Revised Manuscript text only

Accepted for BBRC original article

Title:
Detection of cellular G-quadruplex by using a loop structure as a structural determinant

Authors:
***Hisao Masai, Yutaka Kanoh, Naoko Kakusho and Rino Fukatsu**

Affiliations:
Department of Genome Medicine, Tokyo Metropolitan Institute of Medical Science, Kamikitazawa, Setagaya-ku, Tokyo 156-8506, Japan

Correspondence to
Dr. Hisao Masai
Telephone: +81-3-5316-3220
Fax: +81-3-5316-3145
E-mail: masai-hs@igakuken.or.jp

Highlights

G4 formation in vitro could be analyzed by examining the resistance to restriction enzyme cleavage in the modified loop segment of G4-forming sequences.

G4 generated on duplex DNA by heat denaturation or transcription equally exhibited restriction enzyme resistance.

This method was applied to detect G4 formation at the cellular model G4 sequences, and restriction enzyme resistance was detected at the expected G4 forming sequences.

Abstract

G-quadrupex is now known to play crucial roles in various biological reactions. However, direct evidence for its presence in cells has been limited, due to the lack of versatile and non-biased methodology. We use Rif1 binding sites on the fission yeast genome, which has been shown to adopt G4 structures, as a model to prove that Rif1 BS indeed adopt G4 structure in cells. We take advantage of the presence of a single-stranded loop in the G4 structure. Rif1BS is unique in that they contain unusually long loop sequences, and we replace them with a 18 bp I-SceI restriction site. We show *in vitro* that I-SceI in the loop is not cleaved when G4 is formed on duplex Rif1BS DNA, but is cleaved when G4 is not formed due to a mutation in the G-tracts. This is observed both heat-induced and transcription-induced G4 structure, and gives proof of evidence for this procedure. We apply this strategy for detection of a G4 structure at the same Rif1BS in fission yeast cells. We present evidence that in vivo cleavage of I-SceI can be a measure for the presence of G4 at the target sequence in cells as well. The method described here gives a platform strategy for genome-wide analyses of cellular G4 and their dynamic formation and disruption.

Introduction

G-quadruplex (G4) can be formed on guanine-rich nucleic acids. Typical G4 forming sequences contain four repeats of guanine-tracts, that form layers of G-quartet structures. G4 can be generated both on DNA and RNA [1-5], and sometimes on RNA-DNA hybrid [6-9]. The bioinformatics analyses of G4 predicted 375,000 G4 on the human genome [10,11], and another method "G4-seq", which is based on the G4's ability to physically block the chain elongation by the DNA polymerase, showed the presence of over 710,000 potential G4 [12]. On the other hand, the ChIP analyses with G4-specific antibody, BG4, revealed 1,496 and 10,560 G4 peaks in non-cancer cells (NHEK, normal human epidermal keratinocytes) and in cancer cells (HaCaT, spontaneously immortalized counterparts of NHEK), respectively [13]. Significantly lower numbers of in vivo G4 compared to predicted numbers of G4 on the human genome may indicate that only a subset of the predicted G4-forming sequences actually form G4 in cells. Alternatively, BG4 may be capable of detecting only specific forms of G4. Thus, in order to gain a comprehensive view on the profiles of cellular G4s, new methods and tools for detecting G4 in cells are being awaited.

Rif1 is a G4 binding protein and 35 strong Rif1 binding sequences, that can form G4 in vitro, were identified on the fission yeast genome [14,15]. On the other hand, 446 potential G4-forimg sequences were identified on the fission yeast genome [16]. Unexpectedly, only two of the 35 Rif1 binding sequences overlap with these predicted G4. This indicates that informatics prediction with the typical consensus sequence cannot accurately reflect the *in vivo* occurrence of G4 and the sequences diverged from the consensus can form G4 in cells. This was also indicated by identification of G4 by "G4 seq" that were not predicted from the bioinformatics analyses [12]. Thus, it is an important issue to obtain a precise view on cellular G4 profiles and their dynamics.

A G4 structure contains a loop segment that connects the planar G-quartets and appears as single stranded DNA. The loop sequences and length may affect topology and stability of G-quadruplex [17,18]. The Rif1 binding sequences contain multiple copies of G-tracts (with 5 or 6 consecutive guanines) with potentially long loop sequences (up to 29 nucleotide). This segment indeed appeared as single-stranded DNA, as shown by

mapping by S1 nuclease, in the structure generated on duplex DNA by heat treatment [19]. We surmised that the restriction site present in a loop sequence will not be cleaved by the enzyme if G4 is formed. We took advantage of the long loop of Rif1BS and replaced the loop sequence with a 18 bp I-SceI restriction site. The I-SceI site would not be cleaved by I-SceI when G4 is formed, but would be cleaved when G4 is disrupted. We generated G4 on duplex DNA either by heat treatment or transcription. *In vitro* experiments showed that our prediction was warranted. We then applied this methodology to evaluate the presence of G4 at cellular Rif1BS. By LM-PCR (linker-mediated PCR), we were able to detect the cleavage of I-SceI *in vivo* and have shown that the functional Rif1BS is resistant to the cleavage in spite of the presence of the I-SceI site. We will discuss potential applications of this method to more comprehensive analyses of G4 on the cellular genome.

Results

Replacement of a long loop with I-SceI in Rif1BS does not affect its ability to bind to Rif1 and suppress origins

Rif1 binds to selected intergenic segments on the fission yeast genome (Rif1BS containing G-rich conserved sequences, termed Rif1CS [Rif1 consensus sequence]). Furthermore, we reported that sequences derived from Rif1BS can form G-quadruplex (G4) like structures and Rif1 preferentially binds to G4-like structures generated by heat denaturation/ annealing [15]. Rif1BS are characterized by the presence of long loop sequences which would protrude from G4 as a single-stranded DNA [19].

A representative Rif1BS, $Rif1BS_{I:2663}$, contains 23 nucleotide putative loop sequences (**Fig. 1**). Indeed, we showed this sequence on duplex DNA is sensitive to S1 nuclease upon heat denaturation [19]. We replaced a part of the loop sequences with a 18 bp I-SceI restriction site (Rif1BSI [I-SceI]wt). This modification does not affect the ability of $Rif1BS_{I:2663}$ to form G4 after heat denaturation/ renaturation, as indicated by the fast-migrating form (a monomer G4) and slow-migrating forms (oligomeric forms of G4), and to bind to Rif1 *in vitro* (**Supplementary Fig. S1**).

Heat denaturation of duplex DNA containing Rif1BS causes resistance to I-SceI digestion

The G-tracts in Rif1BS$_{I:2663}$ are essential for formation of G4, and mutations in the G-tracts disrupt G4 formation [15]. For the use as control DNA, the same I-SceI replacement was made on a mutant version of Rif1BS$_{I:2663}$ carrying base substitutions in G-tracts of the two Rif1CS (Rif1BSI[I-SceI]mut). The mutation causes the sequences to be incapable of forming G4, and not to be bound by Rif1 in vitro (**Supplementary Figure S1**). Duplex DNAs of these sequences were heat-denatured or non-treated and digested by I-SceI and run on denaturing polyacrylamide gel.

The non-treated wild-type sequence (Rif1BSI[I-SceI]wt) was cleaved by I-SceI, but was not cleaved after heat denaturation (**Fig. 1**, lanes 2 and 1). On the other hand, the mutant sequence (Rif1BSI[I-SceI]mut) was cleaved by I-SceI with or without heat-denaturation (**Fig. 1**, lanes 3 and 4). These results support the prediction that the loop segment in Rif1BS$_{I:2663}$ -derived G4 is in a single-stranded form and resistant to cleavage by a restriction enzyme.

Transcription of the duplex DNA containing Rif1BS also generates a structure that is resistant to I-SceI digestion.

We have generated a duplex DNA containing Rif1BSI [I-SceI]wt or Rif1BSI[I-SceI]mut sequence under the 19 bp T7 promoter (**Fig. 2A**). This DNA was transcribed *in vitro* by T7 RNA polymerase in the presence and absence of PEG200, which is known to facilitate the G4 formation through its molecular crowding effect [20]. Transcription in the presence of PEG200 reduced the extent of the cleavage by I-SceI, whereas that in the absence of PEG200 did not affect the cleavage (**Fig. 2B**, lanes 3 and 4). The G-tract mutant DNA, similarly transcribed in the presence of PEG200, was digested by I-SceI to a slightly greater extent compared to the wild-type (**Fig. 2D**, lanes 3 and 6). We also made a similar template DNA containing an EcoRI cleavage site in the loop (**Fig. 2A**). We confirmed that the insertion of EcoRI site does not affect the G4 formation and binding of Rif1 (**Supplementary Figure S1**). Cleavage by EcoRI was inhibited on the template transcribed in the presence of PEG200, whereas the G-tract mutant was cleaved to much greater extent even after transcription with PEG200 (**Fig. 2C**, lane 4; **Fig. 2E**, compare lanes 3 and 6). These data suggest that transcription in the presence of PEG200 on G-rich Rif1BS sequence can generate structures similar to G4 in which the putative loop sequence is resistant to restriction digestion.

