A conjugation-specific gene (cnjC) from *Tetrahymena* encodes a protein homologous to yeast RNA polymerase subunits (RPB3, RPC40) and similar to a portion of the prokaryotic RNA polymerase α subunit (rpoA)

Duane W. Martindale

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

Received January 24, 1990; Revised and Accepted April 9, 1990

GenBank accession no. M29447

ABSTRACT

The *cnjC* gene from the protozoan *Tetrahymena thermophila* was completely sequenced. The deduced gene product was found to have significant sequence similarity to the yeast and prokaryotic RNA polymerase subunits involved with subunit assembly. Since *cnjC* is active only during the sexual stage (conjugation) of *Tetrahymena*'s life cycle, these results indicate it may be part of a novel type of transcriptional control. The yeast proteins to which the *Tetrahymena* *cnjC* is homologous are the 40 kd protein of RNA polymerases I and III (coded for by gene *RPC40*) and the third-largest subunit of RNA polymerase II (coded for by gene *RPB3*). The degree of similarity of the *cnjC* protein to the two yeast subunits was found to be greater than the similarity of the two yeast subunits to each other. The α subunit of the core RNA polymerase from prokaryotes (coded for by gene *rpoA*) was found to have regions of similarity to the *cnjC* protein as well as to the subunits encoded by *RPC40* and *RPB3*. Regions of high conservation among the four proteins are noted. The significance of these results is discussed.

INTRODUCTION

The RNA polymerases of eukaryotes are complex (see 1). Three RNA polymerases carry out the transcription duties that are accomplished by one RNA polymerase in eubacteria (see 2). RNA polymerase I (or A) transcribes ribosomal RNA, RNA polymerase II (or B) transcribes mRNA and RNA polymerase III (or C) transcribes small RNAs such as 5S RNA and tRNA. Each enzyme is composed of two large polypeptides (> 120 kd) and 8–11 smaller subunits (< 100 kd). Three of these smaller subunits (27 kd, 23 kd and 14.5 kd) are shared among all three RNA polymerases, while two (40 kd and 19 kd) are shared by RNA polymerases I and III, giving a total of approximately 22 distinct polypeptides making up the three RNA polymerases. The preparation of antibodies against the RNA polymerases (3–5) has led to the isolation of the genes for some of these subunits (6, 7). The largest and second-largest subunits of the three

eukaryotic RNA polymerases have been found to be homologous to the β' and β subunits respectively of the prokaryotic RNA polymerase (8–12).

In eubacteria, it is well established that the sigma subunit is necessary for the initiation of transcription and is not essential once transcription has started. Different sigma subunits are used for the initiation of different sets of genes because they enable the recognition of different sets of promoters (13). In eukaryotes, the ability of a particular RNA polymerase to recognize specific promoters and initiate transcription at the appropriate time and place is thought to come largely from the binding of transcription factors to specific DNA sequences within promoter and enhancer regions (see 14, 15). Recently, Cisek and Corden (16) showed that the cell-cycle control gene from mouse that is homologous to *cdc2* is involved in phosphorylating the unusual carboxyl-terminal domain of the largest subunit of RNA polymerase II. This region contains a heptapeptide repeat (52 times in the mouse) and is known to be highly phosphorylated when it is part of an actively transcribing RNA polymerase II. Thus, modifications of RNA polymerase subunits may be involved in some transcriptional control. There is also evidence that some general initiation factors directly bind RNA polymerase II and are involved in the accurate initiation of transcription (17–19). These factors have not yet been shown to be involved in promoter selection or to be developmentally controlled.

