Enrichment of oligo(dG)·oligo(dC)-containing fragments from human genomic DNA by Mg^{2+}-dependent triplex affinity capture

Naoko Nishikawa, Naotoshi Kanda[1], Michio Oishi[+] and Ryoiti Kiyama*

Institute of Molecular and Cellular Biosciences, University of Tokyo, Yayoi, Bunkyo-ku, Tokyo 113, Japan and [1]Department of Veterinary Anatomy, School of Veterinary Medicine, Tokyo University of Agriculture and Technology, Fuchu-city, Tokyo 183, Japan

ABSTRACT

Oligo(dG)·oligo(dC)- or short poly(dG)·poly(dC)-containing fragments were enriched and cloned by means of Mg^{2+}-dependent triplex affinity capture and subsequent cloning procedures. A library constructed after three cycles of enrichment showed that ~80% of the clones in the supercoiled form formed a complex with labeled oligonucleotide (dG)$_{34}$. However, while the rest of the clones retained the ability to form a complex (type I clones), 90.9% failed to form a complex when they were linearized. This group of DNA was abundant in the genomic DNA, although it showed only ~3-fold enrichment by one cycle of affinity capture. This group was further classified into two species (types II and III) based on complex formation ability after phenol extraction. Type II clones retained the complex formation ability after treatment, while the human telomere [(TTAGGG)$_n$] and telomere-like [(TGGAA)$_n$] or [(TGGAG)$_n$] sequences belonging to type III clones did not. Serial deletion experiments and the binding assays using oligonucleotides confirmed that the repetitive units containing T(G)$_n$T (n= 3–5) tracts or (G)$_n$-motifs (n $\geq$ 3) were the sites of complex formation for type II and III clones. On the other hand, type I clones contained poly(dG)·poly(dC) tracts at least 10 nt long, and DNase I-footprinting analysis indicated that these tracts were the sites of complex formation.

INTRODUCTION

A group of repetitive sequences consisting of short unit sequences, usually up to a few bases, called microsatellites appear very frequently in the genomic DNA of eukaryotes (1–4). Unlike satellite DNA, however, they usually exist in the euchromatic region and the numbers of repeats are, in most cases, 10–100. Therefore, they do not show a satellite profile on ultracentrifugation in the presence of CsCl. Generally, their frequencies in the genomic DNA are higher than expected, suggesting that several mechanisms are responsible for their active formation and stable maintenance (5–7). Furthermore, their presence in regions close to genes has been shown to influence gene expression (8–10). For example, a (CA)$_n$ repeat located in the promoter region of the rat prolactin (rPRL) gene changes the superhelical density of the region thereby inhibiting transcription (11). The recent finding of (CAG)$_n$ repeats in genes responsible for neural genetic diseases especially have focused attention on microsatellites (7,12).

Since genomic DNA of higher eukaryotes is generally A+T rich, guanine and cytosine bases are underrepresented. Therefore, unlike poly(dA)·poly(dT) and poly(dG–dA)·poly(dT–dC), poly(dG)·poly(dC) and other G+C-rich microsatellites appear infrequently. Their frequencies in human genomic DNA are 0.3% for poly(dA)·poly(dT), 0.2% for poly(dG–dA)·poly(dT–dC) or 0.0002% for poly(dG)·poly(dC), respectively (2). The unit sequences containing the dinucleotides CG appear very infrequently due to the transition of cytosine to thymine (13). As a result, TG dinucleotides appear at a higher frequency than the theoretical values.

In previous reports, we enriched two types of microsatellites, poly(dA)·poly(dT) and poly(dG–dA)·poly(dT–dC), from human genomic DNA by Mg^{2+}-dependent triplex affinity capture (14,15). More than half of the clones in these libraries contained perfect repeats of these microsatellites. Their numbers of repetitions ranged from 14 to 37 for poly(dA)·poly(dT) or from 15 to 42 for poly(dG–dA)·poly(dT–dC), indicating that the complexes formed by these repeated sequences are relatively stable. However, we observed polymorphism associated with the repeat length between individuals and, in some cases, between normal and cancer tissues from the same individual (15). Such instability of microsatellites has been reported for colon cancers and other genetic diseases caused by deficiencies in mismatch repair genes (16,17).

While increasing numbers of microsatellites have been utilized for genetic and genome analysis (6,18), it has been laborious to enrich fragments containing these microsatellites. Affinity capture of duplex DNA by triplex DNA formation with biotin-labeled oligonucleotides as the third strand was first reported by Ito *et al.* (19). We and others have modified this method for broader use by introducing Mg^{2+}-dependent association and dissociation of the triplex (14,15,20). While C$^+$GC triplet bases are stabilized at a low pH without metal ions, several triplexes, poly(dT)·poly(dA)·poly(dT), poly(dG–dA)·poly(dG–dA)·poly(dT–dC) and poly(dG)·poly(dG)·poly(dC) for example, require metal ions for

---

*To whom correspondence should be addressed. Tel: +81 3 3812 2111; Fax: +81 3 3818 9437; Email: kiyamar@hgc.ims.u-tokyo.ac.jp

+Present address: National Institute of Bioscience and Human-technology, Tsukuba-shi, Ibaraki 305, Japan
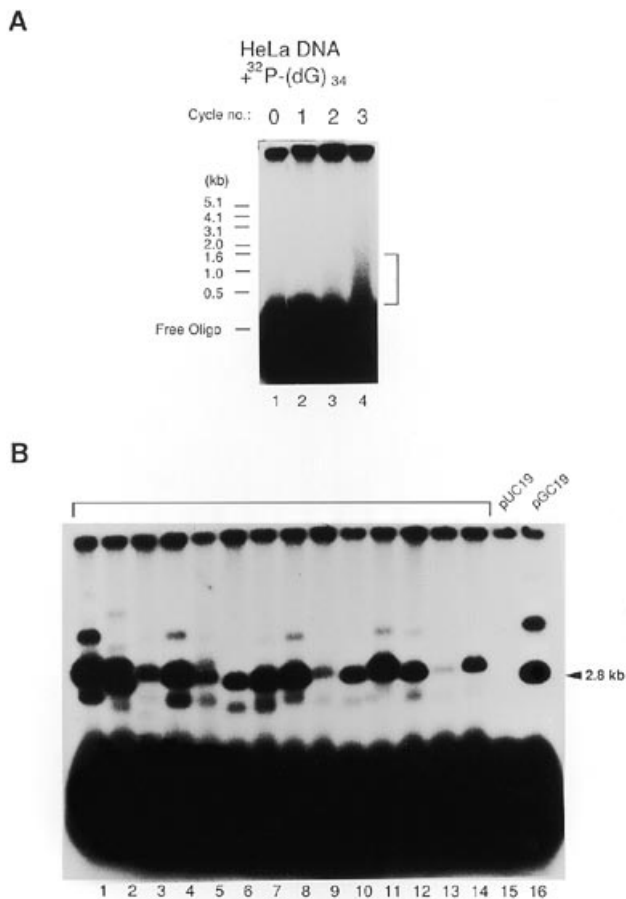
Figure 1. Construction of a library of DNA fragments containing poly(dG)·poly(dC) tracts. (**A**) Affinity enrichment of human genomic DNA fragments by triplex affinity capture. After Mg^{2+}-dependent triplex affinity capture, samples obtained at each cycle (lanes 2–4) or the initial sample (lane 1) were used for triplex DNA gel assay. (**B**) Triplex DNA gel assay with clones from the library. The library was constructed with the sample after the 3rd cycle (A, lane 4). Fourteen randomly selected clones from the library (lanes 1–14) were assayed with ^{32}P-labeled $(dG)_{34}$ as a probe. pUC19, a negative control (lane 15) and pGC19, a positive control (2.8 kb in size; lane 16) are also shown.

