Volume 16 Number 5 1988 Nucleic Acids Research

Multiple introns in a conjugation-specific gene from *Tetrahymena thermophila*

Duane W.Martindale* and Frances M.Taylor

Department of Microbiology, Macdonald College of McGill University, Ste Anne de Bellevue, PQ H9X 1C0, Canada

Received January 7, 1988; Accepted February 2, 1988 Accession no.XO6462

ABSTRACT

Multiple introns have been found in a gene from a ciliated protozoan. This Tetrahymena thermophila gene (cnjB) is large (7.5 kb mRNA) and active only during conjugation, the organism's sexual cycle. Six introns ranging in size from 62 bp to 676 bp were found when we sequenced a 3.1 kb segment of the cnjB gene together with its corresponding cDNA. We estimate, by extrapolation of our current data, a total of approximately 30 introns in this gene with a total gene size (introns plus exons) of 15 kb or more. The number of introns is surprising given the scarcity of introns in ciliate genes examined to date. Our findings constitute the first example of multiple introns in a ciliate gene. Having the sequence of several introns has allowed us to construct consensus sequences for T. thermophila mRNA introns. The 5' and 3' intron junctions resemble those of general nuclear mRNA (GT/AG rule is followed) but differences are seen. In particular, stretches of 10 or more adenines and thymines are found adjacent to the conserved GT and AGs at the junctions. Unusual aspects of the coding region of this gene are discussed.

INTRODUCTION

The ciliated protozoa appear to have emerged as a group soon after the appearance of the eukaryotic progenitor approximately two billion years ago. Evidence for this comes from an examination of several conservative chronometric molecules (both protein and RNA) (see 1, 2). For example, cytochrome C of Tetrahymena was found to more more 'prokaryotic' than the cytochrome of any other eukaryote studied (3) and the ciliate forms of 'eukaryotic proteins' such as histones, actins and calmodulin have been found to be unexpectedly divergent (see 1, 4). Further evidence of ciliate divergence comes from the fact that the genetic code of ciliates has changed from the 'universal' code found in all eubacteria and eukaryotes (4-7). It appears that early in the evolution of the ciliates, a G to A transition occurred in the anticodon of a duplicated glutamine tRNA gene allowing the suppression of the stop codons UAA and UAG (8). This led to the use of UAA and UAG as codons for glutamine and left UGA as the only stop codon.

An examination of the gene structure of this ancient group of organisms might give some insight into the evolution of the eukaryotic genome. One of

the characteristic features of the eukaryotic genome is the presence of intervening sequences or introns within the coding regions of genes. The first report of introns in a nuclear mRNA from ciliates was for the conjugation specific gene cnjB of Tetrahymena thermophila (9, 10). Evidence for multiple introns in this gene was initially obtained through a comparison between the restriction maps of a cDNA clone and the corresponding genomic DNA. Recently, the T. thermophila genes for histone HI (11) and ribosomal protein S25 (12) were each found to contain a single intron. In this report, we present the sequence of six introns from cnjB and a portion of the coding region of this gene. The sequence of these six introns and the introns in the histone HI gene and ribosomal protein S25 gene will be compared and a consensus sequence will be constructed for intron junctions in T. thermophila genes. Unusual aspects of the coding region of cnjB will be discussed.

MATERIALS AND METHODS

Cloning procedures

The construction of the cDNA clone pC2-1 has been described (13) as has the isolation of several recombinant lambda clones (lC2-1, lC2-2 etc) containing T. thermophila genomic DNA with cnjB gene sequences (10). Standard methods were used for agarose gel electrophoresis, lambda DNA isolation, subcloning and restriction enzyme digests (14). Restriction fragments were isolated by electrophoresis from agarose gels onto NA-45 DEAE membranes (Schleicher and Schuell) from which the DNA was recovered as described by the manufacturer. These fragments were often further purified using NACS columns (Bethesda Research Laboratories). Plasmids pure enough for subsequent manipulations were isolated by an alkaline lysis procedure (15). For quick mini-preps of plasmids we used the LiCl-boiling method (16).

