CELLULAR & MOLECULAR BIOLOGY LETTERS
Volume 9, (2004) pp 557 – 566
http://www.cmbl.org.pl

Received 5 April 2004
Accepted 30 July 2004

ISOLATION, CLONING AND CHARACTERISATION OF MOTIFS CONTAINING $(GA/TC)_n$ REPEATS ISOLATED FROM VETCH, *VICIA BITHYNICA*

TOMASZ SAKOWICZ[1]*, RICHARD BOWATER[2]
and PAWEŁ PARNIEWSKI[3]
[1]University of Łódź, Department of Cytogenetics and Plant Molecular Biology, 90-237 Łódź, Banacha 12/16, Poland, [2]University of East Anglia, School of Biological Sciences, Norwich NR4 7TJ, United Kingdom, [3]Centre for Medical Biology, Polish Academy of Sciences, 93-232 Łódź, Lodowa 106, Poland

Abstract: Microsatellites are widely distributed in plant genomes and comprise unstable regions that undergo mutational changes at rates much greater than that observed for non-repetitive sequences. They demonstrate intrinsic genetic instability, manifested as frequent length changes due to insertions or deletions of repeat units. Detailed analysis of 1600 clones containing genomic sequences of *Vicia bithynica* revealed the presence of microsatellite repeats in its genome. Based on the screening of a partial DNA library of plasmids, 13 clones harbouring $(GA/TC)_n$ tracts of various lengths of repeated motif were identified for further analysis of their internal sequence organization. Sequence analyses revealed the precise length, number of repeats, interruptions within tracts, as well as sequence composition flanking the repeat motifs. Representative plasmids containing different lengths of $(GA/TC)_n$ embedded in their original flanking sequence were used to investigate the genetic stability of the repeats. In the study presented herein, we employed a well characterised and tractable bacterial genetic system. Recultivations of *Escherichia coli* harbouring plasmids containing $(GA/TC)_n$ inserts demonstrated that the genetic instability of $(GA/TC)_n$ microsatellites depends highly on their length (number of repeats). These observations are in agreement with similar studies performed on repetitive sequences from humans and other organisms.

Key Words: DNA, Genetic, Instability, Microsatellites, *Vicia bithynica*

* Corresponding author; e-mail: tomeksakowicz@wp.pl

INTRODUCTION

A large proportion of higher eukaryotic genomes consists of different kinds of repetitive DNA [1-2]. Such repeats are often used as informative genetic markers in genome mapping, population genetics, forensics, evolutionary studies and molecular diagnostics [3-4]. Their usefulness in these studies is linked to the high polymorphism of the microsatellites and their frequent occurrence in eukaryotic genomes. The presence of such sequences has been confirmed in plant genomes [5]. Many years of extensive studies on the nature of the genetic instabilities of microsatellites revealed that size alterations of these tracts may be generated via different biochemical mechanisms, including replication, transcription, DNA repair and recombination [2]. Microsatellites are thought to play important functions in chromatin organization and in regulation of many biological processes and genes [2]. Of these repeats, a special class of microsatellites, trinucleotide repeat sequences (TRS), are involved in some human neurodegenerative diseases [6, 7]. Many experiments in bacteria, yeast and mammalian systems suggest that the high frequency of length changes (insertions and deletions) of triplet repeats is caused by their propensity to form unusual secondary structures [7, 8]. Alterations in the size of DNA repeats are not, however, limited to the TRS-dependent disorders, as instabilities in other microsatellites are also observed, e.g. in tumours from patients with hereditary nonpolyposis colorectal cancer [9-11].

In general, a significant feature of direct (tandem) repeats in DNA is their intrinsic ability to form non-B-DNA conformations. Unusual DNA structures such as left-handed Z DNA, cruciforms, triplexes, and tetraplexes may form under physiological conditions. If such structures form *in vivo* they may be hazardous for genome stability if not removed by repair mechanisms. Many experimental lines of evidence have shown that non-B-DNA-forming sequences are genetically unstable and are deleterious to normal cellular function [8, 12]. Tracts of pur·pyr, such as $(GA/TC)_n$, have a propensity to form triplexes, which are among the most thermodynamically stable of unusual DNA structures [13, 14].

