

Association of Poly(CA)·Poly(TG) DNA Fragments into Four-Stranded Complexes Bound by HMG1 and 2

Claire Gaillard and François Strauss*

The tandemly repeated DNA sequence poly(CA)·poly(TG) is found in tracts up to 60 base pairs long, dispersed at thousands of sites throughout the genomes of eukaryotes. Double-stranded DNA fragments containing such sequences associated spontaneously with each other *in vitro*, in the absence of protein, forming stable four-stranded structures that were detected by gel electrophoresis and electron microscopy. These structures were recognized specifically by the nuclear nonhistone high mobility group (HMG) proteins 1 and 2 as evidenced by gel retardation. Such sequence-specific complexes might be involved *in vivo* in recombination or other processes requiring specific association of two double-stranded DNA molecules.

Tandemly repeated sequences of one to six nucleotides, known as microsatellites, are very common in eukaryotic genomes. Among the most frequently found is poly(CA)·poly(TG), which is present in mammals in approximately 10^5 tracts, each up to 60 base pairs (bp) long (1). Although the function of these sequences is still unknown, they are of interest for several reasons. First, their length at any given site in the genome varies, and this length polymorphism has been a useful source of genetic markers for genome mapping (2). Second, they enhance recombination between DNA molecules, as observed in eukaryotes (3) and bacteria (4). Third, their physical conformation changes readily from the orthodox right-handed B-DNA to left-hand-

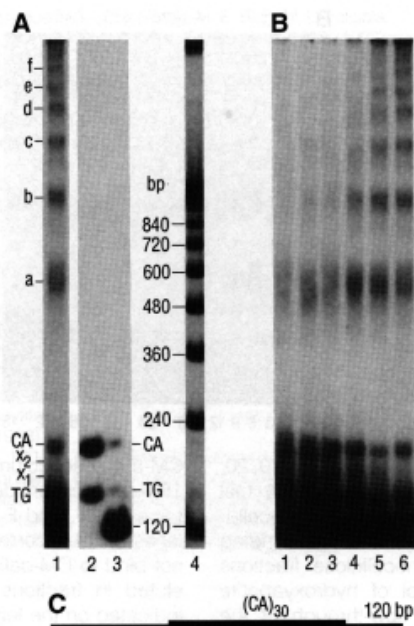
ed Z-DNA with ionic conditions (5) or negative supercoiling (6), providing a useful model to study the function of Z-DNA.

We observed that DNA fragments containing poly(CA)·poly(TG) formed stable multimeric complexes *in vitro* (Fig. 1). By gel electrophoresis, a 120-bp double-stranded DNA fragment containing a 60-bp tract of poly(CA)·poly(TG) showed, in addition to the intense band corresponding to the starting DNA fragment, a ladder of slowly migrating weak bands (Fig. 1A, lane 1). No protein and no reagent known to bind DNA had been added to the sample, however. When DNA from the predominant band was electroeluted and immediately loaded on a new gel, no band ladder was observed (lane 3). But the ladder reappeared when the same sample was analyzed again on a gel a few days later (7). In the group of lower bands, two were readily identified as the single strands of the DNA fragment (lane 2, bands labeled CA and TG);

Institut Jacques Monod, 2 place Jussieu, 75251 Paris 05, France.

*To whom correspondence should be addressed.

