

Structure of the macronuclear polyubiquitin gene in *Euplotes**

Loren J. Hauser, Arthur E. Roberson, and Donald E. Olins

U.T.-Oak Ridge Graduate School of Biomedical Sciences, Biology Division, Oak Ridge National Laboratory,
 P.O. Box 2009, Oak Ridge, TN 37831-8077, USA

Received October 29, 1990

Accepted February 7, 1991 by D. Bazett-Jones

Abstract. The hypotrichous ciliate, *Euplotes eurystomus*, contains both a transcriptionally inactive micronucleus (MIC) and a transcriptionally active macronucleus (MAC) in the same cell. MAC DNA is small (0.5–20 kb), linear and highly amplified. Each DNA fragment consists of two telomeres, a single coding region, and the necessary control elements to regulate gene transcription and replication. The polyubiquitin gene consists of 898 bp, plus 28 bp of double-stranded and 14 bases of single-stranded DNA of the telomeric repeat G_4T_4 at each end. The coding region exists as three copies of the ubiquitin gene (690 bp) fused in a head-to-tail arrangement as in other organisms. The stop codon is TAA, as in other *Euplotes* genes, and is not the rare glutamine codon used in most other ciliates. The 3' non-translated region contains two presumptive poly(A) addition sites; the 5' nontranslated region possesses two putative TATA boxes, several imperfect direct and inverted repeats, and a possible origin of replication. Nucleosome positioning studies reveal four tightly packed nucleosomes and a non-nucleosomal area containing the probable 5' control region as well as part of the coding region. The 5' area does not contain any DNase I hypersensitive sites. Although the telomeres are protected from exonuclease digestion, they are not as well protected as *Oxytricha* telomeres against endonucleases and cleavage by methidium propyl Fe^{2+} EDTA.

* Research sponsored by the Office of Health and Environmental Research, United States Department of Energy, under Contract DE-AC05-84OR21400 with the Martin Marietta Energy Systems, by NSF Research Grant DCB870057 to D.E.O., by NCI Training Grant T32CA09336, and by NIH Genetics Training Grant GM07438 to U.T.-Oak Ridge Graduate School of Biomedical Sciences

Abbreviations: MIC, micronucleus; MAC, macronucleus; MPE- Fe^{2+} , methidium propyl EDTA- Fe^{2+} ; DNase I, deoxyribonuclease I; MNase, micrococcal nuclease; RNase A, ribonuclease A; SDS, sodium dodecyl sulfate; $20 \times$ SSC, 3 M NaCl, 300 mM sodium citrate

Offprint requests to: D.E. Olins

Introduction

Ubiquitin is a small, highly conserved, multi-functional protein that has been found in all eukaryotes examined. This 76 amino acid protein has few variable positions and calculations suggest that it is the most highly conserved protein known (Sharp and Li 1987). Conjugation of ubiquitin to other proteins can target them into an ATP-dependent proteolytic pathway, or modulate the function of the affected protein (for reviews see Hershko 1988; Rechsteiner 1987; Finley and Varshavsky 1985). Ubiquitin has also been found conjugated to various cell surface receptors (Siegelman et al. 1986; Yarden et al. 1986; Leung et al. 1987). Ubiquitinated histones are believed to disrupt nucleosome structure and appear to be preferentially associated with active genes (Levinger and Varshavsky 1982; Ridsdale and Davie 1987). Two of the known E2 proteins in yeast (RAD 6 and CDC 34), which are involved in the conjugation of activated ubiquitin to other proteins, may be involved in the regulation of DNA repair and the cell cycle (Jentsch et al. 1987; Goebel et al. 1988). All thoroughly studied organisms contain at least two distinct ubiquitin genes: one polyubiquitin gene with 2–18 copies of ubiquitin in a tandem array (Sharp and Li 1987; Lee et al. 1988); and at least one other gene with the reading frame of a ribosomal protein (Finley et al. 1989; Redman and Rechsteiner 1989) fused directly to the 3' end of a ubiquitin coding region (Sharp and Li 1987).

Euplotes eurystomus, like other hypotrichous ciliated protozoa, contains both a macronucleus (MAC), and a micronucleus (MIC). The MIC contains normal chromosomes, is transcriptionally inactive, and is the equivalent of a germ cell in multicellular organisms. The MAC is transcriptionally active and contains small gene-sized pieces of DNA that range in size from 500 bp to 20 kb (Kraut et al. 1986). The MAC is derived from MICs through a complex developmental process following sexual conjugation (Kraut et al. 1986).

A number of genes from various hypotrichous ciliates have been cloned and sequenced. These data, together

with that presented in this paper, support the hypothesis that each piece of MAC DNA exists as an autonomous unit, consisting of one coding region, two telomeres, and all essential regulatory sequences (Kraut et al. 1986). Each of the MAC DNA pieces is packaged into mini-chromatin molecules containing a number of nucleosomes and two telomeres (Kraut et al. 1986). Soluble chromatin can be readily isolated from MACs without the use of nucleases, and specific-sized chromatin molecules can be enriched using standard techniques (Cadilla et al. 1986). This isolated soluble chromatin exhibits many of the same properties as chromatin from other eukaryotes (Butler et al. 1984). The chromatin structure of individual MAC genes can be readily probed using indirect end-labeling (Wu 1980; Nedospasov and Georgiev 1980). Previous studies in *Oxytricha* have suggested that nucleosomes are positioned inward from the telomeres (Gottschling and Cech 1984); however, data on the *Euplotes* 5S gene (Roberson et al. 1989) and *Euplotes* polyubiquitin gene (this paper) suggest otherwise. Both *Euplotes* studies reveal that the positions of nucleosomes are characteristic for each MAC gene.

Materials and methods

***Euplotes* cell culture, isolation of MACs, MAC DNA, and chromatin digestions.** The handling of the *Euplotes* cell stocks and isolation of MACs have been described elsewhere (Roberson et al. 1989). DNA was prepared from nuclei or permeabilized cells by proteinase K digestion (Boehringer Mannheim, 0.5 mg/ml), phenol and phenol:chloroform extraction, RNase A digestion (Sigma, 100 µg/ml), phenol and phenol:chloroform extraction and ethanol precipitation. Chromatin digestions for indirect end-labeling were performed on purified nuclei or permeabilized cells. Digestion parameters were as described previously (Roberson et al. 1989), except that methidium propyl EDTA Fe^{2+} (MPE· Fe^{2+}) was used at 20 µM for the chromatin digestions. PHEM buffer permeabilized cells were prepared as previously described (Olins et al. 1989), but were rinsed once with buffer prior to digestion with either DNase I, micrococcal nuclease (MNase), MPE· Fe^{2+} , or Bal31. The buffer for Bal31 digestions was essentially the same as that used for DNase I and MNase, except it contained 2 mM CaCl_2 and MgCl_2 , and 0.2 mM EDTA.