A large amount of slow migrating forms appeared on an acrylamide gel after transcription in the presence of PEG200, but not in its absence (**Supplementary Fig. S2**, lanes 5,6,11,12,17, and 18). The appearance of the slow migrating forms were largely sensitive to RNaseH treatment (**Fig. 3A**, lanes 1 and 4), but largely resistant to RNaseA+RNaseT1 (**Fig. 3A**, lanes 2 and 5). The treatment with RNaseA+T1 did not change the sensitivity to the restriction enzyme, whereas the RNaseH digestion restored the sensitivity (**Fig. 3B**, lanes 1-3, 7-9). These results indicate that transcription of Rif1BS sequence also generates a structure containing RNA-DNA hybrid that may be a part of this putative transcription-induced G4 structure, as has been indicated before [9].

The loop segment of Rif1BS is protected from I-SceI digestion in cells

We next asked whether the above experimental results can be applicable to the detection of G4 in cells. We introduced I-SceI at the same loop segment of $Rif1BS_{I:2663}$ in the wild-type and mutant G-tract background on the chromosome (**Fig. 4A**). We then expressed I-SceI on a plasmid under inducible nmt1 promoter, which is activated by depletion of thiamine in the medium. I-SceI can be identified by western blotting at 12 hrs after induction (**Fig. 4B**). We introduced Rad52-YFP into the assay strains, and counted Rad52 foci as a measure for the I-SceI cleavage at 18 hrs after induction. 42.3% of the cells with the I-SceI site at non-G4 (II:1542kb) were positive for Rad52 foci, while only 15.6% of those with I-SceI at the wild-type $Rif1BS_{I:2663}$ were positive. The value was 33.1 % with I-SceI at the G-tract mutant $Rif1BS_{I:2663}$ (**Fig. 4C and D**). The I-SceI at non-Rif1BS was cleaved as early as at 12 hrs, but that at the wild-type $Rif1BS_{I:2663}$ was not cleaved until 14 hrs after induction (**Supplementary Figure S3**).

Digestion was measured also by linker-mediated PCR (LM-PCR) (**Fig. 4E**). The I-SceI at non-Rif1BS (non-G4) and at $Rif1BS_{I:2663}$ wt started to be cleaved at 12 hrs after induction of I-SceI. The cleavage of I-SceI at $Rif1BS_{I:2663}$ wt started to decrease at 16 hrs, whereas that at the non-Rif1BS (non-G4) stayed high until 18hrs after induction. (**Fig. 4F**). These results indicate that G4 is indeed formed at the $Rif1BS_{I:2663}$ wt in cells and prevents digestion of the I-SceI site in the loop.

Discussion

G-quadruplex is a prevalent non-B type DNA/RNA structure which is now known to play diverse roles in various chromosome transactions and biological reactions [21].

Various studies through mutagenesis of the putative G4 sites and use of G4-specific ligands have strongly indicated the presence of G4 and their critical roles in diverse cellular reactions [22]. However, physical evidence for the presence of G4 within cells is rather limited. The use of a polypeptide antibody for G4 (BG4) indicated the presence of 1,496 and 10,560 G4 in non-cancer cells (NHEK, normal human epidermal keratinocytes) and in cancer cells (HaCaT, spontaneously immortalized counterparts of NHEK)[13]. BG4 was also used to visualize the cellular G4 in human cells [23]. However, topology-specificity of BG4 is rather biased and would not detect all G4 in cells (our unpublished result). Fluorescent G-quadruplex ligands were also used to detect the G4 formation in vivo [24]. Again, in this case, the binding specificity of the G4 ligands used would affect the results. The precise understanding of the presence of G4 and its dynamics within cells is essential for clarifying the entire picture of cellular G4. Considering the diversity of structures/ topology of G4 and their binding proteins, varieties of methods based on different principles need to be employed.

A novel method for evaluation of G4 formation

We describe here a new method for evaluating the presence of G4 in cells that is based on its structural feature. The most notable feature of G4 is the presence of single-stranded loops, which were previously assumed to be 1~7 nucleotide long [10,11]. However, Rif1BS, the G4 structures bound by Rif1 in fission yeast, are unique in that they can have a long loop segment [15,19]. We introduced an I-SceI restriction site in this loop segment and examined whether G4 formation affects the cleavage. The duplex DNA containing the $Rif1BS_{I:2663}$ wt was heat denatured and G4 was formed. The effect of heat treatment on I-SceI digestion was examined. As expected, heat denaturation rendered the I-SceI in the loop resistant to digestion. Disruption of G4 by loss of G-tracts restored I-SceI cleavage. Similarly, transcription of the $Rif1BS_{I:2663}$ wt inhibited the cleavage of the I-SceI. The inhibition depends on the G-tracts and formation of RNA-DNA hybrids, consistent with the transcription-driven formation of hybrid G4 on RNA-DNA hybrid [6-9].

These results give proof of concept for potential use of loop sequences for detection of cellular G4. Indeed, the cellular G4 modified to contain an I-SceI site in the loop could be detected by its resistance to I-SceI digestion. The I-SceI site was cleaved

when it was in non-G4 sequence region or when the G-tract in the Rif1BS was mutated. We call this "Noncleavable-G4-loop" method for G4 detection.

Previously, Zheng Tan's group reported the use of restriction digestion for identification of G-quadruplex formation in supercoiled plasmids [25]. These authors also provided evidence for in vivo formation of G4 by analyzing premature transcription termination caused by the formed G4 in bacterial cells. In this report, we introduced a restriction site specifically in the loop segments without affecting the ability of the target G4-forming sequences to generate G4, and used the restriction cleavage in vivo as a readout for G4 formation in yeast cells. Chemical footprinting using permanganate and S1 nuclease in conjunction with high-throughput sequencing was previously employed to detect endogenous G4 in live human cells [26]. This method would detect single-stranded DNA segments in general, and our method may be more specific to G4 structure.

Potential application of "Noncleavable-G4-loop" method

One of the unanswered questions regarding cellular G4 is how the formation of cellular G4 is regulated. Factors affecting the formation and stability of G4 would include local superhelicity (negative supercoiling) [27], epigenetic state of DNA [28], the levels of RNaseH or G4 helicase, the extent of transcription or DNA replication, cell cycle stage and so on. Now with a relatively easy and reliable method for detecting cellular G4 in hand, one can examine how G4 formation is affected under a specific genetic background or in the presence of various drugs.

The method can be also applicable to monitoring of the dynamics of cellular G4 by using fluorescence-tagged catalytically dead I-SceI which can bind to the I-SceI site but cannot cleave it. Our results show that a 6-bp restriction enzyme can be also used for detection of a loop in G4. We are now trying to develop a method that can be used for detection of cellular G4 without need to incorporate an I-SceI site. Restriction digestion will be conducted on isolated nuclei, in which G4 structures on chromatin will be maintained, and the cleavages will be evaluated by LM-PCR. The comprehensive analyses of cleavages will indicate hard-to-cleave sites and comparison of these sites with potential G4-forming sequences will provide information pertinent to

Materials and Methods

Plasmids and strains

NLS_2-HA-I-SceI was amplified from pCMB-NLS-I-SceI by PCR using primes [pREPp_NLS_I-SceI_Fwd (5'-CTTTGTTAAATCATATGGGATCAAGATCGCCA-3') and pREPt_I-SceI_Rev (5'-TCTAGAGTCGACATATTATTTCAGGAAAGTTTCGGAGGA-3')]. PCR products were cloned at NdeI of pREP41 by In-Fusion® HD Cloning Kit (TAKARA Bio Inc.). Rif1 binding site at 2663kb on chromosome 1 and non-G4 site at 1542kb on chromosome 2 were disrupted by insertion of the Ura4 gene (selected by uracil-auxotroph on SD). DNA fragments containing an I-SceI cleavage site was integrated into the 23 bp loop between G-tracts in the $Rif1BS_{I:2663}$ segment or into a non-G4 segment at 1542kb on the chromosome 2. DNA fragments carrying these segments were transformed into the ura4 insertion strains described above, and surviving cells were selected on 5FOA (that inhibits the growth of ura^+ cells)/SD plate. Fission yeast cells freshly transformed with pREP41-NLS_2-HA-I-SceI were used to induce the expression of I-SceI.

Induction of I-SceI endonuclease in cells and genomic DNA extraction

Cells were grown in EMM medium containing 15 μM thiamine until exponential phase at 25 ºC, and were washed with fresh EMM medium and grown in the same medium at 30 ºC to induce I-SceI. At indicated time points, cells were collected and treated with 0.1% Azide, and were incubated in 200 U/mL Zymolyase/1.2M Sorbitol 0.1 M Potassium Phosphated Buffer pH7.0/ 10 mM EDTA/ 200mM /1% 2-mercaptetanol. The obtained spheroplasts were collected by centrifugation, re-suspended in 0.1 mg/μL RNaseA/ 0.1% TrionX-100/ TE and incubated in 37 ºC for 1 hr. 0.5 mg/μL Proteinase K was added and incubated at 37 ºC overnight. DNA was extracted by phenol/chloroform/isoamylalchole and precipitated by addition of 2x volume of ethanol. Precipitated DNA was washed with 80% ethanol and was dissolved in H_2O.