DNA Sequencing

DNA fragments to be sequenced were subcloned into the multiple cloning site of the plasmid vector pIBI31 from International Biotechnologies, Inc. (IBI). This vector is a pUC19 derivative containing an f1 origin of replication and T3 and T7 RNA polymerase promoters. The bacterial strain 71-18 was used as a host. Series of deletions of the inserted DNAs were created using exonuclease III (Boehringer-Mannheim) digestion as described by Henikoff (17). Single-stranded templates for sequencing were prepared using helper phage (M13K07) as described by the supplying company (IBI). These templates were then sequenced by the dideoxynucleotide method (18) using a modified T7

DNA polymerase (sequenase; ref. 19) supplied in kit form from United States Biochemical Corporation.

RESULTS

The restriction maps of cnjB cDNA (pC2-1) and its corresponding genomic DNA indicated, through restriction fragment size differences, that a minimum of six introns was present within the cnjB gene (10; Fig. 1). DNA sequencing of this cDNA and corresponding genomic DNA was undertaken to confirm that these restriction fragment differences were due to the presence of introns within the gene and not an artifact from the construction of the cDNA clone pC2-1. If these differences were due to introns, we could then examine the structure of mRNA introns in T. thermophila genes.

Subclones containing cnjB sequences were constructed using the vector pIBI31 and appropriate restriction fragments from cnjB cDNA (pFT100) or

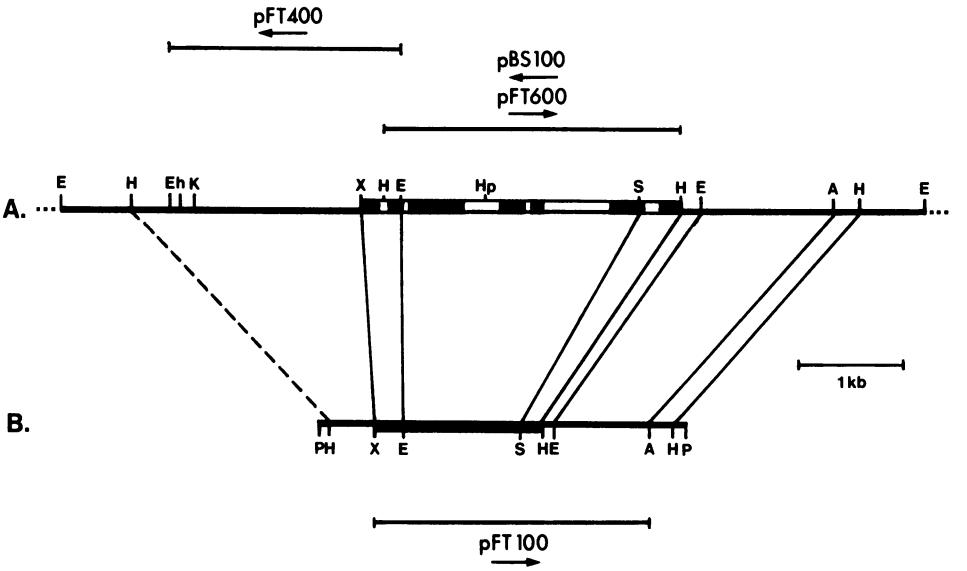
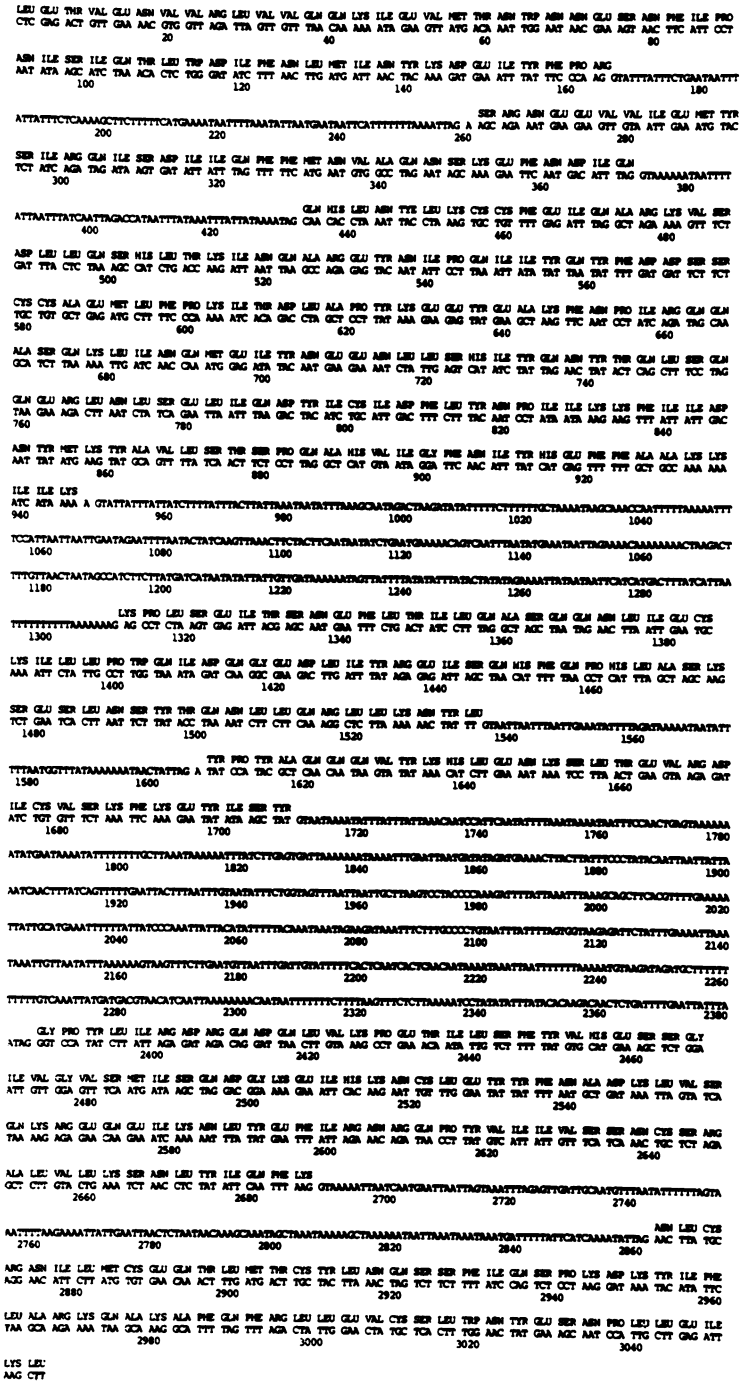


