Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 1593–1596, March 1991
Biochemistry

Eliminated chromatin of *Ascaris* contains a gene that encodes a putative ribosomal protein

A. ETTER, M. ABOUTANOS, H. TOBLER, AND F. MÜLLER

Institute of Zoology, University of Fribourg, Pérolles, CH-1700 Fribourg, Switzerland

Communicated by Joseph G. Gall, November 30, 1990 (received for review September 21, 1990)

ABSTRACT **Chromatin diminution in the nematodes *Parascaris equorum* and *Ascaris lumbricoides* leads to the formation of somatic cells that contain less DNA than the germ-line cells. We present molecular evidence for the coding potential of germ-line-specific DNA. We report on a cDNA clone that codes for a putative ribosomal protein (ALEP-1, for *A. lumbricoides* eliminated protein 1). That the corresponding gene is located in the eliminated portion of the genome indicates a difference in germ-line and somatic ribosomes of *A. lumbricoides* and *P. equorum*. Elimination of the ALEP-1 gene from all somatic cells in its fully active state may represent an alternative way to gene regulation.**

It is generally believed that development and differentiation of a multicellular organism are based on regulation of a constant genome in the different cell types rather than on qualitative or quantitative changes in the genetic content. Chromatin diminution in *Parascaris equorum* and *Ascaris lumbricoides* represents the classic exception to this so-called DNA constancy rule. In both nematodes, this phenomenon is correlated with the separation of the germ line from the somatic cell lines. Chromatin diminution was discovered by T. Boveri in *P. equorum* more than 100 years ago and provided the first direct proof for the early segregation and independent development of the germ line and the somatic cell lines (1). This segregation was predicted by A. Weismann in his famous germ-line theory established in 1885 (2). During the process of chromatin diminution, about 25% (*A. lumbricoides*) or 80–90% (*P. equorum*) of the germ-line DNA—mostly satellite DNA, but also some middle-repetitive and single-copy DNA sequences—is expelled from the presomatic cells (3–7). Earlier studies did not reveal any evidence for the elimination of genetically active material (3, 8). In this paper we show that the germ-line-specific material of *A. lumbricoides* contains coding potential. We present a cDNA clone that codes for a putative ribosomal protein and demonstrate that the corresponding gene is lost from the somatic cell lines.* This finding opens new aspects concerning the possible biological function of the chromatin elimination process.

MATERIALS AND METHODS

Growth of *Ascaris* and Nucleic Acid Preparation. Adult females were dissected and intestines, oocytes, and eggs were isolated. Eggs were incubated at 30°C in water containing 0.1% formaldehyde and the RNA was harvested at days 1, 4, and 11 of incubation. Genomic DNA and both total and poly(A)$^+$ RNA were isolated by standard methods (5, 9, 10).

cDNA Library Screening and Southern and Northern Blot Analyses. All hybridizations were done with nick-translated probes. The cDNA library was constructed by Clontech. cDNA complementary to poly(A)$^+$ mRNA of *A. lumbricoides* four-cell embryos was synthesized by oligo(dT) and random priming and then cloned in phage λgt10. About 600,000 plaque-forming units were plated, transferred to nitrocellulose filters, and hybridized under high-stringency conditions. Southern blots were prepared and high-stringency hybridizations were carried out by standard procedures (10). RNA was glyoxylated and then electrophoresed in a 1.2% agarose gel. Northern blot hybridizations were done as described (11).

Sequencing of DNA. The insert of a randomly selected cDNA clone (λA122) was subcloned in plasmid pBSM13+. For sequencing, smaller fragments from the plasmid were then subcloned in an appropriate M13mp18 or M13mp19 phage vector. Both DNA strands were sequenced by the chain-termination method (12).

Computer-Assisted Data Analysis. Analyses were performed by using an AT-386 machine and the software package PC/GENE commercialized by Genofit. Coding capacities of open reading frames were estimated by the program COD-FICK (13). The homology search of the Swiss-Prot data base (release no. 11) was done with the program FSTPSCAN (14). The best score was analyzed by PCOMPARE (15) using 100 randomized runs. Protein alignment was carried out by PALIGN (16) with the structure–genetic matrix. Open gap cost and unit gap cost were set to 4.

