DNA Sequence—The Journal of Sequencing and Mapping, Vol. 6, pp. 137–146
Reprints available directly from the publisher
Photocopying permitted by license only

© 1996 OPA (Overseas Publishers Association)
Amsterdam B.V. Published in The Netherlands by
Harwood Academic Publishers GmbH
Printed in Malaysia

Identification of three highly expressed replacement histone H3 genes of alfalfa

A.J. ROBERTSON, T. KAPROS, D. DUDITS and J.H. WATERBORG

Division of Cell Biology and Biophysics, University of Missouri-Kansas City, 5100 Rockhill Road, Kansas City MO 64110-2499, USA
Institute of Plant Biology, Biological Research Center, Hungarian Academy of Sciences, H-6701 Szeged, P.O. box 521, Hungary

Database Accession No. U09460, U09645

One genomic and six cDNA clones for the replacement histone H3.2 protein of alfalfa (*Medicago sativa*) were isolated and sequenced. By gene organization they represent 3 distinct genes. PCR methods were used to confirm that only three intron-bearing histone H3.2 genes of this type exist per haploid genome. They co-exist with approximately 56 copies of the previously characterized replication-dependent, intronless histone H3.1 variant gene. Comparison of the relative expression of few constitutive H3.2 genes with the high S phase expression of the abundant cell cycle-dependent H3.1 genes by mRNA levels and protein synthesis measurements revealed that the replacement histone H3.2 genes are very highly expressed. Structural analysis of the genomic replacement H3.2 gene revealed a unique feature. A repeated polypyrimidine sequence motif in the 5′ untranslated region of this gene replaces the ubiquitous intron present in all known replacement H3 genes. A hypothesis is presented that this motif and other, non-randomly distributed polypyrimidine sequences in the introns of replacement histone H3 genes of alfalfa and *Arabidopsis*, may affect nucleosome assembly. Chromatin repression of these replacement genes would be avoided, consistent with the high, constitutive expression of replacement H3 histone genes in plants.

KEY WORDS: Alfalfa, histone H3, intron, polypyrimidine, replacement histone.

INTRODUCTION

Histone H3 is the central protein within the histone octamer that organizes DNA into nucleosomes and, with linker histones, into chromatin (Van Holde, 1989). Study of chromatin in terminally differentiated animal cells has revealed that a minor, cell cycle-independent histone H3.3 variant progressively replaces histone H3 proteins that were produced in a replication-dependent manner in cycling cells (Zweidler, 1980). This replacement occurs preferentially in transcriptionally active chromatin (Ridsdale and Davie, 1987; Hendzel and Davie, 1990). These observations have suggested that transcription of chromatin causes loss of nucleosomes, that histone H3 from displaced nucleosomes is not reused to restore chromatin packaging, and that newly synthesized histone H3.3 supports formation of nucleosomes and restoration of nucleosomal density of chromatin in the absence of replication.

Recently, a distinct histone H3 variant has been identified in alfalfa and several other dicot and monocot plants that is characterized by a high level of post-synthetic acetylation (Waterborg *et al.*, 1989; Waterborg, 1992) and a cell cycle-independent pattern of expression (Kapros *et al.*, 1992). A study of the synthesis and stability of this H3.2 histone variant in alfalfa has demonstrated that it is a functional replacement histone, equivalent to histone H3.3 in animals (Waterborg, 1993). In plants, however, some aspects of the mechanism to maintain chromatin packaging appear distinct. The steady-state level of replacement histone H3 protein in plant chromatin is much higher than in animal cells, especially comparing growing cell populations (Waterborg, 1991, 1992). Consistent with this is the high rate and extent of histone H3.2 turnover which has suggested that loss of nucleosomes dur-

Address for correspondence: Dr. J.H. Waterborg, Division of Cell Biology and Biophysics, University of Missouri-Kansas City, 5100 Rockhill Road, Room 414 BSB, Kansas City, MO 64110-2499, U.S.A.

ing transcription of chromatin may occur at a higher frequency in plants than in animals (Waterborg, 1993). To maintain the steady-state level of 30 to 40 percent replacement histone H3.2 protein in asynchronously growing cell populations (Waterborg *et al.*, 1989; Waterborg, 1990), alfalfa cells must synthesize the H3.2 protein at a very high rate.

In this paper, we describe the sequence analysis of three intron-bearing histone H3.2 genes which produce twice as much histone H3.2 protein (Waterborg, 1993) as the more than 50 other histone H3 genes, previously detected in alfalfa (Wu *et al.*, 1988). We confirm by PCR analysis the high number of replication-dependent, intronless H3.1 genes in alfalfa, and show that the steady-state ratio between the variant mRNAs directly determines the relative rate of *de novo* histone H3 variant synthesis. Finally, we present a comparison between the histone H3 variant genes in alfalfa and *Arabidopsis*. This analysis suggests the testable hypothesis that plant replacement histone H3 genes may have the observed high basal levels of expression because they are less repressed by chromatin.

RESULTS

Isolation and analysis of alfalfa histone H3.2 cDNA sequences

The two histone H3 protein variants of alfalfa (Waterborg *et al.*, 1989; Waterborg, 1992) differ only by 4 amino acids (Waterborg, 1990). This results in cross-hybridization between the coding sequences of the single partial cDNA clone for histone variant H3.2, pH3c11, isolated prior to the start of this project, with all histone H3.1 cDNA clones, isolated from the same library, named LUCA (Wu *et al.*, 1989). However, the 3′UTR (untranslated region) sequences of H3.1 clone pH3c1 and H3.2 clone pH3c11 provide histone H3 variant-specific hybridization probes (Kapros *et al.*, 1992). By using both probes to rescreen the LUCA library, we identified, among 100,000 transformants, 66 positives for H3.1 and 33 positives for H3.2. This ratio resembles closely the relative level of the histone H3 variant proteins (Waterborg, 1990, 1992). Plasmid DNA was prepared from all positives for H3.2 and the DNA sequences for the six clones with the longest inserts were determined by double-stranded sequencing with primer walking along both DNA strands. These results with identified or putative sequence elements are shown in figure 1.