It has been documented that the $(GA/TC)_n$ motif is one of the most abundant microsatellites in plants [15]. Our search through the Plant Repeat Database at TIGR revealed a differential distribution of such motifs in the main plant families. There are no significantly long (more than 10 units) $(GA/TC)_n$ repeats in *Solanaceae*, two such tracts in *Brassicaceae* (*Arabidopsis thaliana*), and eight motifs in *Gramineae* (four motifs found in *Oryza sativa*). There are four (GA/TC) loci comprising at least 10 repeats identified in *Fabaceae*, to which *V. bithynica* belongs. To allow analysis of mechanisms affecting the genetic stability of pur·pyr tracts, we are characterizing such sequences in different organisms. This work describes analysis of polymorphisms of $(GA/TC)_n$ motifs from *V. bithynica* and identifies their genetic instability when cloned in bacterial plasmids.

MATERIALS AND METHODS

Seeds of *V. bithynica* were obtained from the International Center for Agricultural Research in the Dry Areas (ICARDA) in Aleppo (Syria).
Total DNA of *V. bithynica* was isolated according to the procedure developed by Rogers and Bendrich [16]. Four-cm-long shoots were grown in the dark for 6 days and the DNA was isolated from meristem roots. The DNA was enzymatically hydrolysed with Sau3A and the restriction products of 100-1000 bp in length were cloned into the BamHI restriction site of pUC18 and transformed into *E. coli* NM522 competent cells yielding a collection of 1600 clones. All clones were derived from a single transformation experiment. Using a dot blot hybridisation technique, thirteen clones displaying a strong signal against a γ-^{32}P-dATP radiolabelled $(GA)_{15}$ probe were selected. Cloned inserts were sequenced by the primer extension method (SequiTherm Cycle Sequencing Kit, Epicentre Technologies), using BigDye terminator kit and analysed on an automatic DNA sequencer (ABI Prism 377).
Analyses of the genetic instability of clones harbouring $(GA/TC)_n$ motifs contained on plasmids after multiple recultivations of *E. coli* KA796 cells (E. coli Genetic Stock Centre, http://cgsc.biology.yale.edu) were performed on 1,5% agarose gels in TAE buffer [17]. The cells were grown overnight and then subcultured into fresh LB medium with dilution factor of 10^7. Plasmid DNA was purified and restriction analysis with EcoRI and HindIII identified the fragment containing the pUC18 polylinker and the $(GA/TC)_n$ repeats cloned within it. Searches for the $(GA/TC)_n$ microsatellites in plant families were performed at the Plant Repeat Database at TIGR (http://www.tigr.org/tdb/e2k1/plant.repeats).

RESULTS AND DISCUSSION

DNA sequence analyses of plasmids containing $(GA/TC)_n$ inserts

Sequences containing a variety of sizes of $(GA/TC)_n$ were isolated from *V. bithynica* and cloned into bacterial plasmids. The presence and estimation of length of the cloned inserts was confirmed after restriction analyses of all plasmids that showed a strong hybridisation signal with the $(GA)_{15}$ probe (thirteen clones, not shown). Then, the detailed nucleotide sequence of the clones was determined. However, a primer extension sequencing technique employed for such determination revealed the presence of very strong, possibly structural aberrations that caused DNA polymerase to pause at specific sites of the analysed $(GA/TC)_n$ tracts (Fig.1A and 1B).
There were no difficulties in obtaining the sequences that flank both sides of the repeat tract. Moreover, plasmids containing relatively short stretches of uninterrupted $(GA/TC)_n$ (n = 17, 19, 23) could be easily sequenced compared to molecules harbouring longer tracts (n = 29, 31, 41).

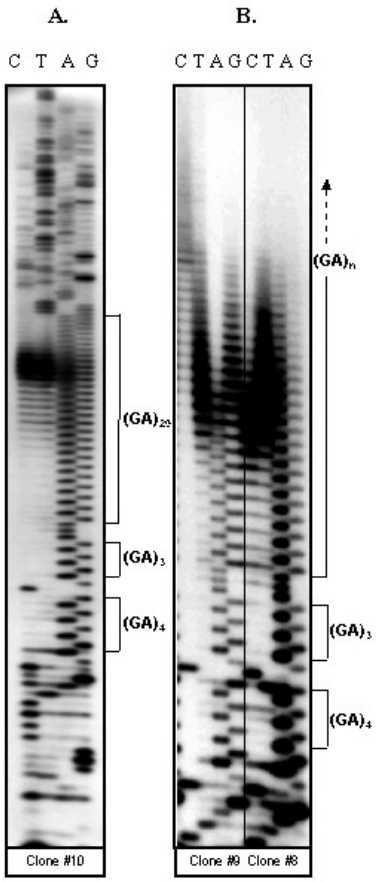

Fig. 1. Sequencing analyses performed using PCR primer extension method. The figure illustrates moderate sequencing aberrations observed within $(GA)_{29}$ tract in clone #10 (A), and strong aberrations in the case of longer $(GA)_n$ motifs in clones #9 and #8 (B).

