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

Emmanuelle Fabre*, Bernard Dujon and Guy-Franck Richard

Unité de Génétique 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

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 repeats, 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 *Pst*I piece of λ DNA into a *Pst*I site located in the *URA3* 3′-UTR of pE186 (Fig. 1A). pTRI626 was constructed by cloning a 700 bp *Pst*I piece of λ DNA into the same site. pTRI627 was constructed by cloning a 300 bp *Bam*HI–*Xba*I DNA fragment containing a $(CAG)_{98}$ trinucleotide repeat from pTRI620 (J.Boyer and B.Dujon, unpublished data) between the *Bgl*II and *Nhe*I sites located in the *URA3* 3′-UTR of pE186. pTRI628 was constructed by cloning a 300 bp *Bam*HI–*Pst*I DNA fragment containing a $(CTG)_{98}$ trinucleotide repeat from pRW3216 (22) between the *Bgl*II and *Pst*I 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 *Pst*I and *Nhe*I sites located in the *URA3* 3′-UTR of pE186. Transformations of plasmids pE186 and pTRI625–637 were performed in the FYBL1-22B strain (*MAT*a *ura3*Δ851 *trp1*Δ63 *his3*Δ200 *lys2*Δ202) (24), using the one-step method (25).

*To whom correspondence should be addressed. Tel: +33 1 40 61 34 54; Fax: +33 1 40 61 34 56; Email: efabre@pasteur.fr
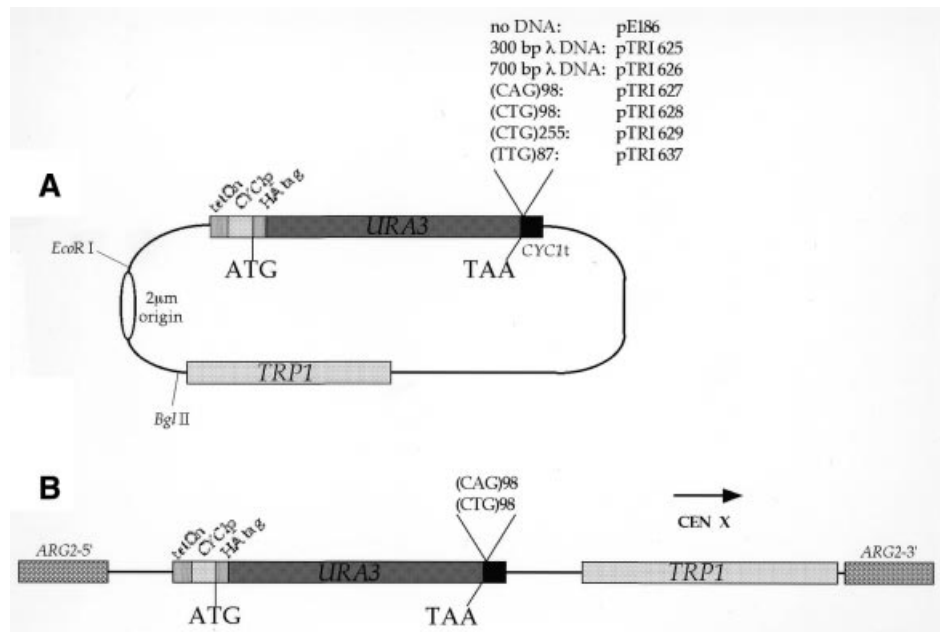
Figure 1. Experimental systems used in the present work. (**A**) λ or trinucleotide repeat DNA was cloned in the 3′-UTR of the *URA3* gene under the control of a regulatable tetOn/*CYC1* promoter. Trinucleotide repeats or control sequences were cloned between the stop codon and the *CYC1* transcription terminator. (**B**) pTRI627 and pTRI628 plasmids shown in (A), deleted for their replication origin, were integrated at the *ARG2* locus (dotted box).