Fig. 1. Spontaneous formation of a band ladder by a DNA fragment containing a tract of poly(CA)·poly(TG). **(A)** A 120-bp fragment containing the sequence (CA)₃₀(TG)₃₀, isolated from a plasmid grown in *E. coli* (19), was end-labeled, purified by electroelution from a polyacrylamide gel, stored for several days at 4°C in 10 mM tris-HCl (pH 7.5), 1 mM EDTA, and analyzed by polyacrylamide gel electrophoresis (20) (lane 1). A series of bands (a to f) were produced that were not present in the starting material. Several lower bands (x₁, x₂, CA, TG) were also produced. Lane 2, starting DNA fragment, heat denatured; bands labeled CA and TG correspond to the single strands containing (CA)₃₀ and (TG)₃₀, respectively; lane 3, double strand DNA fragment immediately after gel purification; lane 4, linear oligomers of the fragment, obtained by incubation with T4 DNA ligase. **(B)** Influence of the DNA fragment concentration on the formation of a ladder. The labeled fragment was incubated in the presence of increasing concentrations of the same fragment, unlabeled. Total DNA fragment concentration: 0.02, 0.05, 0.1, 0.2, 0.5, and 2 ng/μl in lanes 1 to 6, respectively. **(C)** Map of the 120-bp DNA fragment containing a tract of (CA)₃₀(TG)₃₀ used in these experiments.



two others, labeled x₁ and x₂, varied markedly in intensity from one experiment to another (compare Figs. 1, 2, and 4). The upper bands formed a regular ladder (bands a to f), with the first rung of the ladder migrating at the position of linear fragments of 500 to 600 bp. This observation led us to study the possibility that DNA fragments had associated, in the absence of any protein, to form complexes of large steric volume as judged by their low

electrophoretic mobility. This suggestion was strengthened by the effect of varying the DNA fragment concentration. Figure 1B shows an increase in low-mobility species with increasing DNA fragment concentration, a result which would not be expected for the binding of a protein or another molecule to DNA, but which is in agreement with an association of the DNA fragments with each other.

The complexes were sufficiently stable to be purified by preparative electrophoresis. More than 70% of the eluted material migrated at the same position as on the preparative gel (Fig. 2B). The rest of the material migrated as regular double-stranded DNA fragments and may correspond to

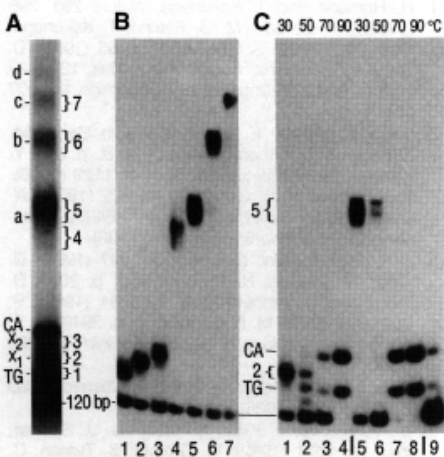


Fig. 2. Purification and stability of complexes. **(A)** Preparative gel of complexes formed by the 120-bp fragment. Gel slices in regions labeled 1 to 7 were cut, and DNA was electroeluted, precipitated with ethanol, and redissolved in 10 mM tris-HCl (pH 7.5), 1 mM EDTA. **(B)** DNA eluted from slices 1 to 7 was analyzed by gel electrophoresis in lanes 1 to 7, respectively. **(C)** Heat dissociation of the complexes. DNA eluted from slices 2 (lanes 1 to 4) and 5 (lanes 5 to 8) was incubated at the temperatures indicated and analyzed by gel electrophoresis. Lane 9, starting 120-bp fragment.

complexes dissociated during the elution process. The stability of the purified structures as a function of temperature was also studied (Fig. 2C). When heated in a buffer of low ionic strength, they dissociated at about 50°C to yield the starting double strand fragment, and only at higher temperature did the double strand DNA dissociate to two single strands. The dissociation temperature increased with the ionic strength (8), in agreement with the fact that complex formation occurred readily at 50°C and high ionic strength (7).

Similar structures could also be formed with a large DNA fragment. Instead of the 120-bp fragment, a linear 2390-bp plasmid containing the same (CA)₃₀(TG)₃₀ insert was used (Fig. 3A). When incubated at 37°C and high ionic strength and analyzed on an agarose gel, retarded bands were observed with an intensity that increased with the total concentration in DNA fragment (Fig. 3B, arrowheads). Like those of the 120-bp fragment, these complexes were stable and could be eluted from the gel. Observed in the electron microscope (Fig. 3C), they appeared as X-shaped structures with two large arms and two short arms, in which the lengths of the two large arms and of the two small arms were constant from one complex to another and in which the position of the junction relative to the ends was compatible with an association of two linear DNA molecules specifically at the level of their poly(CA)·poly(TG) tracts.

