Transcription and nuclear transport of CAG/CTG trinucleotide repeats in yeast

Emmanuelle Fabre*, Bernard Dujon and Guy-Franck Richard

Unité de Généétique Moléculaire des Levures (URA 2171 CNRS and UFR 927 Université Pierre et Marie Curie) Institut Pasteur, 25 rue du Dr Roux, 75724 Paris cedex 15, France

Received May 21, 2002; Revised and Accepted June 26, 2002

ABSTRACT

Trinucleotide repeats are involved in several neurological disorders in humans. DNA sequences containing CAG/CTG repeats are prone to slippage during replication and double-strand break repair. The effects of trinucleotide repeats on transcription and on nuclear export were analyzed in vivo in yeast. Transcription of a CAG/CTG trinucleotide repeat in the 3'-untranslated region of a URA3 reporter gene leads to transcription of messenger RNAs several kilobases longer than the expected size. These long mRNAs form more readily when CAG rather than CTG repeats are transcribed. CAG- or CUG-containing transcripts show a non-homogeneous cellular localization. We propose that long mRNAs result from transcription slippage, and discuss the possible implications for human diseases.

INTRODUCTION

Among human neurological disorders linked to expansions of trinucleotide repeats, the most frequent triplet found to be expanded is CAG and its complementary sequence CTG (1). CAG repeats are almost always found in open reading frames (2), whereas CTG expansions are located in the 3'-untranslated region (3'-UTR) of the DMPK gene in myotonic dystrophy type 1 (3,4) and in SCA8 (5). Transcription of the DMPK expanded allele is not significantly reduced in humans. Yet, it was found that CUG expansions impaired pre-mRNA processing (6) and that CUG-containing RNAs accumulated in the nucleus as foci (7–9). Myotonic dystrophy 2, caused by the expansion of a CCTG tetranucleotide repeat, also exhibits nuclear retention of transcripts (10). Since proteins belonging to a family involved in muscular differentiation specifically bind CUG repeat-containing RNAs, it was suggested that abnormal recruitment of these proteins would be responsible for DM-1 pathology (11). Nuclear retention is specific to CUG expansions impaired pre-mRNA processing, since CAG repeat-containing mRNAs, in other trinucleotide repeat disorders, are apparently correctly exported.

CTG repeats induce transient pauses in RNA elongation in vitro (12). However, it is not known if similar pauses occur in vivo, and could promote slippage of the transcript during elongation. Transcription slippage has been observed in vitro (13–16), in vivo in Escherichia coli (17,18), and in Alzheimer and Down patients (19,20), but was never reported for trinucleotide repeats.

We used the yeast Saccharomyces cerevisiae to investigate the effects of long CAG or CTG repeats on transcription and on nucleocytoplasmic transport of a reporter mRNA containing trinucleotide repeats in its 3'-UTR. We found CAG- and CUG-containing transcripts longer than the expected size. Formation of these long transcripts is sequence and length dependent. In addition, we show that these mRNAs are not evenly distributed in the cell. We discuss the possible implications of these two observations on the human pathologies associated with CAG/CTG repeats.

