Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 6465–6469, November 1982
Biochemistry

A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes

(left-handed helical DNA/dispersed repeated sequence/stretched alternating dT-dG or dC-dG/actin gene)

HIROSHI HAMADA, MARIANNE G. PETRINO, AND TAKEO KAKUNAGA

Cell Genetics Section, Laboratory of Molecular Carcinogenesis, National Cancer Institute, Bethesda, Maryland 20205

Communicated by Igor B. Dawid, July 29, 1982

ABSTRACT **By Southern blotting and hybridization analysis using ^{32}P-labeled poly(dT-dG)·poly(dC-dA) as a probe, we have found, in eukaryotic genomes, a huge number of stretches of dT-dG alternating sequence, a sequence that has been shown to adopt the Z-DNA conformation under some conditions. This sequence was found in all eukaryotic genomes examined from yeast to human, indicating extraordinary evolutionary conservation. The number of the sequence ranged from about 100 in yeast to tens of thousands in higher eukaryotes. Comparison of nucleotide sequences of dT-dG alternating regions and its flanking regions in several cloned genes showed that the repeated element [the Z(T-G) element] consists only of dT-dG alternating sequence with variable length. The presence of another purine–pyrimidine alternating sequence was also surveyed in eukaryotic genomes by Southern blot hybridization using ^{32}P-labeled poly(dG-dC)·poly(dG-dC) as the probe. The stretches of dC-dG alternating sequence [the Z(C-G) element] were found to be moderately repetitive in human, mouse, and salmon genomes. However, a few and no copies of the Z(C-G) element were found in yeast and calf genomes, respectively. These results provide evidence for the abundance of potential Z-DNA-forming sequences in nature.**

Recent physicochemical studies of DNA conformation have shown that some synthetic DNAs with certain primary sequences have a novel conformation, called the Z form (1–4). Although the Z conformation was first observed with poly(dG-dC) and most studies on Z-DNA have been done with it, other synthetic purine–pyrimidine alternating sequences such as poly(dT-dG)·poly(dC-dA) (2, 5, 6) and poly(ds^{4}A-dT) (2) have also been shown to adopt the Z conformation. Until recently, however, there has been little direct evidence that such Z-DNA-forming sequences exist in native DNA. Nordheim *et al.* (7) have shown that a specific antibody against brominated poly(dG-dC)·poly(dG-dC), a polymer that forms a Z-DNA under physiological conditions, reacts with interband regions of *Drosophila* polytene chromosomes. Recently, we have shown that the human genome has approximately 10^5 copies of stretches of dT-dG alternating sequence (8). A tandem block of 17 T-G (9) and 27 T-G dinucleotides (10) were found in human globin and in mouse immunoglobin genes, respectively, but the general occurrence of these sequences in the genomes was not investigated.

Here we report that one of the Z-DNA-forming sequences, a long stretch of dT-dG alternating sequence is the sole unit of a repeated element [designated the Z(T-G) element] that is highly conserved throughout eukaryotic genome evolution. Furthermore, another Z-DNA-forming sequence, a stretch of dC-dG sequence, was found to be at least a part of another repeated element [designated the Z(C-G) element] and is moderately repeated in human, mouse, and salmon genomes but not in yeast or calf DNA.

MATERIALS AND METHODS

Materials. Calf thymus DNA, salmon sperm DNA, and poly(dG-dC)·poly(dG-dC) were purchased from Sigma. Poly(dT-dG)·poly(dC-dA) was obtained from Boehringer Mannheim. High molecular weight DNAs were extracted from nuclei of HeLa cells and mouse L cells and from whole cells of *Saccharomyces cerevisiae*. Nuclei were prepared as described (11). High molecular weight DNAs from *Drosophila melanogaster* and *Xenopus laevis* were kindly provided by M. E. Digan and H. U. Affolter, respectively.

Isolation and characterization of λHa-25, a recombinant phage containing a human cardiac muscle actin gene, have been described (12). λHa-314 is a recombinant phage probably containing a human actin gene, since a cDNA sequence from *Dictyostelium* actin mRNA (13) hybridized to the cloned human DNA in λHa-314. This clone was isolated from a human DNA library constructed from partially *Eco*RI-digested Hut-14 cell DNA, a chemically transformed human fibroblast (14).

