Nucleic Acids Research, Vol. 18, No. 6 © *1990 Oxford University Press*

Isolation and characterization of a *Drosophila hydei* histone DNA repeat unit

Hannie Kremer and Wolfgang Hennig
Department of Molecular and Developmental Genetics, Katholieke Universiteit, Toernooiveld, Nijmegen, NL-6525 ED, The Netherlands

Received October 13, 1989; Revised and Accepted February 14, 1990 EMBL accession no. X17072

ABSTRACT

Histone genes in *D. hydei* are organized in tandemly repeated clusters., accomodating in total 120–140 repeat units. We cloned one of the repeat units and analysed the nucleotide sequence. The repeat unit has a size of 5.1 x 10^3 base-pairs and contains one copy of each of the genes coding for the core histones and one copy coding for the histone H1. In the promoter regions of the genes we identified the presumptive cap sites and TATA boxes. Two additional sequence elements are shared by all five *Drosophila hydei* histone genes in the cluster. The sequence CCCTCT/G[1] is found in the region upstream of the presumptive CAP sites. The sequence element AGTGAA occurs downstream of the presumptive cap sites and is, in contrast to the promoter element, also seen in the histone genes of *Drosophila melanogaster* [1]. Cell-cycle dependent regulation of transcription of the *Drosophila* histone genes may be different from that in other eukaryotes since sequence elements involved in the regulation of cell-cycle dependent transcription are absent. Also other regulatory elements for transcription differ from those of other genes. The highly conserved H1-specific promoter sequence AAACACA and the H2B specific promoter sequence ATTTGCAT, which are involved in the cell-cycle dependent transcription of those histone genes in eukaryotes, are missing in the *Drosophila* genes. However at the 3′ end of the genes the palindrome and the purine-rich region, both conserved sequence elements in histone genes of eukaryotes, are present. The spacer regions show a simple sequence organization. The silent site substitution rate between the coding regions of the *D. hydei* and *D. melanogaster* histone genes is at least 1.5 times higher for *Drosophila* than for sea urchin histone genes.

INTRODUCTION

The genes coding for the highly conserved histones constitute a multigene family. Their structure and organization in the genome have been studied in a variety of species. In higher eukaryotes histone genes generally occur clustered in the genme (for review see [2]). Two types of clustering have been identified. In the first type the genes are arranged in random clusters dispersed throughout the genome. Examples are chicken [3], mouse [4] and man [5, 6]. In the second type of arrangement histone genes are organized as tandemly repeated clusters with each cluster containing all genes for the core histones and for H1. This has been found, for example, for the genes coding for the early histones in sea urchin (for review see [7]), *Notophthalmus* [8] and *Drosophila* [9].

In *Drosophila melanogaster* histone gene organization is known in detail. The majority of the histone genes, about 100 copies per haploid genome, exists in two types of repeat units called L (5.0 kb) and S (4.8 kb). They are localized in region 39 D-E of chromosome 2 [9, 10]. The two types of clusters differ by an insertion of a 240 bp tRNA-derived element in the H1–H3 spacer region of the L unit 1, [9]. Most of the two repeat types occur not intermingled [11]. In addition to the tandem repeats, orphon genes of H3 and H2B may be present [12].

In *Drosophila virilis* histone gene organization seems to be different from that in *D. melanogaster*. There are indications of a type of cluster without an H1 gene [13]. Furthermore, histone genes are located at two different sites as shown by in situ hybridization on polytene chromosomes [14].

We initiated the analysis of the *D. hydei* histone genes since we have evidence for testis-specific histone variants ([15], and unpublished data). We cloned a 5.1 kb histone repeat unit from an Eco RI genomic library. A 5.1 kb EcoRI band is one of the main restriction fragments arising after hybridization with a *D. hydei* histone gene probe.

The 5.1 kb Eco RI fragment was sequenced. It contains the genes coding for the core histones and H1. Sequence analyses of coding regions, promoter regions and the remaining spacer regions including comparisons to the recently published sequence of a *D. melanogaster* repeat unit of the L type [1] are presented. The results of our comparison of the DNA sequences of both species allow to conclude that some of the regulatory elements for transcription of the *Drosophila* histone genes differ from those of other eukaryotes.

MATERIALS AND METHODS

Drosophila strains

The wild-type strain of *D. hydei* from our laboratory collection was used.

Isolation of nucleic acids

For small-scale DNA extraction from flies we followed the procedure described by Huijser and Hennig [16]. DNA isolation

of recombinant phages and plasmids were done according to standard protocols as described by Maniatis *et al.* [17].

Hybridization of nucleic acids

Genomic DNA fragments from restriction digests were separated on horizontal 0.8% agarose gels according to Maniatis et al. [17].

DNA blotting and hybridizations were done on Hybond N+ membranes (Amersham) according to the Amersham protocols [18].

Isolation of clone pDh6/7

A genomic library of *D. hydei* wild type DNA, partially digested with EcoRI and cloned in vector λ641 was probed with nicktranslated DNA of clone p604 [19] containing the genes coding for H1 and the core histones of *D. melanogaster*. One of the selected *D. hydei* clones was designated λ Dh6/7 and studied in detail. The 5.1 kb insert of this clone was recloned in pBR328 and designated as pDh6/7. Screening, subcloning and isolation of recombinant DNA were performed according to protocols of Maniatis *et al.* [17]. A second *D. hydei* clone was designated λDh6/8. It has an insert of 2.3kb. The insert was recloned in pBR328 and designated as pDh6/8.