This study reports the sequence of a gene (*cnjC*) from the ciliated protozoan *Tetrahymena thermophila* whose derived amino acid sequence appears to be homologous to both the 40 kd subunit of RNA polymerase I and III from yeast and to the third largest subunit (44.5 kd) of RNA polymerase II from yeast. Furthermore, these three eukaryotic proteins are shown to be related to the bacterial RNA polymerase α subunit encoded by *rpoA*. Since *cnjC* is transcribed only during the conjugation stage of *Tetrahymena*'s life cycle (20; Laurie Stargell, personal communication), this gene's product may be exchanged or substituted during conjugation for the RNA polymerase subunit present during other stages of growth, possibly to control which genes are being transcribed. To my knowledge, the synthesis of stage-specific RNA polymerase subunits has not been previously described in eukaryotes.

```

-78  ACATTTTCTAATTTAAAAACAGGAATATCAAATAAATATAACAACCTCAAAAAGTTTAAAGGTCAATTATTAAAGTAA
      M   E   G   I   K   L   T   L   T   N   L   E   E   F   K   C   E   F   T   L
1    ATG GAA GGA ATT AAA TTA ACT CTT ACA AAC CTC GAA GAA TTC AAA TGC GAA TTT ACC CTA
      E   N   C   D   I   S   L   A   N   S   L   R   R   I   L   L   A   E   I   E
61   GAA AAT TGC GAT ATC TCA TTA GCT AAT TCA TTA AGA AGA ATT CTT TTG GCT GAA ATT GAA
      T   I   A   I   H   N   V   K   V   Y   E   N   T   S   S   L   P   D   E   Y
121  ACA ATA GCT ATC CAT AAC GTT AAA GTA TAT GAG AAT ACA TCT TCT CTT CCT GAT GAA TAC
      I   A   H   R   L   G   L   I   P   L   N   S   A   K   V   D   D   Y   N   Y
181  ATA GCA CAT AGA TTA GGT TTA ATT CCT TTA AAT TCT GCT AAA GTT GAT GAT TAT AAC TAT
      I   W   N   C   N   C   Q   G   E   C   D   R   C   T   I   H   F   D   L   R
241  ATT TGG AAT TGC AAT TGC CAA GGT GAA TGC GAT AGA TGT ACA ATT CAT TTT GAT TTG AGA
      V   S   N   F   T   N   E   V   L   E   V   T   S   L   D   L   K   Q   R   Q
301  GTG AGT AAC TTT ACA AAT GAA GTC TTA GAA GTT ACT TCT CTT GAT CTT AAA TAA AGA TAG
      D   H   T   D   E   D   A   V   R   P   I   K   V   Y   S   I   L   R   D   E
361  GAC CAC ACT GAT GAA GAT GCC GTT CGC CCT ATA AAG GTA TAT TCA ATT TTA CGT GAT GAA
      Y   G   N   K   R   E   V   G   I   P   I   I   K   L   S   K   G   Q   Q   I
421  TAT GGA AAT AAG AGA GAA GTA GGT ATT CCT ATT ATC AAG CTA TCT AAA GGA CAA TAA ATT
      H   F   E   C   E   A   Q   K   G   I   G   K   M   H   S   K   W   N   P   V
481  CAC TTT GAA TGT GAA GCC TAG AAA GGT ATC GGT AAA ATG CAT TCT AAA TGG AAT CCA GTA
      C   I   S   T   F   S   V   E   P   E   I   I   F   D   Q   E   I   Q   N   L
541  TGT ATT TCA ACT TTC TCA GTT GAA CCT GAA ATT ATT TTC GAT CAA GAA ATT TAA AAT TTA
      D   I   E   Q   K   K   A   L   V   N   A   C   P   T   K   V   F   G   I   N
601  GAT ATT GAA TAG AAG AAA GCT CTA GTA AAT GCA TGT CCT ACT AAA GTA TTT GGT ATC AAT
      P   H   T   N   T   L   E   V   Q   D   H   M   K   C   M   Y   C   E   E   C
661  CCT CAT ACC AAT ACT TTG GAA GTC TAA GAT CAC ATG AAA TGT ATG TAT TGT GAA GAA TGT
      V   Y   K   C   A   E   D   F   Q   K   P   K   L   I   K   I   D   H   K   K
721  GTT TAC AAG TGT GCT GAA GAC TTC TAA AAA CCT AAA TTA ATC AAA ATA GAC CAT AAA AAG
      D   K   F   F   F   K   V   E   T   T   G   V   M   K   P   V   E   V   L   R
781  GAT AAG TTT TTC TTC AAA GTC GAA ACT ACA GGA GTT ATG AAG CCT GTT GAA GTG TTA AGA
      R   G   F   S   T   L   R   Q   K   L   D   F   M   R   E   Q   I   Q   E   N
841  AGA GGT TTT TCA ACC TTG AGA TAG AAG CTT GAT TTT ATG AGA GAA TAA ATC TAA GAA AAT
      S   N   N   Q   F   E   Y   G   N   E   *
901  TCT AAT AAT TAA TTT GAA TAT GGT AAC GAG TGA GCTATTTACAGTCAAGTTTCTAGCTTAAAAATAGT
969  TTTTTTTATTTTAACTGGCTTTATAAAATTATTTATATTAAAAATGTCTTATCTTTGTGGTCAAAAAGTATGCATTTA
1048 TCAATCTAAATATTTTATAA

```

Figure 1. DNA sequence and derived amino acid sequence of the *cnjC* gene from *T. thermophila*. Numbering of the DNA sequence begins at the first codon.