To completely digest genomic DNA with I-SceI, the genomic DNA isolated from cells prior to induction was digested with I-SceI (New England BioLabs) in CutSmart Buffer (New England BioLabs) for 6 hrs at 37 ºC, followed by extraction with phenol/chloroform/isoamylalcohole extraction and ethanol precipitation.

Generation of duplex DNA containing sequences derived from Rif1BS

Two oligonucleotides (500 pmole each) with overlapping 24 nt annealing segments were mixed in 10 mM Tris-HCl (pH7.5), 1mM EDTA and 1 mM LiCl, incubated at 96°C for 3 min and cooled down to room temperature. DNA chain was elongated by addition of Cutsmart buffer, 0.25 mM four deoxynucleotides, and 5 units of the Klenow fragment (New England Biolabs) at 37°C for 20 min. The elongated DNA products were run on 10% polyacrylamide gel, and duplex DNAs were isolated.

G4 formation by heat treatment

Isolated duplex DNA (0.2µg; ~2 pmole) was end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase at 37°C for 30 min, followed by further incubation for 5 min after addition of 1 mM ATP. The labeled DNA was phenol-treated and isolated by G50 spun column. The labeled DNA (20ng; ~200 fmole) was heated at 96°C for 3 min in 50 mM KCl, 40% PEG200, and slowly cooled down to 37°C. The heat-treated DNA (20 fmole) was incubated with a restriction enzyme (I-SceI or EcoRI; 2.5 unit) indicated in CutSmart buffer at 37°C for 2 hr. The same amount of formamide containing 10 mM EDTA was added to the reaction mixes, heated at 97°C for 3 min and applied to 10% PAGE containing 8M urea.

G4 formation by heat and transcription

The labeled template DNA (~5 pmole) was heated at 96°C for 3 min in 50 mM KCl and 40% PEG200 (20 µl) and slowly cooled down to room temperature. The heat-treated oligonucleotide (0.125 pmole) was used for restriction digestion.

The labeled template DNA (20 ng; ~200 pmole) was transcribed in vitro in 40 mM Tris-HCl (pH7.5), 8 mM $MgCl_2$, 10 mM DTT, 5% glycerol, 50 mM KCl, 1mM 4 ribonucleotides, and 2.5 units of T7 RNA polymerase in the presence or absence of 20% PEG200 at 37°C for 60 min. The products were phenol-treated, ethanol precipitated and resuspended in 10 µl TE. The purified products (20 fmole) were digested by a restriction enzyme and analyzed as above.

Digestion by ribonucleases

A portion of the transcription products (20 fmole) were directly digested by RNaseH (0.6 units), or RNaseA (0.1µg) and RNaseT1 (1 unit) at 37°C for 30 min in CutSmart buffer. A portion of the digestion products was mixed with a dye containing 0.2% SDS and run on a gel or digested by a restriction enzyme as above.

Analyses of the heat-induced G4 or transcription-induced G4 on PAGE

Products were run on 10% polyacrylamide gel containing 0.5X TBE, 50 mM KCl and 40 % PEG200 in 0.5X TBE containing 50 mM KCl.

Gel shift analyses of Rif1 binding to oligonucleotide DNA

Purification of fission yeast Rif1 and gel shift assays were conducted as described previously [29,30]. Briefly, ^{32}P-labeled heat-treated oligonucleotides were mixed with purified full-length Rif1 protein in 40 mM Hepes-KOH (pH 7.6), 50 mM KCl, 1 mM EDTA, 10% glycerol, and 0.01% Triton X-100 (in 10µl). After incubation at room temperature for 20 min, the reaction mixtures were applied onto polyacrylamide gel in 40% PEG200, 50 mM KCl, and 0.5x TBE. The gel was run with running buffer containing 0.5x TBE and 50 mM KCl. The gel was dried and autoradiographed.

Counting of Rad52 foci in cells expressing I-SceI.

A plasmid expressing I-SecI under an inducible nmt1 promoter was introduced into the strain carrying an I-SceI cleavage site and Rad52-YFP on its genome. The transformed cells were grown in the absence of thiamine for the indicated time in EMM medium to express I-SceI endonuclease. Rad52-YFP foci at double stranded DNA breaks were observed by Microscope BZ-X700 (KEYENCE), and the foci forming cells were counted. The fractions of Rad52-YFP positive cells were calculated by dividing the numbers of Rad52-YFP $^{+}$ cells with those of the all the cells in the fields.

Adaptor Ligation and LM-PCR

The I-SceI Adaptor (20 nM) was ligated to genomic DNA digested with I-SceI (100 ng) by T4 DNA ligase (1.5 units). The reaction (25µl) was diluted with 175 µl water. The completely digested genomic DNA was diluted serially (10x dilution; 0.01% ~ 10.0%) for calibration curve of real time PCR using primer sets in **Supplementary Table S1**. Real time PCR was performed with TB Green® Premix Ex Taq™ II (Tli RNaseH Plus)

(TAKARA Bio inc.) on LightCycler 480 (Roche). The primer sets were designed at 4282 kb on Chromosome 2 to correct copy number deviation of genomic DNA. The primers used in LM-PCR are listed in **Supplementary Table S1**.

Acknowledgements

HM conceived the idea for the experiments described in this manuscript. HM and NK conducted all the in vitro analyses. YK analyzed G4 formation in fission yeast cells by examining the I-SceI digestion. RF gave various technical assistance during the course of this experiment. This work was supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research (A) [Grant Numbers 23247031 and 26251004] to HM. We also would like to thank the Uehara Memorial Foundation and the Takeda Science Foundation for the financial support of this study. We would like to thank all the members of the laboratory for helpful discussion.

References

1 M. Gellert, M.N. Lipsett, D.R. Davies, Helix formation by guanylic acid, Proc. Natl. Acad. Sci. U. S. A. 48 (1962) 2013-2018.

2 D. Sen, W. Gilbert, Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, Nature 334 (1988) 364-366.

3 W.I. Sundquist, A. Klug, Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops, Nature 342 (1989) 825-829.

4. Kumari S, Bugaut A, Huppert JL, Balasubramanian S An RNA G-quadruplex in the 5’ UTR of the NRAS proto-oncogene modulates translation, Nat Chem Biol 3 (2007) 218–221.

5. Wieland M, Hartig JS RNA quadruplex-based modulation of gene expression, Chem Biol 14 (2007) 757–763.

6. J.Q. Liu, S. Xiao, Y.H. Hao, Z. Tan, Strand-Biased Formation of G-Quadruplexes in DNA Duplexes Transcribed with T7 RNA Polymerase, Angew. Chem. Int. Ed. Engl. 54 (2015) 8992-8996.

7. Y. Zhao, J.Y. Zhang, Z.Y. Zhang, T.J. Tong, Y.H. Hao, Z. Tan, Real-Time Detection Reveals Responsive Cotranscriptional Formation of Persistent Intramolecular DNA and Intermolecular DNA:RNA Hybrid G-Quadruplexes Stabilized by R-Loop, Anal. Chem. 89 (2017) 6036-6042.

8. B.P. Belotserkovskii, S. Tornaletti, A.D. D'Souza, P.C. Hanawalt, R-loop generation during transcription: Formation, processing and cellular outcomes, DNA Repair (Amst) 71 (2018) 69-81.

9. J.Y. Zhang, K.W. Zheng, S. Xiao, Y.H. Hao, Z. Tan, Mechanism and manipulation of DNA:RNA hybrid G-quadruplex formation in transcription of G-rich DNA, J. Am. Chem. Soc. 136 (2014) 1381-1390.

10. J.L. Huppert, S. Balasubramanian, Prevalence of quadruplexes in the human genome, Nucleic Acids Res. 33 (2005) 2908-2916.

11. A.K. Todd, M. Johnston, S. Neidle, Highly prevalent putative quadruplex sequence motifs in human DNA, Nucleic Acids Res. 33 (2005) 2901-2907.

12. V.S. Chambers, G. Marsico, J.M. Boutell, M. Di Antonio, G.P. Smith, S. Balasubramanian, High-throughput sequencing of DNA G-quadruplex structures in the human genome, Nat. Biotechnol. 33 (2015) 877-881.

13. Hansel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Di Antonio M, Pike J, Kimura H, Narita M, et al: G-quadruplex structures mark human regulatory chromatin, Nat Genet 48 (2016) 1267-1272.

14. Hayano M, Kanoh Y, Matsumoto S, Renard-Guillet C, Shirahige K, Masai H, Rif1 is a global regulator of timing of replication origin firing in fission yeast, Genes Dev. 26 (2012) 137-150.

15. Kanoh Y, Matsumoto S, Fukatsu R, Kakusho N, Kono N, Renard-Guillet C, Masuda K, Iida K, Nagasawa K, Shirahige K, Masai H, Rif1 binds to G quadruplexes and suppresses replication over long distances, Nat Struct Mol Biol. 22 (2015) 889-897.

16. Sabouri N, Capra JA, Zakian VA, The essential Schizosaccharomyces pombe Pfh1 DNA helicase promotes fork movement past G-quadruplex motifs to prevent DNA damage, BMC Biol. 12 (2014) 101.

17. Cheng M, Cheng Y, Hao J, Jia G, Zhou J, Mergny JL, Li C. Loop permutation affects the topology and stability of G-quadruplexes, Nucleic Acids Res. 46 (2018) 9264-9275.