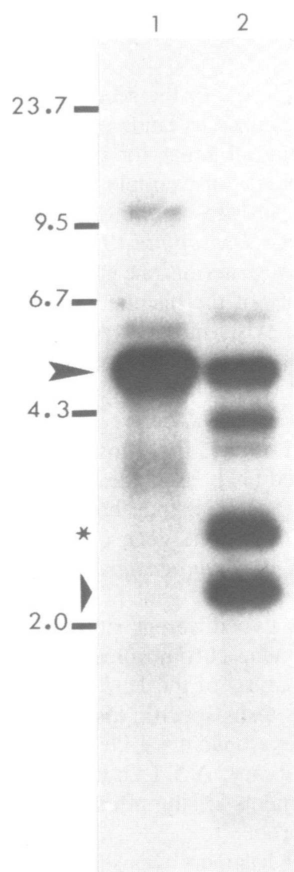

Fig. 1. Hybridization of pDh6/7 insert DNA to blots of BamHI- (lane 1) and EcoRI- (lane 2) digested genomic DNA of male flies. With genomic DNA of female flies the same results were obtained (data not shown). The molecular weight marker was HindIII- digested lambda DNA. large arrowhead, 5.1 kb; asterisk, 2.9 kb; small arrowhead, 2.3 kb.

DNA sequencing

Cloning of selected as well as random restriction fragments of insert DNA of the clone pDhH6/7 in M13mp18-19 [20, 21] and isolation of single-stranded templates were performed according to the Amersham protocols [22]. The nucleotide sequences were determined by the dideoxy chain termination method of Sanger [23]. In the last steps of sequencing oligonucleotide primers synthesized according to known parts of the sequence were used. Parts of the sequence were determined from denatured M13 RF DNA of the subclones using T7 DNA polymerase according to protocols of Pharmacia. The DNA sequence was determined for both strands.

DNA sequence analysis

For assembly of gel readings the computer program of Staden [24] was used. The GCG sequence analysis software package version 5, 1987 (University of Wisconsin, Biotechnology Center), was used for further sequence analyses.

RESULTS AND DISCUSSION

Structure and organization of the histone gene repeat unit

The *D. hydei* histone genes mainly occur in tandemly organized repeat units of 5.1 kb in length as is shown in Bam HI digested genomic DNA hybridized with pDh 6/7 insert DNA (fig. 1, lane 1). EcoRI digested genomic DNA (fig. 1, lane 2) shows heterogeneity within the repeats. The 5.1 kb band represents one of the main bands. The 2.3 kb and 2.9 kb fragments most likely together form one repeat unit comparable to the unit represented in clone pDh6/7. By Southern blot analysis it was determined that H1 coding sequences are present in both fragments, H2A and H2B coding sequences in the 2.3 kb fragment and coding sequences of H3 and H4 in the 2.9 kb fragment. This is in agreement with the organization of genes in clone pDh6/7 (fig 2, see below). Moreover the genomic clone pDh6/8 with an insert of 2.3 kb is in its restriction pattern (BamHI, HhaI and PstI) identical to the region of pDh6/7 containing the genes H2A, H2B and the 3′ region of H1.

The minor bands (fig. 1) might represent orphan genes or genes coding for histone variants [25] located outside the gene cluster (cf [12]), repeat units with either a deletion or with an inserted

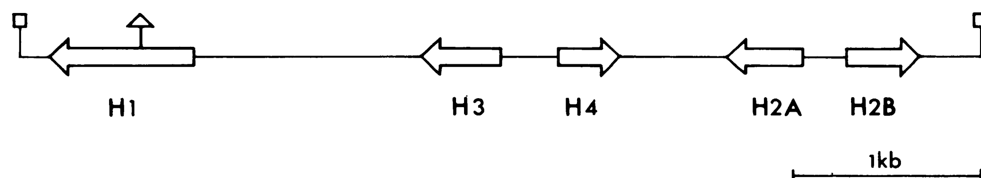

Fig. 2. Organization of the histone genes within the repeat unit pDhH6-7. The direction of transcription is given by the arrows marking the genes. BamH1, EcoRI,