MATERIALS AND METHODS

cDNA

The isolation and Northern analysis of the cDNA clone (pC7-1) corresponding to *cnjC* has been previously described (20, 21) as has its restriction map (22). The cDNA was made from mRNA isolated from a conjugating culture of two B strain derivatives of *Tetrahymena thermophila*. The cDNA was subcloned from pBR322 into pIBI30 and pIBI31 (IBI) which each contain an f1 origin of replication.

Southern

Genomic DNA from *Tetrahymena thermophila* (4 µg per lane) was digested with EcoR1, separated on a 1% agarose gel and transferred to Gene Screen filters (New England Nuclear) as described previously (22). The cDNA clone pC7-1 was nick-

translated and hybridized to the filters as described in 22, with the modification that the stringency of the hybridization was varied by using formamide concentrations of 30%, 40% or 50% in the hybridization buffer; salt concentration was kept constant at 5×SSC.

Sequencing

The cDNA was sequenced using the chain-termination method (23). Overlapping clones were constructed using the exonuclease III (Boehringer-Mannheim) digestion method of Henikoff (24) and single stranded DNAs isolated using helper phage supplied by IBI. Both DNA polymerase I (Pharmacia) and T7 DNA polymerase (US Biochemicals) were used with kits supplied by the respective companies.

Amino Acid	Codon	cnjC	cnjB	TET	Amino Acid	Codon	cnjC	cnjB	TET
Ala	GCT	6	11	128	Lys	AAA	17	22	36
	GCC	2	3	65		AAG	9	15	220
	GCA	2	5	4	Phe	TTT	9	17	10
	GCG	0	0	0		TTC	6	9	76
Arg	CGT	1	0	7	Pro	CCT	9	12	48
	CGC	1	0	0		CCC	0	0	59
	CGA	0	0	0		CCA	1	5	3
	CGG	0	0	0		CCG	0	0	2
	AGA	11	18	130	Ser	TCT	7	16	67
	AGG	0	2	2		TCC	0	2	67
Asn	AAT	16	24	17		TCA	6	8	13
	AAC	5	17	63		TCG	0	0	1
Asp	GAT	15	12	75	Thr	AGT	1	4	6
	GAC	3	7	33		AGC	0	12	4
Cys	TGT	8	5	6	Tyr	ACT	7	7	57
	TGC	5	8	24		ACC	3	2	67
Gln	CAA	3	11	30		ACA	6	4	5
	CAG	0	3	2		ACG	0	1	0
	TAA	8	19	14	Val	TAT	7	25	21
	TAG	4	14	5		TAC	2	10	47
Glu	GAA	29	31	121	Met	GTT	8	12	67
	GAG	2	10	6		GTC	3	1	76
Gly	GGT	8	1	139		GTA	6	8	3
	GGC	0	1	8		GTG	2	3	0
	GGA	4	4	10	Trp	ATG	6	10	51
	GGG	0	0	1		TGG	<u>2</u>	<u>4</u>	<u>19</u>
His	CAT	6	8	6	TOTAL		310	538	2250
	CAC	3	2	40	%GC		30.5	30.0	45.2
Ile	ATT	15	30	61	Fpr		.64	.60	.90
	ATC	7	13	73					
	ATA	4	14	2					
Leu	TTA	11	14	35					
	TTG	4	10	43					
	CTT	6	16	28					
	CTC	1	5	45					
	CTA	3	8	1					
	CTG	0	3	1					

Figure 2. Codon usage in cnjC, cnjB and the *Tetrahymena* genes (TET) analyzed in Martindale (28). The bold numbers under TET indicate class I codons and italicized numbers indicate class II codons (28). Fpr is the frequency with which preferred codons are used (28).

Computing

The DNA sequence was analyzed on a Macintosh Plus (Apple) using the DNA Inspector IIe programs (Textco) and the Pustell programs (IBI). The FastP program (25) was used to screen the National Biological Research Foundation database (release 19.0). Alignments of homologous sequences were done using the ALIGN program (Protein Identification Resource, NBRF).

RESULTS

The cDNA clone corresponding to cnjC (pC7-1) was one of several obtained when poly A⁺ RNA from early conjugation (2 to 3 h after mixing two starved cultures with different mating types) was used to prepare a cDNA library (20). Northern hybridizations using the pC7-1 cDNA against both total RNA and poly A⁺ RNA isolated from different stages of *Tetrahymena*'s life cycle (20, 21) indicated that the corresponding

gene was active only during early conjugation, with the amount of mRNA peaking at the crescent stage of micronuclear elongation. This cytological stage is thought to correspond to the pachytene stage of meiotic prophase. Because it appears to be active only during conjugation, this gene was called cnjC (22). Two other conjugation-specific genes were also analysed in these studies (cnjA and cnjB) and they have transcription patterns that are similar to cnjC (21).