Therefore, like gel electrophoresis, electron microscopy experiments suggest an association of DNA fragments at the poly(CA)·poly(TG) inserts to form four-stranded structures in the absence of protein. The upper bands of the ladders are most probably due to the possibility of multiple

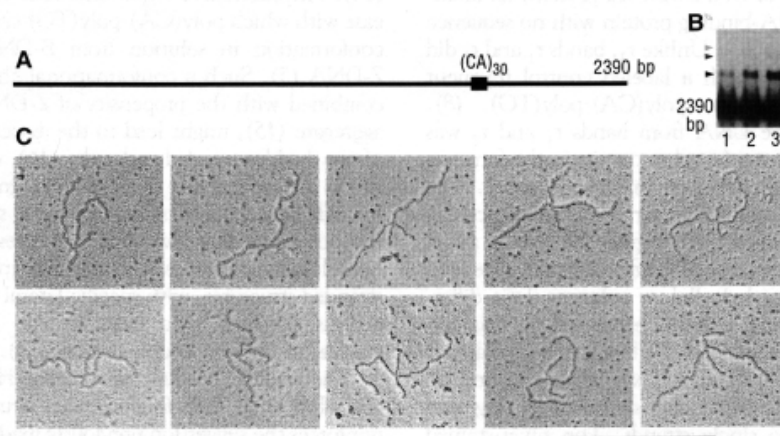
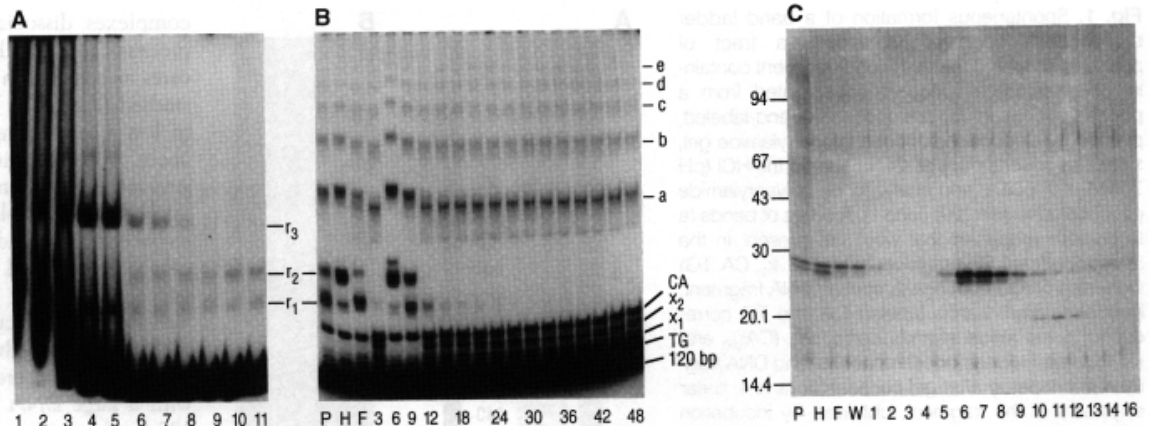


Fig. 3. Complexes formed by a plasmid containing a 60-bp tract of poly(CA)·poly(TG). **(A)** Map of the linearized 2390-bp plasmid. **(B)** Analysis of the complexes on an agarose gel. The end-labeled plasmid was incubated for 16 hours at 37°C in 2.5 M NaCl, 25 mM tris-HCl (pH 7.5), 1 mM EDTA in the presence of increasing amounts of the same unlabeled plasmid and analyzed on a 1% agarose gel. Total plasmid concentration: 0.1, 5, and 25 ng/μl in lanes 1 to 3, respectively. **(C)** X-shaped structures observed in the electron microscope (27). Bar, 1 μm.