Enzymes and radioisotopes. Restriction endonucleases, purchased from Bethesda Research Laboratories or New England Biolabs, were used according to the manufacturer's instructions with a two- to tenfold excess of enzyme units per microgram DNA. DNase I and MNase were purchased from Worthington, and Bal31 from New England Biolabs. γ - ^{32}P ATP (>1,500 Ci/mmol), α - ^{32}P dCTP (3,000 Ci/mmol) and α - ^{32}P ATP (3,000 Ci/mmol) were purchased from New England Nuclear.

Cloning and sequencing. The *Euplotes* library was prepared and screened as previously described (Roberson et al. 1989). The yeast ubiquitin clone, pUB2 (Özkaynak et al. 1987), used to screen the library was kindly provided by Dr. Alexander Varshavsky. A BclI-BstXI fragment was excised from the plasmid and labeled by nick translation for use as the probe. Large scale plasmid preparations were performed using the alkaline lysis procedure (Birnboim and Doly 1979). Cloned inserts were electroeluted from agarose gels using DE-81 paper (Yang et al. 1982); the DNA was eluted from the paper using 1.5 M NaCl. Purified inserts of pU5.3 were digested with EcoRI, Sau3A or HpaII and shotgun cloned into M13 mp18 or mp19. M13 clones were sequenced using a Sequenase kit from

U.S. Biochemicals and α - ^{35}S dATP (>1,000 Ci/mmol) from Amersham.

Blotting and hybridizations. Colony transfers and Southern DNA blotting were performed using standard techniques (Maniatis et al. 1982). All blots were hybridized in $6\times$ SSC, 50 µg/ml denatured salmon sperm DNA, 0.5% SDS, and 0.3% non-fat dry milk at 50° C. ($1\times$ SSC is 0.15 M NaCl, 0.015 M sodium citrate.) Blots were rinsed at: 50° C, 0.1% SDS and $2\times$ SSC (using the yeast probe); 50° C, 0.1% SDS, $0.1\times$ SSC (using the synthetic oligomers); and 60° C, 0.1% SDS, $0.1\times$ SSC (using pU5.3 or one of its subclones).

Copy number determination. Various concentrations of MAC DNA (10 µg, 1 µg, 100 ng, and 10 ng) and purified pU5.3 insert (10 ng, 1 ng, 100 pg, and 10 pg) were electrophoresed on a 1% agarose gel, transferred to nylon and hybridized with a nick translated, purified pU5.3 insert. Autoradiograms were obtained from four different exposure times. Bands in the linear range were scanned using a Zeineh Soft Laser scanning densitometer (Biomedical Instruments). The area under the curves were used to estimate the number of copies of the polyubiquitin gene per cell (assuming 400 pg DNA per MAC; Ammermann and Muenz 1982).

Indirect end-labeling. Indirect end-labeling (Wu 1980; Nedospasov and Georgiev 1980) of Southern blots was performed using end-labeled synthetic oligomers as probes (see Fig. 2). The oligomers were synthesized on a Microsyn 1450 DNA synthesizer (Systec).

Nucleosome positioning. To determine the position of individual nucleosomes, the average size of each of the bands seen on the Southern blots was calculated. Numbers were generated using probes from both ends of the molecules (see Fig. 2). DNA was

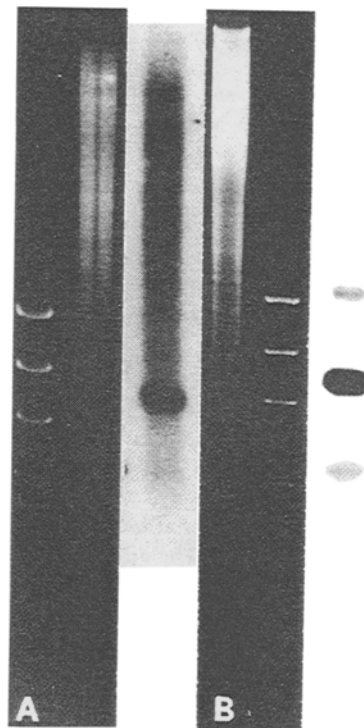


Fig. 1A, B. Identification of the *Euplotes* ubiquitin genes. Southern blots of total macronuclear (MAC) DNA were probed with either the insert from the yeast clone pUB2 (A) or the *Euplotes* polyubiquitin gene pU5.3 (B). In each panel lanes 1 and 2 show the ethidium bromide (EtBr) stained gel with $\phi\text{X174}/\text{HaeIII}$ fragments as markers; lane 3 is the autoradiograph

```

10      20      30      40      50      60      70      80
CCCCAAAACCCCAAAACCCCAAAACCCCGAGACTAGTTTTCTTATGAAGTATAGAGAAAATAT
GGTTTTGGGGTTTTGGGGTTTTGGGGTTTTGGGG*****
90      100     110     120     130     140     150     160
CAGATTCATAAATTGTAGATTATACCGGAAATTTAAAGTTAGACCGGAAATTAGATTTATAAGGAAAATTTTAA
170     180     190     200     210     220     230     240
TCAATTAATAACAAAATAAATATGCAAACTTCGTAAGACCCCTACAGGCAAGACAATTACCTTAGACGTTGAGCAAT
      M Q I F V K T L T G K T I T L D V E Q
250     260     270     280     290     300     310     320
CAGACACCATTGACAACGTTAAGACTAAGATCAAGACAAGGAAGGAATCCCCAGATCAACAAAGACTTATCTTCGCA
S D T I D N V K T K I Q D K E G I P P D Q Q R L I F A
330     340     350     360     370     380     390     400
GGAAAGCAATTAGAAGCGGAAGAACCCCTTGCTGATTATAACATTGAGAAGGAGTCAACTCTCCACTTGGTCTCAGACT
G K Q L E D G R T L A D Y N I Q K E S T L H L V L R L
410     420     430     440     450     460     470     480
TAGAGGAGGTATGCAAACTTCGTCAGACTTTGACCGGAAAGACTATTACCTTGACGTCGAACAATCTGATACTATTG
R G G M Q I F V K T L T G K T I T L D V E Q S D T I
490     500     510     520     530     540     550     560
ATAACGTCAAGACCAAGATTCAAGATAAGGAAGGAATCCACCAGATCAACAAAGACTCATCTTTGCTGGAAAGCAATTG
D N V K T K I Q D K E G I P P D Q Q R L I F A G K Q L
570     580     590     600     610     620     630     640
GAAGACGGAAGAAGCTCTGCTGACTACAACATCCAAAAGGAGTCAACTCTTCACCTCGTCTCAGACTTAGAGGAGGTAT
E D G R T L A D Y N I Q K E S T L H L V L R L R G G M
650     660     670     680     690     700     710     720
CAGCAAACTTCGTCAGACCTTGACTGGAAGACTATTACTCTTGACGTCGAGCAAAGCGACTATTGACAAATGTARAGA
Q I F V K T L T G K T I T L D V E Q S D T I D N V K
730     740     750     760     770     780     790     800
CTAAAATCAAGATAAGGAAGGAATTCACCAGATCAGCAAAGACTCATCTTCGCCGGAAGCAACTCGAGGACGGAGA
T K I Q D L E G I P P D Q Q R L I F A G K Q L E D G R
810     820     830     840     850     860     870     880
ACCCCTTGCTGACTACAACATCCAGAAAGAGTCCACACTTCACTTGGTCTTAGACTCAGAGGAGGATTCTAATCTAACGT
T L A D Y N I Q K E S T L H L V L R L R G G F -
890     900     910     920     930     940     950     960
AATTCATAATAAATTTCAAAAATAAATCTCTTCAATTTTCTATTCTCAAAACAGCTTGGGGTTTTGGGGTTTTGGGG
      *****
970     980     990     1000
TTTTGGGGTTTTGGGGTTTTGGG
AAAACCCC

