Characterizaton of the promoter region of *Tetrahymena* genes

Clifford F.Brunk and Lori A.Sadler
Biology Department and Molecular Biology Institute, University of California, Los Angeles, CA 90024-1606, USA

Received September 19, 1989; Revised and Accepted December 11, 1989

ABSTRACT

The regions between adjacent histone H3 and H4 genes, as well as portions of the genes, from 22 species of *Tetrahymena* have been amplified using the polymerase chain reaction and sequenced. Both histone genes are transcribed divergently with initiation occurring within the intergenic region, thus 2 sets of 22 homologous *Tetrahymena* promoters can be compared. A sequence comparison of these regions reveals a single putative promoter element, with a consensus sequence TATCCAATTCARA, present in front of each gene. This sequence contains a 'CCAAT' box, which also occurs at 8 locations preceeding other ciliate genes. No other putative promoter sequences are found in front of these sets of histone genes. Sequences searched for include 'TATA' boxes, 'GC' boxes and other sequences suggested as putative promoter elements for ciliate genes. The coding strand immediately preceeding ciliate genes is very high in A content and the consensus sequence at the site of protein synthesis is AAAATGG.

INTRODUCTION

At the molecular level the ciliates are among the most divergent eucaryotic organisms yet characterized. In general, proteins that have been characterized from *Tetrahymena* are very divergent from homologues found in higher eucaryotes. For example the histones, actin and cytochrome c of *Tetrahymena* are among the most divergent representatives yet characterized (1,2,3). This divergence is dramatically demonstrated by the fact that many ciliates have a nuclear codon assignment that deviates significantly from the universal nuclear genetic code (4,5,6). A difference in the genetic code is an absolute barrier to the exchange of genetic information indicating a long period of isolation. This strongly implies that these organisms diverged early during the evolution of the eucaryotic lineages (7).

The promoter regions of ciliate genes are poorly characterized at present. One approach to this problem has been to search the upstream regions of the sequenced ciliate genes for DNA sequences that contain 'typical' eucaryotic promoter elements. Currently, at least three distinct types of regions are believed to constitute eucaryotic promoters; 'TATA' boxes, 'CCAAT' boxes and 'GC' boxes (8,9). The 'TATA' box is usually found proximal to the transcription initiation site in higher eucaryotes (25−30 bp upstream), but ranges between 40 and 120 bp upstream in yeast (10). The eucaryotic 'TATA' box appears to be analogous to the procaryotic '−10 sequence' (11). The majority of eucaryotic genes characterized have 'TATA' boxes although it is not an absolute requirement. Transcription of genes from which the 'TATA' box has been deleted or modified suggests that this region plays a major role in setting the site for the initiation of transcription (12,13). Genes that naturally lack a 'TATA' box (3-hydroxy-3-methylglutary coenzyme A reductase, epidermal growth factor and hypoxanthine phosphoribosyl transferase) generally have multiple sites of transcription initiation (14,15,16).

The 'CCAAT' box is a less common promoter element, found in some eucaryotic promoters and is usually 40 to 100 bp upstream from the initiation site for transcription (12,17). These sequences can appear in tandem and in either orientation and appear to modulate the expression of the gene (12,18,19).

'GC' boxes are found in a few promoters over a wide range of sites, from relatively close to the initiation site for transcription to several hundred bp upstream (8,20). These sequences are often in tandem arrays with both orientations present. Sp1 DNA binding protein is thought to interact with 'GC' boxes to control gene expression (9).

Positive identification of these sequences in ciliates has been inconclusive. The number of sequenced ciliate genes is small; most are from *T. thermophila* with several from *T. pyriformis* and *T. pigmentosa* and a few from other genera including *Paramecium*, *Oxytricha* and *Stylonychia*. When the upstream sequences from these genes are examined, some regions have been found that resemble eucaryotic promoter elements but no convincing description of a ciliate promoter has emerged. The identification of promoter elements such as 'TATA' boxes is hampered by the high AT content of most ciliate genomes, which is often as great as 75% in intergenic regions.

Another approach to this problem involves comparing homologous ciliate genes and searching for elements that appear to be conserved among them. Attempts to do this in the past have been hampered by the difficulty in aligning promoter regions from distantly related genes. We have sequenced a set of homologous intergenic regions from 22 species in the *Tetrahymena pyriformis* Complex and compared the conserved portions of these sequences in an attempt to identify essential features of the *Tetrahymena* promoter.

