Detection of triple-helix related structures adopted by poly(dG)-poly(dC) sequences in supercoiled plasmid DNA

T.Kohwi-Shigematsu and Y.Kohwi

La Jolla Cancer Research Foundation, 10901 N.Torrey Pines Rd, La Jolla, CA 92037, USA

Received February 11, 1991; Revised and Accepted June 19, 1991

ABSTRACT

Negative superhelical strain induces the poly(dG)-poly(dC) sequence to adopt two totally different types of triple-helices, either a dG.dG.dC triplex in the presence of Mg^{++} at both neutral and acidic pHs or a protonated $dC^+.dG.dC$ triplex in the absence of Mg^{++} ions at acidic pH (1). To examine whether there are still other types of non-B DNA structures formed by the same sequence, we constructed supercoiled plasmid DNAs harboring varying lengths of the poly(dG) tract, and the structures adopted by each supercoiled plasmid DNA were studied with a chemical probe, chloroacetaldehyde. The potential of a poly(dG)-poly(dC) sequence to adopt non-B DNA structures depends critically on the length of the tract. Furthermore, in the presence of Mg^{++} and at a mildly acidic pH, in addition to the expected dG.dG.dC triplex detected for the poly(dG) tracts of 14 to 30 base pairs (bp), new structures were also detected for the tracts longer than 35 bp. The structure formed by a poly(dG) tract of 45 bp revealed chemical reaction patterns consistent with a dG.dG.dC triplex and protonated $dC^+.dG.dC$ triple-helices fused together. This structure lacks single-stranded stretches typical of intramolecular triplexes.

INTRODUCTION

Homopurine-homopyrimidine sequences are often found in the regulatory regions of various genes (2–9). These DNA sequences may play an important role in gene regulation. For example, a stretch of poly(dG)-poly(dC) sequence is located within 200 base pairs from the cap site of chicken adult β -globin gene, and an erythrocyte specific factor which binds specifically with such a sequence has been reported (10). Homopurine-homopyrimidine sequences have unusual structural properties. When cloned into supercoiled plasmid DNA, these sequences relieve negative supercoiling by adopting non-Watson–Crick type DNA structures (1,9,11,19). The exact structure that forms depends on the sequence and the chemical environment: pH, ionic strength, and type of metal ions (1,9,19).

We previously reported that poly(dG)-poly(dC) sequence, under superhelical strain and in the presence of Mg^{++} , folds sharply at its center forming a poly(dG)-poly(dG)-poly(C) triple helix, with both halves of the poly(dG) strand and the 5' half of the poly(dC) strand leaving the 3' half of the poly(dC) strand base unpaired. This structure containing a

purine.purine.pyrimidine triple-helix forms at pH 5–8 in the presence of Mg^{++} . In the absence of Mg^{++} , on the other hand, the poly(dC)-poly(dG)-poly(dC) triplex forms which requires protonation at cytosine residues to stabilize the $dC^+.dG.dC$ base triplet. Thus this structure is detected only under acidic conditions and is commonly referred to as a H-form (11).

To examine systematically the structural potentials of homopurine-homopyrimidine sequences, we analyzed DNA structures adopted by poly(dG)-poly(dC) sequences of varying lengths with the chemical probe, chloroacetaldehyde. Haloacetaldehydes react with the exocyclic amino group and the N-1 of adenine or N-3 of cytosine to generate 1, N^6 etheno-adenine or 3, N^4 ethenocytosine derivatives (20,21) when these bases are not hydrogen bonded (22,23,24). Although less reactive, guanine residues also react with chloroacetaldehyde at the exocyclic amino group and the N-1 or N-3 position (25,26,27). Haloacetaldehydes are useful probes for detecting unusual DNA structures because their reaction specificities are conserved at least within the pH range of 5–8 and specifically modify those sites that are strictly involved in non-B DNA conformations (1,15,28,29). Chloroacetaldehyde-modified sites can be mapped at a single-base resolution with chemical cleavage methods (1).