Therefore, the emergence of structural aberrations was directly proportional to the number of repeated motifs. Although, this kind of examination is not widely used to confirm the presence of alternative non-B-DNA structures, similar analyses have shown that many DNA polymerases pause within long stretches of CTG/CAG and CGG/CCG *in vitro* [7, 12]. Such pausing may cause a primer-template realignment, which may be related to the propensity of repeats to undergo deletions and expansions. Also, poly-(AG/CT) tracts in plasmids have been shown to form very stable intramolecular triplexes *in vitro* and *in vivo* [13, 14, 18, 19] and long GAA/TTC motifs were demonstrated to form an alternative DNA structure called "sticky DNA" [20-22].

Using an automatic DNA sequencer allowed precise determination of the nucleotide composition of all $(GA/TC)_n$ containing plasmids. Details of the sequence composition of thirteen analysed DNA fragments identified within the genomic mini-library are shown in Tab. 1.

Tab.1. Sequence composition of the $(GA)_n$ repeats identified within the genomic mini-library of *V. bithynica.* RI-RV – repetitive motifs; S1-S4 – spacer regions; 5'-F and 3'-F – flanking sequences.

Class	Clone no.	Name / Sequence											Total no. of $(GA)_n$ repeats
		5'-F GATCACAA AAATTGTGT	RI $(GA)_n$	S1 TA	RII $(GA)_n$	S2 GTGA CA	RIII $(GA)_n$	S3 AA	RIV $(GA)_n$	S4 GC	RV $(GA)_n$	3'-F GC	
A	1	+	24	+	11	+	17	+	3	+	4	+	59
	2	+	24	+	11	+	15	+	3	+	4	+	57
	3	+	63	-	-	+	22	+	3	+	4	+	92
B	4	+	52	-	-	+	22	+	3	+	4	+	81
	5	+	44	-	-	+	22	+	3	+	4	+	73
	6	+	10	-	-	+	22	+	3	+	4	+	39
C	7	+	31	-	-	+	21	-	-	- C	4	+	56
	8	+	41	-	-	-	-	+	3	+	4	+	48
	9	+	31	-	-	-	-	+	3	+	4	+	38
D	10	+	29	-	-	-	-	+	3	+	4	+	36
	11	+	23	-	-	-	-	+	3	+	4	+	30
	12	+	19	-	-	-	-	+	3	+	4	+	26
	13	+	17	-	-	-	-	+	3	+	4	+	24

The nature of the analysed $(GA/TC)_n$ repeats containing inserts is composite. The top of the table shows the sequence of clone 1, which is the most complex and is composed of a 5' flanking region (5'-F), $(GA/TC)_n$ repeats (R), spacer regions (S), and a 3' flanking region (3'-F). Based on their complexity, the clones have been divided into four subclasses (A-D). Class A is most complex. Like the other sequences it contains a 17-nt long 5' flanking sequence, but it also contains all five $(GA/TC)_n$ areas (RI-RV) as well as four spacer regions (S1-S4) and 3' flanking GC nucleotides. There are two clones subsumed into this class differing in their number of $(GA/TC)_n$ repeats within region RIII (seventeen and fifteen units, respectively). Class B includes four DNA motifs, which lack the S1 spacer region and neighbouring RII $(GA/TC)_n$ repeat. Within this class the clones contain the same number of repeats in regions RIII, RIV, and RV (22, 3, and 4, respectively), but they differ in length of their RI region. Clones 3-6 embody 63, 52, 44, and 10 $(GA/TC)_n$ units, respectively, in RI. Class C differs from other classes and has only one clone included in it; clone 7 lacks S1, S3, part of the S4 spacer, RII and RIV repeats. Class D appears to be the most homogeneous group of clones, with six clones falling into this category (clones 8-13). They contain RI, RIV and RV repeat regions and they lack RII and RIII as well as S1 and S2 spacers. Sequences belonging to this class differ in the number of their $(GA/TC)_n$ units within the RI region and contain 41, 31, 29, 23, 19, and 17 units, respectively. A schematic representation of the sequence classes described above is shown in Fig. 2.