18. Hao F, Ma Y, Guan Y. Effects of Central Loop Length and Metal Ions on the Thermal Stability of G-Quadruplexes. Molecules, 24 (2019) pii: E1863.

19. Masai H, Kakusho N, Fukatsu R, Ma Y, Iida K, Kanoh Y, Nagasawa K. Molecular architecture of G-quadruplex structures generated on duplex Rif1-binding sequences, J Biol Chem. 293 (2018) 17033-17049.

20. D. Miyoshi, H. Karimata, N. Sugimoto, Hydration regulates thermodynamics of Gquadruplex formation under molecular crowding conditions, J. Am. Chem. Soc. 128 (2006) 7957-7963.

21. Hansel-Hertsch R, Di Antonio M, Balasubramanian S DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential, Nat Rev Mol Cell Biol 18 (2017) 279–284.

22. Masai, H, Tanaka, T G-quadruplex DNA and RNA: their roles in regulation of DNA replication and other biological functions, Biochem Biophys Res Commun (2020) In press

23. Biffi G, Tannahill D, McCafferty J, Balasubramanian S Quantitative visualization of DNA G-quadruplex structures in human cells, Nat Chem 5 (2013) 182–186.

24. Shivalingam A, Izquierdo MA, Le Marois A, Vysniauskas A, Suhling K, Kuimova MK, Vilar R The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy, Nat Commun 6 (2015) 8178.

25. Wu RY, Zheng KW, Zhang JY, Hao YH, Tan Z. Formation of DNA:RNA hybrid G-quadruplex in bacterial cells and its dominance over the intramolecular DNA G-quadruplex in mediating transcription termination, Angew Chem Int Ed Engl 54 (2015) 2447–2451.

26. Kouzine F, Wojtowicz D, Baranello L, Yamane A, Nelson S, Resch W, Kieffer-Kwon KR, Benham CJ, Casellas R, Przytycka TM, Levens D Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome, Cell Syst 4 (2017) 344–356 e347.

27. Sekibo DAT, Fox KR. The effects of DNA supercoiling on G-quadruplex formation, Nucleic Acids Res. 45 (2017) 12069-12079.

28. Mao SQ, Ghanbarian AT, Spiegel J, Martínez Cuesta S, Beraldi D, Di Antonio M, Marsico G, Hänsel-Hertsch R, Tannahill D, Balasubramanian S, DNA G-quadruplex structures mold the DNA methylome, Nat Struct Mol Biol. 25 (2018) 951-957.

29. Kobayashi S, Fukatsu R, Kanoh Y, Kakusho N, Matsumoto S, Chaen S, Masai H. Both a Unique Motif at the C Terminus and an N-Terminal HEAT Repeat Contribute to G-Quadruplex Binding and Origin Regulation by the Rif1 Protein, Mol Cell Biol. 39 (2019) pii: e00364-18.

30. Masai H, Fukatsu R, Kakusho N, Kanoh Y, Moriyama K, Ma Y, Iida K, Nagasawa K. Rif1 promotes association of G-quadruplex (G4) by its specific G4 binding and oligomerization activities, Sci Rep. 9 (2119) 8618.

Legends to Figures

Figure 1 Replacement of a long loop of RifIBS-derived G4 and its resistance to restriction digestion.

The lower drawing shows the sequences (of the G-rich strand in 5'->3' direction) of $Rif1BS_{I:2663}$ derivative with I-SceI site (upper) and its G-tract mutant (lower). G-tracts are highlighted in pink. The mutated G-tracts are in yellow boxes. The I-SceI sites that replace the 18 bp segment in the 23 bp loop region are highlighted in red. The red arrowheads indicate the cleavage sites of I-SceI (downward showing the cleavage site on G-rich strand, while upward arrowheads that on the C-rich strand). The 122bp duplex DNA was labeled with ^{32}P at both 5'-ends by T4 polynucleotide kinase. DNA was heat-denatured and cut with I-SceI (lanes 1-4) or not (lanes 5-8). The products were analyzed on 10% PAGE containing 8M urea. The 25 nt cleaved fragment derived from the G-rich strand ran off the gel and is not shown.

Figure 2 Transcription-driven Rif1BS G4 also shows resistance to restriction digestion

The sequences of the duplex DNAs derived from Rif1BSI are shown in **A**. G-tracts are highlighted in pink, and the I-SceI and EcoRI sites are highlighted in red. The red arrowheads indicate the cleavage sites of I-SceI and EcoRI, as in Figure 1. The white letters on black arrows indicate the 19 bp T7 promoter sequence. The double-labeled duplex DNA containing I-SecI (**B** and **D**) or EcoRI (**C** and **E**) in the 23 bp loop segment was heated or transcribed *in vitro* with T7 RNA polymerase, were cleaved with the restriction enzyme as shown, and were analyzed on 10% PAGE containing 8M urea. Lanes 1,5,9,13 heated; lanes 2,6,10.14, incubated without T7 RPase; lanes 3,7,11,15, incubated with T7 RPase (no PEG200); lanes 4,8,12,16, incubated with T7 RPase and 20% PEG200. Red asterisks in **C** indicate the single-stranded DNA used to generate the duplex DNA, that is present in the duplex DNA template preparations used for the assay.

Figure 3 RNA-DNA hybrid formation is required for formation of restriction-resistant loop formation.

A. The double end-labeled duplex DNAs as shown at the bottom were transcribed *in vitro* with T7 RNA polymerase, were digested with ribonuclease H (lanes 1 and 3), ribonuclease A+ribonuclease T1 (lanes 2 and 4) or non-treated (lanes 3 and 6), and were analyzed on 10% PAGE containing 50 mM KCl and 40% PEG200 in 0.5X TBE. The arrowhead indicates the RNA-DNA hybrid containing G4. **B**. The transcribed products treated with ribonucleases in **A** were cleaved with a restriction enzyme indicated, and analyzed on 10% PAGE containing 8M urea. The red asterisk in **B** indicate the single-stranded DNA used to generate the duplex DNA, that is present in the duplex DNA template preparations used for the assay.

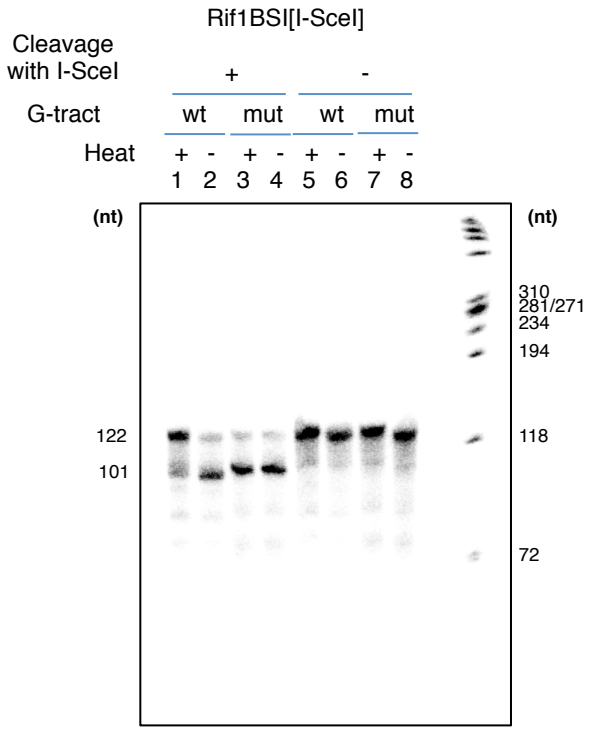
Figure 4. In vivo cleavage of the I-SceI site introduced within the loop

A. I-SceI cleavage sites, indicated in blue, were integrated at Rif1BS ($Rif1BS_{I:2663}$) and non-Rif1BS (II:1524kb) in the genomic DNA on Chromosome I and II, respectively. Residues in lowercase indicate those that changed from the wild-type due to the integration of the I-SceI sites and G-tract mutations. Lower drawings show genomic landscape of the inserted areas. The red and blue boxes indicate the genes and non-coding transcripts, respectively. **B**. The strains (with Rad52-YFP; a marker for double-stranded DNA breaks) with an I-SceI cleavage site at I:2663kb (Rif1BS wild-type; left) or at I:1542kb (non-Rif1BS; center) or at I:2663kb (Rif1BS mut; right) and harboring pREP41-NLS-HA-I-SceI plasmid were transferred to EMM without thiamine and the cells were collected at the indicated time after medium exchange. The whole cell extracts were analyzed by western blot using anti-HA antibody to determine the expression level of I-SceI. **C**. The same cells were grown in EMM without thiamine for 18 hr at 30 ºC, and were observed under fluorescent microscope. **D**. The numbers of Rad52-YFP foci of the cells carrying I-SceI site at the indicated loci were counted and the fractions of YFP-positive (blue) and -negative (red) are presented. **E** and **F**. The genomic DNA was extracted and purified at each timepoint after induction of I-SceI. A portion of genomic DNA at time 0 was digested with I-SceI in vitro. The genomic DNA was treated with BAP to avoid re-ligation between the digested genomic DNA. BAP-treated genomic DNA was ligated with an adaptor DNA. Ligated DNA was diluted and real time PCR was performed to quantify the level of digestion. The levels of digestion at the target I-SceI site were quantified and normalized by the input genomic DNA that was quantified by PCR at II:4282kb site. The cleavage efficiency at

each timepoint was determined by calculating the ratio of digestion at the target site to that of in vitro digestion (complete digestion at time 0).