The region of the *Tetrahymena thermophila* genome that

includes the histone H4II and H3II genes has been completely sequenced (30). These histone genes are transcribed divergently, thus the intergenic region probably contains two promoters. We have sequenced a portion of the genome from each species, which includes parts of the H3II and H4II histone genes as well as the intergenic region. This provides a set of 22 sequences that can be accurately aligned and compared. The essential elements of the promoters should be identifiable as conserved regions among these sequences. With the exception of 'CCAAT' boxes, we find no evidence for the presence of any of the sequences previously suggested as promoter elements in ciliates. A 'CCAAT' box sequence that is comparable with general eucaryotic 'CCAAT' boxes is found not only in the histone H3II and H4II promoters, but preceeding several other ciliate genes as well.

MATERIALS AND METHODS

Species of ciliates and preparation of DNA

In this study we have compared the histone H3II/H4II regions from the following species (strain: EMBL Data Library accession numbers for the H3/H4 sequences); *T. americanis* (HP1: X17126), *T. australis* (GB5: X17127), *T. borealis* (WZ3: X17128), *T. canadensis* (MP69: X17129), *T. capricornis* (AU34: X17130), *T. caudata* (MP49: X17131), *T. elliotti* (Ch: X17125), *T. furgasoni* (W: X17132), *T. hyperangularis* (WB6: X17133), *T. leucophyrs* (TUR: X17134), *T. malaccensis* (MP75: X17135), *T. mimbres* (HS: X17136), *T. nanneyi* (WXO: X17137),*T. nippisingi* (X31R: X17138), *T. paravorax* (RP: X17139), *T. patula* (LI: X17140), *T. pigmentosa* (IL3: X17143),*T. pyriformis* (GL: X17141), *T. rostrata* (30770: X17144), *T. sonnneborni* (XQ10: X17142), *T. thermophila* (CU401), *T. tropicalis* (TC3: X17145). All of the species except *T. thermophila* and *T. rostrata* were a generous gift from Drs. E. Simon and D. Nanney (University of Illinois). *T. thermophila* was a gift from Dr. P. Bruns (Cornell University) and *T. rostrata* was purchased from the American Type Culture Collection (Princeton, N J). These organisms were grown under standard conditions (21). DNA was isolated and purified as previously described (22).

Amplification and cloning of the histone H3II/H4II regions

Using the published sequence from *T. thermophila*, oligonucleotides complementary to opposite strands in the histone H3II and histone H4II genes were synthesized (30). As shown in figure 1, the H3II oligonucleotide sequence has a naturally occurring Kpn I restriction site, while a single base change was introduced into the H4II oligonucleotide to create a Hind III restriction site. These oligonucleotides were used as primers to amplify the histone H3II/H4II region from the DNA of each species using the polymerase chain reaction (PCR) and*Taq* DNA polymerase as described by the manufacturer (Perkin-Elmer Cetus)(23). The amplified histone H3II/H4II regions were isolated by 3% agarose electrophoresis, electroeluted from the gel and restricted with Kpn I and Hind III. These amplified regions were then ligated into pUC19 and used to transform *E. coli* MC1061 (24). DNA from at least four colonies containing the amplified regions was prepared and combined in equal molar amounts for sequence determination. The inherent nucleotide misincorporation rate for the PCR amplification and cloning procedure is about 0.25% (25). Thus the combination of four independent clones avoids any potential errors due to misincorporation during the PCR amplification.

DNA sequence determination and sequence analysis

The nucleotide sequences of the combined clones containing the histone H3II/H4II regions from the various species were determined by dideoxynucleotide sequencing procedures using Sequenase (U.S. Biochemicals) as described by the manufacturers (26). The sequences were aligned by pairwise application of a Needleman-Wunsch alignment protocol (27). Further refinement of the relative alignments was done by inspection. Conserved sequences within the region were identified by calculating the number of different symbols (nucleotides or gaps) at each position and then passing a triangular smoothing function (window width 13) over these data (28). A window width of 13 is the minimum length that produces a smooth curve. The adenine (A) profile of the regions was determined in a similar fashion, by calculating