Figure 1. Restriction map of cnjB genomic DNA (A) corresponding to the cDNA (B) from pC2-1. The restriction fragments that were subcloned into pIBI31 for sequencing are indicated along with the names given for the resulting plasmids. The direction that the restriction fragments were inserted relative to the sequencing primer region of PIBI31 is indicated by arrows underneath the plasmid names. The thickest lines in A and B are the sequenced regions of the cDNA and its corresponding genomic DNA (see Fig. 2; the filled regions indicate exons while the open regions indicate introns). The PstI sites (P) on the ends of the cDNA were constructed during cloning (13). A, AccI; E, EcoRI; H, HindIII; Hp, HpaI; K, KpnI; P, PstI; S, SstI; X, XhoI .



genomic DNA (pBS100, pFT400 and pFT600) (see Fig. 1). Ordered deletions in the plasmids pFT100, pFT600 and pBS100 were constructed (17) and the resulting plasmids were sequenced (18). The 1.6 kb XhoI-HindIII region of the cDNA was completely sequenced in one direction, and the 2.9 kb HindIII-HindIII region of the genomic DNA was sequenced in both directions. The genomic sequence 5' to the insert in pFT600 was obtained by sequencing one end of pFT400 (see Fig. 1). Although the cDNA sequence (approximately 150 bp) immediately adjacent to the XhoI site was not confirmed by sequencing the corresponding genomic DNA, restriction mapping indicates that no intron sequences are present in this region. The sequence obtained is presented in Fig. 2. Note that the HpaI site (GTTAAC) previously reported in the cDNA (see Fig. 2 in ref. 10) was not found. We believe we were misled by partial cutting of the cDNA at the sequence GCTAAC, located at position 1444 in Fig. 2 which is where we had calculated the HpaI site to be by restriction mapping of the cDNA.