Hybridization. *Eco*RI-digested DNAs were fractionated on a 0.7% agarose gel and blotted to nitrocellulose filters as described by Southern (15). Cellular DNAs used for Southern blotting are of high molecular weight [30–50 kilobases (kb)] except salmon sperm DNA (10–20 kb). For spot hybridization, DNAs were denatured and spotted directly on nitrocellulose papers as described by Kafatos *et al.* (16).

Filters, in all cases, were prehybridized and hybridized to a ^{32}P-labeled probe as described (17). ^{32}P-Labeled probes were prepared by nick-translation of poly(dT-dG)·poly(dC-dA) or poly(dG-dC)·poly(dG-dC). The specific activities of these probes were 1.0×10^8 cpm/μg and 0.8×10^8 cpm/μg, respectively; 0.5 μg of probe was used for each hybridization. It should be noted that denatured *Escherichia coli* DNA at 25 μg/ml was always present in the prehybridization and hybridization mixtures. After hybridization, filters were washed three times with 0.15 M NaCl/0.015 M Na citrate/0.1% $NaDodSO_4$ at 50°C and twice with 15 mM NaCl/1.5 mM Na citrate/0.1% $NaDodSO_4$ at 50°C. Autoradiography was carried out at −80°C for the indicated time.

Estimation of Copy Number. The haploid genome size of each organism was taken from ref. 18. λHa-25 is about 50 kb and contains one copy of $d(T\text{-}G)_{25}$ in an intron (12). Therefore, for example, a single copy of $(dT\text{-}dG)_{25}$ in 100 ng of λHa-25 DNA is equivalent to 10^5 copies of $(dT\text{-}dG)_{25}$ in 60 ng of the human haploid genome: i.e., $(10^2 \text{ ng}/5 \times 10^4) \times (3 \times 10^9/60 \text{ ng}) = 10^5$. The approximate copy number of $(dT\text{-}dG)_n$ in the genome of each organism was estimated by comparison of the hybrid-

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.

Abbreviations: kb, kilobase(s); bp, base pair(s).

ization intensity of each cellular DNA with that of λHa-25, after spot hybridization (see Table 1).

RESULTS

The Z(T-G) Element Is Highly Repeated and Evolutionarily Conserved. First, the presence of the $(dT\text{-}dG)_n$ sequence in various eukaryotic cells from yeast to human was surveyed. We used Southern blot hybridization and ^{32}P-labeled poly(dT-dG)·poly(dC-dA) as a probe. The specificity of the probe was demonstrated by its hybridization only to the fragment containing a stretch of dT-dG in restriction digests of λHa-25 DNA (Fig. 1). λHa-25, a clone containing a human cardiac muscle actin gene, has only one copy of $(dT\text{-}dG)_{25}$ in one of its introns (12). Therefore, the intensity of the hybridized band of λHa-25 (Fig. 1, lane 1) is approximately equivalent to 6×10^2, 6×10^2, 6×10^2, 2×10^2, and 2 copies of $(dT\text{-}dG)_{25}$ in the human, calf, mouse, chicken, salmon, and yeast genome, respectively. This probe gave a "smear" pattern in the *Eco*RI-digested total DNA from all species examined. It is likely that $(dT\text{-}dG)_n$ is randomly dispersed in these genomes. The intensity of the hybridization observed in each cellular DNA suggests the presence of many copies of the $(dT\text{-}dG)_n$ sequence in each genome. The approximate copy number of the Z(T-G) element in various genomes was also determined by spot hybridization (Fig. 2*a*). As summarized in Table 1, human, calf, mouse, chicken, *Xenopus*, salmon, *Drosophila*, and yeast have approximately 5×10^4, 3×10^4, 10^5, 4×10^3, 10^5, 2×10^5, 2×10^3 and 10^2 copies of the Z(T-G) element, respectively, assuming that the average dT-dG alternation in the Z(T-G) element is 25. The organisms with a larger genome seem to have more copies of the Z(T-G) element, except that the chicken genome has fewer.