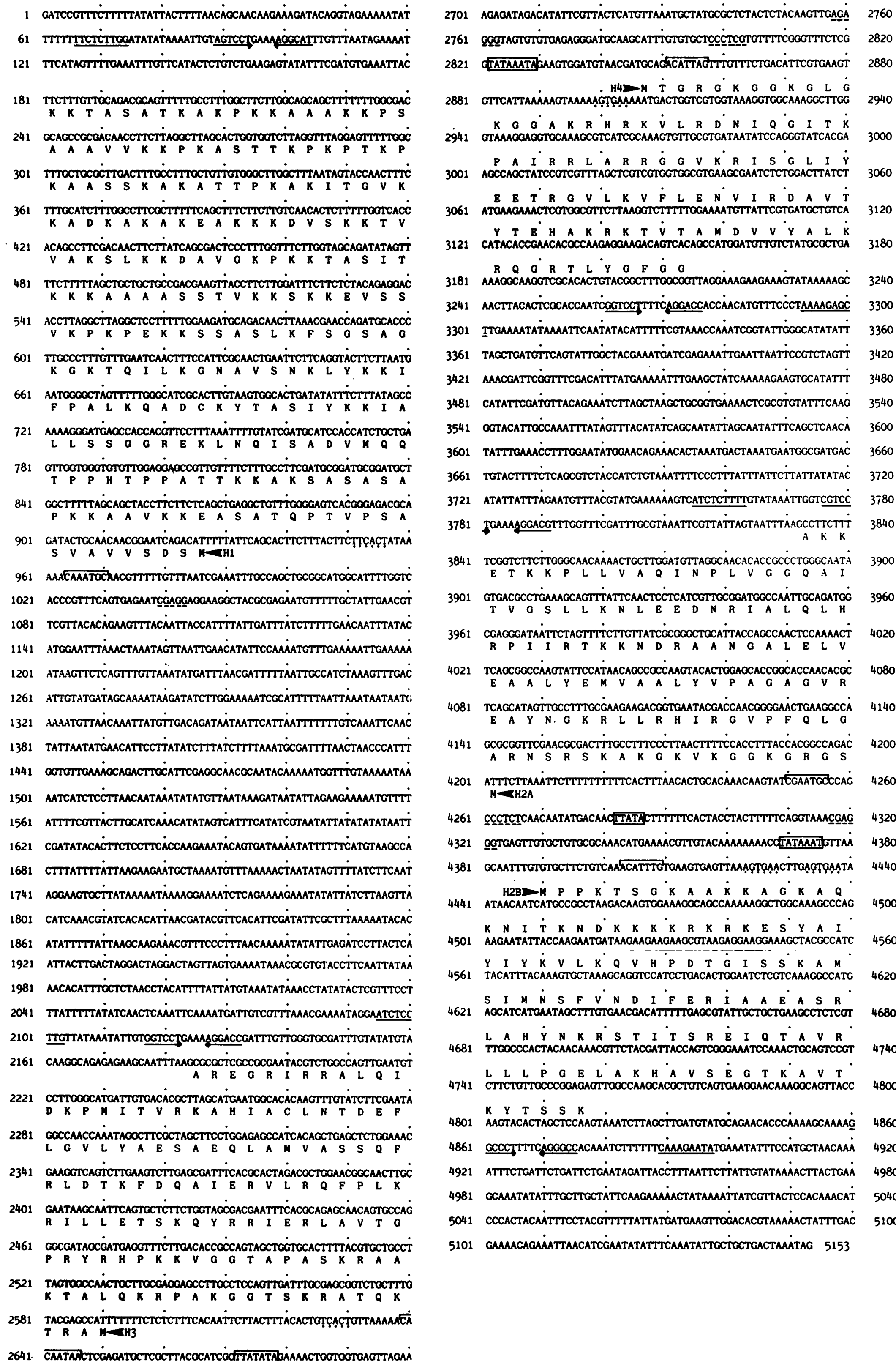

Fig. 3. Complete DNA sequence of clone pDhH6-7 and the deduced amino acid sequence of the histones. The start of the sequence is at the unique BamHI site in the H2B−H1 spacer region (see Fig. 4). The direction of transcription is marked with arrowheads. The following sequence elements are marked: cap site, ┌──┐; TATA box ▭, AGTGAA, ...; CCCTGT/G − − − −; palindrome, ←; purine-rich region, ______.

transposable element (cf [1, 13]) or fragments at the border of the tandemly arranged repeat clusters.

The copy number of the histone gene cluster per haploid genome is determined to range between 120 and 140. By *in situ* hybridization it was determined that the cluster is located on chromosome 3 in band 50 A which is in agreement with the results of Boender [26].

The general structure of the *D. hydei* histone repeat unit present in clone pDhH6/7 is shown in figure 2. The total length of the *D. hydei* repeat unit is 5153 bp (base-pairs). The sequence is presented in figure 3. After sequencing, the genes were located with aid of the *D. melanogaster* sequence of an L repeat unit [1]. The arrangement and orientation of the genes within the cluster is the same as in *D. melanogaster* [1, 9].

Upstream sequence elements

The 5′ upstream regions of the histone genes were examined for sequence elements relevant to regulation and initiation of transcription using data from the literature for upstream sequence elements of histone genes.

Cap site and TATA box: We searched upstream of the ATG codon for TATA boxes and presumptive cap sites. These elements are marked in Fig. 3. For determining the presumptive cap sites we used the consensus sequence for initiation of transcription as defined by Breathnach and Chambon [27], the consensus sequence for cap sites of *Drosophila* genes, TCAGTT/C, deduced from sequences given by Snyder *et al.*, [28], the distance between TATA box and cap site (25 to 30 bp in eukaryotic genes transcribed by RNA polymerase II [27]) and the sequence similarity between the upstream regions of the coding regions of the *D. melanogaster* histone genes [1].

The cap site of the H1 gene of *D. melanogaster* has been determined by S1 mapping at -34 ± 2 bp from the start codon [29]. A homologous sequence of 6 bp is found in the *D. hydei* H1 upstream sequence at -38 bp from the ATG codon (pos. 964). A clear TATA box is not found within 150 bp upstream of this presumptive cap site. No sequence similarity to the TATA box of *D. melanogaster* H1 gene [1], located at -82 bp from the cap site, is found in the respective region of *D. hydei*. It is not a common phenomenon that H1 genes lack a TATA box or have a TATA box at a significantly larger distance from the cap site (see 30, 31). Alternatively, a 5 bp sequence (GCAAA) (pos. 992) may serve as a TATA box. GCAAA occurs in the H1 promoter region of both *D. hydei* (-22 bp from the cap site) and *D. melanogaster* (-25 bp from the cap site).

The transcription initiation site of many histone genes is located within the consensus sequence PyCATTCPu [32–34]. This also holds true for the most likely candidates for cap site sequences of *D. hydei* and *D. melanogaster* histone genes. The presumptive cap sites resemble this sequence better than the consensus sequence for the *D. melanogaster* cap sites as given above.

Generally accurring upstream sequence elements: CAT box and Spl binding site: The CCAAT box and the Spl binding site (GGGCGG or GC box) are generally found in histone genes ([35, 36], see also [30]). Neither *D. hydei* nor *D. melanogaster* gaster histone genes contain upstream sequences identical to either the CAT or the GC boxes. However, an exact match to the CAT box and GC box sequences seems to be necessary for function [37–39]. Since the Spl and CCAAT binding factors are absent in *Drosophila* culture cells, [40] these factors are probably not general transcription factors in *Drosophila*.

Promoter elements specific for histone genes: Most histone H1 genes contain the H1-specific promoter element ([41], see also [30]). The H1 genes of both *D. hydei* and *D. melanogaster* have similar elements in their 5′ upstream regions as do the upstream regions of the core histone genes if one mismatch is accepted.

The H2B-specific octamer ATTTGCAT is generally highly conserved (see [30]). In the H2B genes of *D. hydei* and *D. melanogaster* such a octamer sequence can only be identified if 4 and 3 mismatches respectively are accepted. It is unlikely that they are funtional since mutation of two nucleotides in the octamer eliminates its function in a human H2B gene [42].