MATERIALS AND METHODS

Plasmids and strains

Plasmid pCMha191 was derived from pCM184 (21) by replacing the CEN-ARS fragment with a PCR-generated 2 µm replication origin and by cloning an oligonucleotide containing the HA tag downstream of the CYC1 promoter (J.Boyer, unpublished results). The URA3 gene was amplified by PCR using Pfu DNA polymerase (Stratagene) and cloned in frame with the HA tag in pCMha191 to give plasmid pE186. pTRI625 was constructed by cloning a 300 bp PstI piece of λ DNA into a PstI site located in the URA3 3'-UTR of pE186 (Fig. 1A). pTRI626 was constructed by cloning a 700 bp PstI piece of λ DNA into the same site. pTRI627 was constructed by cloning a 300 bp BamHI–XbaI DNA fragment containing a (CAG)98 trinucleotide repeat from pTRI620 (J.Boyer and B.Dujon, unpublished data) between the BglII and NheI sites located in the URA3 3'-UTR of pE186. pTRI628 was constructed by cloning a 300 bp BamHI–PstI DNA fragment containing a (CTG)98 trinucleotide repeat from pRW3216 (22) between the BglII and NheI sites located in the URA3 3'-UTR of pE186. pTRI629 was constructed by cloning a 780 bp DNA fragment containing a (CTG)255 trinucleotide repeat from pRW3222 (22) between the same sites in the same plasmid. pTRI637 was constructed by cloning a 300 bp (TTG)87 trinucleotide repeat from pTRI622 (23) between the PstI and NheI sites located in the URA3 3'-UTR of pE186. Transformations of plasmids pE186 and pTRI625–637 were performed in the FYBL1-22B strain (MATa ura3Δ851 trp1Δ63 his3Δ200 lys2Δ202) (24), using the one-step method (25).
Plasmids pE186, pTRI627 and pTRI628 were also transformed in strain FYEF95 (MATa ura3Δ851 trp1Δ63 leu2Δ1 nup145Δ nup145N-LEU2) (26), using the same protocol. pTRI644 and pTRI645 have been respectively constructed by cloning an EcoRI±BglII fragment from pTRI627 or pTRI628 into the pTRI643 plasmid containing the ARG2 yeast gene flanked by two NotI sites to allow linearization before transformation of the yeast. Plasmids pTRI644 and pTRI645 were transformed in strain FYBL1-22B, after NotI linearization, to give strains GFR108 and GFR109, respectively, using a protocol derived from Gietz et al. (27).

Total RNA extraction and northern blotting
The procedure used is as described by Richard et al. (28). RNA samples are denatured in 80% formamide for 5 min at 95°C before loading on a denaturing agarose gel containing 0.6% formaldehyde. Probes were randomly labelled using [α-32P]dATP for the URA3 probe and [α-32P]dCTP for the CAG/CTG probe. All quantifications were performed using a Phosphorimager (Molecular Dynamics) and the ImageQuant software 5.1.

Preparation of polyadenylated RNAs
Oligo(dT)-cellulose type 7 (Pharmacia) was washed in 0.1 M NaOH for 30 min and equilibrated in buffer ‘H’ (0.5 M NaCl, 10 mM Tris pH 7.5, 2 mM EDTA, 0.1% SDS). An aliquot of 500 μg of total RNAs was denatured for 10 min at 65°C, resuspended in 300 μl of buffer ‘H’ and incubated in the presence of 125 mg of oligo(dT)-cellulose for 30 min at room temperature with gentle shaking. Three washes (5 min each at room temperature) were performed in buffer ‘M’ (0.15 M NaCl, 10 mM Tris pH 7.5, 2 mM EDTA and 0.1% SDS). Poly(A)+ RNAs were eluted in buffer ‘L’ (2 mM EDTA and 0.1% SDS) at 65°C for 5 min. An aliquot of 1 μg of poly(A)+ RNAs was loaded on gels for subsequent northern blot.

RNase and DNase treatments
For RNase treatment, 5 μg of total RNAs was incubated in 40 mM Tris–HCl, pH 7.5, 6 mM MgCl2, with 0.25 μg of DNase free RNase A (Boehringer) in a final volume of 6 μl, for 20 min at 37°C. Alternatively, 5 μg of total RNAs was incubated in 800 mM NaOH (6 μl final volume) for 20 min at 65°C, for alkaline hydrolysis. Neutralization was achieved by adding 5 μl of 1 M Tris–HCl, pH 7.5, and 3 μl of 2 N HCl. DNase treatment was performed on 5 μg of total RNAs incubated in 40 mM Tris–HCl, pH 7.5, 6 mM MgCl2, with 5 U FPLC-pure DNase I (Pharmacia) and 5 U RNase inhibitor (Promega) in a final volume of 6.5 μl for 20 min at 37°C.

Fluorescence in situ hybridization (FISH)
CUG and CAG mRNAs were detected as described by Jensen et al. (29). Each cell sample was hybridized with 0.5 pmol Cy3-labelled CAG or CTG 21mer primers. Nuclear DNA was labelled with 0.1 μg/ml DAPI. A Cy3-dT probe (30mer) was used for ratio normalization of nuclear signals. CAG, CTG and oligo-dT probe signals were quantified (ImageQuant 5.1). Nuclear signals detected with each probe were quantified by using DAPI staining. Ratios of nuclear signal/total signal were not significantly different between the probes for each strain.