Including the previously isolated cDNA, pH3c11, three distinct classes could be identified on the basis of 3′UTR sequence repeat differences, supported by single residue heterogeneity at 7 sites. None of these differences affected the histone H3.2 protein sequence. The only heterogeneity within classes was the location of polyadenylation. Plasmid pH3c11 is the only representative of class I. It used the previously identified AAUUGAA polyadenylation signal sequence at 188 bp beyond the stop codon (Wu *et al.*, 1989). Class II plasmids pH3c110 (GenBank accession number U09460), pH3c118 (U09461), pH3c126 (U09462) and pH3c130 (U09464) are characterized by a direct 12 bp repeat at 167 bp beyond the stop codon. Class III plasmids pH3c127 (U09463) and pH3c131 (U09465) contain the result of an apparent recombinational event that did not change the length of the 3′UTR but that resulted in a repeat of 3′UTR sequences 99–108 and 135–148 as 124–133 and 110–123, respectively. This event created a direct 25 bp tandem repeat between sequence positions 1444 and 1493. Plasmids pH3c126 of class II and pH3c131 of class III start their polyadenylation prior to the signal sequence used by all other H3.2 cDNAs. They may have used a second plant polyadenylation consensus sequence, GAUG-AA (Wu *et al.*, 1989; Mikami and Iwabuchi, 1993), located at or just upstream of the start of polyadenylation. Alternatively, an 8 bp inverted repeat sequence that might form a GT-hyphenated hairpin structure of unknown significance, might be involved (Fig. 1).

The 5′UTR sequences of class II pHc110 (55 bp) and pH3c118 (41 bp) clones contained a direct 17 bp repeat in a GAG-interrupted series of polypyrimidine sequences. The presence of these elements could not be confirmed for class I and III histone H3.2 genes because pH3c11 plasmid was incomplete (Wu *et al.*, 1989), and because the longest analyzed class III clone, pH3c131, contains only 18 bp upstream sequence (Fig. 1).

Isolation and analysis of one alfalfa H3.2 gene

Screening of 60,000 recombinants in a genomic library of *M. sativa* cv. Chief with histone H3 variant-specific hybridization probes (Kapros *et al.*, 1992) yielded 30 positive plaques. Direct phage DNA sequencing of 10 isolates with a sense primer, common between known H3.2 and H3.1 sequences (nucleotides 963–982 in Figure 1) (Wu *et al.*, 1988, 1989), identified a majority by codons 87 and 90 as histone H3.1 variant genes (Waterborg, 1990).

```
(-482) ac ttnactaacg gagtctgcat ttaggtacta aaatgactaa tatagtctac attcagggac tattttgcaa tttacctgca ttcagggact aaagtgacga cttctttcct attcagagac 122
msH3g1
                CCAAT box                                                                                   AC-box              nonamer CAATC-- CAACgyy
   (-360) taaagtgaCC AATctctcaa aatgaagata ttttgttttg ttttggcgag ataagttgca cacgatttac actcacaaaa gaaacacaaa ttgtccacgc tggCAATCcg CAACtTTaca 242

              CCAAT box   AC-box  CGCGGATC octamer
   (-240) aaccaaCCAA Tcagaaacaa acaCaCGGAT Cgcacttaat attttcactt aaaaaactca tcattaccgt tgaagcattc aaagtccacg atctctctac tctaattaac cttccttaat 322

                       TATA box  AC-box                                                                    cap site                                 (-1)
   (-120) catcattaac ccttgcaTAT ATAAAcacac ttctcttcaa caaccctcat tacacatttc ttctctttcg ctaaatctaa TCAATCTttc cctctcttcg agctttctct ctccgattcc 482
5'UTR:                                                                           class II   : pH3c110     aatctttc cctctcttcg agctttctct ctc-┐
                                                                                                                         └ ttcg agctttctct ctccgattcc
                                                                                 class III  : pH3c131                                 ctttctct ctccgattcc

EXON 1:   ATG GCT CGT ACC AAG CAA ACC GCT CGC AAA TCC ACT GGT GGT AAG GCT CCA AGG AAG CAG CTC GCC ACC AAG 554
      I                                                                       ... ... ... ... ... ... ...
     II   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
    III   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
    (0)    M   A   R   T   K   Q   T   A   R   K   S   T   G   G   K   A   P   R   K   Q   L   A   T   K  (23)

intron 1: GT aaccaccg ttcaccgccg taacggtttt ttcttcttta tgttttcttg atcttagggt ttcgttttct tcaattcgaa tttttttgatt gatttcatcg atttttttggt tc AG  668
                                                                                                                                        (114)
EXON 2:   GCT GCT AGG AAA TCT GCT CCT ACT ACT GGA GGA GTC AAG AAA CCT CAC CGA TAC CGC CCT GGA ACT GTC GCT CTT CG 745
      I   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..t ... ... ... ... ..
     II   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..c ... ... ... ... ..
    III   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..t ... ... ... ... ..
   (24)    A   A   R   K   S   A   P   T   T   G   G   V   K   K   P   H   R   Y   R   P   G   T   V   A   L   R (49)

intron 2: GT aatttcct tttccccaat ttttaggttt tcggagtttt gcagtttcta ttattaattt tttttaggtt ttcgttgtgt tttgaatatt ctattgaatt ttatgttnng          855
                     aatttgaatt ttggattcat aatttttagg aatttggagt tttgtattct cggtttatga tttttaggtt ttcgtagttg ttaattttca attgttgtga tttac AG 962
                                                                                                                                     (217)
EXON 3:    T GAG ATC CGT AAG TAC CAG AAG AGT ACC GAG CTT TTG ATC CGC AAG CTT CCA TTT CAG CGT CTT GTC CGT GAA ATT GCT CAA GAT TTC AAG 1053
      I    . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..t ... ... ... ... ...
     II    . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..t ... ... ... ... ...
    III    . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..c ... ... ... ... ...
   (50)      E   I   R   K   Y   Q   K   S   T   E   L   L   I   R   K   L   P   F   Q   R   L   V   R   E   I   A   Q   D   F   K  (79)

intron 3: GT aaatattg ttacttagtt tcataattga ttttgtgtga aatcttgttc ctctttggtt tatttaattt gtaattgttg ttttttgatat ttttctgatt aatttgctgt tgttgttat AG 1174
                                                                                                                                                (121)
EXON 4:   ACG GAT CTG AGA TTC CAG AGC CAT GCA GTT CTT GCA CTT CAG GAA GCA GCT GAG GCT TAC CTG GTT GGA TTG TTT GAG GAC ACC AAT 1261
      I   ... ... c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
     II   ... ... t.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
    III   ... ... t.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
   (80)    T   D   L   R   F   Q   S   H   A   V   L   A   L   Q   E   A   A   E   A   Y   L   V   G   L   F   E   D   T   N  (108)

          CTG TGT GCA ATT CAT GCT AAG AGG GTG ACA ATT ATG CCT AAG GAC ATT CAG CTT GCT CGT CGC ATT CGC GGT GAA CGT GCT TAG 1345
      I   ... ... ... ... ... ..c ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
     II   ... ... ... ... ... ..t ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
    III   ... ... ... ... ... ..t ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
  (109)    L   C   A   I   H   A   K   R   V   T   I   M   P   K   D   I   Q   L   A   R   R   I   R   G   E   R   A  stp (136)

3'UTR:(+1) ggtggtgaag gcgcttttag cgttatggtg gattagtatt ttggaaggat ttagggtttt atgaattgaa ttttctttta tgcgttgtat agttctgaac ctataatgtt 1455
      I    .......... .......... c......... .......... .......... .......... .......... t......... .......... .......... ..........
     II    .......... .......... g......... .......... .......... .......... .......... c......... .......... .......... .........
    III    .......... .......... g......... .......... .......... .......... .......... t......... .......... .......... ........gt (+110)

                                                                                      GAUG-AA          AAUUGAA
                                                                                         >>>>>>>>() <<<<<<<<
           caatctttaa caacagacat attttggatt atgattagtt ttttgcggac aaatttg             tga tgtaattggt caattacaat tgaagtctct gcaactattt tacttatatc 1565
      I    .......... .......... .......... t......... .......... .......             ... .......... .......... .......... .......... ......■   (+217)
     II    .......... .......... .......... c......... .......... ....... cggacaaatttg ... .....■.... .......... .......... .......... ....■     (+235)
    III    ggattatgat tagacctata atgg...... t......... .......... .......             ... .......... ....■..... .......... .......... ....■     (+215)

           tccattgctt cctgatttcg ttatgcgctt ttgatgatgc gactgtggtt tctagctctg aattcaattt gtgatgcgtt tctctctatg gtagtttgtc tatcgctgac a  1676
```