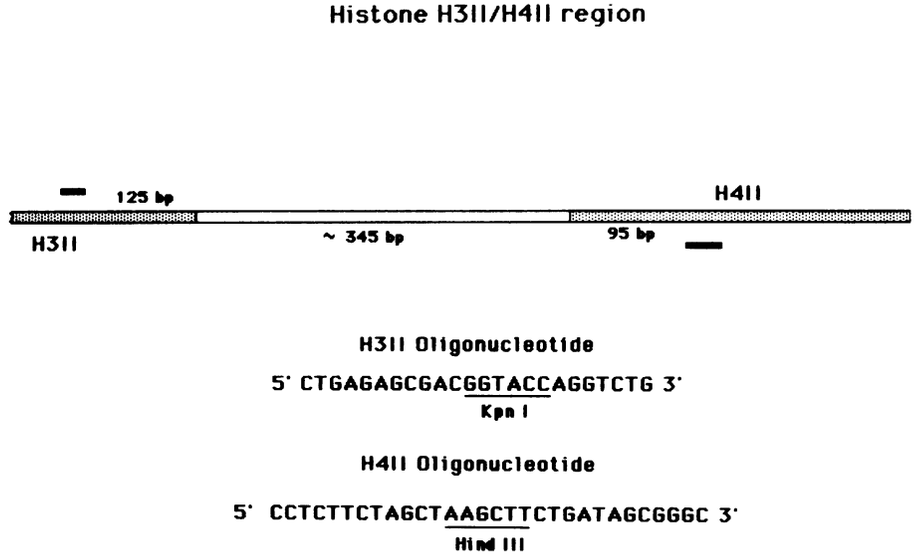

Figure 1. The portion of the *Tetrahymena* genome including the intergenic region between histones H3II and H4II is depicted. The histone coding regions are shown as stippled. Both genes are transcribed divergently from the intergenic region. The placement of the primer oligonucleotides for PCR amplification is shown as black boxes and the sequences of the oligonucleotides are shown in the lower portion of the figure.

the percentage of adenine at each position and then smoothing the data. Sequence matching and dot plots were performed using the sequences analysis packages of the University of Wisconsin programs (29).

RESULTS

Analysis of histone intergenic regions

The consensus sequence for the histone H3II/H4II intergenic region is shown in figure 2. The sequence, which is 75% AT, begins immediately adjacent to the histone H3II gene and concludes immediately adjacent to the histone H4II gene. Positions at which the nucleotides in all of the species are greater than 90% identical are underlined in figure 2. These conserved regions are an aid in the appropriate alignment of the sequences.

Histone H3II/H4II Intergenic Consensus Sequence

```
1                                                    50
IICIITTTAA GTGTTTTAAA AGGAGIIGTC IIIITGACTT TTCIIIIGAA

51                                                  100
GGTIAITGAI IIIIIIIIAA IAAAAIICII IAIGAACAAC GATTAAGGCG

101                                                 150
AGAICAIIIG AAAIGIIIGG AAITAICCTG GAATGCGAAG AIATGCCCGA

151                                                 200
TIITGAIIGG AIGAIIIGAA AGGAAAICAG AIIIITGAGA TTCIAICCAA

201                                                 250
ICAGAAICGA AAICIGAATT GAAIIITGGAI AATATGATAA TAIIITAAAA

251                                                 300
AAAAAAAGAT ATCTTTCCCC AAAGACTAIA AICAIIAAAA CAAAAAATAA

301                                                 346
AAAAACTAAT IAAAAATAAA AATAAAAAAA AAAAAACCA GCAAAA
```

Figure 2. The consensus sequence for the intergenic region between the histone H3II and H4II genes is shown. Positions at which the nucleotides in all of the species are more than 90% identical are underlined.

In general, the majority of the insertions/deletions required to align the species occur within 50 nucleotides of the histone H3II gene and 100 nucleotides of the histone H4II gene, while the central portion of the intergenic region is more conserved. In order to more clearly identify the conserved regions, the number of symbols at each position was calculated and these data were smoothed to produce a profile of the variation, shown in figure 3. If all of the nucleotides were identical at a given position the height of the profile would be 1, similarly if all four nucleotides appeared at a given position the height of the profile would be 4. In figure 3, the regions corresponding to histone H3II and H4II are very conserved with a profile height only slightly above 1. The intergenic region is much less conserved with portions of the profile approaching a height of 3. The regions adjacent to the genes display substantial variation, while conserved sequences are identified by dips in the profile. There are about a dozen such dips in the intergenic region profile. The sequences (length 13) associated with each of these dips were analyzed.

If the intergenic region contains two promoters in opposite orientations, common elements in these promoters should appear as conserved regions that have sequence identity with conserved regions in the reverse complement of the consensus sequence. Each of the conserved sequences identified by the profile (figure 3) was matched to the reverse complement. Only the two deep dips in the profile, identified in figure 3 as a heavy line, showed significant sequence identity with the reverse complement. Most of the other dips are regions of high A or T content and thus any similarity with other regions can be accounted for simply by the high AT content of the intergenic region.