their formation. Triplexes with mixed bases could also be stabilized with metal ions (21–24). However, the efficiencies of the procedure for other microsatellites and the types of microsatellites that could be obtained by specific probes have not been examined in detail. Therefore, for broader use of triplex affinity capture, it has been a prerequisite to characterize the method by first using simple repeats. In this study, we explored the method using poly(dG)·poly(dC)-containing sequences by forming poly(dG)·poly(dG)·poly(dC) triplexes in the presence of Mg^{2+}.

MATERIALS AND METHODS

Materials

A pair of oligonucleotides, oligo A (GATCCGCGGCCGCCCGAT) and oligo B (ATCGGGCGGCCGCG), for the adaptor were annealed and then ligated to *Alu*I-digested human DNA. Oligo A was used as a primer for PCR. Biotinylated $(dG)_{34}$ used for Mg^{2+}-dependent triplex affinity capture was synthesized by Greiner (Japan). Streptavidin-coated magnetic beads were

purchased from Promega (USA). Plasmid pGC19 was constructed from pUC19 by inserting a $(dG)_{34}$ sequence between *Sac*I and *Eco*RI sites. The clone containing the telomere sequence was TT6 (44). The oligonucleotides used for the binding assays were CTGTTCCAGGCTGTCAGATGCTCACCTGGGGGGTGTGGG-TG and $(TGGGGGTGTGGGTG)_2TGGGGGT$ from pHGC32, and $(GGAGT)_6GGAG$ from pHGC39, which were annealed with their complementary oligonucleotides.

Mg^{2+}-dependent triplex affinity capture

This procedure was performed as described previously (15), except that oligonucleotide $(dG)_{34}$ was used as the third strand. Approximately 1 µg of plasmid DNA or genomic DNA was mixed with 200 nM biotinylated $(dG)_{34}$ in 25 µl of triplex buffer (10 mM Tris–HCl, pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl) and incubated at 37°C for 30 min. The sample was adsorbed onto 0.3 mg of streptavidin-coated magnetic beads. After incubation with 0.5 ml of the triplex buffer at room temperature for 30 min, the beads were separated with magnets. After extensive washing with the triplex buffer (seven times with 0.5 ml: FT fractions), the bound DNA was eluted with 2.5 ml of the elution buffer (10 mM Tris–HCl, pH 8.0, 5 mM EDTA, 50 mM NaCl: EDTA eluates).

Construction of the library

Chromosomal DNA from HeLa cells was digested with *Alu*I and ligated with the adaptor. After PCR amplification (a total of 29 cycles with a 1/10 dilution after the 17th cycle) with the primer (oligo A), the amplified fragments were purified using a Wizard PCR Preps kit (Promega, USA). One µg of the amplified DNA was used for the first cycle of affinity capture. The DNA fragments recovered from the first cycle were amplified by PCR and used for the next cycle. After the 3rd cycle, the amplified DNA was digested with *Not*I and ligated into pBluescript SK(–). The ligation products were used for transformation of *Escherichia coli* DH5α. The clones of the library were screened by triplex DNA-gel assay for complex formation with $(dG)_{34}$. PCR was performed for appropriate cycles (a total of 38 cycles with a 1/20 dilution after the 20th cycle for the first, 31 cycles with a 1/20 dilution after the 15th cycle for the second and 30 cycles with a 1/20 dilution after the 15th cycle for the third cycle of the enrichment) consisting of 94°C for 1 min, 40°C for 2 min and 72°C for 3 min, with a final 10 min extension at 72°C.

