Gene Ther Mol Biol Vol 7, 291-298, 2003

Unusual chemical hypersensitivity of the $d(GA)_n \cdot d(TC)_n$ repeat in vivo dependent on functional lactose repressor

Research Article

Gerald L. Buldak and Sergei M. Mirkin*

Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607

***Correspondence**: Sergei M. Mirkin, Tel: (312)-996-9610; fax: (312)-413-0353; e-mail: mirkin@uic.edu

Key Words: $d(GA)_n \cdot d(TC)_n$, microsatellites, primers, operator deletions, repressor deletion,

Abbreviations: Chloroacetaldehyde, (CAA); Escherichia coli, (E. coli); transcription start site, (TSS); isopropyl-D-thiogalactoside, (IPTG);

Received: 17 November 2003; Accepted: 23 December 2003; electronically published: December 2003

Summary

A microsatellite, $d(GA)_n \cdot d(TC)_n$, was inserted upstream of an inducible promoter in an Escherichia coli plasmid and its structure was probed by chemical footprinting in vivo. Hyper-reactivity to the single-strand DNA specific chemical, chloroacetaldehyde, was observed within the repeat, pointing to a structural transition within it. Surprisingly, hyper-reactivity of the $d(GA)_n \cdot d(TC)_n$ repeat diminished upon increased negative supercoiling caused by transcription. Furthermore, the fine modification pattern of the repeat was inconsistent with H-DNA or other known conformations that it adopts in vitro. Finally, functional lactose repressor appeared to be required for chemical hyper-reactivity of the repeat. We believe, therefore, that unanticipated binding of the lactose repressor to the $d(GA)_n \cdot d(TC)_n$ repeat, which is non-homologous to its regular binding sites, leads to elevated chemical sensitivity of the repeat in vivo.

I. Introduction

Simple tandem DNA repeats are of great interest because of their prevalence within various genomes (Cox and Mirkin, 1997), associations with various hereditary disorders (Kunkel, 1993) and ability to form various non-canonical structures both in vitro and in vivo (Sinden, 1994). Among the microsatellites that have drawn considerable interest is the repeat, $d(GA)_n \cdot d(TC)_n$. This sequence is known to comprise approximately 0.4-0.5% of the human genome, making it among the most highly represented microsatellites (Manor et al, 1988). This repeat has been found to play roles in important molecular transactions, most notably in areas associated with transcriptional regulation (Gilmour et al, 1989; Glaser et al, 1990; Wilkins and Lis, 1997; Leibovitch et al, 2002; Lu et al, 2003) and at replication borders (Baran et al, 1987; Rao et al, 1988; Baran et al, 1991; Rao, 1994; Krasilnikova et al, 2001). Destabilization of the $d(GA)_n \cdot d(TC)_n$ repeats has also been implicated in melanoma thus demonstrating its importance from the clinical perspective (Chakraborty, 2000).

The $d(GA)_n \cdot d(TC)_n$ repeat is capable of forming various non-canonical DNA structures in vitro. These structures include intramolecular triplex DNA structures, i.e. H-DNA (Mirkin, 1999), parallel-stranded DNA (Germann et al, 1998) and GA hairpins (Kalisch et al, 1998; Ortiz-Lombardia et al, 1998). The formation of these structures is dependent upon negative DNA supercoiling (Sinden, 1994). H-DNA can form in two possible ways, depending upon which strand of the repeat serves as the third strand (either the polypurine strand or the polypyrimidine strand). The H-r conformation, where the purine strand is the third strand, is stable at physiological pH in the presence of divalent cations, e.g. Mg^{2+}. The H-y conformation, where the polypyrimidine strand is the third strand, requires protonation of the cytosine residues, and is, thus, unfavorable under physiological pH (Mirkin, 1999).