Fig. 4. Specific interaction of proteins HMG1 and HMG2 with the complexes formed by a DNA fragment containing a tract of poly(CA)·poly(TG). **(A)** Interaction of the labeled 120-bp fragment with a nuclear extract of cultured monkey cells (CV1 line), as analyzed by gel retardation (9). Labeled DNA fragment was incubated with protein in the presence of increasing amounts of unlabeled competitor



DNA from *E. coli*. Competitor DNA concentration: 0, 0.6, 1.3, 2.5, 5, 10, 20, 40, 80, 160, and 320 ng/ μ l in lanes 1 to 11, respectively. **(B)** Gel retardation assay of fractions from a carboxymethyl-cellulose (CM-cellulose) column, the fourth step of purification of the specific DNA-binding activity detected in (A). Lane P, pool of phosphocellulose fractions containing the DNA-binding activity; lane H, pool of hydroxyapatite fractions containing the binding activity; lane F, flow-through of the CM-cellulose column; lanes 3 to 48, corresponding fractions of the

CM-cellulose column, analyzed every third fraction. **(C)** Silver-stained 18% polyacrylamide-SDS gel of fractions from the CM-cellulose column. Lanes P, H, and F are the same as in (B); lane W, CM-cellulose wash; lanes 1 to 16, corresponding CM-cellulose fractions. HMG1, which does not bind to CM-cellulose, is found in fractions F and W, whereas HMG2 eluted in fractions 5 to 9. Molecular size markers (Pharmacia) are indicated on the left (in kilodaltons).

associations along the 60 bp of poly(CA)·poly(TG). The detailed structure of such complexes will have to be defined precisely, but the following results may already shed some light on this question.

The 120-bp fragment was used to search for proteins binding specifically to the poly(CA)·poly(TG) sequence. In a gel retardation experiment in which a nuclear extract from cultured monkey cells (9) was added to the labeled fragment in the presence of increasing amounts of nonradioactive competitor DNA from *Escherichia coli*, three bands were visible (Fig. 4A), two of which, labeled r_1 and r_2 , remained present at the highest concentration of competitor DNA. Band r_3 , which disappeared when the amount of competitor was increased, was identified as the complex of DNA with Ku protein, an abundant DNA-binding protein with no sequence specificity (10). Unlike r_3 , bands r_1 and r_2 did not appear with a labeled control fragment containing no poly(CA)·poly(TG) (8). When the DNA from bands r_1 and r_2 was eluted, it was found to consist in the forms x_1 and x_2 already seen in Figs. 1 and 2. The proteins responsible for the binding activity were purified (9). Figure 4B shows a gel retardation assay of the fractions from the final carboxymethylcellulose column. Two different binding activities separated on this column and shifted all the bands of the ladders as well as bands x_1 and x_2 , while showing no binding to the double strand DNA fragment or to the single strands. The silver-stained SDS-polyacrylamide gel of Fig. 4C shows the purity of the corresponding proteins, which were identified as proteins HMG1 and HMG2 according to several criteria, including their abundance, their chromatographic characteristics, their appar-

ent molecular weight of about 29,000 in SDS, and their solubility in 5% trichloroacetic acid (8), which is a characteristic property of "high mobility group" proteins (11).

This result must be linked to the observation that proteins HMG1 and HMG2, and more precisely their HMG box, bind specifically to four-way helical junctions of DNA (12). The hypothesis may thus be made of an association of poly(CA)·poly(TG) sequences in a structure resembling a four-way junction (13), but with no strand exchange between the two double helices. Such crossed structures with no strand exchange have been observed in certain crystals of oligonucleotides, where two B-DNA helices are associated by reciprocal groove-backbone interactions (14). An alternative hypothesis rests on the ease with which poly(CA)·poly(TG) changes conformation in solution from B-DNA to Z-DNA (5). Such a conformational change, combined with the propensity of Z-DNA to aggregate (15), might lead to the association of two double-stranded molecules (16), which would in turn stabilize the Z-DNA segments. In addition, theoretical models have shown the possibility that two double helices with homologous sequences may associate in four-stranded structures, with specific lateral interactions between base pairs in the major grooves of the two double helices (17).