Figure 1 Alfalfa replacement histone H3.2 gene sequences. Genomic clone msH3g1 sequence is shown aligned with Class I cDNA clone pH3c11, with class II cDNA clone pH3c110 and with class III cDNA clone pH3c131. All invariant nucleotide positions are shown by dots. In addition to the msH3g1 nucleotide numbering (1 to 1676), reverse numbering of 5′ upstream sequences (–482 to –1), numbering of 3′ downstream nucleotides (+1 to +217, +235 and +215 for classes I, II and III, respectively), and numbering of the codons (**0** to **136**) from start to stop codon (***stp***) with single letter amino acid codes. The amino acids residues that differ among alfalfa histone H3 expression variants are double underlined. In the upstream sequence, putative CCAAT boxes, consensus nonamer and octamer sequences, TATA box and cap site are shown as capitals, AC boxes are underlined and polypyrimidine sequences are shown throughout the sequences by bold-italics. The 5′UTR 17 bp repeat of class II is graphically marked. 5′ and 3′ splice boundaries of all introns are marked by capitalized consensus GT and AG signals, respectively. The putative PyNPyPyPuAPy branch point consensus sequences of all introns are shown by single and double underlining. The downstream sequences, repeated between 1454 and 1478 in class III, are marked in msH3g1 and pH3c131 by single and double underlining. The downstream 12 bp tandem repeat in class II is marked in pH3c1 and pH3c110 by underlined italics. Consensus polyadenylation signals are shown in bold-italic capitals above the msH3g1 sequence. The last mRNA nucleotide, prior to the polyA sequence in the cDNA clones, is marked by a solid square. The putative GT-hyphenated stem-loop structure is marked by the > > 0 < < symbol.

Isolate msH3g1 (GenBank accession number U09458) produced a histone H3.2 variant-specific sequence at these codons. Combination of the sense sequencing primer with an anti-sense primer (nucleotides 1520–1500 in Fig. 1), specific for the 3'UTR of H3.2 variant sequences (Wu *et al.*, 1989), produced a 558 bp PCR product only with msH3g1 DNA, predicting the presence of an intron in this H3.2 gene. Primer walking confirmed the presence of this and two additional introns—after codon 23, in codon 49 and after codon 79—and produced 0.5 kb upstream and 0.3 kb downstream sequence (Fig. 1). All introns contain consensus splice junctions with PyNPyPyPuAPy higher eukaryote branch-site consensus sequences (Lewin, 1994), upstream of the 3' junctions. Comparison with the three cDNA classes identified the msH3g1 H3.2 gene as class I, only differing at wobble positions of codons 44 and 114 between Chief and Regen S. alfalfa cultivars. A search for putative regulatory consensus sequences (Tabata *et al.*, 1987; Dalton and Wells, 1988; Chaboute *et al.*, 1993; Ohtsubo *et al.*, 1993) revealed a typical regulatory plant histone nonamer and octamer sequence, two CCAAT boxes, three AC boxes, typical of histones genes, and a consensus TATA box and transcription cap site, 103 and 40 bp upstream of the start codon, respectively (Fig. 1). The 5' terminus of pH3c110 coincided with the preferred start position in the cap site, and, thus, pH3c110 is a complete cDNA.

Determination of the number of histone H3 variant genes in alfalfa

Combinations of sequencing primers were utilized in PCR reactions to determine the number of histone H3.2 genes in the alfalfa genome. Genomic DNA produced 3 PCR products of distinct size but nearly equal amounts (Fig. 2A) with primers between sequence positions −184 and +33, relative to the start of translation (Fig. 1). The intermediate size product was identical to that produced from class I genomic clone msH3g1, as predicted (216 bp). The larger size product matches a class II gene, longer by its characteristic 17 bp 5'UTR repeat. The 190 bp PCR product suggests that the third histone H3.2 gene, assumed to generate class III transcripts, has a shorter upstream sequence. Whether this indicates a divergent 5'UTR or upstream sequence or a defined deletion, relative to msH3g1, awaits sequencing of a class III genomic clone. These PCR results also demonstrated that the short 5'UTR of these three histone H3.2 genes of alfalfa is devoid of introns.

This analysis of histone H3.2 copy number is independent of the estimated size of the alfalfa genome. However, it might miss H3.2 genes with divergent sequences at the upstream PCR primer site. Therefore, a second PCR method was used to confirm the presence of 3 histone H3.2 genes. A coding region primer at sequence position 496, common for all histone H3 clones, was matched with an antisense primer at position 1520, specific for the 3'UTR of all H3.2 sequences (Fig. 1). All three introns of genomic clone msH3g1 are located between these primers. The intron-bearing msH3g1 DNA and intronless pH3c110 cDNA produced the expected products of 1024 and 572 bp, respectively (Fig. 2B). Genomic DNA gave a single band of 1 kbp. This showed that all alfalfa H3.2 genes have introns, identical to msH3g1. Addition of increasing amounts of the homologous pH3c110 DNA as competitor in known molar ratios, assuming a haploid genome size for alfalfa of 1.2 pg (Winicov *et al.*, 1988), decreased the amount of genomic PCR product. Equimolar amounts of PCR product, based on ethidium bromide fluorescence with correction for size differences, was obtained when 3 plasmid DNA molecules were added per complete haploid alfalfa genome (Fig. 2B). This confirmed that the initial, genome size-independent PCR method had not missed any histone H3.2 genes.