Figure

Figure 1



10% PAGE with 8M urea in 0.5X TBE

Rif1BSIwt

GTGGGATGGGTATCTAGGAATGAAAGGCAATGTGGGGGGATAGTGGGCATCCCTTTTCAAACCATCACTTATGACTTGGGAAGAGCAACAGTGGAGAAAATAAACCCGATGTGGGGGGAAAG

Rif1BSI[I-SceI]wt

GTGGGATGGGTAATTACCCTGTTATCCCTATGTGGGGGGATAGTGGGCATCCCTTTTCAAACCATCACTTATGACTTGGGAAGAGCAACAGTGGAGAAAATAAACCCGATGTGGGGGGAAAG

Rif1BSI[I-SceI]mut

GTAGAATAGATAATTACCCTGTTATCCCTATGTAGAGAGATAGTAGACATCCCTTTTCAAACCATCACTTATGACTTGGGAAGAGCAACAGTGGAGAAAATAAACCCGATGTGGGGGGAAAG

Figure 2

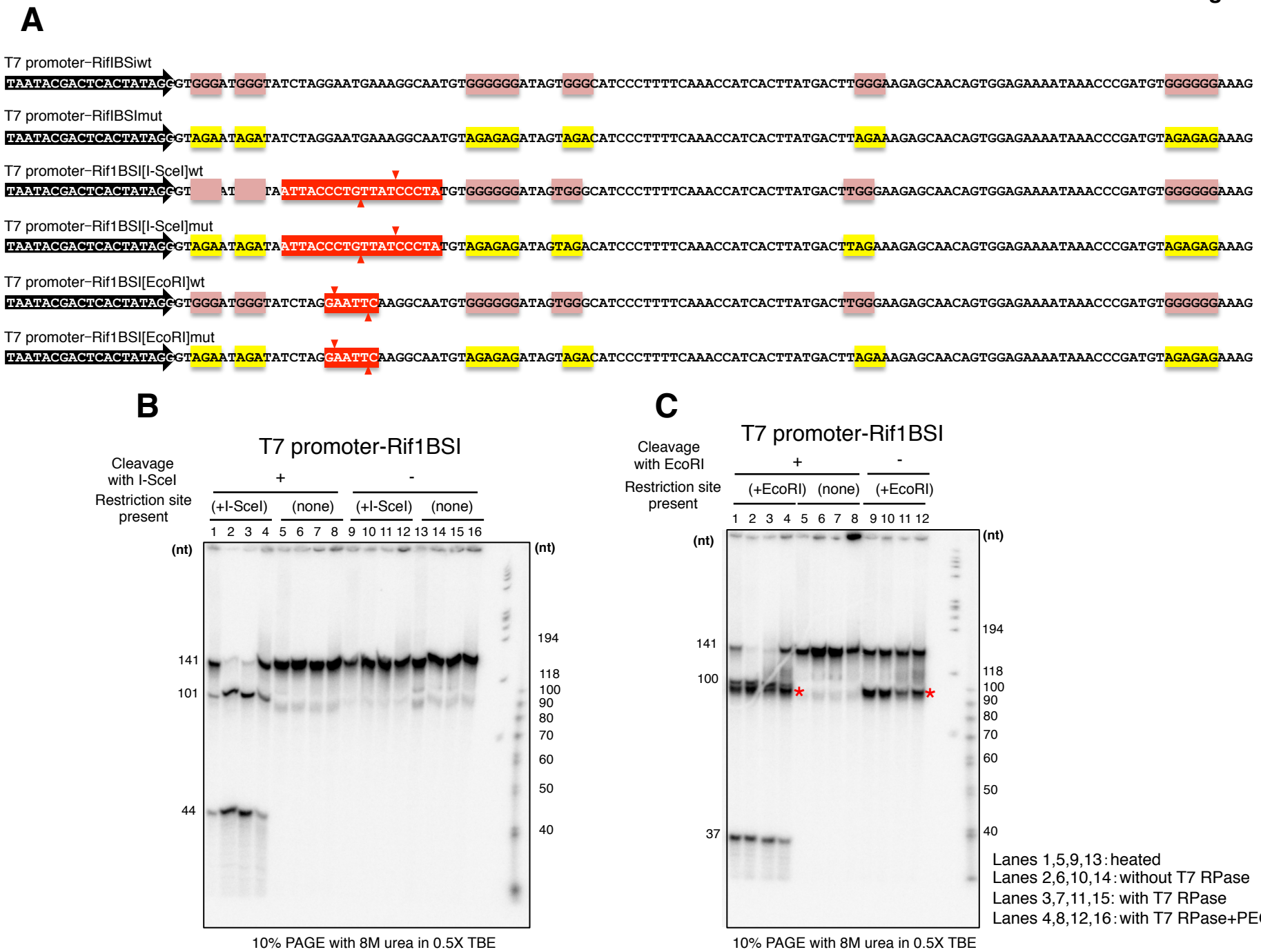
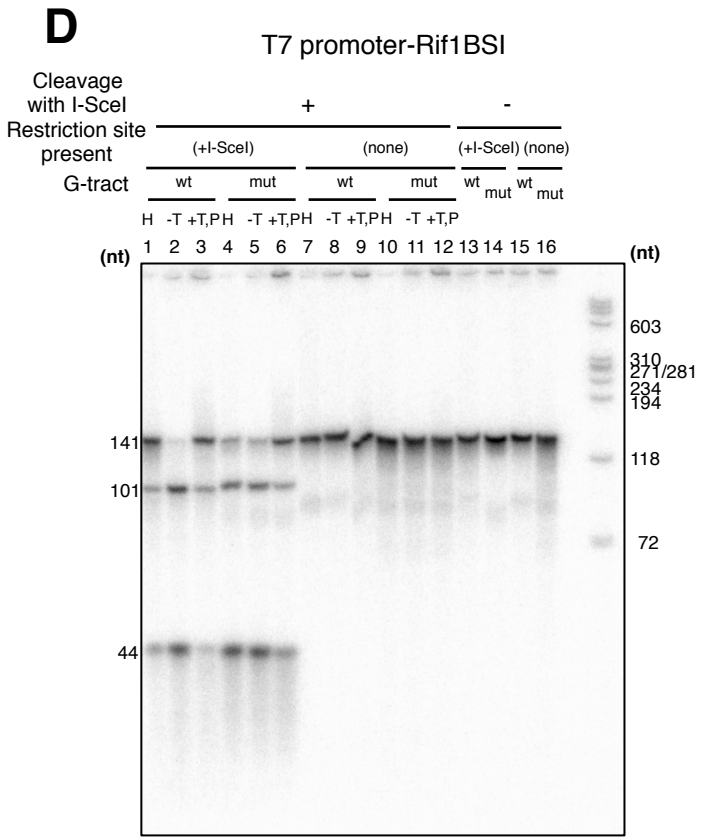
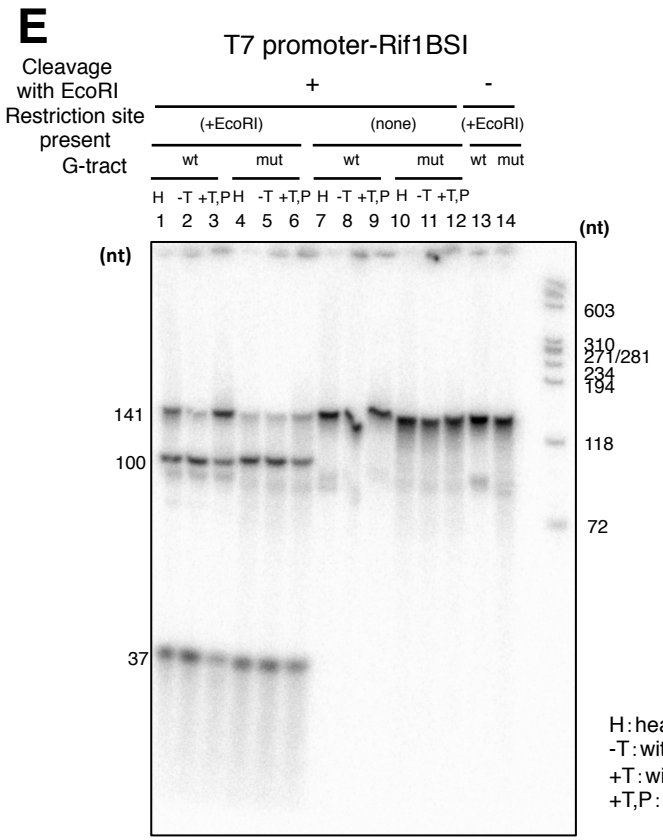