The specific binding of proteins HMG1 and HMG2 to such four-stranded structures reinforces the suggestion previously made that the role of HMG-box proteins is linked to their binding to structures formed in the course of wrapping or looping DNA (12, 18), which could be implicated in the structure of chromosomal domains, in recombination and repair, or in looping mechanisms involved

in gene activation. Independently of the function of HMG proteins, however, the known recombinogenic properties of poly(CA)·poly(TG) sequences, and the sequence-specific association of double helices shown here suggest that the structures we observe might be implicated in a "pairing first" mechanism of recombination or repair.

REFERENCES AND NOTES

1. H. Hamada and T. Kakunaga, *Nature* **298**, 396 (1982); H. Hamada, M. G. Petrino, T. Kakunaga, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6465 (1982); D. Tautz and M. Renz, *Nucleic Acids Res.* **12**, 4127 (1984); R. L. Stallings *et al.*, *Genomics* **10**, 807 (1991).
2. J. L. Weber and P. E. May, *Am. J. Hum. Genet.* **44**, 388 (1989); M. Litt and J. A. Luty, *ibid.*, p. 397; Y. Kashi *et al.*, *Nucleic Acids Res.* **18**, 1129 (1990); T. J. Hudson *et al.*, *Genomics* **13**, 622 (1992); W. Dietrich *et al.*, *Genetics* **131**, 423 (1992); J. Weissenbach *et al.*, *Nature* **359**, 794 (1992).
3. J. R. Stringer, *Mol. Cell. Biol.* **5**, 1247 (1985); D. Treco, B. Thomas, N. Arnheim, *ibid.*, p. 2029; D. Treco and N. Arnheim, *ibid.* **6**, 3934 (1986); P. Bullock, J. Miller, M. Botchan, *ibid.*, p. 3948; W. P. Wahls, L. J. Wallace, P. D. Moore, *ibid.* **10**, 785 (1990).
4. K. E. Murphy and J. R. Stringer, *Nucleic Acids Res.* **14**, 7325 (1986).
5. M. Vorlickova, J. Kypr, S. Stokrova, J. Sponar, *ibid.* **10**, 1071 (1982); C. Zimmer, S. Tyman, C. Marck, W. Guschlbauer, *ibid.*, p. 1081; A. Woisard and G. V. Fazakerley, *Biochemistry* **25**, 2672 (1986); J. P. Ridoix, J. Liqueur, E. Taillandier, *Nucleic Acids Res.* **15**, 5813 (1987); J. H. Riazance, W. C. Johnson, L. P. McIntosh, T. M. Jovin, *ibid.*, p. 7627; V. P. Antao, R. L. Ratliff, D. M. Gray, *ibid.* **18**, 4111 (1990).
6. D. B. Haniford and D. E. Pulleyblank, *Nature* **302**, 632 (1983); A. Nordheim and A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1821 (1983).
7. The incubation conditions had a marked importance in this phenomenon. The band ladder appeared spontaneously but very slowly at low or physiological ionic strength. Its formation was strongly accelerated by high ionic strength (2.5 M NaCl, or 3 M CsCl, or 1.6 M ammonium sulfate) and by raising the

incubation temperature to 37°C or up to 50°C. Under such conditions, the ladders shown here could be obtained within 3 hours rather than within days.