Quantitative Southern analysis has been used in the past to estimate that one haploid alfalfa genome may contain 40 histone H3 genes, assuming a genome size of 1.7 pg (Wu *et al.*, 1988). This is equivalent to 55 copies, using our genome size value (Winicov *et al.*, 1988). This estimate was confirmed by competitive PCR with a sense primer that started in codon 49 of all H3 sequences and an antisense primer, specific for the 3'UTR of histone H3.1 clone pH3c1 and all known genomic and cDNA H3.1 sequences (Wu *et al.*, 1988, 1989). pH3c1Δ65 DNA, a derivative of pH3c1 with an internal 65 bp deletion, was designed to produce a 272 bp PCR product. It was used in 24 to 379 fold molar excess as a competitor for histone H3.1 variant genes in genomic DNA, which gave the predicted 337 bp product. Five independent titration experiments showed by equimolar cross-over point calculations that alfalfa contains 56 ± 6 H3.1 genes per haploid genome (Fig. 3). None of these contained introns.

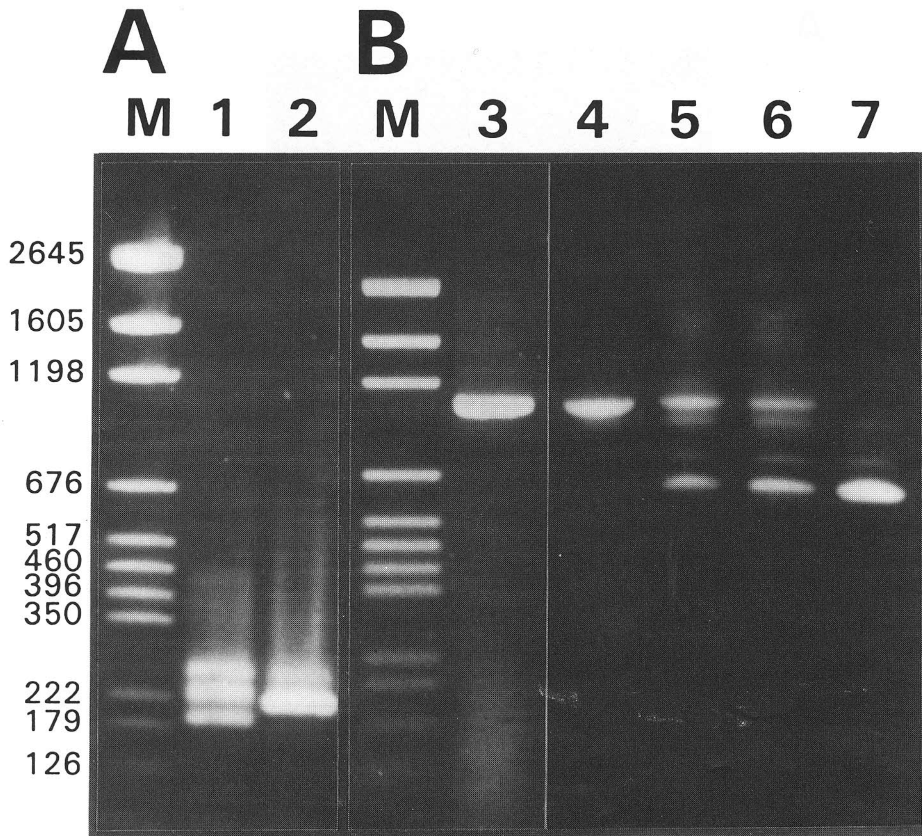

Figure 2 PCR-based determination of the number of histone H3.2 genes in alfalfa. Agarose gel electrophoretic separation of PCR products, in parallel to marker DNA fragments (M), visualized by ethidium bromide fluorescence and quantitated by densitometry of Polaroid negatives. The size of the DNA size markers, in bp, is shown along the side. Template DNAs were cloned msH3g1 DNA (lanes 2 and 3) and genomic DNA of *M. sativa* cv. Chief (lanes 1 and 4) alone, or with pH3c110 plasmid DNA as internal competitor in a molar ratio of 3 (lane 5), 5 (lane 6) and 50 (lane 7). Primer pairs were between nucleotide positions 299 and 515 (panel A) and between 496 and 1520 (panel B), as numbered on msH3g1 in Fig. 1.

Comparison of histone H3 mRNA levels and protein synthesis rates

Previous studies have revealed that expression of the histone H3.2 genes is cell cycle-independent, as determined by relatively constant mRNA levels in partially synchronized alfalfa suspension cultures (Kapros *et al.*, 1992). Moreover, the rate of H3.2 protein synthesis is the same in exponentially growing cultures and in cells, arrested in S phase by hydroxyurea treatment (Fig. 4) (Kapros *et al.*, 1995). In contrast, histone H3.1 mRNA levels and protein synthesis rates are high only in S phase cells (Kapros *et al.*, 1992, 1995). The rate of histone H3.1 protein synthesis in growing cultures increases from 40%, relative to the stable rate of H3.2 protein synthesis, to 220% when increasing numbers of alfalfa cells are collected and arrested in S phase by hydroxyurea (Fig. 4). It should be noted that the response of plant cells to treatment with hydroxyurea is different from that in animals. Inhibition of DNA replication does not lead to the rapid destruction of existing histone mRNA and cessation of histone protein synthesis which is seen in animal cells. Plant histone transcripts lack the 3′UTR stem-loop structure of cell cycle-regulated animal histones that is responsible for this feedback response (Kapros *et al.*, 1995).

Northern analysis of the histone H3 variant genes has demonstrated for each of the variant H3 histone mRNAs individually that changes in mRNA levels

