Histone gene organization of fission yeast: a common upstream sequence

Seiji Matsumoto and Mitsuhiro Yanagida

Department of Biophysics, Faculty of Science, Kyoto University, Sakyo-ku, Kyoto 606, Japan

Communicated by D.Gallwitz

Histone genes of the fission yeast Schizosaccharomyces pombe were cloned from Charon 4A and cosmid gene libraries by hybridization, and their nucleotide sequences were determined. The genome of S. pombe has a single, isolated H2A, a pair of H2A-H2B and three pairs of H3-H4 (one H2B, two H2A and three each of H3 and H4). This non-assorted histone gene organization is distinct from that of the budding yeast which has two pairs of H2A-H2B and H3-H4. The predicted amino acid sequences of S. pombe histone H2As, H3s and H4s were identical except for three residue changes in H2As. Compared with those of S. cerevisiae and human, variable residues were clustered near the NH₂- and COOH-terminal regions of H2A and H2B. Sequence homologies to the two organisms were roughly the same in H2A (79 - 83%), H3 (92 - 93%) and H4 (91%), but differed in H2B (82% to S. cerevisiae and 68% to human). The coding sequences in the pairs of S. pombe histone genes were divergently directed. A 17-bp long highly homologous sequence (AACCCT box) that had internal 6-bp direct repeats was present in the intergene spacer sequences or in the 5' upstream region of all the cloned histone genes. A possible regulatory role of the common upstream sequence for histone gene expression is discussed.

Key words: histone genes/*Schizosaccharomyces pombe*/common upstream sequence

Introduction

Histone genes are highly reiterated in a wide range of organisms (reviewed by Kedes, 1979; Hentschel and Birnstiel, 1981; Old and Woodland, 1984). About 600-800 copies are present in newt, 300-600 in sea urchin and 100 in *Drosophila*. Mammals have fewer histone genes, of the order of 10- to 40-fold reiteration. Histone genes have been divided into a variety of subtypes; certain variants are expressed only in the S phase (DNA synthesis) of the cell cycle and other variants are expressed throughout the cycle. Tissue-specific subtypes of histones are also known, such as H5 in the nucleated erythrocytes of birds and sperm-specific histones in a range of organisms.

In lower eukaryotes, histone genes are few in number (Hereford *et al.*, 1979; Smith and Andresson, 1983; Woudt *et al.*, 1983). Yeast *Saccharomyces cerevisiae* contains two copies of each of the H2A and H2B, H3 and H4 histone genes. The H1 gene has not yet been found in yeast. When either of the two H2B genes is disrupted, the cells are still viable, indicating that neither of these two variants plays an essential role (Rykowski *et al.*, 1981). Disruption of both of the H2B genes is, however, lethal. The H2A and H2B genes are adjacent to one another and there are two genetically unlinked copies of this pair per haploid genome (Hereford *et al.*, 1979; Wallis *et al.*, 1980; Choe *et al.*, 1982). The H3 and H4 genes similarly are paired and there are

two copies of the pair that are not contiguous with H2A and H2B genes (Smith and Andresson, 1983; Smith and Murray, 1983). The genes at each locus are divergently transcribed and the coding sequences are flanked by a non-homologous sequence several hundred base pairs long. In *Neurospora crassa*, the number of histone genes is even fewer; only one copy of H3 and H4 genes is present in the haploid genome (Woudt *et al.*, 1983).

We undertook to isolate the histone genes of the fission yeast *Schizosaccharomyces pombe* and to determine the histone gene organization. *S. pombe* has a well defined genetic system (Kohli *et al.*, 1977) and is suitable for cell cycle analyses (Mitchison and Carter, 1975; Nurse *et al.*, 1976). The DNA content per haploid genome is small, equivalent to that of *S. cerevisiae* and 3-4 times that of *Escherichia coli*. The cell cycle consists of distinct S and M phases interspersed with short G1 and long G2 phases (Mitchison, 1970; Beach *et al.*, 1982). Chromosomes condense during mitosis (Toda *et al.*, 1981; Hiraoka *et al.*, 1984) and individual chromosomes could be seen in the cells of a β -tubulin mutant under the restrictive condition (Umesono *et al.*, 1983; Hiraoka *et al.*, 1984).

We report here the histone gene organization of S. pombe with complete nucleotide sequences of two H2A, one H2B, and three H3 and three H4 genes. The histone genes were cloned from Charon 4A or cosmid libraries of S. pombe genomic DNAs using the probes of S. cerevisiae histone H2A genes and of the Drosophila histone gene repeating unit. S. pombe histone genes thus obtained were employed as the probes to complete cloning of the whole histone genes. Our results show that the genome of S. pombe has a single, isolated H2A, a pair of H2A and H2B and three copies of a pair of H3 and H4. The gene organization is not assorted, that is, the copy numbers of each histone gene are not equal. Predicted amino acid sequences of the S. pombe histone genes differ by about the same extent from S. cerevisiae and mammals except in the case of H2B, supporting the notion that S. pombe is evolutionarily very distinct from S. cerevisiae. Computer sequence analyses indicate that all of the cloned histone genes have a highly homologous upstream sequence that contains internal direct repeats. A possible regulatory role of this sequence in the expression of the S. pombe histone genes is discussed.

Results

Cloning of an H2A gene

 α -³²P-Labelled probe E' (0.4 kb in length) containing the coding region of *S. cerevisiae* histone H2A2 (Hereford *et al.*, 1979) was used for Southern blots of *S. pombe* genomic DNA digests electrophoresed in agarose gel. Hybridizing bands were obtained at 5.5 and 8.6 kb for *Eco*RI and 1.2 and 14 kb for *Hind*III digests (Figure 1a, lanes 1–2). The probe E (1.4 kb) containing *S. cerevisiae* H2A1, protein 1 and flanking sequences (Hereford *et al.*, 1979) gave the same bands with a few additions (data not shown).

We constructed a Charon 4A library containing 15-23 kb of S. pombe genomic DNA as described in Materials and methods.

S.Matsumoto and M.Yanagida

Four thousand plaques from the library were screened with ³²Plabeled E'. Seven weakly positive clones were obtained and designated Chl-1, Chl-2, --- Chl-7. These were restricted with EcoRI or HindIII and hybridized with ³²P-labeled E'. Only Chl-7 produced the intense hybridizing bands (at 8.6 kb with EcoRI and 1.2 kb with *Hind*III) expected from the results of genomic Southern blots. The others showed very faint non-specific bands. Therefore, we analyzed Chl-7 to determine whether it contained the H2A gene. The 1.2-kb HindIII fragment of Chl-7 was subcloned into pBR322 (designated pSJM211; Figure 2) and probed for S. pombe genomic DNA restricted with EcoRI or Hind-III. Two sets of bands were obtained (Figure 1a, lanes 3-4): a set of intense bands (self-hybridized) at 8.6 kb with EcoRI and 1.2 kb with HindIII, and another weaker set at 14 kb with HindIII and 5.5 kb with EcoRI. The sizes of these hybridized bands were identical to those produced by the S. cerevisiae H2A probe E'. Nucleotide sequence determination (described below) showed that the 1.2-kb HindIII fragment of Chl-7 encoded an H2A gene (designated H2A.2). The remaining six Ch clones did not hybridize at all with the 1.2-kb S. pombe H2A clone and



Fig. 1. Genomic Southern hybridization of *S. pombe* DNA digested with EcoRI (lanes 1 and 3 in a and lanes 1 and 2 in b) or *Hind*III (lanes 2 and 4 in a) and probed with ³²P-labeled histone H2A and H2B probes. (a) probed with *S. cerevisiae* H2A2 (lanes 1 and 2) and with *S. pombe* H2A.2 (lane 3 and 4). (b) Probed with *S. cerevisiae* H2B1 (lane 1) and with *S. pombe* H2B.1 (lane 2).

were not investigated further. Attempts to clone the histone H2B gene of *S. pombe* by the G probe of *S. cerevisiae* H2B1 (Hereford *et al.*, 1979) were not successful, although genomic Southern blots of *Eco*RI-digested *S. pombe* DNA gave a faint band at 5.5 kb (Figure 1b, lane 1).