A
T7 promoter-RifIBSiwt
TAATACGACTCACTATAGGGTGGGATGGGTATCTAGGAATGAAAGGCAATGTGGGGGGATAGTGGGCATCCCTTTTCAAACCATCACTTATGACTTGGGAAGAGCAACAGTGGAGAAAATAAACCCGATGTGGGGGGAAAG
T7 promoter-RifIBSImut
TAATACGACTCACTATAGGGTAGAATAGATATCTAGGAATGAAAGGCAATGTAGAGAGATAGTAGACATCCCTTTTCAAACCATCACTTATGACTTAGAAAGAGCAACAGTGGAGAAAATAAACCCGATGTAGAGAGAAAG
T7 promoter-Rif1BSI[I-SceI]wt
TAATACGACTCACTATAGGGTATTACCCTGTTATCCCTATGTGGGGGGATAGTGGGCATCCCTTTTCAAACCATCACTTATGACTTGGGAAGAGCAACAGTGGAGAAAATAAACCCGATGTGGGGGGAAAG
T7 promoter-Rif1BSI[I-SceI]mut
TAATACGACTCACTATAGGGTAGAATAGATATTACCCTGTTATCCCTATGTAGAGAGATAGTAGACATCCCTTTTCAAACCATCACTTATGACTTAGAAAGAGCAACAGTGGAGAAAATAAACCCGATGTAGAGAGAAAG
T7 promoter-Rif1BSI[EcoRI]wt
TAATACGACTCACTATAGGGTGGGATGGGTATCTAGGAATTCAAGGCAATGTGGGGGGATAGTGGGCATCCCTTTTCAAACCATCACTTATGACTTGGGAAGAGCAACAGTGGAGAAAATAAACCCGATGTGGGGGGAAAG
T7 promoter-Rif1BSI[EcoRI]mut
TAATACGACTCACTATAGGGTAGAATAGATATCTAGGAATTCAAGGCAATGTAGAGAGATAGTAGACATCCCTTTTCAAACCATCACTTATGACTTAGAAAGAGCAACAGTGGAGAAAATAAACCCGATGTAGAGAGAAAG
B
T7 promoter-Rif1BSI
Cleavage with I-SceI
Restriction site present
+
-
(+I-SceI)
(none)
(+I-SceI)
(none)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
(nt)
141
101
44
194
118
100
90
80
70
60
50
40
10% PAGE with 8M urea in 0.5X TBE
C
T7 promoter-Rif1BSI
Cleavage with EcoRI
Restriction site present
+
-
(+EcoRI)
(none)
(+EcoRI)
1 2 3 4 5 6 7 8 9 10 11 12
(nt)
141
100
37
194
118
100
90
80
70
60
50
40
10% PAGE with 8M urea in 0.5X TBE
Lanes 1,5,9,13 : heated
Lanes 2,6,10,14 : without T7 RPase
Lanes 3,7,11,15: with T7 RPase
Lanes 4,8,12,16 : with T7 RPase+PEG20

Figure 2



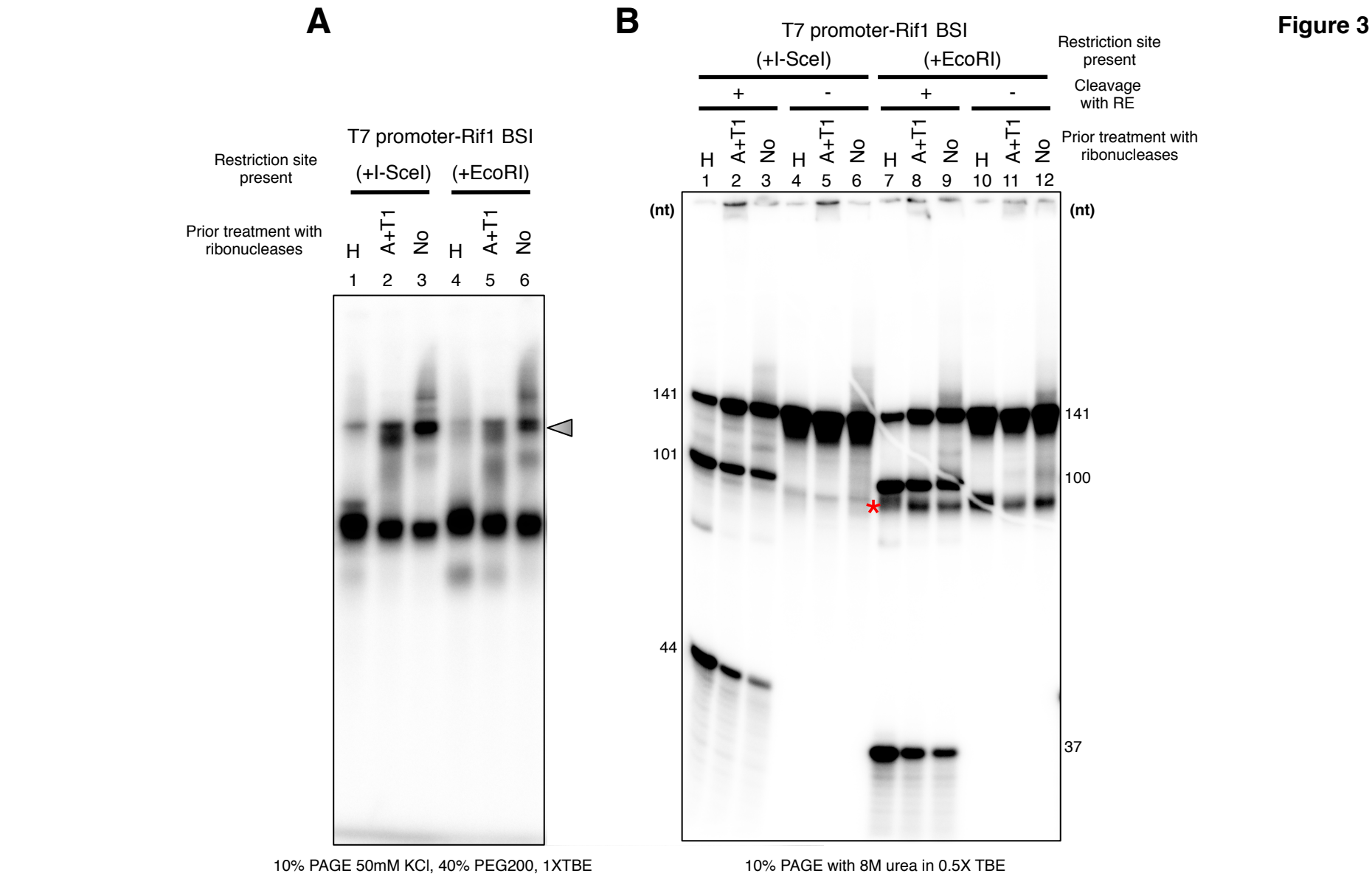
10% PAGE with 8M urea in 0.5X TBE



10% PAGE with 8M urea in 0.5X TBE

H : heated
-T : without T7 RPase
+T : with T7 RPase
+T,P : with T7 RPase+PEG200

Figure 3


A
T7 promoter-Rif1 BSI
Restriction site present
(+I-SceI)
(+EcoRI)
Prior treatment with ribonucleases
H
A+T1
No
1 2 3 4 5 6
B
T7 promoter-Rif1 BSI
(+I-SceI)
(+EcoRI)
Restriction site present
Cleavage with RE
+
-
Prior treatment with ribonucleases
1 2 3 4 5 6 7 8 9 10 11 12
(nt)
141
101
44
100
37
10% PAGE 50mM KCl, 40% PEG200, 1XTBE
10% PAGE with 8M urea in 0.5X TBE
T7 promoter−Rif1BSI[I-SceI]wt
TAATACGACTCACTATAGGGTGGGATGGGTATTACCCTGTTATCCCTATGTGGGGGGATAGTGGGCATCCCTTTTCAAACCATCACTTATGACTTGGGAAGAGCAACAGTGGAGAAAATAAACCCGATGTGGGGGGAAAG
T7 promoter−Rif1BSI[EcoRI]wt
TAATACGACTCACTATAGGGTGGGATGGGTATCTAGGAATTCAAGGCAATGTGGGGGGATAGTGGGCATCCCTTTTCAAACCATCACTTATGACTTGGGAAGAGCAACAGTGGAGAAAATAAACCCGATGTGGGGGGAAAG

Supplementary Information for

Detection of cellular G-quadruplex by using a loop structure as a structural determinant

By
***Hisao Masai, Yutaka Kanoh, Naoko Kakusho and Rino Fukatsu**

Affiliations:
Department of Genome Medicine, Tokyo Metropolitan Institute of Medical Science, Kamikitazawa, Setagaya-ku, Tokyo 156-8506, Japan

Correspondence to
Dr. Hisao Masai
Telephone: +81-3-5316-3220
Fax: +81-3-5316-3145
E-mail: masai-hs@igakuken.or.jp

Contents

Legends for Supplementary Figures

Supplementary Figure S1

Gel shift assays of Rif1 protein with various Rif1BS-derived oligonucleotides: Insertion of a restriction site in the loop does not affect G4 formation and Rif1 binding in vitro.