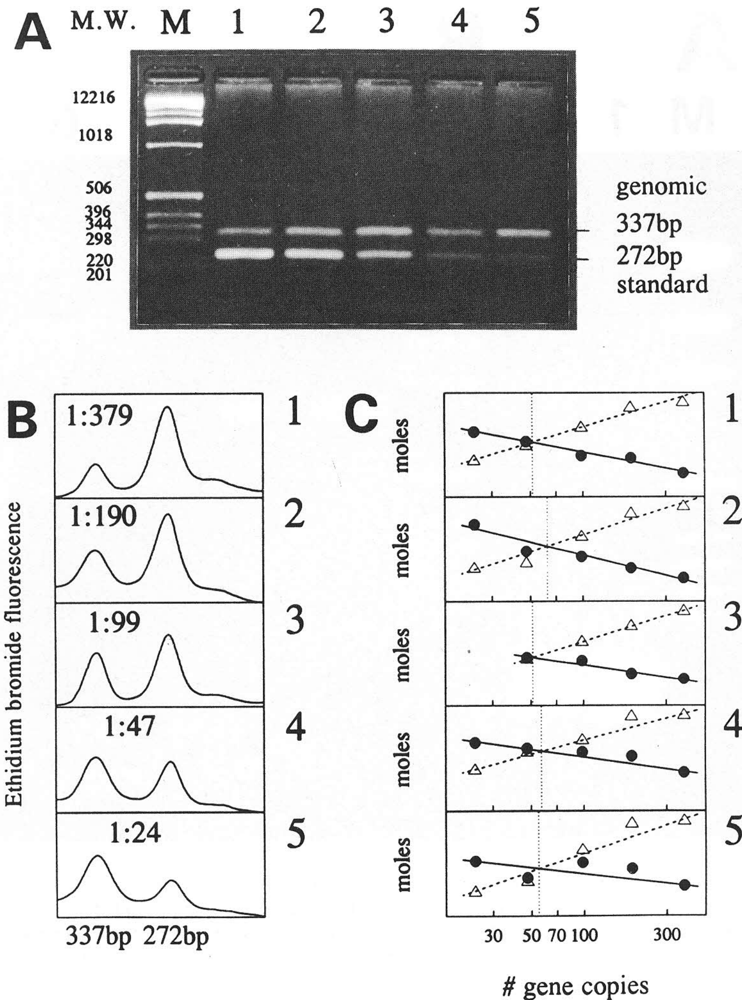

Figure 3 PCR-based titration of histone H3.1 genes in alfalfa. Panel A. Agarose gel electrophoretic separation of PCR products, visualized by ethidium bromide. The molecular weight (M.W.) of the markers in lane M are shown along the side in bp. Competitive DNA templates were genomic alfalfa DNA and pH3c1Δ65 plasmid DNA in the molar ratios for lanes 1 to 5, shown in panel B in subpanels 1 to 5. The predicted size of genomic and competitor DNA ('standard') for the set of histone H3.1 variant-specific primers is shown. Panel B. Densitometric analysis of Polaroid negatives for lanes 1–5 of panel A. Panel C. Size-corrected ethidium bromide fluorescence of 337 bp (line with solid circles) and 227 (broken line with triangles) PCR products were plotted on an arbitrary scale (moles) *versus* the log of the molar ratio of the template DNAs. Five independent experiments are shown. The point of equimolarity is marked for each experiment by a dotted line.

can occur when cells progress throught the cell cycle in synchronized cultures or when growth conditions are changed (Kapros et al., 1992, 1995). So far, differences in hybridization probe efficiencies for the two gene types has prevented a comparison of the absolute mRNA levels for the two proteins. Inclusion of known amounts of *in vitro* transcripts of both genes in Northern analysis has made it possible to determine the abolute levels of both mRNA types and to correlate these with the observed rates of protein synthesis. In hydroxyurea-treated cultures, when both histone H3 variant gene types are expressed (Kapros et al., 1992, 1995), the steady-state level of H3.1 mRNA was 90 ± 12 pg (n=4) per μg of total cellular RNA. The level of H3.2 mRNA was 65 ± 14 pg (n=4). Thus, histone H3.1 mRNA represented 59 ± 4 % (n=4) of total H3 mRNA when newly synthesized histone H3.1 protein was 60 ± 7% (n=3) of total *de novo* histone H3 synthesis (Fig. 4B). The relative level of mRNA, therefore, directly determined the relative rate of protein synthesis for the histone H3 variant genes.

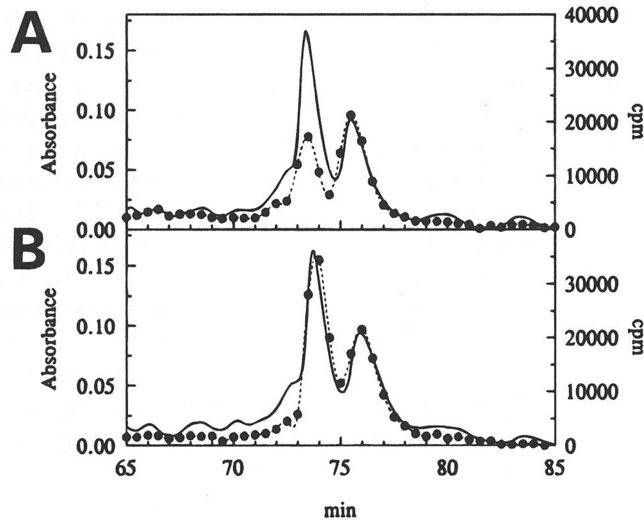

Figure 4 Histone H3 variant protein synthesis. Reversed-phase HPLC elution of histone H3.1 (near 74 min) and H3.2 (near 76 min) variant proteins, prepared from alfalfa A2 cell suspension cultures in exponential growth (panel A) or from cultures treated with 10 mM hydroxyurea for 24 h (panel B) and incubated with tritiated lysine, is shown by absorbance of the eluate at 214 nm and the radioactivity, determined by liquid scintillation counting, as cpm per 0.5 ml fraction.

DISCUSSION

Expression of H3.1 and H3.2 genes compared

The measured ratio between H3.1 *versus* H3.2 variant genes of 56:3 in alfalfa is remarkably different from the observed relative amounts of steady-state H3 variant proteins of 2:1 (Waterborg *et al.*, 1989; Waterborg, 1990), and even more different from the relative ratio of 1:2 for *de novo* H3 variant protein synthesis in exponentially growing alfalfa suspension cells (Fig. 4A) (Waterborg, 1993). To date, one factor, the difference in cell cycle-dependent expressions of the H3 variants genes, has been identified as contributing to this difference. The steady-state level of replacement histone H3.2 mRNA has been shown to be fairly constant and independent of cell cycle phases (Kapros *et al.*, 1992, 1995). In contrast, the level of H3.1 mRNA rises in S phase cells (Kapros *et al.*, 1992) and drops sharply upon completion of DNA replication (Kapros *et al.*, 1993), characteristic of a replication-dependent histone. Considering the time required for alfalfa cells to progress through S phase, relative to the total length of the cell cycle (Kapros *et al.*, 1992), one would predict a three-fold rise, approximately, in the level of histone H3.1 mRNA and in the rate of H3.1 protein synthesis if all cells were in S phase. Cell cycle arrest of alfalfa A2 suspension cells with hydroxyurea approaches this condition (Kapros *et al.*, 1995) and was used to test this prediction. Relative to the level of histone H3.2 protein synthesis (Fig. 4), the level of H3.1 synthesis rose 2.2-fold. The level of H3.1 mRNA and the rate of H3.1 protein synthesis both became 60 percent of the total amount of histone H3 mRNA and of the total rate of histone H3 protein synthesis (Fig. 4), respectively. Under these conditions, when both variant gene types are expressed, it is possible to compare their relative levels of expression. The 56 active replication-dependent H3.1 genes in S phase cells make 60 percent of the total H3 mRNA and new protein. Only 3 constitutive H3.2 genes produce 40 percent, i.e. they appear to be more highly expressed by a factor 14, on a 'per gene' basis. This does not translate directly into a difference in the rate of gene expression. Analysis of mRNA stability of histone H3 variant transcripts has shown that, depending on culture conditions and when measured under conditions of inhibition of transcription by actinomycin D, histone H3.2 mRNA is 3- to 5-fold more stable (Kapros *et al.*, 1995). This suggests that the transcription rate of the 3 H3.2 genes is, on average, 5- to 3-fold higher than that of the 56 H3.1 genes during their active state of transcription in S phase cells. We have no information yet on the possibility that a fraction of the H3.1 genes are not transcribed, or only inefficiently. However, even should this be the case, our experimental evidence strongly suggests that the rate of transcription of the 3 replacement histone H3 genes of alfalfa is high.