RESULTS

cDNA A122 Codes for a Putative Protein Homologous to the Ribosomal Proteins S16 of Yeast and S12 from *Halobacterium marismortui*. A differential screening of a germ-line genomic library was performed by using first-strand cDNA made from poly(A)$^+$ RNA of four-cell embryos versus cDNA derived from 11-day-old larvae. Clones that gave a positive signal with four-cell cDNA only were further selected and analyzed for their genomic organization. One phage, λ94121, turned out to contain only single-copy sequences that are lost during the elimination process. Two different regions were identified that are transcribed in early development (A.E., unpublished data). A restriction fragment containing one of the two regions was used to screen a four-cell-stage cDNA library. Eight independent cDNA clones, with the same insert length, were isolated and one of them, λA122, was randomly selected for complete sequencing (Fig. 1). The insert is 522 base pairs (bp) long and shows all the expected features of a translatable eukaryotic poly(A)$^+$ mRNA. It contains a single, long open reading frame, potentially encoding a protein of 148 amino acids, dubbed ALEP-1 (for *A. lumbricoides* eliminated protein 1). The start codon is surrounded by the sequence motif A^{-3}/G^{+4} which is found to be conserved in the initiation region of many eukaryotic mRNAs and is known to favor the initiation of translation by eukaryotic ribosomes (17–19). A canonical poly(A)$^+$ signal (AATAAA) is found 17 bp from the poly(A)$^+$ tail. Since the sequence upstream of the start codon

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. M59417).

```
        start
GAGA ATG GTAAAAGCGACTAGCGTGAAGGATGTCGATCAACATGAAATTGTTCAGCATATCGCCAAGTTC   70
      M  V  K  A  T  S  V  K  D  V  D  Q  H  E  I  V  Q  H  I  A  K  F

TTAAAGAAGAGTGGTAAGGTGAAGGTGCCGGAATGGTCAGATGTTACTAAGATGGGTATTTCGAAGGAGC  140
 L  K  K  S  G  K  V  K  V  P  E  W  S  D  V  T  K  M  G  I  S  K  E

TAGCGCCTTTAAACAGCGACTGGTATTATGTGAGAACGGCGAGCATCGCACGACGCCTTTATGTGCGATC  210
L  A  P  L  N  S  D  W  Y  Y  V  R  T  A  S  I  A  R  R  L  Y  V  R  S

GCCAACGGGCGTTGATGCACTGAGGCTTGTCTATGGTGGCAGCAAACGACGTGGCGTGGTCCCTAATCAT  280
  P  T  G  V  D  A  L  R  L  V  Y  G  G  S  K  R  R  G  V  V  P  N  H

TTTGCCAAGGCTTCCGGCTCGGTAATCCGTAAAGCCCTGCAGACACTCGAGGCAATCAAATGGGTGCAGA  350
 F  A  K  A  S  G  S  V  I  R  K  A  L  Q  T  L  E  A  I  K  W  V  Q

AACATCCAGATGGTAACGGACGCGTACTTACCAAACAGGGAAGGAAGGATCTCGATCGAATTGCTTCTCA  420
K  H  P  D  G  N  G  R  V  L  T  K  Q  G  R  K  D  L  D  R  I  A  S  Q

GATGCGTCAGAACGATCGGTTCACTGCG TAG CAGCAGAAATATTTCTGTGTTGTCTGCGTGG AATAAA AA 490
  M  R  Q  N  D  R  F  T  A stop                                  poly A signal

TGTTTGTTGTTTGGCAAAAAAAAAAAAAAAAA
                  poly A tail
```

FIG. 1. Nucleotide sequence of the ALEP-1 cDNA clone λA122 and the deduced amino acid sequence (single-letter code). The polyadenylylation signal as well as the putative start and stop codons of translation are boxed. The motif A^{-3}/G^{+4} which surrounds the start codon is shown in bold letters.

is only 4 bp long, we assume that the leader sequence of the cDNA is not complete.

