

## A New Trick for an Old Dog: TraY Binding to a Homopurine-homopyrimidine Run Attenuates DNA Replication

Maria M. Krasilnikova, Ekaterina V. Smirnova, Andrei S. Krasilnikov and Sergei M. Mirkin\*

Department of Molecular Genetics, University of Illinois at Chicago, Chicago IL 60607, USA

The effects of the  $d(GA)_n \cdot d(TC)_n$  repeat on plasmid replication in *Escherichia coli* cells were analyzed using electrophoretic analysis of replication intermediates. This repeat appeared to stall the replication fork progression in *E. coli* strains carrying  $F'$  episomes. The potency of replication stalling increased with the repeat's length but did not depend on its orientation relative to the replication origin, or transcription through the repeat. Treatment of *E. coli* cells with the protein synthesis inhibitor chloramphenicol abolished replication blockage, indicating that protein binding might be responsible for the repeat-caused replication blockage. Concordantly, dimethylsulfate footprinting *in vivo* revealed methylation protection of all guanine residues within the  $d(GA)_n \cdot d(TC)_n$ . Gel retardation assays with crude cell extracts confirmed the presence of a  $d(GA)_n \cdot d(TC)_n$ -binding activity in  $F'$ , but not  $F^-$ , strains. Further, strains cured from the  $F'$  episome lost this activity, while  $F^-$  strains that acquired the  $F'$  factor *via* conjugation, acquired  $d(GA)_n \cdot d(TC)_n$ -binding activity as well. Thus, this  $d(GA)_n \cdot d(TC)_n$ -binding protein is encoded by the  $F'$  factor. Purification of this protein by affinity chromatography revealed a single polypeptide with an apparent molecular mass of 15.2 kDa. Microsequencing of its two tryptic peptides revealed two perfect matches with the TraY protein, which is encoded by the F factor. Overexpression of an individual TraY protein in the  $F^-$  *E. coli* strain conveyed  $d(GA)_n \cdot d(TC)_n$ -binding activity *in vitro* and replication stalling at  $d(GA)_n \cdot d(TC)_n$  repeats *in vivo*. We conclude that TraY binding to a homopurine-homopyrimidine repeat is responsible for stalling DNA replication. Biological applications of this phenomenon are discussed.

© 2001 Academic Press

**Keywords:** TraY protein; homopurine-homopyrimidine repeat; DNA replication; termination; *Escherichia coli*

\*Corresponding author

### Introduction

Replication of simple DNA repeats attracted serious attention in recent years for several reasons. First, massive accumulation of genomic data revealed that inverted, direct and mirror DNA repeats are enormously overrepresented in natural DNAs, particularly in eukaryotes.<sup>1,2</sup> This points to the existence of a mechanism(s) responsible for their expansion in the course of evolution, with anomalous repeat replication being a prime candi-

date to explain expansion. Second, already substantial length polymorphisms of mono- and dinucleotide repeats, such as  $d(A)_n \cdot d(T)_n$ ,  $d(CA)_n \cdot d(GT)_n$  and  $d(GA)_n \cdot d(TC)_n$ , characteristic for various organisms, are additionally increased in human cancers, such as hereditary non-polyposis colorectal cancer.<sup>3</sup> The latter is attributed to mutations in the mismatch repair system and subsequent lack of repair of misaligned daughter and template strands during the replication of repeats.<sup>4</sup> Third, more than a dozen human neurological disorders are caused by large-scale expansions of simple trinucleotide repeats, such as  $(CGG)_n \cdot (CCG)_n$ ,  $(CTG)_n \cdot (CAG)_n$  and  $(GAA)_n \cdot (TTC)_n$ .<sup>5</sup> It is widely assumed that the

Abbreviations used: DMS, dimethylsulfate.  
E-mail address of the corresponding author:  
mirkin@uic.edu

large-scale nature of this phenomenon could be best explained by abnormalities in trinucleotide repeats replication.<sup>6-8</sup>

The effects of simple DNA repeats on DNA synthesis *in vitro* is sufficiently well studied. They commonly block DNA polymerization due to their propensity to form unusual DNA structures. For example, inverted repeats attenuate DNA polymerases by forming stable hairpins within template DNA strands;<sup>9-14</sup> homopurine-homopyrimidine mirror repeats block DNA synthesis due to the formation of triplexes in the course of DNA polymerization,<sup>15-18</sup> and (CNG)<sub>n</sub>-composed trinucleotide repeats prevent DNA polymerization due to the formation of unorthodox hairpins or intramolecular quadruplexes.<sup>19-22</sup> It has been suggested that length polymorphisms of such repeats might be generated while DNA polymerase is trying to overcome this blockage.<sup>22,23</sup>

The next question, i.e. whether the replication fork stalls at simple DNA repeats *in vivo*, similarly to purified DNA polymerases *in vitro*, is far less clear. The replication fork is equipped with DNA helicases and a single-strand binding (SSB) protein that are known to disrupt unusual secondary structures and diminish repeat-caused polymerization blockage *in vitro*.<sup>24-28</sup> Thus, structural properties of DNA repeats may not necessarily be as detrimental for replication *in vivo* as they are for polymerization *in vitro*. Yet, blockage of DNA replication *in vivo* by different simple repeats, including d(GA)<sub>n</sub>·d(TC)<sub>n</sub>,<sup>29-31</sup> d(CGG)<sub>n</sub>·d(CCG)<sub>n</sub>,<sup>23</sup> d(G)<sub>n</sub>·d(C)<sub>n</sub><sup>32</sup> and d(GAA)<sub>n</sub>·d(TTC)<sub>n</sub>,<sup>33</sup> was observed in various cell cultures. Note, however, that inhibitory effects of DNA repeats on replication *in vivo* could arise from different reasons: structure of DNA template, protein binding to repeated runs, repeat-mediated changes in chromatin structure, transcription-mediated replication arrest etc.

Our lab concentrated on replication of various repeats in bacterial cells using electrophoretic analysis of replication intermediates. We have described two mechanisms responsible for the repeat-caused replication stalling *in vivo*. Expandable trinucleotide repeats appear to inhibit replication elongation due to an unusually stable secondary structure formed by a repeated stretch at the lagging strand template.<sup>23</sup> d(G)<sub>n</sub>·d(C)<sub>n</sub> repeats severely impede replication, but this inhibition depends entirely on the transcription through a repeated run. Stalling RNA polymerase at the repeat and formation of a multistranded complex containing both DNA and RNA chains is likely responsible for the subsequent replication blockage.<sup>32</sup>

Here, we studied the effects of another repeat, d(GA)<sub>n</sub>·d(TC)<sub>n</sub>, on DNA replication in *Escherichia coli* cells. This repeat is interesting for several reasons. First, this is one of the most common microsatellites in eukaryotic DNA.<sup>34</sup> Second, structural transitions of this repeat are well characterized; under superhelical stress, it forms triple-

helical H-DNA *in vitro* and sometimes *in vivo*.<sup>35,36</sup> Third, DNA polymerization throughout this repeat *in vitro* is blocked due to H-DNA formation.<sup>16</sup> Fourth, it attenuates the replication fork progression in mammalian cells.<sup>29,31</sup> Finally, length polymorphism, characteristic for this repeat, was attributed to irregularities in the repeat's replication, and it is additionally increased by mismatch repair mutations associated with several human cancers.<sup>37</sup>

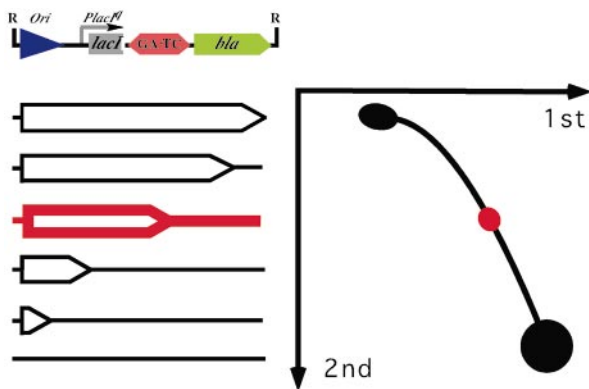
Using electrophoretic analysis of replication intermediates,<sup>38</sup> we have found that replication fork progression is attenuated at this repeat in *E. coli* plasmids, so that the strength of attenuation increases with the repeat's length. Surprisingly, however, this blockage depended entirely on a protein binding to the repeated stretch. Upon purification and microsequencing, this protein was identified as the renowned TraY protein, encoded by the F-factor that existed in our cells. In *E. coli* strains without TraY protein, there is no repeat-binding activity and the replication fork progresses through the repeat smoothly. We conclude that binding of multiple TraY protomers to the d(GA)<sub>n</sub>·d(TC)<sub>n</sub> repeat, rather than the structure of repeated DNA *per se*, impedes the replication fork progression. This constitutes yet another mechanism of repeat-caused replication blockage.