DNase I footprinting

DNase I footprinting was performed as described previously (25). After the plasmid DNA was digested with *Hin*dIII followed by treatment with bacterial alkaline phasphatase, its 5′ ends were labeled with ^{32}P by T4 polynucleotide kinase. The labeled DNA was digested with *Sca*I and the 293 bp fragment was recovered from 1% low-melting agarose with Wizard PCR Preps. After aliquots of 40 µl were incubated at 37°C for 30 min in the presence of 0, 0.2, 2 or 20 µM $(dG)_{34}$ in triplex buffer (10 mM Tris–HCl, pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl), 60 µl of a solution containing 10 mM Tris–HCl (pH 8.0), 2.5 mM $CaCl_2$ and 10 mM $MgCl_2$ was added and the reaction was initiated with 2 µl of 1.4×10^{-3} U of DNase I. After incubation for 2 min at room temperature, the samples were precipitated with ethanol and electrophoresed on a 6% polyacrylamide–7 M urea gel under denaturing conditions. The G+A marker was prepared according to the Maxam–Gilbert method (26).
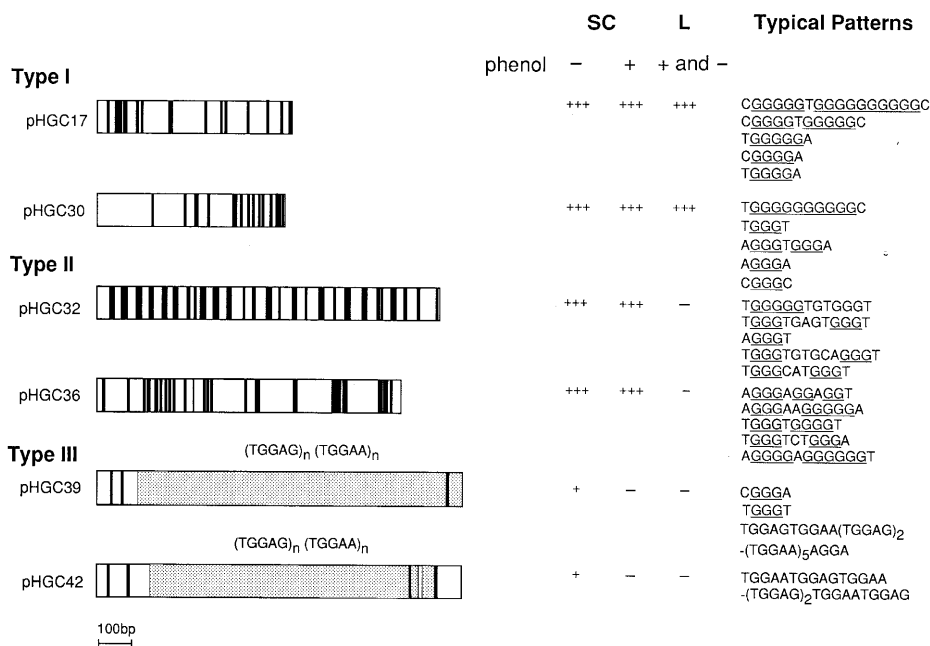
Figure 2. Classification of the clones of the library. Three types (I–III) of clones were classified by complex formation ability before or after phenol treatment in different DNA forms. The positions of $(G)_n$ ($n \geq 3$) are indicated by vertical bars. The regions of satellite DNA are shadowed. The relative abilities of triplex formation of each clone in the supercoiled (SC) or linear (L) DNA before and after phenol treatment are shown by the numbers of '+', and typical sequences containing $(G)_n$-motifs (underlined) are shown on the right. The accession numbers of the clones in the DDBJ/GenBank/EMBL databases are: D88106 for pHGC17, D88107 for pHGC30, D88108 for pHGC32, D88109 for pHGC36, D88110 for pHGC39, and D88111 and D88112 for pHGC42.

Triplex DNA gel assay

500 ng of plasmid DNA was incubated with 100 nM ^{32}P-labeled $(dG)_{34}$ in triplex buffer for 30 min at 37°C (27). Samples were electrophoresed on a 1% agarose gel at 1 V/cm for 17 h at 4°C. The gel was fixed with 10% (w/v) trichloroacetic acid, dried and autoradiographed.

Triplex DNA assay using oligonucleotides

One nM of labeled-duplex oligonucleotides derived from type II clone pHGC39 and type III clone pHGC32 were incubated with various concentrations of $(dG)_{34}$ in triplex buffer for 30 min at 37°C. The duplex DNA and the complex formed with $(dG)_{34}$ were resolved by 10% polyacrylamide gels. Duplex oligonucleotides were labeled with T4 polynucleotide kinase.

Construction of the deletion mutants

Each clone, pHGC32 or pHGC39, was digested with *Pst*I and *Xba*I. After ethanol precipitation, the samples were incubated with 1.8 U/µl of exonuclease III (Takara, Kyoto) in 50 mM Tris–HCl (pH 8.0), 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol at 37°C. Aliquots of 10 µl were removed from the reaction mixture at 1 min intervals up to 10 min and, after phenol extraction, 0.5 U/µl mungbean nuclease (Takara) was added to the reaction mixture containing 10 mM Tris–HCl (pH 8.0), 10 mM $MgCl_2$, 50 mM NaCl, 1 mM dithiothreitol, 1 mM $ZnSO_4$ and the mixture was incubated for 1 h at 37°C. After phenol extraction and ethanol precipitation, the samples were incubated with 0.04 U/µl of Klenow fragment (New England Biolabs) in 10 mM Tris–HCl (pH 8.0), 5 mM $MgCl_2$, 0.5 mM dithiothreitol, 1.25 mM dNTPs for 30 min at 37°C. They were then self-ligated with T4 DNA ligase (New England Biolabs) and used for transformation.

RESULTS

Enrichment of poly(dG)·poly(dC)-containing fragments by Mg^{2+}-dependent triplex affinity capture

Figure 1 shows a summary of enrichment of poly(dG)·poly (dC)-containing fragments from human genomic DNA by Mg^{2+}-dependent triplex affinity capture. In a separate experiment using pUC19 and pGC19, a pUC19 derivative containing a $(dG)_{34}$·$(dC)_{34}$ tract, ~30-fold enrichment was achieved by a single cycle of the procedure (data not shown). This method was applied for multiple cycles to human genomic DNA. Figure 1A shows the results of a gel assay for triplex formation using ^{32}P-labeled $(dG)_{34}$. After three cycles of the procedure, fragments of up to 2–3 kb in length, but mainly <1 kb, were amplified (Fig. 1A, lane 4). Since there was little enrichment after the third cycle (data not shown), we constructed a library with the products after the third cycle. The low intensity of the signals was presumably due to the inefficient complex formation of certain sequences under the conditions of affinity capture (discussed below).

In the library, 33 of 42 clones examined (78.6%, data not shown) including 14 randomly selected clones shown in Figure 1B (lanes 1–14) formed complexes with ^{32}P-labeled $(dG)_{34}$ when in the form of supercoiled DNA. Since none of 14 clones randomly selected from the fragments before enrichment (cloned from the sample used in Fig. 1A, lane 1) showed complex formation ability (data not shown), the procedure indeed enriched the fragments of interest. Table 1 summarizes the clones which formed a complex with $(dG)_{34}$ and whose sequences were determined. Most of these clones contained multiple stretches of $(G)_n$ ($n \geq 3$) tracts, while others were derived from satellite DNA containing the human telomere-like sequences. When these clones were linearized, 30 of 33 clones examined (90.9%) failed to form a complex with the oligonucleotide (details below).