BLAST searches performed on plant sequences available at the NCBI database revealed sequences with some similarity to those discussed above. However, homology of the recognized fragments did not exceed 17 nucleotides within the 5'-flanking sequence and the number of subsequent $(GA)_n$ motifs was smaller than 10. Thus, these classes of repeats have not previously been characterized in the level of detail reported here.

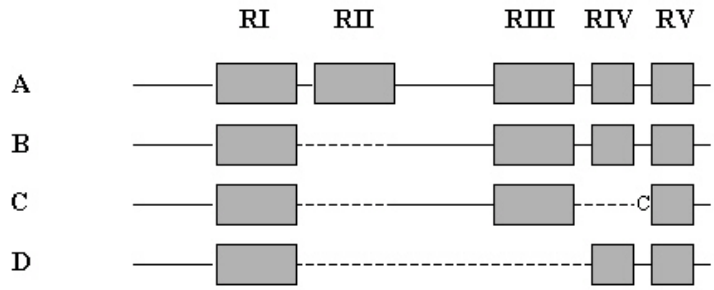

Fig. 2. Schematic representation of four classes of the $(GA)_n$ motifs from *V. bithynica.* Grey boxes correspond to the RI-RV (GA) repetitive motifs, solid lines represent sequences surrounding the R blocks, dashed lines correspond to sequences not present in classes B, C, and D as compared to class A.

The fact that all of the analysed motifs contain identical regions at the 5'-end (i.e. 5'-flank, and first $(GA/TC)_n$ block) as well as at the 3'-end of the sequence (i.e. last $(GA/TC)_n$ block and flanking GC nucleotide) and to a different extent share the regions in between, may indicate that they have a common origin. There are differences in internal regions, but recombination processes could have produced partial deletions of these parts. Intra- and intermolecular recombination studies showed that the frequency of recombination between similar, GAA/TTC, tracts was 15 times higher than for non-repeating control sequences [20]. The same consequence could have been due to erroneous replication of the internal regions as a result of formation of non-B-DNA structures within a poly-$(GA)_n$ stretch.

This could have caused DNA polymerase to pause, resulting in various rearrangements of the sequence. Since all clones have the same flanking sequence, it is possible that the different repeat tracts have arisen during cloning procedures in *E. coli*. However, it is also possible that the common ancestry may have arisen due to duplications in the plant genome. Identification and characterisation of sites of $(GA/TC)_n$ in the plant genome should elucidate whether the plant cell has generated a diversity of repeats among common flanking sequences.

Analysis of the genetic stability of plasmids containing $(GA/TC)_n$ inserts

To analyse the genetic stability of the DNA repeats, we used *E. coli* KA796. This strain is wild-type for mismatch repair and has been shown to induce high rates of genetic instability in repeating DNA sequences. [17]. Two plasmids

belonging to class D differing in length of the $(GA/TC)_n$ tract in region RI were transformed into *E. coli* KA796. Clone no. 8 consisted of 41 units of the repeat and clone no. 11 harboured 23 $(GA/TC)_n$ units (Table 1). These plasmids were chosen for further analyses because of the lower internal complexity of the sequence, as compared to sequences from other classes.

Bacterial cells harbouring appropriate plasmids were subjected to multiple recultivations into liquid media. This type of analysis [23] has proven its usefulness in the determination of the genetic stability of trinucleotide repeat sequences [17, 24, 25]. After each overnight growth, which is equivalent to 22-24 generations (cell divisions), the genetic instability of the TRS increased as manifested by accumulation of deletion products. Therefore, in this study the same assay was performed to evaluate the genetic stability of $(GA/TC)_n$ motifs (Fig. 3).