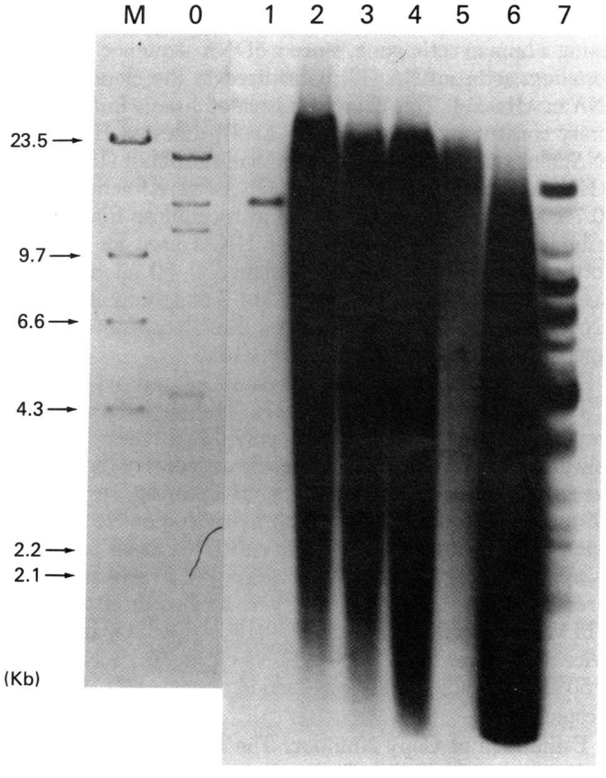

FIG. 1. Detection of the dT-dG alternating sequence by Southern blotting. One hundred nanograms of λHa-25 (lanes 0 and 1) or 10 μg of human (HeLa cell) (lane 2), calf (lane 3), mouse (L cell) (lane 4), chicken (lane 5), salmon (lane 6), or yeast (lane 7) DNA was completely digested with *Eco*RI. Each digest was electrophoresed on a 0.7% agarose gel, blotted on nitrocellulose paper, and hybridized to ^{32}P-labeled poly(dT-dG)·poly(dC-dA). Autoradiography (lane 1–7) was carried out for 18 hr. Lanes M and 0 are ethidium bromide staining patterns. Lane M: *Hin*dIII digests of λDNA served as molecular size standards.

Table 1. Estimation of approximate copy number of the Z(T-G) element in various eukaryotic genomes

	Haploid genome size, bp	Approximate copy number
Human	3×10^9	5×10^4
Calf	3×10^9	3×10^4
Mouse	3×10^9	10^5
Chicken	10^9	4×10^3
Xenopus	2×10^{10}	10^5
Salmon	6×10^9	2×10^5
Drosophila	2×10^8	2×10^3
Yeast	10^7	10^2

Approximate copy numbers were estimated by comparison of the hybridization intensities of the spots in Fig. 2. For example, to estimate the copy number in the human genome, the intensities of spots A, B, and C are equivalent to 10^5, 10^4, and 10^3 copies of $(dT\text{-}dG)_{25}$ in the human haploid genome, respectively, and comparison of the hybridization intensity of human DNA (spot 1) with those of spots A, B, and C gives a copy number of 5×10^4.