Both the H1-specific as well as the H2B-specific element are assumed to be required for cell-cycle-dependent regulation of transcription in human cells [42, 43] The lack of these two elements in the corresponding *Drosophila* histone genes that have been studied so far, suggests that factors involved in cell-cycle-dependent regulation of transcription of histone genes in *Drosophila* may be different from those in other eukaryotes.

To find promoter sequence elements shared by all five *D. hydei* histone genes we looked for perfect matches of at least 5 bp in the 5′ upstream regions. One element (CCCTCT/G)occurs at -104 to -184 (pos. 1039, 2758, 2798, 4261 and 4317) from the ATG of all the *D. hydei* histone genes (see below). In *D. melanogaster* such an element is only found upstream of the H1 gene as part of the 16 bp conserved element (see fig. 4). The functional significance of the CCCTCT/G element, needs to be analysed experimentally. Possible promoter elements are indicated below.

Conserved sequence elements in the 3′ region

In the 3′ non-protein-coding regions of cell-cycle regulated histone genes two conserved sequence blocks are found, namely a region of about 24 bp including a palindromic sequence and a purine-rich sequence 13–17 bp downstream of the palindrome ([30, 44] for review see [45]).

Both sequence elements are found in the 3′ downstream region of the *D. hydei* histone genes (fig. 3). The consensus sequence which can be deduced from the *D. hyde*i and *D. melanogaster* genes confirms the consensus given by Wells [30]. The palindrome sequences of the H1 genes of both species differ from those of the core histone genes: the position of the inverted repeats, the imperfectness of the inverted repeats and divergence of their sequences between *D. hydei* and *D. melanogaster*. The distance between the repeat and the stop codon is significantly longer in the H1 gene than in the core histone genes (73 bp in *D. hydei*, 85 bp in *D. melanogaster*). The distance in most histone genes given by Wells is about 30 bp with a maximum of 59 bp. Furthermore in both *D. melanogaster* and *D. hydei* the inverted repeat is imperfect which is also an exception (see 30). While the palindrome is perfectly conserved between the core histone genes of *D. melanogaster* and *D. hydei*, for the H1 gene 3 out of 16 nucleotides are different. Among *Drosophila* species the purine-rich region is less well conserved than in sea urchin, thus resembling the situation in vertebrates [45]. Since the purine-rich region, together with the U_7 snRNA, is necessary for 3′ end processing [46], it would be worthwhile to determine for the *Drosophila* U_7 snRNA whether the larger sequence divergence in the purine-rich region finds its complement in a longer region of complementarity to the U_7 snRNA as was found in mammals [47, 48].

For mouse histone H3 it has been shown that the palindrome sequence is involved in posttranscriptional, cell-cycle-dependent

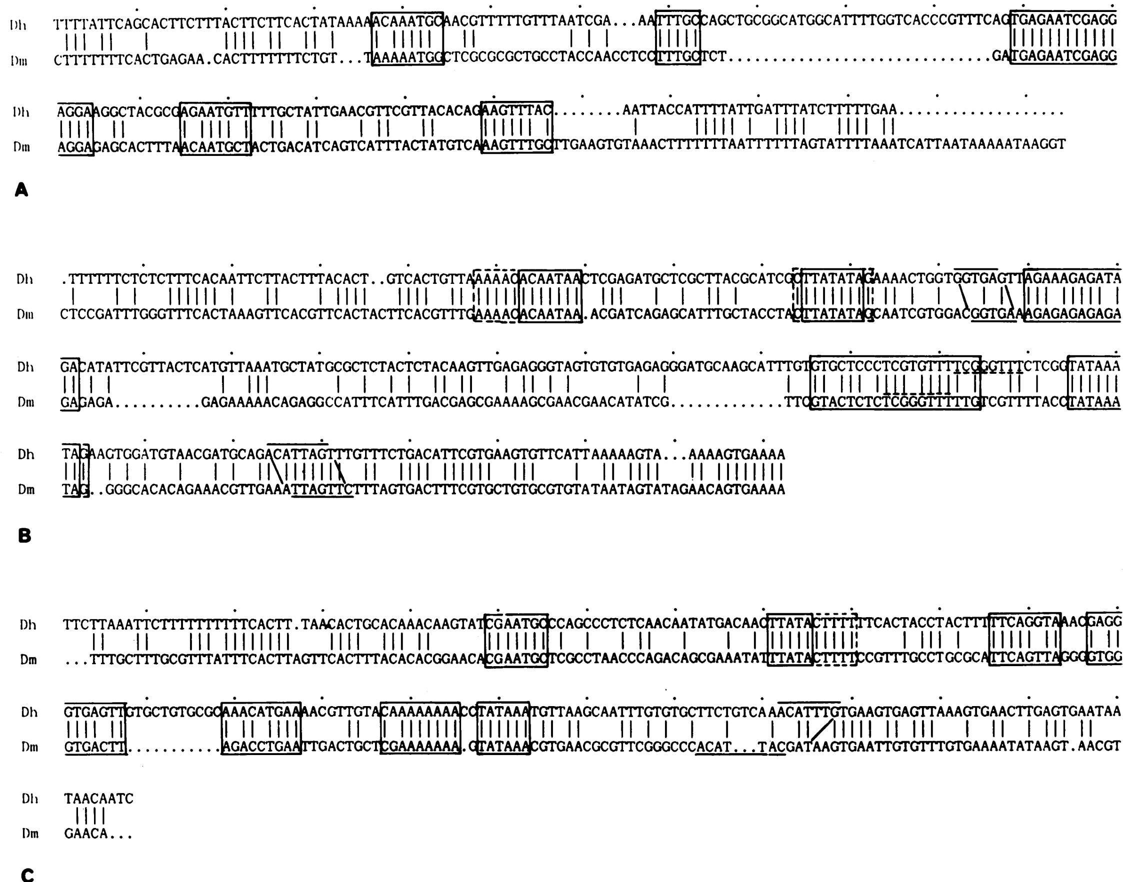

Fig. 4. Alignment of the 5′ regions of the histone genes of *D. hydei* and *D. melanogaster*. Sequences were aligned using the program GAP of the GCG programs with a gap weight of 3.0 and a length weight of 0.2. For the 5′ end of the H1 gene corrections were made. Cap site, TATA box and regions of similarity upstream of the cap site are marked. Sequences are presented in the same sorientation as in figure 1. (a) 5′ region of H1 (204 bp, pos. 927–1130 in fig. 2), (b) H3–H4 spacer region (pos. 2591–2906 in fig. 2), (c) H2A–H2B spacer region (pos. 4203–4449 in fig. 3). Gaps are indicated by dots, identical nucleotides are marked by dashes.

regulation of the mRNA level. The conservation of the 3′ downstream sequence elements suggests that this part of the cell-cycle-dependent regulation of histone mRNA levels in *Drosophila* is similar to that in vertebrates. This is in contrast to sequence diversity in the elements involved in the regulation of transcription (see above).