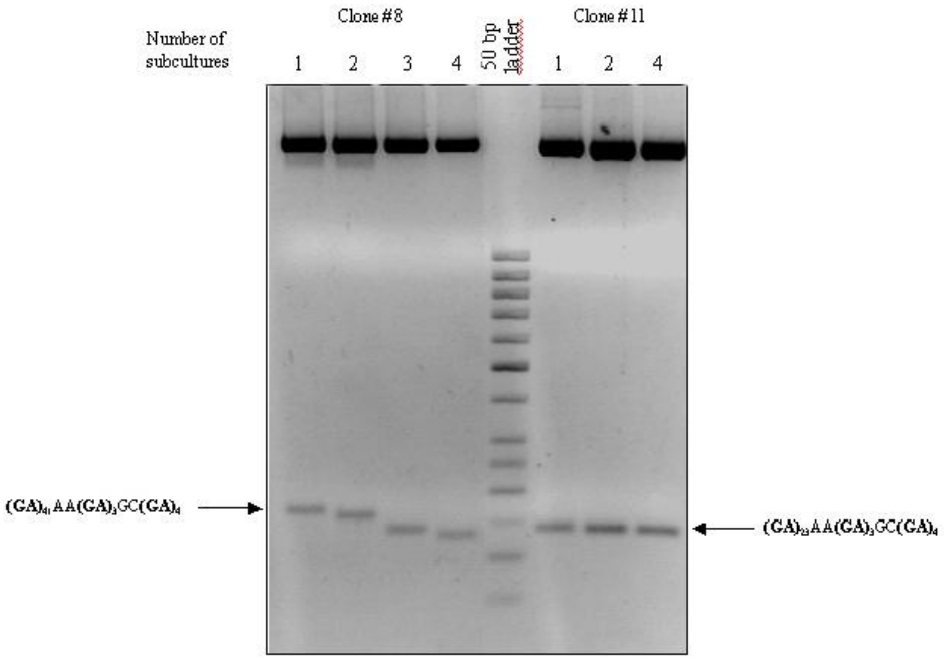

Fig. 3. Genetic instability of repeats harboured on plasmids after a number of recultivations. Two clones harbouring $(GA)_n$ motifs of different lengths on plasmids were subjected to multiple recultivations. EcoRI and HindIII double digestion products of plasmids isolated after the number of recultivations indicated were separated on a 1,5% agarose gel run in TAE buffer. Arrows show initial lengths of $(GA)_n$ containing sequences from clones #8 and #11.

The plasmid harbouring the $(GA)_{41}AA(GA)_3GC(GA)_4$ insert, with the longer uninterrupted tract of the (GA/TC) motif (41 units in RI) displayed high instability as manifested by gradual shortening of the insert in subsequent recultivations. During the same experimental conditions, the plasmid containing

the shorter repetitive sequence (23 GA/TC units in RI) remained stable. Therefore, the length of the repetitive motif was essential for the genetic instability of (GA/TC) motifs in plasmids in *E. coli*. In similar experiments in *E. coli*, the frequency of deletions in plasmid-based TRS was also observed to increase in larger repeat tracts [17, 23-25].

The question remains as to whether (GA/TC) motifs are also genetically unstable in the genome of *V. bithynica*. Currently, there is no direct evidence for this phenomenon. The similarities in sequence composition of the different clones containing $(GA/TC)_n$ motifs suggest that they have a common origin. These observations could reflect changes induced during cloning, but they could indicate that related sequences exist in many regions of the genome of *V. bithynica*. Future investigations will focus on the precise localization of the motifs in the genome, which might be helpful in understanding the functional significance of such repetitive motifs. Also, further study will analyse whether differences in the internal complexity of the $(GA/TC)_n$ motifs could have resulted from numerous processes acting on DNA including replication, recombination, repair or transcription of such motifs.

Acknowledgements. We thank Iwona Kawecka for assistance with these experiments. Funding was provided by the University of East Anglia and the Centre for Medical Biology & Microbiology PAS.

REFERENCES

1. Katti, M.V., Ranjekar, P.K. and Gupta, V.S. Differential distribution of simple sequence repeats in eukaryotic genome sequences. **Mol. Biol. Evol.** 18 (2001) 1161-1167.
2. Li, Y.C., Korol, A.B., Fahima, T., Beiles, A. and Nevo, E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: A review. **Mol. Ecol.** 11 (2002) 2453-2465.
3. Charlesworth, B., Sniegowski, P. and Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature** 371 (1994) 215-220.
4. Symonds, V.V. and Lloyd, A.M. An analysis of microsatellite loci in *Arabidopsis thaliana*: mutational dynamics and application. **Genetics** 165 (2003) 1475-1488.
5. Arabidopsis genome initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. **Nature** 408 (2000) 796-815.
6. Bowater, R.P. and Wells, R.D. The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. **Prog. Nucleic Acid Res. Mol. Biol.** 66 (2001) 159-202.
7. Mitas, M. Trinucleotide repeats associated with human disease**. Nucleic Acids Res.** 25 (1997) 2245-2254.
8. Sinden, R.R., Potaman, V.N., Oussatcheva, E.A., Pearson, C.E., Lyubchenko, Y.L. and Shlyakhtenko, L.S. Triplet repeat DNA structures