```

Fig. 2. Structure of the *Euplotes* polyubiquitin gene molecule. The telomere sequences have been left in the double-stranded form to show the single-stranded overhangs. Putative TATA boxes and poly(A) addition sites are underlined. A conserved pentanucleotide that may be part of the telomere structure is underlined with asterisks. The 30-mers used as probes in the indirect end-labeling experiments are heavily underlined.

derived from both starved and freshly fed cells and digested with three different cutting agents (MPE·Fe²⁺, DNase I, and MNase).

Results

Identification and cloning of the *Euplotes* ubiquitin genes

In order to determine the number and size(s) of the *Euplotes* ubiquitin genes and optimal conditions for screening the library, three strips from a Southern blot of MAC DNA were hybridized, under low stringency, with a nick translated, gel-purified insert of the pUB2 yeast gene. Each strip was washed at 50° C with either 0.1 ×, 1 × or 5 × SSC. The 1 × SSC wash gave the best signal with the least background and is shown in Fig. 1A. A single predominant band of approximately 950 bp in size was observed. The *Euplotes* MAC library was screened using these conditions. Plasmids from 12 second round positive colonies were screened a third time on a Southern blot after digesting with PstI to excise the inserts. All 12 yielded identical size inserts of 950 bp. Six of these were further tested for restriction fragment length polymorphisms and all yielded identical pat-

Euplotes

Trypanosoma

Saccharomyces

plants

vertebrates & insects

Caenorhabditis

```

10      20
MQIFVKLTGKTTITLVEQSDTIDNVK
      A E S E
      E S
      E S
      E P E
      E A E
30      40      50      60      70
TKIQDKEGIPPDQQLIFAGKQLEDGRTLADYNIQKESTLHLVLRGG
A
S
A
A
A

```

Fig. 3. Comparative ubiquitin protein sequences. The animal sequences include data from human (Wiborg et al. 1985; Lund et al. 1985), chicken (Bond and Schlesinger 1985), bovine (Schlesinger and Goldstein 1975), pig (Einspanier et al. 1987), frog (Dworkin-Rastil et al. 1984), trout (Watson et al. 1978), fruit fly (Izquierdo et al. 1984), and medfly (Gavilanes et al. 1982). The plant sequence data include oat (Vierstra et al. 1986) and barley (Gausling and Barkardottir 1986). The other sequences are referenced as follows: *Trypanosoma cruzi* (Kirchhoff et al. 1988; Swindle et al. 1988), and yeast (Özkaynak et al. 1987). The extra amino acid on the end of the last ubiquitin repeat is not included here because it is not known for many of the species.