## Results

### d(GA)<sub>n</sub>·d(TC)<sub>n</sub> repeat blocks DNA replication in some *E. coli* strains

d(GA)<sub>n</sub>·d(TC)<sub>n</sub> inserts of varying lengths were first cloned into described vectors pTrc99Δ<sup>23</sup> (Figure 1). In these vectors, the *bla* gene is controlled by the lacI<sup>q</sup> promoter, while the multiple cloning site (in two orientations) is situated between this promoter and the coding part of the *bla* gene. The two orientations of the multiple cloning site (MCS) allowed us to clone d(GA)<sub>n</sub>·d(TC)<sub>n</sub> runs in both orientations relative to the replication *ori*. Since the directionality of DNA replication for the pTrc99Δ plasmid is well established, we knew for every recombinant plasmid whether d(GA)<sub>n</sub> or d(TC)<sub>n</sub> runs were situated in the lagging strand template for DNA replication. All of our plasmids were named accordingly.

To study the effects of d(GA)<sub>n</sub>·d(TC)<sub>n</sub> repeats on DNA replication *in vivo*, we used two-dimensional neutral/neutral gel electrophoresis of replication intermediates.<sup>39</sup> Our approach is schematically presented in Figure 1. Intermediate products of plasmid replication are Θ-shaped. Since our plasmid replicates unidirectionally, digestion of these intermediates with a restriction enzyme upstream of the replication origin converts them into bubble-shaped molecules, where the size of the bubble correlates with the duration of replication. Bubble intermediates differ in their molecular mass (ranging from one to two plasmid masses) and shape. They are separated in two dimensions:



**Figure 1.** Electrophoretic analysis of replication intermediates for plasmids containing  $d(G-A)_n \cdot d(T-C)_n$  repeats. Upper left panel shows the structure of our plasmids.  $d(G-A)_n \cdot d(T-C)_n$  inserts in two orientations were cloned between the truncated *lacI* gene and the promoterless *bla* gene. This whole cassette is transcribed from the *lacI*<sup>tr</sup> promoter. The restriction enzyme *AlwNI* (designated R) cleaves this plasmid at a unique site upstream of the *ori*. The lower left panel shows that upon restriction digestion, most of the replicative intermediates are bubble-shaped and the size of a bubble increases during replication. Stalling of the replication fork at the  $d(G-A)_n \cdot d(T-C)_n$  repeat results in the accumulation of a specific bubble intermediate (shown in red). The right panel diagrams the separation of bubble-like intermediates by 2D neutral/neutral gel electrophoresis. They are resolved by mass in the first dimension and by mass and shape in the second dimension. The bubble-arc is revealed by Southern hybridization. Stalling of the replication fork at the repeat leads to the appearance of a bulge (red spot) on the otherwise smooth arc.

first by mass (low percentage agarose) and second by mass and shape (high percentage agarose with ethidium bromide). Southern-blotting hybridization with a radioactive plasmid probe reveals a so-called "bubble arc". If there are no impediments during DNA replication, this arc is smooth. Stalling the replication fork at the  $d(G-A)_n \cdot d(T-C)_n$  repeats,

however, should lead to the accumulation of an intermediate of a given size and shape, generating a bulge on the otherwise smooth arc.

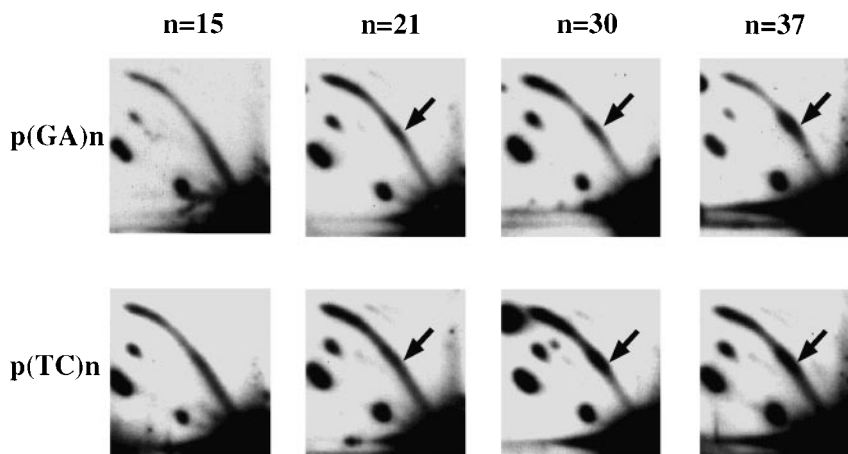
The effects of different  $d(G-A)_n \cdot d(T-C)_n$  repeats on the replication fork progression in *E. coli* XL1-Blue cells are presented in Figure 2. One can see that long repeats cause the appearance of bulges on replication arcs reflecting the replication fork stalling. This stalling clearly depends on the repeat's length: it is negligible at 15 repeats, more pronounced at 21 repeats and very prominent at 37 repeats. At the same time, the strength of the replication stalling does not seem to depend on the repeat's orientation relative to the replication *ori*, which contrasts  $d(G-A)_n \cdot d(T-C)_n$  data from results for expandable trinucleotide repeats.<sup>23</sup>

We have also found that for some simple DNA repeats, such as  $d(G)_n \cdot d(C)_n$ , replication fork stalling depends on transcription through the repeated stretch.<sup>32</sup> To understand whether this is true for  $d(G-A)_n \cdot d(T-C)_n$  repeats, we have cloned those repeats into the *EcoRI* site of the mini-pBR plasmid,<sup>23</sup> where they are situated upstream of the *amp* promoter in the non-transcribed area. This differentiates the latter constructs from those used in Figure 2, where  $d(G-A)_n \cdot d(T-C)_n$  runs are in the transcribed area. Replication stops caused by  $d(G-A)_n \cdot d(T-C)_n$  repeats situated in the non-transcribed region appeared to be very similar to those shown in Figure 2, indicating that transcription is not a factor here (data not shown).

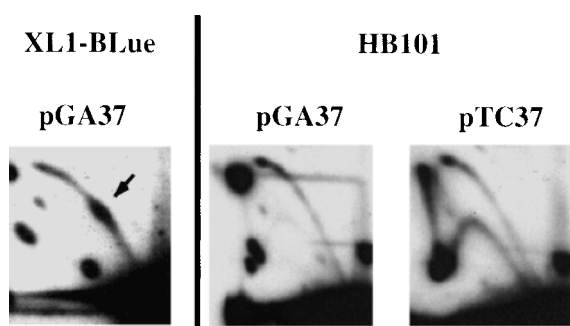
We noticed that  $d(G-A)_n \cdot d(T-C)_n$ -caused replication blockage is strain-dependent. Figure 3 shows that the same length repeat causes profound replication arrest in the XL1-Blue strain, but no replication inhibition at all in the HB101 strain. This indicates that some strain-specific factor is responsible for the repeat-caused replication blockage.

#### $d(GA)_n \cdot d(TC)_n$ -caused replication blockage is due to a protein bound to the repeat

One obvious possibility explaining the strain-dependent nature of the repeat-caused replication



**Figure 2.** Electrophoretic analysis of replication intermediates of plasmids with various  $d(G-A)_n \cdot d(T-C)_n$  inserts. Plasmids are named according to the repeated sequence in the template for the lagging strand DNA synthesis. The arrows show replication stops.

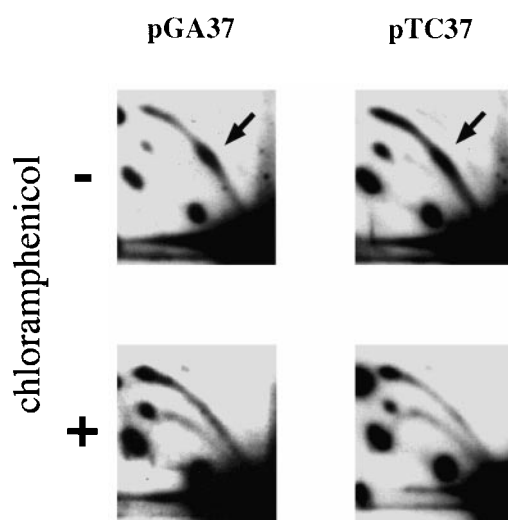


**Figure 3.**  $d(G-A)_n \cdot d(T-C)_n$  repeats block the replication fork progression in a strain-dependent manner. Plasmids are designated as in Figure 2. The arrows show replication stops.

blockage is that there is a protein, expressed in certain *E. coli* strains, that binds to  $d(G-A)_n \cdot d(T-C)_n$  runs, obstructing the replication fork progression. This model is consistent with the lack of orientation dependency and the position dependency in the effects of  $d(G-A)_n \cdot d(T-C)_n$  repeats on DNA replication. Further, augmented replication blockage with an increase in the repeat's length can easily be explained by more protein molecules binding to longer repeats. This hypothesis was addressed in several ways.

