A repetitive DNA sequence in *Paramecium* macronuclei is related to the β subunit of G proteins

James Forney and Karyn Rodkey

Department of Biochemistry, Purdue University, West Lafayette, IN 479079, USA

Received July 7, 1992; Revised and Accepted September 17, 1992

GenBank accession no. M96642

ABSTRACT

A repeated DNA sequence has been identified in the macronucleus of several *Paramecium* species. In *P.tetraurelia* the repeat was identified in the subtelomeric region of four randomly selected telomere clones, as well as downstream of the A type variable surface protein gene. The complete sequence of the A gene linked repeat consists of 15 tandem repeats of exactly 126 nucleotides that contain an open reading frame with significant similarity to the β subunits of trimeric G proteins. The most striking consensus feature is the amino acid sequence DX Ω WD where X is any amino acid and Ω is I, L, or V spaced at precise 42 amino acids intervals. This sequence and spacing are found in G-protein beta subunits and other members of this protein motif family. Analysis of the five cloned telomeric restriction fragments showed the repeats can be found in either orientation with respect to the telomere. Poly(A) RNA transcripts containing this sequence have been identified in *Paramecium tetraurelia*. The conserved presence of this sequence in several species of *Paramecium* suggests an important physiological function, and the study of this repeat may reveal information about the evolution of this common protein motif.

INTRODUCTION

Ciliated protozoa possess two functionally and morphologically distinct types of nuclei within a single cell. Both types, micronuclei and macronuclei, arise from a single diploid fertilization nucleus. In *Paramecium tetraurelia*, the 2 micronuclei remain diploid and the DNA in the transcriptionally active macronucleus is amplified to about 800C. The formation of the new macronucleus involves the loss and rearrangement of sequences as well as chromosomal fragmentation (reviewed in 1). The resulting macronuclear DNA fragments are approximately 450 kb in size (2, 3). Estimates of the genome size of *P.tetraurelia* (about 9×10^4 kb) (4) and the number of micronuclear chromosomes (40–45) (5), suggest that each micronuclear chromosome contains roughly 2,000 kb. The estimated sizes of micronuclear and macronuclear DNA suggest that micronuclear chromosomes are fragmented into roughly 4 macronuclear DNA molecules, and many of the macronuclear telomeres are derived from internal regions of micronuclear chromosomes. The linear DNA fragments in the macronucleus are not true chromosomes; macronuclear division occurs

amitotically, and DNA fragments presumably do not contain centromeres. Various models for maintaining genic balance in the macronucleus have been proposed, but the molecular mechanism that maintains DNA copy number is unknown (reviewed in 2). Efforts to understand macronuclear gene expression and copy number control will require basic information about the sequence organization of the *Paramecium* macronuclear genome. Thus far studies of chromosomal organization in *Paramecium* have been limited to ribosomal RNA genes and variable surface protein genes (3, 6, 7). Investigations of other ciliates, such as *Tetrahymena* and *Oxytricha* have identified micronuclear specific repetitive DNA sequences, but repetitive macronuclear sequences have not been characterized, with the notable exceptions of ribosomal RNA genes and telomere sequences (reviewed in 1).

In this paper we characterize a repetitive DNA sequence found in the macronucleus of several *Paramecium* species that encodes a predicted polypeptide related subunit of trimeric G proteins. This motif was first observed in the β subunit of mammalian transducin (reviewed in 1) and has since been identified in a large number of proteins (11). The consensus sequence includes several repeats of approximately 43 amino acids that contain a number of conserved amino acids including a tryptophan–aspartic acid pair (11, 12). The precise biochemical function of the motif remains unknown, but proteins containing it participate in a wide variety of functions such as RNA splicing (*PRP4*, 11), transcriptional repression in yeast (*TUP1*, 13), microtubule-dependent processes (*CDC20*, 14), and *Drosophila* neurogenesis (Enhancer of split, 15). It has been suggested by Goebel and Yanagida (16) that proteins containing the β -transducin repeat may interact with partner proteins containing a 34 amino acid repeat called TPR. In support of this theory, Ssn6 (a TPR protein) and Tup1 have been shown to exist as a functional complex in yeast (17, 18). Individual proteins in the β -transducin repeat family may have evolved to interact with specific members of the TPR family. Analysis of this repeated sequence in *Paramecium* may yield information about the evolution of this common protein motif as well as its function in *Paramecium*.