Comparison of the predicted ALEP-1 amino acid sequence with sequences stored in the Swiss-Prot data base (release no. 11) revealed a significant similarity to only one protein, the ribosomal protein S16A of *Saccharomyces cerevisiae* (20). A second computer analysis, done with the ribosomal data base of T. Tanaka, School of Medicine, University of the Ryukyus, Okinawa, Japan) showed an additional homology to ribosomal protein S12 of *H. marismortui* (hmas-12) (21). The alignment of ALEP-1 with the two protein sequences is shown in Fig. 2. Between ALEP-1 and S16A, 42% of the amino acids are identical and 17.5% of them have undergone conservative changes. The hmas-12 sequence, on the other hand, has 34.2% identity with ALEP-1 and 17.8% conservative amino acid substitutions. Sequence conservation among evolutionarily widely separated species such as nematodes, yeast, and archaebacteria implies an important biological function, and we therefore suggest that ALEP-1 is a ribosomal protein. This argument is further supported by the small size of the ALEP-1 transcript, because eukaryotic ribosomal proteins are generally known to be encoded by small mRNAs (22–25).

Genomic Organization of the ALEP-1 Gene. On Southern blots, the ALEP-1 cDNA probe hybridizes predominantly to a single restriction fragment in the germ-line DNA of *A. lumbricoides* (Fig. 3*A*) and also to a prominent band in the germ-line DNA of *P. equorum* (Fig. 3*B*). In both species, this band corresponds to single-copy DNA. An additional, fainter band is observed in the genomic DNAs of both nematodes. This represents either a distantly related gene or, more likely, because of its co-elimination, a less-frequent allele of the same gene. Since the genomic DNAs were isolated from pools of animals, we cannot distinguish between these two alternatives. The *A. lumbricoides* cDNA probe cross-hybridizes under high-stringency conditions to the *P. equo-*

```
ALEP-1  M V K A T S V K D V D Q H E I V Q H I A K F L K K S G K V K
S16A    M P - G V S V R D V A A Q D F I N A Y A S F L Q R Q G K L E
hmas-12 - - - - A T L Y D V P P E E L I E A L T E T - - - - - L A

ALEP-1  V P E W S D V T K M G I S K E L A P L N S D - W Y Y V R T A
S16A    V P G Y V D I V K T S S G N E M P P Q D A E G W F Y K R A A
hmas-12 D E D W A E F T K T G V D R E L P P E Q E D - F W T R R A A

ALEP-1  S I A R R L Y V R S P T G V D A L R L V Y G - - - - G S K R
S16A    S V A R H I Y M R K Q V G V G K L N K L Y G - - - - G A K S
hmas-12 S L L R K V A V D G P V G V N A L R S E Y G T S K Q G T T R

ALEP-1  R G V V P N H F A K A S G S V I R K A L Q T L E A I K W V Q
S16A    R G V R P Y K H I D A S G S I N R K V L Q A L E K I G I V E
hmas-12 Y R V R P H Q K T K G S G N I I R T A L Q Q L E D A G Y V E

ALEP-1  K H P D G N G R V L T K Q G R K D L D R I A S Q M R Q N - D
S16A    I S P K G - G R R I S E N G Q R D L D R I A A Q T L E E - D
hmas-12 T S E - - N D R R V T G D G R S L L D D T A G D L L T E L D

ALEP-1  R F T - - - - - A 148 Aa
S16A    - - - - - - - - E 143 Aa
hmas-12 R P E E L E R Y A 146 Aa
```

FIG. 2. Computer alignment of the predicted amino acid sequence (single-letter code) of ALEP-1 with the ribosomal proteins S16A of yeast (20) and S12 of *H. marismortui* (hmas-12) (21). Identical amino acids are boxed and similar residues are symbolized by bold letters. Amino acids said to be similar are as follows: A, S, T/D, E/N, Q/R, K/I, L, M, V/F, Y, W. Dashes indicate gaps inserted to optimize alignment with ALEP-1.