Attempts to detect intramolecular triplexes formed by $d(G\text{-}A)_n \cdot d(TC)_n$ repeats DNA structures in vivo were made using chemical probing of intracellular DNA in a model Escherichia coli system (Karlovsky et al, 1990; Ussery and Sinden, 1993). The conclusions from these studies were that intramolecular triplexes could be formed in E. coli when two requirements were met: (i) superhelicity of intracellular DNA was artificially

increased, usually by growing cells in the presence of chloramphenicol, and (ii) ambient conditions of cell growth favored H-DNA formation by either the presence of divalent cations or decreased pH of the media.

Another group of studies focused on a d(GA)$_n$· d(TC)$_n$ repeat found within the promoter region of the hsp26 gene in Drosophila. Both the GA repeat and a protein factor that recognizes it, GAGA-factor, are required for the proper functioning during the heat-shock response in Drosophila. While this repeat forms H-DNA in vitro, in vivo studies, using mutational analysis and chemical footprinting, indicate that protein binding to this repeat rather than its triplex-forming potential, is essential for the promoter function (Glaser et al, 1990; Lu et al, 2003).

One can conclude, therefore, that the formation of H-DNA by a d(GA)$_n$· d(TC)$_n$ repeat can be induced in E. coli cells in principle, but is not readily detectable in a natural setting, such as the Drosophila genome. We decided, therefore, to look at the structure of the d(GA)$_n$· d(TC)$_n$ repeats cloned into an E. coli plasmid under physiological growth conditions. In this setting, one can elevate negative DNA supercoiling in the upstream promoter region by inducing transcription (Liu and Wang, 1987; Wu et al, 1988). We have previously demonstrated that transcription does, in fact, elevate DNA supercoiling in vivo inducing cruciform formation at d(AT)$_n$· d(TA)$_n$ repeats as far as 1 kbp upstream of the promoter (Dayn et al, 1992; Krasilnikov et al, 1999).

We inserted the d(GA)$_n$· d(TC)$_n$ repeat in two possible orientations upstream of an inducible promoter and probed its structure by chemical modification with chloroacetaldehyde (CAA) in vivo. While we observed chemical hyperreactivity within the repeat, its pattern was inconsistent with H-DNA or other known structures. Furthermore, this chemical reactivity was repressed, rather than enhanced, upon induction of transcription. Finally, hyper-modification of the d(GA)$_n$· d(TC)$_n$ repeat required functional lactose repressor. We believe, therefore, that lactose repressor could contribute to a structural alteration within the d(GA)$_n$· d(TC)$_n$ repeat, leading to its chemical hyper-reactivity in vivo.

II. Materials and methods

A. Primers

Sequencing primers used were: Sequencing primer Pr1 (5'ACGGTGCACCAATGCTTCTG3') and Pr2 (5'CCGGCTCGTATAATGTGT3'). Primers for PCR-mediated deletion of lactose operators were as follows: del1 (5'AAAAAGCTTCACTGCCCGCTTTC3'); del1' (5'GCCGTCAACCACCATC3'); del2 (5'AAAAAGCTTAGCGCGAATTGATC3'); del2' (5'CGGATAAAACTTGTGC3'); del3 (5'AAAACTCGAGTTCCACACATTATAC3'); del3' (5'AAAACTCGAGTCACACAGGAAACAGAC3'); del4 (5'TAGGTACATTGAGCAAC3'); del5 (5'TGTGACTCGAGTATTCGCTTGCTTATACGAGCCGGATG3').

B. Plasmid construction

Plasmid pTrc99A (Amann et al, 1988) was obtained from Pharmacia. Plasmid pTrcCat/Pst, carrying the unique PstI site at the -50 position relative to the trc promoter, was described in (Krasilnikov et al, 1999). Plasmid pTrcCat/Cla, carrying the unique ClaI site at the -180 position relative to the trc promoter, was a gift from Dr. Andrey Krasilnikov. Inserts containing d(GA)$_{30}$· d(TC)$_{30}$ and d(GA)$_{37}$· d(TC)$_{37}$ repeats were obtained by the EcoRI/HindIII digestion of the pGA30 and pGA37 plasmids, respectively, described in (Krasilnikova et al, 2001). These repeats were cloned into either the PstI site of pTrcCat/Pst plasmid, or the ClaI of the pTrcCat/Cla plasmid.