RESULTS
CAG and CTG trinucleotide repeats generate long RNAs
We designed an experimental system in yeast, in which various repeats were cloned in the 3’-UTR of the URA3 gene.
(Fig. 1). Several episomal constructs containing CAG_{98}, CTG_{98}, CTG_{255} and TTG_{87}, as well as 300 or 700 bp pieces of λ DNA as controls, were transformed into yeast. Total RNAs from transformant strains were examined by northern blot (Fig. 2A). For each strain, a URA3 transcript of the expected size was detected. Surprisingly, smears extending several kilobases above the expected transcript sizes were also detected in the three strains containing CAG or CUG repeats (Fig. 2A). All RNA species detected in total RNAs were also detected in a poly(A)^+ fraction (Fig. 2B). It was found that
both the mature size RNAs and the high molecular weight smears hybridized with a CAG/CTG probe indicating that the corresponding RNA molecules contain the CAG or the CUG repeats (Fig. 2C). Hybridization with a DNA probe covering the region immediately after the transcription termination site showed no detectable signal (data not shown), indicating that CAG- or CUG-containing smears do not extend past the normal termination site. Bands corresponding to normal size mRNAs and to the smears were not detected after treatment with RNase or alkaline hydrolysis but were still visible after treatment with DNase I, confirming that the smears were not due to contaminating DNA (see Materials and Methods). RNA molecules detected as high molecular weight smears were not transcribed from expanded CAG/CTG DNA repeat tracts since no expansion of CAG/CTG repeats was detected in the different strains analyzed by Southern blot (Supplementary Material). Smears visible below the main band at the expected size of 1.4 kb are either due to degradation of the mRNAs (as is commonly observed with abundant transcripts) (30) or, alternatively, to repeat tract contractions at the DNA level (Supplementary Material). From this series of experiments, we concluded that the long RNAs detected as smears correspond to poly(A)⁺ RNAs, transcribed from a normal size DNA template, containing CAG or CUG repeats. They will hereafter be referred to as L-RNAs.

The amount of L-RNAs was quantified (Fig. 2D). For strains containing CAG₉₈ or CUG₉₈ repeats, 40–50% of the total URA₃ mRNAs were detected as L-RNAs and up to 69% for the strain containing a CUG₂₅₅ repeat (Fig. 3A).

To determine if the nature of the repeated sequence played a role in the process generating L-RNAs, we expressed a TTG₈₇ repeat at the same position (Fig. 1A). Contrary to CAG and CTG repeats, TTGₙ repeats (or their complementary CAAₙ repeats) do not form stable secondary structures in vitro. In strains containing the URA₃-TTG₈₇ construct the level of L-RNA was reduced to 19%, significantly lower than the levels observed with the URA₃-CAG₉₈ or URA₃-CTG₉₈ constructs (Figs 2E and 3A). We thus concluded that the nature of the repeated sequence influences L-RNA level and that, most probably, in vivo stable DNA secondary structures promote L-RNA formation.

To determine whether expression from a chromosomal location has an influence on L-RNA formation compared with a multicopy plasmid, the URA₃ gene containing CAG₉₈ or CTG₉₈ repeats in its 3′-UTR was integrated in yeast chromosome X at the ARG2 locus (Fig. 1B). The total amount of URA₃ transcripts detected with this construct was lower than with the episomal constructs, yet a striking reduction in the amount of L-RNA was observed when URA₃-CUG₉₈ mRNAs were produced (strain GFR109) compared with the URA₃-CAG₉₈ mRNAs (strain GFR108) (Fig. 4). The difference in relative abundance of L-RNAs between the two

![Figure 3.](image)