In the first set of experiments, we utilized a phenomenon of plasmid amplification.<sup>40</sup> While protein synthesis *de novo* is required for the initiation of bacterial DNA synthesis, it is not necessary for ColE1-type plasmids. Consequently, in the presence of the protein synthesis inhibitor chloramphenicol, plasmids amplify rapidly. This gross increase of the plasmid copy number occurs in cells where the protein content is, at best, stagnant. Thus, if protein binding is responsible for the replication blockage, one would expect that replication stops would decrease in intensity upon chloramphenicol treatment. Figure 4 shows that this is indeed the case. Prolonged (seven hours) chloramphenicol treatment abolishes replication stalling at  $d(G-A)_{37} \cdot d(T-C)_{37}$  runs, notwithstanding their orientation relative to the *ori*. Note that in two cases, expandable trinucleotide repeats<sup>23</sup> and  $d(G)_n \cdot d(C)_n$  runs,<sup>32</sup> chloramphenicol treatment enhanced rather than abolished replication stops, indicating that protein binding was an unlikely scenario there.

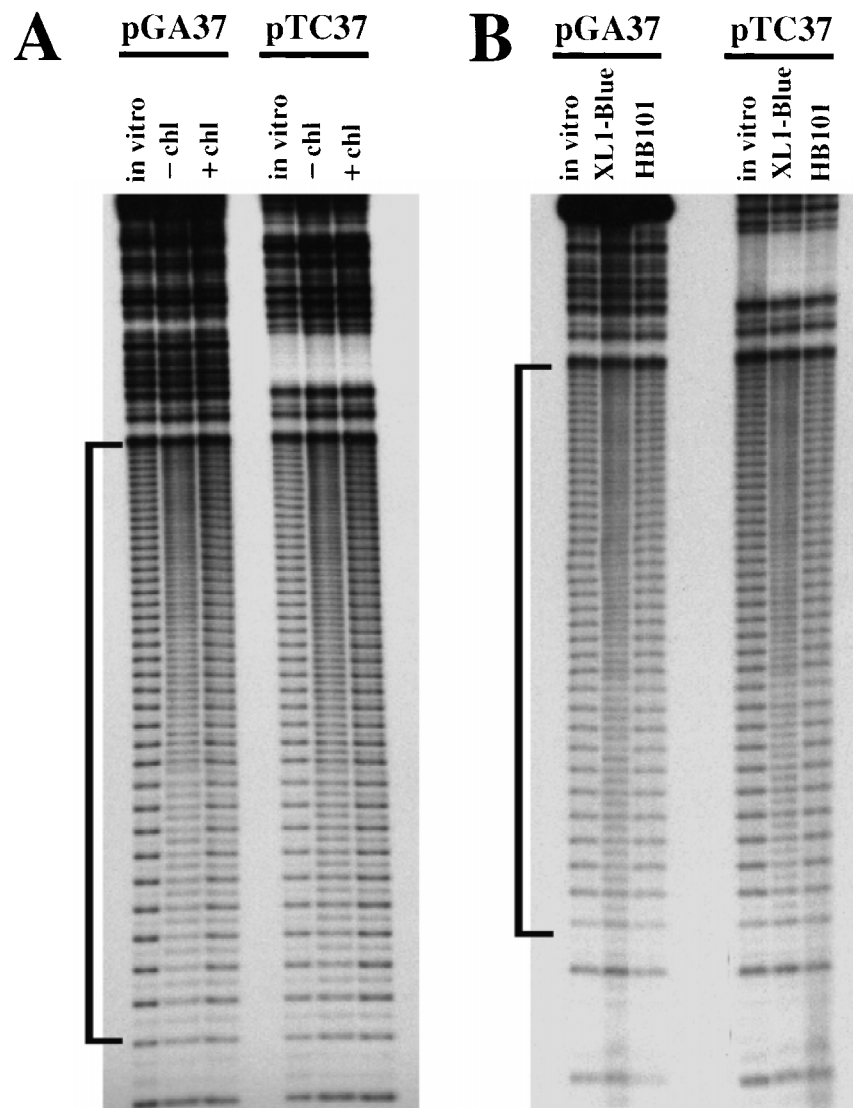
Second, a profound nature of replication blockage indicates that the replication fork experiences problems at  $d(G-A)_n \cdot d(T-C)_n$  runs in a substantial fraction of an intracellular plasmid population. If a protein is responsible for this effect, it must be bound to  $d(G-A)_n \cdot d(T-C)_n$  runs in a substantial fraction of intracellular plasmids. The latter can be detected by *in vivo* footprinting. Cells containing plasmids with the  $d(G-A)_{37} \cdot d(T-C)_{37}$  repeat in both orientations were incubated with dimethylsulfate



**Figure 4.** Chloramphenicol abolishes repeat-caused replication blockage. Plasmids are designated as in Figure 2. Cells were grown to logarithmic phase without chloramphenicol (-) followed by seven hours incubation with chloramphenicol (+). The arrows show replication stop sites.

(DMS) followed by rapid cooling and plasmid DNA isolation. DMS-modified guanine residues were detected by the standard Maxam-Gilbert protocol. Figure 5(b) shows that all guanine bases within the  $d(G-A)_{37} \cdot d(T-C)_{37}$  run are protected against methylation in exponentially growing XL1-Blue, but not HB101, cells. Further, treatment of cells with chloramphenicol prior to the DMS footprinting completely abolishes methylation protection in the XL1-Blue strain (Figure 5(a)). This proves that there is a protein bound to this repeat inside cells. Interestingly, together with guanine protection, there is a visible increase in DMS modification of adenine bases within the repeated run. The latter modification intensifies when the N3-position in adenine bases becomes available for the DMS, i.e. when the DNA minor groove widens.<sup>41</sup> We believe, therefore, that our protein recognizes the  $d(G-A)_n \cdot d(T-C)_n$  repeat *via* the major groove, simultaneously protecting guanine bases against methylation and making the N3 atom of adenine more accessible by widening the minor groove.

Third, we studied whether the  $d(G-A)_n \cdot d(T-C)_n$ -binding activity is present in the cell extracts of various *E. coli* strains. To this end, we used a standard gel retardation assay with linear, end-labeled  $d(G-A)_{37} \cdot d(T-C)_{37}$  fragment as a probe. Figure 6 shows that  $d(G-A)_n \cdot d(T-C)_n$  binding is detected in the XL1-Blue, but not in the HB101, cell extracts. At intermediate amounts of cell extract, one can clearly see a characteristic ladder of retarded DNA fragments. This ladder is likely caused by binding of an increasing number of protein molecules to the same  $d(G-A)_{37} \cdot d(T-C)_{37}$  run. The total number of bands in the ladder corresponds to seven. It is



**Figure 5.** DMS footprinting of the  $d(G-A)_{37} \cdot d(T-C)_{37}$  run *in vivo*. Cells carrying pGA37 or pTC37 plasmids were treated with DMS followed by plasmid DNA isolation and Maxam-Gilbert sequencing. (a) XL1-Blue cells incubated with or without chloramphenicol. (b) Comparison of XL1-Blue and HB101 strains. The reference ladder, i.e. DMS treatment of naked linear pGA37 or pTC37 DNA, is designated *in vitro*. Parentheses show  $d(G-A)_{37} \cdot d(T-C)_{37}$  runs.

plausible to speculate, therefore, that seven protomers of a hypothetical  $d(G-A)_n \cdot d(T-C)_n$ -binding protein can cover the run consisting of 37 repeats at saturation. This leaves us with one protein molecule per 10.6 bp of the repeat, i.e. per DNA helical turn (see also Discussion).

#### **$d(GA)_n \cdot d(TC)_n$ -binding protein is the TraY protein encoded by the F factor**

One major difference between the XL1-Blue and HB101 strains is that the former contains an  $F'$  episome carrying a *lacZ*-complementation cassette. It was suggested, therefore, that a  $d(G-A)_n \cdot d(T-C)_n$ -binding protein might be encoded by the F factor.