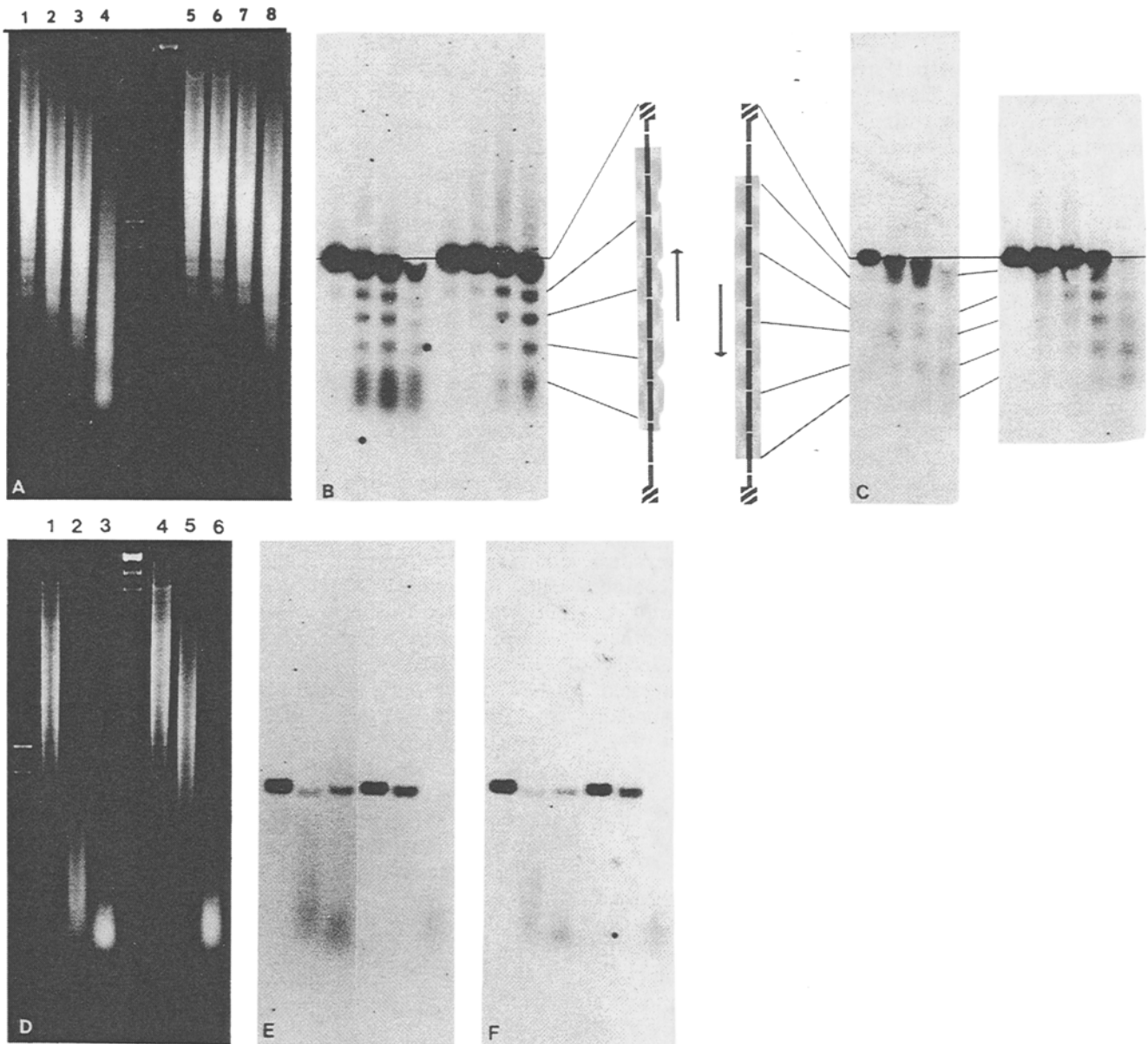


Fig. 4A-F. Nucleosome positioning gels. All gels were 1% agarose. **A, B** and **C** Results of chromatin digestions. **D, E** and **F** Digestions of naked DNA. **A, D** EtBr-stained gels. **B, E** probed with the 5' end 30-mer (see Fig. 2). **C, F** Probed using the 3' end 30-mer (see Fig. 2). **A, B** and **C, lanes 1-4**, nuclei were digested with DNase I using 0, 1, 3, and 10 units/ml for 3 min at room temperature. **A, B, lanes 5-8**, nuclei were digested with 5 μ M methidium propyl EDTA \cdot Fe²⁺ (MPE \cdot Fe²⁺) for 0, 2, 5, or 10 min at room temperature; **C, lanes 5-9**, permeabilized cells were digested with 20 μ M MPE for 0, 1, 2, 5 and 10 min at room temperature. **D, E** and

F, lanes 1-3, DNA was digested with DNase I using 0.1, 0.3 or 1.0 units/ml for 30 s at room temperature; **lanes 4-6**, DNA digested with 5 μ M MPE \cdot Fe²⁺ for 2, 7, or 20 min at room temp. Between panels **B** and **C** are schematic diagrams of the polyubiquitin gene. The striped boxes at each end denote the telomere sequences; the arrow denotes the direction of transcription; the gray-shaded box contains the coding region; and each space in the solid control line denotes 100 bp starting at the 5' end. Lines have been drawn from the schematic to the photographs to relate the approximate position of the bands on the gel to a position in the gene

terns. Restriction fragments from one clone (pU5.3) were subcloned into M13 for sequence analysis.

After identifying pU5.3 as the polyubiquitin gene by sequence analysis (see below), purified insert from pU5.3 was labeled and used to probe a Southern blot of MAC DNA. This probe revealed (Fig. 1B) two other genes not observed using the yeast probe. These two genes are 600 and 1,600 bp in size and are probably the fusion genes (Özkaynak et al. 1987; Wiborg et al. 1985; Lund et al. 1985; Kirchhoff et al. 1988) seen in other eukar-

yotes, whose function has recently been identified (Finley et al. 1989; Redman and Rechsteiner 1989).

Purified insert was also hybridized to a Southern blot containing four known amounts of MAC DNA and four known amounts of purified insert. Autoradiograms from various exposures were scanned with a densitometer and areas under the curves used to estimate the number of copies of the polyubiquitin gene per cell. This gave an estimate of 2×10^5 copies per cell, about 20 times the copy number for an average *Euplotes* gene.

Sequence analysis of the *Euplotes polyubiquitin* gene

EcoRI, Sau3A, and HpaII restriction fragments from the pU5.3 clone were cloned into M13 and sequenced. The sequence of the entire MAC DNA molecule is shown in Fig. 2. The entire molecule contains 954 bp plus 14 bases of the telomeric 3' overhang on each end. There are 56 bp of double-stranded telomere sequence (28 at each end), 690 bp in the coding region, 68 bp 3' of the coding region and 140 bp 5' of the coding region. The telomere structure is identical to that found in *Euplotes aediculatus* (Klobutcher et al. 1982). The consensus pentanucleotide TTGAA (indicated by asterisks in Fig. 2), which is observed 17 bp in from the internal edge of the C₄A₄ repeats in all *Euplotes* genes sequenced so far, is seen in the polyubiquitin gene, although there is a single change in the 3' pentanucleotide. The coding region contains three copies of the ubiquitin gene (228 bp) followed by a single extra codon and the stop codon TAA.

Putative regulatory sequences in this gene are hard to define. There are two consensus poly(A) addition sites (underlined in Fig. 2) within the 68 bp 3' of the coding region. In the 140 bp 5' upstream region there are two weak TATA boxes (underlined in Fig. 2) but no other recognizable transcriptional regulatory sequences. In the 5' noncoding region, there is a pair of imperfect inverted repeats closely associated with short palindromes similar to putative origins of replication (Soede et al. 1977; Stinchcomb et al. 1980).

Conservation of the ubiquitin protein sequence

Evolutionarily, ubiquitin is the most highly conserved protein known (Sharp and Li 1987). Analysis of Fig. 3 reveals only 6 variable positions (14, 15, 19, 24, 28, 57) out of 76 amino acids. Although *Euplotes* ubiquitin contains unique amino acids at positions 16, 19 and 28, only *Euplotes* contains a variation at position 16.

Chromatin structure of the polyubiquitin gene

In an attempt to observe modulations in the chromatin structure of the polyubiquitin gene under different metabolic states, nucleosome positioning experiments, using indirect end-labeling (Wu 1980; Nedospasov and Georgiev 1980), were performed on cells that were either fed the previous day or were starved for at least 4 days. Digestions were performed with PHEM buffer permeabilized cells or with isolated nuclei. Both methods gave identical results.

Blots probed using an end-labeled synthetic 30-mer homologous to the 5' end (see Fig. 2), revealed four closely spaced protected regions and five susceptible regions (Fig. 4B). Blots probed with a 30-mer homologous to the 3' end (see Fig. 2), also showed four protected and five susceptible regions; however, the pattern was slightly different. Based on these data, it is postulated that the four protected regions correspond to four nucleosomes.

The three well-defined susceptible regions in the middle are short internucleosomal spacers. The long susceptible region at the bottom of the gel, when probed from the 5' end (Fig. 4B), is the spacer between the last nucleosome and the smallest piece of DNA that can be detected (~100 bp). This same region is also observed when probed from the 3' end (Fig. 4C); it then appears as a lighter intensity "tail" below the main intense band. The fifth susceptible region is near the ends of the molecule. The intact polyubiquitin gene can be seen in the undigested lane (1 and 5) of each set. In Fig. 4B, C a line has been drawn through the middle of the band representing the intact molecule. As the digestion proceeds, this intense band continuously decreases in size, revealing digestion near the telomere. Also, there is no well-defined telomeric complex as seen in *Oxytricha* (Gottschling and Cech 1984).

In order to position the nucleosomes, the average size of each band on the autoradiographs was determined. The average size of each fragment was then used to estimate the average repeat length for each nucleosome. This repeat length and a value of 100 bp for the telomeric region (determined from other sources, since it was not well delineated in these experiments) was used to place the nucleosomes in the schematic diagram seen in Fig. 5.

