

Benchmarks

Interruption of Perfect CAG Repeats by CAA Triplets Improves the Stability of Glutamine-Encoding Repeat Sequences

BioTechniques 33:976-978 (November 2002)

Several disorders are caused by the genomic expansion of a trinucleotide repeat beyond a specific critical size in the coding region of genes. The largest group of these diseases, which includes Huntington's disease and the spinocerebellar ataxias of type 1, 2, 3, 6, and 7, are caused by expansion of a CAG repeat resulting in proteins with an expanded polyglutamine stretch (14). A correlation between the lengths of the repeat and the age of onset of the diseases has been found. However, functional studies and the generation of cell and animal models expressing very long repeats have been hampered by the significant instability of long perfect CAG repeats in *E. coli* (2,9,10,13) and the constraints on the size of homogeneous CAG repeats that can be efficiently and accurately produced using PCR (1,13). Several PCR-based methods for cloning long repeats and introducing these repeats in target genes have, nevertheless, been described, but these methods have not resulted in the generation of stable clones with very long repeats (5,6,11). We now report on a strategy that exploits both the redundancy of the genetic code and the special properties of type II restriction enzymes, which cleave at an exact distance outside their recognition site to generate constructs expressing long homopolymeric amino acid repeats. The method described here is fast, easy to perform, and results in stable constructs encompassing very long repeats.

First, we tested whether the stability problems of homogeneous CAG repeats could be solved by the interruption of a perfect repeat at the DNA level by another triplet encoding the same amino acid (i.e., for polyglutamine interruption of a CAG repeat with CAA). Mixed CAG/CAA repeats were constructed using two complementary synthetic DNA

oligonucleotides 5'-GCAGCAACAGC-AGCAGCAACA-3' and 5'-TGCT-GTTGCTGCTGCTGTTGC-3'. Oligonucleotides were phosphorylated with T4 polynucleotide kinase (Amersham Biosciences, Piscataway, NJ, USA), hybridized and ligated using T4 DNA ligase (Roche Applied Science, Indianapolis, IN, USA). Ligation yields a 21-bp repeat with the basic sequence [(CAG)²-CAA-(CAG)³-CAA]_n. The ligated fragments were treated with AmpliTaq[®] DNA polymerase (Applied Biosystems, Foster City, CA, USA) to create an A extension, followed by electrophoresis on a 2% agarose gel, and the isolation of fragments in the 300–400 bp size range using the Nucleotrap extraction kit (Macherey-Nagel, Düren, Germany). The purified fragments were subcloned in the pGEM[®]-T Easy vector containing T extensions (Promega, Madison, WI, USA) and transformed to STBL-2 cells (Invitrogen, Carlsbad, CA, USA). The insert lengths of the clones obtained were determined with

PCR (buffer: 16.6 mM [NH₄]SO₄, 67 mM Tris-HCl, pH 8.8, 10 mM β-mercaptoethanol, 6.7 μM EDTA, and 170 μg/mL BSA) using M13-primers, and both strands of the three largest inserts were sequenced with the T7-primer and the M13-reverse primer, respectively. One clone, designated p98Gln, contained a mixed repeat encoding a perfect Gln-repeat of 98 residues. The two other clones had one or more variations that caused an interruption of the encoded Gln-repeat sequence.

To test the stability of the insert in *E. coli*, p98Gln was transformed to a bacterial strain often used in cloning experiments, XL-1 Blue[®] (Stratagene, La Jolla, CA, USA). Plasmid DNA was isolated from independent colonies, and the insert lengths were determined with restriction enzyme digestion. Nine out of 10 contained an insert of the appropriate length; one clone contained a deletion of about 20 bp (probably one 21-mer repeat unit). To determine the insert lengths of the first nine clones more ac-