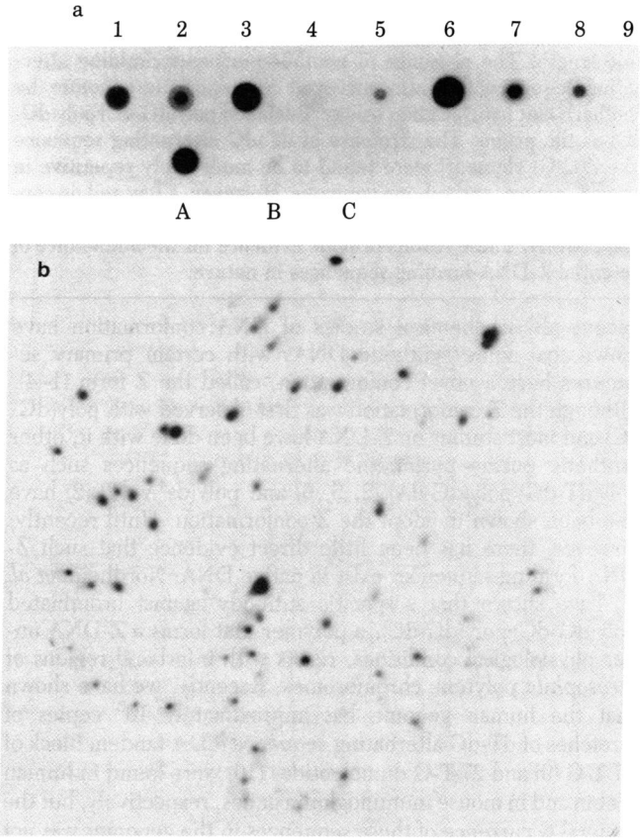

FIG. 2. (*a*) Quantification of the (dT-dG) alternating sequence by spot hybridization. One hundred nanograms (spot A), 10 ng (spot B), or 1 ng (spot C) of λHa-25 or 60 ng of human, calf, mouse, chicken, *Xenopus*, salmon, *Drosophila*, yeast, or *Eco*RI-digested pBR322 (spots 1–9, respectively) DNA was denatured, spotted on nitrocellulose paper, and hybridized to ^{32}P-labeled poly(dT-dG)·poly(dC-dA). Autoradiography was carried out for 36 hr. (*b*) Screening of the human DNA library using poly(dT-dG)·poly(dC-dA) as a probe. Four hundred recombinant phages from the human DNA library (19) were plated on an agarose dish. All phages were transfered to nitrocellulose filters (20) and hybridized to ^{32}P-labeled poly(dT-dG)·poly(dC-dA). Autoradiography was for 20 hr. Approximately 80 phages showed a positive signal on the autoradiogram.

The calf genome contains several classes of satellite DNA, some of which give rise to discrete fragments after *Eco*RI digestion (21). The poly(dT-dG)·poly(dC-dA) probe detected two of these satellite DNA bands in the Southern blot hybridization experiment (Fig. 1, lane 3). Thus, in the calf genome, dispersal of the Z(T-G) element extends even to satellite DNAs.

The repeated existence of the Z(T-G) element was also indicated by the following observation. When the human DNA library was screened using poly(dT-dG)·poly(dC-dA) as a probe, about 20% (80/400) of the recombinant phages showed positive hybridization to the probe (Fig. 2*b*). Since the average size of the inserted human DNA in each recombinant phage is 20 kb, the human haploid genome [3×10^9 base pairs (bp)] has 3×10^4 copies of the Z(T-G) element; i.e., $3 \times 10^9 \times 0.2/2 \times 10^4 = 3 \times 10^4$, which is roughly consistent with the value (5×10^4) estimated by spot hybridization.

The Structure of the Z(T-G) Element. We have previously determined the nucleotide sequence of the $(dT\text{-}dG)_n$-containing region and its flanking region of a human cardiac muscle actin gene (12). To understand the more general structure in other occurrences of the Z(T-G) element [i.e., whether the Z(T-G) element consists only of $(dT\text{-}dG)_n$ sequence or can contain $(dT\text{-}dG)_n$ plus other sequences], the primary sequences of the Z(T-G) element in other genes were compared with those in a human cardiac muscle actin gene. λHa-314, a clone that seems to contain another human actin gene, was also shown to contain a Z(T-G) element by hybridization with ^{32}P-labeled poly(dT-dG)·poly(dC-dA) (Fig. 3*a*). It is noteworthy that the probe again hybridized to a single restriction fragment among a number of fragments, indicating the high specificity of the hybridization. The locations of the presumptive coding sequence and of the Z(T-G) element in λHa-314 are shown in Fig. 3*b*. DNA sequence analysis of the $(dT\text{-}dG)_n$-containing region has shown a 15 times dT-dG precisely alternating sequence, which is the Z(T-G) element (Fig. 3*c*). The Z(T-G) elements of the human cardiac muscle actin gene and of λHa-314 share no homology