Blots of MPE or DNase I digested naked DNA showed no banding pattern (Fig. 4E, F). No differences in chromatin structure could be observed between starved and refed cells.

To determine if the ends of the molecules are protected, permeabilized cells were digested with Bal31 nuclease, the DNA purified, digested with AvaII, electrophoresed, transferred to nylon and probed with the entire gene. AvaII digests the polyubiquitin gene into 130

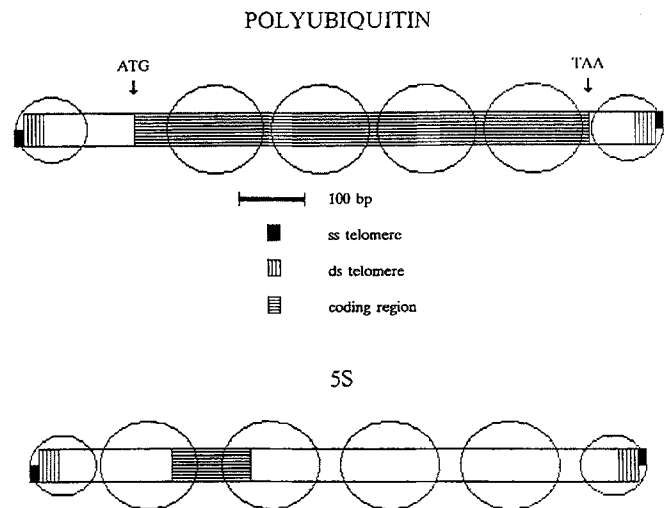


Fig. 5. Chromatin structure of the *Euplotes* polyubiquitin and 5S genes. The diagrams are drawn to scale. Each telomeric complex occupies approximately 100 bp and is represented by the *small circles* drawn at the ends of the genes. Each nucleosome covers 146 bp (one core particle). The nucleosome positions were determined by measuring the average repeat length of each nucleosome and centering the core particle within that repeat. The basic length for the polyubiquitin gene is approximately 160 bp, whereas the basic repeat length for the 5S gene is approximately 180 bp

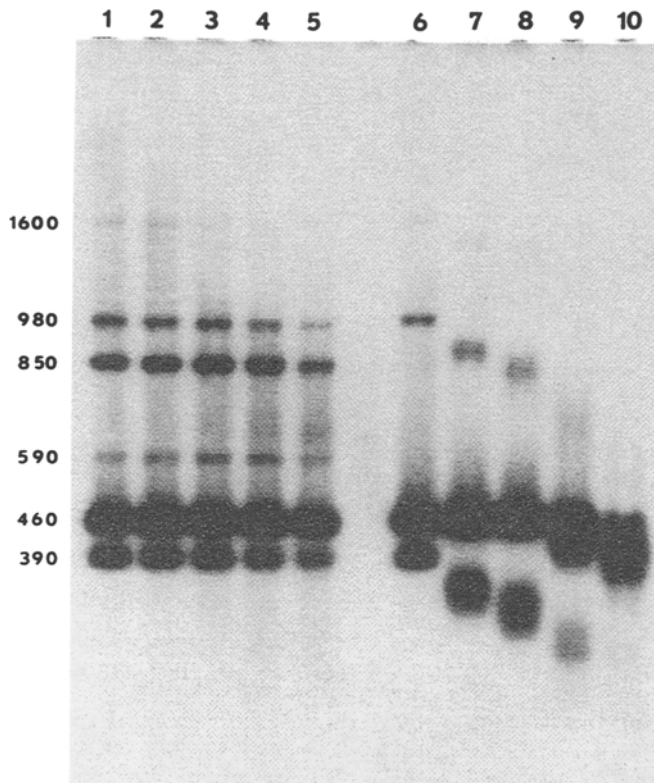


Fig. 6. Bal31 nuclease protection. Permeabilized cells or naked DNA were digested with Bal31 nuclease, the DNA was purified, digested with AvaII, electrophoresed on a 1.5% agarose gel, transferred to nylon, and hybridized to the purified ubiquitin cloned insert. AvaII digests the polyubiquitin gene into 130 and 390 bp end fragments and a 460 bp internal fragment. Lanes 1–5, DNA from permeabilized cells digested with Bal31 at 10 units/ml for 0, 15, 30, 60, and 90 min. Lanes 6–10, naked DNA digested with Bal31 at 1 unit/ml for 0, 15, 30, 60, and 90 min. The numbers at the left represent the size in base pairs of the various bands. The 1,600 bp band is one of the ubiquitin-ribosomal protein fusion genes; the 980 bp band is the intact polyubiquitin gene. The 850 and 590 bp bands are incomplete AvaII digestion products. The 130 bp fragment is not visible

and 390 bp end fragments and a 460 bp internal fragment. The 130 bp fragment is not visible in Fig. 6. If the ends are protected none of the fragments should disappear with time. If the ends are not protected, the end fragments would disappear before the internal fragment. Figure 6 shows that the end fragments are not digested. The ends of the molecules must be protected, although the same degree of protection against endonucleolytic attack as provided by the telomeric proteins of *Oxytricha* is not observed.

Digestion of naked DNA with Bal31 is illustrated in lanes 6 through 10 of Fig. 6. Bal31 worked well under the digestion conditions used; naked DNA was digested with tenfold less enzyme than used with permeabilized cells.

Discussion

The organization of the *Euplotes* ubiquitin genes is similar to that seen in most other eukaryotes: one polyubi-

quitin gene and one or more ubiquitin-ribosomal protein fusion genes. The structure of the *Euplotes* polyubiquitin gene is extremely compact, which is not unusual for genes from the hypotrichous ciliates. Tubulin and actin genes from *Oxytricha*, *Stylonychia*, and *Euplotes* are all under 2 kb in length (Roberson 1988; Greslin et al. 1988), while some members of the 70 kDa HSP gene family and one of the histone H3 genes from *E. eurystomus* have less than 200 bp of noncoding DNA (L. Hauser et al., unpublished data). The lack of extraneous DNA in these compact genes should allow for ready analysis of the putative regulatory sequences; however, few if any of the standard regulatory signals are present (Greslin et al. 1988; Roberson et al. 1989). The polyubiquitin gene lacks the heat shock consensus element (HSE) seen in the yeast gene (Özkaynak et al. 1987). There is a 7 out of 10 bp match to the HSE (Fig. 2, bp 108–121), but all known functional elements have a dimer (i.e., two elements in tandem) and most have multiple dimers. The absence of an obvious HSE is somewhat surprising since ubiquitin is heat shock inducible in yeast (Finley et al. 1987), *Drosophila* (Lee et al. 1988), chicken (Bond and Schlesinger 1985) and *Dictyostelium* (Müller-Taubenberger et al. 1988), although not in *Caenorhabditis elegans* (Graham et al. 1989) nor in trypanosomes (Kirchhoff et al. 1988; Swindle et al. 1988). The lack of regulatory signals may indicate that these genes are continuously transcribed. Further analysis of the expression of these and other genes in the hypotrichs is needed in order to establish mechanisms for regulation of gene expression.