C. Operator deletions

The primary operator site, lacO1, was deleted by using the following PCR-based approach. Plasmid pTrcCat/Pst-GA37 was used as a template for independent PCR reactions with two pairs of primers. Using primers del1' and del3, a fragment of the plasmid was amplified that included the repeat, and the region immediately upstream of the O1 site relative to the transcription start site. A second fragment was generated using primers del3' and del4 that included the region immediately downstream of the O1 site. The two fragments were designed such that an XhoI restriction site replaced the O1 operator upon XhoI-digestion and co-ligation.

Pseudo-operator O3 was deleted by a similar approach using two PCR fragments from the pTrcCat/Pst-GA37 template. One of the fragments was amplified with primers del1 and del1', corresponding to the region immediately upstream of the O3 site. The other fragment was produced with primers del2 and del4, and contained the region immediately downstream of the O3 site. These fragments were generated such that a HindIII restriction site substituted for the O3 operator upon HindIII digestion and co-ligation.

To construct the plasmid without both operator sites, ΔO1-ΔO3, we deleted the O3 operator from the ΔO1 plasmid using the PCR approach described above.

In all these cases, the ligated PCR products contained NarI and KpnI sites at their 5'- and 3'-ends, respectively. Upon digestion with these two restriction enzymes, they were used to replace the corresponding NarI-KpnI fragment of the original pTrcCat/Pst plasmid.

D. Repressor deletion

Deletion of the lacI^q gene was achieved by excising the 0.9 kbp BssHII fragment, containing the coding part of the repressor, from the pTrcCat/Pst-GA$_{37}$ plasmid.

E. Bacterial strains

We used either an XL1-Blue strain (Invitrogen) or its derivative XL1-BluES, which was cured of the F'-factor by us previously (Krasilnikova et al, 2001).

F. Chemical probing of intracellular DNA

Intracellular plasmid DNA was modified with chloroacetaldehyde as described in (Krasilnikov et al, 1999).

G. Radiolabeling of DNA fragments and Maxam-Gilbert sequencing reactions

Labeling and sequencing reactions were carried out according to standard protocols (Sambrook et al, 1989).

III. Results

To look at the structure of $d(GA)_n \cdot d(TC)_n$ repeats under the influence of superhelical stress induced by transcription, we placed this repeat upstream of the inducible trc promoter in the multicopy pTrc-derived plasmid. DNA footprinting was carried out by CAA in vivo under conditions of both promoter induction and repression. CAA interacts specifically with base pairing positions of cytosines and adenines. Consequently, the stability of phosphodiester bonds decreases, and, after Maxam-Gilbert DNA sequencing, one can see additional bands corresponding to modified cytosines on the purine ladder and to modified adenines on the cytosine ladder (Kohwi and Kohwi-Shigematsu, 1988). Since in double stranded DNA, these positions are inaccessible, CAA modification reveals single-standed DNA segments or other significant DNA distortions.

The experimental results of chemical modification of both strands are presented in **Figure 1**. One can clearly observe modification within both the $d(GA)_n \cdot d(TC)_n$ repeat and the trc promoter. The size of the repeat here was 30 units and its subsequent increase to 37 units did not yield any difference in modification patterns. The scheme of modification within the $d(GA)_n \cdot d(TC)_n$ insert and the trc promoter region is presented in **Figure 2**. As can be seen in the figure, adenine residues within the 3' third of the polypurine strand were modified above background when the insert was placed into the plasmid in such a way that the purines were on the top strand as drawn. In each repeat length, when IPTG was absent from the growing culture, the 13 3'adenine (A) residues were modified. When IPTG was added, the length of the modification shrank. No modifications were observed within the polypyrimidine strand.

These results ran contrary to what was expected, based on the previous studies of different structure-prone repeats. First, the modification was more profound in the absence of transcription. Second, the modification was only evident on one DNA strand. Third, the modification was observed in only one orientation of the repeat relative to the promoter. Furthermore, modification results are inconsistent with H-DNA formation by the $d(GA)_n \cdot d(TC)_n$ repeat because of the lack of modification of the polypyrimidine strand.