We report that the formation of either triple-helix, dG.dG.dC in the presence of Mg^{++} at neutral pH or $dC^+.dG.dC$ in the absence of Mg^{++} at acidic pH, depends strongly on the length of the poly(dG) tract. For longer tracts in the presence of Mg^{++} and at pH 5, new structures have been detected.

MATERIALS AND METHODS

Chemicals

Chloroacetaldehyde was purchased from Fluka and doubly distilled (boiling point, 78–80°C). The purified chloroacetaldehyde was aliquoted in small fractions and stored at –20°C until used. Hydrazine, formic acid and piperidine were purchased from Sigma.

Plasmid DNA

Varying lengths of poly(dG) tract were inserted into the Sma 1/Sac 1 sites in the pUC 13 plasmid by the G-C tailing method employing terminal transferase. Plasmid pUC 13 was first digested with Sma 1, and dG tailing was added with terminal transferase and dGTP. The plasmid was further digested with Sac 1, and dC tailing was added with terminal transferase and

dCTP. The reaction time terminal transferase was varied from 10–30 min to obtain a wide range of the poly(dG) tract length. After the second terminal transferase reaction, the plasmids were transferred into competent *E. coli* cells, and the resulting colonies were grown to determine the length of the tract. Supercoiled plasmid DNA for non-B DNA structural analysis was prepared by the alkaline lysis procedure (30) and purified by centrifugation to equilibrium in cesium chloride-ethidium bromide gradients twice.

Mapping of chloroacetaldehyde-modified sites

Supercoiled plasmid DNA containing varying lengths of poly(dG) tracts were reacted with 2 μ l of chloroacetaldehyde in 100 μ l-reaction volume at 37°C under the conditions described in the figure legends. Chloroacetaldehyde-modified DNAs were purified by successive precipitations with ethanol. Two micrograms of DNA was digested with Hind III and radiolabeled either at the 3' end with Klenow large fragment of DNA polymerase I or at the 5' end with kinase. The end-labeled DNA was further cleaved at a distal site with the second restriction enzyme, Bgl I, so that the resulting DNA fragment contains the poly(dG)-poly(dC) region. The DNA fragment was isolated from a native acrylamide gel, and the DNA was eluted from the gel in an elution buffer (0.5 M (NH₄)OAc, 1 mM EDTA, 0.1% SDS) at 42°C

overnight. The DNA was then subjected to hydrazine or formic acid reaction followed by piperidine reaction as described in Maxam and Gilbert (31), except that hydrazine or formic acid reaction was performed at 15°C instead of 20°C for 5 min. When unpaired DNA bases are modified with chloroacetaldehyde, the DNA becomes susceptible to cleavage by piperidine at these sites. For chloroacetaldehyde-modified cytosine residues, the cleavages at these sites can be detected by the appearance of new bands on a sequencing gel after the DNA is treated with formic acid and piperidine reactions which normally cleave DNA at adenine and guanine residues. Similarly, the cleavages of DNA due to chloroacetaldehyde-modified adenine and guanine residues can be detected as new bands within a cytosine ladder generated after the hydrazine/piperidine reactions.

RESULTS

The dG.dG.dC triplex forming potential of poly(dC) tracts of varying length at neutral pH

To study the effect of the poly(dG)-poly(dC) sequence length on its potential for adopting an altered DNA structure, supercoiled plasmid DNAs containing a poly(dG)-poly(dC) sequence of 9, 14, 18, 21, 28, 33, 35, or 45 base pairs (bp) were constructed. These plasmid DNA constructs were reacted with chloroacetaldehyde in a buffer consisting of 50mM Tris-maleic acid at pH 7 in the presence of 2mM Mg⁺⁺. The chloroacetaldehyde modified bases within and surrounding the poly(dG) tracts were examined. The end-labeled DNAs were isolated from a native polyacrylamide gel and subjected to formic

Figure 1. The length effect of the poly(dG) tract on a dG.dG.dC triplex formation at neutral pH with Mg⁺⁺. Supercoiled pUC13 plasmids containing varying number of dG residues, as indicated on top of each lane, are chloroacetaldehyde modified in a buffer containing 50mM Tris-maleic acid pH 7 and 2mM Mg⁺⁺. The mapping of chloroacetaldehyde-modified sites were performed as described in the text and Materials and Methods. Control DNA untreated with chloroacetaldehyde; lanes 1, 3, 5, 7, 9, 13, and 15. Chloroacetaldehyde-modified DNA; lanes 2, 4, 6, 8, 10, 12, 14, and 16. The DNA sequences immediately surrounding the poly(dC) tract are 5' TGTAACGACGCGCCAGTGAATTC-GAGCT (C)_n GGGGATCCTCTAGAGTCGA 3'.