The use of TAA as a stop codon and not as a rare glutamine codon, as in most other ciliates (Helftenbein 1985; Conzelmann and Helftenbien 1987; Caron and Meyer 1985; Preer et al. 1985; Horowitz and Gorovsky 1985; Hirono et al. 1987), correlates with its use in all species of *Euplotes* studied (T. Reddy et al., unpublished data; Harper and Jahn 1989; Miceli et al. 1989). The evolution of this phenomenon is discussed elsewhere (Harper and Jahn 1989). In Table 1 the codon usage of the polyubiquitin gene has been compared with two other *Euplotes crassus* genes (Harper and Jahn 1989). Since the polyubiquitin gene uses only 41 codons it has a bias similar to other ciliates and the *E. crassus* β -tubulin gene (Harper and Jahn 1989). The bias among the codons for each amino acid differs from either of the other two *Euplotes* genes, although the overall pattern is similar. In general, the codon bias for these three *Euplotes* genes is very similar to other hypotrich genes, except for two instances (Martindale 1989). One, mentioned above, is the use of TAA as a stop codon and the lack of its use along with TAG as a glutamine codon. The second is the use of GGA as the primary glycine codon, while other hypotrich genes rely primarily on GGT (Martindale 1989).

The presence of inverted repeats in close proximity to short palindromes is thought to be the structure of origins of replication (Soede et al. 1977; Stinchcomb et al. 1980). There are at least two of these structures in the 5' noncoding region of the gene. Whether these are true origins of replication is purely speculative due

Table 1. *Euplotes* codon usage

Codon	Amino acid	Act	Tub	Ubi	Codon	Amino acid	Act	Tub	Ubi
TTT	F	2	0	1	TAT	Y	6	2	1
TTC	F	12	23	5	TAC	Y	11	14	2
TTA	L	2	1	2	TAA	—	1	1	1
TTG	L	9	17	5	TAG	—	0	0	0
CTT	L	4	0	11	CAT	H	2	4	0
CTC	L	7	15	9	CAC	H	4	6	3
CTA	L	3	0	0	CAA	Q	11	10	18
CTG	L	2	0	0	CAG	Q	3	11	3
ATT	I	8	3	13	AAT	N	5	1	1
ATC	I	15	15	8	AAC	N	5	20	5
ATA	I	0	0	0	AAA	K	9	2	2
ATG	M	12	20	3	AAG	K	20	13	19
GTT	V	11	8	2	GAT	D	16	6	8
GTC	V	11	19	8	GAC	D	10	21	13
GTA	V	4	3	2	GAA	E	16	27	6
GTG	V	1	2	0	GAG	E	16	9	6
TCT	S	2	8	1	TGT	C	4	3	0
TCC	S	4	16	1	TGC	C	6	5	0
TCA	S	8	3	3	TGA	—	0	0	0
TCG	S	0	1	0	TGG	W	4	4	0
CCT	P	1	0	0	CGT	R	0	0	0
CCC	P	0	0	1	CGC	R	1	0	0
CCA	P	16	19	5	CGA	R	0	0	0
CCG	P	0	0	0	CGG	R	1	0	0
ACT	T	10	10	12	AGT	S	2	0	0
ACC	T	7	17	9	AGC	S	3	0	1
ACA	T	4	1	3	AGA	R	13	22	12
ACT	T	0	0	0	AGG	R	2	0	0
GCT	A	10	10	4	GGT	G	9	7	2
GCC	A	3	16	1	GGC	G	1	1	1
GCA	A	9	1	1	GGA	G	20	30	15
GCG	A	1	0	0	GGG	G	1	0	0

Act, *Euplotes crassus* actin (Harper and Jahn 1989); Tub, *E. crassus* tubulin (Harper and Jahn 1989); Ubi, *Euplotes eurystomus* ubiquitin (this paper)

to the lack of information about replication in this organism.

All three tandem copies of the ubiquitin coding region code for an identical protein, even though they vary by 30–35 bp from one another. This degree of variation is comparable to the variability in the repeats in the yeast polyubiquitin locus. Both yeast and *Euplotes* contain substantially greater variation than other species (Sharp and Li 1987). There is no evidence for any gene conversion or unequal crossing over events within the locus; the single base pair substitutions that occur are relatively random over the entire coding region. This implies that the number of repeats within the gene has been fixed for a long period of time. The presence of only three tandem copies in the polyubiquitin gene could decrease the chances of recombinational events, since close proximity of the ends of an array seem to reduce recombination frequency (Sharp and Li 1987).

An analysis of the rate of evolution of the ubiquitin protein sequence has recently been published (Sharp and Li 1987). The data presented in Fig. 3 reveal only six variable positions in a range of species representing many of the major eukaryotic groups. At positions 16 and 24, D and E can be interchanged; while, at positions 14, 28 and 57, A, S and T can be exchanged. These

represent conservative (D↔E) or, at worst, neutral (T↔S↔A) changes. Only at position 19 are the changes truly non-conservative (P↔S↔Q↔A); even then, there is no alteration of charge. Three of these changes are specific to *Euplotes*: D at position 16, Q at 19, and T at 28. All six positions seem to be at the surface of the molecule (Ecker et al. 1987), and variations at positions 19, 24, and 28 do not affect ubiquitin function (Ecker et al. 1987).

The chromatin structure of the polyubiquitin gene has some unique features. This gene contains 954 bp (plus the 14 base 3' telomeric overhang on each end), thus greatly restricting the number and placement of nucleosomes. If one assumes 100 bp for each telomeric complex, the remainder, divided equally (754/4) yields an average nucleosome repeat length of 189 bp, in agreement with a previously determined value for *Euplotes* bulk chromatin of 190 bp (Cadilla et al. 1986). In contrast, the observed average repeat length for the four nucleosomes on the polyubiquitin gene is less than 165 bp and all are spaced inward from the 3' telomere. This leaves about 120–150 bp free of nucleosomes at the 5' end. Nucleosome positions in three other hypotrich genes have been determined: the 5S gene from *E. eurystomus* (Roberson et al. 1989), and the rDNA and C2

genes from *Oxytricha* (Gottschling and Cech 1984). In all three of these genes the nucleosomes are spaced approximately evenly over the available DNA. In addition, bulk chromatin from *Oxytricha* seems to be phased in from the telomeres (Gottschling and Cech 1984).