Cloning of other histone genes

To obtain other histone genes of fission yeast, the histone gene repeating unit of *Drosophila* (Lifton *et al.*, 1977) was alternatively employed as the probe. Genomic Southern blots probed with pKSS100 (a gift from K. Saigo; Saigo *et al.*, 1981) that has an insert of the 4.8-kb *Bam*HI repeating unit (Figure 7) produced several intense hybridizing bands with *S. pombe* DNA digested with *Eco*RI or *Hind*III (data not shown). From the Charon 4A library, four positive clones were obtained by hybridization with pKSS100 (Ch2-2, Ch2-3, Ch2-4 and Ch11-6).

To identify the histone genes contained in these clones, fragments of the *Drosophila* repeating unit were subcloned into pBR322 as described in Materials and methods (Figure 7). Resulting plasmids pSJM102, pSJM103 and pSJM104 contained H3, H4 (plus a part of H2A) and H2B (plus the other part of H2A) genes, respectively.

Five Ch clones (Ch1-7, Ch2-2, Ch2-3, Ch2-4 and Ch11-6) were digested with *Eco*RI or *Eco*RI/*Hind*III, electrophoresed in



Fig. 2. *S. pombe* genomic DNA sequences containing histone genes. DNA fragments obtained from Charon 4A and cosmid libraries by hybridization were subcloned into pBR322. Restriction sites are indicated: E. *Eco*RI; H, *Hind*III; P, *Pst*I; B, *Bam*HI. The arrow represents the coding region and the direction of transcription for each histone gene. pSJM210 has an insert of the 8.6-kb *Eco*RI fragment of Ch1-7 (Contains H2A.2); pSJM211, the 1.2-kb *Hind*III fragment of Ch1-7 (H2A.2); pSJM220, the 5.5-kb *Eco*RI fragment of Ch1-6 (H2A.1-H2B.1); pSJM221, the 2.6-kb *PstI-Eco*RI fragment of Ch1-6 (H2B.2); pSJM230, the 2.9-kb *Eco*RI fragment of Ch2-3 (H3.1-H4.1); pSJM231, the 1.2-kb *Hind*III-*Pst*I fragment of Ch2-3 (H4.1); pSJM232, the 0.8-kb *PstI-Eco*RI fragment of Ch2-3 (H3.1); pSJM240, the 3.0-kb *Eco*RI fragment of Ch2-4 (H3.2-H4.2); pSJM250, the 5.3-kb *Eco*RI fragment of c13 (H3.3-H4.3).

Table I.	Hybridization of S	. pombe genom	c DNA	fragments	by the	probes of S.	cerevisiae,	Drosophila a	ınd S.	pombe	histone	genes
----------	--------------------	---------------	-------	-----------	--------	--------------	-------------	--------------	--------	-------	---------	-------

Clones	<u>EcoRI digest</u> pKSS100 (kb)	EcoRI/HindIII double digest							
		pKSS100 (kb)	E' (H2A) (kb)	pSJM211 (H2A) (kb)	pSJM104 (H2B+H2A) (kb)	pSJM102 (H3) (kb)	pSJM103 (H4+H2A) (kb)		
 Ch1-7	8.6	1.2	1.2	1.2	1.2	_	1.2		
Ch11-6	5.5	3.3, 1.5	3.3	3.3	3.3, 1.5	_	3.3		
Ch2-2	10	5	-	_	_	5	_		
Ch2-3	2.9	2.1		-	_	2.1	2.1		
Ch2-4	3.0	1.0	_	-	-	1.0	1.0		

Sizes of the hybridizing bands are shown with the probes as follows: pKSS100, containing a five histone gene repeating unit of *Drosophila*. E', containing the *S. cerevisiae* H2A2 gene. pSJM211, containing *S. pombe* H2A.2. pSJM104, pSJM102 and pSJM103 contain the *Drosophila* histone genes specified in parenthesis (see Figure 7). -, no hybridization.