Figure 3. (A) Relative amount of L-RNA molecules compared with the total amount of URA₃ transcripts detected in each strain with the URA₃ probe. Quantification was performed as shown in Figure 2D. Ratios are calculated as [light grey area/total (light + dark) grey areas] (in %). (B) Ratios of the L-RNA signals detected with a CAG/CTG probe over the same signals detected with a URA₃ probe. The expected size URA₃ mRNA was used as an internal standard and the ratio of the URA₃ signal detected with the CAG/CTG probe over the same signal detected with the URA₃ probe was arbitrarily set to 1 (shown in white: URA₃). Ratios for the CAG and the CUG constructs are shown in light and dark grey, respectively. Hybridization kinetics on a membrane, in the presence of an excess of probe (approximately 10⁶ RNA molecules on the membrane versus 4.5 × 10⁴ probe molecules) depends on probe concentration (40). Since the same probe concentrations were used, hybridization kinetics are similar and the resulting signals are only proportional to the number of RNA molecules on the membrane. Every CAG/CTG molecule also carries URA₃ but the ratio of CAG or CTG over URA₃ is higher than 1. This means that there are more CAG/CTG probe molecules bound to the RNAs than URA₃ probe molecules, therefore that the CAG/CTG part is longer than the URA₃ part. Note that the higher ratio with CAG₉₈ suggests that CAG-containing L-RNAs are longer than those produced in CUG-containing strains. The moderate increase in the ratio of CUG₂₅₅ compared with CUG₉₈ is due to a higher signal of the mature size CUG₂₅₅ RNA with the CAG/CTG probe (compare the signal of the mature size URA₃ RNA in the URA₃-CUG₂₅₅ lane in Fig. 2C and D).
chromosomal constructs is more pronounced than with the episomal constructs. We concluded that orientation of the repeated sequence compared with the direction of transcription was also important to generate L-RNAs.

Since CAG/CTG repeats have a propensity to form stable secondary structures in vitro (31–34), and CUG repeats can also form similar stable secondary structures (35), we could not exclude that L-RNAs are structured RNAs that migrate aberrantly in agarose gels. However, strain GFR108 (URA3-CAG98) exhibits a much higher proportion of L-RNAs than strain GFR109 (URA3-CUG98). Since hairpins formed by CUG repeats are more stable than hairpins formed by CAG repeats (35), we would expect the opposite result if L-RNAs were indeed secondary structures. In addition, we found that the L-RNA signal detected with a CAG/CTG probe is higher than the same signal detected with a URA3 probe (Fig. 3B), of comparable specific activity (CAG/CTG probe, 9.2 × 10⁷ c.p.m./µg; URA3 probe, 7.7 × 10⁸ c.p.m./µg). This difference is more important between the CAG and the URA3 signals than between the CUG and the URA3 signals, suggesting that CAG-containing molecules are, indeed, longer molecules than CUG-containing molecules. This is visible in Figure 2A, B and E (Fig. 2C shows the same gel as Fig. 2B, rehybridized with the CAG/CTG probe), where most of the CAG L-RNA molecules have higher molecular weight than CUG L-RNA molecules.

Measurements have been performed on two independent experiments and standard error calculation shows that the difference between the two ratios is statistically significant (Fig. 3B). In an attempt to clone and sequence these long RNA molecules, we did not manage to reverse transcribe and amplify the full-length cDNA (Supplementary Material).

**CAG and CTG repeat-containing mRNAs are exported but form clusters in yeast cells**

In order to determine if trinucleotide repeat-containing mRNAs were exported to the cytoplasm, we analyzed URA3 mRNA localization by FISH, using Cy3-labelled CAG or CTG oligonucleotides. The probe’s ability to enter the nucleus was assayed in a strain known to impair mRNA nuclear export (Δnup145C) (26). When CUG repeats were expressed in this mutant, a strong nuclear signal was detected with the CAG probe, but not with the CTG probe (Fig. 5A, lower part). Conversely, a nuclear accumulation was seen only with the CTG probe when CUG repeats were expressed (Fig. 5A, upper part). This indicates that CNG probes can enter the nucleus, are strand specific and do not hybridize with DNA molecules containing trinucleotide repeats.