The most relevant comparison to the *Euplotes* polyubiquitin gene is the *Euplotes* 5S gene, since it is approximately the same size. Figure 5 is a comparison of the two genes and illustrates the dramatic difference in their chromatin structure. It is probable that these two *Euplotes* genes employ different polymerases and their chromatin structure could reflect this difference. The compacting of the nucleosomes toward the 3' end of the polyubiquitin gene could be due to a number of reasons. RNA polymerase II might also be sitting in the non-nucleosomal area in all the genes, but the lack of any protection from nucleases or $MPE \cdot Fe^{2+}$ appears to argue against this. A small trans-acting factor may be bound in this region; however, the resolution of these experiments is not sufficient to detect small proteins bound in this region.

Despite the seeming lack of protection of the non-nucleosomal DNA at the 5' end it does not contain a DNase I hypersensitive site which many trans-acting factors can induce. The rate of cleavage of this DNA appears to be comparable to that of nucleosomal spacer DNA in the rest of the gene. Attempts at *in vivo* footprinting of this region have been unsuccessful. Indirect end-labeling using a probe from the middle of the gene is not feasible because of the repetitive nature of the coding region.

The nucleosome positioning experiments do not provide any reliable evidence for protection of the telomeres. The relatively rapid, small decrease in size of the intact gene during these digestions implies poor protection of the telomeres. Bal31 digestion shows that, at least, the very ends of the molecules are protected. Close examination of Fig. 4B, C does not reveal any distinct separation between the intact molecule and the 150 bp unprotected region at the 5' end, or the first nucleosomal region at the 3' end, arguing against a telosome like that seen in *Oxytricha* (Gottschling and Cech 1984).

The telosomes of *Oxytricha* are formed by two different proteins. *Euplotes* appears to possess only one telomere binding protein (D. Olins and A. Herrmann, unpublished observations) which may be protecting the single-stranded overhang and closely neighboring double-stranded DNA. It may also affect neighboring DNA by generating a DNase I hypersensitive site (Levinger and Varshavsky 1982) (which is also sensitive to $MPE \cdot Fe^{2+}$), effectively removing the extreme ends of the molecules during digestion, generating the rapid decrease in size of the intact molecule. Alternatively, the single *Euplotes* telomeric protein may only protect the single-stranded overhang, and the rapid decrease in size reflects the removal of each single-stranded overhang in a single cleavage event. It is unlikely that the single telomeric protein would protect only the double-stranded portion of the telomere, since there is no decrease in size of the end fragments during Bal31 digestion. Further experimentation is necessary to determine the chromatin

structure of the polyubiquitin gene and its relation to the regulation of expression.

Acknowledgements. The authors would like to thank J. Finch, L. Cacheiro, and A. Herrmann for technical assistance; A. Olins, R. Mural, F. Larimer, D. Gottschling and C. Price for helpful discussions; R. Foote for synthesizing the oligomers; A. Varshavsky for the gift of the yeast pUB200 clone; C. Rains for help in the preparation of the manuscript, and A. Olins, M. Hauser and R. Mural for critical reading of the manuscript. Research was sponsored by the Office of Health and Environmental Research, United States Department of Energy, under Contract DE-AC05-84OR21400 with the Martin Marietta Energy Systems; NSF Research Grant DCB870057 to D.E.O. A. Roberson was supported by the NIH Genetics Training Grant GM07438. L. Hauser was supported by the NCI Training Grant T32CA09336.

References

- Ammermann D, Muenz A (1982) DNA and protein content of different hypotrich ciliates. *Eur J Cell Biol* 27:22-24
- Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7:1513-1523
- Bond U, Schlessinger MJ (1985) Ubiquitin is a heat shock protein in chicken embryo fibroblasts. *Mol Cell Biol* 5:949-956
- Butler AP, Laughlin TJ, Cadilla CL, Henry JM, Olins DE (1984) Physical structure of gene-sized chromatin from the protozoan *Oxytricha*. *Nucleic Acids Res* 12:3201-3217
- Cadilla CL, Roberson AE, Harp J, Olins AL, Olins DE (1986) Subfractionation of soluble macronuclear chromatin and enrichment of specific genes as chromatin from *Euplotes eurystomus*. *Nucleic Acids Res* 14:8501-8512
- Caron F, Meyer E (1985) Does *Paramecium primaurelia* use a different genetic code in its macronucleus? *Nature* 314:185-188
- Conzelmann KK, Helftenbien E (1987) Nucleotide sequence and expression of two β -tubulin genes in *Stylonchia lemnae*. *J Mol Biol* 198:643-653
- Dworkin-Rastl E, Shrutkowski A, Dworkin MB (1984) Multiple ubiquitin mRNAs during *Xenopus laevis* development contain tandem repeats of the 76 amino acid coding sequence. *Cell* 39:321-325
- Ecker DJ, Butt TR, Marsh J, Sternberg EJ, Margolis N, Monia BP, Jonnalagaada S, Ishaqkhan M, Weber PL, Mueller L, Crooke ST (1987) Gene synthesis, expression, structures, and functional activities of site-specific mutants of ubiquitin. *J Biol Chem* 262:14213-14221
- Einspanier R, Sharma HS, Scheit KH (1987) An mRNA encoding poly-ubiquitin in porcine corpus luteum: Identification by cDNA cloning and sequencing. *DNA* 6:395-400
- Finley D, Varshavsky A (1985) The ubiquitin system: functions and mechanisms. *Trends Biochem Sci* 10:343-347
- Finley D, Ozkaynak E, Varshavsky A (1987) The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses. *Cell* 48:1035-1046
- Finley D, Bartel B, Varshavsky A (1989) The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature* 338:394-401
- Gausing K, Barkardottir R (1986) Structure and expression of ubiquitin genes in higher plants. *Eur J Biochem* 158:57-62
- Gavilanes JG, Gonzalez de Buitrago G, Perez-Castells R, Rodriguez R (1982) Isolation, characterization, and amino acid sequence of a ubiquitin-like protein from insect eggs. *J Biol Chem* 257:10267-10270
- Goebel MG, Yochem J, Jentsch S, McGrath JP, Varshavsky A (1988) The yeast cell cycle gene CDC34 encodes a ubiquitin-conjugating enzyme. *Science* 241:1331-1335
- Gottschling DE, Cech TR (1984) Chromatin structure of the molecular ends of *Oxytricha* macronuclear DNA: Phased nucleosomes and a telomeric complex. *Cell* 38:501-510