Arabidopsis thaliana contains 6 histone H3 genes. Four histone H3 genes, represented by two histone H3.1 genomic clones (Chaboute *et al.*, 1987; Chaubet *et al.*, 1987), produce a single histone H3 protein, different only at residue 90 from alfalfa H3.1 (Waterborg, 1992), in a meristem-dependent and presumably replication-dependent pattern of expression (Chaboute *et al.*, 1993). Two tandemly linked and intron-bearing histone H3-III genes (Chaubet *et al.*, 1992) produce a histone H3 protein (Waterborg, 1992), identical to alfalfa H3.2 (Waterborg, 1990, 1992) in a meristem-independent pattern of expression (Chaubet *et al.*, 1992). This apparent replacement histone H3 is also highly acetylated (Waterborg, 1992), supportive of its preferential assembly into replacement nucleosomes within transcriptionally active chromatin (Waterborg, 1993). The steady-state level of the *Arabidopsis* replacement histone H3 is 55 percent of total histone H3 in proliferating cells (Waterborg, 1992). This, compared with the gene numbers of replication-dependent and -independent histone H3 genes, suggests that the constitutive level of expression of the replacement histone H3 genes in *Arabidopsis* is high, as in alfalfa.

Comparisons among replacement histone H3 genes

Replacement histone H3 genes have been characterized by two gene features, a cell cycle-independent pattern of expression and the presence of introns (Van Holde, 1989; Wells *et al.*, 1989). Differences in transcript polyadenylation, observed between replacement and replication-dependent histone types in animals, are not characteristic for plants. All plant histone genes, irrespective of their pattern of gene expression, produce polyadenylated mRNA (Chaboute *et al.*, 1988, 1993; Wu *et al.*, 1989).

Comparison of intron locations within known replacement histone H3.3 genes of animals (Table 1) with those in alfalfa and *Arabidopsis*, the only plant species from which genomic replacement histone H3 genes have been cloned (Chaubet *et al.*, 1992) (Fig. 1), suggests a large evolutionary distance between these groups of genes and supports the notion that each arose independently (Thatcher *et al.*, 1994). A comparison among the three plant replacement H3 genes (Table 1) suggests that the initial gain of 4 introns in the primordial plant replacement gene has been followed by instances of loss, e.g., of the 5′UTR intron in the evolution leading to alfalfa.

The ubiquitous occurrence of an intron in the 5′URT of all other known replacement H3 genes, including animals, suggests that it provides an essential function. By inference, one would expect to see such a function retained in the alfalfa msH3g1 gene, despite the loss of the 5′UTR intron. A distinctive sequence feature of the very short 5′UTR of the alfalfa H3 replacement genes is an abundance of polypyrimidine sequences (Fig. 1).

Repetitive polypyrimidine sequences have the potential to form intramolecular triplex DNA structures (Maher III, 1992) and these could potentially affect or regulate gene expression (Maher III, 1992; Sarkar and Brahmachari, 1992; Firulli *et al.*, 1994). However, attempts to detect the occurrence of triplex DNA in the full-length 5′UTR of pH3c110 cDNA plasmid by digestion of the excluded fourth DNA strand by single-strand-specific nucleases has consistently failed (results not shown).

Polypyrimidine sequences are found at a higher than statistical frequency in the coding strand of the msH3g1 gene, exclusively in non-coding sequences (Fig. 1), and they are absent in the non-coding strand. The same is true for the two *Arabidopsis* replacement histone H3 genes (Chaubet *et al.*, 1992). This localization and asymmetry is remarkable in the light of the existence of ubiquitous polypyrimidine-binding proteins (Kolluri *et al.*, 1992; Desjardins and Hay, 1993; Morris *et al.*, 1993), such as GAGA factor in *Drosophila melanogaster* (Lu *et al.*, 1993), that

Table 1 Intron locations in replacement histone H3 genes

Species	*H3 gene*	*Intron locations*[a]				*GenBank*
Homo sapiens (man)	H3.3	5′UTR		42(3)	94(1)	M11353, X05854
Gallus gallus (chicken)	H3.3	5′UTR		42(3)	94(1)	M11392, M11393
Drosophila melanogaster	H3.3	5′UTR			94(1)	X81207
Drosophila hydei	H3.3	5′UTR			94(1)	X81208
Medicago sativa	H3.2		24(1)	49(3)	80(1)	U09458
Arabidopsis thaliana	H3-III #1	5′UTR	24(2)	49(3)	80(1)	X60429
Arabidopsis thaliana	H3-III #2	5′UTR	24(2)	49(3)		X60429

[a]The information of intron positions was obtained from GenBank sequences, which are identified by their accession numbers. The intron position within coding sequences is defined as the number of the codon that immediately follows the intron. Within brackets is shown the place within this codon which is the first nucleotide of the exon. The numbers of the codons refer to the amino acid residues within the mature histone H3 protein.

have the potential to exclude nucleosomes (Tsukiyama *et al.*, 1994; Wallrath *et al.*, 1994) and facilitate gene transcription (Lu *et al.*, 1993; O'Donnell and Wensink, 1994). This could be a factor in the constitutive, high basal level of expression that was observed for these genes. Analysis of the chromatin conformation of the alfalfa msH3g1 replacement histone gene is in progress to evaluate this speculation.