MATERIALS AND METHODS

Strains

Paramecium tetraurelia stock 51 s, *P.biaurelia* stock 7s, *P.multimicronucleatum* stock 203 and *P.sonneborni* were obtained from the Indiana University collection (John Preer,

Bloomington, IN). *P. caudatum* (Nasco #0926) was provided by Thomas Cole (Wabash College, Crawfordsville IN).

Isolation of RNA and DNA from *Paramecium*

RNA was isolated using a guanidine·HCl method previously described by Preer, *et al.* (19). DNA was isolated as follows. Packed cells (0.1–0.4 ml) were resuspended in 0.7 ml of culture fluid and squirted into 2.1 ml of lysing solution (0.01 M Tris, 0.05 M sodium EDTA, 1% SDS pH 9.5) at 65°C. After 10 minutes, 7 ml of saturated CsCl was added and the solution was centrifuged in a VTi65.1 rotor at 55,000 rpm for 10–24 hours. DNA containing fractions were collected and dialysed overnight against TE.

Telomere clones

Construction of telomere clones was previously described by Forney and Blackburn (20).

Sequence analysis

Restriction fragments were subcloned into pUC119 or 118 and exonuclease III was used to construct a set of nested deletions according to the method of Henikoff (21). The resulting plasmids were transformed into *E. coli* strain JM101 and single strand DNA was produced according to standard protocols (22). Sequencing reactions were performed using the Sequenase version 2.0 DNA sequencing kit (US Biochemicals, Cleveland, OH). DNA sequence was determined from both strands and analyzed using the programs of the University of Wisconsin GCG Sequence analysis software package Version 6.2 Copyright(c) 1989 John Devereaux (23). The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the accession number M96642.

Southern and Northern analysis

Southern and Northern blots were prepared according to Maniatis (22). Filters were washed in 10×Denhardt, solution, 0.1% SDS, 0.2 M phosphate buffer and 5×SET (1×SET = 0.15 M NaCl, 0.03 M Tris, 2 mM EDTA) at 65°C for one hour. The filters were then incubated with hybridization solution (0.2 M phosphate buffer, 1×Denhardt's solution, 5×SET, 0.25% SDS) and after one hour the labeled probe was added. Unless stated otherwise in the figure legends, filters were washed three times, 30 minutes each in 5×SET, 0.1% SDS, 0.1% sodium pyrophosphate, 0.025 M phosphate buffer at 65°C.

RESULTS

Identification of repeated DNA in subtelomeric regions

As part of a study of d48, a mutant cell line that is defective in its ability to process the A surface protein gene into the macronucleus (isogenic with *P. tetraurelia* stock 51), terminal macronuclear HindIII restriction fragments were cloned from the mutant and used to compare telomeric fragments from mutant and wild type cells (S.Fong and J.Forney, unpublished data). Subtelomeric restriction fragments were purified and hybridized to Southern blots containing wild type and d48 total cellular DNA. Some of these fragments hybridized to multiple bands on a Southern blot containing HindIII digested genomic DNA although theoretically only a single band was expected (for an example see Figure 4A, lane 1). Hybridization experiments using a HindIII–BglII fragment from one of the clones (pME1 6) showed that of the 7 telomeric clones with inserts greater than 4 kb, 5