Six introns were found within the 3056 bp XhoI-HindIII region of the genomic DNA. The average intron size of 240 bp (range 62 bp to 676 bp) was very close to the average exon size of 251 bp (range 103 bp to 514 bp). The size and distribution of the exons and introns is graphically illustrated in Figure 3. The introns were found to be very AT rich. They had a GC content of 16% compared to 30% for the coding region. When the GC content of the introns was plotted against the size of the introns (Fig. 4) the introns were found to be increasingly AT rich as their size decreased. As the intron size increased, its GC content approached 25% which is the average for T. thermophila DNA (20). Having the sequences of multiple introns allowed us to search for consensus sequences (see discussion; Fig. 5 and 6).

There was an open reading frame throughout the cDNA sequenced (1611 bp; Fig. 2) This coding region is unusually AT rich (30% GC). Frequent use is made of the ciliate specific glutamine codons TAA and TAG (19 and 14 times respectively). Furthermore, two of the codons used in this gene (GTG and ACG) of this cDNA and corresponding genomic DNA was undertaken to confirm that these restriction fragment differences were due to the presence of introns within the gene and not an artifact from the construction of the cDNA clone pC2-1. If these differences were due to introns, we could then examine the structure of mRNA introns in T. thermophila genes.

Subclones containing cnjB sequences were constructed using the vector

Figure 2. Partial DNA sequence of the cnjB gene from T. thermophila (3.1 kb XhoI-HindIII region). Six complete introns are included in this sequence.

Figure 3. Intron and exon sizes in base pairs over the sequenced region of cnjB.

pIBI31 and appropriate restriction fragments from cnjB cDNA (pFT100) or genomic DNA (pBS100, pFT400 and pFT600) (see Fig. 1). Ordered deletions in the plasmids pFT100, pFT600 and pBS100 were constructed (17) and the resulting plasmids were sequenced (18). The 1.6 kb XhoI-HindIII region of the cDNA was completely sequenced in one direction, and the 2.9 kb HindIII-HindIII region of the genomic DNA was sequenced in both directions. The genomic sequence 5' to the insert in pFT600 was obtained by sequencing one end of pFT400 (see Fig. 1). Although the cDNA sequence (approximately 150 bp) immediately adjacent to the XhoI site was not confirmed by sequencing the corresponding genomic DNA, restriction mapping indicates that no intron sequences are present in this

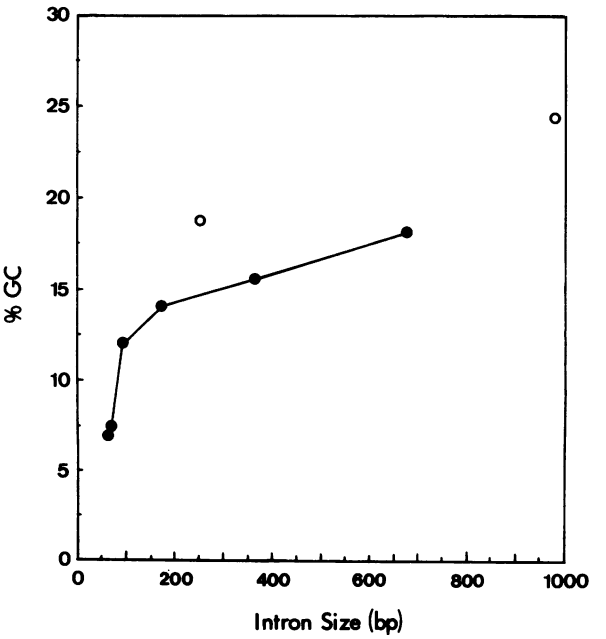


Figure 4. Intron size versus % GC. The filled circles represent the introns found in cnjB, while the open circles represent T. thermophila introns from the histone H1 gene (254 bp, ref. 11) and the ribosomal protein S25 gene (979 bp, ref. 12).