The stage-specific transcription of cnjC has recently been confirmed by another method. Laurie Stargell, working in the laboratory of Martin Gorovsky at the University of Rochester, has done run-on transcription experiments using nuclei isolated from *Tetrahymena* and various cDNA probes including pC7-1 (personal communication). This technique detects genes that are being actively transcribed whereas the Northern technique detects the presence of an RNA whether or not it is being actively transcribed. Her results confirmed the results that we obtained with Northern analysis: for pC7-1 she found that nuclei from

cnjC: 1- M E G I K L T L T N L E E F K C E F T L L S N U D L A M A N S L R R I L L A E I E
 RPB3: 3- E E G P Q U K I I R E A S K D N U D F I L L S N U D L A M A N S L R R I U M I A E I P
 RPC40: 37- K K D F E U N I S S L D A R E A N F D L I N I D T S I A N A F R R I M I S E U P

cnjC: 41- T I A I H N U K U V E N T S S L P D E V I A H R L G L I P L N S A K U D D Y N V Y
 RPB3: 43- T L A I D S U E U E T N T T U L A D E F I A H R L G L I P L Q S M D I E Q L E Y
 RPC40: 77- S U A A E Y U Y F F N N T S U I Q D E U L A H R I G L U P L - - K U D P D M L

cnjC: 81- I W N C N C Q G E C D R C T I H F D L R U S N - F T N E U L E U T S L D L K - Q
 RPB3: 83- S R D C F C E D H C D K C S U U L T L Q A F G - E S E T T N U V Y S K D L U - I
 RPC40: 114- T W - - - - - - - - - U D S N L P D D E K F T D E N T I U L S L N V K C T

cnjC: 119- R Q D H T D E D A U R P I K U Y S I L R D E - - - - - - - - - - - - - - -
 RPB3: 121- U S N L M G R N I G H P I -
 RPC40: 142- R N P D A P K G S T D P K E L Y N N A H U Y A R D L K F E P Q G R Q S T T F A D

cnjC: 141- - Y G N K R E U G I P I I K L S K G Q Q I H F E C E A Q K G I G K M H S K W N P
 RPB3: 134- - I Q D K E G N G U L I C K L R K G Q E L K L T C U A K K G I A K E H A K W G P
 RPC40: 182- C P U U P A D P D I L L A K L R P G Q E I S L K A H C I L G I G G D H A K F S P

cnjC: 180- U C I S T F S U E P E I I F D Q E I Q N L D I E Q K K A L U N A C P T K U F G I
 RPB3: 173- A A G I E F E Y D P W N K L K H T D Y W Y E Q D S A K E W P Q S K N C E Y E D P
 RPC40: 222- U S T A S Y R L L P Q I N I L Q P I K G - - - E S A R R F Q K C F P P G U I G I

cnjC: 220- N P H T N T L E U Q D H M K C M Y C E E C U - Y K C A E D F Q K P K L I K I D H
 RPB3: 213- P N E G D P F D Y K A Q A -
 RPC40: 259- D E G S D E A Y U K D A R K D T U S R E U L R Y - - - E E F A D K - - U K L G R

cnjC: 259- K K D K F F F K U E T T G U M K P U E U L R R G F S T L R Q K L - - - - -
 RPB3: 226- - - D T F Y M N U E S U G S I P U D Q U U U R G I D T L Q K K U A S I L L A L T
 RPC40: 294- U R N H F I F N U E S A G A M T P E E I F F K S U R I L K N K A - - - - -

cnjC: 291- - - D F M R E Q I Q E N S N N Q F E Y G N E *
 RPB3: 264- Q M D Q D K U N F A S G D N N T A S N M L G S N E D U M M T G A E Q D P Y...(<+18>
 RPC40: 326- - - E Y L K N C P I T Q *

Figure 3. Alignment of the protein products of *cnjC*, RPB3 and RPC40. Identical amino acids are boxed.

rapidly growing cells have no detectable signal above background; nuclei from starved cells have either no detectable signal or, in one experiment, very slight signal (10 cps above background); and nuclei isolated from conjugating cells (four hours after mixing two starved strains with different mating types) had at least a 13-fold increase over the possible signal seen in starved cells.