Identification of 'CCAAT' boxes

The regions identified by the boxes in figure 3 are almost as conserved as the histone coding regions. The consensus sequence between position 195 and 205 is ATCCAATCAGA, while the reverse complement consensus sequence between 162 and 152 is ATCCAATCAAA. The identity of these sequences is significant, 10 of 11 nucleotides match within highly conserved

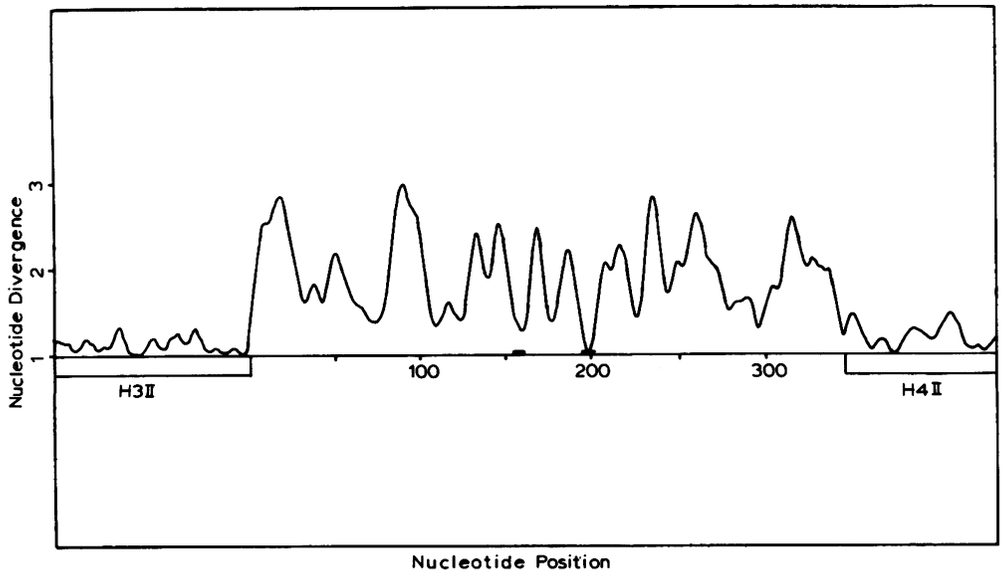

Figure 3. The smoothed (window 13) nucleotide divergence for the histone H3II/H4II intergenic region is displayed. The black boxes indicate the locations of the 'CCAAT' boxes.

regions that contain a substantial number of Cs. These conserved sequences also contain a canonical 'CCAAT' box sequence which further suggests that they may indeed be general promoter elements.

The putative 'CCAAT' box, proximal to the histone H4II gene, is 141 nucleotides upstream from the ATG at which translation of the protein starts and 82 or 87 nucleotides upstream from the two major initiation sites for transcription of the histone H4II gene (30). The putative 'CCAAT' box proximal to the histone H3II gene is 151 nucleotides upstream from the ATG at which translation of the protein starts. Unfortunately the initiation sites for transcription of this gene have not been mapped. The locations of these putative 'CCAAT' boxes are consistent with the general placement of 'CCAAT' boxes in eucaryotic promoters.

If these 'CCAAT' boxes are general elements of ciliate promoters then they should be found upstream of other ciliate genes. To date, about 20 ciliate genes have been sequenced, including representatives of several different genera. The sequence of the 'CCAAT' box present in *Tetrahymena* histone H3II and H4II genes was matched to all of the other ciliate sequences available. Table 1 shows that good matches to this sequence are found at 8 other sites in ciliate genes. The consensus sequence of this putative ciliate 'CCAAT' box is TATCCAATCARA. Table 2 lists the ciliate genes searched and gives the position of putative 'CCAAT' box sequences relative to the ATG at which protein synthesis initiates as well as the sites of transcription initiation, for those genes in which they have been mapped. The lengths of the sequenced 5′ region flanking the genes are also given in Table 2. Most of the 'CCAAT' boxes identified are at an appropriate distance from the gene.

The 'CCAAT' box in front of the *T. thermophila* histone H4I gene is in a reverse orientation from the conventional sequence. There are three sequences matching the 'CCAAT' consensus found upstream of the *T. thermophila* histone H2BI gene.

Search for other putative promoter elements

The other conserved sequences in the histone H3II/H4II intergenic region are not found either in the reverse complement of this sequence or in the upstream regions of other ciliate genes. At about position 270 most of the species we have sequenced have a series of 3 to 5 consecutive Cs. This initially suggested itself as an important promoter element and it may indeed be a element in the histone H4II genes promoters, however we do not find this pattern upstream of histone H3II genes or any of the other ciliate genes investigated. Thus this series of Cs is probably not a general feature of ciliate promoters.