Plasmids pE186, pTRI627 and pTRI628 were also transformed in strain FYEF95 (*MAT*a *ura3*Δ851 *trp1*Δ63 *leu2*Δ1 *nup145*Δ *nup145N-LEU2*) (26), using the same protocol. pTRI644 and pTRI645 have been respectively constructed by cloning an *Eco*RI–*Bgl*II fragment from pTRI627 or pTRI628 into the pTRI643 plasmid containing the *ARG2* yeast gene flanked by two *Not*I sites to allow linearization before transformation of the yeast. Plasmids pTRI644 and pTRI645 were transformed in strain FYBL1-22B, after *Not*I 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-^{32}P]dATP for the *URA3* probe and [α-^{32}P]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)$^+$ mRNAs 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 MgCl$_2$, 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 MgCl$_2$, 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
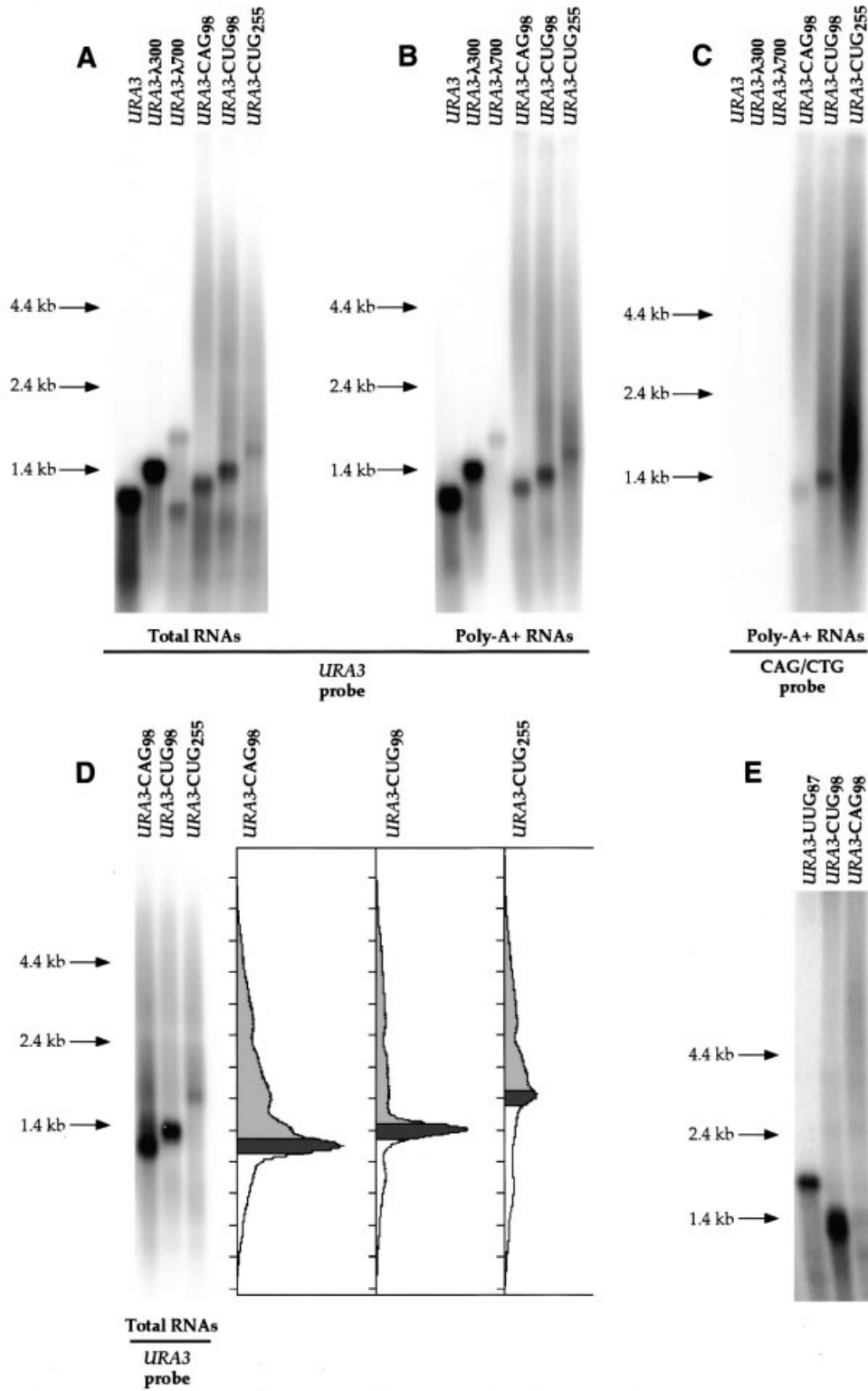
both the mature size RNAs and the high molecular weight smears hybridized with a CAG/CTG probe indicating that the corresponding RNA molecules contain the CAG or the CUG repeats (Fig. 2C). Hybridization with a DNA probe covering the region immediatly after the transcription termination site showed no 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 refered to as L-RNAs.