Figure 2. The length effect of the poly(dG) tract on dC⁺.dG.dC triplex formation. Chloroacetaldehyde reaction was performed with supercoiled plasmids containing varying lengths of poly(dG) tracts in the buffer containing 50mM Tris-maleic acid at pH 5 without Mg⁺⁺. Subsequent analysis was performed as described in Figure 1 legend.

acid followed by piperidine reaction. These chemical reactions cleave DNA at adenine and guanine residues generating a purine ladder. When certain cytosine residues are chloroacetaldehyde-modified, DNA is also cleaved at those cytosines. Thus chloroacetaldehyde modified cytosine can be easily detected by the appearance of new bands within the purine ladder. The altered conformation adopted by a poly(dG)-poly(dC) sequence that contains a dG.dG.dC triplex can be detected by the chloroacetaldehyde reactivity over the 5' half of poly(dC) tract.

As shown in Fig. 1, the typical chloroacetaldehyde reactivity for poly(dG) tracts was seen at the dG length of 18 bp but not at shorter lengths examined here. Even after long exposure of the X-ray film, no chloroacetaldehyde reactivity over 5' half of the dC tract was revealed for dG₉ and dG₁₄ (data not shown). Thus, there is a length requirement between the dG tract of 14 and 18 base pairs for stabilization of the dG.dG.dC triplex structure under this condition (i.e. at bacterial superhelical density and the ionic condition employed). Furthermore, a marked increase in intensity was detected between dG₂₁ (Fig. 1, lane 8) and dG₂₈ (Fig. 1, lane 10) for the new bands generated from DNA cleavages within the dC strand at the chloroacetaldehyde-modified cytosine residues. This indicates a substantial increase in the population of plasmids that contains a dG.dG.dC triplex structure for dG tracts of 28 and longer.

The difference in the chloroacetaldehyde reaction pattern among dG tracts of varying length is not due to variations in the

superhelical density of the plasmid DNAs. Several independent DNA preparations gave identical results.

The length effects of poly(dG) tracts on protonated dC⁺.dG.dC triplex structure

A dC⁺.dG.dC triplex forms only at acidic pH in the absence of Mg⁺⁺. To study this triplex formation potential of poly(dG) tracts of varying length, a chloroacetaldehyde reaction pattern for the dC strand was obtained (Fig. 2). This triplex formation can be detected by the chemical reaction at the center of the dC strand and its 3' neighboring region. A typical chloroacetaldehyde reaction pattern indicating dC⁺.dG.dC triplex was seen for poly(dG) tracts of 28 bp and longer. For the poly(dG) tract of 21 bp, we observed reproducibly more widely dispersed chloroacetaldehyde reaction of cytosines in the poly(dC) tract. For poly(dG) tracts of 9 to 18 bp, there was no chloroacetaldehyde reactivity detected at all. These shorter poly(dG) tracts are apparently incapable of adopting the triplex.