Table 1: Ciliate "CCAAT Boxes"

H4II	ATCAGATTTT	TGAGATTCTA TCCAATCAGA	ATCGAAATCT GAATTGAATT
H3II	TCTGATTTCC	TTTCAAATCA TCCAATCAAA	ATCGGGCATA TCTTCGCATT
H4I *	TCCTTAGCGC	CGAAAAATTA TCCAATCAGA	ATCAGTCTTT CTAGAGATTC
H1	CTTATTTGGA	GAGGAGATTA TCCAATCAGA	TTTCAGATTA TTTTTAAGAG
H2BI (A)	GATAAATAAT	AAATAAATTA TCCAATTTAA	AAATAAATTA TCCAATCAGA
H2BI (B)	TCCAATTTAA	AAATAAATTA TCCAATCAGA	ACGCAGAAAA AGAAGATTAT
H2BI (C)	AACGCAGAAA	AAGAAGATTA TCCAATCAGA	TGCTTTTTTA ATAGAAGATA
H2BII	TCCGTATCTT	GGAAGATTTA TCCAATCAAA	TCACAGAAAT AGGTAGATAC
TP S25I	ATATATGGAT	GTTTTCAATA TCCATTCAAA	CTTAAAAGCT AAGACAAAGA
SL ATUB	AATTAAGAGG	GTCCTTCATA TCCATTCAAA	TATATCTTAA ATACTTAAAC
Consensus		TA TCCAATCARA	

* reverse complement

Table 2: Ciliate 5' sequences

Gene		Initiation site	"CCAAT Box"	5' length	Ref
T. thermophila	histone H1	-	-132	315	36
T. thermophila	histone H2BI	-45,-49	-127,-156,-176	230	31
T. thermophila	histone H2BII	-28,-39,-42	-127	230	31
T. thermophila	histone H3II	-	-151	340	30
T. thermophila	histone H4I	-45,-51,-55	-181	885	1
T. thermophila	histone H4II	-54,-59	-141	340	30
T. thermophila	actin	-55,-67,-74,-79,-86,-93,-97	-	248	2
T. thermophila	r protein S25	-70,-95	-	207	37
T. pigmentosa	r protein S25I	-70	-84	168	2
T. pigmentosa	r protein S25II	-70	-	169	32
T. pyriformis	alpha-tublin	-98,-103	-	300	33
T. pyriformis	beta-tublin I	-66,-70,-78	-	130	33
T. pyriformis	beta-tublin II	-84,-89,-93	-	128	33
P. primaurelia	G surface protein	-8	-	298	38
P. tetraurelia	A surface protein	-	-	195	4
O. fallax	actin	-	-	185	39
S. lemnea	alpha-tubulin	-	-72	192	34
S. lemnea	beta-tublin I	-46	-	154	35
S. lemnea	beta-tublin II	-38	-	190	35

The consensus sequence for the histone H3II/H4II region was analyzed relative to the reverse complement of itself in a dot plot analysis using relatively relaxed conditions (length 12, mismatches 5) to reveal any potential sequence identities. The only sequences that were identified by this analysis, in addition to the putative 'CCAAT' boxes, were several short palindromes and runs of As and Ts. None of the sequences other than the putative 'CCAAT' boxes were found upstream from any of the other ciliate genes.

In the reports of other ciliate gene sequences several putative promoter elements have been suggested (2,31,30,32,33,34, 35,36,37, 38,39). All of these sequences have been matched to the consensus and reverse complement of the consensus sequence for the histone H3II/H4II intergenic region. None of these sequences (except the 'CCAAT' box described here) show significant identity with the H3II/H4II intergenic region.

The consensus sequence for a 'TATA' box is TATA(T or A)A(T or A)(8). There are no identical matches to this sequence in any of the H3II/H4II intergenic regions. When the appropriate portion (between the position of the 'CCAAT' box and the initiation site) of the 5′ flanking regions of the other ciliate genes are searched only 3 genes have an identical match to the 'TATA' box consensus sequence. The probability of finding 3 or more identical matches in these regions is about 85%. This points out the difficulty in identifying 'TATA' boxes in regions with high AT content.

The core of the 'GC' boxes, GGGCGG or CCGCCC, found in some eucaryotic promoters is not found in the histone H3II/H4II intergenic consensus nor in the upstream regions of the other sequenced ciliate genes. In view of the high AT content of upstream regions of ciliate genes, if 'GC' boxes were present they should be readily detected. The purine-rich sequences AGGA and AGAGA have been found about 25 bp upstream of the initiation site in several *Tetrahymena* histone genes (39). These sequences are not found as a regular feature in the H3II/H4II intergenic regions we have analyzed. For example, the AGAGA sequence is found at the appropriate position in 4 species, while 9 of the species have AGATC at this position and the rest other combinations.