No difference was seen between the restriction maps of the pC7-1 cDNA and the genomic DNA containing *cnjC* (22). Thus, it is likely no introns are present in this gene. The *Eco*R1 site beginning at basepair 98 in Figure 1 was not cut in any DNA isolated from macronuclei of different strains, but was cut in micronuclear DNA. We thus proposed that this site was completely and specifically methylated at A100 in the macronucleus only (22).

The cDNA present in pC7-1 was completely sequenced as described in Methods. The DNA sequence as well as the deduced amino acid sequence is shown in Figure 1. The poly A+ tail was present in the cDNA so the complete 3' end of the mRNA is represented (132 bases in the untranslated tail). The cDNA insert, including a poly A tail of approximately 50 base pairs, 15 Gs at the 5' end and 22 Cs at the 3' end was 1230 bp. We estimated the mRNA to be 1300 bases (20), therefore the cDNA

may be close to full length given the uncertainty in determining RNA size. The coding region appears to begin at the first ATG, 79 bp from the 5' end. The putative protein is 310 amino acids in length with a calculated molecular weight of 36.0 kd and an estimated pI of 5.4. The coding region shown is supported by three different criteria: 1. The regions flanking the putative coding region are AT rich as in other *Tetrahymena* genes (see 26; the 5' untranslated region is 18% GC, the coding region is 30.5% GC and the 3' untranslated region is 19% GC). 2. The 5' flanking region is A-rich (52% A) as found in other *Tetrahymena* genes (26; Martindale, unpublished observations). 3. This region was shown to be a likely coding region by the codon usage program (27) using data from 2788 codons of *Tetrahymena* (28) and the positional base preference program (27).

The codon usage of the putative protein was examined (Figure 2) and found to resemble that of another conjugation-specific gene *cnjB* (29) with a similar transcription pattern to *cnjC*. The *cnjB* gene was found to differ significantly in its codon usage from the other *Tetrahymena* genes which had been sequenced (28). It was proposed that stage-specific genes would not have the same strong selection pressure to use optimal codons (codons corresponding to most abundant tRNAs) as would genes which

a.

Figure 4. Three conserved regions found in proteins with sequence similarity to the protein product of *cnjC*. a. Identical amino acids are boxed; identical amino acids between the *rpoA* products were not boxed unless they were also found in a eukaryotic protein. a,b. The amino acid sequences of the products of *Tetrahymena* *cnjC* (Tetrah; this paper), *S. cerevisiae* RPB3 (yeast or Sc; 31), *S. cerevisiae* RPC40 (yeast or Sc; 30) and of the *rpoA* genes from *E. coli* (Ec; 32), *Bacillus subtilis* (*B. subt* or Bs; 33) and Spinach chloroplast (plastd or So; 34) are aligned. The *B. subtilis* *rpoA* has been only partially sequenced. b. Region I is indicated by a solid box, region II by vertical stripes and region III by diagonal stripes. The numbers on the right (C-terminal end) are the lengths in amino acids of the various proteins.

are highly expressed in normal exponential growth (28). Without this pressure a selection or drift towards codons that are more AT rich may occur. Alternatively, a different set of codons are optimal during conjugation. This change to AT richness is seen in both *cnjB* (30% GC) and *cnjC* (30.5% GC); all other *Tetrahymena* genes examined are greater than 40% GC (average 45% GC; 28). Also a reduction in optimum codon usage in *cnjB* and *cnjC* is evident when the frequency of preferred codons (Fpr), which is thought to be a reflection of optimum codon usage (28), is compared between these genes (Fpr of 0.60 and 0.64 respectively) and the highly expressed *Tetrahymena* genes that were analyzed previously (28) (Fpr for 2250 codons = 0.90). Numerous examples can be seen of *cnjB* and *cnjC* using a more AT rich synonymous codon (e.g., Asn, His, Lys, Phe, Pro, Thr, Tyr) than the other *Tetrahymena* genes examined.

Using the derived amino acid sequence of *cnjC*, I initiated a computer search of the National Biological Research Foundation database using the FASTP program (25). To simplify the following discussion I will refer to the protein products of the genes *cnjC*, RPB3, RPC40 and *rpoA* as CnjC, Rpb3, Rpc40 and RpoA respectively. Rpb3 from yeast (30) was identified as having significant sequence identity to CnjC (Z value greater than 10; 25). Rpb3 is a subunit of RNA polymerases I and III and is approximately 40 kd (37.6 kd). The yeast gene (RPB3) for the third largest subunit (45 kd) of RNA polymerase II (Rpb3) was cloned and sequenced recently (31) and found to be homologous to RPC40. When the sequence of CnjC was compared to that of Rpb3, significant identity was again seen. Figure 3 shows an alignment between CnjC, Rpb3 and Rpb3: Rpb3 has 84 amino acids identical to the 310 amino acids of CnjC

(27%) while Rpc40 has 80 (26%). If conservative amino acid changes are included, the similarity between the sequences is even more striking. When regions are examined in which all three sequences overlap (245 amino acids), CnjC is found to be 31% identical to Rpb3 and 30% identical to Rpc40. Rpb3 and Rpc40 share only 25% identity over this region.