contained related sequences in their subtelomeric regions (data not shown). Of those telomere clones with inserts less than 4 kb only one clone out of 10 hybridized to this fragment. Four of the clones that hybridized to the fragment were selected for further study. The restriction map of each clone is unique and none share any obvious similarity for the six base recognition enzymes that were used (Figure 1). In addition to these randomly selected telomere clones, the telomeric region beyond the 3' end of the A surface protein gene hybridized to the pME16 fragment (20 and see Figure 1). This observation strengthens the correlation between the repeated sequence and a telomeric location, since for this example the identification of the repeated sequence was not dependent on its initial selection as a telomere clone. This small sample of telomeric clones suggests that as many as 75% of the macronuclear telomeres (6 of 8) contain the repeated sequence within 4 kb of the end. Unfortunately, without additional information concerning the number of locations the repeat is found in the genome, we cannot conclude the repeated sequence is preferentially located near telomeres (see Discussion).

The repeated sequence consists of tandem repeats related to the subunit of trimeric β proteins

A portion of each telomere clone as well as the region 3' to the A gene was sequenced and found to contain tandem repeats of precisely 126 bp. The primary sequence of the repeat unit is conserved, although some regions of the repeat have more variation than others. The region downstream of the A gene contained a total of 15 repeats located within a 2.8 kb open reading frame (data not shown, Genbank accession number M96642). The nucleotide sequence for a segment that includes 3 complete repeats is shown in Figure 2. Although the repeats are highly conserved, individual nucleotide differences can be detected in each repeat (for example, see the end of the repeat sequence shown in Figure 2). About 2 1/2 repeats of the sequence downstream of the A surface protein gene was used to search the Genbank database. Since the tandem repeats were precisely 126 bp in size it was deduced that significant similarity to a protein product could be confirmed if identical residues were spaced every 42 amino acids (126 bp/3). Interestingly, a large number of proteins met this criterion. They included the Prp4 protein, a component of the yeast RNA splicing complex (11), Tup1 a yeast protein involved in transcriptional repression (13), and the β subunits of trimeric G proteins (12). A representative sample and their alignment with the predicted amino acid sequence of the *Paramecium* repeats (hence forth referred to as P126 repeats) is shown in Figure 3. The most conspicuous feature of the identity is the conserved sequence DX₄ΩWD, where X is any amino acid and Ω represents leucine, isoleucine or valine. Other conserved amino acids include a glycine–histidine pair, and a proline residue in each repeat (boxed in Figure 3). As found in other β transducin (the G protein involved in visual excitation) related proteins, the amino acid following the tryptophan–aspartate pair is a hydrophobic residue. Analysis of the amino acid sequence of the P126 repeats showed that most positions within the repeats highly conserved, one exception is the residue preceding the first aspartate in the sequence DX₄ΩWD where 10 different amino acids are found in the 15 repeats (data not shown).

Selected restriction fragments were subcloned from each pME clone and sequenced to confirm the presence of the tandem 126 bp repeats and determine their orientation with respect to the telomere. All clones contained the tandem repeats (data not

Figure 1. Maps of cloned genomic fragments containing repetitive DNA. The top line shows a map of the A surface protein gene region, the dashed line with an arrow indicates the direction of transcription of the A gene. The location of the 2.1 kb EcoRI–XhoI fragment (used as a probe in Figures 4–6) is indicated and this region is shown below at the same scale as four pME (Macronuclear End) clones (16, 18, 19, 21). The solid lines with arrows indicate the region sequenced from each clone and the orientation of the repeats with respect to the telomere (the arrow is drawn to indicate the 5' to 3' direction of transcription if the repeats are transcribed to produce the predicted product in Figure 3). E, EcoRI; H, HindIII; B, BglII; X, XbaI; Xh, XhoI; S, SacI; BL, BamHI linker ligated to the telomeric repeats.