A set of histone H3 and H4 genes are divergently transcribed from an intergenic region in yeast (41). An analysis of these yeast histone promoters suggests that the consensus sequence GCGAAAAANTNRGAAC functions as an element in cell-cycle-dependent transcription of these genes (42). In spite of the similarity of this gene organization with *Tetrahymena*, neither the consensus sequence nor the core element GCGAAA are found in the *Tetrahymena* histone H3II/H4II intergenic region.

High A content of coding strands

The intergenic region between histones H3II and H4II is not only high in AT content; the coding strand immediately adjacent to the genes has a very high A content (30). This can be seen in figure 4, which displays the A content of the regions upstream of the histone H4 II and H3 II genes as well as the *T. thermophila* actin gene. The A profile is displayed both as a smoothed function (window 13 bp) and as 50 bp blocks. Within the 100 nucleotides proximal to the gene the A content is a very high , peaking as high as 75% on the smoothed profile. The sites for the initiation of transcription are shown for histone H4II and actin indicating that the 5′ untranslated region of these messenger RNAs are very A rich. Peaks and dips in the A content occur in the immediate vicinity of the transcription initiation sites, particularly evident in the histone H4II and actin sequences. The majority of the other ciliate genes display similar high A profiles in the vicinity of the genes.

Initiation site for protein synthesis

The sequences flanking the initiation site for protein synthesis have been analyzed for a number of eucaryotic organisms and yield a consensus sequence of ACCATGG, while a similar analysis for plants gives a consensus sequence of AACATGG (43,44). Table 3 summarizes the sequence composition of 61

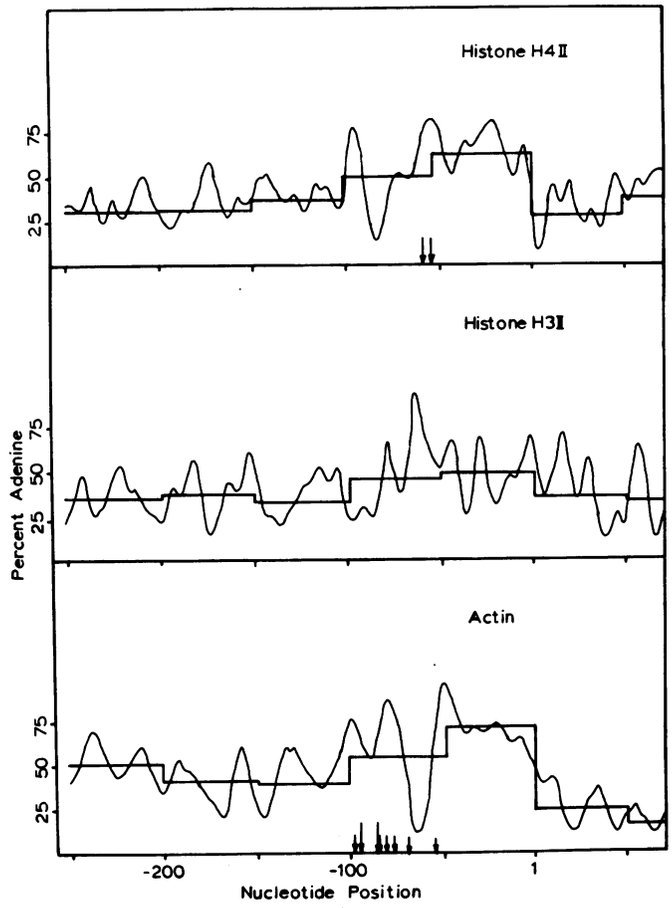

Figure 4. The percent adenine as a smoothed (window 13) and a block function is displayed for the upstream regions for histone H4II, histone H3II and actin genes from *Tetrahymena thermophila*. Long arrows indicate major transcription initiation sites and short arrows indicate minor transcription initiation sites. Transcription initiation sites have not been mapped for histone H3II.

Table 3: Sequence at site of Initiation of Protein Synthesis for Ciliate Genes

	-3	-2	-1	1	2	3	4
A	43	56	57	61	0	0	8
C	1	2	2	0	0	0	0
G	15	0	1	0	0	61	52
T	2	3	1	0	61	0	1
consensus	A	A	A	A	T	G	G

ciliate genes in the region flanking the initiation site of protein synthesis. The upstream nucleotides are predominantly As while a G is found in the +4 position for the majority of the genes.

DISCUSSION

The 22 homologous histone H3II/H4II regions we have compared can be accurately aligned and share a substantial degree of sequence identity. Within the conserved regions we have identified a putative promoter element that is also found at 8 sites upstream of other ciliate genes. The consensus sequence for this putative promoter element, TATCCAATCARA, contains a 'CCAAT' box sequence which is found in many eucaryotic promoters (8,12,17).