Intron	
62	TAG↓GTAAAAATAATT
71	TTT↓GTAATTAATTTA
95	AAG↓GTATTTATTTCT
174	AAG↓GTAAAAATTAAT
364	AAA↓GTATTATTTATT
676	TAT↓GTAATAAAATAT
H1	AGC↓GTAAGAATTTAA
S25	AAG↓GTGAAATATTTT
5' Consensus	$\begin{matrix}A\\T\end{matrix}$AG↓GTAA$(\begin{matrix}A\\T\end{matrix})_{n\geq 8}$

Figure 5. Determination of the consensus sequence for T. thermophila's 5' intron junction . The cnjB introns are distinguished by their lengths in base pairs. H1 and S25 identify introns from the histone H1 gene (11) and ribosomal protein S25 gene (12). The 5' exon/intron junctions (or splice sites) are denoted by arrows.

region. The sequence obtained is presented in Fig. 2. Note that the HpaI site (GTTAAC) previously reported in the cDNA (see Fig. 2 in ref. 10) was not found. We believe we were misled by partial cutting of the cDNA at the sequence GCTAAC, located at position 1444 in Fig. 2 which is where we had calculated the HpaI site to be by restriction mapping of the cDNA.

Six introns were found within the 3056 bp XhoI-HindIII region of the genomic DNA. The average intron size of 240 bp (range 62 bp to 676 bp) was

	Intron
TATTATAAAATAG↓CAA	62
AAATAACTATTAG↓ATA	71
TTTTTAAAATTAG↓AAG	95
ATCAAAATATTAG↓AAC	174
TTTTTAAAAAAAAG↓AGC	364
AATTATTTAATAG↓GGT	676
ATTTAAAAATTAG↓GAT	H1
TTTAAATAAAAAG↓ATG	S25
3' Consensus	$(\begin{matrix}A\\T\end{matrix})_{n\geq 8}A\begin{matrix}A\\T\end{matrix}$TAG↓A

Figure 6. Determination of the consensus sequence for T. thermophila's 3' intron junction. The cnjB introns are distinguished by their lengths in base pairs. H1 and S25 identify introns from the histone H1 gene (11) and ribosomal protein S25 gene (12). The 3' intron/exon junctions (or splice sites) are denoted by arrows.

very close to the average exon size of 251 bp (range 103 bp to 514 bp). The size and distribution of the exons and introns is graphically illustrated in Figure 3. The introns were found to be very AT rich. They had a GC content of 16% compared to 30% for the coding region. When the GC content of the introns was plotted against the size of the introns (Fig. 4), the introns were found to be increasingly AT rich as their size decreased. As the intron size increased, its GC content approached 25% which is the average for T. thermophila DNA (20). Having the sequences of multiple introns allowed us to search for consensus sequences (see discussion; Fig. 5 and 6).

There was an open reading frame throughout the cDNA sequenced (1611 bp; Fig. 2) This coding region is unusually AT rich (30% GC). Frequent use is made of the ciliate specific glutamine codons TAA and TAG (19 and 14 times respectively). Furthermore, two of the codons used in this gene (GTG and ACG) have not been seen in the T. thermophila genes reported previously in the literature (see ref. 21, 12). There is a preference for A over T in the cnjB coding region examined (596 versus 534) but not in the introns (607 versus 601). This preference for A over T is also found in another coding region of cnjB that we have sequenced (382 versus 318) (DWM and FMT, unpublished data). A slight preference for C over G may exist in the introns (122 versus 112) but they are approximately equal in the coding region (241 versus 243).

DISCUSSION

The evidence presented here clearly demonstrates the presence of multiple introns in a Tetrahymena thermophila gene that is transcribed into mRNA. This is the only case so far reported of multiple introns in a ciliate gene although the sample size of sequenced ciliate genes is small. To our knowledge the sequences or partial sequences of 11 ciliate genes have been reported (4-7, 11, 12, 21-26). Two have been found to contain single introns, the gene for histone HI (11) and the gene for the ribosomal protein S25 (12). We have compared the restriction maps of cDNA to genomic DNA for five genes active during conjugation (10, DWM, unpublished data). Only the gene described in this paper has introns. By combining this information with that reported in the literature we find that three out of 16 ciliate genes (19%) or three out of 11 T. thermophila genes (27%) contain introns. Since the nuclear RNA complexity of T. thermophila is only about 50% greater than that of the mRNA (27), the frequency of intron containing genes seen so far may be close to that of ciliate genes as a whole. Introns are clearly less common in the ciliates than in higher eukaryotes, but they may be present in approximately the same

frequency as in other lower eukaryotes (28, 29). It has been argued that the relative lack of introns in unicellular eukaryotes is due to a selection for lower DNA content in organisms that have a rapid cell division and DNA replication cycle (30, 31).