AGCACTCCAACAGTAACAGCTTTACCTTCCGACACCTTTTGAGTAGTCGACTGATTTTCTGCTAACGTGTTAGCACGCTAAGCG AAT GAC TCG TCT TCT TAT GAA CCA Stop Gln Ala Ser Ser Ser Tyr Lys Thr 120 125 CTG CCT AAA CCA TGG AAG CCA TTG CCG CAC GAA CCG ATT GAG TGG CCC TTC TTA GTT TGC TTG TCG TCA GAC CTA AAG TGC CCT TCT TTA Val Ser Lys Thr Gly Glu Thr Val Ala His Lys Ala Leu Glu Gly Pro Leu Ile Leu Arg Val Ala Thr Gln Ile Glu Arg Ser Ser Ile 90 110 100 TCA CCT GAA GAA CAA CAT TCG TCG TTC GAA CCT TCG GAG TCA CCG TTA TGC GAG TTT TTA TAG CAA CTG CTT TCT CAA GTT CTA TGC GTA Thr Ser Lys Lys Asn Tyr Ala Ala Leu Lys Ser Ala Glu Thr Ala Ile Arg Glu Phe Ile Asp Asn Val Phe Ser Asn Leu Ile Arg Met 70 60 80 CCG AAC CAA CCT TTA TGG TCA TAG TCC CAC TTG AAC GAA GTT GTG GAA CAT TTA TAT CCT ACT TAT TCA AAG GAA AGA CAA AAA TGG TGC Ala Gin Asn Ser Ile Gly Thr Asp Pro His Val Gin Lys Leu Val Lys Tyr Ile Tyr Ser Ser Tyr Thr Glu Lys Arg Asn Lys Gly Arg 30 50 40 Ω TTT GTT TTC AGC AGC AGA CAT TTTTTAC GAA GAA TAG TCG TCT GAA GTA CCA TAG GGA TCC TCG GAA TGG CCG CCC TCG GAA CCT TCG CCC AAA GAA AAG TCG TCG TCT CTA H2B.1 Lys Lys Asp Ala Ser Lys Met Thr Asp Arg Pro Ala Lys Gly Ala Pro Ala Lys Ser Ala Pro Lys Lys Glu Ala Ala Ser 20 10 -120-20 -40 -60 -80 -100GATTTAAAAATTGGTTGTCGAGAGAATCTCAAAATTGATGGACTGAACTAAGTAATTGCAATTGCAATTGTAGGAATTTGATAATCCTAAACTTTCGAGGAATATCAAGTTGTTGGTATTAACTA.-140 -167 .-370 -340 .-320 -300 .-280 .-260 .-240 .-220 .-200 AAGA'I'I'AGTATATATATGAAGGGTGCGCGCCCCCTATGTCCCTTTTCCCTATATCGCGACTTTTCGTTGTTTAGGATACCATTTGATATGAAATTGATATAAAAATTG -180. -80 .-160 .-140 .-120 . - 100 GATCCTTAGTTGATAAGGAGATTGATTGTCCGTAAAGGCTGAAATTGCGAGAAAATTCTGTTAATTTCACCAAATTTGATCCATTCGCTGCTTCATCGTTGACTGTTATCTTTTTTCAT -60 .-40 .-20 .0 H2A.1 Ser Gly Gly Lys Ser Gly Gly Lys Ala Ala Val 10 GCC AAA TCT GCT CAA TCC CGT TCT GCC AAG GCC GGT TTG GCT TTC CCC GTT GGT CGT GTC CAT CGT TTG TTA CGT AAG GGT AAC TAT GCT Ala Lys Ser Ala Gln Ser Arg Ser Ala Lys Ala Gly Leu Ala Phe Pro Val Gly Arg Val His Arg Leu Leu Arg Lys Gly Asn Tyr Ala 20 30 40 CAA CGT GTT GGT GCT GGT GCT CCC GTT TAC TTG GCC GCC GTT TTG GAA TAT TTG GCT GCT GAA ATT CTC GAA TTG GCT GGA AAC GCT GCT Gin Arg Val Gly Ala Gly Ala Pro Val Tyr Leu Ala Ala Val Leu Glu Tyr Leu Ala Ala Glu Ile Leu Glu Leu Ala Gly Asn Ala Ala 50 60 70 CGT GAC AAG AAG AAG ACT CGT ATC ATT CCC CGT CAT CTT CAA CTC GCC ATC CGT AAC GAT GAA GAG TTG AAC AAA CTT CTT GGT CAC GTT Arg Asp Asn Lys Lys Thr Arg Ile Ile Pro Arg His Leu Gln Leu Ala Ile Arg Asn Asp Glu Glu Leu Asn Lys Leu Leu Gly His Val 90 80 100 ACC ATT GCT CAA GGT GGT GTT GTT CCC AAC ATC AAC GCC CAT CTT TTG CCC AAA ACT TCT GGT CGC ACT GGA AAG CCT AGT CAG GAG CTG Thr Ile Ala Gin Giy Giy Val Val Pro Asn Ile Asn Ala His Leu Leu Pro Lys Thr Ser Giy Arg Thr Giy Lys Pro Ser Gin Giu Leu 120 110 131 . 50 .100 Stop .-480 .-440 -420 -400 -380 ,-460 AAGCTTTTTGTACTTTTGCACGTCCTGTTATATTTAGTTTAAACTACGATGGTCGATATTTAGTAGCAGACTTTTCTTTTCCGTAAAAATTAAGAGTATACTTAGTTTAATAAAATCAT b .-280 .-360 .-340 .-320 .-300 -260 .-140 .-240 .-220 .-200 .-180 .-160 .-80 .-20 .-60 .-40 .-120 .-100 AAACCCTAACCCTGTATATTGCTGTATCGAAAGAGCTTGTCGCTGCATCATCTTCTTCTTCTAAATATTGGGCAATTGAGCAATCCTGCATACATTATTCCAATTTAAAACTTAAGT .0 ATG, TCT GGA GGT AAA TCT GGT GGT AAG GCC GCA GTC GCC AAG TCT GCT CAA TCT CGT TCC GCT AAG GCT GGT TTG AATTTCTTTATCAAGC H2A.2⁹Ser Gly Gly Lys Ser Gly Gly Lys Ala Ala Val Ala Lys Ser Ala Gln Ser Arg Ser Ala Lys Ala Gly Leu 10 20 GCC TIC CCT GTC GGT CGT GTT CAT CGT TTG TTG CGT AAG GGT AAT TAT GCT CAA CGT GTT GGT GCT GCT CCC GTT TAC TTG GCT GCC Ala Phe Pro Val Gly Arg Val His Arg Leu Leu Arg Lys Gly Asn Tyr Ala Gln Arg Val Gly Ala Gly Ala Pro Val Tyr Leu Ala Ala 30 40 50 GTT CTT GAG TAT TTA GCT GCC GAA ATC CTC GAA TTG GCC GGT AAT GCT GCT GGT GAT AAC AAG AAG ACT CGT ATC ATT CCC CGT CAT CTT Val Leu Glu Tyr Leu Ala Ala Glu Ile Leu Glu Leu Ala Gly Asn Ala Ala Arg Asp Asn Lys Lys Thr Arg Ile Ile Pro Arg His Leu 60 70 80 CAA TTG GCT ATT CGC AAT GAC GAA GAA TTG AAC AAA CTA CTT GGT CAT GTT ACT ATT GCC CAG GGT GGT GTT GTA CCT AAT ATC AAC GCT Gin Leu Ala Ile Arg Asn Asp Glu Glu Leu Asn Lys Leu Leu Gly His Val Thr Ile Ala Gin Gly Gly Val Val Pro Asn Ile Asn Ala

90 100 110 50. CAT CTC TTG CCC AAA CAA TCT GGT AAG GGC AAG CCT AGC CAA GAG CTT TAA ACTCTTAAATGAATGATAAATTGTATATTAAGATTATTTGTCGTTTATACT His Leu Leu Pro Lys GIn Ser Gly Lys Gly Lys Pro Ser GIn Glu Leu Stop 120 130

.100 . .150 TAACTATTATTATCCACACTAATTGCTTGAACTGTTGGTTTCGTGACTGATATTCGTTCACACCCTTTTTTTACGCAAATTTCTCTATCAGAAATTTAGCTTTATCATTATCAGTGATA

Fig. 3. Nucleotide sequences of *S. pombe* histone H2A and H2B genes. Predicted amino acids are shown in italics below the nucleotide sequences. (a) H2A.1 and H2B.1 encoded in the 5.5-kb *Eco*RI fragment (pSJM220) of Ch11-6. (b) H2A.2 encoded in the 1.2-kb *Hind*III fragment (pSJM211) of Ch1-7. Arrow indicates the direction of transcription. The underlined sequences represent the AACCCT box (see text).

agarose gel and probed with either *Drosophila* pKSS100, *S. cerevisiae* E', *S. pombe* H2A (pSJM211) or one of the three *Drosophila* subfragments (pSJM102-104). The results are shown in Table I; sizes of the hybridized bands in the five Ch clones are indicated.

containing probes but not with H3-containing probe. Similarly the 5.5-kb *Eco*RI fragment of Ch11-6 was hybridized with H2A-containing probes, but an additional 1.5-kb band was obtained only with the H2B-containing probe, indicating that the 1.5-kb fragment contained H2B (the 5.5-kb fragment was later found to have a pair of H2A-H2B). The 10-kb *Eco*RI fragment of Ch2-2