The amount of L-RNAs was quantified (Fig. 2D). For strains containing CAG$_{98}$ or CUG$_{98}$ repeats, 40–50% of the total *URA3* mRNAs were detected as L-RNAs and up to 69% for the strain containing a CUG$_{255}$ repeat (Fig. 3A).

To determine if the nature of the repeated sequence played a role in the process generating L-RNAs, we expressed a TTG$_{87}$ repeat at the same position (Fig. 1A). Contrary to CAG and CTG repeats, TTG$_n$ repeats (or their complementary CAA$_n$ repeats) do not form stable secondary structures *in vitro*. In strains containing the *URA3*-TTG$_{87}$ construct the level of L-RNA was reduced to 19%, significantly lower than the levels observed with the *URA3*-CAG$_{98}$ or *URA3*-CTG$_{98}$ 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 CAG$_{98}$ or CTG$_{98}$ 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*-CUG$_{98}$
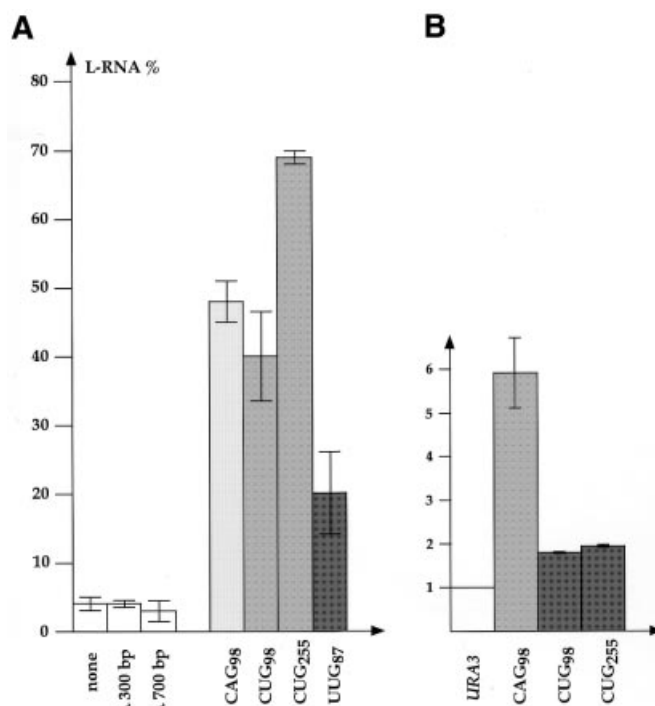
Figure 3. (**A**) Relative amount of L-RNA molecules compared with the total amount of *URA3* transcripts detected in each strain with the *URA3* probe. Quantification was performed as shown in Figure 2D. Ratios are calculated as [light grey area/total (light + dark) grey areas] (in %). (**B**) Ratios 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 10^8 RNA molecules on the membrane versus 4.5×10^{11} 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 CAG$_{98}$ suggests that CAG-containing L-RNAs are longer than those produced in CUG-containing strains. The moderate increase in the ratio of CUG$_{255}$ compared with CUG$_{98}$ is due to a higher signal of the mature size CUG$_{255}$ RNA with the CAG/CTG probe (compare the signal of the mature size *URA3* RNA in the *URA3*-CUG$_{255}$ lane in Fig. 2C and D).