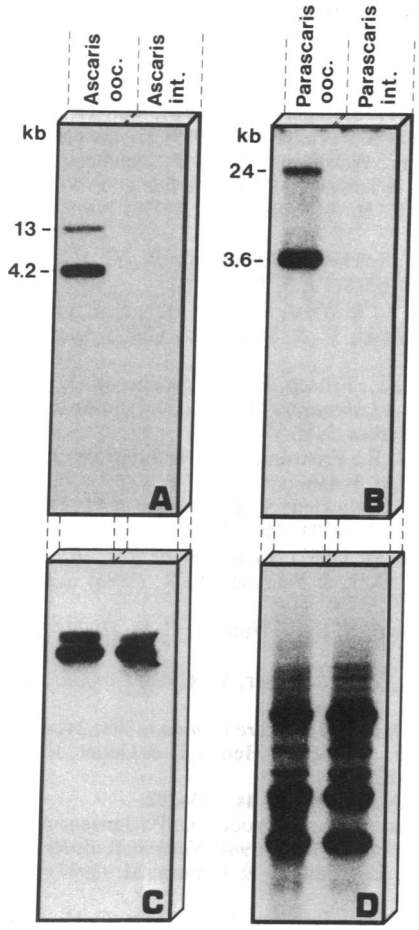

FIG. 3. Genomic distribution of the ALEP-1 gene. The ALEP-1 cDNA clone A122 (Fig. 1) was hybridized to Southern blots of germ-line DNA from oocytes (ooc.) and intestinal DNA (int.) of the two chromatin-eliminating nematodes *A. lumbricoides* and *P. equorum*. As a size marker we used λ DNA digested with *Hin*dIII. kb, Kilobases. (*A*) *A. lumbricoides* DNA (3 μg per lane) digested with *Hin*dIII. (*C*) Rehybridization of the filter in *A* with the nick-translated clone pAlr8, containing an *A. lumbricoides* rDNA repeat, confirms that the same amounts of *Ascaris* DNA were loaded in the two lanes. The probe hybridizes to the major organizational form of the *Ascaris* rDNA as well as to a fainter band, representing the minor rDNA form (26). (*B*) Total *P. equorum* DNA digested with *Eco*RI. Whereas the left lane contains 10 μg of oocyte DNA, only 3 μg of somatic DNA was loaded in the right lane in order to compensate for the large amount of highly repetitive germ-line-specific material, which is lost from the somatic genome of *P. equorum* during the process of chromatin elimination. (*D*) Rehybridization of the filter in *B* with the *Ascaris* rDNA probe as control confirms that roughly the same somatic genome equivalent was loaded in each slot. The *P. equorum* rDNA repeats are cut several times by the restriction enzyme *Eco*RI, resulting in a complicated hybridization pattern.

rum DNA, indicating a high degree of sequence conservation between the two nematodes. No signal is detected in the somatic DNA of both species, demonstrating the elimination of the ALEP-1 gene from all somatic cells in both nematodes. As a control, the blot was retested with cloned *Ascaris* rDNA (26).

Expression Pattern of ALEP-1. On Northern blot analysis with *A. lumbricoides* RNA, a single type of transcript can be detected (Fig. 4). Its length, 650 nucleotides, corresponds well to the size of the analyzed cDNA clone. The ALEP-1 transcript is present in large amounts in oocytes. Only a weak signal can be seen in fertilized eggs, a dauer stage in which development is arrested until the eggs are exposed to an appropriate environment (27). Fertilized eggs are transcriptionally inactive and contain the ALEP-1 transcripts probably in the form of stably stored mRNAs. In developing early embryos, the relative amount of ALEP-1 transcripts increases dramatically, indicating transcription of the ALEP-1 gene (four-cell stage). During these developmental stages, ALEP-1 expression parallels general polymerase II transcription activity in *A. lumbricoides* embryos (28). This is exactly what is to be expected for housekeeping-function genes, such as those which code for ribosomal proteins. After the four-cell stage, all presomatic cells undergo chromatin diminution. Therefore, the ALEP-1 gene will be lost from the somatic cells and no further transcripts can be formed. Indeed, no signal is detected in 11-day-old larvae, which are composed of roughly 500 somatic cells but only a few germ cells.