Structure of the spacer regions

The spacer regions of the histone repeat of *D. hydei* were screened for the occurrence of repeats, simple sequences and purine-pyrimidine stretches and they were compared with each other. The designation 'spacer region' is used for those regions of the repeat unit not coding for proteins.

All spacer regions are AT-rich. Short (3–6 bases) homopolymers of dA or dT are often found, but rarely runs of dG or dC. This is regularly observed in spacer regions of histone genes, for example in *D. melanogaster* [1], sea urchin [49–52], in *Xenopus* [35], trout [53], and in the sea star [54]. The short homopolymers of dA or dT might have a function in nucleosome positioning since homopolymeric runs longer than 4 or 5 bp have preferential rotational settings on the nucleosome supercoil [55].

Short simple repeating sequence elements of two or three nucleotides in length and longer repetitive elements are frequent in histone gene spacers ([35, 50, 53, 56]). In the histone gene spacers of *D. hydei* neither was found. No simple repeating sequence elements with more than three repeats are present. The fact that the (GA)10 element of H3-H4 spacer of *D. melanogaster* is not found in *D. hydei*, although the region in which the GA repeat starts is conserved between both species, in combination with the differences which are generally found in histone gene clusters with regard to simple sequence elements [35, 50, 51] support the idea that these elements in general have no specific function in gene expression [57]).

Alternating purine-pyrimidine stretches longer than 6 bp are also not found in the *D. hydei* histone gene spacers.

The different spacers of the *D. hydei* histone repeat show no extensive homology with each other. Analyses were done with dot matrices (window-stringency 11–8). The similarities immediately upstream and downstream of the protein coding region are somewhat higher than in the spacer regions in between.

The leader regions between the presumptive cap sites and the ATG codons of the different histone genes seem to be constructed of homologous AT-rich elements. Their positions differ however in the different leaders. Most of these elements are related to the sequence AGTGAA. AGTGAA elements are found once or twice in each of the leaders of the five histone genes in *D. hydei* (fig. 4) as well as in *D. melanogaster* [1]. Only the H3 leader in *D. hydei* shows the sequence AGTGA. Matsuo and Yamazaki [1] suggested a function of this element specific for *Drosophila* histone gene expression. This is supported by the conservation of this element in the histone genes of *D. hydei*.

At the 3′ ends of the genes homology is restricted to the palindrome and the purine-rich region (see above).

SAR: The existence of a scaffold attachment region (SAR) in the H1−H3 spacer of the *D. melanogaster* histone repeat unit has been reported [58−60]. It is suggested that SARs may border functional units in the chromatin (for review see [61]). Because of the basic similarity in organization of the histone genes we analysed the *D. hydei* histone gene cluster for the existence of a SAR with the same criteria 59, 60 . They are: (1) enrichment of topoisomerase II cleavage site consensus sequences (GTNA/TAC/TATTNATNNG) [62] (2) the occurrence of A boxes (AATAAAT/CAAA) and T boxes (TTA/TTT/ATTT/ATT). We repeated these analyses for the sequence of the *D. melanogaster* [1]. According to table 1, in *D. melanogaster* the H1−H3 spacer shows a clear enrichment of sequences homologous to the topoisomerase II cleavage site and both A and T boxes occur. In *D. hydei*, however, the enrichment in topoisomerase II cleavage sites is less prominent. In addition, both the H2B−H1 and to a lesser extent the H4−H2A spacer regions show more topoisomerase II cleavage site homologies per 100 bp than the H1 - H3 spacer region and in addition contain A and T boxes. This picture of topoisomerase II cleavage site enrichment did not change for nucleotides 1280−1990 which are in position comparable to the SAR in the *D. melanogaster* H1−H3 spacer. In contrast to the situation in *D. melanogaster*, the H2B−H1 spacer in *D. hydei*, would be the most likely candidate for having a SAR since it contains 1.4 topoisomerase II sites. This means that either the SAR in the H1−H3 spacer is not conserved in *D. hydei* or the criteria given [59, 60] are not generally applicable to detect SAR regions.

Table 1. Topoisiomerase II sites, A and T boxes.

	TopoisomeraseII sites/100 bp		A boxes /100 bp		T boxes /100 bp	
	Dh	Dm	Dh	Dm	Dh	Dm
H1 - H3	1.2	0.9*	2.2	2.6	9.3	11.4
H3 - H4	0.0	0.0	1.3	0.3	3.2	1.0
H4 - H2A	1.3	0.0	0.6	0.8	3.7	4.0
H2A - H2B	0.8	0.4	0.4	0.4	8.1	2.6
H2B - H1	1.8	0.7	1.0	2.2	6.6	7.0
H1	0.1	0.1	0.4	0.4	1.7	1.8
H3	0.0	0.0	0.5	0.0	0.7	0.0
H4	0.0	0.0	0.0	0.0	0.0	0.0
H2A	0.0	0.3	0.0	0.0	0.5	0.0
H2B	0.8	0.3	0.3	0.0	0.5	1.1

Occurrence of topoisomerase II sites (6 bp surrounding the cleavage site conserved, 70% overall similarity). A boxes and T boxes (2 mismatches allowed) in both orientations in the protein and non protein coding regions of the histone repeat unit of *D. hydei* (Dh) and *D. melanogaster* (Dm) as published by Matsuo and Yamazaki.
The number of sites are given per 100 bp. *, in contrast to Gasser and Laemmli [60] we find only 10 topoisomerase II sites in the HinfI−EcoRI fragment (including the SAR region) by using the criteria defined by the authors. They give the number of 18.