The 8.6-kb *Eco*RI fragment of Chl-7 was hybridized with H2A-

S.Matsumoto and M.Yanagida

.150

. 50 a gtccccattitttccttaattcaccatttattattattcatatgttigttcctctttttttagttacattatcctca aat tgg tgg ctt tgg tat tta cca tgc cgg aac Stop Gly Gly Phe Gly Tyr Ile Thr Arg Gly Gln 102 100 TGC GAA GTT TCT TAT CTG TTG CAG GTT TCT TCA CTG TCA GAA TGC GAA CCG TAC GAG TCA TAT CCA CTG CCG TAG CGC TTA TTG TAA AAG Arg Lys Leu Ser Tyr Val Val Asp Leu Ser Thr Val Thr Lys Arg Lys Ala His Glu Thr Tyr Thr Val Ala Asp Arg Ile Val Asn Glu ٥n 80 70 ATT CTT CTC GAA CTC TTG TCG TCC TCA GAG AAG TAT TTG GTT TCG CCT TTA TCC GAA GTG TGG TGG TGC TCC TCC TTC TGC TGC CTA TCG Leu Phe Leu Lys Leu Val Ala Arg Thr Glu Glu Tyr Val Leu Ala Ser Ile Arg Lys Val Gly Arg Arg Ala Leu Arg Arg Ile Ala 60 50 40 TCC GAA TCA TTA TGG AAC TTA CAA CAG TGC TTC TTA AAA TGC CAC TGC GAA TCG TGG TGG AAA AGG GTT TGG AAA TGG TGG AAA TGG TGC Pro Lys Thr IIe Gly Gln IIe Asn Asp Arg Leu IIe Lys Arg His Arg Lys Ala Gly Gly Lys Gly Leu Gly Lys Gly Gly Lys Gly Arg 30 10 20 .-20 . -40 .-60 .-80 -100 .0 ACC AGA CAT GTTAAGAAGTAGAAAAAATTTAAAAATTAAAAATCTACGCAAAAAAACGCGTAATTCTTTAAGTATAAAGATAAAGGGTTTTTCCGTTGACTAACAAAGGAACAATGAATTT ACC AGA CAL TGG TCT GTA Gly Ser H4.1 .-140 .-160 .-180 .-200 -220 ATAGATGTCAACAGATGTTAACGGTAGGATAAAAGTCAATGCAAATGGAAATGGAAAGGATTTTATCCGTATTCAAAAACAATTAAAAGAAATGATTGTCTGAAGAGAGGGGAAGTTGGGGCGA .-250 .-166 -140 .-120 -100 CGCTGTCTATTTGTTTTGAATCTGCATCAAAACCCTAACCCTGCTTTTGATGAGATCGGGTGGATTCAAAACATATATAGCAAATATCTGCAGTACGCTTGCGTTTCCATTAATGTC . -80 .-60 -40 .-20 HA ATG GCT CGT ACT AAA CAA ACA H3.10*Ala Arg Thr Lys Gin Thr* .0 GCT CGT AAG TCT ACC GGT GGT AAG GCA CCC CGT AAG CAA TTG GCC TCT AAG GCC GCT CGT AAG GCC GCT CCC GCT ACC GGA GGT GTT AAG Ala Arg Lys Ser Thr Gly Gly Lys Ala Pro Arg Lys Gln Leu Ala Ser Lys Ala Ala Arg Lys Ala Ala Pro Ala Thr Gly Gly Val Lys 30 20 10 AAG CCT CAT CGT TAT CGT CCT GGT ACT GTC GCT CTT CGT CGT GAG ATT CGT CGT TAT CAA AAG TCT ACT GAA CTT TTA ATT CGT AAG CTA CCT Lys Pro His Arg Tyr Arg Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Arg Tyr Gln Lys Ser Thr Glu Leu Leu Ile Arg Lys Leu Pro 40 50 60 TTC CAA CGT TTG GTC CGT GAA ATC GCC CAA GAT TTC AAG ACT GAC TTG CGT TTC CAA TCT TCT GCC ATT GGT GCT CTC CAA GAA GCT GTT Phe Gln Arg Leu Val Arg Glu Ile Ala Gln Asp Phe Lys Thr Asp Leu Arg Phe Gln Ser Ser Ala Ile Gly Ala Leu Gln Glu Ala Val 70 80 90 GAG GCC TAC CTT GTC TCT CTA TTT GAG GAC ACT AAC TTG TGT GCT ATT CAC GGA AAA CGT GTT ACG ATT CAA CCC AAG GAT ATG CAG TTG Glu Ala Tyr Leu Val Ser Leu Phe Glu Asp Thr Asn Leu Cys Ala Ile His Gly Lys Arg Val Thr Ile Gln Pro Lys Asp Met Gln Leu 110 100 120 50 Ala Arg Arg Leu Arg Gly Glu Arg Ser Stop 130 135 100 150

 ${\tt CCTTTGATATGTTGCTACCAATAAGTGAAAAGCAGGCTTCGTGAATAATTGTTGTTTTTGTTCTACATTACCAAGAGCTAGATGAAC}$

ATTCAGTTCACTCTTTCGTTTTTGTATTTTAATTATACGATTTGGGCTGTAATTTGGTTCCTTTATATGTACTGTTACTAACTTTTTTAATCCTTTTGAATTTTGGGTTATTT GTGTTATATGGTTTATTATAATCCGTTAGTTACGTA AAT TCT CGC AAG CGG TGC TTC TGC CCG GTT AAC GTA TAG GAA CCC AAC TTA CCA TTG Stop Ser Arg Glu Gly Arg Leu Arg Arg Ala Leu Gln Met Asp Lys Pro Gln Ile Thr Val 135 130 120 TGC GAA CGG TAC CTA TCG CGT TTC TAA CCA TAG AAG CTT CTC TCT CTG TTC TAT TCG GAG CTG TCG AAG AAC TTC TCG TGG TTA ACG CCT Arg Lys Cly His Ile Ala Cys Leu Asn Thr Asp Clu Phe Leu Ser Val Leu Tyr Ala Clu Val Ala Clu Cln Leu Ala Cly Ile Ala Ser 100 110 90 CCT AAC CTT TGC GTT CAG TCA GAA TTT TAG AAC CCG TTA AAG TGC CTG GTT TGC AAC CTT CCC TTC AAA CGC TTA GTT GTT AAG TCA TCT Ser Gin Phe Arg Leu Asp Thr Lys Phe Asp Gin Ala Ile Giu Arg Val Leu Arg Gin Phe Pro Leu Lys Arg Ile Leu Leu Giu Thr Ser 80 70 60 GAA AAC CAT TGC TGC CTA AAG TGC TTC TCG CTG TCA TGG TCC TGC TAT TGC TAC TCC GAA GAA CTG TGG CGG TCA TCG CCC TCG CCG GAA Lys Gln Tyr Arg Arg Ile Glu Arg Leu Ala Val Thr Gly Pro Arg Tyr Arg His Pro Lys Lys Val Gly Gly Thr Ala Pro Ala Ala Lys 50 40 30 .0 TTT ACG AGC AGT TTG CTT GGT ACG AGC CAT AACGTAATTCG TGC CCG TCG GAA TCT CCG GTT AAC GAA TGC CCC ACG GAA TGG TGG CCA TCT AAA TGC TCG TCA AAC GAA CCA TGC TCG GTA Arg Ala Ala Lys Ser Ala Leu Gin Lys Arg Pro Ala Lys Gly Gly Thr Ser Lys Arg Ala Thr Gin Lys Thr Arg Ala H3.2 20 10 .-60 -20 .-40 .-81 . - 237 -220 -200 .-180 .-160 .-140 .-120 .-100 CCTGAAAAAATGTATATATATATATGGAGATGGACGCCTAACTACAGCACTTACACAGTAACTACTACGACTAGGAGGCGCATTGAAATTCAAAATCATAGGAATTTCAGGATTTCAGTGTTTGC . -80 .-60 -40 .-20 .0 AAA GGA TTG GGA AAG GGT GGT GGT AAG CGC CAT CGT AAG ATC CTT CGT GAT AAC ATT CAA GGT ATT ACT AAG CCT GCC ATT CGT CGT CTT Lys Gly Leu Gly Lys Gly Gly Ala Lys Arg His Arg Lys Ile Leu Arg Asp Asn Ile Gln Gly Ile Thr Lys Pro Ala Ile Arg Arg Leu 20 10 30 GCT CGT CGT GGC GGT GTT AAG CGT ATT TCT GCT TTG GTT TAC GAA GAG ACT CGT GCC GTT CTC AAG CTT TTC TTG GAA AAC GTT ATC CGT Ala Arg Arg Gly Gly Val Lys Arg Ile Ser Ala Leu Val Tyr Glu Glu Thr Arg Ala Val Leu Lys Leu Phe Leu Glu Asn Val Ile Arg 50 40 60 GAT GCA GTT ACC TAC ACT GAA CAC GCC AAG CGT AAG ACT GTC ACT TCC TTG GAC GTT GTC TAC TCT TTG AAG CGT CAA GGC CGT ACC ATT Asp Ala Val Thr Tyr Thr Glu His Ala Lys Arg Lys Thr Val Thr Ser Leu Asp Val Val Tyr Ser Leu Lys Arg Gln Gly Arg Thr Ile 70 80 90 TAT GGT TTC GGT GGT TAA ACTGGTTGCACACTTTTCGA Tyr Gly Phe Gly Gly Stop 100 102