The T. thermophila mRNA introns most resemble general nuclear mRNA introns as opposed to class I and class II self-splicing introns (Table 1). The consensus sequence at the 5' junction (A/T A G : G T A A $(A/T)_n$, where n equals 8 or more, Fig. 5) conforms reasonably well with the vertebrate nuclear mRNA intron consensus sequence (A/C A G : G T A A G T) (35, 36) except that there is an AT rich segment in Tetrahymena introns (Table 1). At the 3' intron junction there is less of a relationship between the consensus sequence of Tetrahymena ($(A/T)_n$ A $A/_T$ T A G : A, where n equals 8 or more) and that of vertebrates ($(T/C)_n$N $C/_T$ T A G : G) (36). Both contain the absolutely conserved AG (Table 1), but instead of the stretch of approximately 10 pyrimidines seen in vertebrates, 11 or more A/T bases are found next to the conserved AG in Tetrahymena.

Introns contain sequences that are necessary for that particular intron's splicing to occur. One would predict that the smaller the intron, the greater the proportion of the intron occupied by these essential sequences. Thus, the observation that smaller introns are increasingly more AT rich than larger introns (Fig. 4) would also suggest that AT richness is important for intron processing in Tetrahymena. As the intron's size increased, its GC content also increased, approaching the average GC content (25%) of the Tetrahymena genome (20) as might be expected of non-coding regions of DNA. The two introns reported previously (11, 12) were within conserved genes (histone H1 and ribosomal protein S25) with coding regions containing a higher GC content (49.5% GC and 44% GC respectively) than cnjB's coding region (30% GC). The GC

Table 1. Intron junctions

Intron type	5' junction	3' junction	Reference
Class I	T↓	G↓	32
Class II	↓GTG$\frac{C}{T}$G	$(\frac{C}{T})_n$NA$\frac{T}{C}$↓	33,34
General nuclear mRNA	$\frac{C}{A}$AG↓GT$\frac{A}{G}$AGT	$(\frac{T}{C})_n$N$\frac{C}{T}$AG↓G	35,36
Yeast nuclear mRNA	↓GTATGT	$\frac{A}{T}$N$\frac{C}{T}$AG↓	37,38
Tetrahymena nuclear mRNA	$\frac{A}{T}$AG↓GTAA$(\frac{A}{T})_{n\geq 8}$	$(\frac{A}{T})_{n\geq 8}$A$\frac{A}{T}$TAG↓A	

content of their introns was also somewhat higher than would be expected from the curve drawn using only cnjB intron size and GC content data (Fig. 4). This implies a relationship between the GC content of these introns and their surrounding DNA. The introns in Dictyostelium, another lower eukaryote whose genome is AT rich, are also extremely AT rich although the sequences at their intron junctions (28, 39) differ somewhat from Tetrahymena's consensus sequences.

The resemblance between T. thermophila's intron junctions and that of nuclear mRNAs from other organisms (GT/AG rule maintained and other sequence similarities) leads us to speculate that the processing of T. thermophila's introns will be similar to them in many respects. One would expect that small nuclear RNAs will be involved in the splicing events and lariat formation will occur (40, 41). The T. thermophila introns do not contain obvious internal conserved sequence elements that may act like branch points for lariat formation. In higher eukaryotes there is significant sequence variability at the branch point but a preference is shown for the sequence Py Py Pu A Py and all branch points are found 18 to 37 nucleotides upstream from the 3' splice site (40). In yeast the branch site is the highly conserved sequence TACTAAC (note that it conforms to the vertebrate sequence; ref. 40). The sequence $A_{7/8}$ $T_{8/8}$ $A_{8/8}$ $A_{8/8}$ $T_{7/8}$ $T_{7/8}$ $T_{6/8}$, where the subscripts indicate the frequency of occurrence of the individual nucleotides, is found on average 33 nucleotides upstream from the 3' splice site in the T. thermophila introns examined, but this may not be significant considering the AT richness of these introns. Given that the ciliates are one of the earliest groups to branch off of the eukaryotic mainstream during evolution, an examination of RNA splicing in Tetrahymena should give some insight into the evolutionary relationship between the various forms of RNA splicing (42).