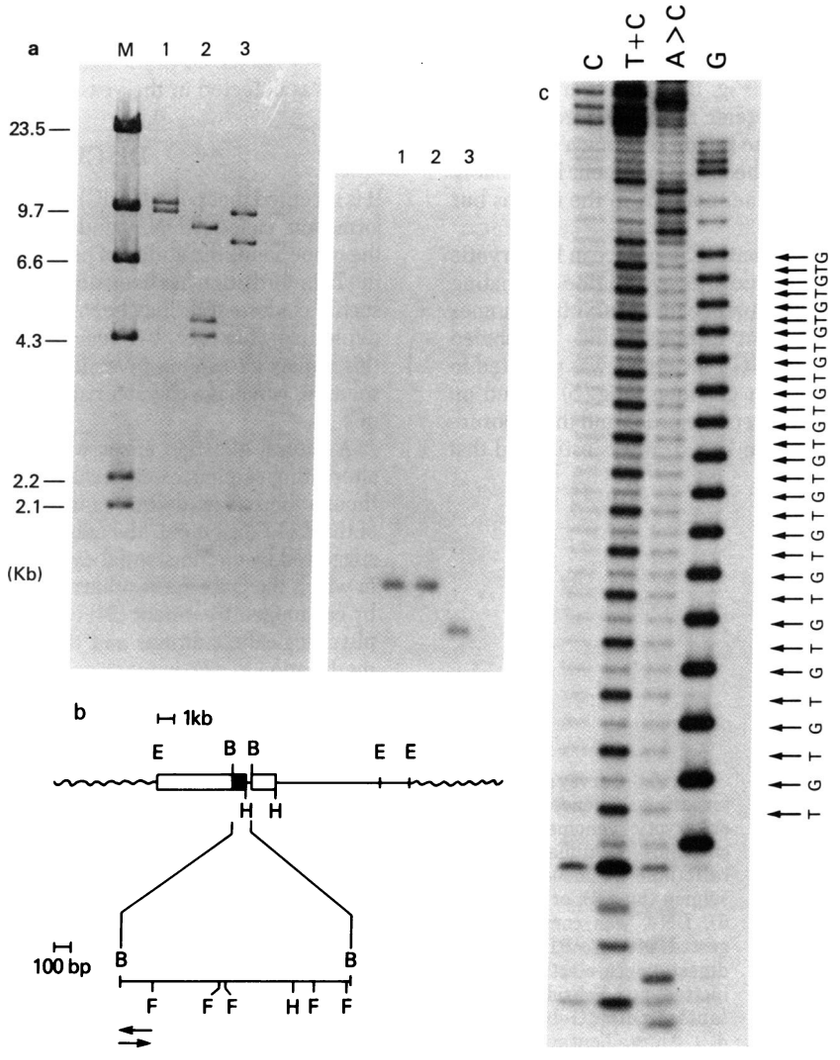

FIG. 3. (*a*) Mapping of the Z(T-G) element in λHa-314. The 15-kb *Eco*RI fragment cloned in λHa-314 was subcloned at the *Eco*RI site of pBR325. The subcloned DNA was digested with *Bam*HI (lane 1), *Bam*HI/*Eco*RI (lane 2), or *Bam*HI/*Hin*dIII (lane 3), and the resulting restriction fragments were separated on a 1% agarose gel, blotted, and hybridized to ^{32}P-labeled poly(dT-dG)·poly(dC-dA). Lane M: *Hin*dIII digests of λDNA served as molecular size standards. (*Left*) Staining pattern visualized with ethidium bromide. (*Right*) Autoradiogram. (*b*) Restriction map of λHa-314. E, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; F, *Hinf* I. The wavy lines show the right and left arms of charon 4A phage. The other part is the cloned human gene. The open box indicates the region hybridized to ^{32}P-labeled pcD DNA, which is a recombinant DNA containing a cDNA copy of *Dictyostelium* actin mRNA. The black box denotes the region hybridized to ^{32}P-labeled poly(dT-dG)·poly(dC-dA). (*c*) dT-dG alternating sequence in λHa-314. The DNA sequence of the region hybridized to poly(dT-dG)·poly(dC-dA) was determined (22) by the strategy shown by the arrows in *b*. This gel shows the primary sequence from the *Bam*HI site, although the nucleotide sequence was confirmed by sequence analysis of both strands.

1: ----GGATCCAGCCTGTAACACATTCG——— $(TG)_{15}$ ——TAGAGATGGGGTCTCACTATGTTGACCAGG----

2: ----CCTCTACCTTGACCTGAATGCAC—$(TG)_2A(TG)_{25}$—ACTCGTTCCCAGGTATGGAATCTGCTGGCA----

3: ----TCGCTGTCTCTTTTTTGAGGACT—$(TG)_{17}T(TG)_2$—GTCAGTGGGGCTGGAATAAAAGTAGAATAG----

4: ----TCTTAGCAATATAACTTAAGATA—$(TG)_{27}TA(TG)_4$—GGACAAGTTGTTAAATGAATCCCAGCCATT----

FIG. 4. Comparison of (dT-dG) alternating sequences and flanking sequences in various genes: 1, λHa-314; 2, human cardiac actin gene (λHa-25); 3, human globin gene (9); 4, mouse immunoglobulin κ light chain gene (10) [the nucleotide sequence of the $(dT\text{-}dG)_n$-containing strand is shown].