.100

50

.100 . 50 TAAACACTAATTATGTAAATAACTTACTTACTAATTATATTATCCGTTTCATTTACTCCCTTCCTAATGTTACCCCTATCTGCTGGTTTTCTAAGGTTTAGTGTGGTAGTCTACCGTTG 0 GTGTTAAACCATTTCAACG AAT ACT CGC AAG TGG CGC TTC TGC TGC TCG GTT AAC GTA TAG GAA CCC AAC TTA TCA ATG TGC GAA AGG CAC TTA Stop Ser Arg Glu Gly Arg Leu Arg Arg Ala Leu Gln Met Asp Lys Pro Gln Ile Thr Val Arg Lys Gly His Ile 135 130 120 TCG TGT GTT CAA TCA CAG AAG TTT ATC TCT CTG TTC TAT CCG GAG TTG TCG AAG AAC CTC TCG TGG TTA CCG TCT TCT GAC CTT TGC GTT Ala Cys Leu Asn Thr Asp Glu Phe Leu Ser Val Leu Tyr Ala Glu Val Ala Glu Gln Leu Ala Gly Ile Ala Ser Ser Gln Phe Arg Leu 110 100 90 CAG TCA GAA CTT TAG AAC CCG CTA AAG TGC CTG GTT CGC AAC CTT TCC ATC GAA TGC TTA GTT TTC AAG TCA TCT GAA AAC TAT TGC TGC Asp Thr Lys Phe Asp Gin Ala Ile Glu Arg Val Leu Arg Gin Phe Pro Leu Lys Arg Ile Leu Glu Thr Ser Lys Gin Tyr Arg Arg 80 70 60 TTA AAG TGC TTC TCG CTG ACA TGG CCC TGC TAT TGC TAC TCC GAA GAA TTG TGG TGG TCA TCG CCC TCG TCG GAA TGC CCG TCG GAA TCT Ile Glu Arg Leu Ala Val Thr Gly Pro Arg Tyr Arg His Pro Lys Lys Val Gly Gly Thr Ala Pro Ala Ala Lys Arg Ala Ala Lys Ser 50 40 30 .0 -20 AGT TTG CTT AGT ACG AGC CAT TATTTTCTAGTGTTCTGATAATTGCCCTGTG CCG TTC AAC GAA TGC TCC TCG GAA TGG TGG TCA ACT AAA TGC ACG TCA AAC GAA TCA TGC TCG GTA Ala Leu Gin Lys Arg Pro Ala Lys Gly Gly Thr Ser Lys Arg Ala Thr Gin Lys Thr Arg Ala **413.3** 20 10 . -40 -60 -80-92 -128-100 $TAGATAAAAAACAAATTATAGGGAGATTACGCTGTATAAATATGTCCCTACATCCAACTTTG\underline{ATCACAACCCTAACCCT}TAATTAAATTGACGTAAGTTAATAGGGACGGCGCATTAAT$.-80 .-60 .-40 .-20 .0 ATATATTGTGAAATGCCGCATACCCTACTAGCAATAGAAGCACTGTTTATAGTAGTAACTCTATTTCGTTAATATTAATAGTGCAAT ATG, TCT GGC CGT GGT AAA GGT GGA H4.3^{Ser} Gly Arg Gly Lys Gly Gly AAA GGT TTA GGA AAG GGT GGT GCC AAG CGT CAC CGT AAG ATC CTT CGT GAC AAT ATT CAA GGT ATT ACT AAG CCT GCC ATT CGT CGT CTT Lys Gly Leu Gly Lys Gly Gly Ala Lys Arg His Arg Lys Ile Leu Arg Asp Asn Ile Gln Gly Ile Thr Lys Pro Ala Ile Arg Arg Leu 10 20 30 GCT CGT CGT GGT GGT GTT AAG CGT ATT TCC GCT TTA GTT TAC GAA GAG ACT CGT GCT GTC CTC AAG CTC TTC TTG GAA AAT GTT ATT CGT Ala Arg Arg Cly Cly Val Lys Arg Ile Ser Ala Leu Val Tyr Glu Glu Thr Arg Ala Val Leu Lys Leu Phe Leu Glu Asn Val Ile Arg 40 50 60 GAT GCT GTC ACC TAT ACT GAG CAT GCC AAG CGT AAG ACT GTT ACA TCT TTG GAT GTT GTC TAC TCT TTG AAG CGT CAA GGC CGT ACC ATT Asp Ala Val Thr Tyr Thr Glu His Ala Lys Arg Lys Thr Val Thr Ser Leu Asp Val Val Tyr Ser Leu Lys Arg Gln Gly Arg Thr Ile 80 90 . 50 TAT GGT TTC GGT GGT TAA ATTCGATATCAATTGCCGACTCCGAATTTTTTGTTTCGAGTCTATCAATTTTATGTGTACACTATAACTATTCTATATACATGCCAAACGAGA Tyr Gly Phe Gly Gly Stop 100 102

Fig. 4. Nucleotide sequences of *S. pombe* histones H3 and H4. (a) H3.1 and H4.1 encoded in the 2.9-kb *Eco*RI fragment (pSJM230) of Ch2-3. (b) H3.2 and H4.2 encoded in the 3.0-kb *Eco*RI fragment (pSJM240) of Ch2-4. (c) H3.3 and H4.4 encoded in the 5.3-kb *Eco*RI fragment (pSJM250) of c13. The arrow indicates the direction of transcription. The underlined sequences represent the AACCCT box (see text).

was hybridized weakly only with the *Drosophila* H3 probe, but later turned out not to contain the H3 gene; the fragment did not hybridize at all with the *S. pombe* H3 gene. The remaining two clones Ch2-3 and Ch2-4 seemed to contain H3 and H4 but not H2A or H2B.

Two copies of H2A and one copy of H2B

The 1.2-kb *Hind*III fragment of Ch1-7 and the 5.5-kb *Eco*RI fragment of Ch11-6 were subcloned into pBR322 (pSJM211 and pSJM220, respectively), and their nucleotide sequences were determined mainly by the dideoxy method (Figures 2 and 3).

As expected, the 1.2-kb *Hind*III fragment of pSJM211 contained a coding region for histone H2A (designated H2A.2), as shown in Figure 3b. No other open reading frame was present in the fragment. The 5.5-kb *Eco*RI fragment of pSJM220 (Figure 2) contained coding sequences and their predicted amino acids were highly homologous to histones H2A and H2B (Figure 3a), and were designated H2A.1 and H2B.1, respectively. The coding sequences resided at the middle of *Eco*RI fragment and their directions for transcription were divergent. The two coding regions were separated by the 554-bp flanking sequence. Genomic DNA fragments hybridized with *S. pombe* H2A.2 (pSJM211) were recloned from the Charon 4A library and were found to be identical to both of the two clones described above. Therefore, *S. pombe* appears to have two H2A genes.