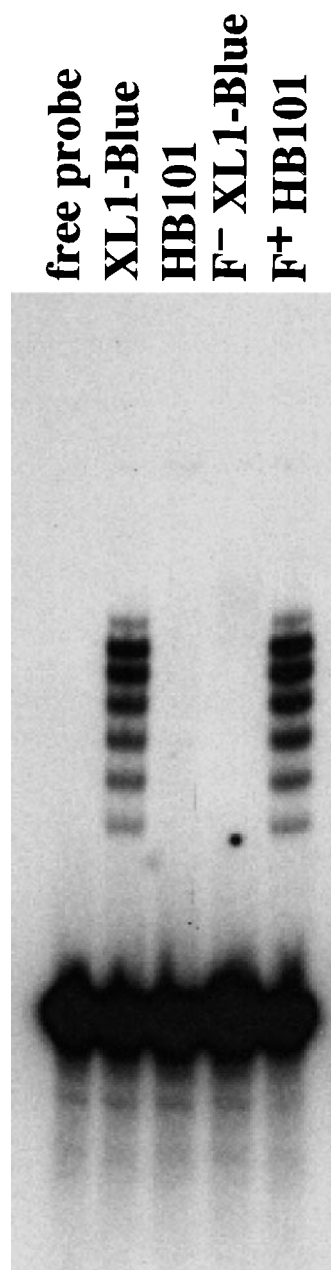
In order to confirm this, we constructed two bacterial strains. The first one was a derivative of the XL1-Blue that was cured from the  $F'$  episome upon

growth in the presence of novobiocin as described.<sup>42</sup> The second strain was a derivative of the HB101 containing  $F'$  episome transferred by conjugation from the XL1-Blue strain. The data on the  $d(G-A)_n \cdot d(T-C)_n$ -binding activity in cell extracts from those strains are shown in Figure 7. It is obvious that the  $F^-$  derivative of XL1-Blue lost the  $d(G-A)_n \cdot d(T-C)_n$ -binding, while the  $F'$  derivative of HB101 acquired this activity. These data decisively prove that the  $d(G-A)_n \cdot d(T-C)_n$ -binding protein is indeed encoded by the F factor.

To determine the nature of the  $d(G-A)_n \cdot d(T-C)_n$ -binding protein, it was purified to near homogeneity from the XL1-Blue strain. Briefly, the crude cell extract was fractionated on the HiTrap heparin column and further purified by two cycles of affinity chromatography on the  $d(G-A)_n \cdot d(T-C)_n$ -coupled CNBr-activated Sepharose. Figure 8 shows



**Figure 6.** Gel-retardation assay reveals  $d(G-A)_n \cdot d(T-C)_n$ -binding activity. End-labeled  $d(G-A)_{37} \cdot d(T-C)_{37}$  probe (1 ng) was incubated with 0.5  $\mu$ l crude cell extracts of XL1-Blue or HB101 strains followed by separation by 5% PAGE.



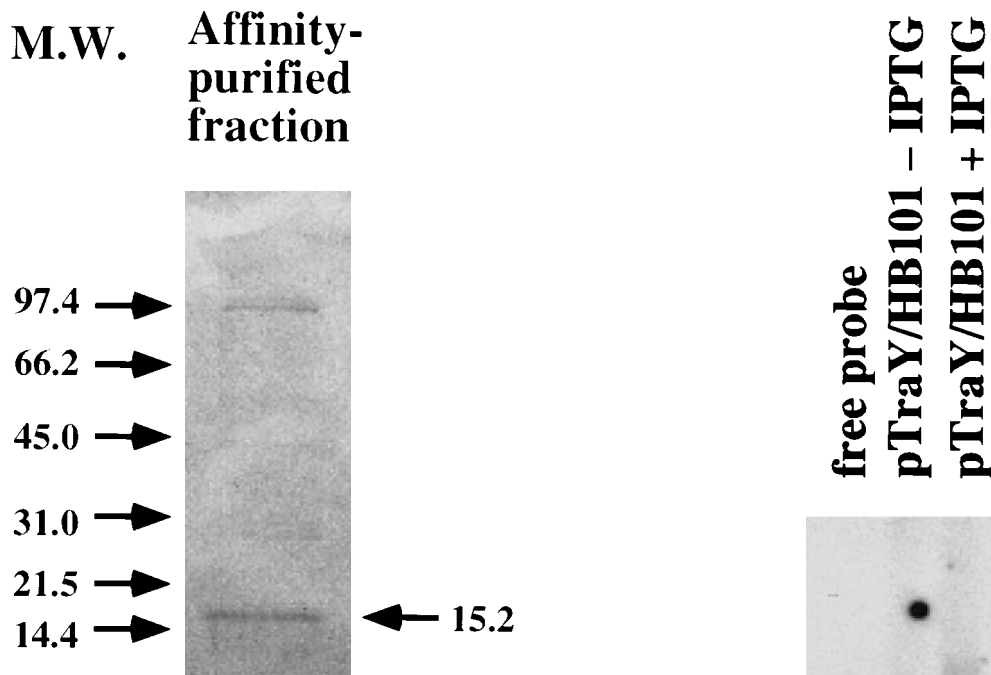
**Figure 7.** Gel-retardation assay confirms that the  $d(G-A)_n \cdot d(T-C)_n$ -binding activity is encoded by the F factor.

that the result of this purification is a major protein band with an approximate molecular mass of 15 kDa. Using mass spectrometric analysis, we were able to refine its molecular mass as 15.2 kDa (data not shown).

Two tryptic peptides of this protein were isolated and microsequenced by Dr Carol Beach at the Macromolecular Structure Analysis Facility at the University of Kentucky. The sequences of these peptides matched precisely the known sequence of the TraY protein that is encoded by the F factor (Figure 9(a)). We conclude, therefore, that our  $d(G-A)_n \cdot d(T-C)_n$ -binding protein is actually the TraY protein.

To confirm this, we PCR-cloned the *traY* open reading frame into the specially designed pACYC 84-derivative under the control of an inducible *trc* promoter. There is certain ambiguity about the translation start site of the TraY protein: an AUG

codon deduced from the DNA sequence or a UUG codon situated 36 bp upstream of it. It is generally believed, however, that the UUG codon is the primary one.<sup>43</sup> A predicted molecular mass for the protein starting from the UUG codon is 15.18 kDa, which is fairly close to our value of 15.2 kDa. Thus, we designed our PCR primers to incorporate the primary UUG codon. Figure 10 shows that overexpression of the recombinant TraY in the HB101 cells supplies this strain with marked  $d(G-A)_n \cdot d(T-C)_n$ -binding activity.



**Figure 8.** Protein content of the  $d(G-A)_n \cdot d(T-C)_n$ -binding fraction. After two cycles of affinity chromatography, proteins in the  $d(G-A)_n \cdot d(T-C)_n$ -binding fraction were separated by gradient (4-20%) SDS-PAGE.

To prove that recombinant TraY alone could cause the replication fork stalling at  $d(G-A)_n \cdot d(T-C)_n$  repeats, we co-transfected HB101 strain

**A.**

```

MKRFGTRSATGKMKVCLKLPVDVESLLIEASNRS
  RSFEAVIR
  ::::::::::
GRSRSFEAVIRLKDHLHRYPKFNRAGNIYGKSL

  LDDETNQLLIAAK
  ::::::::::
KYLTMLDDETNQLLIAAKNRSGWCKTDEAADR

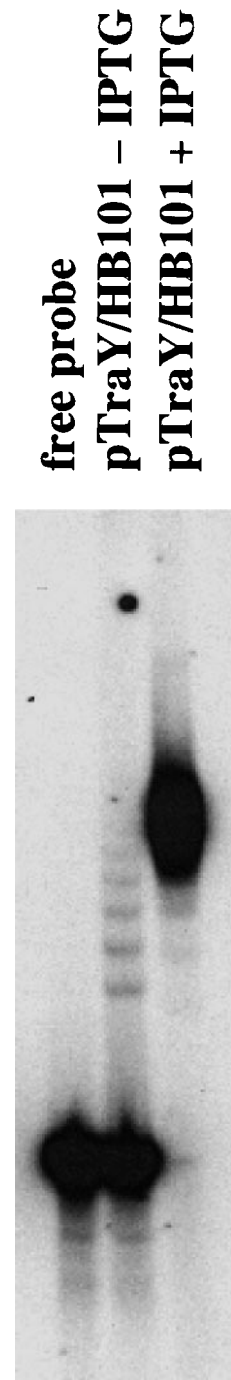
VIDHLIKFPDFYNSEIFREADKEEDITFNTL
    
```

**B.**

```

5' -TAGTTCTCTTACTCTCTTTA-3'
      | | | | |
5' -CTCTCTCTCTCTCTCTCTCTC-3'
    
```

**Figure 9.** Microsequencing of the  $d(G-A)_n \cdot d(T-C)_n$ -binding protein. (a) The complete amino acid sequence of the TraY protein contains two precise matches with the two sequenced tryptic peptides of the  $d(G-A)_n \cdot d(T-C)_n$ -binding protein. (b) The TraY-binding site within the *oriT* shares homology with  $d(G-A)_n \cdot d(T-C)_n$  run.