Three regions of CnjC and the yeast proteins appeared to be particularly conserved (regions I, II and III; see Figure 4). Over the 116 amino acids in these regions, CnjC and Rpb3 share 53% identity, while CnjC and Rpc40 as well as Rpb3 and Rpc40 have 41% identical amino acids. Regions I and III also share some sequence identity with the α subunit of the prokaryotic RNA polymerase (RpoA); out of a total of 77 amino acids in this region, the *E. coli* RpoA (32) shares 26 with CnjC (34%), 17 with Rpb3 (22%) and 20 with Rpc40 (26%). The partial sequence of RpoA from *Bacillus subtilis* (33) and the sequence of RpoA from Spinach chloroplast (34) are also included in Figure 4. The similarity of Rpb3 to RpoA was first noted by Kolodziej and Young in a recent preliminary report (35).

To determine whether there are other *Tetrahymena* genes homologous to cnjC, blots of *Tetrahymena* genomic DNA were hybridized with a cnjC cDNA probe (pC7-1; 1230 bp) at different stringencies. Stringency was varied by changing the percent formamide (see 36); hybridization temperature (42°C) and salt concentration ($5 \times \text{SSC}$) were unchanged. At concentrations of 40% formamide or higher, only restriction fragments corresponding to cnjC were detected (Figure 5). At 30% formamide, extra bands were detected. The three extra bands that are most prominent are denoted with arrows in Figure 5 and correspond to EcoRI fragments of 3.6 kb, 2.2 kb and 2.0 kb.

DISCUSSION

The coding region of cnjC was completely sequenced. The derived amino acid sequence (CnjC) suggests it is homologous to RNA polymerase subunits from yeast and *E. coli* (Figures 3 and 4). The yeast subunits to which it is similar are the 40 kd subunit found in both RNA polymerases I and III (Rpc40, encoded by the gene RPC40), and the third largest (45 kd) subunit of RNA polymerase II (Rpb3, encoded by the gene RPB3). There is also similarity to the α subunit of the eubacterial (and chloroplast) RNA polymerases (RpoA, encoded by the rpoA genes from various organisms) (Figure 4). What makes this observation surprising is that cnjC appears to be active only during the sexual cycle (early conjugation) of *Tetrahymena* (20; L. Stargell personal communication) and is thus unlikely to code for the RNA polymerase subunits present at life cycle stages other than conjugation.

The putative cnjC protein was aligned with the RPB3 and RPC40 gene products (Figure 3). CnjC was found to be about equally related to both by looking at the overall alignment (27% identical to Rpb3; 26% identical to Rpc40) or by examining only the regions of overlap between the three proteins (245 amino acids; CnjC is 31% identical to Rpb3 and 30% identical to Rpc40). Rpb3 and Rpc40 share only 25% identity in the region where all three overlap. Thus, the unexpected situation exists in which a stage-specific protozoan protein has more similarity to two homologous RNA polymerase subunits from yeast than the yeast subunits have with each other (Figure 6). One possible explanation for this would be that the *Tetrahymena* cnjC gene product functions in all three RNA polymerases whereas the two yeast subunits have adapted to their specific RNA polymerases

Figure 5. Southern hybridizations at different stringencies of cnjC cDNA (pC7-1) against *Tetrahymena* genomic DNA. The lane designated 50 had 50% formamide in its hybridization buffer, while the lane designated 30 had 30% formamide. The genomic DNA was cut with EcoRI. The restriction fragments designated 1.4 and 0.5 correspond to 1.4 kb and 0.5 kb EcoRI fragments containing the cnjC gene. The arrows point to fragments detectable only after hybridization at low stringency (30% formamide) and correspond to EcoRI fragments of 3.6 kb, 2.2 kb and 2.0 kb.

Figure 6. Percent amino acid identity over the regions of overlap (245 amino acids) between the three proteins (see Figure 3).