The non-B DNA structure formation by poly(dG)-poly(dC) sequences at low pH and in the presence of Mg⁺⁺

A similar experiment to the one described above was repeated at 50mM Tris-maleic acid at pH 5 in the presence of 2mM Mg⁺⁺. As shown in Fig 3A, at pH 5, the minimal length of poly(dG) tract that showed the typical chloroacetaldehyde reaction pattern for the dG.dG.dC triplex was 14 bp, shorter than the 18

Figure 3. The length effect of the poly(dG) tract on a dG.dG.dC triplex formation at acidic pH with Mg⁺⁺. A) The same experiment as described in Figure 1 legend was repeated except that the buffer used for chloroacetaldehyde modification contained 50mM Tris-maleic acid at pH 5 and 2mM Mg⁺⁺. B) Chloroacetaldehyde-modified supercoiled plasmids containing either a 35 or 45 bp-poly(dG) tract was cleaved with Hind III and its 5' end was labeled with γ -ATP and kinase. The end-labeled DNA was further digested with Bgl I and the Hind III-Bgl I fragment was isolated from a native acrylamide gel. The DNA was subjected to hydrazine reaction (lanes 17–20) or formic acid reaction (lanes 21–24) followed by the cleavage reaction with piperidine. Control DNA untreated with chloroacetaldehyde, lanes 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23. Chloroacetaldehyde modified DNA; lanes 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24. The DNA sequences immediately surrounding the poly(dG) tract are 3' ACATTGCTGCCGGTCACTTAAGCTCGA (G)_n CCCCTAGGAGATCTCAGCT 5'. The 15 bp repeats within the dC₄₅ tract are marked with solid circle on side.

Figure 4. Model proposed for an altered triple-helix structure. Hydrogen bonds within triple-helices are shown by horizontal lines connecting three bases.

bp observed at neutral pH. At pH 5, longer poly(dG) tracts revealed yet another reaction pattern where the intensity of the cytosine-new bands is not uniform but localized in discrete regions in a periodic fashion. The poly(dG) tract of 35 bp revealed that the chloroacetaldehyde reaction with cytosines was confined to four discrete regions within the 5' half of poly(dC) tract. This trend first appeared over the uniform background for 33 bp, suggesting two plasmid populations with different structures. For the poly(dG) tract of 45 bp, the chloroacetaldehyde reaction is limited to two major regions, which occur in intervals of 15 bp: the first region centers at 15 bp, the second region at 30 bp from the 3' end of the poly(dC) tract.

Next, the chloroacetaldehyde reactivity with guanine residues within the poly(dG) tracts on the complementary strand was examined. The chloroacetaldehyde modifications of guanine residues can be detected as new bands, usually with low intensity, within the cytosine ladder generated after the DNA is reacted with hydrazine followed by piperidine. For a typical dG.dG.dC triplex, the guanine residues at the center of the poly(dG) tract are hyperreactive with chloroacetaldehyde, which is the case for the poly(dG) tracts of 14 to 28 bp (data not shown). For the poly(dG) tracts of 35 and 45 bp, the chloroacetaldehyde reactivity of their complementary dG strands are shown in Fig. 3B. For the poly(dG) tract of 35 bp, the chloroacetaldehyde reactivity is again confined to the center guanines of the poly(dG) tract (Fig. 3B, lane 18). On the other hand, for the poly(dG) tract of 45 bp, the chloroacetaldehyde-modified guanine residues were mapped to the guanines complementary to the chloroacetaldehyde-modified cytosines (Fig. 3B, lane 20). In addition, the center guanine reactivity to chloroacetaldehyde was also detected and most likely shows that there still exists a subpopulation of plasmid DNA within a plasmid DNA mixture

of different superhelical densities that are forming the dG.dG.dC triplex. The chloroacetaldehyde reactions at these guanine residues for poly(dG) tracts of 35 and 45 bp were confirmed by enhanced cleavages at the corresponding sites within the purine ladders generated by formic acid/piperidine reactions (Fig. 3B, lanes 22 and 24). In addition, the two new bands that appeared in the 3' neighboring regions of the poly(dG) tracts reflect adenine residues reacted with chloroacetaldehyde.

These altered chemical reaction patterns generated for longer poly(dG) tracts indicate that yet new DNA structures are formed (see Discussion for a proposed model for the structure formed by the poly(dG) tract of 45 bp).