As far as trc promoter structure, we observed prominent chemical hyperreactivity in both repressed and activated state. When IPTG was absent, distinct modifications could be observed at positions -8, -9, -11, -14, -16, -19, -20, -40, -43, -44, -45, -49 and -50 relative to the transcription start site (TSS), designated +1. Modification patterns in all instances described changed dramatically upon addition of IPTG and promoter activation. All hypermodifications present from -11 to -50 disappeared. At the same time, new modifications that were not present in the absence of IPTG appeared at positions +1 and +2. These results are also summarized in **Figure 2**.

To confirm that these modifications within the promoter were not occurring as a result of the presence of a $d(GA)_n \cdot d(TC)_n$ insert, we performed the same type of footprint analysis on the empty vector, pTrcCat/Pst.

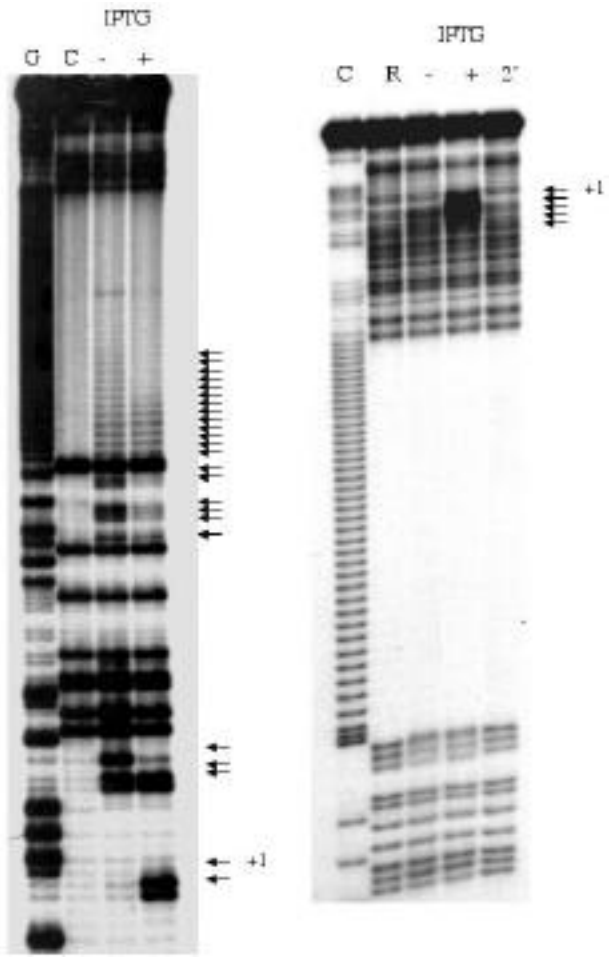

Figure 1. CAA modification in vivo of the polypurine (left) and polypyrimidine (right) strands of a $d(GA)_{37} \cdot d(TC)_{37}$ repeat in the pTrcCat/Pst plasmid. G, R and C represent Maxam-Gilbert sequencing reactions of plasmid DNA in vitro. IPTG - and + indicate CAA modifications in vivo in the absence or presence of IPTG, respectively. IPTG 2' is a control in which the plasmid DNA was CAA-modified in vivo for two minutes, instead of the normal 20 minutes. Arrows indicate modified adenine and cytosine bases.

Chemical probing of this plasmid in the presence and absence of IPTG revealed modification pattern of the promoter region that was identical to that of repeat-containing plasmids (**Figure 3**). Altogether, the promoter data support the model in which RNA polymerase interacts with the DNA at positions from -10 to -50 when the promoter is in the repressed state. Upon induction, RNA polymerase moves forward such that RNA polymerase-DNA interaction is now readily seen around the TSS instead of the 5'-part of the promoter.