An H2B-specific probe was made (pSJM221; a 2.5 kb *PstI-EcoRI* fragment derived from the 5.5-kb *EcoRI* fragment and inserted into pBR322, Figure 2). It did not hybridize with any restriction fragment of Ch1-7, indicating that H2A.2 was not paired with H2B. The H2A.2-coding sequence resided in the middle of the 8.6-kb *EcoRI* fragment of Ch1-7 (pSJM210,

Figure 2), and even if another H2B copy was present, it would not exist within 3 kb from the two ends of H2A.2. To determine the copy number of the H2B gene, a genomic Southern blot was done using pSJM221 as the probe. Only a single hybridized band was observed at 5.5 kb for the genomic DNA restricted with *Eco*RI (Figure 1b; lane 2); Thus, the fission yeast has only one copy of the H2B gene paired with one of the two H2A genes.

Three pairs of H3 and H4 genes

The 2.9-kb *Eco*RI fragment of Ch2-3 that was presumed to contain the sequences for H3 and H4 was subcloned into pBR322 (designated pSJM230, Figure 2), and its nucleotide sequence was determined by the dideoxy method (Figure 4a). The two coding regions flanked by a 433-bp sequence were found, encoding H3.1 and H4.1. The directions of transcription were divergent. Similarly, a 3.0-kb *Eco*RI fragment of Ch2-4 subcloned into pBR322 (pSJM340, Figure 2) encoded H3.2 and H4.2 separated by a 335-bp flanking sequence (Figure 4b). Nucleotide sequence homologies were 89% between the coding regions of H3.1 and H3.2, and 91% between H4.1 and H4.2. Flanking sequences, however, were mostly non-homologous (see below).

0.8-kb *PstI-Eco*RI (pSJM232) and 1.4-kb *Hind*III-*PstI* (pSJM2321) subfragments of the 2.9-kb *Eco*RI fragment were made and used as specific probes for H3 and H4, respectively (Figure 2). Genomic Southern hybridization patterns employing either of these two probes were identical, showing three sets of hybridizing bands at 2.9, 3.0 and 5.4 kb in the *Eco*RI digest (data not shown). The 2.9- and 3.0-kb bands corresponded to H3.1-H4.1 and H3.2-H4.2 pairs, respectively (Figure 2). The remaining band at 5.3 kb was equally intense, suggesting the



Fig. 5. Predicted amino acid sequences of *S. pombe* histones. Amino acids are indicated by single letters: A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; Y, Tyr. (a) H2A: H2A.1 (131 residues, upper line) and H2A.2 (130 residues, lower line) with two replacements and one deletion (indicated by -) near the COOH end. H2B: H2B.1 (125 residues). H3: H3.1, H3.2 and H3.3 (135 residues) identical in sequence. H4: H4.1, H4.2 and H4.3 (102 residues) identical in sequence. Amino acid replacements in comparison with human and *S. cerevisiae* histones are shown by the overlined letters (replaced in human) or underlined letters (replaced in *S. cerevisiae*). To simplify the figure, alterations at the NH₂ and COOH termini of H2A and H2B that include deletions and additions are shown in (b). (b) Comparison of the NH₂- and COOH-terminal sequences of H2A and H2B with human and *S. cerevisiae*. Identical residues are boxed.

presence of an additional H3-H4 pair in the genome of S. pombe.

Attempts were made to clone the presumed third pair of H3-H4 from the Charon 4A library by using H3.1-containing pSJM232 as the probe, but these were not successful. Then, a cosmid library of *S. pombe* genomic DNA was screened with the same probe and 11 positives were obtained. Among those, two had the same restriction fragments as those of Ch2-3 (containing H3.1-H4.1), four had the H3.2-H4.2 fragments and the remaining five clones had the 5.3-kb *Eco*RI fragment hybridizing with both H3 and H4 probes. The reason for the failure to clone the 5.3-kb *Eco*RI fragment by the Charon vector is not understood.

The 5.3-kb *Eco*RI fragment in a cosmid clone (c13) was subcloned into pUC18 (pSJM250, Figure 2). Nucleotide sequence determination showed that the fragment contained a pair of H3 and H4 (designated H3.3 and H4.3, respectively) as shown in Figure 4c. The coding regions were separated by a 237-bp flanking sequence and were divergently transcribed. About 90% of the sequences in the coding regions were homologous among the three paired H3-H4 genes, but most of the upstream and downstream sequences were non-homologous. Thus, the genome of *S. pombe* has three pairs of H3 and H4 genes all of which have been cloned and their nucleotide sequence determined.

Predicted amino acid sequences of S. pombe histones

Predicted amino acid sequences of *S. pombe* histones are shown in Figure 5a by single letter designation. H2A.1 and H2A.2 contain 131 and 130 residues, respectively, and differ at three positions near the carboxy end, one deletion and two replacements $(T \rightarrow Q, R \rightarrow K)$. The single copy of H2B.1 has 125 residues, identical in number to human H2B and five residues less than *S. cerevisiae* H2B. All of the three H3s have 135 residues and have the same amino acid sequences. Similarly, three H4s contain 102 residues and are identical. S. pombe H2A.1 differs from S. cerevisiae and human H2A at 22 and 28 positions, respectively. Variable residues are clustered near the ends of H2A polypeptides (Figure 5b). The internal region is highly conserved. Homology between S. pombe and S. cerevisiae H2A is 83%, whereas homology between S. pombe and human is 79%. S. pombe H2B differs greatly from human H2B in the NH₂-terminal region; the first 24 residues are least homologous. On the other hand, the 13 NH₂-terminal residues are identical to the 2nd to 14th residues of S. cerevisiae. A variable cluster between S. pombe and S. cerevisiae is found in the positions between 18 and 25. The other regions of H2Bs are highly conserved. Hence, S. pombe H2B.1 is more homologous to that of S. cerevisiae (82%) than that of human (68%).

S. pombe H3 and H4 are highly conserved, compared with those of other organisms. They are identical to S. cerevisiae and human H3 and H4 in polypeptide length. Variable amino acid residues were scarce. Compared with S. cerevisiae and human H3, 11 and 10 replacements are found, respectively. In H4, nine replacements of each are found. Homology among the three organisms is >90% in H3 and H4. Thus, H3 and H4 of S. pombe appear to be equidistantly related to those of S. cerevisiae and human.

A common upstream sequence, the AACCCT box

By computer sequence analyses, we found a 17-bp long homologous nucleotide sequence (designated the AACCCT box) commonly present in the flanking regions of all the five *S. pombe* histone clones. Their locations are underlined in Figures 3 and 4 and are summarized in Figure 6. In the case of H2A.2, the common sequence is located \sim 130 bp upstream from the initiation codon. In the pair of H2A.1 and H2B.1, the AACCCT box is present in the intergene spacer sequence, \sim 380 bp upstream from the initiation codon of H2A.1. In all three pairs of H3-H4,



b AACCCT box

а

H2B.1 H4.1 H3.2 H3.3	TGAATGCCCATAAACATCC GGTTATATACAACTTGGGC CTATTTGTTTTGAATCTGC GCTATATATACCTCAGTCA GTCCCTACATCCAACTTTG	АТСАЛААСССТААСССТ АТСАСААСССТААСССТ АТСАЛААСССТААСССТ АТСАЛААСССТААСССТ АТСАСААСССТААСССТ АТСАСААСССТААСССТ	GTATATTGCTGTATCGAAAGA GGTGTTTAGCATGCGTCGCGG GCTTTTGATGAGATCGGGTGG GATTTAAGATTGCGATTTCGG TTAATTAAATTGACGTAAGTT	H2A. H2A. H3.1 H4.2 H4.3
		•		

c ATCA^CAACCCTAACCCTG

Fig. 6. Homologous upstream sequences (AACCCT box) in *S. pombe* histone genes. (a) Locations of the 17 bp long homologous sequence in the 5' upstream region of H2A.2 and in the intergene spacer region of the four, paired histone genes. The small filled arrow indicates the position and direction of the AACCCT box. Coding regions for each histone gene are shown by the boxed arrows. (b) Nucleotide sequences of the AACCCT box and surrounding regions. There is only one alteration at the 5th position. The consensus sequence is shown below. The 17-bp sequence is represented by larger letters. In H2A.2, H2A.1-H2B.1 and H3.1-H4.1, the consensus sequence extends to 19 bp long (shown by the small letters).

the box was also found in the intergene spacer regions, located ~ 170 bp, 240bp and 130 bp upstream from the initiation codons of H3.1, H4.2 and H4.3, respectively.