MATERIALS AND METHODS

DNA libraries and procedures

The LUCA cDNA library, made available for this study, had been prepared from somatic embryos of the RA3 line of *Medicago sativa*, cv. (cultivar) Regen S. (Wu *et al.*, 1989). The alfalfa genomic library, purchased from Clontech Laboratories, was prepared from 18 day post-emergence seedlings of *Medicago sativa*, cv. Chief. Histone H3.2 variant clones were identified by standard colony and plaque screening procedures (Sambrook *et al.*, 1989) by hybridization with the *Pvu*II-*Hin*dIII 3'UTR (untranslated region) sequence of pH3c11, a partial cDNA for alfalfa histone H3.2 (Wu *et al.*, 1989), which has been shown to be histone variant gene specific (Kapros *et al.*, 1992). All hybridization probes were labeled by random priming of purified plasmid DNA fragments with the digoxigenin Genius system, following procedures recommended by Boehringer, and detected by chemiluminescence with CSPD (Tropix). DNA sequences were determined by Sequenase 2.0 double-strand sequencing with DMSO for plasmids (Seto, 1990; Schuurman and Keulen, 1991) and by a modified protocol for phage DNA (Manfioletti and Schneider, 1988). Initially, sequencing primers were designed from the pH3c11 sequence (Wu *et al.*, 1989), and subsequently from newly acquired sequences.

Polymerase chain reactions

Polymerase Chain Reactions (PCR) for histone H3 variant genes were performed on alfalfa genomic DNA, prepared from RA3 callus cultures of *M. sativa* cv. Regen S., as described before (Winicov *et al.*, 1988), and from sprouted seeds of *M. sativa* cv. Chief, kindly provided by Agrigenetics Research, by the O-ethylxanthic acid method described by Jhingan (Jhingan, 1992a, 1992b). Optimized PCR conditions for the MJ Research Minicycler with primers for histone H3 gene copy determinations and for mapping of the upstream region of the H3.2 gene were 10 mM Tris.HCl, pH 8.3, 50 mM KCl, 0.2 mM of 4 dNTPs, 3.5 mM $MgCl_2$, with 10% (v/v) DMSO as a specificity enhancer (Mody and Paul, 1990; Smith *et al.*, 1990) and Amplitaq *Taq* DNA polymerase (Perkin Elmer), for 30 cycles of 45 sec 92 °C, 45 sec 58 °C and 45 sec 72 °C. Plasmid ph3c1Δ65 was created as an internal standard for gene copy titration of the H3.1 gene by linearizing H3.1 cDNA plasmid pH3c1 (Wu *et al.*, 1989) with *Eco*RV, directed removal of 65 bp by nuclease Bal 31 and religation.

Quantitation of histone H3 transcript levels

Run-off transcript RNA (570 n) for histone H3.2 was produced by SP6 polymerase transcription in the Promega Riboprobe Gemini II System of plasmid pH3c11Δ5', linearized at the single *Mun*I restriction site, very close to the polyA tail. Plasmid pH3c11Δ5' was produced by digestion of pH3c11 DNA with *Bam*HI and *Sty*I and religation. This removes a polyG sequence, a result of the cDNA cloning procedure (Wu *et al.*, 1989). Run-off transcript for histone H3.1 (600 n) was produced from *Hin*dIII-linearized pH3c1Δ5' DNA, derived from pH3c1 by deletion of the *Bam*HI-*Bsa*AI fragment. Dilutions of these standard RNAs between 5 and 200 pg, based on absorbance at 260 nm, were co-electrophoresed in a single formaldehyde agarose gel (Sambrook *et al.*, 1989) with samples of 12 μg total RNA, as determined by absorbance at 260 nm, prepared from alfalfa A2 suspension cell cultures as described elsewhere (Kapros and Waterborg, 1995). Northern analysis was performed with digoxigenin-substituted histone H3 variant-specific hybridization probes, prepared from equal-sized dimers of the 3'UTR sequences of plasmids pH3c1 and pH3c11, as described elsewhere (Kapros *et al.*, 1994). Hybridization was quantitated by densitometry of films after chemiluminescent detection of digoxigenin with CSPD (Tropix) (Kapros *et al.*, 1995).

Histone H3 protein synthesis

Alfalfa A2 suspension cell cultures in exponential growth or cultures enriched in S phase cells by growth for 24 h at 10 mM hydroxyurea (Kapros *et al.*, 1995) were incubated for 1 h with 0.125 mCi of L-[4,5-^{3}H]lysine (102.4 Ci/mmol) (New England Nuclear) in 50 ml growth medium. Histone proteins were extracted from cells and fractionated by reversed-phase HPLC, as described previously (Waterborg, 1993). The rate of *de novo* synthesis of the separated peaks of histone H3.1 and H3.2 variants was determined from the amount of radioactivity incorporated per amount of protein absorbance at 214 nm in the HPLC eluate.

ACKNOWLEDGEMENTS

The assistance of Dr. Janos Györgyey in the rescreening of the LUCA library is greatly appreciated. This work was supported by National Science Foundation grant DCB 9118999 to JHW.

(Received 5 October 1995)

REFERENCES

Chaboute M.E., Chaubet N., Philipps G., 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.

Chaboute M.E., Chaubet N., Gigot C. and Philipps G. (1993). Histones and histone genes in higher plants: structure and genomic organization. *Biochimie* **75**, 523–531.

Chaubet N., Chaboute M.E., Philipps G. and Gigot C. (1987). Histone genes in higher plants organization and expression. *Dev. Genet.* **8**, 461–474.

Chaubet N., Clement B. and Gigot C. (1992). Genes encoding a histone H3.3-like variant in Arabidopsis contain intervening sequences. *J. Mol. Biol.* **225**, 569–574.

Dalton S. and Wells J.R. (1988). A gene-specific promotor element is required for optimal expression of the histone H1 gene in S-phase. *EMBO J.* **7**, 49–56.

Desjardins E. and Hay N. (1993). Repeated CT elements bound by zinc-finger proteins control the absolute and relative activities of the two principal human c-myc promoters. *Mol. Cell. Biol.* **13**, 5710–5724.

Firulli A.B., Maibenco D.C. and Kinniburgh A.J. (1994). Triplex formation ability of a c-myc promoter element predicts promoter strength. *Arch. Biochem. Biophys.* **310**, 236–242.

Hendzel M.J. and Davie J.R. (1990). Nucleosomal histones of transcriptionally active competent chromatin preferentially exchange with newly synthesized histones in quiescent chicken erythrocytes. *Biochem. J.* **271**, 67–73.

Jhingan A.K. (1992a). Efficient procedure for DNA extraction from lyophilized plant material. *Meth. Mol. Cell. Biol.* **3**, 185–187.

Jhingan A.K. (1992b). A novel technology for DNA isolation. *Meth. Mol. Cell. Biol.* **3**, 15–22.

Kapros T. and Waterborg J.H. (1995). Rapid extraction of multiple RNA samples from plant suspension cells. *Focus* **17**, 20–21.