**Figure 10.** The expression of the recombinant TraY protein supplies the HB101 strain with the  $d(G-A)_n \cdot d(T-C)_n$ -binding activity. A pACYC184-derived plasmid containing the *traY* gene under the control of an inducible *trc* promoter was transformed into HB101 strain. Exponentially grown cells were incubated with or without IPTG followed by crude cell extract preparation and gel-retardation assay with the  $d(G-A)_{37} \cdot d(T-C)_{37}$  probe.

with two plasmids: pACYC 84-derivative containing *traY* gene under the control of the inducible *trc* promoter and pTrc99Δ-derivative containing  $d(G-A)_{37} \cdot d(T-C)_{37}$  repeat. The two plasmids are

compatible, i.e. can co-exist in the same cell. We analyzed the replication fork progression through the repeat in the presence or absence of the recombinant TraY protein. Figure 11 demonstrates that induction of the TraY expression leads to the replication attenuation. We conclude, therefore, that  $d(G-A)_n \cdot d(T-C)_n$ -binding activity of the TraY protein is responsible for the replication blockage at  $d(G-A)_n \cdot d(T-C)_n$  repeat.

## Discussion

Our data unambiguously show that blockage of replication fork progression at the  $d(G-A)_n \cdot d(T-C)_n$  repeat is caused by a protein binding to this repeat, and this protein is TraY. This came as an unexpected result, since TraY-DNA binding was reasonably well characterized.<sup>43–47</sup>

TraY is a DNA-binding protein essential for the F factor conjugal transfer.<sup>48</sup> It binds to the major *tra* operon promoter, up-regulating its transcription.<sup>44</sup> It also binds to the plasmid origin of transfer (*oriT*) where, together with the bacterial integration host factor (IHF), it stimulates the *Tral* protein to nick the *oriT* DNA and initiate conjugal transfer.<sup>44,49</sup> Together with *Arc*, *Mnt* and *Met* repressors, TraY belongs to a family of the so-called ribbon-helix-helix DNA-binding proteins.<sup>50</sup> The crystal structure of repressor/DNA complexes revealed that base-specific contacts are made by two antiparallel  $\beta$ -sheets through the major groove of DNA.<sup>51</sup> The crystal structure of the TraY protein is unknown, but a variety of biochemical and biophysical data indicate that it also recognizes the DNA major groove *via*  $\beta$ -sheet residues.<sup>52</sup>

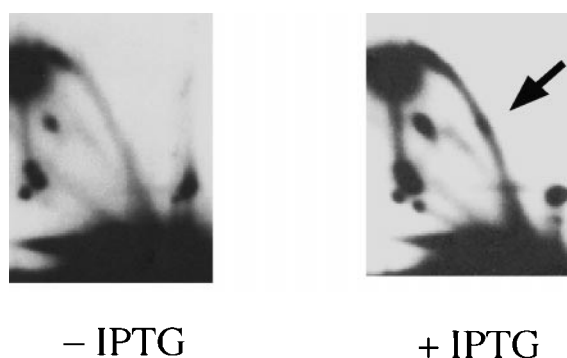
One poorly understood characteristic of TraY-DNA binding is that there is virtually no homology between its two binding sites in the F-plasmid, *tra* promoter and *oriT*. Thus, the TraY

consensus binding sequence is unknown and there is probably more than one. With regard to our studies, it is important to note that the TraY binding site within the *oriT*, as revealed by chemical footprinting, contains the sequence TCTCTTACTCTCT that is fairly close to our  $d(T-C)_n$  repeat (Figure 9(b)). We hypothesize, therefore, that one consensus binding sequence for the TraY protein could correspond to phased  $d(T-C)_n$  runs.

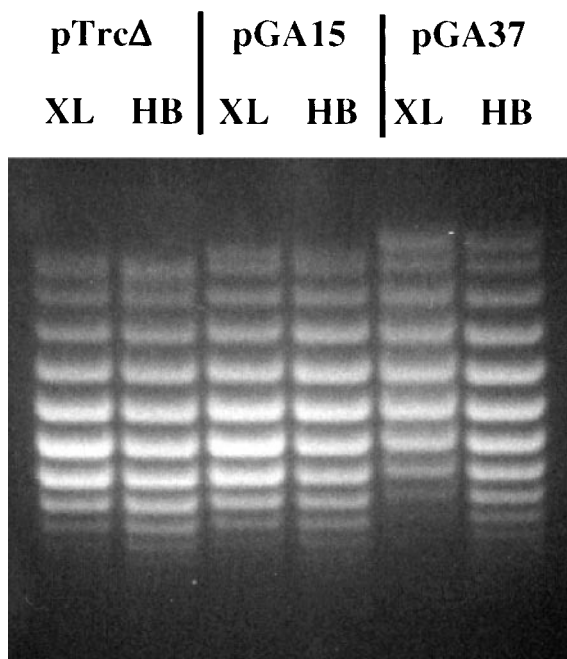
What could be the mechanisms of replication blockage induced by TraY binding to  $d(T-C)_n$  runs? Ribbon-helix-helix repressor proteins bend DNA upon binding.<sup>50</sup> TraY protein bends DNA at an angle of  $50^\circ$ – $55^\circ$ .<sup>45</sup> Concerted binding of multiple TraY protomers to  $d(T-C)_n$  runs could then result in the formation of nucleosome-like structures.

In order to address this opportunity, we performed the following experiments. Plasmids pTrc99 $\Delta$ , pGA15 or pGA37 were isolated from either XL1-Blue, or HB101 strains, and their topological states were analyzed using electrophoretic separation of DNA topoisomers in the presence of chloroquine. We reasoned that, in the HB101 strain, lacking  $d(G-A)_n \cdot d(T-C)_n$ -binding TraY protein, the topoisomer distribution for all three plasmids should be similar. In the XL1-Blue strain the situation could be different. If TraY protein induces wrapping of the repeated DNA, this should affect DNA supercoiling of the whole plasmid. Since DNA supercoiling is tightly regulated by DNA gyrase and Topo I in *E. coli* cells, TraY-mediated changes in plasmid supercoiling would be instantly compensated. Consequently, the distribution of topoisomers for plasmids with and without repeats should differ. Figure 12 shows that our data are in accord with this reasoning. The distribution of topoisomers for all three plasmids is very similar in the HB101 strain, but differs in the XL1-Blue strain. The most pronounced difference was for the plasmid containing a  $d(G-A)_{37} \cdot d(T-C)_{37}$  run, where the center of distribution of topoisomers is shifted 1.0 supercoil in the positive direction. This could be interpreted as the  $d(G-A)_{37} \cdot d(T-C)_{37}$  run wraps around the TraY multimer forming one positive supercoil. Our gel-shift data indicate that seven TraY protomers are bound to this run. Since each of them bends DNA  $50^\circ$ – $55^\circ$ , they could cumulatively curve DNA by  $350^\circ$ – $375^\circ$ , which is fairly closed to one supercoil ( $360^\circ$ ). Figure 13 shows our models where the replication fork stalls upon encountering the nucleosome-like TraY complex with a  $d(G-A)_n \cdot d(T-C)_n$  repeat.

Homopurine-homopyrimidine repeats block DNA synthesis *in vitro* due to the formation of triplexes prior to or during DNA polymerization.<sup>16–18</sup> It has been previously reported that  $d(G-A)_n \cdot d(T-C)_n$  repeats attenuate the replication fork progression in cultured mammalian cells,<sup>29,31</sup> the effect that was also attributed to triplex formation. In light of our data, however, the latter interpretation should be taken with caution. We show that concerted protein binding to the

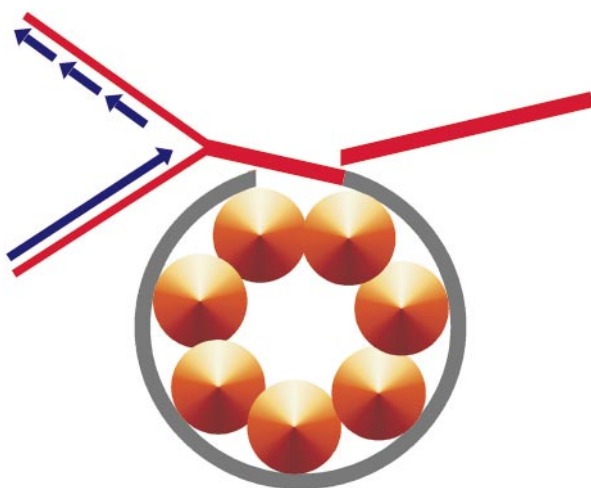


**Figure 11.** The expression of an individual TraY stalls DNA replication at the  $d(G-A)_n \cdot d(T-C)_n$  repeat. HB101 cells were co-transfected with two compatible plasmids: pACYC184, a derivative containing the *traY* gene under the control of an inducible *trc* promoter; and the pTrc-derived pGA37 plasmid. Replication intermediates were isolated from cells growing in the presence or absence of IPTG followed by 2D gel electrophoresis. The arrow shows a TraY-induced replication stop.