other than the stretch of dT-dG sequence. Furthermore, the sequences flanking $(dT\text{-}dG)_n$ were compared with these four different genes: human cardiac muscle actin gene, λHa-314, human globin gene, and mouse immunoglobulin gene. No significant homology was observed in their flanking sequences (Fig. 4). We, therefore, conclude that the unit of the Z(T-G) element is $(dT\text{-}dG)_n$ without any other sequence, although n is variable.

Location of the Z(T-G) Element with Respect to the Coding Sequence. In the human cardiac muscle actin gene, a Z(T-G) element is located in an intron (12). Mapping of the coding sequence in λHa-314 suggests that the Z(T-G) element in λHa-314 is also located in an intron (Fig. 3*a*). However, $(dT\text{-}dG)_n$ sequence was found in the intergenic region between δ and β human globin genes (9) and in the region flanking a mouse immunoglobulin gene (10). Thus, the Z(T-G) element is probably randomly dispersed in the genome, not only in the introns but also in other regions.

The Presence of dC-dG Alternating Sequence in Eukaryotic Genomes. The presence of other purine–pyrimidine alternating sequence (dC-dG alternating sequence) in eukaryotic genomes was surveyed by Southern blot analysis using ^{32}P-labeled poly(dC-dG) as the probe. $(dC\text{-}dG)_n$ sequence was detected in the human, mouse, and salmon genomes (Fig. 5). Based on comparison of the specific activity of the probe and the exposure time used in Fig. 1 with those in Fig. 5, it was estimated that $(dC\text{-}dG)_n$ is not as highly repeated as the Z(T-G) element but is moderately repeated in these genomes. The repeated element containing $(dC\text{-}dG)_n$ sequence is designated the Z(C-G) element, although its structural unit was not determined.

The "smear" hybridization pattern observed in *Eco*RI-digested mouse DNA suggests random dispersion of the Z(C-G) element in the mouse genome. However, a discrete hybridized band is seen in the *Eco*RI-digested human DNA, suggesting sequence homology of the region flanking $(dC\text{-}dG)_n$ or clustering of the Z(C-G) element in the human genome.

On the other hand, no significant copies of the Z(C-G) element were detected in the calf genome and only a single weak band was detected in the yeast genome.

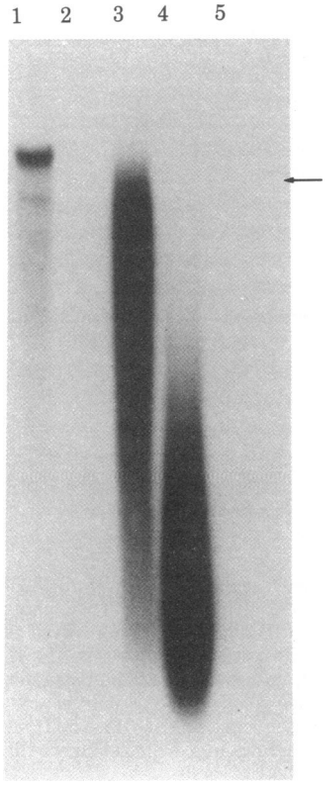

FIG. 5. Survey of dC-dG alternating sequences in various eukaryotic genomes. Ten micrograms of human (lane 1), calf (lane 2), mouse (lane 3), salmon (lane 4), or yeast (lane 5), DNA was completely digested with *Eco*RI, and these digests were electrophoresed, blotted, and hybridized to ^{32}P-labeled poly(dG-dC)·poly(dG-dC). Autoradiography was carried out for 6 days. It should be noted that this filter was exposed for a much longer time than that in Fig. 1, though the same amount of probe with approximately the same specific activity was used in both experiments. The arrow on the right indicates a hybridized band in lane 5.

DISCUSSION

It is accepted that poly(dG-dC) adopts a left-handed helical conformation, called Z-DNA, under certain condition *in vitro*. On the other hand, the ability of poly(dT-dG)·poly(dC-dA) to adopt the Z conformation has been questioned until recently, although such a conformation has been observed in solid fibers of this duplex (2). However, two recent studies (5, 6) have found that this duplex also shows inversion of the dichroism spectrum in solution, which is a characteristic of the Z form seen in poly(dG-dC).