8. C. Gaillard and F. Strauss, data not shown.
9. Cell culture (CV1 line), nuclei purification, protein extraction, gel retardation, as well as phosphocellulose, DNA-sephacryl, and hydroxyapatite chromatography were performed as described by F. Strauss and A. Varshavsky [*Cell* **37**, 889 (1984)]. Because HMG1 and HMG2 copurified on these columns, a fourth chromatographic step on carboxymethyl-cellulose was performed as described by G. H. Goodwin and E. W. Johns [in *Methods in Cell Biology*, D. M. Prescott, Ed. (Academic Press, New York, 1977), vol. 16, chap. 15]. In appropriate salt conditions, HMG1 does not bind to this column, whereas HMG2 does.
10. S. Paillard and F. Strauss *Nucleic Acids Res.* **19**, 5619 (1991); T. Mimori, J. A. Hardin, J. A. Steitz, *J. Biol. Chem.* **261**, 2274 (1986).
11. E. W. Johns, Ed., *The HMG Chromosomal Proteins* (Academic Press, London, 1982); M. Bustin, D. A. Lehn, D. Landsman, *Biochim. Biophys. Acta* **1049**, 231 (1990).
12. J. M. Wright and G. H. Dixon *Biochemistry* **27**, 576 (1988); M. E. Bianchi, M. Beltrame, G. Paonessa, *Science* **243**, 1056 (1989); M. E. Bianchi, L. Falciola, S. Ferrari, D. M. J. Lilley, *EMBO J.* **11**, 1055 (1992).
13. D. M. J. Lilley and R. M. Clegg, *Annu. Rev. Biophys. Biomol. Struct.* **22**, 299 (1993).
14. Y. Timsit, E. Westhof, R. P. P. Fuchs, D. Moras, *Nature* **341**, 459 (1989); Y. Timsit and D. Moras, *J. Mol. Biol.* **221**, 919 (1991).
15. D. B. Haniford and D. E. Pulleyblank, *J. Biomol. Struct. Dyn.* **1**, 593 (1983); H. Castleman, L. Specthrie, L. Makowski, B. F. Erlanger, *ibid.* **2**, 271 (1984).
16. J. A. Blaho and R. D. Wells, *Prog. Nucleic Acid Res. Mol. Biol.* **37**, 107 (1989).
17. S. McGavin, *J. Mol. Biol.* **55**, 293 (1971); J. H. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3641 (1979).
18. D. M. J. Lilley, *Nature* **357**, 282 (1992); K. Giese, J. Cox, R. Grosschedl, *Cell* **69**, 185 (1992); S. Ferrari *et al.*, *EMBO J.* **11**, 4497 (1992); T. T. Paull, M. J. Haykinson, R. C. Johnson, *Genes Dev.* **7**, 1521 (1993).
19. A 120-bp Cla I-Eco RI restriction fragment that contained a 60-bp tract of poly(CA)-poly(TG) was used. The nucleotide sequence of the poly(CA)-containing strand of the fragment was as follows: CGATAAGCTTCTAGAGATCCC-(CA)₃₀-GGG-ATCCGTCGACTCTAGAGGATCCCCGGGC-GAGCTCG.
20. Four-percent polyacrylamide gels were used. Two different electrophoresis buffers were used, with identical results [either 40 mM tris-acetate, 20 mM sodium acetate, 1 mM EDTA (pH 7.8), or 6.7 mM tris-acetate, 3.3 mM sodium acetate, 1 mM EDTA (pH 7.8) with buffer recirculation]. Identical results were obtained with gels run at 4°C or 25°C.
21. DNA spreading was done with cytochrome c on an aqueous hypophase, followed by rotary shadowing with platinum. A control, in which the same DNA was spread at a 1000-fold higher concentration without preincubation at high ionic strength to form complexes, failed to show X-shaped structures, confirming that these are not due to random crossing of DNA molecules on the electron microscope grid.
22. We thank L. Jonk, who participated in an early stage of this work, C. Antony for help with the electron microscopy facility, and S. Elsevier for critical reading of the manuscript. Supported by grants from the Association Française des Myopathes, the Association pour la Recherche contre le Cancer, and the Ligue Nationale Française Contre le Cancer.

14 December 1993; accepted 7 March 1994