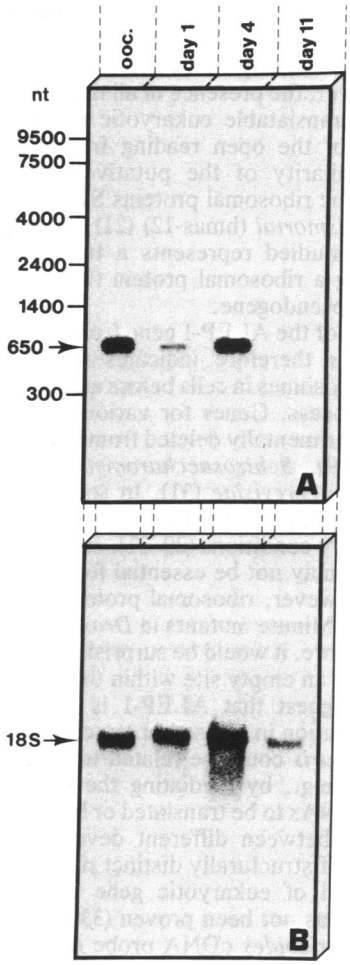

FIG. 4. Northern blot with total RNA from oocytes (ooc.), fertilized eggs (day 1), four-cell-stage larvae (day 4), and 11-day-old larvae (day 11). (*A*) ALEP-1 cDNA clone A122 (Fig. 1) was used as probe. Numbers at left show the molecular size in nucleotides (nt) of the RNA ladder from BRL. Arrow indicates the deduced length of the mRNA that codes for ALEP-1. (*B*) The same filter was rehybridized with a nick-translated, cloned *Ascaris* 18S rDNA fragment (26), showing that the same amount of RNA (5 μg) was loaded in the first three lanes (oocyte, day 1, and day 4). The day-11 lane, however, contained less RNA, but even after very long exposure time, no hybridization signal with the ALEP-1 probe could be detected.

DISCUSSION

We have shown that the germ-line-limited DNA from the chromatin-eliminating nematode *A. lumbricoides* contains protein-coding potential. We have isolated and analyzed a cDNA clone whose corresponding genomic sequence is eliminated from all somatic cells. This sequence is strongly

transcribed in oocytes and in early embryonic stages up to its elimination, but so far there is no direct evidence that the RNA product is actually translated to produce a functional protein. However, the presence of all the expected structural features of a translatable eukaryotic mRNA (Fig. 1), the completeness of the open reading frame, and the highly significant similarity of the putatively encoded protein ALEP-1 with the ribosomal proteins S16A of yeast (20) and S12 of *H. marismortui* (hmas-12) (21) strongly suggest that the sequence studied represents a functional gene, most likely encoding a ribosomal protein (Fig. 2), rather than a nonfunctional pseudogene.

The absence of the ALEP-1 gene from the somatic cells of *A. lumbricoides*, therefore, indicates a structural difference between the ribosomes in cells before and after the chromatin elimination process. Genes for various ribosomal proteins have been experimentally deleted from the genomes of *Escherichia coli* (29), *Schizosaccharomyces pombe* (30), and *Saccharomyces cerevisiae* (31). In some cases, no altered phenotype has been observed upon deletion, at least not under laboratory conditions (29–31), suggesting that some of these proteins may not be essential for these organisms. In metazoans, however, ribosomal proteins are thought to be essential (e.g., Minute mutants in *Drosophila melanogaster*; ref. 32). Therefore, it would be surprising if the elimination of ALEP-1 leaves an empty site within the ribosomal complex. Rather, we suggest that ALEP-1 is replaced by another protein. A variation in ribosomal structure during the development of *Ascaris* could be related to translational control mechanisms—e.g., by mediating the selection of specific families of mRNAs to be translated or by modulating the rate of translation between different developmental stages. A potential role of structurally distinct ribosomes in the translational control of eukaryotic gene expression has been proposed but has not been proven (33–35).

The *A. lumbricoides* cDNA probe cross-hybridizes even under high-stringency conditions to the *P. equorum* DNA, indicating a high degree of sequence conservation of this gene between the two nematodes. In *Parascaris*, as in *Ascaris*, this sequence is eliminated from all somatic cells. Obviously, not only the sequence of this gene but also its behavior in the elimination process have been maintained during evolution of the two nematodes. This is not a trivial result, in light of their very different genomic organization (3), and indicates that the elimination process might have an important function in the development of the two nematodes. That the ALEP-1 gene is expelled from all somatic cell lineages of both nematodes, most likely during its fully active state (Fig. 4), opens the intriguing possibility that the process of chromatin diminution may serve as an alternative way to gene regulation.

A.E. thanks Dr. P. Milasevic for many excellent and most stimulating discussions. We thank Drs. M. L. Birnstiel, W. J. Gehring, J. G. Gall, S. Ward, and all colleagues in our laboratory for helpful comments on the manuscript. We are grateful to Dr. T. Tanaka for the analysis done on ALEP-1 with his ribosomal protein data base. This research was supported by Swiss National Science Foundation Grants 3.113-0.85 and 3.141-0.88.