P126-825	G A T A A T T C T A	T T T T T T T T T T	G G A T T T T T T T	A C A T T T T T T T	A A A A A T T G A A	874
P126-951	G A T T T A A T T A	T T T T T T T T T T	T T T T T T T T T T	A A T T T T T T T T	A A A A A T T G A A	1000
P126-1077	G A T A A A A T T A	T T T T T T T T T T	T T T T T T T T T T	A A T T T T T T T T	A A A A A T T G A A	1126
P126-825	A A T A A T T T T T	C C G C T T T T T T	T T T T T T T T T T	A T T T T T T T T T	A A T T T T T T T T	924
P126-951	T T T T A T T T T T	C C G C T T T T T T	T T T T T T T T T T	A T T T T T T T T T	A A T T T T T T T T	1050
P126-1077	A T T A A A T T T T	T T T T T T T T T T	T T T T T T T T T T	A T T T T T T T T T	T G T T T T T T T T	1176
P126-825	G T A A A A A T T T	A A T T T T T T T T	A A T T T T T T T T	A T T T T T T T T T	A A T T T T T T T T	A T 950
P126-951	G T A T T A A T T T	A A T T T T T T T T	A A T T T T T T T T	A T T T T T T T T T	A A T T T T T T T T	T A 1076
P126-1077	G T A G T T A A T T	A A T T T T T T T T	A A T T T T T T T T	A T T T T T T T T T	A A T T T T T T T T	T G G 1202

Figure 2. Nucleotide sequence of the 126 bp tandem repeats from the A gene linked repeat region of *P.tetraurelia* stock 51. Alignment of three consecutive tandem repeats is shown, the nucleotide numbering corresponds to the sequence submitted to Genbank (accession number M96642).

shown), but they were found in different orientations. Figure 1 shows the orientation of the repeats such that the arrows point in the 5' to 3' direction assuming these genomic sequences are transcribed to produce the protein product related to β transducin (we have not shown that these particular regions are transcribed). Thus, the A gene linked P126 repeat and pME18 face away from the telomere and pME16, 19 and 21 face toward the telomere.

Macronuclear distribution and conservation between species

If the repetitive nature of this *Paramecium* sequence is physiologically important, then the repeats should be found in many species of *Paramecium*. Alternatively, a narrow distribution among species would suggest that a recent amplification event had occurred, perhaps by its insertion into a transposable element. A genomic Southern blot of total HindIII digested DNA from *P.biaurelia*, *P.sonneborni*, *P.multimicronucleatum* and *P.caudatum* was probed with a 2.1 kb EcoRI–XhoI fragment downstream of the A gene that contains the repeat DNA (Figure 4A). After a 1 hour exposure multiple bands as well as a

background smear could be detected from *P.tetraurelia*, *P.biaurelia* and *P.sonneborni* (members of the *P.aurelia* complex). The background smear of hybridization is not surprising since many copies of this sequence are expected on macronuclear telomeres, which are heterogeneous in size. A 24 hour exposure detected at least 22 distinguishable bands in the *P.caudatum* lanes and a similar number in the *P.multimicronucleatum* DNA (Figure 4B). Hybridization of the same probe to total DNA from d12, a mutant that lacks the A surface protein gene and downstream sequences (24), shows that the multiple copies of this sequence cannot be the result of an unusual DNA processing event that overamplifies the copy downstream of the A gene. It should be noted that the deleted copy of repeat DNA downstream of the A gene could not be identified on a Southern blot, presumably due to other bands that obscure that area of the autoradiogram. This emphasizes that the number of distinguishable bands on the Southern is a conservative estimate of the number of locations this sequence is found in the genome. Since the 2.1 kb fragment used as a probe for this Southern

P126	G	W	K	L	I	L	V	K	G	Q	K	V	K	L	D	S	Q	T	V	Y	N	F	S	N	G	T	L	A	S	G	N	L	T	I	50		
Betatrans	S	G	T	T	C	A	I	E	G	Q	T	T	T	F	T	T	G	D	V	M	S	L	A	D	T	R	F	V	A	C	A	S	A	K	241		
Betapro	S	G	T	T	C	A	I	E	G	Q	K	T	V	F	V	T	G	D	C	M	A	V	S	D	F	N	F	I	A	C	A	S	A	K	212		
Prp4	A	H	M	T	W	R	A	S	H	E	L	L	L	Q	E	D	K	G	V	F	S	F	Q	C	D	G	S	V	C	G	M	S	L	S	M	332	
Tup1	G	A	E	R	L	I	R	E	I	E	N	R	K	I	V	M	I	Q	D	I	Y	D	Y	F	S	G	D	K	L	V	S	G	R	T	V	R	510
cons	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	9	