**Figure 12.** Comparative electrophoretic analysis of DNA topoisomers of pGAn plasmids isolated from XL1-Blue and HB101 strains. Plasmid DNA was isolated and separated by agarose gel electrophoresis in the presence of 5  $\mu$ g/ml of chloroquine as described.<sup>63</sup>

$d(G-A)_n \cdot d(T-C)_n$  run attenuates replication in bacterial cells. Since there are proteins binding to these runs in mammalian cells as well, for example GAGA-factors,<sup>53</sup> their concerted binding to the repeats could result in replication blockage.



**Figure 13.** Model of replication blockage caused by the TraY binding to the  $d(G-A)_n \cdot d(T-C)_n$  repeat.  $d(G-A)_{37} \cdot d(T-C)_{37}$  run is shown as gray ribbon and flanking DNA as red ribbons. Blue arrows correspond to newly synthesized DNA strands; leading strand, continuous; lagging strand, discontinuous. The repeat is wrapped around seven TraY protomers (gold circles) forming one positive supercoil.

Finally, our data clarify a relatively old story on *H*-DNA formation *in vivo*. Parniewski and co-workers studied a recombinant plasmid where the GATC site was situated between the two  $d(G-A)_7$  runs.<sup>54</sup> *In vitro*, formation of *H*-DNA prevented its methylation by Dam methyltransferase. In some *E. coli* strains, this site was undermethylated *in vivo* as well. It was suggested, therefore, that *H*-DNA formation *in vivo* is responsible for this undermethylation.<sup>54</sup> There were two problems, however, with this interpretation. First, treatment of *E. coli* cells with chloramphenicol completely abolished protection *in vivo*.<sup>54</sup> Second, the Dam undermethylation was observed only in cells carrying *F'* episome, such as JM101, JM105 and SURE, and not in *F*<sup>-</sup> cells, such as HB101, MO611 and RR1.<sup>55</sup> In light of our data, it is reasonable to believe that binding of the TraY protein to  $d(G-A)_7$  runs surrounding the *dam* site physically prevented Dam methyltransferase from reaching its target.

## Materials and Methods

### Plasmids

Plasmids pGAn and pTCn were obtained by cloning different  $d(G-A)_n \cdot d(T-C)_n$  inserts into pTrc99 $\Delta$ 1 and pTrc99 $\Delta$ 2 vectors that differ only in the orientation of a multiple cloning site.<sup>23</sup> The  $d(G-A)_{37} \cdot d(T-C)_{37}$  repeat was cut off the plasmid by *Eco*RI and inserted into the *Eco*RI site of pTrc $\Delta$  vectors.<sup>56</sup> The  $d(G-A)_{21} \cdot d(T-C)_{21}$  repeat was cut off the plasmid as described<sup>57</sup> by *Eco*RI and *Hind*III and inserted between *Eco*RI and *Hind*III sites of pTrc $\Delta$  vectors.  $d(G-A)_{15} \cdot d(T-C)_{15}$  and  $d(G-A)_{30} \cdot d(T-C)_{30}$  were chemically synthesized with cohesive ends for *Eco*RI and *Hind*III and inserted into the corresponding sites of pTrc $\Delta$  vectors.  $d(G-A)_{30} \cdot d(T-C)_{30}$  repeat was also cloned into a non-transcribed area (*Eco*RI site) of the mini-pBR derivative as described.<sup>23</sup>

The TraY coding sequence was obtained by PCR of the XL1-Blue total DNA using the two primers: 5'-CCGGGTACCTAGAGTGTATTAAATGTTATATC-3' and 5'-AGGAATTCGGGAGGTGTTATTGAAAAG-3'. The PCR product was digested with *Eco*RI and *Kpn*I and cloned into *Eco*RI/*Kpn*I sites of the pTrc99A vector.<sup>58</sup> Subsequently, an *Sph*I-fragment of the resultant plasmid, containing the *lacI*<sup>q</sup> gene and the *traY* ORF controlled by the *trc* promoter, was fused to the *Bsa*AI fragment of pACYC184, containing p15A replication ori and *tet*-resistance gene.

### Bacteria

Plasmids were maintained in XL1-Blue (*F*<sup>+</sup>::Tn10, *proA*<sup>+</sup>*B*<sup>+</sup>, *lacI*<sup>q</sup>,  $\Delta(lacZ)M15/recA1$ , *endA1*, *gyrA96*, *thi*, *hsdR17*, *supE44*, *rel1*, *lac*) or HB101 (*F*<sup>-</sup>  $\Delta(gpt-proA)62$ , *leuB6*, *supE44*, *ara-14*, *galk2*, *lacY1*,  $\Delta(mcrC-mrr)$ , *rpsL20*, *xyl-5*, *mtl-1*, *recA13*) *E. coli* K12 strains were obtained from New England Biolabs. For isolation of replication intermediates and DMS footprinting, bacteria were grown in LB medium with 100  $\mu$ g/ml of ampicillin at 37°C until the mid-logarithmic phase ( $A_{600} \sim 0.6$ ). For the experiments with protein synthesis inhibition, cells were grown under the same conditions until early logarithmic stage ( $A_{600} \sim 0.2$ ) followed by the addition of

chloramphenicol (170 µg/ml) and subsequent incubation for four to seven hours.

### Curing XL1-Blue strain from the F'-factor<sup>42</sup>

XL1-Blue cells were grown overnight in LB medium without tetracycline in the presence of various concentrations of novobiocin (Sigma). The presence of 50 µg/ml of novobiocin in the medium led to the partial inhibition of cell growth. Cell cultures grown at this antibiotic concentration were plated on LB agar without tetracycline and individual colonies were screened for tetracycline-sensitivity. Approximately 2% of colonies were *tet*-sensitive due to the F' plasmid loss. They were additionally confirmed to carry chromosomal markers of XL1-Blue, specifically *gyrA96* (nalidixic acid resistance), *lac*<sup>-</sup> (white colonies in the presence of X-Gal and IPTG) and *recA*<sup>-</sup> (UV-sensitivity).

### Bacterial conjugation<sup>59</sup>

Overnight cultures of HB101 and XL1-Blue strains were diluted 100-fold and grown for three hours in LB medium at 37°C with or without shaking, respectively. One milliliter of each culture were mixed together and incubated at 37°C without shaking for one hour. Serial dilutions of this mixture were plated on LB agar containing 10 µg/ml of tetracycline and 100 µg/ml of streptomycin.

### Electrophoretic analysis of replication intermediates

Isolation of replication intermediates and their 2D gel-electrophoretic analysis were performed as described.<sup>32</sup>

### DMS footprinting *in vivo*

Exponential cell cultures were incubated with 0.025% DMS for five minutes at 37°C, cooled, centrifuged and washed with TES buffer, followed by isolation of plasmid DNA with miniprep kit (Promega). Modifications *in vitro* were performed under the same conditions as *in vivo*. Plasmid DNA (0.2 µg) was digested with *Clal* and *BsrDI* (for pGA37) or with *Clal* and *DraI* (for pTC37) and end-labeled with <sup>32</sup>P-dCTP using the Klenow fragment of DNA polymerase I, followed by the isolation of an end-labeled fragment from 5% (w/v) polyacrylamide gel electrophoresis. Labeled DNA fragments were cleaved with 20% (v/v) piperidine (Fisher) for 30 minutes at 100°C. The products were separated by denaturing PAGE (8% (w/v) polyacrylamide) followed by autoradiography and quantitative analysis carried out on the 445 SI Phosphorimager (Molecular Dynamics).

### Preparation of crude cell extracts for gel retardation assay<sup>60</sup>

Cells were grown in 1 ml of LB medium until  $A_{600} \sim 2.0$ , centrifuged and resuspended in 0.2 ml of lysis buffer (20 mM Hepes (pH 7.9), 100 mM KCl, 1 mM EDTA). Cell suspension was frozen for one hour at -70°C and then thawed. Lysozyme (1 mg/ml) was added, the suspension was incubated for five minutes at 37°C and frozen for five minutes at -70°C. This procedure was repeated several times and the lysate was cleared by centrifugation in the Eppendorf Centrifuge at 13,000 g for 15 minutes at 4°C.