Kapros T., Bögre L., Németh K., Bakó L., Györgyey J., Wu S.-C. and Dudits D. (1992). Differential expression of histone H3 gene variants during cell cycle and somatic embryogenesis in alfalfa. *Plant Physiol.* **98**, 621–625.

Kapros T., Stefanov I., Magyar Z., Ocsovszky I. and Dudits D. (1993). A short histone H3 promotor from alfalfa specifies expression in S phase cells and meristems. *In Vitro Cell. Dev. Biol.* **29**, 27–32.

Kapros T., Robertson A.J. and Waterborg J.H. (1994). A simple method to make better probes from short DNA fragments. *Mol. Biotechnol.* **2**, 95–98.

Kapros T., Robertson A.J. and Waterborg J.H. (1995). Histone H3 transcript stability in alfalfa. *Plant Mol. Biol.* **28**, 901–914.

Kolluri R., Torrey T.A. and Kinniburgh A.J. (1992). A CT promoter element binding protein: definition of a double-strand and a novel single-strand DNA binding motif. *Nucleic Acids Res.* **20**, 111–116.

Lewin B. (1994) *Genes V.* Oxford University Press, Oxford.

Lu Q., Wallrath L.L., Granok H. and Elgin S.C. (1993). (CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila hsp26 gene. *Mol. Cell. Biol.* **13**, 2802–2814.

Maher L.J. (1992). DNA triple-helix formation: an approach to artificial gene repressors ? *Bioessays* **14**, 807–815.

Manfioletti G. and Schneider C. (1988). A new and fast method for preparing high quality lambda DNA suitable for sequencing. *Nucleic Acids Res.* **16**, 2873–2884.

Mikami K. and Iwabuchi M. (1993) Regulation of cell cycle-dependent gene expression. In Verma D.P.S. (ed.), *Control of Plant Gene Expression* (Boca Raton, FL: CRC Press), pp. 51–68.

Mody R. and Paul S. (1990). A specificity enhancer for Polymerase Chain Reactions that reduces non-specific priming by degenerate primers. *Clontech* **5**, 1–2.

Morris D.R., Kakegawa T., Kaspar R.L. and White M.W. (1993). Polypyrimidine tracts and their binding proteins: regulatory sites for posttranscriptional modulation of gene expression. *Biochemistry* **32**, 2931–2937.

O'Donnell K.H. and Wensink P.C. (1994). GAGA factor and TBF1 bind DNA elements that direct ubiquitous transcription of the α1-tubulin gene. *Nucleic Acids Res.* **22**, 4712–4718.

Ohtsubo N., Nakayama T., Terada R., Shimamoto K. and Iwabuchi M. (1993). Proximal promoter region of the wheat histone H3 gene confers S phase-specific gene expression in transformed rice cells. *Plant Mol. Biol.* **23**, 553–565.

Ridsdale J.A. and Davie J.R. (1987). Chicken erythrocyte polynucleosomes which are soluble at physiological ionic strength and contain linker histones are highly enriched in beta-globin gene sequences. *Nucleic Acids Res.* **15**, 1081–1096.

Sambrook J., Fritsch E.F. and Maniatis T., Eds. (1989) *Molecular Cloning. A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sarkar P.S. and Brahmachari S.K. (1992). Intramolecular triplex potential sequence within a gene down-regulates its expression in vivo. *Nucleic Acids Res.* **20**, 5713–5718.

Schuurman R. and Keulen W. (1991). Modified protocol for DNA sequence analysis. *BioTechniques* **10**, 185.

Seto D. (1990). An improved method for sequencing double stranded plasmid DNA from minipreps using DMSO and modified template preparation. *Nucleic Acids Res.* **18**, 5905–5906.

Smith K.T., Long C.M., Bowman B. and Manos M.M. (1990). Using cosolvents to enhance PCR amplification. *Amplific.* **5**, 16–17.

Tabata T., Terayama C., Mikami K., Uchiyama H. and Iwabuchi M. (1987). An accurate transcription of wheat histone genes in sunflower cells. *Plant Cell Physiol.* **28**, 73–82.

Thatcher T.H., MacGaffey J., Bowen J., Horowitz S., Shapiro D.L. and Gorovsky M.A. (1994). Independent evolutionary origin of histone H3.3-like variants of animals and Tetrahymena. *Nucleic Acids Res.* **22**, 180–186.

Tsukiyama T., Becker P.B. and Wu C. (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**, 525–532.

Van Holde K.E. (1989) Chromatin. (New York: Springer Verlag), pp. 1–497.

Wallrath L.L., Lu Q., Granok H. and Elgin S.C. (1994). Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures. *Bioessays* **16**, 165–170.

Waterborg J.H. (1990). Sequence analysis of acetylation and methylation of two histone H3 variants of alfalfa. *J. Biol. Chem.* **265**, 17157–17161.

Waterborg J.H. (1991). Multiplicity of histone H3 variants in wheat, barley, rice and maize. *Plant Physiol.* **96**, 453–458.

Waterborg J.H. (1992). Existence of two histone H3 variants in dicotyledonous plants and correlation between their acetylation and plant genome size. *Plant Mol. Biol.* **18**, 181–187.

Waterborg J.H. (1993). Histone synthesis and turnover in alfalfa. Fast loss of highly acetylated replacement histone H3.2. *J. Biol. Chem.* **268**, 4912–4917.

Waterborg J.H., Harrington R.E. and Winicov I. (1989). Differential histone acetylation in alfalfa (*Medicago sativa*) due to growth in NaCl: Responses in salt stressed and salt tolerant callus cultures. *Plant Physiol.* **90**, 237–245.

Wells J.R.E., Coles L.S. and Robins A.J. (1989) Organization of histone genes and their variants. In Hnilica L.S., Stein G.S. and Stein J.L. (eds), *Histones and Other Basic Proteins.* (Boca Raton, FL: CRC Press), pp. 253–267.

Winicov I., Maki D.H., Waterborg J.H., Riehm M.R. and Harrington R.E. (1988). Characterization of the alfalfa (*Medicago sativa*) genome by DNA reassociation. *Plant Mol. Biol.* **10**, 369–371.

Wu S.-C., Bögre L., Vincze E., Kiss G.B. and Dudits D. (1988). Isolation of an alfalfa histone H3 gene: structure and expression. *Plant Mol. Biol.* **11**, 641–650.

Wu S.-C., Györgyey J. and Dudits D. (1989). Polyadenylated H3 histone transcripts and H3 histone variants in alfalfa. *Nucleic Acids Res.* **17**, 3057–3063.

Zweidler A. (1980). Nonallelic histone variants in development and differentiation. *Dev. Biochem.* **15**, 47–56.