P126	L	V	E	M	C	R	Q	F	T	E	S	D	I	N	A	I	C	F	F	N	N	A	F	T	D	88
Betatrans	L	V	E	M	C	R	Q	F	T	E	S	D	I	N	A	I	C	F	F	N	N	A	F	T	D	279
Betapro	L	V	E	M	C	R	Q	F	T	E	S	D	I	N	A	I	C	F	F	N	N	A	F	T	D	250
Prp4	L	V	E	M	C	R	Q	F	T	E	S	D	I	N	A	I	C	F	F	N	N	A	F	T	D	370
Tup1	L	V	E	M	C	R	Q	F	T	E	S	D	I	N	A	I	C	F	F	N	N	A	F	T	D	548
cons	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	17

Figure 3. Alignment of the predicted amino acid sequence of the P126 repeats with members of the β transducin repeat family. The predicted protein sequence from the *Paramecium* 126 bp repeats (P126) is aligned with the amino acid sequence of β transducin (Betatrans), the β subunit of a human G protein (Betapro), Prp4, and Tup1. The numbering corresponds to the amino acid sequence in references, 12, 26, 11, and 13, for bovine β transducin, the subunit of a human G protein, Prp4, and Tup1, respectively. The consensus line (cons) indicates residues that are conserved in all examples of the repeat not just those shown in this figure.

contains some sequence that is not P126 repeat sequence, an additional Southern containing *P.tetraurelia* wild type and d12 DNA was probed with an oligonucleotide corresponding to a conserved portion of the P126 repeat. The resulting autoradiogram showed the same type of smear and banding pattern as the 2.1 kb fragment (data not shown).

The presence of multiple bands on a Southern blot of restricted DNA does not necessarily indicate that the repeats are present on many different macronuclear DNA molecules. Alternatively, many copies could be located on a few macronuclear DNA fragments. To distinguish between these possibilities, total undigested DNA from *P.tetraurelia* was electrophoresed on a pulsed field gel, blotted to nitrocellulose, and probed first with a fragment of A surface protein gene then with the 2.1 kb fragment containing repeats downstream of the A gene (Figure 5). As expected, the A gene hybridized to a single band of approximately 320 kb in size (Figure 5A and 3). The repetitive probe hybridized to at least 15 distinguishable bands as well as a smear of hybridization throughout the gel (Figure 5B). Due to the large number of macronuclear DNA molecules (estimated at about 300) and variations in macronuclear DNA processing (20), ethidium stained pulsed field gels reveal only a few bands within a background smear. In general, the hybridization signal of the repeat DNA corresponded to the ethidium bromide stained pattern (Figure 5C).

The repeated DNA hybridizes to polyadenylated transcripts

To determine if any of the repeated DNA motif is transcribed, a Northern blot containing total RNA from a *Paramecium* culture in log phase was probed with the 2.1 kb repeat fragment (Figure 6). Autoradiography revealed a main band centered at a size of 3,500 nucleotides and a heterogeneous signal both above and below this region. Polyadenylated RNA isolated from an oligo(dT) cellulose column gave a stronger signal and hybridization to the unbound RNA fraction was almost undetectable. These results demonstrate that at least some of the repeat sequences are transcribed into poly(A⁺) RNA. Hybridization of a Northern blot containing RNA from the d12 mutant cell line resulted in a signal indistinguishable from wild type cells (data not shown). Since d12 does not contain the A gene linked repeats, they either contribute to only a small fraction of the signal or they are not transcribed at all.