In most of the original reports of ciliate genes an identification of putative promoter elements is included (2,30,31,33,34, 35,39,40). In our analysis, with the exception of the 'CCAAT' boxes, we do not find any of these elements represented in the H3II/H4II regions of the 22 *Tetrahymena* species we have analyzed. In fact several of the 'CCAAT' boxes were not explicitly identified in the original reports. Clearly it is difficult or impossible to reliably identify putative promoter elements on the basis of a few sequence comparisons. Having a substantial number of well aligned sequences for comparison greatly enhances these identifications.

A search for putative 'TATA' boxes in the histone H3II/H4II region and among other sequenced ciliate genes has yielded only a few examples which match the canonical sequence. There are, however, numerous suggested 'TATA' boxes in the original reports. We have attempted to match most of the these suggested 'TATA' sequences to the *Tetrahymena* H3II/H4II sequences without identifying convincing matches. In view of the high AT content of the flanking regions of ciliate genes most putative 'TATA' boxes could occur by chance. The multiple initiation sites for transcription found in most of the ciliate genes mapped are consistent with the general absence of 'TATA' boxes in ciliate promoters. Probably a conventional eucaryotic 'TATA' box is not a regular element in ciliate promoters.

No clear examples of 'GC' box cores, GGGCGG or CCGCCC, were found in any of the ciliate gene upstream sequences nor did any of the original reports suggest the presence of putative 'GC' boxes. In view of the high AT content of the upstream sequences these core elements should be readily apparent if they are present. Most probably, eucaryotic 'GC' boxes are also not common elements in ciliate promoters.

The sequences immediately upstream of the genes in most ciliates have a very high A content in the coding strand. In many cases the 5′ untranslated region has an A content as high as 70%. There are also peaks and dips in the A content in the vicinity of the initiation of transcription. It is possible that this fluctuation in A content acts as a signal for the initiation of transcription. Such poly dA-dT sequences may affect the nucleosome pattern and act as regulatory sequences (45). In any case, the 5′ end of the messenger RNAs have a very high A content, which has been suggested as a possible explanation for the recovery of *Tetrahymena* histone messages in a poly A+ fraction (31). The high A content immediately upstream of ciliate genes is a common feature but its function is unclear.

The ribosome 'scanning' model suggests that the nucleotides in the immediate vicinity of the initiation codon may play an important role in initiation of protein synthesis in eucaryotes (43,44). In the vast majority of cases the nucleotide at the +4 position is a G. This is the case with ciliate genes; 85% of the nucleotides at the +4 position are Gs. The nucleotides found at the −3, −2 and −1 positions in ciliates are usually As;70%, 92% and 93% respectively. Generally in eucaryotes, the most common nucleotide at the −3 position is A, while either A or C is found at the −2 position and C is found at the −1 position (43,44). Thus, ciliates genes differ from most eucaryotic genes by generally having an A at the −1 position while C is found at this position in other eucaryotes. The ciliate genes analyzed are predominantly histones (48 of 61) and the majority are from *Tetrahymena* species (55 of 61). If the non histone genes are considered alone (13 genes), a similar pattern is observed except at the +4 position which has 61% As.

Several ciliate promoters appear to have typical 'CCAAT' boxes as elements, however few if any canonical 'TATA' boxes are observed and no 'GC' boxes are found. No other common elements of ciliate promoters have been identified. A comparison of a substantial number of well aligned homologous sequences allows a convincing identification of the 'CCAAT' boxes, but due to the high AT content the presence or absence of 'TATA' boxes is less certain. Apparently, ciliate genes share some of the elements common to eucaryotic promoters. An accurate definition of the essential elements of ciliate promoter elements will require *in vitro* modification of putative promoter and reintroduction of the gene by transformation to establish function.

ACKNOWLEDGEMENTS

We gratefully acknowledge the assistance given by Dr. David Nanney's laboratory and in particular Dr. Ellen Simon in providing the various species of *Tetrahymena*. We also thank C. A. Brunk for assisting with computer analysis of the data and R.W. Kahn and H. Sarner for excellent laboratory assistance. This work was supported by a grant from the National Science Foundation; BSR-8800805 to CFB and a USPHS National Research Service Award (GM-07104) to LAS.