75mer oligonucleotides derived from$Rif1BS_{I:2663}$ containing G-tracts and their derivatives carrying a restriction site in the 23bp loop were end-labeled with ^{32}P and heat-treated and renatured under G4-forming condition. Theses oligonucleotides were used for gel shift assays with purified SpRif1 protein, as indicated. The amount of SpRif1 protein added was 4.5 nM amd 14 nM. The labeled oligonucleotides were present at 0.5 μM. DNA-protein complexes were analyzed on 8% PAGE (29:1) containing 40% PEG200, 50mM KCl and 0.5X TBE. The open arrow indicates the monomeric form of G4 structures generated from the oligonucleotide. Due to its shape, it runs faster in PAGE compared to G-tract mutant oligonucleotide. The slow-migrating DNA species indicated by the lower bracket represent the oligomeric forms of G4. The sequences of oligonucleotide used for assays and their features are shown at the bottom of the figure with the same legend as for **Fig. 1**. The extent of DNA binding was assessed by dividing the intensity of the shifted bands (indicated by the upper bracket) by that of the sum of the unbound and shifted bands.

Supplementary Figure S2

Product analyses of in vitro transcription reactions on duplex Rif1BS DNA templates: Putative G4 structures are formed on the Rif1BS templates both in the presence and absence of an inserted restriction site.

Double-end-labeled duplex DNA shown at the bottom were heated or transcribed in vitro by T7 RNA polymerase and were analyzed on a neutral polyacrylamide gel containing PEG200 and KCl. Lanes 1,7,and 13, heat-treated; lanes 2, 8 and 14, incubated without T7 RNA polymerase; lanes 3,4,9,10,15,16; transcribed by T7 RNA polymerase in the absence of PEG200; lanes 5,6,11,12,17,18; transcribed by T7 RNA polymerase in the presence of PEG200. Lanes 2,4,6,8,10,12,14,16,18; applied directly on the gel; lanes 3,5,9,11,25 and 17; applied after addition of loading buffer containing 0.1% SDS. The bands shown by gray arrow indicate a slow mobility RNA-DNA hybrid

containing G4. This band is not observed in the samples transcribed in the absence of PEG200. The method for loading samples did not affect the result.

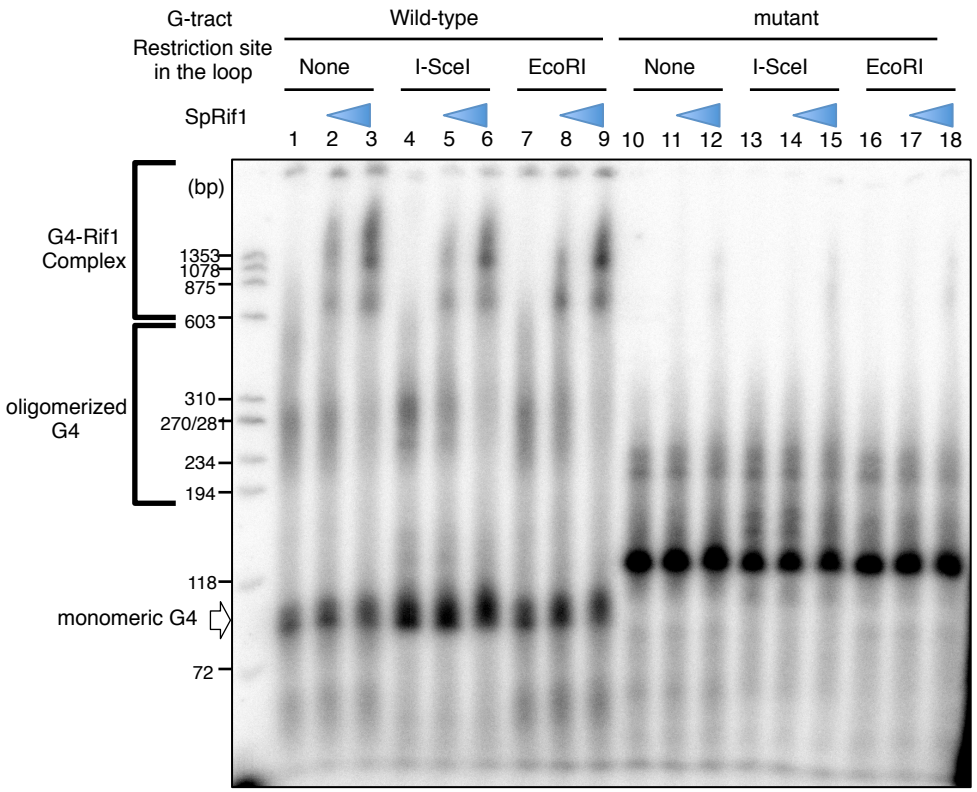
Supplementary Figure S3

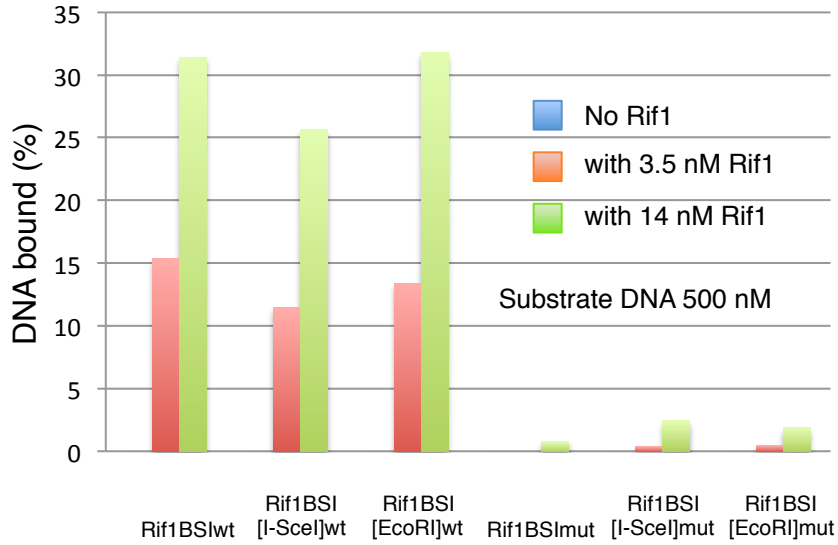
Time course of Rad52 foci formation after I-SceI induction: DBS formation is delayed in the strain containing the wild-type Rif1BS[I-SceI] sequences compared to that carrying I-SceI in non-G4 or mutant Rif1BS.

A. The Rad52-YFP strain containing $Rif1BS_{I:2663}$ with (lower) or without (upper) I-SceI cleavage sites were transferred to EMM without thiamine, cultured for the indicated time (hr) at 30 ºC, and Rad52-YFP was observed by fluorescent microscope. **B**. Rad52-YFP containing cells without I-SceI site and those with I-SceI inserted at the wild-type Rif1BS, G-tract mutant Rif1BS or non-G4 sequence were grown and were transferred to EMM without thiamine at time 0. The Rad52 foci formation rates (representing the cleavage) were calculated in each cell population at each time point.

Supplementary Table S1 Oligonucleotides used for cellular G4		
Primer annealing to genomic DNA for copy number deviation	II:4282kb_Fwd	5'-ATGCTTCACATATGGTAGGCTCAGA-3'
	II:4282kb_Rev	5'-GACCAAGGCAAGAAGTTAGTTTACGA-3'
Primer for LM-PCR	LMPCR_Primer	5'-TGGTACGATTACTGGTGTAGCTGATGG-3'
	I:2663kb_Rev5.4	5'-CAGGCTTCCGTACTGCACTTCA-3'
	II:1542kb_Fwd5	5'-TGGATAGTCACCTTCGGCTTAGCAA-3'
Oligo for I-SceI Adaptor	I-SceI_AdaptorS1	5'-GCTGGTACGATTACTGGTGTAGCTGATGGTTACAGCATTAT-3'
	I-SceI_AdaptorAS1	5'-TGCTGTAACCATCAGCTACACCAGTAATCGTACCAGC-3'
	I-SceI_AdaptorS2	5'-AATCAAGACTGACCCATCGAATGGCTATCCGACGTATATAA-3'
	I-SceI_AdaptorAS2	5'-ATACGTCGGATAGCCATTCGATGGGTCAGTCTTGATT-3'

Supplementary Figure S1





8% PAGE (29:1) in 40% PEG200, 50mM KCl and 0.5X TBE.