Comparison of the *D. hydei* and *D. melanogaster* histone repeat unit

The protein coding regions of the core histone genes of both species are of the same length. The H1 protein coding region of *D. melanogaster*, however, codes for a protein which is 7 amino acids longer than that of *D. hydei* (see below). Otherwise the difference in length between the repeat units of both species is due to length differences in the spacer regions, mainly in the spacers between the H1−H2B and H2−H4 genes. The high degree of length conservation of the spacers of the H2A−H2B and H3−H4 genes, which are transcribed in opposite directions, is barely reflected in sequence conservation. Also the H1−H3 spacer shows a high degree of length conservation which might be correlated to the functions of a SAR (see above).

Protein coding sequences: We compared the amino acid sequence of the histones as it can be deduced from the sequence of clone pDhH6/7 of *D. hydei* (fig. 3) with those for *D. melanogaster* [1]. The amino acid sequences of the core histones are identical (H4 and H2A) or nearly identical (H3 and H2B). In H3 there is one conservative amino acid substitution in position 117 which is a valine in *D. hydei* and an isoleucine in *D. melanogaster*. In H2B there is a nonconservative amino acid substitution from asparagine in *D. hydei* by threonine in *D. melanogaster* in the N terminal region of the protein at position 22. The N terminal region is the most variable region of H2B [63].

For all organisms H1 is the least conserved of all histones [64]. The N and C terminal parts generally show a high degree of sequence variation in comparison to the globular part which includes a central stretch of about 80 amino acids [63−66]. An alignment of the H1 amino acid sequence of both species was made (fig. 5). The amino acids 44−128 form a central highly

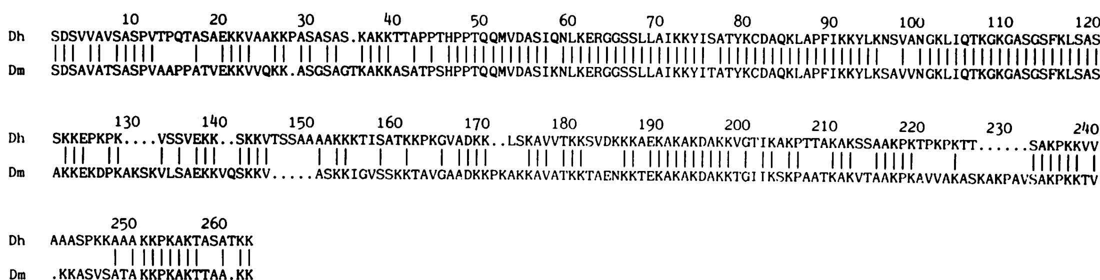

Fig. 5. Alignment of the amino acid sequence of H1 of *D. hydei* and *D. melanogaster*. Gaps are indicated by dots, identical residues are marked by dashes.

conserved region of 85 amino acids. The homology (amino acid level) in that region is 89% while in the N (43 amino acids) and C terminal part (127 and 120 amino acids in *D. hydei* and *D. melanogaster*, respectively) it is 63% and 61%, respectively. The alignment shows that the length difference of 7 amino acids between H1 of *D. hydei* and *D. melanogaster* is located in the C terminal region of the protein.

Non protein-coding regions: We analysed the conservation of the non protein coding-regions of the *D. hydei* and *D. melanogaster* [1] repeats by alignment. In general the homology between corresponding non coding regions is very limited. We found that the regions upstream (= total H2A−H2B and H3−H4 spacer, 200 bp for the H1 gene) and downstream (150 bp) of the protein coding regions are slightly more conserved than the remaining spacer sequences (about 57% and about 50% respectively, calculated by the GCG program GAP, gap weight 3.0; length weight, O.2). The 5′ regions show short perfectly matching regions of 5−10 bp (a 16 bp region for H1), in addition to the TATA boxes and the presumptive cap sites. These sequence elements are marked in the alignment (fig. 4). Because the core histone genes are divergently transcribed the promoter regions of H2A and H2B genes might be overlapping (cf. chicken H2A−H2B spacer region [67]). So we cannot assign specific elements to specific genes. The same situation exists for the H3 and H4 genes.

The H1 upstream regions are difficult to align. A conserved sequence element of 16 bp (TCCTCCTCGATTCTCA) (pos. 1032) is found which has a different location upstream of the presumptive cap site of *D. hydei* H1 in comparison to *D. melanogaster*. We marked conserved sequence elements in an alignment based on the 16 bp conserved region and the cap sites. Whether or not the marked sequence elements (fig. 4) are functional promoter elements has to be determined experimentally. Sequences of further *Drosophila* species which could give additional information are unknown.

Silent site substitution rate: Histones, with the exception of H1, are that highly conserved [64] that amino acid substitution rates can only be calculated over relatively long periods of divergence. However, the phylogenetic distance between *D. hydei* and *D. melanogaster* results in a significant amount of silent site substitutions at the DNA level. The highly repeated sea urchin histone genes, coding for H3 and H4 necessary in early embryogenesis, evolved at the silent sites with a constant rate of 0.5−0.6% base changes-Myr (million years) [68] (method: [69]). We used the same method (GCG Sequence Analysis Software Package) to calculate the silent site substitution rate for the *Drosophila* histone genes (table 2). The substitution rate for H4 of *Drosophila* is about 1.3−3 times higher than for sea urchin. The rate for *Drosophila* H3 is 1.7−4.2 times higher than that for sea urchin H3. The values for H2A and H2B are comparable to those for H3.