The AACCCT box was AC (TG) rich (82%), and had a consensus sequence of ATCAC(A)AACCCTAACCCT with a C or A substitution at the 5th position (Figure 6). The remaining 16 nucleotides were identical in the five sequences. In H2A.2 and two pairs of H2A.1-H2B.1 and H3.1-H4.1, the homologous sequences extended to a length of 19 bp. The AACCCT box had no palindrome but contained internal 6-bp long direct repeats. The repeating element AACCCT is arranged tandemly without interruption and is perfectly conserved.

As will be described in the Discussion, the AACCCT box is located between two initiation sites of divergent transcription in the paired H3.1-H4.1 genes and may play an important role in the regulation of histone gene expression.

Discussion

The histone gene organization of the fission yeast is not symmetrical in gene copy number. The genome has a pair of H2A-H2B, three pairs of H3-H4 and an unpaired, single H2A gene. In total, there is one copy of H2B, two of H2A and three each of H3 and H4. These are all the histone genes isolated on the basis of sequence homologies to the probes used, initially S. cerevisiae H2A, then Drosophila and lastly S. pombe histone genes cloned in this study. Thus, the histone gene organization of fission yeast is distinct from that of S. cerevisiae which consists of two sets each of the paired H2A-H2B and H3-H4. The presence of the histone H1 gene is not certain. We constructed the Drosophila H1 probe and obtained a few weakly positive clones, but the nucleotide sequences of hybridized restriction fragments did not have a coding frame corresponding to H1. An H1-like protein may possibly be present in S. pombe, but its encoding gene may not be cloned by hybridization with a probe of low sequence homology. The significance of the non-assorted histone gene organization of fission yeast is not clear, but a number of explanations are possible. For example, more gene



Fig. 7. Subcloning of *Drosophila* histone gene repeating unit. pKSS100 that contains an insert of the 4.8-kb *Bam*HI five histone gene sequence (Lifton *et al.*, 1977) was restricted with *HpaI/AvaI*, *AvaI*, or *AvaI/BgII*. The resulting fragments were subcloned into pBR322 and designated pSJM102, pSJM103 and pSJM104.

products of H3 and H4 than that of H2B gene may be required in the fission yeast. Multi-copied histone variants may be expressed differently or have different functional roles. Alternatively, more histone H2A and H2B genes may be present which were not cloned on the basis of sequence homology. Recently, *S. pombe* genomic DNA, of which the predicted amino acid sequence is 56% homologous to H2A and highly homologous to chicken H2A.F, was found (J. Hindley and P. Nurse, personal communication). Such variants, which might also be present for H2B, would give structural and functional diversity to nucleosomes of fission yeast.

In Northern blots, the H2B.1-probe (pSJM221) showed a single hybridizing mRNA, whereas the H2A.2-probe (pSJM211) showed two mRNA bands, indicating that H2A.1, H2A.2 and H2B.1 were all transcribed (unpublished result). When the probe (pSJM220) containing H2A.1 and H2B.1 was used, the band intensity for H2B.1 was much higher than those of H2A. The amount of H2B.1 mRNA may be high so as to balance the total amount of H2A transcripts. The probe of the H3.1-H4.1 pair showed multiple hybridizing mRNA bands. In collaboration with P. Nurse, we found that these histone transcripts are synthesized highly synchronously during the S phase (Aves *et al.*, 1985).

Transcriptional regulation of histone genes during the cell cycle has been extensively studied in S. cerevisiae (Hereford et al., 1981, 1982; Osley and Hereford, 1982). Using fusions of one of the H2A genes to the bacterial gene encoding β -galactosidase, a DNA segment required for periodic transcription of the H2A gene was identified in the 3'-flanking region of H2B. This DNA sequence appears to coincide with an ARS sequence that supports autonomous replication, indicating a close physical connection between DNA replication and histone gene transcription. To examine whether cloned S. pombe histone gene sequences have ARS, the 1.2-kb HindIII fragment containing H2A.2 and the 1.5-kb XbaI-PstI fragment containing H2B.1 were ligated with YIp33 (Botstein et al., 1979), and the resulting plasmids were used for transformation of S. pombe. The high frequency transformation by these plasmids indicated the presence of ARSs, although their exact locations remain to be determined.

All of the *S. pombe* histone clones contain the highly homologous sequence in the 5'-flanking or the inter-gene spacer regions. The consensus sequence is ATCAC(A)AACCCTAACCCT (abbreviated as AACCCT box). This 17 bp long sequence is located 100-380 bp upstream from the initiation codon (ATG) of each histone gene. S1 mapping experiments of the H3.1-H4.1 pair indicated that the initiation sites for their transcription were 40-65 bp downstream from the AACCCT box (Y. Asakura, unpublished result). TATA-like sequences (CAAATA) are present between the box and the initiation sites. Therefore, the AACCCT box seems not to be transcribed, but is proximate to putative RNA polymerase binding sites. The nucleotide sequence

of the AACCCT box is AC(TG) rich, has no palindrome but has internal direct repeats of the 6-bp element, AACCCT. This 6-bp sequence repeated twice without interruption is strictly conserved. Note that the 17-bp sequence and the 6-bp repeating element are 58 Å and 20 Å long, respectively, in the duplex DNA, lengths appropriate as binding sites for medium-sized proteins.

The existence of an homologous 5' sequence has not been reported in the histone genes of other organisms. The possibility is raised that the AACCCT box may be a species-specific sequence abundant in the genome of S. pombe. A computer search of the nucleotide sequence of several other S. pombe cloned genes including α -, β -tubulins and *cdc10*, however, did not show such a sequence homologous to the AACCCT box. Therefore, the existence of the common sequence in the upstream of S. pombe histone genes is not coincidental and may well play an important role in transcriptional regulation. The sequence is possibly the binding site for a specific protein and may function as a structural basis for the concerted transcriptional behavior of a set of the histone genes during the S phase of the cell cycle.

Comparing the amino acid sequences of S. pombe histones with S. cerevisiae and human, striking homology between S. pombe and S. cerevisiae is found at the carboxy end of H2A and at the amino end of H2B, in contrast to low homology to human (Figure 5). The last four residues of H2A are SQEL for the two yeasts but are AKGK for human. The first 13 residues of S. pombe H2B are identical to the 2nd to 14th residues of S. cerevisiae H2B but are least homologous to human. Because in the other parts the three organisms differ to similar extents, the common yeast sequences localized at the polypeptide ends of H2A and H2B may suggest a role specific for the yeast histones or the nucleosome structures, for example, the short length of the yeast nucleosomal DNA repeat (Thomas and Furber, 1976).