The promoter alone is modified up to position -50 relative to the TSS. The $d(GA)_n \cdot d(TC)_n$ insert was inserted immediately upstream of this point. It is conceivable therefore, that modification of the repetitive run is caused by a mere extension of the promoter modification. To address this possibility, repeats were moved back to a position of -180 bp relative to the TSS. The unusual modifications within the repeat remained when placed at this position as well (**Figure 4**). Furthermore, the inhibitory effect of IPTG remained unchanged.

Figure 2. Summary of CAA modifications observed between positions -70 and +5 relative to the TSS. Modified cytosine and/or adenine residues shown in red occurred only in the absence of IPTG. Modifications shown in green occurred only in the presence of IPTG. Modifications shown in turquoise occurred in both states of promoter induction.

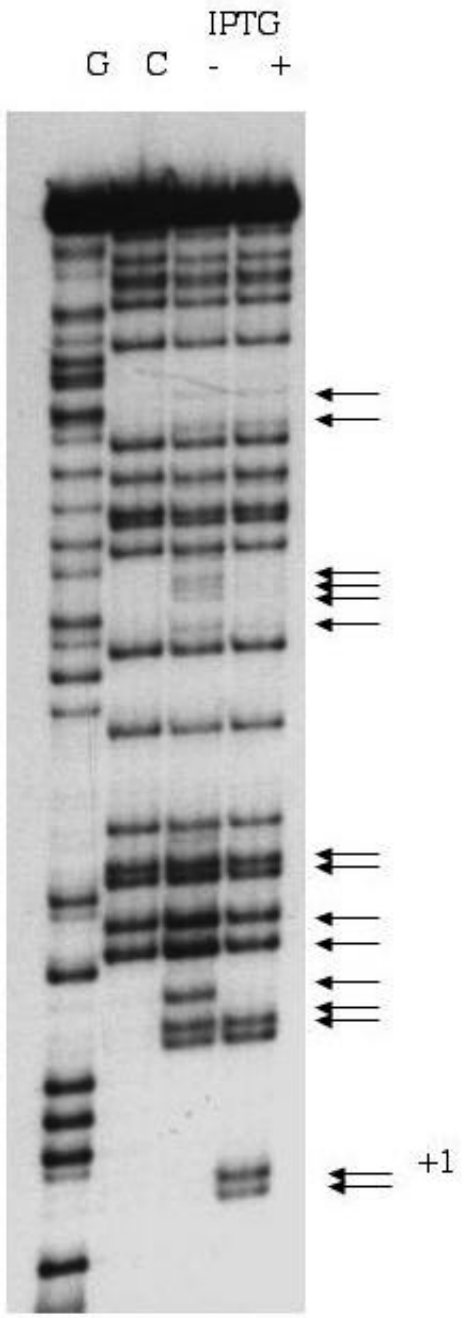

Figure 3. CAA modification in vivo of the non-transcribed strand of pTrcCat/Pst. G, C, Maxam-Gilbert sequencing reactions of plasmid DNA in vitro. IPTG - and + indicate C reactions for in vivo CAA-modified plasmid DNAs. Arrows indicate modified adenine and cytosine bases.

These data, therefore, rule out the possibility that the modification of the repeat results from the deformation at the promoter region.

Though modification of the d(GA)$_n$· d(TC)$_n$ insert does not seem to be directly linked to the trc promoter activity, it is downgraded in the presence of IPTG. Since the only known effect of IPTG is to block DNA-binding activity of the lactose repressor, we looked at the role of the lactose repressor in the d(GA)$_n$· d(TC)$_n$ insert modification. To this end, we first deleted lacI^q gene from our test plasmid and transformed the resultant construct into the lacI$^-$ background. The data on chemical probing in vivo presented in **Figure 5** show that there are no modifications of the d(GA)$_n$· d(TC)$_n$ insert in this setting. At the same time, the modifications within the promoter region were identical to those seen in previous experiments in the presence of IPTG, showing that trc promoter was in a constantly induced state. We conclude, therefore, that functional lactose repressor is necessary for a structural alteration of the d(GA)$_n$· d(TC)$_n$ repeat leading to its chemical hyperreactivity in vivo.