1. Boveri, T. (1887) *Anat. Anz.* **2**, 688–693.
2. Weismann, A. (1885) *Die Continuität des Keimplasmas als Grundlage einer Theorie der Vererbung* (Fischer, Jena, F.R.G.).
3. Tobler, H. (1986) in *Problems in Cell Differentiation*, ed. Hennig, W. (Springer, New York), Vol. 13, pp. 1–69.
4. Pimpinelli, S. & Goday, C. (1989) *Trends Genet.* **5**, 310–315.
5. Müller, F., Walker, P., Aeby, P., Neuhaus, H., Felder, H., Back, E. & Tobler, H. (1982) *Nucleic Acids Res.* **10**, 7493–7510.
6. Moritz, K. B. & Roth, G. E. (1976) *Nature (London)* **259**, 55–57.
7. Aeby, P., Spicher, A., de Chastonay, Y., Müller, F. & Tobler, H. (1986) *EMBO J.* **5**, 3353–3360.
8. Bennett, K. & Ward, S. (1986) *Dev. Biol.* **118**, 141–147.
9. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
11. Natzle, J. E., Fristrom, D. K. & Fristrom, J. W. (1988) *Dev. Biol.* **129**, 428–438.
12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
13. Ficket, J. W. (1982) *Nucleic Acids Res.* **10**, 5303–5318.
14. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441.
15. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48**, 443–453.
16. Meyers, E. W. & Miller, W. (1988) *Comput. Appl. Biosci.* **4**, 11–17.
17. Kozak, M. (1984) *Nature (London)* **308**, 241–246.
18. Morlé, F., Lopez, B., Henni, T. & Godet, J. (1985) *EMBO J.* **4**, 1245–1250.
19. Kozak, M. (1986) *Cell* **44**, 283–292.
20. Molenaar, C. M. T., Woudt, L. P., Jansen, A. E. M., Mager, W. H. & Planta, R. J. (1984) *Nucleic Acids Res.* **12**, 7345–7358.
21. Kimura, J., Arndt, E. & Kimura, M. (1987) *FEBS Lett.* **224**, 65–70.
22. Fried, H. M., Pearson, N. J., Kim, C. H. & Warner, J. R. (1981) *J. Biol. Chem.* **256**, 10176–10183.
23. Woolford, J. L., Jr., Hereford, L. M. & Rosbash, M. (1979) *Cell* **18**, 1247–1259.
24. Pierandrei-Amaldi, P. & Campioni, N. (1982) *Cell* **30**, 163–171.
25. Dudov, K. P. & Perry, R. P. (1984) *Cell* **37**, 457–468.
26. Back, E., Van Meir, E., Müller, F., Schaller, D., Neuhaus, H., Aeby, P. & Tobler, H. (1984) *EMBO J.* **3**, 2523–2529.
27. Poinar, G. O., Jr. (1983) *The Natural History of Nematodes* (Prentice-Hall, Englewood Cliffs, NJ), pp. 203–243.
28. Cleavinger, P. J., McDowell, J. W. & Bennett, K. L. (1989) *Dev. Biol.* **133**, 600–604.
29. Dabbs, E. R. (1986) in *Structure, Function, and Genetics of Ribosomes*, eds. Hardesty, B. & Kramer, G. (Springer, New York), pp. 733–748.
30. Beltrame, M. & Bianchi, M. E. (1990) *Mol. Cell. Biol.* **10**, 2341–2348.
31. Remacha, M., Santos, C. & Ballesta, J. P. G. (1990) *Mol. Cell. Biol.* **10**, 2182–2190.
32. Kay, M. A. & Jacobs-Lorena, M. (1987) *Trends Genet.* **3**, 347–351.
33. Gunderson, J. H., Sogin, M. L., Wollett, G., Hollingdale, M., de la Cruz, V. F., Waters, A. P. & McCutchan, T. F. (1987) *Science* **238**, 933–937.
34. Selker, E. U., Stevens, J. N. & Metzenberg, R. L. (1985) *Science* **227**, 1340–1343.
35. Peterson, R. C., Doering, J. L. & Brown, L. D. (1980) *Cell* **20**, 131–141.