and human genetic disease: dynamic mutations from dynamic DNA. **J. Biosci.** 27 (2002) 53-65.
9. Fishel, R. and Kolodner, R.D. Identification of mismatch repair genes and their role in the development of cancer. **Curr. Opin. Genet. Dev.** 5 (1995) 382-395.
10. Komarova, N.L., Lengauer, C., Vogelstein, B. and Nowak, M.A. Dynamics of genetic instability in sporadic and familial colorectal cancer. **Cancer Biol. Ther.** 1 (2002) 685-692.
11. Rajagopalan, H., Nowak, M.A., Vogelstein, B. and Lengauer, C. The significance of unstable chromosomes in colorectal cancer. **Nat. Rev. Cancer** 3 (2003) 695-701.
12. Bzymek, M. and Lovett, S.T. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. **Proc. Natl. Acad. Sci. U.S.A.** 98 (2001) 8319-8325.
13. Kohwi, Y., Malkhosyan, S.R. and Kohwi-Shigematsu, T. Intramolecular dG.dG.dC triplex detected in *Escherichia coli* cells. **J. Mol. Biol.** 223 (1992) 817-822.
14. Parniewski, P., Kwinkowski, M., Wilk, A. and Klysik, J. Dam methyltransferase sites located within the loop region of the oligopurine-oligopyrimidine sequences capable of forming H-DNA are undermethylated *in vivo*. **Nucleic Acids Res.** 18 (1990) 605-61.
15. Powell, W., Machray, G.C. and Prowan, J. Polymorphism revealed by simple sequence repeats. **Trends Plant Sci.** 1 (1996) 215-222.
16. Rogers, S.O. and Bendrich, A. DNA extraction. **Plant Molecular Biology Manual** (S.B.Gelvin, S.B. and Schilperoort, R.A. Eds). Kluwer Academic Publishers, Dordrecht-Belgium, 1988, A6, 1-10.
17. Bowater, R.P., Jaworski, A., Larson, J.E., Parniewski, P. and Wells, R.D. Transcription increases the deletion frequency of long CTG.CAG triplet repeats from plasmids in *Escherichia coli*. **Nucleic Acids Res.** 25 (1997) 2861-2868.
18. Rustighi, A., Tessari, M.A., Vascotto, F., Sgarra, R., Giancotti, V., and Manfioletti, G. A polypyrimidine/polypurine tract within the Hmga2 minimal promoter: a common feature of many growth-related genes. **Biochemistry** 41 (2002) 1229-1240.
19. Vigneswaran, N., Thayaparan, J., Knops, J., Trent, J., Potaman, V., Miller, D.M. and Zacharias, W. Intra- and intermolecular triplex DNA formation in the murine c-myb proto-oncogene promoter are inhibited by mithramycin. **Biol. Chem.** 382 (2001) 329-342.
20. Napierala, M., Dere, R., Vetcher, A. and Wells, R.D. Structure-dependent recombination hot spot activity of GAA.TTC sequences from intron 1 of the Friedreich's ataxia gene. **J. Biol. Chem.** 279 (2004) 6444-6454.
21. Sakamoto, N., Chastain, P.D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J.D., and Wells, R.D. Sticky DNA: self-association properties of

long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. **Mol. Cell** 3 (1999) 465-475.
22. Vetcher, A.A. and Wells, R.D. Sticky DNA formation *in vivo* alters the plasmid dimer/monomer ratio. **J. Biol. Chem.** 279 (2004) 6434-6443.
23. Bowater, R.P., Rosche, W.A., Jaworski, A., Sinden, R.R. and Wells, R.D. Relationship between *Escherichia coli* growth and deletions of CTG.CAG triplet repeats in plasmids. **J. Mol. Biol.** 264 (1996) 82-96.
24. Jaworski, A., Rosche, W.A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R.P., Sinden, R.R. and Wells, R.D. Mismatch repair in *Escherichia coli* enhances instability of (CTG)n triplet repeats from human hereditary diseases. **Proc. Natl. Acad. Sci. U.S.A.** 92 (1995) 11019-11023.
25. Rosche, W.A., Jaworski, A., Kang, S., Kramer, S.F., Larson, J.E., Geidroc, D.P., Wells, R.D. and Sinden, R.R. Single-stranded DNA-binding protein enhances the stability of CTG triplet repeats in *Escherichia coli*. **J. Bacteriol.** 178 (1996) 5042-5044.