Table 1. Summary of the clones that showed triplex formation

Clone no.	Insert length (bp)	Triplex formation[a]	Category[b]	Pattern of repeats[c]
2	550	+	II	A**GGGG**CGT**GGG**C C**GGGG**TGTG A**GGGG**ACC**GGGG**AG
7	150	+	III	$(TCCAC)_n(TGGAG)_m$
11	1050	++	II	C**GGGGG**C**GGGG**AT**GGG**T T**GGGG**C**GGGG**A AGC**GGG**A**GGGGG**A
14	550	+	II	T**GGGG**T**GGG**C A**GGGG**A A**GGGG**C
15	950	+++	II	A**GGGGGGGGGGG**A A**GGG**CA**GGGG**A C**GGGG**A
17	598	+++	I	C**GGGG**T**GGGGGGGGGG**C C**GGGGG**C T**GGGGG**A
19	1000	++	II	A(**GGG**CC)₃**GGGGGG**A C**GGG**A
30	581	+++	I	T**GGGGGGGGGGG**C T**GGG**C A**GGG**T**GGG**A
32	1048	+++	II	T**GGGGG**TGT**GGG**T T**GGGG**TGCA**GGG**T T**GGGGG**TGTGT**GGG**T
34	750	++	III	T**GGGGGG**A T**GGGGG**C
36	849	+++	II	A**GGGG**A**GGGGGG**T T**GGGG**T**GGG**T A**GGG**TCTGT**GGGG**A
37	1050	+	III	$(TGGAA)_n(TGGAG)_m$
39	1020	++	III	$(TGGAA)_n(TGGAG)_m$
40	900	+	III	A**GGGGG**T A**GGGGG**C
42	1050	+	III	$(TGGAA)_n(TGGAG)_m$

[a]Relative intensity of triplex formation detected by triplex DNA gel assay with the cloned DNA in the supercoiled form.
[b]Classified as types I, II or III (see text).
[c]Typical sequences containing poly(dG) tracts. The repeats of $(G)_n$ ($n \geq 3$) are underlined.

Characterization of the cloned DNA fragments

The nucleotide sequence data as well as the gel assay indicated that (i) most of the clones contained relatively short stretches of poly(dG) tracts, (ii) some were apparently derived from satellite DNA, containing multiple repeats of short (mostly 5 bp) sequences, (iii) there was no apparent correlation between the intensity of the signals and the length of the fragments, and (iv) some lost the signal when they were linearized. These observations indicated that the clones in the library could be classified into several types. As summarized in Figure 2, they were tentatively classified into three types according to complex formation ability under various conditions. The first type (type I; pHGC17 and pHGC30, for example) of clones formed complexes with labeled $(dG)_{34}$ in both supercoiled and linear forms, and these clones contained relatively long stretches of poly(dG) tracts. The second type (type II; pHGC32 and pHGC36) contained short, usually 3 bp, stretches of poly(dG) tracts which were sandwiched by Ts. These clones lost the signal when they were linearized. The third type (type III; pHGC39 and pHGC42) were derived from satellite DNA and contained multiple repeats of $(TGGAG)_n$ and/or

$(TGGAA)_n$. The clones belonging to this type also lost the signal after linearization (Fig. 2). Types II and III were separated according to the differences after phenol treatment (details below). The frequencies of each type in the library were 8.3% (for type I), 25.0% (for type II) and 66.7% (for type III).

Figure 3 shows the results of triplex DNA gel assay of the clones from the library. Complex formation was assayed before and after *Pvu*II digestion (Fig. 3A). Among eight clones examined, five (pHGC5, pHGC15, pHGC22, pHGC32 and pHGC36) did not show signals in the linear form, while the remaining three (pHGC9, pHGC17 and pHGC30) did (shown by arrows in Fig. 3A). The latter belonged to type I while the former belonged to types II or III. The effects of phenol treatment were examined and the results are shown in Figure 3B. All of the four clones (pHGC11, pHGC34, pHGC40 and pHGC42) tested lost the signal when they were treated with phenol. Therefore, these clones were classified as type III. These clones also lost the signal when they were linearized (see the *Eco*RI digests in Fig. 3B). A clone from human telomere DNA, containing multiple repeats of (TTAGGG) tract, showed the same results (Fig. 3B, lanes 1–3).
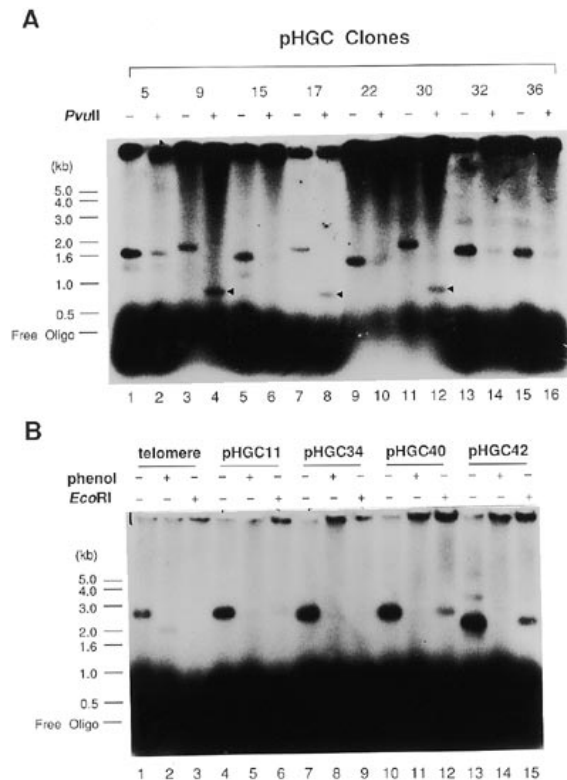
Figure 3. Triplex DNA gel assay with clones in the form of linear or supercoiled DNA or after phenol treatment. (**A**) Effects of linearization. pHGC DNA digested with *Pvu*II or uncut DNA was subjected to triplex DNA gel assay with 100 µM ^{32}P-labeled (dG)$_{34}$. The signals detected in the linear DNA (*Pvu*II fragments; lanes 4, 8 and 12) were indicated by arrowheads. The sizes of clones were 4.0 kb, pHGC5; 3.6 kb, pHGC9; 3.9 kb, pHGC15; 3.6 kb, pHGC17; 3.7 kb, pHGC22; 3.5 kb, pHGC30; 4.0 kb, pHGC32; 3.8 kb, pHGC36. (**B**) Effects of phenol treatment. The plasmid DNA before (lanes 1, 4, 7, 10 and 13) or after (lanes 2, 5, 8 and 11) phenol treatment were used for triplex DNA gel assay. Aliquots without phenol treatment were linearized by *Eco*RI and used for the assay (lanes 3, 6, 9 and 12). The sizes of clones were 3.3 kb, telomere clone; 2.8 kb, pHGC11; 3.7 kb, pHGC34; 3.9 kb, pHGC40; 4.0 kb, pHGC42.