DISCUSSION

Strong length dependency for triple-helix formation of poly(dG)-poly(dC) sequence

We examined the potential of adopting either a dG.dG.dC triplex or a dC⁺.dG.dC triplex by poly(dG) tracts of varying lengths. For either triplex formation under a given ionic condition, only a small difference in the length of dG tracts (within 2–7 bp) is sufficient to confer totally different DNA structures, such as double-stranded as opposed to triple-stranded or triple-stranded to yet another type of non-B form DNA structure. The experiments were performed using supercoiled DNA of bacterial superhelical density, and the ionic condition was kept the same at 50mM Tris-maleic acid. Under other ionic conditions, we predict that the pattern of the length dependency for the formation of triplex structures would vary. However, in any case, under varying ionic conditions or superhelical strain, it is likely that there exists a distinct border between the lengths of poly(dG) tracts that separate formation of different non-B DNA structures.

Our data demonstrate that a small change in the length of the poly(dG)-poly(dC) sequence is sufficient to induce completely different structures. Therefore this sequence can be used as a potential probe to examine structure/function relationships of DNA, if a biological activity associated with this sequence is found. In fact, the poly(dG)-poly(dC) sequence has been found to enhance gene expression in mammalian cells (submitted for publication).

Models for DNA structures adopted by poly(dG)-poly(dC) sequences of 45 bp at pH 5 and with Mg⁺⁺

Based on chloroacetaldehyde reaction patterns obtained for the dC and dG strands, a possible structure formed by a 45 bp-poly(dG) tract in supercoiled plasmid DNA is proposed (Fig. 4). We assumed that the guanine residues hyperreactive with CAA reflect the sites of sharp DNA folding based on our previous results for dG.dG.dC and dC⁺.dG.dC triple-helices. The CAA reactivity revealed that the entire poly(dG)-poly(dC) sequence is folded every 15 bp, since specific cytosines residues as well as their complementary guanine residues are reactive with CAA and the centers of CAA reactive regions are exactly 15 bp apart. This implies that the structure contains dG.dG.dC and dC⁺.dG.dC triple-helices immediately adjacent to each other as shown in the model. Note that this structure has inversion symmetry. The formation of any non-B DNA structure would cost thermodynamic energy because it requires formation of junctions with adjoining B-DNA. For an intramolecular triplex formation, it would require additional energy due to sharp folding of the sequence at the center. For the double-triplex, this folding

occurs twice and therefore would cost higher thermodynamic energy than a single-triplex. It was previously reported that the superhelical strain required to drive formation of a protonated non-B DNA structure by $(dG-dA)_n : (dC-dT)_n$ sequences is inversely related to the length of the sequence that is involved in structure formation (18). Apparently the 45-bp poly(dG) tract is sufficient in length to stabilize the double-triplex structure.

When a poly(dG)-poly(dC) tract folds at the center to form either an intramolecular dG.dG.dC or a $dC^+.dG.dC$ triplex, the fourth strand remains continuously base-unpaired. However, the double-triplex structure lacks these stretches of unpaired region. A single-triplex structure brings neighboring DNA sequences 5' and 3' of a poly(dG)-poly(dC) tract in close contact with each other (see the three-dimensional model in Ref. 18), while a double-triplex structure keeps them apart from each other (Fig. 4). For regulatory sequences, this could affect the assembly of the transcriptionally active complex. Therefore, if intramolecular triplex structures form *in vivo* within or in close proximity to regulatory regions, there might be a difference in the biological activity depending on which triplex structure is formed.

A similar symmetrical structure that contains two identical pyrimidine.purine.pyrimidine triplexes was suggested to form by a $(dT-dC)_{18}:(dA-dG)_{18}$ in H-DNA and $(dT-dC)_{18}$ in single-stranded phage M13 DNA (16) as well as linear, protonated oligonucleotides containing palindromic $d(TC)_n-d(GA)_n$ sequences (32). In either case, a double-triplex was formed intermolecularly. We did not expect to find that an intramolecular double-triplex could actually form in supercoiled plasmid DNA. We suspect that for the 35-bp poly(dG) tract there is a mixture of plasmid DNAs with different non-B DNA structures presumably related to intramolecular triplexes. However, no clear model can be envisioned at this point which satisfies the unusual chloroacetaldehyde reaction pattern.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (RO1-CA39681 to T. K.-S. and RO1-CA51377 to Y.K.) and by funds provided by the Cigarette and Tobacco Surtax Fund of the State of California through the Tobacco-Related Disease Research Program of the University of California (1KT98 to Y.K.) and American Cancer Faculty Award to T.K.-S.