Figure 4. Hybridization of several species of *Paramecium* with a probe containing the P126 repeats from *P.tetraurelia*. Genomic DNA (5 μ g) was digested with HindIII restriction enzyme, electrophoresed on a 0.8% agarose gel, blotted to nitrocellulose and probed with nick-translated 2.1 kb EcoRI-XbaI fragment (see Figure 1). The fragment was labeled to a specific activity of approximately 5×10^7 cpm/ μ g. Panel A shows a one hour exposure with an intensifying screen, panel B shows a 24 hour exposure of lanes 5 and 6. Lanes: 1, *Paramecium tetraurelia*; 2, d12, a mutant strain of *P.tetraurelia*; 3, *P.biaurelia*; 4, *P.sonneborni*; 5, *P.multimicronucleatum*; 6, *P.caudatum*. Size markers are indicated in kilobases.

DISCUSSION

Genomic characterization of repeats

Genomic Southern blots suggest that the P126 repeat is present in at least 15–20 locations in the macronuclear genome. This is most likely a conservative estimate. Since the sequence is present as tandem repeats, determining the copy number of the repeat alone cannot be used to determine the total number of genomic locations. A rough estimate of the total locations could

Figure 5. Pulsed field gel analysis of P126 repetitive sequences. Duplicate samples of total *Paramecium tetraurelia* DNA (250:1, macronuclear:micronuclear) were electrophoresed on a CHEF (clamped homogeneous electric field) apparatus for 48 hours at 150 volts with a pulse time of 50 seconds, then blotted to nitrocellulose. **A.** Hybridization of the filter with an A gene specific probe (3 day exposure). **B.** The same filter in **A** was rehybridized with the 2.1 kb EcoRI-XhoI fragment. (4 hour exposure.) **C.** Photograph of the ethidium stained gel. Lambda ladder size markers are indicated in kilobases.

Figure 6. Northern blot hybridization analysis of the P126 repeat. RNA from log growth cells was electrophoresed on a 1% formaldehyde gel and blotted to nitrocellulose, then probed with nick-translated 2.1 kb EcoRI-XhoI fragment. Lane 1, three micrograms of poly(A) RNA; lane 2, three micrograms of total RNA, lane 3, three micrograms of RNA that was not bound to the oligo(dt) cellulose column. RNA size markers are shown in 1×10^{-3} nucleotides.

be made if the total number of repeats is known as well as the average number of tandem repeats per site, and these data are currently being determined.

Interestingly, all the cloned examples of this repeat are near the ends of macronuclear DNA molecules. Since all but one of the clones were initially selected because they contained

telomeres, this cannot be used to argue that the sequence is preferentially located near macronuclear telomeres. Bal31 digestion of genomic DNA only identified two preferentially sensitive bands (data not shown), but the organization of *Paramecium* macronuclear DNA could make it difficult to easily demonstrate a preferential telomeric location. Discrete yet variable sites of telomere addition have been demonstrated in *Paramecium*. For example, the repeat located downstream of the A gene can be 2.5, 7.5, or 24.5 kb from the macronuclear telomere depending on how the individual molecule has been processed (20). This heterogeneity could place some repeats too far from an end to demonstrate Bal31 sensitivity and at the same time obscure other Bal31 sensitive molecules. In fact, since only random telomere clones with inserts larger than 4 kb contained repeat DNA it is unlikely that the repeat is directly adjacent to the telomeric sequence. A more striking example of variable telomere location has recently been described by Caron in *P. primaurelia* (25). Analysis of a specific region of the macronuclear genome showed that variable DNA rearrangements create overlapping macronuclear chromosomes in which a region can be telomeric on one macronuclear DNA molecule and over 100 kb away from a telomere on an alternatively processed molecule. Obviously, this type of variable DNA rearrangement would complicate experiments to determine Bal31 sensitivity.