Lactose repressor is known to bind to at least two operator sites in our plasmid. The next question is therefore, whether repressor-operator binding is essential for the repeat's modification. To answer this question, we deleted the operator and pseudo-operator, either individually or together. When the primary operator (O1) was deleted, the modification of the d(GA)$_n$· d(TC)$_n$ repeat was similar to that observed in the presence of the normal operator site (**Figure 6**). The modification of the promoter region, as one would expect, became equivalent to that of the induced wild-type promoter even in the absence of IPTG. When the O3 pseudo-operator was deleted individually, leaving O1 intact, modifications of both the insert and of the promoter regions (**Figure 7A**) were identical to those seen in the original plasmid as summarized in **Figure 2**. Furthermore, when both operator sites were deleted (ΔO1-ΔO3 construct), the results of modification were identical to those seen in the case of the ΔO1 (**Figure 7B**). We conclude, therefore, that lactose repressor per se, rather than its binding to the operator sites, is necessary for the d(GA)$_n$· d(TC)$_n$ hyperreactivity.

IV. Discussion

We found that a d(GA)$_n$· d(TC)$_n$ repeat does undergo structural changes, accompanied by chemical modification

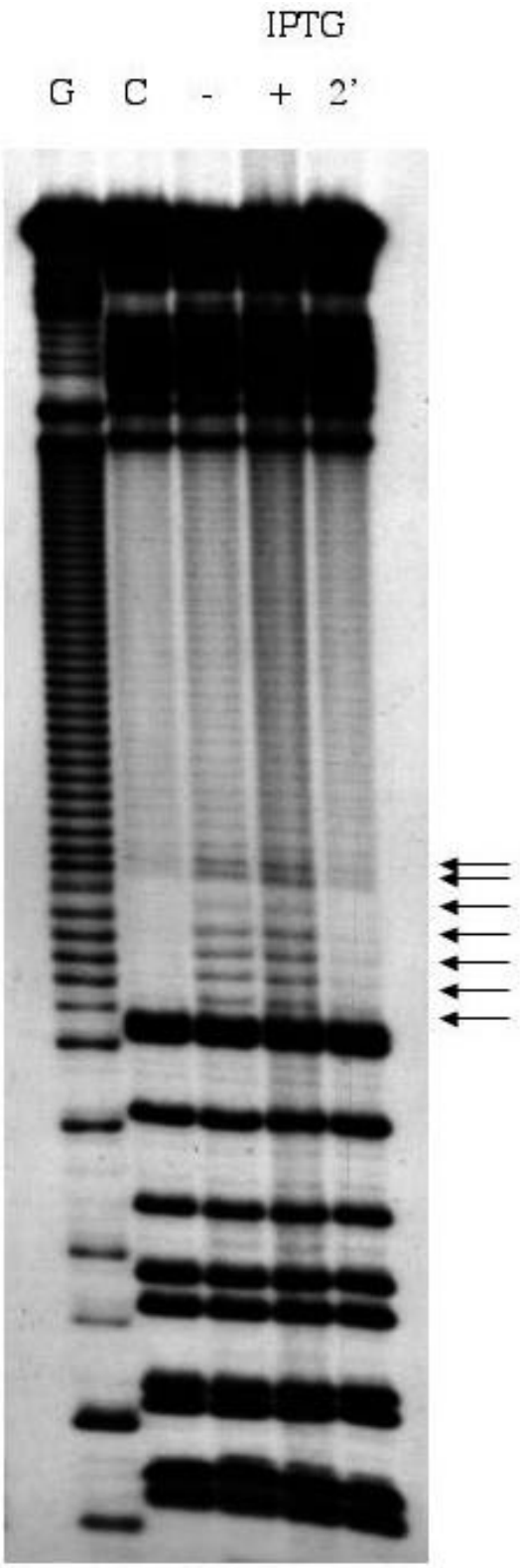

Figure 4. CAA modification in vivo of the polypurine strand of a $d(GA)_{37}\cdot d(TC)_{37}$ repeat at position -180 relative to the TSS in pTrcCat/Cla-GA37. G, C, Maxam-Gilbert sequencing reactions of plasmid DNA in vitro. IPTG - and + indicate CAA modifications in vivo in the absence or presence of IPTG, respectively. IPTG 2' is a control in which the plasmid DNA was CAA-modified in vivo for two minutes, instead of the normal 20 minutes. Arrows indicate modified adenine residues.