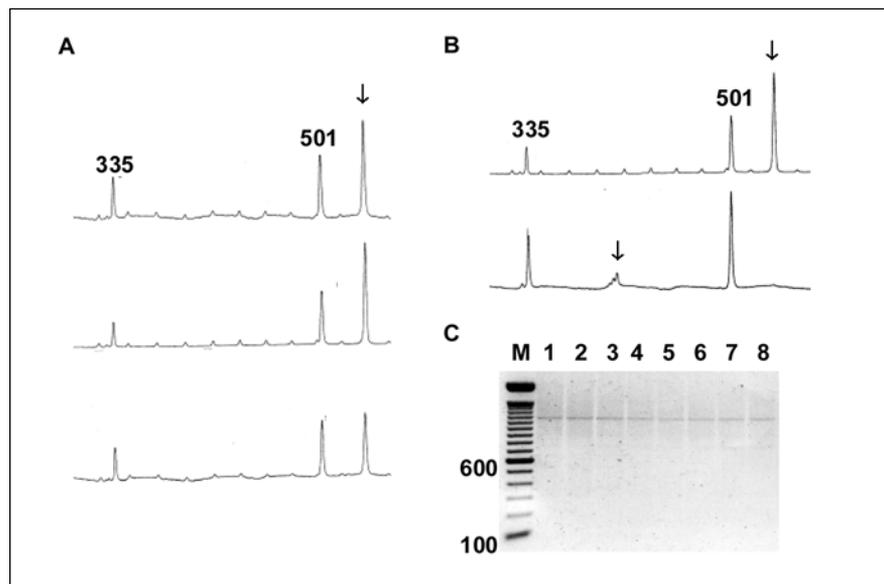


Figure 1. Comparison of perfect CAG and mixed CAG/CAA repeats. (A) Analysis of lengths of the 98 mixed repeat present in independent plasmid preparations of p98Gln. Insert sequences present in the independent preparations of p98Gln were amplified by PCR with fluorescent primers. The lengths of the resulting fragments were determined with ALF sequencer analysis. An arrow indicates the position of the peaks derived from the p98Gln fragments. The other two peaks indicated with numbers correspond to the internal standards used for sizing measuring 335 and 501 bp respectively. (B) Comparison of PCR amplification of a perfect and a 98 mixed repeat. PCR fragments containing a perfect 73 CAG repeat and a 98 mixed repeat were compared. The relevant peaks for both fragments are indicated with an arrow. The two other peaks indicated with numbers correspond to the internal standards used for sizing. (C) Analysis of lengths of approximately 300 mixed repeats present in independent plasmid preparations of p291Gln. Insert sequences present in the independent preparations of p291Gln were amplified by PCR (lane 1–8). The lengths of the resulting fragments were determined with agarose gel electrophoresis. The positions of the 100- and 600-bp fragments of the DNA molecular weight marker are indicated (lane M).

curately, repeat sizes were determined via automated laser fluorescence (ALF) sequencer analysis (Amersham Biosciences) after PCR amplification using fluorescent oligonucleotide primers. Figure 1A shows an example of such an experiment. The three plasmids shown contain exactly the same length of insert, as visualized by peaks at exactly the same position. The other six clones show exactly the same pattern (results not shown). Perfect repeats in the same size range, including 98-repeat sequences, can already be unstable (10).

In addition, we compared the pattern obtained after ALF analysis for a mixed repeat and a perfect CAG repeat (Figure 1B). In agreement with previous studies, PCR amplification across perfect CAG repeats results in a peak at the expected position with a typical "stutter", which is probably caused by trinucleotide slippage during PCR (1,12,13). In contrast, mixed repeats gave rise to very sharp peaks at the expected position without any stutter (Figure 1B). Moreover, fragments containing long mixed repeats always give rise to much higher peaks in ALF experiments than fragments containing long perfect CAG repeats (Figure 1B). A similar difference in yield is also observed when PCR fragments were analyzed using agarose gel electrophoresis. These results suggest that interruption of a CAG repeat by CAA sequences significantly alters the accuracy and efficiency of its replication by *Taq* DNA polymerase.

Since the mixed-repeat approach was successful for an approximately 100-repeat sequence, we decided to clone even longer mixed-repeat stretches. Plasmid p98Gln was used as the starting point. PCR forward primers 5'-AAGGATCCGGATGGAATTCGATT-TGCTGTTG-3' and reverse primers 5'-AAGGATCCGGATGTAGTGATT-AGCAACAGCAG-3' were designed. These primers contain the recognition sequence for the type II restriction enzyme *FokI*, which cleaves at an exact distance outside its recognition site (5'-GGATG-(N)9-3'/5'-(N)13CATCC-3'), followed by sequences complementary to vector sequences of p98Gln allowing a specific priming on sequences flanking the mixed-repeat sequences. Thus, PCR results in a fragment encompassing the 98-repeat sequence flanked by vec-

tor sequences, which can be removed by *FokI* digestion. For the generation of plasmids encompassing longer repeats, PCR was performed as described above, and the resulting fragments were subsequently cleaved with *FokI*, purified, and ligated, followed by treatment with AmpliTaq DNA polymerase and ligation into the pGEM-T Easy vector. This procedure resulted in various clones containing three times the original insert size (about 300 repeats). Again we did not observe any instability in *E. coli*, as detected by PCR amplification of the inserts (Figure 1C). The very high stability of the mixed approximately 300 repeats is noteworthy, since perfect repeats in the 140 bp size range (data not shown) and in the 180–230 bp size range (9,13) have already been found to be very unstable. An independent approach to generate mixed CAG/CAA repeats and to clone these repeats in target genes has also been reported (4,8), although no data on DNA stability in *E. coli* or results of PCR analysis were presented. The method described here has the advantage that the repeats can be cloned without unwanted flanking sequences by simple digestion via a type II restriction enzyme.