The average sizes of the exons and of the introns in the region of cnjB examined were similar (251 bp versus 240 bp). If these sizes are typical, then the total size of the cnjB gene will be approximately 15 kb (the cnjB mRNA is 7.5 kb) with the number of introns approaching 30. This large number of introns is in stark contrast to the situation in yeast, where intron-containing genes contain only one or two introns (29). The following observation suggests that caution must be exercised in generalizing intron/exon size and distribution in cnjB from the region which we have sequenced and which represents only about 20% of the total cnjB mRNA. Restriction mapping indicates that the HindIII-XhoI cDNA fragment (445 bp) immediately upstream of the region that we have sequenced corresponds to at

least a 2.2 kb HindIII-XhoI genomic DNA fragment (Fig. 1). This fragment may be larger if the genomic HindIII site shown turns out to be within an intron. Thus, in this region of cnjB, 445 bp of exon corresponds to at least 1755 bp of intron DNA, a ratio of approximately one to four instead of the one to one ratio seen over the sequenced region. We will attempt to synthesize cDNA representing the complete mRNA of cnjB and use it as a probe against genomic DNA so that we may determine the actual size of the cnjB gene and have a better estimate of the number and size of the introns in this gene.

The coding region of cnjB that we examined was found to be unusually AT rich (30% GC). This percent GC has been found over the 2.6 kb of the cnjB cDNA we have so far sequenced (Fig. 2; DWM and FMT unpublished data). All other Tetrahymena gene sequences reported are above 42% GC (see 43, 12). This significant difference in GC content may be related to the fact that the other T. thermophila genes reported are highly conserved evolutionarily while cnjB does not appear to be (based upon hybridization of cnjB cDNA to genomic DNA from other species; DWM, unpublished data). Perhaps the GC content of cnjB is typical of a T. thermophila gene under fewer evolutionary constraints than conserved genes and thus able to use more freely codons that reflect the overall AT richness of T. thermophila's genome. The next lowest GC contents reported for ciliate genes are 35 to 38% GC for the Paramecium surface antigen genes (5, 6, 25). These intronless genes are large (approximately 8 kb mRNA; ref. 44) and, like cnjB, are probably not evolutionarily conserved. There is an overall preference shown in cnjB for the more AT rich codons of a particular amino acid. An interesting example of this is the use of the more AT rich ciliate specific glutamine codons TAA and TAG (see 8) in preference to than the 'normal' glutamine codons CAA and CAG in this gene (34 times versus 24 times).

The cnjB gene, with its multiple introns, should be well suited for further studies on intron processing in T. thermophila. The small size of some of the introns may facilitate mutagenic studies and the identification of the critical sequences for intron splicing. Complete sequencing of the coding region of cnjB and analysis of its protein product will eventually tell us whether introns separate functional or physical domains of this protein. We also hope that further study of this unusually large gene and its presumably conjugation-specific protein product will allow us to determine its function. It is likely to be involved in meiosis or related events since the cnjB is transcriptionally active only during conjugation and maximally during the meiotic prophase (45) of this process.

ACKNOWLEDGEMENTS

The authors would like to thank Frances Stejskal for her expert technical assistance and Bettina Sanders for constructing and making a deletion series from the plasmid pBS100. We also acknowledge the assistance given by Marlene Parkinson in preparing the manuscript (especially Fig. 2). FMT was supported by a graduate fellowship from the province of Quebec (Fonds pour la Formation de Chercheurs et l'aide a la recherche). This work was supported by Grant A3032 from the Natural Sciences and Engineering Council of Canada to DWM.