We hybridized the FYBL1-22B strain transformed either with pE186, pTRI627 or pTRI628 with labelled CAG or CTG oligonucleotides. In the control strain (FYBL1-22B/pE186), a faint staining of the entire cells was visible with both probes, corresponding to the background level (Fig. 5B, none). In the FYBL1-22B/pTRI627 strain, as expected, a signal was visible only when probed with the CTG oligonucleotide, whereas in the FYBL1-22B/pTRI628, a signal was detected only with the CAG probe (Fig. 5B, CAG98 and CUG98). In all labelled cells, trinucleotide repeat-containing mRNAs were not evenly distributed over the cell but formed aggregates. This uneven distribution has no effect on cellular mRNAs that do not contain such repeats since the same strains, labelled with an oligo-dT that hybridizes with the poly(A)-tail of mRNAs, do not show these clusters (Fig. 5C). The Δnup145C mutant strain is shown as a positive control to visualize accumulation of mRNAs in the nucleus and to estimate cytoplasmic background. We also tried to follow URA3 mRNAs with the chromosomal integrated constructs (GFR108 and GFR109) but the signals were too low to be detected. To detect nuclear accumulation of transcripts, nuclear signals in both FYBL1-22B/pTRI627 and FYBL1-22B/pTRI628 strains were quantified and compared with signals obtained in a strain labelled with the oligo-dT probe (see Materials and Methods). Intensity of the nuclear staining is not significantly increased in such strains compared with control cells, suggesting that trinucleotide repeat-containing mRNAs are not specifically accumulated in the nuclear compartment. We concluded that although both CAG and CUG transcripts form visible intracellular clusters, there is no obvious nuclear export defect.

**DISCUSSION**

In the present work, we present evidence that trinucleotide repeat expansions occur in vivo, in yeast, at the RNA level when CAG or CTG trinucleotide repeats are cloned in the 3′-UTR of a Pol II transcribed gene. Molecules up to 3 kb longer than expected and terminated at the normal transcription terminator site are observed. The amount of L-RNA detected
is higher when CAG repeats are transcribed, showing sequence dependence. We propose that these large RNA molecules are formed by transcription slippage, a phenomenon reported to be usually stimulated by short mononucleotide or dinucleotide repeats, both in vitro (14–16,36), and in vivo in E. coli (17,18). Alzheimer and Down patients exhibit transcriptional slippage of the β amyloid precursor mRNA or of the ubiquitin-B mRNA, leading to deletions of a mono- or di-nucleotide repeat within or near a repeat stretch (19,20). It is generally assumed that during transcription, transient pausing of the RNA polymerase complex promotes backward slippage of the polymerase complex associated with the mRNA being elongated along the DNA template and leads to resynthesis of the same RNA sequence (37). CAG/CTG trinucleotide repeats form secondary structures in vitro (31–34), and it was shown that CTG repeats induce transient pauses in RNA elongation, after the first, second and sixth to ninth triplet repeat unit, in vitro (12). In vivo, CTG repeats stall replication forks in E. coli (38) and impair recombination associated-DNA synthesis in yeast (23). We observed a significant decrease of the L-RNA level with the URA3-TTG87 construct (Figs 2E and 3A), supporting the notion that stable in vivo secondary structures are important for transcription slippage. We also detected a significantly higher level of L-RNAs with the URA3-CTG255 construct (Fig. 3A), again consistent with the role of secondary structures in L-RNA formation. Therefore, we propose that stalling of the RNA Pol II complex by secondary structures formed on the template DNA strand would promote slippage and give rise to longer transcripts, containing reiterations of the repeated sequence. Several rounds of stalling/slippage/synthesis would probably be required to reach the size of the L-RNAs detected (Fig. 2). If transcription does not resume after pausing, the mRNA being elongated may be degraded, leading to a reduction in the amount of mature URA3 mRNA. Indeed, the amount of URA3 mRNA detected in strain GFR108 (URA3-CAG98) was 40% of the amount detected in strain GFR109 (URA3-CUG98) (Fig. 4). However, we cannot rule out the possibility that relative stabilities of the CAG- and CTG-containing transcripts are different.