A natural question arises whether such purine–pyrimidine alternating sequences can really adopt the Z form *in vivo*. Although there is no definite answer at this time, at least in case of the Z(C-G) element, the adoption to Z form *in vivo* has been suggested by circumstantial evidence such as that poly(dG-dC) in which the cytosine residues are methylated (23) or modified by carcinogen treatment (24) can adopt the Z form even under physiological conditions and that cytosine is the major site of methylation *in vivo* in many eukaryotes (25). There may be similar forces that stabilize the Z-form of Z(T-G) elements *in vivo*.

The results presented here indicate that DNA sequences that have the potential to adopt the Z or a Z-like form, stretches of dT-dG alternating sequence designated the Z(T-G) element and stretches of dC-dG alternating sequence designated the Z(C-G) element, are scattered throughout a diverse evolutionary spectrum of eukaryotic genomes. It is not certain that all genomic DNA fragments that hybridized to the probe used here contain the precisely alternating sequence. It is nevertheless likely that most of hybridized DNAs contain considerable lengths of alternating sequence. First, we have surveyed the Z(T-G) element in seven different human actin genes. Two clones (λHa-25 and λHa-314) that were positive by the hybridization proved to contain $(T\text{-}G)_{15}$ or $(T\text{-}G)_{25}$ by DNA sequence analysis (Figs. 3 and 4; ref. 8). Second, the hybridization condition was stringent and the probe seemed to be very specific in that only $(T\text{-}G)_n$-containing restriction fragments hybridized among the restriction fragments of λHa-25 and λHa-314 (Figs. 1 and 3*a*). Furthermore, the probe did not hybridize to pBR322 DNA (Fig. 1), which contains $(T\text{-}G)_3$, indicating that sequence homology with considerable base length is necessary for the hybridization under the conditions used.

Since *E. coli* DNA was used here as carrier DNA during hybridization and did not compete with the ^{32}P-labeled probe, both elements seem to be absent in this prokaryotic genome. The finding by others that antibody against Z-DNA did not react with *E. coli* DNA (26) is compatible with our data. It is obvious that previous failures to detect the Z(T-G) or the Z(C-G) element in eukaryotic DNA can be ascribed to the use of eukaryotic DNA such as salmon sperm or calf thymus DNA as the carrier during hybridization.

A number of highly or moderately repeated DNA elements have been identified in various eukaryotic genomes (27, 29). Their biological functions, however, are still unknown, though some possible functions have been proposed (30). The Z(T-G) and Z(C-G) elements we have described here are an especially interesting class of repeated DNA element because of their potential capability to adopt the Z form. In this light, their high evolutionary conservation (at least in the case of the Z(T-G) element) may reflect an essential function for Z-DNA sequence. At present, it is not certain that the Z(T-G) and Z(C-G) elements studied here are the only major Z-DNA sequence in eukaryotic genomes. There could be other sequences with ability to form Z-DNA *in vivo*. Other types of simple repetitive sequence such as $(T\text{-}C)_n$ and $(G\text{-}A\text{-}T\text{-}A)_n$ have been reported in cloned genes (31, 32). Furthermore, a high copy number of poly(dG)·poly(dC) sequence was detected by hybridization in the calf and chicken genomes but many fewer in human and mouse genome (unpublished data). Such simple repetitive sequences could also have an unique DNA conformation other than Z form.

In any case, the abundant occurrence and evolutionary conservation of Z(T-G) and Z(C-G) elements could have important biological implications: they could be involved in the regulation of gene expression, they could be hot spots for gene recombination or rearrangement, or they could be especially reactive with chemical reagents such as mutagens and carcinogens. The most interesting implication, however, is that the Z(T-G) and Z(C-G) elements may be involved in the regulation of gene expression, especially at transcriptional level; e.g., if these sequences can reversibly interconvert between the B and the Z form *in vivo*, such interconversion would change the distortion of DNA at a proximal or distal site and result in activation or inactivation of the gene. If Z-DNA-forming sequences are involved in such a fundamental function, the Z(T-G) element may be the major functional Z-DNA sequence, since the Z(T-G) element is highly repeated and conserved throughout eukaryotic genomes while the Z(C-G) element is moderately repeated and conserved only in limited eukaryotic genomes.