REFERENCES

- Kohwi, Y. and Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3781–3785.
- Hentschel, C.C. (1982) *Nature* **295**, 714–716.
- Goding, C.R. and Russell, W.C. (1983) *Nucl. Acids Res.* **11**, 21–36.
- Mace, H.A.F., Pelham, H.R.B. and Travers, A.A. (1983) *Nature* **304**, 555–557.
- Siegfried, E., Thomas, G.H., Bond, U.M. and Elgin, S.C.R. (1986) *Nucl. Acids Res.* **14**, 9425–9444.
- Boles, T.C. and Hogan, M.E. (1987) *Biochemistry* **26**, 367–376.
- McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D. and Felsenfeld, G. (1981) *Cell* **27**, 45–55.
- Allan, M., Lanyon, W.G. and Paul, J. (1983) *Cell* **35**, 187–197.
- Kohwi, Y. (1989) *Nucl. Acids Res.* **17**, 4493–4502.
- Lewis, C.D., Clark, S.P., Felsenfeld, G. and Gould, H. (1988) *Genes Dev.* **2**, 863–873.
- Lyamichev, V.I., Mirkin, S.M. and Frank Kamenetskii, M.D. (1985) *J. Biomol. Struct. Dyn.* **3**, 327–338.
- Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A. and Frank Kamenetskii, M.D. (1987) *Nature* **330**, 495–497.
- Voloshin, O.N., Mirkin, S.M., Lyamichev, V.I., Belotserkovskii, B.P. and Frank Kamenetskii, M.D. (1988) *Nature* **333**, 475–476.
- Vojtisková, M. and Paleček, E. (1987) *J. Biomol. Struct. Dyn.* **5**, 283–296.
- Hanvey, J.C., Klysik, J. and Wells, R.D. (1988) *J. Biol. Chem.* **263**, 7386–7396.
- Htun, H. and Dahlberg, J.E. (1988) *Science* **241**, 1791–1796.
- Johnston, B.H. (1988) *Science* **241**, 1800–1804.
- Htun, H. and Dahlberg, J.E. (1989) *Science* **243**, 1571–1576.
- Bernués, J., Beltrán, R., Casanovas, J.M. and Azorín, F. (1989) *EMBO J* **8**, 2087–2094.
- Kochetkov, N.K., Shibaev, V.N. and Kost, A.A. (1971) *Tetrahedron Lett.* **22**, 1993–1996.
- Barrio, J.R., Secrist, J.A., III and Leonard, N.J. (1972) *Biochem. Biophys. Res. Comm.* **46**, 597–604.
- Kayasuga-Mikado, K., Hashimoto, T., Negishi, T., Negishi, K. and Hayatsu, H. (1980) *Chem. Pharm. Bull.* **28**, 932–938.
- Kimura, K., Nakanishi, M., Yamamoto, T. and Tsuboi, M. (1977) *J. Biochem. (Tokyo)* **81**, 1699–1703.
- Kohwi-Shigematsu, T., Enomoto, T., Yamada, M., Nakanishi, M., and Tsuboi, M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4689–4693.
- Sattang, P.D. Leonard, N.J. and Frihart, C.R. (1977) *J. Org. Chem.* **42**, 3292–3296.
- Oesh, F. and Doerj, G. (1982) *Carcinogenesis* **3**, 663–665.
- Singer, B., Spengler, S.J. Chavez, F and Kusmier, J.T. (1987) *Carcinogenesis* **8**, 745–747.
- Kohwi-Shigematsu, T., Manes, T. and Kohwi, Y. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2223–2227.
- Lilley, D.M.J. (1983) *Nucl. Acid Res.* **11**, 3097–3112.
- Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory.
- Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
- Glover, J.N.M. and Pulleyblank, D.E. (1990) *J. Mol. Biol.* **215**, 653–663.