Interestingly, we detected more L-RNAs in a strain that uses the CTG strand as template during transcription than in a strain using the CAG strand as template (Fig. 4). It is possible that secondary structures formed by CTG repeats in vivo are more stable than structures formed by CAG repeats, thus increasing the chance of polymerase pausing and subsequently of transcription slippage. However, the difference between CAG and CTG repeats was reduced when they were carried by a high copy number plasmid (Fig. 3A), suggesting that increasing the number of copies of the template DNA modifies
the amount of slippage, or alternatively that chromatinnian factors are assembled differently on a plasmid and on a chromosome, possibly leading to different properties of the transcription machinery.

We investigated whether natural trinucleotide repeats in yeast were subject to transcription slippage too. There is no natural perfect trinucleotide repeat sequence of length comparable to CAG\textsubscript{98} in yeast (39). \textit{YPL009c} contains a stretch of 33 imperfect GAA trinucleotide repeats (one of the longest trinucleotide repeats in the yeast genome), encoding either glutamic acid or lysine. A normal size mRNA but no L-RNA was detected on northern blots for this gene (data not shown), which suggests that degenerate repeats tend to reduce slippage, or that there is a length threshold below which slippage cannot occur, or that GAA repeats are less prone to slippage. These three hypotheses are not mutually exclusive.

Transcription slippage at trinucleotide repeats might also occur in mammalian systems. Analysis of CTG-containing transcripts in normal and myotonic myoblasts showed large size variation of the mutant \textit{DMPK} mRNA (8). This was attributed to repeat size heterogeneity at the DNA level, among the cultured fibroblasts. However, it is possible that some of the different mRNA molecules observed resulted from transcription slippage.

In previous studies where authors looked at the cellular distribution of long trinucleotide repeat-containing transcripts, it was shown that although the long transcripts were correctly spliced and polyadenylated they were not correctly exported to the cytoplasm. CUG-containing mRNAs have been reported to accumulate in the nuclei of DM cultured myoblasts (7,8,11) and in the nuclei of muscle cells of transgenic mice expressing untranslated CUG repeats (9). More recently, it was shown that myotonic dystrophy 2, caused by the expansion of a CCTG tetranucleotide repeat in the first intron of the \textit{ZNF9} gene, also exhibited nuclear retention of transcripts (10). In yeast, we found that L-RNAs co-purify with the poly(A) + gene, also exhibited nuclear retention of transcripts (10). In previous studies where authors looked at the cellular distribution of long trinucleotide repeat-containing transcripts, it was shown that although the long transcripts were correctly spliced and polyadenylated they were not correctly exported to the cytoplasm. CUG-containing mRNAs have been reported to accumulate in the nuclei of DM cultured myoblasts (7,8,11) and in the nuclei of muscle cells of transgenic mice expressing untranslated CUG repeats (9). More recently, it was shown that myotonic dystrophy 2, caused by the expansion of a CCTG tetranucleotide repeat in the first intron of the \textit{ZNF9} gene, also exhibited nuclear retention of transcripts (10). In yeast, we found that L-RNAs co-purify with the poly(A)+ fraction, suggesting they are polyadenylated, but CAG- and CUG-containing mRNAs did not show any obvious defect in nuclear export, although they tend to form clusters in yeast cells, apparently not specific to any subcellular compartment. Both CAG- and CUG-containing transcripts aggregate in yeast, whereas in humans only CUG-containing transcripts clusters have been documented so far. What is the nature of these clusters? It has been proposed that interactions between expanded CUG repeats and double stranded RNA binding proteins, like homologs of the \textit{Drosophila muscleblind} protein or the PKR protein (the last one having yeast homologs), could explain the nuclear retention of mutant \textit{DMPK} transcripts (11,35). Alternatively, they could occur when trinucleotide repeat-containing transcripts are expressed due to intermolecular interactions between such RNAs, both hypotheses not being mutually exclusive.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at NAR Online.

**ACKNOWLEDGEMENTS**

We thank J. Boyer for the generous gift of plasmid pCMha191, M. Richetti and G. Fischer for critical reading of the manuscript, and our colleagues for fruitful discussions. We also thank B. Arcangiolli and A. Jacquier for pertinent technical advice and thoughtful comments on the present work. B.D. is a member of the Institut Universitaire de France.

**REFERENCES**