- Graham RW, Jones D, Candido EPM (1989) UbiA, the major polyubiquitin locus in *Caenorhabditis elegans*, has unusual structural features and is constitutively expressed. *Mol Cell Biol* 9:268-277
- Greslin AF, Loukin SH, Oka Y, Prescott DM (1988) An analysis of the macronuclear actin genes of *Oxytricha*. *DNA* 7:529-536
- Harper DS, Jahn CL (1989) Differential use of termination codons in ciliated protozoa. *Proc Natl Acad Sci USA* 86:3252-3256
- Helftenbein E (1985) Nucleotide sequence of a macronuclear DNA molecule coding for α -tubulin from the ciliate *Stryltonchia lemnae*. Special codon usage: TAA is not a translation termination codon. *Nucleic Acids Res* 13:415-433
- Hershko A (1988) Ubiquitin-mediated protein degradation. *J Biol Chem* 263:15237-15240
- Hirono M, Endoh H, Okada N, Numata O, Watanbe Y (1987) *Tetrahymena* actin. Cloning and sequencing of the *Tetrahymena* actin gene and identification of its gene product. *J Mol Biol* 194:181-192
- Horowitz S, Gorovsky MA (1985) An unusual genetic code in nuclear genes of *Tetrahymena*. *Proc Natl Acad Sci USA* 82:2452-2455
- Izquierdo M, Arribas C, Galceran J, Burke J, Cabrera VM (1984) Characterization of a *Drosophila* repeat mapping at the early-ecdysone puff 63F and present in many eucaryotic genomes. *Biochim Biophys Acta* 783:114-121
- Jentsch S, McGrath JP, Varshavsky A (1987) The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature* 329:131-134
- Kirchhoff LV, Kim KS, Engman DM, Donelson JE (1988) Ubiquitin genes in trypanosomatidae. *J Biol Chem* 263:12698-12704
- Klobutcher LA, Swanton MT, Donini P, Prescott DM (1982) All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc Natl Acad Sci USA* 78:3015-3019
- Kraut H, Lipps HJ, Prescott DM (1986) The genome of hypotrichous ciliates. *Int Rev Cytol* 99:1-28
- Lee H, Simon JA, Lis JT (1988) Structure and expression of ubiquitin genes of *Drosophila melanogaster*. *Mol Cell Biol* 8:4727-4735
- Leung DW, Spencer SA, Cachianes G, Hammonds RJ, Collins C, Henzel WJ, Barnard R, Waters MJ, Wood WI (1987) Growth hormone receptor and serum binding protein: purification, cloning and expression. *Nature* 330:537-543
- Levinger L, Varshavsky A (1982) Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the *Drosophila* genome. *Cell* 28:375-385
- Lund PK, Moats-Staats BM, Simmons JG, Hoyt E, D'Ercole J, Martin F, Van Wyk JJ (1985) Nucleotide sequence analysis of a cDNA encoding human ubiquitin reveals that ubiquitin is synthesized as a precursor. *J Biol Chem* 260:7609-7613
- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Martindale DW (1989) Codon usage in *Tetrahymena* and other ciliates. *J Protozool* 36:29-34
- Miceli C, La Terza A, Melli M (1989) Isolation and structural characterization of cDNA clones encoding the mating pheromone Er-1 secreted by the ciliate *Euplotes raikovi*. *Proc Natl Acad Sci USA* 86:3016-3020
- Müller-Taubenberger A, Hagmann J, Noegel A, Gerisch G (1988) Ubiquitin gene expression in *Dictyostelium* is induced by heat and cold shock, cadmium, and inhibitors of protein synthesis. *J Cell Sci* 90:51-58
- Nedospasov SA, Georgiev GP (1980) Non-random cleavage of SV40 DNA in the compact minichromosome and free in solution by micrococcal nuclease. *Biochem Biophys Res Commun* 92:532-539
- Olins DE, Olins AL, Cacheiro LH, Tan EM (1989) Proliferating cell nuclear antigen/cyclin in the ciliate *Euplotes eurystomus*: Localization in the replication band and in micronuclei. *J Cell Biol* 109:1399-1410
- Özkaynak E, Finley D, Solomon MJ, Varshavsky A (1987) The yeast ubiquitin genes: a family of natural gene fusions. *EMBO J* 6:1429-1439
- Preer Jr JR, Preer LB, Rudman BM, Barnett AJ (1985) Deviation from the universal code shown by the gene for surface protein 51A in *Paramecium*. *Nature* 314:188-190
- Rechsteiner M (1987) Ubiquitin-mediated pathways for intracellular proteolysis. *Annu Rev Cell Biol* 3:1-30
- Redman KL, Rechsteiner M (1989) Identification of the long ubiquitin extension as ribosomal protein S27a. *Nature* 338:438-440
- Ridsdale JA, Davie JR (1987) Chicken erythrocyte polynucleosomes which are soluble at physiological ionic strength and contain linker histones are highly enriched in β -globin gene sequences. *Nucleic Acids Res* 15:1081-1096
- Roberson A (1988) The 5S ribosomal RNA gene of *Euplotes*. Ph.D thesis, University of Tennessee, Knoxville
- Roberson AE, Wolffe AP, Hauser LJ, Olins DE (1989) The 5S RNA gene minichromosome of *Euplotes*. *Nucleic Acids Res* 17:4699-4712
- Schlesinger DH, Goldstein G (1975) Molecular conservation of 74 amino acid sequence of ubiquitin between cattle and man. *Nature* 255:423-424
- Sharp PM, Li W-H (1987) Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. *J Mol Evol* 25:58-64
- Siegelman M, Bond MW, Gallatin WM, St John T, Smith HT, Fried VA, Weissman IL (1986) Cell surface molecule associated with lymphocyte homing is a ubiquitinated branched-chain glycoprotein. *Science* 231:823-829
- Soede E, Miura K, Nakaso A, Kimura G (1977) Nucleotide sequence around the replication origin of polyoma virus DNA. *FEBS Lett* 79:383-389
- Stinchcomb DT, Thomas M, Kelly J, Selker E, Davis RW (1980) Eukaryotic DNA segments capable of autonomous replication in yeast. *Proc Natl Acad Sci USA* 77:4559-4563
- Swindle J, Ajioka J, Eisen H, Sanwal B, Jacquemot C, Browder Z, Buck G (1988) The genomic organization and transcription of the ubiquitin genes of *Trypanosoma cruzi*. *EMBO J* 7:1121-1127
- Vierstra RD, Langan SM, Schaller GE (1986) Complete amino acid sequence of ubiquitin from the higher plant *Avena sativa*. *Biochemistry* 25:3105-3108
- Watson DC, Levy-Wilson B, Dixon GH (1978) Free ubiquitin is a non-histone protein of trout testis chromatin. *Nature* 276:196-198
- Wiborg O, Pedersen MS, Wind A, Berglund LE, Marcker KA, Vuust J (1985) The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences. *EMBO J* 4:755-759
- Wu C (1980) The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* 286:854-860
- Yang TP, Hansen SK, Oishi KK, Ryder OA, Hamkalo BA (1982) Characterization of a cloned repetitive DNA sequence concentrated on the human X chromosome. *Proc Natl Acad Sci USA* 79:6593-6597
- Yarden Y, Escobedo JA, Kuang WJ, Yang-Feng TL, Daniel TO, Tremble BM, Chen EY, Ando ME, Harkins RN, Franke U, Fried VA, Ullrich A, Williams LT (1986) Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors. *Nature* 323:226-232