The increased stability of mixed-repeat sequences should facilitate the generation of transgenic animals and cell models carrying considerably larger and more stable repeats than currently used (3,7,14). The use of such larger sequences may accelerate the disease process, a feature that may be especially advantageous in transgenic animals with a short lifespan (e.g., *Drosophila melanogaster*) and in cell models. The use of larger sequences may cause more severe (lethal) phenotypes allowing, for example, high-throughput drug testing studies (14). Furthermore, the mixed repeat at the DNA level results in a perfect repeat at the protein level and thus provides a unique tool to dissect the deleterious effects at the nucleotide level (DNA instability and RNA effects) from those at the protein level. The approach of mixed DNA repeat sequences encoding the same amino acid is also applicable to generate proteins containing repeats of any other amino acid, with the exception of methionine and tryptophan, thereby facilitating studies to dissect their potential deleterious effects.

REFERENCES

1. Bates, G.P., L. Mangiarini, A. Mahal, and S.W. Davies. 1997. Transgenic models of Huntington's disease. *Hum. Mol. Genet.* 6:1633-1637.
2. Bowater, R.P., A. Jaworski, J.E. Larson, P. Parniewski, and R.D. Wells. 1997. Transcription increases the deletion frequency of long CTG/CAG triplet repeats from plasmids in *Escherichia coli*. *Nucleic Acids Res.* 25:2861-2868.
3. Gomes-Pereira, M., M.T. Fortune, and D.G. Monckton. 2001. Mouse tissue culture models of unstable triplet repeats: in vitro selection for larger alleles, mutational expansion bias and tissue specificity, but no association with cell division rates. *Hum. Mol. Genet.* 10:845-854.
4. Kazantsev, A., E. Preisinger, A. Dranovsky, D. Goldgaber, and D. Housman. 1999. Insoluble detergent-resistant aggregates form between pathological and nonpathological lengths of polyglutamine in mammalian cells. *Proc. Natl. Acad. Sci. USA* 96:11404-11409.
5. Laccone, F., R. Maiwald, and S. Binge-mann. 1999. A fast polymerase chain reaction-mediated strategy for introducing repeat expansions into CAG-repeat containing genes. *Hum. Mutat.* 13:497-502.
6. Lin, C.-H., S. Tallaksen-Greene, W.-M. Chien, J.A. Cearley, W.S. Jackson, A.B. Crouse, S. Ren, X.-J. Li, et al. 2001. Neurological abnormalities in a knock-in mouse model of Huntington's disease. *Hum. Mol. Genet.* 10:137-144.
7. Mangiarini, L., K. Sathasivam, R. Mahal, M. Mott, M. Seller, and G.P. Bates. 1997. Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat. Genet.* 15:197-200.
8. Michalik, A., A. Kazantsev, and C. van Broeckhoven. 2001. Method to introduce stable, expanded, polyglutamine-encoding CAG/CAA trinucleotide repeats into CAG repeat-containing genes. *BioTechniques* 31:250-254.
9. Ordway, J.M. and P.J. Detloff. 1996. In vitro synthesis and cloning of long CAG repeats. *BioTechniques* 21:609-612.
10. Parniewski, P., A. Jaworski, R.D. Wells, and R.P. Bowater. 2000. Length of CTG/CAG repeats determines the influence of mismatch repair on genetic instability. *J. Mol. Biol.* 299:865-874.
11. Peters, M.F. and C.A. Ross. 1999. Preparation of human cDNAs encoding expanded polyglutamine repeats. *Neurosci. Lett.* 275:129-132.
12. Petruska, J., M.J. Hartenstine, and M.F. Goodman. 1998. Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *J. Biol. Chem.* 273:5204-5210.
13. Sathasivam, K., I. Amaechi, L. Mangiarini, and G. Bates. 1997. Identification of an HD patient with a (CAG)180 repeat expansion and the propagation of highly expanded CAG repeats in lambda phage. *Hum. Genet.* 99:692-695.
14. Zoghbi, H.Y. and H.T. Orr. 2000. Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* 23:217-247.

Benchmarks

This work was supported in part by the Netherlands Organization for Scientific Research (NWO) through the Foundation for Medical Sciences in the Netherlands (MW) and by the European Community (PL 960244) and by the Princess Beatrix Fund (99-0114). Address correspondence to Dr. Josephine C. Dorsman, MCB1/HCG, Sylvius Laboratory, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands. e-mail: j.c.dorsman@lumc.nl

Received 20 August 2001; accepted 22 July 2002.