REFERENCES

1. Bannon, G. A., Bowen, J. K., Yao, M.-C. and Gorovsky, M. A. (1984) Nucleic Acids Res. **12**, 1961−1975.
2. Cupples, C. G. and Pearlman, R. E. (1986) Proc. Natl. Acad. Sci. USA **83**, 5160−5164.
3. Tarr, G. E. and Fitch, W. M. (1976) Biochem. J. **159**,193−197.
4. Preer, J. R., Preer, L. B., Rudman, B. M. and Barnett, A. J. (1985) Nature **314**,188−190.
5. Horwitz, S. and Gorvosky, M.A. (1985) Proc. Natl. Acad. Sci. USA **82**, 2452−2455.
6. Caron, F. and Meyer, E. (1985) Nature **314**,185−188.
7. Sogin, M. L., Elwood, H. J. and Gunderson, J. H. (1986) Proc. Natl. Acad. Sci. USA **83**, 1383−1387.
8. Lewin, B. (1987) Genes III. John Wiley and Sons, New York.
9. McKnight, S. and Tijan, R. (1986) Cell **46**, 795−805.
10. Struhl, K. (1987) Cell **49**, 295−297.
14. Goldberg, M. (1979) Ph.D Thesis. Stanford University.
12. McKnight, S. L. and Kingsbury, R. (1982) Science **217**, 316−324.
13. Benoist, C. and Chambon P. (1981) Nature **290**, 304−309.
14. Luskey, K. L. (1987) Mol. Cell. Biol. **7**,1881−1893.
15. Ishii, S., Xu, Y.-H., Stratton, R., Roe, B. A., Merlino, G. T. and Pastan, I. (1985) Proc. Natl. Acad. Sci. USA **82**, 4920−4924.
16. Melton, D. W., McEwan, C., McKie, A. B. and Reid, A. M. (1986) Cell **44**, 319−328.
17. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Bioch. **50**, 349−383.
18. Graves, B. J., Johnson, P. F. and McKnight, S. L. (1986) Cell **44**, 565−576.
19. Barberis, A., Superti-Furga, G. and Busslinger, M. (1987) Cell **50**, 347−359.
20. McKnight, S. L. (1982) Cell **31**, 355−365.
21. Orias, E. and Bruns, P. J. (1976) Methods in Cell Biol. **13**, 247−282.

22. Brunk, C. F. and Hanawalt, P. C. (1967) Science **158**, 663–664.
23. Mullis, K. B. and Faloona, F. A. (1987) Meth. Enz. **155**, 335–350.
24. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Mannual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
25. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H.A. (1988) Science **239**, 487–491.
26. Sanger, F. and Coulson, A. R. (1975) J. Mol. Biol. **94**, 441–448.
27. Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. **48**, 443–453.
28. Randall, R. B. (1987) Frequency Analysis. Bruel and Kjaer, Naerum, Denmark.
29. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. **12**, 387–395.
30. Horwitz, S., Bowen, J. K., Bannon, G. A. and Gorovsky, M. A. (1987) Nucleic Acids Res. **15**,142–160.
31. Nomoto, M., Imai, N., Sagiga, H., Matsui, T. and Mita, T. (1987) Nucleic Acids Res. **15**, 5681–5697.
32. Engberg, J., Bojsen, K. and Nielsen, H. (1989) UCLA Symposium in press.
33. Barahona, I., Soares, H., Cyrne, L., Penque, D., Denoulet, P. and Rodrigues-Pousada (1988) J. Mol. Biol. **202**, 365–382.
34. Helftenbein, E. (1985) Nucleic Acids Res. **13**, 415–433.
35. Conzelmann, K. K. and Helftenbein, E. (1987) J. Mol. Biol. **198**, 643–653.
36. Wu, M., Allis, C. D., Richman, R., Cook, R. G., and Gorovsky, M. A. (1986) Proc. Natl. Acad. Sci. USA **83**, 8674–8678.
37. Nielson, H., Andreasen, P. H., Dreisig, H., Kristiansen, K. and Engberg, (1986) J. EMBO J. **5**, 2711–2717.
38. Prat, A., Katinka, M., Caron, F. and Meyer, E. (1986) J. Mol. Biol. **189**, 47–60.
39. Kaine, B. P. and Spear, B. B. (1982) Nature **295**, 430–432.
34. Nomoto, M., Matsui, T., Saiga, H. & Mita, T. (1988) Oxford Survey on Eukaryotic Genes **5**, 251–278.
41. Smith, M. M. & Andresson, O. S. (1983) J. Mol. Biol. **169**, 663–690.
42. Osley, M. A., Gould, J., Kim, S., Kane, M. and Hereford, L. Cell **45**, 537–544.
43. Kozak, M. Cell (1986) **44**, 283–292.
44. Lutcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F. and Scheele, G. A. EMBO J. **6**, 43–48.
45. Struhl, K. (1985) Proc. Natl. Acad. Sci. USA **82**, 8419–8423.