Materials and methods

Strains

A wild-type S. pombe 972 h⁻ (Gutz et al., 1974) was used for isolation of genomic DNA. The following E. coli strains were used: HB101 and KH802 for transformation and preparations of plasmids; LE392 for preparation of Charon 4A; BHB2688 and BHB2690 for preparation of packaging extracts. Plasmids containing S. cerevisiae histone H2A and H2B genes were used as the probes (E, E', and G; Hereford et al., 1979). Plasmid pKSS100 containing a D. melanogaster histone gene repeating unit was a gift from Dr. K. Saigo (Lifton et al., 1977; Saigo et al., 1981). pKSS100 consists of pBR322 and a 4.8-kb BamHI fragment containing a whole repeating unit of Drosophila histone genes (Figure 7). The plasmid was restricted with HpaI, AvaI and Bg/II which produced distinct fragments containing specific histone coding sequences: 0.9-kb HpaI-AvaI for H3; 1.0-kb Aval for H4 and a COOH domain of H2A; 1.2-kb Aval-BglII for H2B and a NH₂-terminal domain of H2A (Figure 7). These fragments were extracted from agarose gel and each ligated with pBR322. The resulting plasmids were used as the specific probes for histone genes.

Charon 4A library for S pombe genomic DNA

S. pombe genomic DNA was prepared according to Cryer et al. (1975) with modifications. Isolated DNA was dialyzed against TE buffer, digested with DNasefree RNase and extracted with phenol. This procedure was repeated at least twice, and DNA was precipitated with ethanol. DNA fragments produced by partial EcoRI digestion were centrifuged by linear sucrose gradient centrifugation and were recovered from the fractions containing 15-22 kb fragments. Charon 4A phage particles were purified (Yamamoto et al., 1970) and DNA was extracted and digested with EcoRI. The arms of the vector were purified by electrophoresis through low-melting-temperature agarose gel and ligated with the genomic DNA partially cleaved with EcoRI. Subsequent in vitro packaging was carried out as described by Hohn and Murray (1977). The cosmid library for S. pombe genomic DNA was constructed by Y. Nakaseko and was a gift from him.

Plaque hybridizations

The standard procedures described by Maniatis et al. (1982) were followed. Plasmids used as the probes were ³²P-labeled by nick-translation; the plasmids E and E' of the S. cerevisiae histone H2A gene (Hereford et al., 1979) were hybridized at 50°C in 6 × SSC. pKSS100 containing a repeat unit of five histone genes of D. melanogaster was also hybridized at 50°C in 6 × SSC. For the probes containing S. pombe histone H2A, H2B, H3 or H4, the inserted fragments were isolated from the plasmids and hybridized at 65°C in 6 \times SSC.

Gel transfer hybridization and nucleotide sequence determination

Procedures for gel transfer hybridization were those described by Maniatis et al. (1982). Nucleotide sequence determination was done by the Maxam-Gilbert method (Maxam and Gilbert, 1980) and by the dideoxy method (Sanger et al., 1977) using M13 (Messing et al., 1981) or pUC plasmids (Yanisch-Perron et al., 1985).

Acknowledgements

We thank Drs. L. Hereford, M. Osley, and K. Saigo for plasmids, Drs. K. Yasuda and P. Nurse for discussions and suggestions. This work was supported by grants from the Ministry of Education, Science and Culture of Japan.

References

- Aves, S.J., Durkacz, B.W., Carr, A. and Nurse, P. (1985) EMBO J., 4, 457-463.
- Beach, D., Durkacz, B. and Nurse, P. (1982) Nature, 300, 706-709.
- Botstein, D., Falco, S.C., Stewart, S.E., Brennan, M., Scherer, S., Stinchcomb, D.T., Struhl, K. and Davis, R.W. (1979) Gene, 8, 17-24.
- Choe, J., Kolodrubetz, D. and Grunstein, M. (1982) Proc. Natl. Acad. Sci. USA, 79, 1484-1487.
- Cryer, D., Eccleshall, R. and Marmur, J. (1975) in Prescott, D.M. (ed.), Methods in Cell Biology, Vol. 12, Academic Press, NY, pp. 39-44.
- Gutz, H., Heslot, H., Leupold, U. and Loprieno, N. (1974) in King, R.C. (ed.), Handbook of Genetics, Vol. 1, Plenum Press, NY, pp. 395-446.
- Hentschel, C.C. and Birnstiel, M.L. (1981) Cell, 25, 301-313.
- Hereford, L., Fahrner, K., Woolford, J., Jr. and Rosbash, M. (1979) Cell, 18, 1261-1271.
- Hereford, L.M., Osley, M.A., Ludwig, J.R., II. and McLaughlin, C.S. (1981) Cell, 24, 367-375.
- Hereford, L., Bromley, S. and Osley, M.A. (1982) Cell, 30, 305-310.
- Hiraoka, Y., Toda, T. and Yanagida, M. (1984) Cell, 39, 349-358.
- Hohn, B. and Murray, K. (1977) Proc. Natl. Acad. Sci. USA, 74, 3259-3263. Kedes, L.H. (1979) Annu. Rev. Biochem., 48, 837-870.
- Kohli, J., Hottinger, H., Muntz, P., Strauss, A. and Thuriaux, P. (1977) Genetics, 87, 471-489.
- Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977) Cold Spring Harbor Symp. Quant. Biol., 42, 1047-1051.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual published by Cold Spring Harbor Laboratory Press, NY. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol., 65, 499-560.
- Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res., 9, 309-321. Mitchison, J.M. (1970) in Prescott, D.M. (ed.), Methods in Cell Physiology, Vol.
- 4, Academic Press, NY, pp. 131-165. Mitchison, J.M. and Carter, B.L.A. (1975) in Prescott, D.M. (ed.), Methods in Cell Biology, Vol. 11, Academic Press, NY, pp. 201-219.
- Nurse, P., Thuriaux, P. and Nasmyth, K. (1976) Mol. Gen. Genet., 146, 167-178. Old, R.W. and Woodland, H.R. (1984) Cell, 38, 624-626.
- Rykowski, M.C., Wallis, J.W., Choe, J. and Grunstein, M. (1981) Cell, 25, 477-487.
- Saigo, K., Millstein, L. and Thomas, C.A., Jr. (1981) Cold Spring Harbor Symp. Quant. Biol., 45, 815-827.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
- Smith, M.M. and Andresson, O.S. (1983) J. Mol. Biol., 169, 669-690.
- Smith, M.M. and Murray, K. (1983) J. Mol. Biol., 169, 641-661.
- Thomas, J.O. and Furber, V. (1976) FEBS Lett., 66, 274-280.
- Toda, T., Yamamoto, M. and Yanagida, M. (1981) J. Cell Sci., 52, 271-287.
- Umesono, K., Hiraoka, Y., Toda, T. and Yanagida, M. (1983) Curr. Genet., 7, 123-128.
- Wallis, J.W., Hereford, L. and Grunstein, M. (1980) Cell, 2, 799-805.
- Woudt, L.P., Pastinak, A., Kempes-Veenstra, A.E., Jansen, A.E.M., Mager, W.H. and Planta, R.J. (1983) Nucleic Acids Res., 11, 5347-5360.
- Yamamoto, K.R., Alberts, B.M., Benzinger, R., Lawhorne, L. and Treiber, C. (1970) Virology, 40, 734-744.
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene, 33, 103-119.

Received on 4 September 1985