### Gel retardation assay

Repeat-containing probe was obtained upon digestion of pGA37 plasmid with *EcoRI*, end-labeling with [<sup>32</sup>P]-dCTP and the Klenow fragment of DNA polymerase I followed by gel purification. Protein binding reactions were performed with 1 ng of end-labeled probe and 0.5 µl of crude extract in 20 mM Hepes (pH 7.9), 40 mM KCl, 0.05 mM DTT, 1 mM EDTA, 0.01% NP-40, 0.1 mg/ml BSA in the presence of 0.5 µg of double-stranded poly(dI-dC) competitor DNA. DNA fragments were resolved by 5% PAGE and visualized by autoradiography.

### DNA affinity column

Synthetic 5'-GATC(GA)<sub>30</sub>-3' and 5'-GATC(TC)<sub>30</sub>-3' oligonucleotides (100 µg) were purified by 8% PAGE. Upon elution they were phosphorylated with phage T4 polynucleotide kinase. Phosphorylated complementary oligonucleotides were co-precipitated, annealed and multimerized by phage T4 DNA ligase. Ligation products were purified by phenol/chloroform-extraction followed by gel-filtration on ProbeQuant G50 Micro column (Amersham Pharmacia Biotech). Coupling of multimerized d(GA)<sub>37</sub> fragments to the CNBr-activated 4B Sepharose (Amersham Pharmacia Biotech) was performed as described.<sup>61</sup>

### Protein purification

An overnight XL1-Blue culture (1 l) was pelleted, washed in 50 ml of ice-cold 20 mM Hepes (pH 7.9), 0.2 mM EDTA, 100 mM KCl, and resuspended in ~3 ml of the same buffer supplemented with 2 mM DTT and one-third of a tablet of complete mini-protease inhibitor (Boehringer Mannheim). Cells were lysed by sonication using microtip in Sonabox (Artek). The lysate was cleared for 15 minutes at 13,000 g in a cold Eppendorf Centrifuge and supernatant was loaded onto a 1 ml HiTrap heparin column (Amersham Pharmacia Biotech) equilibrated with purification buffer (20 mM Hepes (pH 7.9), 0.2 mM EDTA, 10% (w/v) glycerol, 2 mM DTT) containing 0.1 M KCl. Upon washing the column with three volumes of the purification buffer with 0.1 M KCl and three volumes of the purification buffer with 0.25 M KCl, the protein was eluted with purification buffer containing 0.5 M KCl and protease inhibitors: 0.5 ml fractions were collected and analyzed for d(GA)<sub>n</sub>·d(TC)<sub>n</sub>-binding activity by gel retardation assay. Fractions containing the GA-binding activity were mixed together, diluted four times in purification buffer and incubated for 30 minutes at 4°C with 500 µg of double-stranded poly(dI-dC) competitor DNA (Sigma). One milliliter of the affinity resin was then added, the mixture was incubated overnight, the resin was transferred into the column, washed with 6 ml of 0.1 M KCl purification buffer, 3 ml of 0.2 M KCl purification buffer and the protein was eluted with 0.4 M KCl purification buffer: 0.2 ml fractions were collected and analyzed for d(GA)<sub>n</sub>·d(TC)<sub>n</sub>-binding activity by gel-retardation assay. Proteins in 100 µl from the final d(GA)<sub>n</sub>·d(TC)<sub>n</sub>-binding fraction were separated by gradient (4-20%) SDS-PAGE (BioRad). The gel was stained with Coomassie brilliant blue as described.<sup>62</sup> The major protein band was excised from the gel and sent for protein microsequencing.

## Acknowledgments

We are indebted to Elena Siyanova, who was the first to suggest that our replication data could be explained by protein binding rather than triplex formation. We thank Elena Davydova for her invaluable help in TraY purification, George Samadashwily for his input in the early stage of these experiments, Pradip Raychadhury, Pavel Shiyonov and Tatiana Shiyonova for their help with chromatography, Carol Beach for expert protein sequencing, Igor Panyutin for helpful discussions and Gerald Buldak for proofreading the manuscript. This work was supported by grant GM60987 from the NIH to S.M.M. E.V.S. is supported by a graduate fellowship from the University of Illinois at Chicago.

## References

- Schroth, G. P. & Ho, P. S. (1995). Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucl. Acids Res.* **23**, 1977-1983.
- Cox, R. & Mirkin, S. M. (1997). Characteristic enrichment of DNA repeats in different genomes. *Proc. Natl Acad. Sci. USA*, **94**, 5237-5242.
- Modrich, P. (1994). Mismatch repair, genetic stability, and cancer. *Science*, **266**, 1959-1960.
- Kunkel, T. A. (1993). Slippery DNA and diseases. *Nature*, **365**, 207-208.
- Wells, R. D. & Warren, S. T. (1998). *Genetic Instabilities and Hereditary Neurological Disorders*, Academic Press, San Diego.
- McMurray, C. T. (1995). Mechanisms of DNA expansion. *Chromosoma*, **104**, 2-13.
- Wells, R. D. (1996). Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem.* **271**, 2875-2878.
- Siyanova, E. Y. & Mirkin, S. M. (2001). Expansion of trinucleotide repeats. *Mol. Biol. (Moscow)*, **35**, 168-182.
- Sherman, L. A. & Gefter, M. L. (1976). Studies of the mechanism of enzymatic DNA elongation of by *Escherichia coli* DNA polymerase II. *J. Mol. Biol.* **103**, 61-76.
- Chalberg, M. D. & Englund, P. T. (1979). The effect of template secondary structure on vaccinia DNA polymerase. *J. Biol. Chem.* **254**, 7820-7826.
- Huang, C. C. & Hearst, J. E. (1980). Pauses at positions of secondary structure during in vitro replication of single-stranded fd bacteriophage DNA by T4 DNA polymerase. *Anal. Biochem.* **103**, 127-139.
- Kaguni, L. S. & Clayton, D. A. (1982). Template-directed pausing in in vitro DNA synthesis by DNA polymerase  $\alpha$  from *Drosophila melanogaster* embryos. *Proc. Natl Acad. Sci. USA*, **79**, 983-987.
- Weaver, D. T. & DePamphilis, M. L. (1982). Specific sequences in native DNA that arrest synthesis by DNA polymerase alpha. *J. Biol. Chem.* **257**, 2075-2086.
- Abbotts, J., SenGupta, D. N., Zon, G. & Wilson, S. H. (1988). Studies on the mechanisms of *Escherichia coli* DNA polymerase I large fragment. Effect of template sequence and substrate variation on termination of synthesis. *J. Biol. Chem.* **263**, 15094-15103.
- Lapidot, A., Baran, N. & Manor, H. (1989). (dT-dC)<sub>n</sub> and (dG-dA)<sub>n</sub> tracts arrest single stranded DNA replication *in vitro*. *Nucl. Acids Res.* **17**, 883-900.
- Baran, N., Lapidot, A. & Manor, H. (1991). Formation of DNA triplexes accounts for arrests of DNA synthesis at d(TC)<sub>n</sub> and d(GA)<sub>n</sub> tracts. *Proc. Natl Acad. Sci. USA*, **88**, 507-511.
- Dayn, A., Samadashwily, G. M. & Mirkin, S. M. (1992). Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization. *Proc. Natl Acad. Sci. USA*, **89**, 11406-11410.
- Samadashwily, G. M., Dayn, A. & Mirkin, S. M. (1993). Suicidal nucleotide sequences for DNA polymerization. *EMBO J.* **12**, 4975-4983.
- Woodford, K. J., Howell, R. M. & Usdin, K. (1994). A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J. Biol. Chem.* **269**, 27029-27035.
- Usdin, K. & Woodford, K. J. (1995). CGG repeats associated with DNA instability and chromosome fragility from structures that block DNA synthesis *in vitro*. *Nucl. Acids Res.* **23**, 4202-4209.
- Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S. & Wells, R. D. (1995). Pausing of DNA synthesis *in vitro* at specific loci in CTG and CGG triplet repeats from human hereditary disease genes. *J. Biol. Chem.* **270**, 27014-27021.
- Ohshima, K. & Wells, R. D. (1997). Hairpin formation during DNA synthesis primer realignment *in vitro* in triplet repeat sequences from human hereditary disease genes. *J. Biol. Chem.* **272**, 16798-16806.
- Samadashwily, G. M., Raca, G. & Mirkin, S. M. (1997). Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.* **17**, 298-304.
- Samadashwily, G. M. & Mirkin, S. M. (1994). Trapping DNA polymerases using triplex-forming oligodeoxyribonucleotides. *Gene*, **149**, 127-136.
- Kopel, V., Pozner, A., Baran, N. & Manor, H. (1996). Unwinding of the third strand of a DNA triple helix, a novel activity of the SV40 large T-antigen helicase. *Nucl. Acids Res.* **24**, 330-335.
- Baran, N., Pucshansky, L., Marco, Y., Benjamin, S. & Manor, H. (1997). The SV40 large T-antigen helicase can unwind four stranded DNA structures linked by G-quartets. *Nucl. Acids Res.* **25**, 297-303.
- Fry, M. & Loeb, L. A. (1999). Human Werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)<sub>n</sub>. *J. Biol. Chem.* **274**, 12797-12802.
- Kamath-Loeb, A. S., Loeb, L. A., Johansson, E., Burgers, P. M. & Fry, M. (2001). Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitates copying of tetraplex and hairpin structures of the d(CGG)<sub>n</sub> trinucleotide repeat sequence. *J. Biol. Chem.* **276**, 16439-16446.
- Rao, S., Manor, H. & Martin, R. G. (1988). Pausing in simian virus 40 DNA replication by a sequence containing (dG-dA)<sub>27</sub>·(dT-dC)<sub>27</sub>. *Nucl. Acids Res.* **16**, 8077-8094.
- Brinton, B. T., Caddle, M. S. & Heintz, N. H. (1991). Position and orientation-dependent effects of a eukaryotic Z-triplex DNA motif on episomal DNA replication in COS-7 cells. *J. Biol. Chem.* **266**, 5153-5161.
- Rao, B. S. (1994). Pausing of simian virus 40 DNA replication fork movement *in vivo* by (dG-dA)<sub>n</sub>·(dT-dC)<sub>n</sub> tracts. *Gene*, **140**, 233-237.
- Krasilnikova, M. M., Samadashwily, G. M., Krasilnikov, A. S. & Mirkin, S. M. (1998). Transcrip-