J.C. Dorsman, M. Bremmer-Bout, B. Pepers, G.-J.B. van Ommen, and J.T. Den Dunnen
Leiden University Medical Center
Leiden, The Netherlands

For reprints of this or any other article, contact Reprints@BioTechniques.com

Gap Repair Transformation in Fission Yeast to Exchange Plasmid-Selectable Markers

Biotechniques 33:978-982 (November 2002)

The fission yeast *Schizosaccharomyces pombe* is an attractive model organism for studying eukaryotic biology due to the ease of genetic and molecular manipulations (3). In recent years, several selectable markers and promoters have been developed to enhance our ability to manipulate gene expression in this yeast. Recently, Invitrogen (Carlsbad, CA, USA) has developed the SpECTRA™ *S. pombe* Expression System that consists of three TOPO® cloning vectors that allow for rapid cloning of PCR products, with the option to epitope-tag the gene product (adding a carboxy-terminal V5-6his tag). Transcription of cloned open reading frames is under the control of the thiamine-repressible promoters *nmt1*, *nmt41*, and *nmt81*. (Promoters *nmt41* and *nmt81* are TATA box mutant derivatives of *nmt1* that allow for different

levels of transcription.) However, as all three plasmids utilize the same selectable marker (simian virus 40-driven *Saccharomyces cerevisiae* *LEU2*, which complements an *S. pombe leu1* mutant allele), use of these vectors limits the choice of host strains and the possibility to transform strains with more than one plasmid construct.

To enhance the use of the SpECTRA vectors by allowing for the transformation of *Leu*⁺ strains or for the introduction of two distinct plasmids into a host strain, we have developed a protocol that facilitates the exchange of the *LEU2* marker for another selectable marker. This protocol takes advantage of the robust homologous recombination machinery found in yeasts, which has been used in the past to rescue chromosomal mutations onto plasmids by gap repair (8) or to screen mutant alleles created by PCR by co-transformation of the PCR product with a linearized plasmid (5,6). In both situations, the plasmid DNA is linearized before transformation and is repaired and recircularized by DNA that bears homology to the ends of the linearized plasmid. Using a protocol described here, we can readily replace the *LEU2*-selectable

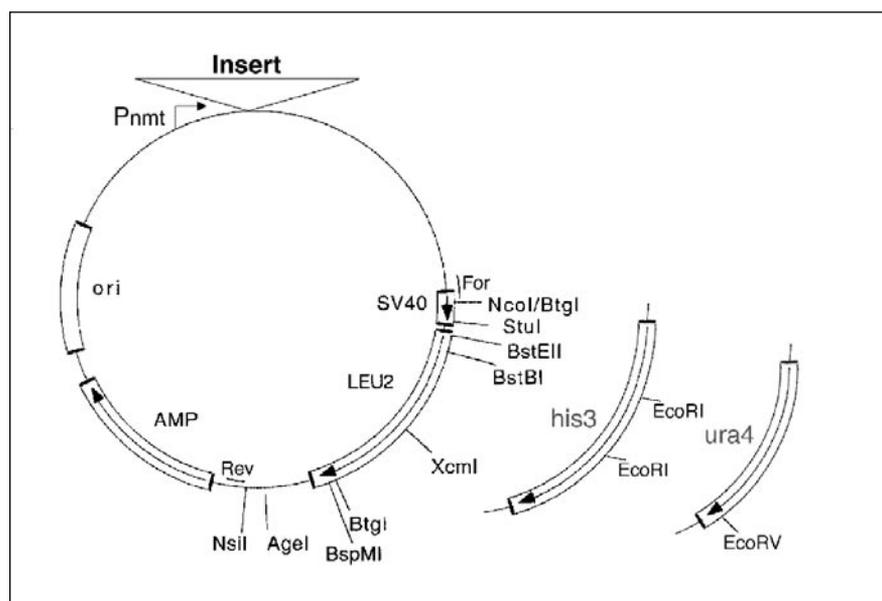


Figure 1. Schematic diagram of SpECTRA plasmid. The three SpECTRA plasmids differ only within the TATA box of the *nmt1*, *nmt41*, and *nmt81* promoters. The regions homologous to the PCR oligonucleotides used for the marker exchange are indicated as "For" and "Rev". Restriction sites that are unique to this region, and thus suitable for gapping the plasmid within the *LEU2*-selectable marker, are indicated. Also shown are schematics of the PCR products carrying either the *his3* gene (containing a unique *EcoRI* site indicated) or the *ura4* gene (containing two *EcoRI* sites). Not indicated are two *EcoRI* sites in the SpECTRA plasmid that flank the cloning site.