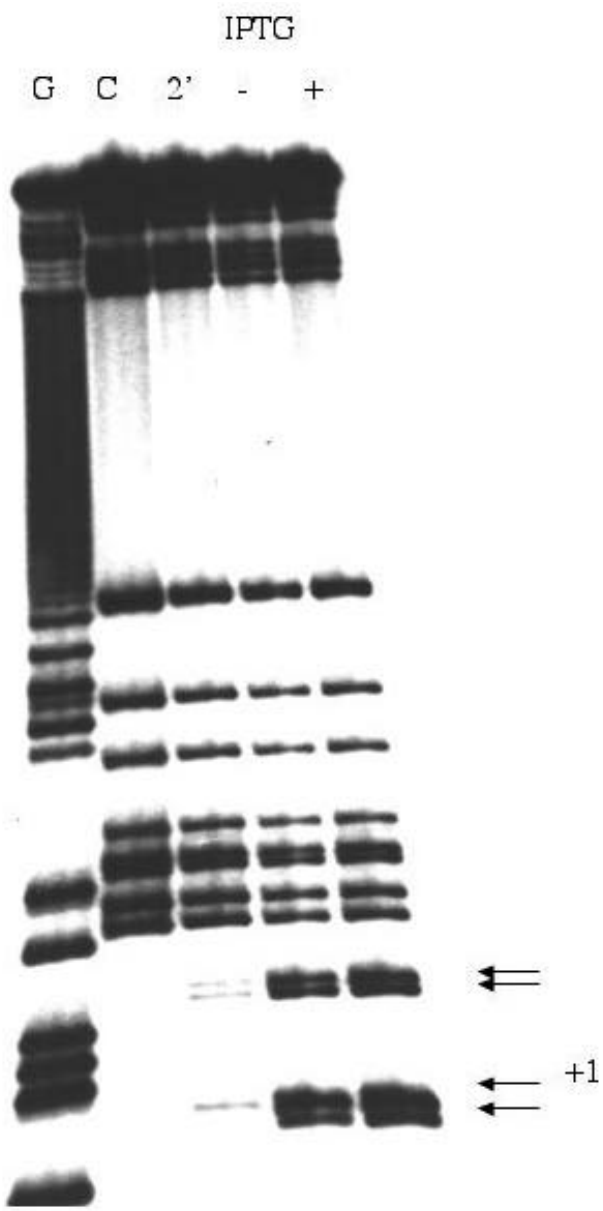

Figure 5. CAA modification in vivo of the polypurine strand of a $d(GA)_{37}\cdot d(TC)_{37}$ repeat, inserted 50 bp upstream of the trc TSS in pTrcCat/Pst-GA37ΔlacI^q. G, C, Maxam-Gilbert sequencing reactions of plasmid DNA in vitro. IPTG - and +, C reactions for plasmid DNAs that were CAA-modified in vivo in the absence or presence of IPTG, respectively. The left panel shows the $d(GA)_{37}\cdot (d(TC)_{37}$ repeat with the promoter, while right panel only the repeat. Arrows indicate modified adenine and cytosine residues.

of the 3'-part of the polypurine strand, when situated in the upstream promoter area. Contrary to expectations, based on previous studies by us and others, this modification pattern is not immediately consistent with previously described types of non-canonical structures adopted by this repeat.

Based on our previous studies (Dayn et al, 1992; Krasilnikov et al, 1999), we assumed that transcriptional supercoiling would stimulate structural transitions within the $d(GA)_n\cdot d(TC)_n$ repeat.