*To whom reprint requests should be sent

REFERENCES

1. Nanney, D.L. (1986) In Gall, J.G. (ed.), The Molecular Biology of Ciliated Protozoa, Academic Press, New York, pp. 1-26.
2. Sogin, M.L., Elwood, H.J. and Gunderson, J.H. (1986) Proc. Natl. Acad. Sci. USA **83**, 1383-1387.
3. Baba, M.L., Dargy, L.L., Goodman, M. and Czelusniak, J. (1981) J. Mol. Evol. **17**, 197-213.
4. Horowitz, S. and Gorovsky, M.A. (1985) Proc. Natl. Acad. Sci. USA **82**, 2452-2455.
5. Caron, F. and Meyer, E. (1985) Nature (London) **314**, 185-188.
6. Preer, J.R., Jr., Preer, L.B., Rudman, B.M. and Barnett, A.J. (1985) Nature (London) **314**, 188-190.
7. Helftenbein, E. (1985). Nucleic Acids Res. **13**, 415-432.
8. Hanyu, N., Kuchino, Y., Nishimura, S. and Beier, H. (1986) EMBO J. **5**, 1307-1311.
9. Martindale, D.W., Goldberg, M., Martindale, H. and Bruns, P.J. (1984) Genetics **107**(suppl), S68-S69 (abstract).
10. Martindale, D.W., Martindale, H.M. and Bruns, P.J. (1986) Nucleic Acids Res. **14**, 1341-1354.
11. Wu, M., Allis, C.D., Richman, R., Cook, R.G. and Gorovsky, M.A. (1986) Proc. Natl. Acad. Sci. USA. **83**, 8674-8678.
12. Nielsen, H., Andreasen, P.H., Dreisig, H., Kristiansen, K. and Engberg, J. (1986) EMBO J. **5**, 2711-2717.
13. Martindale, D.W. and Bruns, P.J. (1983) Mol. Cell. Biol. **3**, 1857-1865.
14. Maniatis, T.F., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
15. Hattori, M. and Sakaki, Y. (1986) Anal. Biochem. **152**, 232-238.
16. Wilimzig, M. (1985) Trends in Genet. **1**, 158.
17. Henikoff, S. (1984) Gene **28**, 351-359.
18. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
19. Tabor, S. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA **84**, 4767-4771.
20. Allen, S.L. and Gibson, I. (1973) In Elliott, A.M. (ed.), Biology of Tetrahymena, Dowden, Hutchinson and Roth, Stroudsburg, Pennsylvania, pp. 307-373.
21. Horowitz, S., Bowen, J.K., Bannon, G.A. and Gorovsky, M.A. (1987) Nucleic Acids Res. **15**, 141-160.
22. Kaine, B.P. and Spear, B.B. (1982). Nature (London) **295**, 430-432.

23. Bannon, G.A., Bowen, J.K., Yao, M.-C. and Gorovsky, M.A. (1984) Nucleic Acids Res. **12**, 1961-1975.
24. Cupples, C.G. and Pearlman, R.E. (1986) Proc. Natl. Acad. Sci. USA **83**, 5160-5164.
25. Godiska, R. (1987) Mol. Gen. Genet. **208**, 529-536.
26. Nomoto, M., Imai, N., Saiga, H., Matsui, T. and Mita, T. (1987) Nucleic Acids Res. **15**, 5681-5697.
27. Calzone, F.J., Stathopoulos, V.A., Grass, D.G., Gorovsky, M.A. and Angerer, R.C. (1983) J. Biol. Chem. **258**, 6899-6903.
28. Kimmel, A.R. and Firtel, R.A. (1980) Nucleic Acids Res. **8**, 5599-5610.
29. Fink, G.R. (1987) Cell **49**, 5-6.
30. Doolittle, W.F. (1978) Nature (London) **272**, 581-582.
31. Gilbert, W., Marchionni, M. and McKnight, G. (1986) Cell **46**, 151-154.
32. Waring, R.B. and Davies, R.W. (1984) Gene **28**, 277-291.