(I and III, or II), and do not need to retain certain regions necessary for them to function in the other. This argument could also accommodate the findings seen with the α subunit of the *E. coli* RNA polymerase (RpoA). Over the regions in which the eukaryotic proteins share similarity with RpoA (77 amino acids, Figure 4), the putatively undifferentiated cnjC product shares 26 amino acids (34%) with RpoA while Rpb3 shares 17 (27%) and Rpc40 shares 20 (26%). Although these arguments are very speculative, they make predications that can be tested experimentally.

The α subunit of the prokaryotic RNA polymerase is involved in the assembly of the core enzyme (see 37). RNA polymerase contains two α subunits whereas the other core subunits are present once. The ability of the prokaryotic RNA polymerase to recognize specific promoters comes from the sigma subunit and a change in the type of sigma subunit will change which promoters are recognized.

The RPB3 and RPC40 gene products also appear to be involved in the assembly of the RNA polymerase complex. Mann et al. (30), through *in vitro* mutagenesis, showed that the RPC40 gene product (~40 kD protein shared by RNA polymerases I and III) is essential for cell viability and the assembly of RNA polymerases I and III. Similarly, Kolodziej and Young (31) showed RPB3 to be essential and described the isolation of a ts mutant of the RPB3 gene that prevented assembly of RNA polymerase II at the restrictive temperature. As with the α subunit of the prokaryotic RNA polymerase, the RNA polymerase II subunit coded for by RPB3 was found to be present at twice the molar ratio of the two largest subunits (31).

If the *cnjC* gene product is an RNA polymerase subunit, then the fact that it is active only during early conjugation (20, 21; L. Stargell, personal communication) would suggest that it may be involved in the regulation of transcription at this time. Several possible functions can be envisioned: 1. It may be essential for the expression of some genes active only at this time, possibly by allowing the recognition of their promoters either directly or indirectly. 2. It may stabilize the three RNA polymerases so that they may function in the cellular environment of meiosis. 3. It may allow the assembly of RNA polymerases within the micronucleus and thus allow transcription to occur within this nucleus. The micronucleus is normally transcriptionally silent, but appears to be transcriptionally active at approximately the same time that the *cnjC* gene is expressed (21, 38). 4. It may be part of a stage-specific RNA polymerase made during conjugation. If my earlier reasoning is correct and the *cnjC* protein is found in all three RNA polymerases, then possibility 1 is unlikely since there should not be a need to change the promoter specificity of RNA polymerase I which transcribes a single type of gene. Support for possibilities 2 and 3 comes from the findings that the homologous proteins from yeast and bacteria are primarily involved in RNA polymerase assembly (30, 31, 37). If possibility 4 is true, then multiple genes for other subunits should exist.

There are likely to be at least two *Tetrahymena* RNA polymerase subunit genes with homology to *cnjC* that are expressed at life cycle stages other than conjugation; these would correspond to the genes for the subunit shared by RNA polymerase I and III and the subunit for RNA polymerase II. The Southern hybridizations seen in Figure 5 are compatible with there being at least two *Tetrahymena* genes with sequence similarity to *cnjC*. The fact that the stringency of hybridization had to be reduced significantly to detect these other sequences suggests that the *Tetrahymena* genes homologous to *cnjC* will not have a high degree of similarity to *cnjC* (probably less than 75% DNA sequence identity; 36). This may not be too surprising given the relatively low sequence identity found between the two homologous yeast subunits Rpb3 and Rpc40.

Efforts are now being made to obtain antibodies to the *cnjC* gene product and thus determine whether it is an RNA polymerase subunit and, if so, which polymerase(s) it is found in. *In situ* hybridizations with such antibodies should reveal whether there is localization to the micronucleus (possibility 3).

Attempts will be made to isolate the *Tetrahymena* RNA polymerase subunit gene(s) homologous to *cnjC* and compare their sequence both to *cnjC* and the homologous yeast genes and the *rpoA* gene from prokaryotes. To my knowledge, this study is the first to present evidence suggesting that RNA polymerase subunit exchange occurs in eukaryotes in a stage-specific manner.

ACKNOWLEDGEMENTS

I would like to thank Peter Kolodziej for communicating the sequence of RPB3 before publication. This work was supported by an Operating grant from the Natural Sciences and Engineering Council of Canada.