To explain the enrichment of the fragments that did not contain a long stretch of poly(dG)·poly(dC), we examined the degrees of enrichment of plasmids pHGC17 and pHGC32 from types I or II, respectively, and pH8, which showed no complex formation, as a control, from the mixture with pUC19 (Table 2). The clone pHGC17 was enriched to ~16.6-fold after one cycle of affinity capture. In contrast, pHGC32 was enriched to ~3-fold even in the form of supercoiled DNA. No enrichment was observed for the control clone pH 8. This suggests that type II and III clones represented by pHGC32 comprised a large portion (90.9%) of the library because, although they were enriched inefficiently, they were abundant in the genomic DNA. On the other hand, type I clones represented by pHGC17 appeared infrequently in the genomic DNA, although they can be enriched efficiently.

Complex formation of poly(dG)·poly(dC)-containing sequences

The region of complex formation was determined for the clone pHGC17 by DNase I footprinting assay. As shown in Figure 4, a protection of the poly(dG)·poly(dC) tract and its flanking region was observed with as low as 0.2 µM (dG)$_{34}$. Thus, this poly(dG)·poly(dC) tract is indeed the site of complex formation.
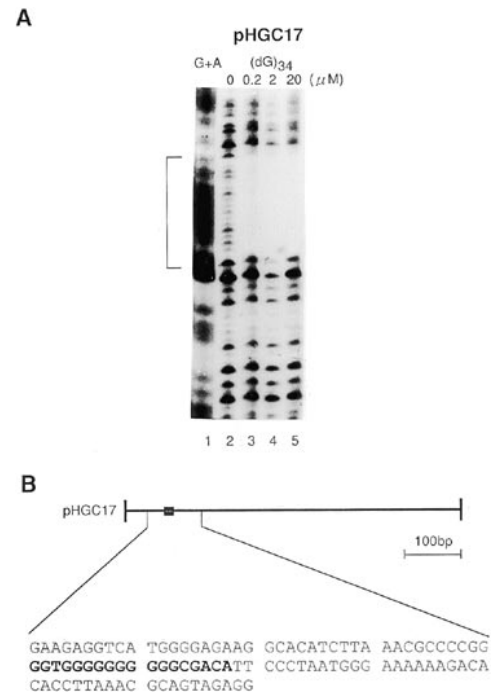
Figure 4. DNase I footprinting analysis of the clone pHGC17. (**A**) DNase I footprinting assay. pHGC17 containing a (G)$_{10}$ tract was digested with DNase I in the presence of 0 µM (lane 2), 0.2 µM (lane 3), 2.0 µM (lane 4) or 20 µM (lane 5) (dG)$_{34}$. The samples along with the sequence ladder G+A (lane 1) were electrophoresed on a 6% polyacrylamide gel under denaturing conditions. (**B**) The nucleotide sequence of the protected region is shown in bold characters.

On the other hand, no protection pattern was seen with type II or III clones with (dG)$_{34}$ at concentrations as high as 20 µM (data not shown), suggesting either that the dissociation constants of these complexes are higher than 20 µM or that DNase I footprinting somehow cannot detect the complex formation of these clones. Footprinting assays with dimethylsulfate were also unsuccessful (data not shown).

Table 2. Enrichment of clones by affinity capture

	Number of colonies[a]		Degree of enrichment[b]		Recovery
	Blue	White	%	Ratio	%
(A) pHGC17					
Original mixture	4.8×10^7	2.4×10^6	4.8	1.0	100
FT fraction	3.4×10^7	1.7×10^6	4.8	1.0	0.8
EDTA eluate	6.3×10^4	3.0×10^5	83.0	16.6	0.7
(B) pHGC32					
Original mixture	5.3×10^7	4.7×10^6	8.0	1.0	100
FT fraction	1.4×10^7	2.6×10^6	8.6	1.0	28.8
EDTA eluate	4.9×10^4	1.2×10^4	20.0	2.9	0.1
(C) pH8[c]					
Original mixture	1.4×10^8	5.7×10^6	4.0	1.0	100
FT fraction	1.0×10^8	2.9×10^6	3.0	0.8	70.6
EDTA eluate	6.6×10^5	3.0×10^3	0.4	0.1	0.4