- tion through a simple DNA repeat blocks replication elongation. *EMBO J.* **17**, 5095-5102.
33. Ohshima, K., Montermini, L., Wells, R. D. & Pandolfo, M. (1998). Inhibitory effects of expanded GAA·TTC triplet repeats from intron I of the Friedreich's ataxia gene on transcription and replication *in vivo*. *J. Biol. Chem.* **273**, 14588-14595.
  34. Manor, H., Sridhara-Rao, B. & Martin, R. G. (1988). Abundance and degree of dispersion of genomic d(GA)<sub>n</sub>·d(TC)<sub>n</sub> sequences. *J. Mol. Evol.* **27**, 96-101.
  35. Mirkin, S. M. & Frank-Kamenetskii, M. D. (1994). H-DNA and related structures. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 541-576.
  36. Mirkin, S. M. (1999). Structure and biology of H DNA. In *Triple Helix Forming Oligonucleotides* (Malvy, C., Harel-Bellan, A. & Pritchard, L. L., eds), pp. 193-222, Kluwer Academic Publishers, Boston.
  37. Perucho, M., *et al.* (1994). Defects in replication fidelity of simple repeated sequences reveal a new mutator mechanism for oncogenesis. *Cold Spring Harbor Symp. Quant. Biol.* **59**, 339-348.
  38. Friedman, K. L. & Brewer, B. J. (1995). Analysis of replication intermediates by two-dimensional agarose gel electrophoresis. *Methods Enzymol.* **262**, 613-627.
  39. Brewer, B. J. & Fangman, W. L. (1987). The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell*, **51**, 463-471.
  40. Clewell, D. B. (1972). Nature of Col E1 plasmid replication in *Escherichia coli* in the presence of chloramphenicol. *J. Bacteriol.* **110**, 667-676.
  41. Kim, M. G. & Camerini-Otero, R. D. (1997). An alteration in the structure of the minor groove of duplex DNA induced by the formation of the intermolecular d(GA):d(GA)·d(TC)<sub>n</sub> triplex. *Mol. Cells*, **7**, 641-647.
  42. Wolfson, J. S., Hooper, D. C., Swartz, M. N., Swartz, M. D. & McHugh, G. L. (1983). Novobiocin-induced elimination of F<sup>+</sup>lac and mini-F plasmids from *Escherichia coli*. *J. Bacteriol.* **156**, 1165-1170.
  43. Lahue, E. E. & Matson, S. W. (1990). Purified *Escherichia coli* F-factor TraY protein binds oriT. *J. Bacteriol.* **172**, 1385-1391.
  44. Nelson, W. C., Morton, B. S., Lahue, E. E. & Matson, S. W. (1993). Characterization of the *Escherichia coli* F factor traY gene product and its binding sites. *J. Bacteriol.* **175**, 2221-2228.
  45. Luo, Y. & Deonier, R. C. (1994). Mutational and physical analysis of F plasmid traY protein binding to oriT. *Mol. Microbiol.* **11**, 459-469.
  46. Nelson, W. C. & Matson, S. W. (1996). The F plasmid traY gene product binds DNA as a monomer or a dimer: structural and functional implications. *Mol. Microbiol.* **20**, 1179-1187.
  47. Schildbach, J. F., Robinson, C. R. & Sauer, R. T. (1998). Biophysical characterization of the TraY protein of *Escherichia coli* F factor. *J. Biol. Chem.* **273**, 1329-1333.
  48. Lanka, E. & Wilkins, B. M. (1995). DNA processing reactions in bacterial conjugation. *Annu. Rev. Biochem.* **64**, 141-169.
  49. Howard, M. T., Nelson, W. C. & Matson, S. W. (1995). Stepwise assembly of a relaxosome at the F plasmid origin of transfer. *J. Biol. Chem.* **270**, 28381-28386.
  50. Raumann, B. E., Brown, B. M. & Sauer, R. Y. (1994). Major groove DNA recognition by β-sheets: the ribbon-helix-helix family of gene regulatory proteins. *Curr. Opin. Struct. Biol.* **4**, 36-43.
  51. Raumann, B. E., Rould, M. A., Pabo, C. O. & Sauer, R. T. (1994). DNA recognition by β-sheets in the Arc repressor-operator crystal structure. *Nature*, **367**, 754-757.
  52. Lum, P. L. & Schildbach, J. F. (1999). Specific DNA recognition by F factor TraY involves β-sheet residues. *J. Biol. Chem.* **274**, 19644-19648.
  53. Wilkins, R. C. & Lis, J. T. (1997). Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation. *Nucl. Acids Res.* **25**, 3963-3968.
  54. Parniewski, P., Kwinkowski, M., Wilk, A. & Klysik, J. (1990). Dam methyltransferase sites located within the loop region of the oligopurine-oligopyrimidine sequences capable of forming H-DNA are undermethylated *in vivo*. *Nucl. Acids Res.* **18**, 605-611.
  55. Klysik, J. (1996). Both an altered DNA structure and cellular proteins are involved in protecting a triplex forming an oligopurine-rich sequence from Dam methylation in *E. coli*. *Biochem. Genet.* **34**, 165-178.
  56. Panyutin, I. G. & Wells, R. D. (1992). Nodule DNA in the (GA)<sub>37</sub>·(CT)<sub>37</sub> insert in superhelical plasmids. *J. Biol. Chem.* **267**, 5495-5501.
  57. Lyamichev, V. I., Mirkin, S. M., Kumarev, V. P., Baranova, L. V., Vologodskii, A. U. & Frank-Kamenetskii, M. D. (1989). Energetics of the B-H transition in supercoiled DNA carrying d(CT)<sub>x</sub>·d(AG)<sub>x</sub> and d(C)<sub>n</sub>·d(G)<sub>n</sub> inserts. *Nucl. Acids Res.* **17**, 9417-9423.
  58. Aman, E., Ochs, B. & Abel, K.-J. (1988). Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene*, **69**, 301-315.
  59. Miller, J. H. (1992). *A Short Course in Bacterial Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
  60. Boiteux, S., O'Connor, T. R. & Laval, J. (1987). Formamidopyrimidine-DNA glycosylase of *E. coli*. *EMBO J.* **6**, 3177-3183.
  61. Kerrigan, L. A. & Kadonaga, J. T. (1989). DNA protein interactions. In *Current Protocols in Molecular Biology* (Ausubel, F. M., *et al.*, eds), pp. 12.10.1-12.10.8, Green Publishing Associates and Wiley-Interscience, New York.
  62. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
  63. Krasilnikov, A. S., Podtelezchnikov, A., Vologodskii, A. & Mirkin, S. M. (1999). Large-scale effects of transcriptional DNA supercoiling *in vivo*. *J. Mol. Biol.* **292**, 1149-1160.

Edited by I. Tinoco

(Received 2 July 2001; received in revised form 30 August 2001; accepted 31 August 2001)