAA ACAATAAAAC AGGTAAATCT CCAACTGCCA AACAAGAGTA ATAGTCCACA GAAAGGTAAA ACTGCAAGAC CAAATCACGC CTGTCAAAAA AATTGGTTGG

      2650       2660       2670       2680       2690       2700
ATCAAATGAT CGAATAAGTG GATGTACAGG TATTCAGTCA GCGGCTTAAT ACAACTAATT A
```

FILE ALH3-11.NUM, ALFALFA HISTONE H3 GENE, ALH3-1.1, 03.02.1989, BY S.-C.WU

```
      10         20         30         40         50         60         70         80         90        100        110
GGATCCTCAT CACACAAACA AAACACATCC ACACGCCACG TCATCGATCC GCGTGTCGCA AATACTCCAA ATAAACGACA CCCGTCGATT AACCAATAAT CAACGGCCAC

     120        130        140        150        160        170        180        190        200        210        220
AATTACACCC CATTCACCCA CTTCTCAAAT TTCAAAACCC GCACAAAATC ATATCACTAT ATAACTCACC CCTTTCATCT CTTCTTCCTC ATCACTCTCA ATTCTTCAAA

     230        240        250        260        270        280        290        300        310        320        330
GCACAAAAAC AACCTTCAAG TTTCTCTGTT TGATACTGTT CTTTTCAGTT ATTTTTCATA ATCAATGGCA CGTACCAAGC AAACCGCTCG CAAATCCACC GGTGGCAAAG
                                                          XXX
     340        350        360        370        380        390        400        410        420        430        440
CTCCAAGGAA GCAACTCGCC ACAAAAGCCG CTCGCAAATC TGCTCCGGCC ACCGGCGGAG TGAAGAAACC TCACCGTTTC AGGCCAGGAA CCGTCGCTCT CCGTGAGATC

     450        460        470        480        490        500        510        520        530        540        550
CGCAAGTACC AGAAGAGCAC TGAGCTCCTC ATCCGCAAAC TCCCCTTCCA GCGTCTTGTC CGTGAGATCG CTCAGGATTT CAAGACTGAT CTCCGTTTCC AGAGCTCCGC

     560        570        580        590        600        610        620        630        640        650        660
CGTGTCGGCT TTGCAAGAAG CGGCCGAGGC TTATCTCGTC GGTCTCTTTG AGGATACTAA CCTCTGCGCC ATTCATGCTA AGCGTGTCAC TATCATGCCT AAGGATATCC

     670        680        690        700        710        720        730        740        750        760        770
AGCTCGCTAG GCGTATCCGT GGCGAGCGTG CTTGATCTTG TTGATTCGCT TTGTTAGGGT TTGTGTAGAT AGGTTCATGA TGTAGTTAAA TCACAAACCG TTGCTATAAG
                            XXX
     780        790        800        810        820        830        840        850        860        870        880
TTTCTCTATG GATTTTGTTA TATTGTAATG TGCTTAACGC TTAATCAATG AAATCGATCA TCTTTTGTTA AACTCTTTGT TCAATTACTT ATGCTTTTTT TTTATCTTTT

     890        900        910        920        930        940        950        960        970        980        990
CTTAACCCTA ATTTTCTGTC ATTTTATTAC ACTTTCCGAA CTTTTGTTAT CCCTAATTGG ATTTGAAATC AAAATTAGGG TTGATAAAGG CATATTGTAT AATGTTGAAA

    1000       1010       1020       1030       1040       1050       1060       1070       1080
TTCTTGTTAA ATATATTGAT AAATACAATT ATTTCAAAGT AAAAATTAAA TGCTGTTGAT TACATCTTGA AATGTAAAAA TAATTGCATG CATAAGCTT
```

FILE PH3C1.NUM, ALFALFA HISTONE H3 cDNA, pH3c-1, 03.02.1989, BY S.-C.WU

```
      10         20         30         40         50         60         70         80         90        100        110
CACCAAAACA ACCTTCAAAC TTTTCTGCTA GATACCGTTT TTTTCAGTTA TTTTTCAGCA TCAATGGCAC GTACCAAGCA AACCGCTCGC AAATCCACCG GTGGCAAAGC
                                                          XXX
     120        130        140        150        160        170        180        190        200        210        220
TCCAAGGAAG CAACTCGCCA CAAAAGCCGC TCGCAAATCT GCTCCGGCCA CCGGCGGAGT GAAGAAACCT CACCGTTTCA GGCCAGGAAC CGTCGCTCTC CGTGAGATCC

     230        240        250        260        270        280        290        300        310        320        330
GCAAGTACCA GAAGAGCACT GAGCTCCTCA TCCGCAAACT CCCCTTCCAG CGTCTTGTCC GTGAGATCGC TCAGGATTTC AAGACTGATC TCCGTTTCCA GAGCTCCGCC

     340        350        360        370        380        390        400        410        420        430        440
GTGTCGGCAC TGCAAGAAGC GGCCGAGGCT TATCTCGTCG GTCTCTTTGA GGATACTAAT CTCTGCGCCA TTCATGCTAA GCGTGTCACT ATCATGCCCA AGGATATCCA