[a]Approximately 1 µg of DNA mixture containing each plasmid DNA and pUC19 DNA (~1:20 ratio) was used as a source of affinity capture (original mixture). Numbers of blue (pUC19) or white (pUC19 derivatives) colonies were calculated from the averages of at least two plates.
[b]The degrees of enrichment were calculated as the numbers of white colonies divided by the total number of colonies (percentage). The relative degrees of enrichment of pUC19-derivatives relative to the original mixture are shown as the 'Ratio'.
[c]Clone pH8 contained an 800 bp fragment that did not form a complex with (dG)$_{34}$.
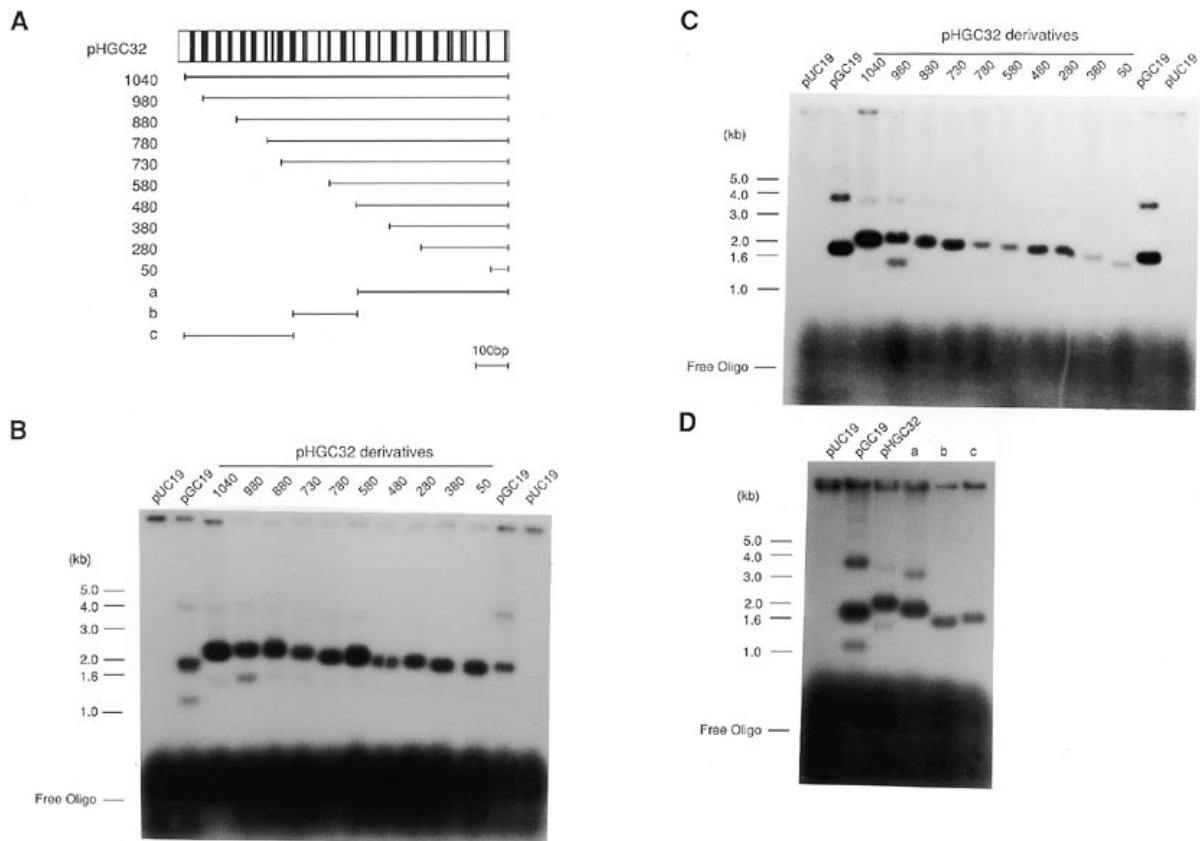
Figure 5. Triplex DNA gel assay with deletion mutants derived from pHGC32. (**A**) Map of pHGC32 and deletion mutants containing inserts of various lengths. The lengths of the inserts (bp) are shown on the left. The lengths of the cloned fragments a, b and c are 480, 210 and 350 bp, respectively. The plasmid containing the fragment a has additional 400 bp vector sequence. (**B**) Results of gel assay with the plasmid DNA before phenol treatment. The numbers shown above the figure correspond to the lengths of the insert. (**C**) Results of gel assay with the plasmid DNA after phenol treatment. (**D**) Results of gel assay with clones containing parts (a–c) of pHGC32 DNA. Plasmid DNA was assayed without phenol treatment.

Since complex formation was detected by the gel assay (Fig. 1), we used this assay for the clones pHGC32 and pHGC39 and their deletion constructs to examine the regions of complex formation (Figs 5 and 6). pHGC32 was a type II clone and contained multiple repeats of a 40 bp consensus sequence containing $(G)_n$-motifs ($n \geq 3$) (shown by vertical lines in Fig. 5A). When pHGC32 and its derivatives containing 50–1040 bp of its sequence were assayed, all of the clones exhibited signals with roughly identical intensities (Fig. 5B). However, when they were treated with phenol, their intensities became dependent on their lengths (Fig. 5C). As this was a type II clone, phenol treatment did not completely abolish the signal. When this plasmid was divided into three parts (a–c), each exhibited the signal (Fig. 5D). Therefore, the repeating unit was likely to be responsible for complex formation.

The same assay was performed with pHGC39, a type III clone containing satellite DNA with TGGAG and TGGAA motifs, and its deletion constructs (Fig. 6A). In contrast to pHGC32, the intensities of the signals were decreased as their lengths shortened (Fig. 6B). The clone containing 150 bp of the sequence showed a very low signal (~1/20) compared with that with 950 bp of the repeat region. This indicates that the repeating unit is the site of complex formation.

The affinity of $(dG)_{34}$ to pHGC32 and pHGC39 was confirmed by the binding assay with duplex oligonucleotides containing the repeat units (Fig. 7). Complex formation was observed for pHGC32 in the presence of 100 µM $(dG)_{34}$ (Fig. 7A, lane 6). Repeats of the

G-rich sequence from the 40 bp unit exhibited shifted bands both at 10 and 100 µM (Fig. 7A, lanes 11 and 12), indicating that this region was indeed the site of complex formation. A similar result was obtained with the clone pHGC39 (Fig. 7B, lane 6). Therefore, although the affinity is low, the interrupted short $(G)_n$ motifs can form a complex with the probe $(dG)_{34}$.

To examine the chromosomal localization of the cloned fragments in pHGC32 and pHGC39, we performed the FISH analysis (45) with biotin-labeled oligonucleotides containing the unit sequences (data not shown). The 5 bp repeat from pHGC39 was hybridized to the centromeric region of several chromosomes including chromosomes 13, 14, 15, 21 and 22, and the Y chromosome, while the 40 bp repeat from pHGC32 was hybridized to chromosome 1p36.3.