REFERENCES

1. Sentenac, A. (1985) CRC Crit. Rev. Biochem. **18**, 31–91.
2. Burgess, R.R., Erickson, B., Gentry, D., Gribskov, M., Hager, D. Lesley, S., Strickland, M. and Thompson, N. (1987) in Reznikoff, W.S. et al. (eds), RNA Polymerase and the Regulation of Transcription. Elsevier Science Publication Co., Inc., pp. 3–15.
3. Buhler, J.M., Huet, J., Davies, K.E., Sentenac, A. and Fromageot, P. (1980). J. Biol. Chem. **255**, 9949–9954.
4. Vilamitjana, J., Baltz, T., Baltz, D. and Barreau, C. (1983) FEBS Lett. **158**, 343–348.
5. Huet, J., Riva, M., Sentenac, A. and Fromageot, P. (1985) J. Biol. Chem. **260**, 15304–15310.
6. Young, R.A. and Davis, R.W. (1983) Science **222**, 778–782.
7. Riva, M., Memet, S., Micouin, J.Y., Huet, J., Treich, I., Dassa, J., Young, R., Buhler, J.M., Sentenac, A. and Fromageot, P. (1986) Proc. Natl. Acad. Sci. USA **83**, 1554–1558.
8. Allison, L.A., Moyle, M., Shales, M. and Ingles, C.J. (1985) Cell **42**, 599–610.
9. Biggs, J., Searles, L.L. and Greenleaf, A.L. (1985) Cell **42**, 611–621.
10. Faust, D.M., Renkawitz-Pohl, R., Falkenburg, D., Gasch, A., Bialojan, S., Young, R. and Bantz, E. (1986) EMBO J. **5**, 741–746.
11. Sweetser, D., Nonet, M. and Young, R.A. (1987) Proc. Natl. Acad. Sci. USA **84**, 1192–1196.
12. Memet, S., Gouy, M., Marck, C., Sentenac, A. and Buhler, J.-M. (1988) J. Biol. Chem. **263**, 2830–2839.
13. Helmann, J.D. and Chamberlin, M.J. (1988) Annu. Rev. Biochem. **57**, 839–872.
14. Johnson, P.F. and McKnight, S.L. (1989) Annu. Rev. Biochem. **58**, 799–839.
15. Struhl, K. (1989) Annu. Rev. Biochem. **58**:1051–1077.
16. Cisek, L.J. and Corden, J.L. (1989) Nature **339**, 679–684.
17. Burton, Z.F., Ortolan, L.G. and Greenblatt, J. (1986) EMBO J. **5**, 2923–2930.
18. Zeng, X.-M., Moncollin, V., Egly, J.-M. and Chambon, P. (1987) Cell **50**, 361–368.
19. Sopta, M., Burton, Z.F. and Greenblatt, J. (1989) Nature **341**, 410–414.
20. Martindale, D.W. and Bruns, P.J. (1983) Mol. Cell. Biol. **3**, 1857–1865.
21. Martindale, D.W., Allis, C.D. and Bruns, P.J. (1985) J. Protozool. **32**, 644–649.
22. Martindale, D.W., Martindale, H.M. and Bruns, P.J. (1986) Nucl. Acids Res. **14**, 1341–1353.
23. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
24. Henikoff, S. (1984) Gene **28**, 351–359.
25. Lipman, D.J. and Pearson, W.R. (1985) Science **227**, 1435–1441.
26. Horowitz, S., Bowen, J.K., Bannon, G.A. and Gorovsky, M.A. (1987) Nucl. Acids Res. **15**, 141–160.
27. Staden, R. (1984) Nucl. Acids. Res. **12**, 551–567.
28. Martindale, D.W. (1989) J. Protozool. **36**, 29–34.
29. Martindale, D.W. and Taylor, F.M. (1988) Nucl. Acids Res. **16**, 2189–2199.
30. Mann, C., Buhler, J.-M., Treich, I. and Sentenac, A. (1987) Cell **48**, 627–637.
31. Kolodziej, P. and Young, R.A. (1989) Mol. Cell. Biol. **9**, 5387–5394.
32. Meek, D.W. and Hayward, R.S. (1984) Nucl. Acids Res. **12**, 5813–5821.
33. Suh, J.-W., Boylan, S.A. and Price, C.W. (1986) J. Bacteriol. **168**, 65–71.
34. Sijben-Müller, G., Hallick, R.B., Alt, J., Westhoff, P. and Herrmann, R.G. (1986) Nucl. Acids Res. **14**, 1029–1044.

35. Kolodziej, P. and Young, R. (1989) Abstracts from Yeast Genetics and Molecular Biology Meeting, Atlanta, Georgia. p. 153.
36. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press.
37. Ishihama, A. (1988) *Trends in Genetics* **4**, 282–286.
38. Sugai, T. and Hiwatashi, K. (1974) *J. Protozool.* **21**, 542–548.