We thank Drs. Che-Hung Lee and Dezider Grunberger for their valuable suggestions, Dr. Francine Eden for her critical review and advice on the manuscript, and Dr. Tom Maniatis for a human DNA library. H.H. is grateful to Dr. M. E. Digan for *Drosophila* DNA, Dr. H. U. Affolter for *Xenopus* DNA, and Drs. P. Hensley and E. Eisenstein for yeast. Thanks are also due to Ms. Linda Nischan for typing the manuscript.

1. Wang, A. H. J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., Van Boom, J. H., Van der Marel, G. & Rich, A. (1979) *Nature (London)* **282**, 680–686.
2. Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. & Ratliff, R. L. (1980) *Nature (London)* **283**, 743–745.
3. Drew, H., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980) *Nature (London)* **286**, 567–573.
4. Crawford, J. L., Kolpak, F. J., Wang, A. H. J., Quigley, G. J., Van der Marel, G. & Rich, A. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4016–4020.
5. Vorlickova, M., Kypr, J., Stokrova, S. & Sponar, J. (1982) *Nucleic Acids Res.* **10**, 1071–1080.
6. Zimmer, C., Tymen, S., Marck, C. & Guschlbauer, W. (1982) *Nucleic Acids Res.* **10**, 1081–1091.
7. Nordheim, A., Pardue, M. L., Lafer, E. M., Moller, A., Stollar, B. D. & Rich, A. (1981) *Nature (London)* **294**, 417–422.
8. Hamada, H. & Kakunaga, T. (1982) *Nature (London)* **298**, 396–398.
9. Miesfeld, R., Krystal, M. & Arnheim, N. (1980) *Nucleic Acids Res.* **9**, 5931–5937.
10. Nishioka, Y. & Leder, P. (1980) *J. Biol. Chem.* **255**, 3691–3694.
11. Hamada, H., Muramatsu, M., Urano, Y., Onishi, T. & Kominami, R. (1979) *Cell* **17**, 163–173.
12. Hamada, H., Petrino, M. G. & Kakunaga, T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5901–5905.
13. McKeown, M. E. & Firtel, R. A. (1981) *J. Mol. Biol.* **151**, 593–606.
14. Kakunaga, T. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1334–1338.
15. Southern, J. (1975) *J. Mol. Biol.* **38**, 503–517.
16. Kafatos, F. C., Jones, C. W. & Efstratiadis, A. (1980) *Nucleic Acids Res.* **7**, 1541–1552.
17. Hamada, H., Leavitt, J. & Kakunaga, T. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3634–3438.
18. Fasman, G. D. (1976) *Handbook of Biochemistry and Molecular Biology* (CRC Press).
19. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1980) *Cell* **19**, 959–972.
20. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
21. Macaya, G., Cortadas, J. & Bernardi, G. (1978) *Eur. J. Biochem.* **57**, 55–68.
22. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
23. Behe, M. & Felsenfeld, G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1619–1623.
24. Santella, R. M., Grunberger, D., Weinstein, I. B. & Rich, A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1451–1455.
25. Vanyyushin, B. F., Tkachera, S. G. & Belozersky, A. N. (1970) *Nature (London)* **225**, 948–950.
26. Lafer, E., Moller, A., Nordheim, A., Stollar, B. D. & Rich, A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3546–3550.
27. Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T. & Schmid, C. W. (1980) *Nature (London)* **284**, 372–374.
28. Adams, J. W., Kaufman, R. E., Kretschmer, P. J., Harrison, M. & Nienhuis, A. W. (1980) *Nucleic Acids Res.* **8**, 6113–6128.
29. Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayev, A. A. & Georgiev, G. P. (1980) *Nucleic Acids Res.* **8**, 1201–1215.
30. Jelinek, W. R., Toomey, R. P., Leinwand, L., Kuncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L. & Schmid, C. W. (1980) *Proc. Natl. Acad. Sci. USA* **76**, 5095–5099.
31. Sures, I., Lowry, J. & Kedes, L. H. (1978) *Cell* **15**, 1033–1044.
32. Epplen, J. T., McCarrey, J. R., Sutou, S. & Ohno, S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3798–3802.