DISCUSSION

Enrichment of poly(dG)·poly(dC)-containing fragments

We reported previously the enrichment of microsatellites poly(dA)·poly(dT) and poly(dG–dA)·poly(dT–dC) from human genomic DNA by Mg^{2+}-dependent triplex affinity capture (14,15). In both cases, the libraries exhibited substantial enrichment of these microsatellites whose lengths were 14–84 bp. In contrast, the library of poly(dG)·poly(dC)-containing fragments constructed after three cycles of affinity enrichment consisted of two groups
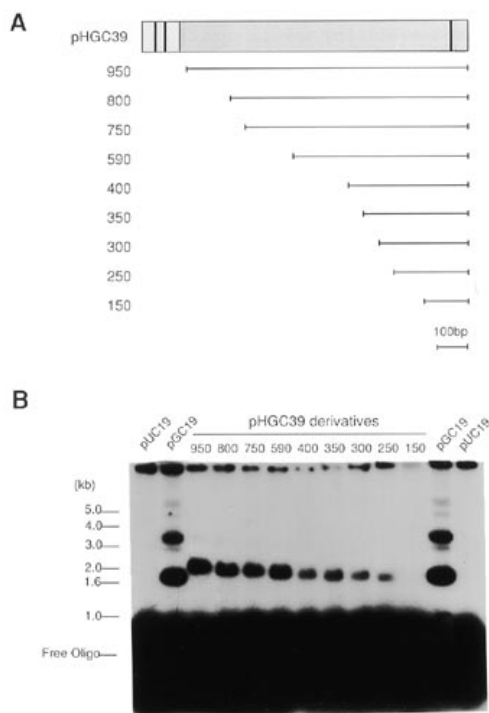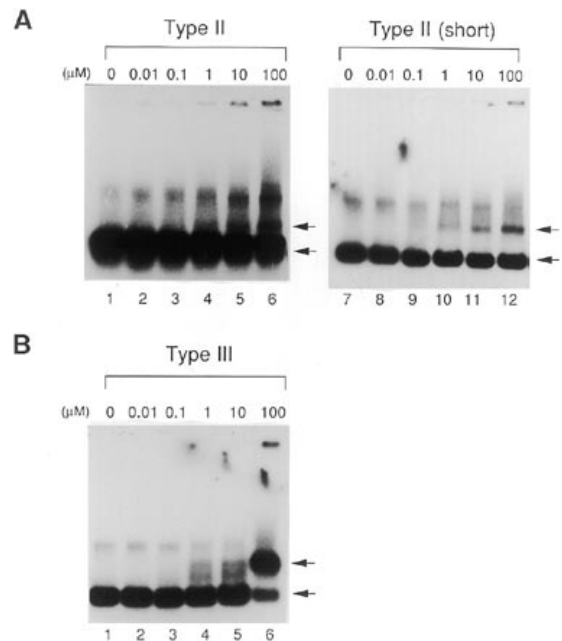
Figure 6. Triplex DNA gel assay with deletion mutants derived from pHGC39. (**A**) Map of pHGC39 and deletion mutants containing (TGGAG)$_n$ and (TGGAA)$_n$ tracts of various lengths. The lengths of the inserts (bp) are shown on the left. (**B**) Results of gel assay. The numbers correspond to the lengths of the insert.

Figure 7. The binding assays using oligonucleotides. (**A**) The assay with the duplex oligonucleotides containing the 40 bp repeats from pHGC32 (lanes 1–6) or with the 35 bp sequence containing 2.5 repeats of the G-rich region from pHGC32 (lanes 7–12). The ^{32}P-labeled duplex oligonucleotides (1 nM) were mixed with indicated amounts of (dG)$_{34}$ and after complex formation they were electrophoresed on 10% polyacrylamide gels. Arrows indicated the positions of the complex formed (upper) and the duplex oligonucleotides (lower). (**B**) The assay with the 34 bp oligonucleotides containing the 5 bp consensus repeat unit from pHGC39 (lanes 1–6).

of clones: one containing perfect poly(dG)·poly(dC) sequences (type I clones) and another containing either the consensus T(G)$_n$T ($n = 3$–5) or telomere-like sequences. The latter group was further classified into two types (types II and III) according to the complex formation ability after phenol treatment. Although the library contained 8.3, 25.0 and 66.7% type I, II and III clones, respectively, the degrees of affinity enrichment varied among these sequences. Affinity enrichment using two clones from each group indicated that those with a perfect poly(dG)·poly(dC) sequence (type I) showed 16.6-fold enrichment, while those with T(G)$_n$T (types II and III) showed only 3-fold enrichment. Therefore, we assumed that sequences with T(G)$_n$T are fairly abundant in human genomic DNA although they could have been lost during affinity enrichment because of their low affinity to (dG)$_{34}$. Although we used 34-base long oligonucleotides for affinity enrichment, it might be possible to increase the stringency for perfect poly(dG)·poly(dC) sequences by raising the temperature at the affinity capture or by shortening the length of the probe.

Complex formation of enriched fragments

DNase I footprinting experiments (Fig. 4) revealed that the poly(dG)·poly(dC) tract was the site of complex formation between the type I clones and ^{32}P-labeled (dG)$_{34}$. However, we could not determine the site of interaction between T(G)$_n$T-containing fragments and the labeled probe by this approach (data not shown). This was probably because the interaction between them was weak. However, since the specific sequence feature appearing in both pHGC32 and pHGC36 was T(G)$_n$T, this tract was most likely to be the site of interaction. Especially, the sequence was repeated

throughout the cloned pHGC32 fragment which consisted of multiple copies of a 40 bp consensus, the repetition of this sequence could stabilize the complex in the gel assay. This was also observed for the other clone pHGC36, where T(G)$_n$T appeared ~30 times. It has been suggested that DNA fragments with interrupted (G)$_n$-motifs still can interact with single-stranded poly(dG) (28). Thus, the repetition of sequences containing such motifs could also stabilize the complex. A large excess of (dG)$_{34}$ (at least 10–100 μM) was needed to see the complex formation with the repeat units containing these motifs (Fig. 7). However, note that there should be some interaction between the type II or III fragments and the probe (dG)$_{34}$ even at 200 nM probe concentration.

A structural transition caused by phenol treatment was reported for A+T-rich large inverted repeats (29). This transition occurs through partial melting of duplex DNA facilitated by hydrophobic interaction which could be energetically favored by releasing the torsional stress caused by superhelicity of supercoiled DNA. This transition should be completed by forming a cruciform or hairpin structure when there is an (AT)$_n$ repeat or an A+T-rich inverted repeat in the cloned fragment, or by D-loop formation in the presence of interacting oligonucleotides. Complex formation with (dG)$_{34}$ would, therefore, release torsional stress, and could be stabilized extensively by forming hydrogen bonds between the G-rich sequences and (dG)$_{34}$ as shown here.

Biological significance of poly(dG)·poly(dC)

As G+C-rich sequences are underrepresented in human genomic DNA, most of the sequences obtained here by affinity capture did

not contain long and perfect poly(dG)·poly(dC) tracts. Most consisted of guanines and other bases and they contained sequences with short $(G)_n$-motifs and/or G-rich sequences. We tentatively classified these sequences into three types (types I–III). Since these three types of clones showed different biophysical behaviors, they may be related to different biological phenomena *in vivo*.

Apparently, some of these sequences were satellite DNAs, a group of DNA elements consisting of a unit sequence of a few to hundreds of base pairs long and repeats of a hundred times or more. One of the satellite DNAs (pHGC39) which had a homology to the telomere sequence $(TTAGGG)_n$ was localized at the centromeric heterochromatin region of several chromosomes. The other satellite DNA represented by the clones pHGC32 was localized at the telomeric region of chromosome 1. These satellite DNAs would be associated with functions that have been postulated for other satellite DNAs (5,30,31).