Our results indicate that the silent site substitution rate for histone genes in *Drosophila* is generally higher than in sea urchins and thus that histone genes in different organisms do not mutate with a constant rate at silent sites. This is in agreement with data, showing that the rates of silent site substitutions vary among groups of animals and among genes [70−72].

The differences in the rates of silent site substitution might result from differences in the generation times, which differ largely between *Drosophila* and sea urchin, and/or selection

Table 2. Silent site substitutions in the histone genes of *D. hydei* and *D. melanogaster*.

	% SSS	%SSS/Myr
H3	125	1.0-2.1
H4	90	0.8 1.5
H2A	132	1.1 2.2
H2B	135	1.1-2.3

The percentages of silent site substitutions were calculated using the program diverge of the GCG software (method: [69]). Because of restrictions in the program it was not possible to calculate a percentage of silent site substitutions corrected for multiple hits for the highly diverged H1 genes. The time of divergence between *D. hydei* and *D. melanogaster* is estimated to range between 30 and 60 million years [73, 74]. The two values given in column 2 represent the percentage silent site substitution per million years using the two extreme values of the divergence time. Myr, million years; sss, silent site substitutions.

pressure on codon usage as discussed by Britten [70] and Li and coworkers [71].

CONCLUSIONS

The genomic organization of the histone genes of *D. hydei* resembles closely that of *D. melanogaster*. The DNA sequence similarity in the 5′ and 3′ regions of the individual histone genes between *D. hydei* and *D. melanogaster* indicate that the regulation of transcription of the histone genes in the genus *Drosophila*, and possibly in insects in general, involves regulatory elements different from other higher eukaryotes.

ACKNOWLEDGEMENTS

We appreciate the gift of oligonucleotid primers by Dr. Renate Renkawitz-Pohl and of histone gene probes by Drs. Martin Blumenfeld, Jan Burckhardt and Eric Weinberg. Judith Mosterd is thanked for her experimental contribution, Drs. Z. Matsuo and T. Yamazaki for making their data available before publication and Dr. J. Leunissen for his advice on computer analyses. We wish to thank our colleagues Drs. Rein Brand, Johannes Hackstein, Ron Hochstenbach, Peter Huijser, Dirk-Henner Lankenau, Susanne Lankenau, Koos Miedema and Fan Wang for discussions and critical reading of the manuscript. To Wieli Janssen we are indebted for technical support.

REFERENCES

1. Matsuo, Z. and Yamazaki, T. (1989). Nucl. Acids Res. 17: 225−238.
2. Maxson, R., Cohn, R. and Kedes, L. (1983). Annu. Rev. Genet. 17: 239−277.
3. Engel, J. D. and Dodgson, J. B. (1981). Proc. Nat. Acad. Sci., U.S.A. 78: 2856−2860.
4. Sittman, D. B., Chiv, I-M., Pan, C-J, Cohn, R. H., Kedes, L. H. and Marzluff, W. F. (1981). Proc. Nat. Acad. Sci., U.S.A. 78: 4078−4082.
5. Heintz, N., Yernik, M. and Roeder, R. G. (1981). Cell, 24: 661−668.
6. Sierra, F., Lichtler, A., Marashi, F., Rickles, R., Van Dyke, T., Clark,