Sequence, expression and conservation of the repeat

The sequence of the tandem repeats in *Paramecium* DNA has clear identity to the β -transducin subunit. The most compelling evidence for this is the alignment of the amino acids DX₄ΩWD, where Ω represents either L, I or V, at precise 42 residue spacing with the β subunits of G proteins as well as the *PRP4*, and *TUP1* gene products. Not all examples of this motif in other organisms are located at precise 42 amino acid intervals, 43 amino acids separate some of the examples of this motif in human G-protein β subunits (26) and 63 amino acids separate DX₄WD repeats in the developmentally regulated AAC rich RNA from *Dictyostelium discoideum* (27). Degenerate versions of the sequence can also be identified, for example in human G-protein β subunits phenylalanine is substituted for the conserved tryptophan at amino acid 253 and the sequence DX₄IYN is located roughly halfway between the first two DX₄ΩWD motifs, a distance of 88 amino acids (26). The full extent of sequence variation is difficult to determine since it is not yet clear what constitutes a functional repeat.

The identification of polyadenylated transcripts that have similarity to the P126 repeat suggests that at least some of these repeat regions are transcribed and potentially translated into protein products. The reproducible size heterogeneity of transcripts observed on Northern blots is consistent with the transcription of more than one repeat region, but confirmation of this will require analysis of multiple cDNA clones. Previously isolated A- mutant cell lines, such as d48 and d12, also lack the macronuclear copies of the A gene linked P126 repeats. Since neither of these mutations have any obvious effect on viability or replication of the remainder of the macronuclear chromosome, at least this set of repeats is not essential for viability. It is not possible from the current data to evaluate if the A gene linked repeats are transcribed, but if so, it is clear that they are not the only transcribed repeats since Northern blots containing RNA from d12 cells give a signal identical to wild type RNA (data not shown).

Since the biochemical function of β -transducin repeats is not known, its presence does not immediately suggest a function of the predicted polypeptide. In fact, it appears that the members of the β -transducin family have many diverse functions such as signal transduction, RNA processing and transcriptional control (discussed in 11). Recently it has been proposed that proteins containing the β -transducin motif may interact with members of the TPR family of proteins (16). The TPR motif is a 34 amino acid repeat postulated to form a helix-turn domain. Direct evidence for the association of Cyc8 (Ssn6) and Tup1, containing TPR and β -transducin motifs respectively, has been provided by Williams and Trumbly (18).

Genomic Southern blots indicate that the P126 repeat is present in multiple copies in diverse species of *Paramecium*. The conservation of this repeat does not prove that multiple copies are functionally important, but current evidence suggests that non-essential DNA sequences near macronuclear telomeres can be eliminated during macronuclear development. The d48 mutation in *Paramecium tetraurelia* creates a macronuclear deletion of the A surface protein gene and downstream sequences, yet contains a complete copy of these sequences in the micronucleus (24, 28). The defect is caused by a macronuclear mutation that controls the processing of the A gene region during the formation of the next macronucleus (28, 29). A similar mutation affecting the G type surface protein gene was induced in *P. primaurelia* by high copy number macronuclear transformation with the cloned G gene followed by formation of a new macronucleus (30). Thus, mechanisms exist in *Paramecium* to eliminate germline sequences from the macronucleus, yet these repeats have been maintained in many telomeric positions. Other ciliates such as *Oxytricha* have evolved mechanisms to eliminate 90% of the micronuclear sequence complexity during macronuclear development, and few repetitive micronuclear sequences are incorporated into the macronucleus (31). The extent of sequence diminution is not known in *Paramecium*, but the recent isolation and analysis of micronuclear DNA has shown that specific elimination events occur within and adjacent to the A gene in *P. tetraurelia* (32).

If the multiple P126 repeats in the *Paramecium* genome are functionally important, then they may have a role unique to the physiology of *Paramecium* or perhaps ciliates in general. Obviously, the formation and maintenance of the macronucleus is one possibility, but others include secretion of trichocysts, and components of the cell cortex. We are now producing antibodies directed against the predicted polypeptide encoded by the repeat to aid in determining its function in the cell.