Poly(dG)·poly(dC)-containing sequences tend to adopt non-B-DNA conformations (21,28,32,33). This is mainly due to the hydrogen bonds formed between guanine bases in the tract and guanine or other bases in other regions (34). In duplex DNA, guanine can accommodate three hydrogen donor or acceptor sites (N_1, N_2 and O_6 positions) for base pairing, and the N_7 position of purines can participate in an additional hydrogen bond with other bases while maintaining the G·C pair. Therefore, guanine can afford multiple, stable interactions with other bases and thus the G-rich sequences tend to adopt 'unusual' DNA conformations. These unusual structures containing G-tracts have been reported in association with many biological functions (35–41). For example, telomere sequences contain $(G)_n$ motifs and they adopt unusual DNA conformations, quadruplex structures (42,43). We obtained such sequences, $(TGGAA)_n$ and $(TGGAG)_n$, by affinity capture with $(dG)_{34}$ (Table 1). Although they do not contain long stretches of poly(dG)·poly(dC), the complexes can be stabilized by hydrogen bonds between perfect or interrupted $(G)_n$ motifs located in other chromosomal regions.

ACKNOWLEDGEMENT

REFERENCES

1 Hamada,H., Petrino,M.G. and Kakunaga,T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6465–6469.
2 Tautz,D. and Renz,M. (1984) *Nucleic Acids Res* **12**, 4127–4138.
3 Tautz,D., Trick,M. and Dover,G.A. (1986) *Nature* **322**, 652–656.
4 Beckmann,J.S. and Weber,J.L. (1992) *Genomics* **12**, 627–631.
5 Vogt,P. (1990) *Hum. Genet.*, **84**, 301–336.
6 Bruford,M.W. and Wayne,R.K. (1993) *Curr. Opin. Genet. Dev.* **3**, 939–943.
7 Richards,R.I. and Sutherland,G.R. (1994) *Nature Genet.* **6**, 114–116.
8 Struhl,K. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8419–8423.
9 Ellison,M.J., Fenton,M.J., Ho,P.S. and Rich,A. (1987) *EMBO J.* **6**, 1513–1522.
10 Bhandari,R. and Brahmachari,S.K. (1995) *J. Biosci.* **20**, 613–627.
11 Naylor,L.H. and Clark,E.M. (1990) *Nucleic Acids Res.* **18**, 1595–1601.
12 Sutherland,G.R. and Richards,R.I. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3636–3641.
13 Burge,C., Campbell,A.M. and Karlin,S. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1358–1362.
14 Kiyama,R., Nishikawa,N. and Oishi,M. (1994) *J. Mol. Biol.* **237**, 193–200.
15 Nishikawa,N., Oishi,M. and Kiyama,R. (1995) *J. Biol. Chem.* **270**, 9258–9264.
16 Fishel,R., Lescoe,M.K., Rao,M.R.S., Copeland,N.G., Jenkins,N.A., Garber,J., Kane,M. and Kolodner,R. (1993) *Cell* **75**, 1027–1038.
17 Leach,F.S., Nicolaides,N.C., Papadopoulos,N., Liu,B., Jen,J., Parsons,R., Peltomaki,P., Sistonen,P., Aaltonen,L.A., Nystrom-Lahti,M., *et al.* (1993) *Cell* **75**, 1215–1225.
18 Hearne,C.M., Ghosh,S. and Todd,J.A. (1992) *Trends Genet.* **8**, 288–295.
19 Ito,T., Smith,C.L. and Cantor,C.R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 495–498.
20 Fox,K.R. (1992) *Nucleic Acids Res.* **20**, 1235–1242.
21 Kohwi,Y. and Kohwi-Shigematsu,T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3781–3785.
22 Malkov,V.A., Voloshin,O.N., Soyfer,V.N. and Frank-Kamenetskii,M.D. (1993) *Nucleic Acids Res.* **21**, 585–591.
23 Kang,S. and Wells,R.D. (1992) *J. Biol. Chem.* **267**, 20887–20891.
24 Martinez-Balbas,A. and Azorin,F. (1993) *Nucleic Acids Res.* **21**, 2557–2562.
25 Kiyama,R. and Oishi,M. (1995) *Nucleic Acids Res.* **23**, 452–458.
26 Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.* **65**, 499–560.
27 Kiyama,R. and Camerini-Otero,R.D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10450–10454.
28 Suda,T., Mishima,Y., Asakura,H. and Kominami,R. (1995) *Nucleic Acids Res.* **23**, 3771–3777.
29 Hyrien,O. (1989) *Nucleic Acids Res.* **17**, 9557–9569.
30 Weiler,K.S. and Wakimoto,B.T. (1995) *Annu. Rev. Genet.* **29**, 577–605.
31 Dernburg,A.F., Broman,K.W., Fung,J.C., Marshall,W.F., Philips,J., Agard,D.A. and Sedat,J.W. (1996) *Cell* **85**, 745–759.
32 Usdin,K. and Furano,A.V. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4416–4420.
33 Fry,M. and Loeb,L.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4950–4954.
34 Saenger,W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, Germany.
35 Nickol,J.M. and Felsenfeld,G. (1983) *Cell* **35**, 467–477.
36 Kohwi,Y. and Kohwi-Shigematsu,T. (1991) *Genet. Dev.* **5**, 2547–2554.
37 Woodford,K.J., Howell,R.M. and Usdin,K. (1994) *J. Biol. Chem.* **269**, 27029–27035.
38 Brahmachari,S.K., Meera,G., Sarkar,P.S., Balagurumoorthy,P., Tripathi,J., Raghavan,S., Shaligram,U. and Pataskar,S. (1995) *Electrophoresis* **16**, 1705–1714.
39 Kremer,E.J., Pritchard,M., Lynch,S., Yu,S., Holman,K., Baker,E., Warrnen,S.T., Schlessinger,D., Sutherland,G.R. and Richards,R.I. (1991) *Science* **252**, 1711–1714.
40 Gardiner-Garden,M. and Frommer,M. (1987) *J. Mol. Biol.* **196**, 261–282.
41 Lindsay,S. and Bird,A.P. (1987) *Nature* **327**, 336–338.
42 Sen,D. and Gilbert,W. (1988) *Nature* **334**, 364–366.
43 Jin,R., Breslauer,K.J., Jones,R.A. and Gaffney,B.L. (1990) *Science* **250**, 543–546.
44 Van der Ploeg,L.H.T., Liu,A.Y.C. and Borst,P. (1984) *Cell* **36**, 459–468.
45 Moyzis,R.K., Buckingham,J.M., Cram,L.S., Dani,M., Deaven,L.L., Jones,M.D., Meyne,J., Ratliff,R.L. and Wu,J.-R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6622–6626.