S., Wells, J., Stein, G. and Stein, J. (1982). Proc. Nat. Acad. Sci., U.S.A. 79: 1795–1799.
7. Kedes, L. H. (1979). Annu. Rev. Biochem. 48, 837–870.
8. Stephenson, E. C., Erba, H. P. and Gall, J. G. (1981). Cell, 24: 639–647.
9. Lifton, R. P., Goldberg, M. L., Karp, R. W. and Hogness, D. S. (1977). Cold Spring Harbor Symp. Quant. Biol. 42: 1047–1051.
10. Pardue, M. L., Kedes, L. H., Weinberg, E. S. and Birnstiel, M. L. (1977). Chromosoma, 63: 135–151.
11. Saigo, K., Millstein, L. and Thomas Jr, C. A. (1980). Cold Spring Harbor Symp. Quant. Biol. 45: 815–827.
12. Childs, G., Maxson, R., Cohn, R. H. and Kedes, L. (1981). Cell, 23: 651–663.
13. Domier, L. L., Rivard, J. J., Sabatini, L. M. and Blumenfeld, M. (1986). J. Mol. Evol. 23: 149–158.
14. Anderson, K.V. and Lengyel, J.A. (1983). In Histone Genes (Stein, G. S., Stein, J. L. and Marzluff, W. F., eds), pp 135–159, John Wiley and Sons, New York.
15. Kremer, H., Hennig, W. and Dijkhof, R. (1986). Chromosoma, 94: 147–161.
16. Huijser, P and Hennig W. (1987). Mol. Gen. Genet. 206:441–451.
17. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
18. Amersham (1985). Membrane transfer and detection methods, Amersham International, Amersham U.K.
19. Chernyshev, A. I., Bashkirov, V. N., Leibovitch, B. A. and Khesin, R. B. (1980). Mol. Gen. Genet. 178: 663–668.
20. Messing, J. and Vieira, J. (1982). Gene, 19: 269–276.
21. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Gene, 33: 103–119.
22. Amersham (1984). M13 Cloning and Sequencing Handbook, Amersham International, Amersham, U.K.
23. Sanger, F., Nicklen, S. and Coulson, A. R. (1977). Proc. Nat. Acad. Sci., U.S.A. 74: 5463–5467.
24. Staden R. (1982). Nucl. Acids Res. 10: 4731–4751.
25. van Daal, A., White, E. M., Gorovsky, M. A. and Elgin, S. C. R. (1988). Nucl. Acids Res. 16: 7487–7497.
26. Boender, P. J. (1983). Ph.D. thesis, University of Nijmegen.
27. Breathnach, R. and Chambon, P. (1981). Annu. Rev. Biochem. 50: 349–383.
28. Snyder, M., Hunkapiller, M., Zuen, D., Silvert, D., Fristrom, J. and Davidson, N. (1982). Cell, 29: 1027–1040.
29. Murphy, T. J. and Blumenfeld, M. (1986). Nucl. Acids Res. 14: 5563.
30. Wells, D. E. (1986). Nucl. Acids Res. 14 (suppl.): r119–r149.
31. Lai, Y. C. and Childs, G. (1988). Mol. Cell. Biol. 8: 1842–1844.
32. Busslinger, M., Portmann, R., Irminger, J. C. and Birnstiel, M. L. (1980). Nucl. Acids Res. 8: 957–977.
33. Hentschel, C. C., Irminger, J. C., Bucher, P. and Birnstiel, M. L. (1980). Nature (London), 285: 147–151.
34. Sures, I., Levy, S. and Kedes, L. H. (1980). Proc. Nat. Acad. Sci. U.S.A. 77: 1265–1269.
35. Perry, M., Thomsen, G. H. and Roeder, R. G. (1985). J. Mol. Biol. 185: 479–499.
36. Hentschel, C. C. and Birnstiel, M. L. (1981). Cell, 25: 301–313.
37. Dynan, W. S. and Tjian, R. (1985). Nature (London), 316: 774–778.
38. Bucher, P. and Trifonov, E. N. (1988). J. Biomol. Struct. Dyn. 5: 1231–1236.
39. Graves, B. J., Johnson, P. F. and McKnight, S. L. (1986). Cell, 44: 565–576.
40. Santoro, C., Mermod, N., Andrews, P.C., and Tjian, R. (1988). Nature 334:218–224.
41. Coles, L. S. and Wells, J. R. E. (1985). Nucl. Acids Res. 13: 585–594.
42. Labella, F., Sive, H. L., Roeder, R. G. and Heintz, N. (1988). Genes Develop. 2: 32–39.
43. Dalton, S. and Wells, J. R. E. (1988). EMBO J. 7: 49–56.
44. Busslinger, M., Portmann, R. and Birnstiel, M. L. (1979). Nucl. Acids Res. 6: 2997–3008.
45. Birnstiel, M. L., Busslinger, M. and Strub, K. (1985). Cell 41: 349–359.
46. Cotten, M., Gick, O., Vasserot, A., Schaffner, G. and Birnstiel, M. L. (1988). EMBO J. 7: 801–808.
47. Mowry, K. L. and Steitz, J. A. (1987). Science, 238: 1682–1687.
48. Soldati, D. and Schümperli, D. (1988). Mol. Cell. Biol. 8: 1518–1524.
49. Roberts, S. B., Weisser, K. E. and Childs, G. (1984). J. Mol. Biol. 174: 647–662.
50. Sures, I., Lowry, J. and Kedes, L. H. (1978). Cell, 15: 1033–1044.
51. Schaffner, W., Kunz, G. Daetwyler, H., Telford, J., Smith, H. O. and Birnstiel, M. L. (1978). Cell, 14: 655–671.
52. Kaumeyer, J. F. and Weinberg, E. S. (1986). Nucl. Acids Res. 14: 4557–4576.
53. Winkfein, R. J., Connor, W., Mezquita, J. and Dixon, G. H. (1985). J. Mol. Evol. 22: 1–19.
54. Cool, D., Banfield, D., Honda, B. M. and Smith, M. J. (1988). J. Mol. Evol. 27: 36–44.
55. Satchwell, S. C., Drew, H. R. and Travers, A. A. (1986). J. Mol. Biol. 191: 659–675.
56. Tönjes, R., Munk, K. and Doenecke, D. (1989). J. Mol Evol. 28: 200–211.
57. Tautz, D. and Renz, M. (1984). Nucl. Acids Res. 12: 4127–4138.
58. Mirkovitch, J, Mirault, M-E and Laemmli, U. K. (1984). Cell, 39: 223–232.
59. Gasser, S. M. and Laemmli, U. K. (1986a). Cell, 46: 521–530.
60. Gasser, S. M. and Laemmli, U. K. (1986b). EMBO J. 5: 511–518.
61. Gasser, S. M. and Laemmli, U. K. (1987). Trends Genet. 3: 16–22.
62. Sander, M. and Hsieh, T-S. (1985). Nucl. Acids Res. 13: 1057–1072.
63. Isenberg, I. (1979). Annu. Rev. Biochem. 48: 159–191.
64. Elgin, S. C. R. and Weintraub, H. (1975). Annu. Rev. Biochem. 44: 725–774.
65. Cole, R. D. (1984). Anal. Biochem. 136: 24–30.
66. Von Holt, C., Strickland, W. N., Brandt, W. F. and Strickland, M. S. (1979). FEBS Letters, 100: 201–218.
67. Sturm, R. A., Dalton, S. and Wells, J. R. E. (1988). Nucl. Acids Res. 16: 8571–8586.
68. Busslinger, M., Rusconi, S. and Birnstiel, M. L. (1982). EMBO J. 1: 27–33.
69. Perler F., Efstratiadis A., Lomedico P., Gilbert W., Kolodner R. and Dodgson J. (1980) Cell, 20: 555–566.
70. Britten, R. J. (1986). Science, 231: 1393–1398.
71. Li, W-H, Tanimura, M., Sharp, P. M. (1987). J. Mol. Evol. 25: 330–342.
72. Sharp, P. M., Li W-H. (1989). J. Mol. Evol. 28: 398–402.
73. Beverly, S. M. and Wilson, A. C. (1984). J. Mol. Evol. 21: 1–13.
74. Throckmorton, L. H. (1975). In Handbook of Genetics (King, R. C., ed), pp. 421–469.

[1] Sequences in the text are given in the 5′–3′ direction.

[2] All positions given in the text relate to Figure 3. We refer always to the position of the first nucleotide of a sequence element relative to the beginning of the sequence of the repeat unit. Distances between two sequence elements are given as the number of nucleotides from the last nucleotide of one sequence element to the first nucleotide of the second element relative to the beginning of the sequence as shown in Figure 3.