Restriction site in the loop

Rif1BSIwt
None GTGGGATGGGTATCTAGGAATGAAAGGCAATGTGGGGGGATAGTGGGCATCCCTTTTCAAACCATCACTTATGAC

Rif1BSI[I-SceI]wt
I-SceI GTGGGATGGGTATTACCCTGTTATCCCTATGTGGGGGGATAGTGGGCATCCCTTTTCAAACCATCACTTATGAC

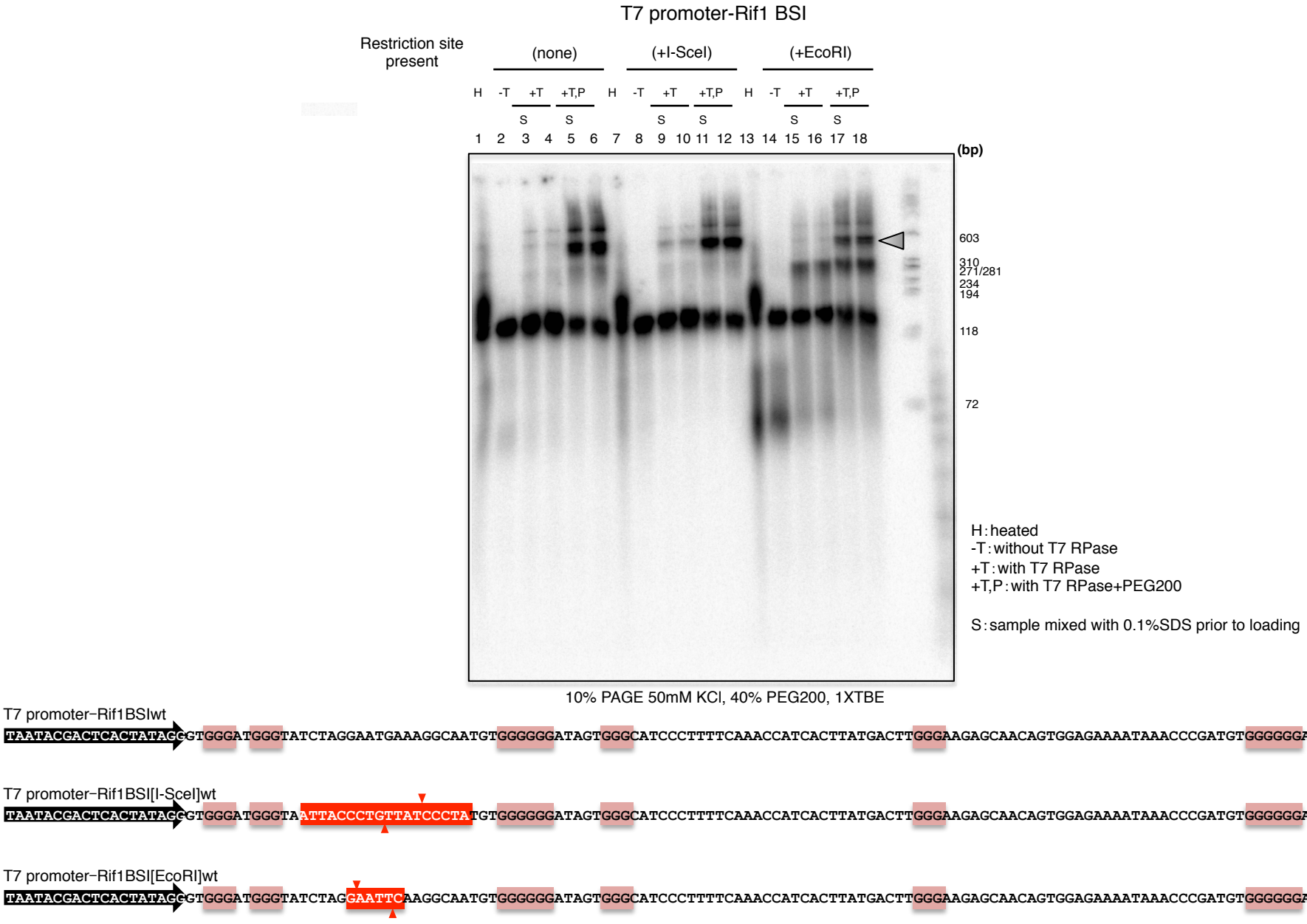
Rif1BSI[EcoRI]wt
EcoRI GTGGGATGGGTATCTAGGAATTCAAGGCAATGTGGGGGGATAGTGGGCATCCCTTTTCAAACCATCACTTATGAC

Rif1BSImut
None GTAGAATAGATATCTAGGAATGAAAGGCAATGTAGAGAGATAGTAGACATCCCTTTTCAAACCATCACTTATGAC

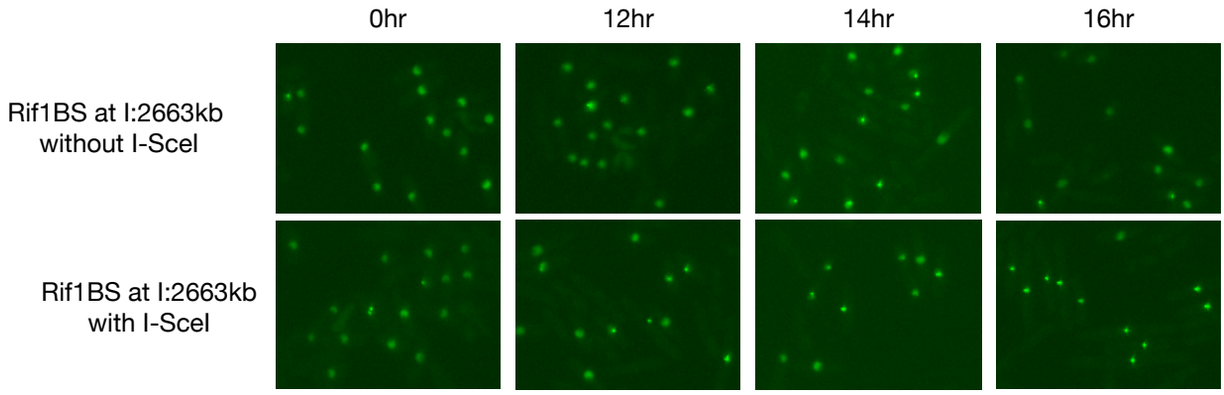
Rif1BSI[I-SceI]mut
I-SceI GTAGAATAGATATTACCCTGTTATCCCTATGTAGAGAGATAGTAGACATCCCTTTTCAAACCATCACTTATGAC

Rif1BSI[EcoRI]mut
EcoRI GTAGAATAGATATCTAGGAATTCAAGGCAATGTAGAGAGATAGTAGACATCCCTTTTCAAACCATCACTTATGAC

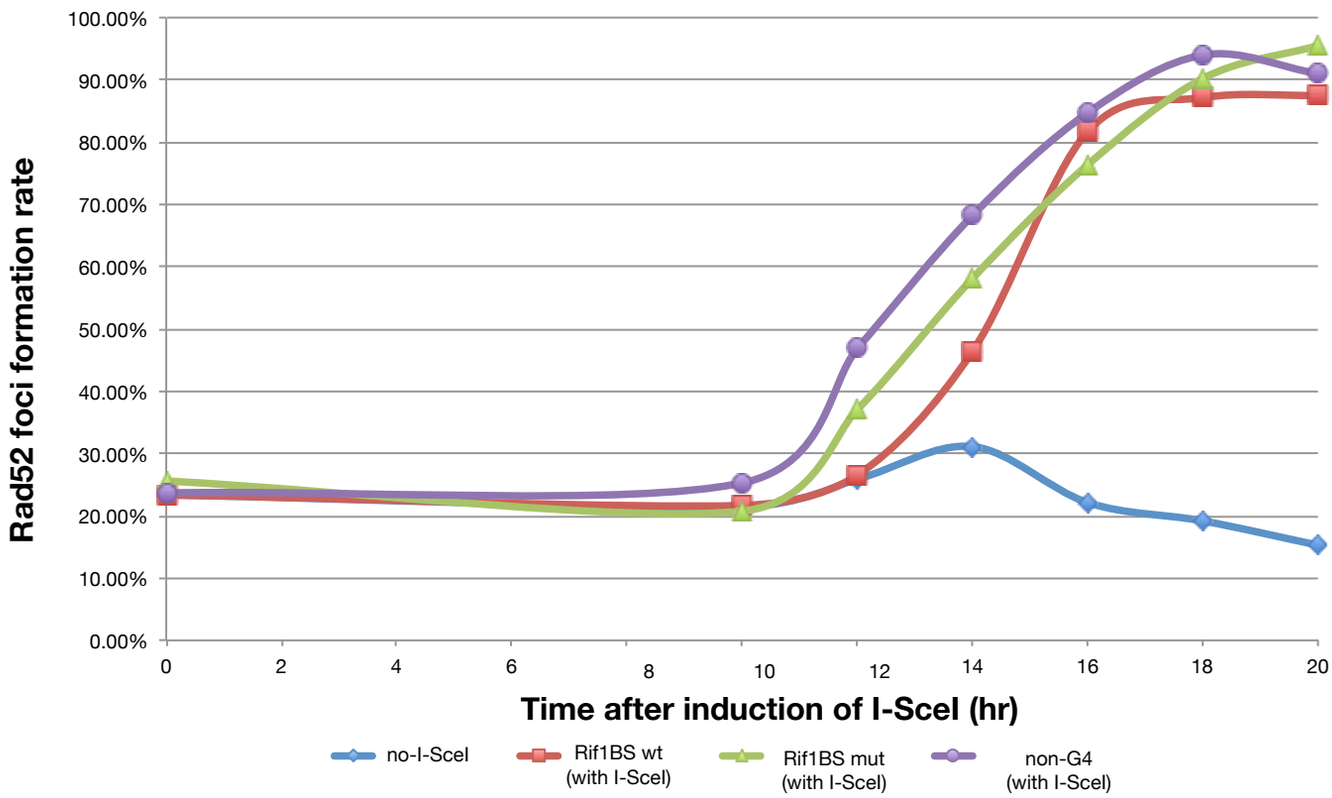
Supplementary Figure S2



A



B



Supplementary Figure 3