ACKNOWLEDGMENTS

We thank Dr. Elizabeth Blackburn in whose laboratory the initial identification of the repeat was made and Susan Fong who participated in some of the early experiments. This work was supported by National Institutes of Health grant GM43357 and A127713, as well as a Junior Faculty Award from the American Cancer Society to J.F. Additional support was provided by an Indiana Elks Grant for Cancer Research. This is journal paper number 13476 from the Purdue Ag. Exp. Station.

REFERENCES

- Blackburn, E.H. and Karrer, K.M. (1986) *Ann. Rev. Genet.* **20**, 501–521.
- Preer, J.R. Jr and Preer, L.B. (1979) *J. Protozool.* **26**, 14–18.
- Phan, H.L., Forney, J. and Blackburn, E.H. (1989) *J. Protozool.* **36**, 402–408.
- Soldo, A.T. and Godoy, G.A. (1972) *J. Protozool.* **19**, 673–678.
- Dippell, R.V. (1954) *Caryologia* **6** (suppl.), 1109–1111.
- Findly, R.C. and Gall, J.G. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3312–3316.
- Caron, F. and Meyer, E. (1989) *Ann. Rev. Microbiol.* **43**, 23–42.
- Cherry, J.M. and Blackburn, E.H. (1985) *Cell* **43**, 747–758.
- Herrick, G., Cartinhour, S., Dawson, D., Ang, D., Sheets, R., Lee, A. and Williams, K. (1985) *Cell* **43**, 759–768.
- Simon, M.I., Strathmann, M.P. and Gautam, N. (1991) *Science* **252**, 802–808.
- Dalrymple, M.A., Petersen-Bjorn, S., Friesen, J.D. and Beggs, J.D. (1989) *Cell* **58**, 811–812.
- Fong, H.K., Hurley, J.B., Hopkins, R.S., Miale-Lye, R., Johnson, M.S., Doolittle, R.F. and Simon, M.I. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2162–2166.
- Williams, F.E. and Trumbly, R.J. (1990) *Mol. Cell. Biol.* **10**, 6500–6511.
- Sethi, N., Monteagudo, M.C., Koshland, D., Hogan, E. and Burke, D.J. (1991) *Mol. Cell. Biol.* **11**, 5592–5602.
- Hartley, D.A., Preiss, A. and Artavanis-Tsakonas, S. (1988) *Cell* **55**, 785–795.
- Goebel, M. and Yanagida, M. (1991) *TIBS* **16**, 173–177.
- Williams, F.E., Varanasi, U. and Trumbly, R.J. (1991) *Mol. Cell. Biol.* **11**, 3307–3316.
- Keleher, C.A., Redd, M.J., Schultz, J., Carlson, M. and Johnson, A.D. (1992) *Cell* **168**, 709–719.
- Preer, J.R. Jr, Preer, L.B. and Rudman, B.M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6776–6778.
- Forney, J.D. and Blackburn, E.H. (1988) *Mol. Cell. Biol.* **8**, 251–258.
- Henikoff, S. (1987) *Methods Enzymol.* **155**, 156–165.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Devereaux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
- Rudman, B., Preer, L.B., Polisky, B. and Preer, J.R. Jr (1991) *Genetics* **129**, 47–56.
- Caron, F. (1992) *J. Mol. Biol.* **225**, 661–678.
- Levine, M.A., Smallwood, P.M., Moen, P.T., Helman, L.J. and Ahn, T.G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2329–2333.
- Shaw, D.R., Richter, H., Giorda, R., Ohmachi, T. and Ennis, H. (1989) *Mol. Gen. Genet.* **218**, 453–459.
- Kobayashi, S. and Koizumi, S. (1990) *J. Protozool.* **37**, 489–492.
- Harumoto, T. (1986) *Mol. Cell. Biol.* **6**, 3498–3501.
- Meyer, E. (1992) *Genes Dev.* **6**, 211–222.
- Lauth, M.R., Spear, B.B., Heumann, J. and Prescott, D.M. (1976) *Cell* **7**, 67–74.
- Preer, L.B., Hamilton, G. and Preer, J.R. Jr (1992) *J. Protozool.* in press.