     450        460        470        480        490        500        510        520        530        540        550
GCTCGCTAGG CGTATCCGTG GCGAGCGTGC TTGATCTTGT TGATTCGCTT TGTTAGGGTT TGTGTAGATA GGTTAATGAT GTAGTTAAAT CACAAACCGT TGCTATAAGT
                            XXX
     560        570        580        590        600        610        620        630
TTCTCTATGG ATCTTGTTAT ATTGTAATGT GCTTAACGCT TAATCAATGA AATCGATCAT CTTTTGTTAA AAAAAAAAAA AAA
```

FILE PH3C3.NUM, ALFALFA HISTONE H3 cDNA, pH3c-3, 03.02.1989, BY S.C-.WU

```
      10         20         30         40         50         60         70         80         90        100        110
GGCAAAGCTC CAAGGAAGCA ACTCGCCACA AAAGCCGCTC GCAAATCTGC TCCGGCCACC GGCGGAGTGA AGAAACCTCA CCGTTTCAGG CCTGGAACCG TCGCTCTCCG

     120        130        140        150        160        170        180        190        200        210        220
TGAGATCCGC AAGTACCAGA AGAGCACTGA GCTCCTCATC CGCAAACTCC CCTTCCAGCG TCTTGTCCGT GAGATCGCTC AGGATTTCAA GACTGATCTC CGTTTCCAGA

     230        240        250        260        270        280        290        300        310        320        330
GCTCCGCCGT GTCGGCACTG CAAGAAGCGG CCGAGGCTTA TCTCGTCGGT CTCTTTGAGG ATACTAACCT CTGCGCCATT CATGCCAAGC GTGTCACTAT CATGCCTAAG

     340        350        360        370        380        390        400        410        420        430        440
GATATCCAGC TCGCTAGGCG TATCCGTGGC GAGCGTGCTT GATCTTGTTG ATTCGCTTTG TTAGGGTTTG TGTAGATAGG TTCATGATGT AGTTATATCA CAAACCGTTG
                                 x xx
     450        460        470        480        490        500        510
CTATAAGTTT CTCTATGGAT CTTGTTATAT TGTAATGTGC TTAACGCTTA ATCAATGAAA TTAATCATCT TTT
```

FILE PH3C11, ALFALFA HISTONE H3 cDNA, pH3c-11, 03.02.1989, BY S.-C.WU

```
      10         20         30         40         50         60         70         80         90        100        110
CAAGGAAGCA GCTCGCCACC AAGGCTGCTA GGAAATCTGC TCCTACTACT GGAGGAGTCA AGAAACCTCA CCGATACCGC CCTGGAACTG TCGCTCTTCG TGAGATCCGT

     120        130        140        150        160        170        180        190        200        210        220
AAGTACCAGA AGAGTACCGA GCTTTTGATC CGCAAGCTTC CATTTCAGCG TCTTGTCCGT GAAATTGCTC AAGATTTCAA GACGGATCTG AGATTCCAGA GCCATGCAGT

     230        240        250        260        270        280        290        300        310        320        330
TCTTGCACTT CAGGAAGCAG CTGAGGCTTA CCTGGTTGGA TTGTTTGAGG ACACCAATCT GTGTGCAATT CATGCCAAGA GGGTGACAAT TATGCCTAAG GACATTCAGC

     340        350        360        370        380        390        400        410        420        430        440
TTGCTCGTCG CATTCGCGGT GAACGTGCTT AGGGTGGTGA AGGCGCTTTT AGCGTTATGG TGGATTAGTA TTTTGGAAGG ATTTAGGGTT TTATGAATTG AATTTTCTTT
                          x xx
     450        460        470        480        490        500        510        520        530        540        550
TATGCGTTGT ATAGTTCTGA ACCTATAATG TTCAATCTTT AACAACAGAC ATATTTTGGA TTATGATTAG TTTTTTGCGG ACAAATTTGT GATGTAATTG GTCAATTACA

     560        570        580        590
ATTGAAGTCT CTGCAACTAT TTTACTTATA AAAAAAAAAA AAAA
```

FILE PH3C12.NUM, ALFALFA HISTONE H3 cDNA, pH3c-12, 03.02.1989, BY S.-C.WU

```
      10         20         30         40         50         60         70         80         90        100        110
TTCAAGACTG ATCTCCGTTT CCAGAGCTCC GCCGTGTCGG CACTGCAAGA AGCGGCCGAG GCTTATCTCG TCGGTCTCTT TGAGGATACT AACCTCTGCG CCATTCATGC

     120        130        140        150        160        170        180        190        200        210        220
CAAGCGTGTC ACTATCATGC CTAAGGATAT CCAGCTCGCT AGGCGTATCC GTGGCGAGCG TGCTTGATCT TGTTGATTCG CTTTGTTAGG GTTTGTGTAG ATAGGTTCAT
                                                           xxx
     230        240        250        260        270        280        290        300        310        320        330
GATGTAGTTA TATCACAAAC CGTTGCTATA AGTTTCTCTA TGGATCTTGT TATATTGTAA TGTGCTTAAC GCTTAATCAA TGAAATTAAT CATCTTTTGT TTAAAAAAAA

AAAAAAA
```