mRNAs were produced (strain GFR109) compared with the *URA3*-CAG$_{98}$ mRNAs (strain GFR108) (Fig. 4). The difference in relative abundance of L-RNAs between the two

Figure 2. (Opposite) Northern blots showing mature *URA3* mRNAs and L-RNAs. (**A**) Total RNAs of yeast strains containing constructs shown in Figure 1A, hybridized with a *URA3* probe. RNA size markers are indicated on the left. The CUG$_{98}$ transcript is longer than the CAG$_{98}$ transcript, due to an extra piece of pBluescript polylinker. In the FYBL1-2B/pTRI626 control strain, the lower signal possibly corresponds to an alternative site of termination or processing located in the 700 bp of λ DNA. L-RNAs are visible as smears extending from the mature *URA3* transcript (~1.4 kb) up to >4.4 kb. (**B**) Poly(A)$^+$ mRNAs purified from the same strains and hybridized with the same probe. The lower band in FYBL1-2B/pTRI626 strain is not detected in the poly(A)$^+$ fraction. (**C**) The same blot as in (B), stripped and rehybridized with the CAG/CTG probe. (**D**) Quantification of radioactivity for total RNAs hybridized with a *URA3* probe. The dark grey area corresponds to *URA3* mRNA molecules of the expected size. The light grey area corresponds to the L-RNAs. Smears lower than expected sizes correspond to RNA degradation. Local decreases of the signal correspond to competition by the large amount of rRNA molecules. (**E**) Northern blot comparing RNA samples extracted from strains containing either the *URA3*-UUG$_{87}$ episomal construct, the *URA3*-CAG$_{98}$ chromosomal construct or the *URA3*-CUG$_{98}$ chromosomal construct. The UUG$_{87}$ transcript is longer than the CUG$_{98}$ transcript, due to an extra 150 bp piece of DNA coming from the original vector containing the TTG$_{87}$ repeat tract. In this particular experiment, amounts of L-RNA were, respectively, 16% (UUG$_{87}$), 18% (CUG$_{98}$) and 85% (CAG$_{98}$).
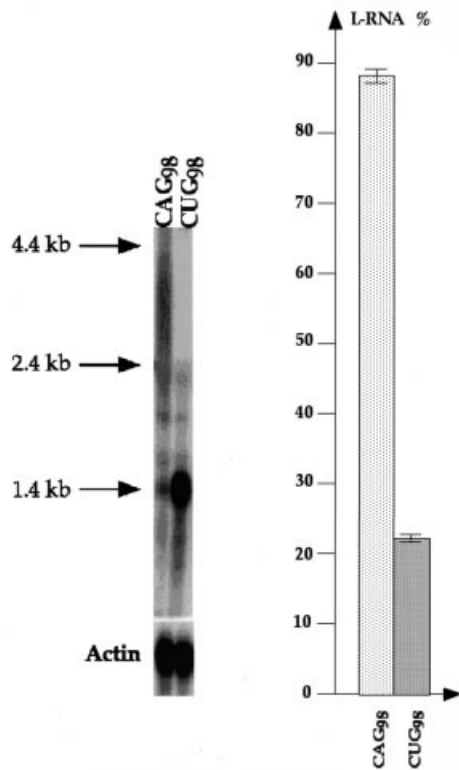
Figure 4. Northern blot of total RNAs extracted from strains GFR108 and GFR109, expressing the chromosomal constructs *URA3*-CAG$_{98}$ or *URA3*-CUG$_{98}$, respectively, hybridized with either a *URA3* or an *ACT1* probe. The actin signal was used to compare the total amount of mRNAs in both lanes (see text). Quantifications of L-RNAs are performed as described in Figure 3A. Discrete bands correspond to competition by the large amount of rRNA molecules, leading to a local decrease of the signal.

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*-CAG$_{98}$) exhibits a much higher proportion of L-RNAs than strain GFR109 (*URA3*-CUG$_{98}$). 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, 9.2×10^7 c.p.m./µg; *URA3* probe, 7.7×10^8 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 (Δ*nup145*C) (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 CAG 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, CAG$_{98}$ and CUG$_{98}$). 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 Δ*nup145*C 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
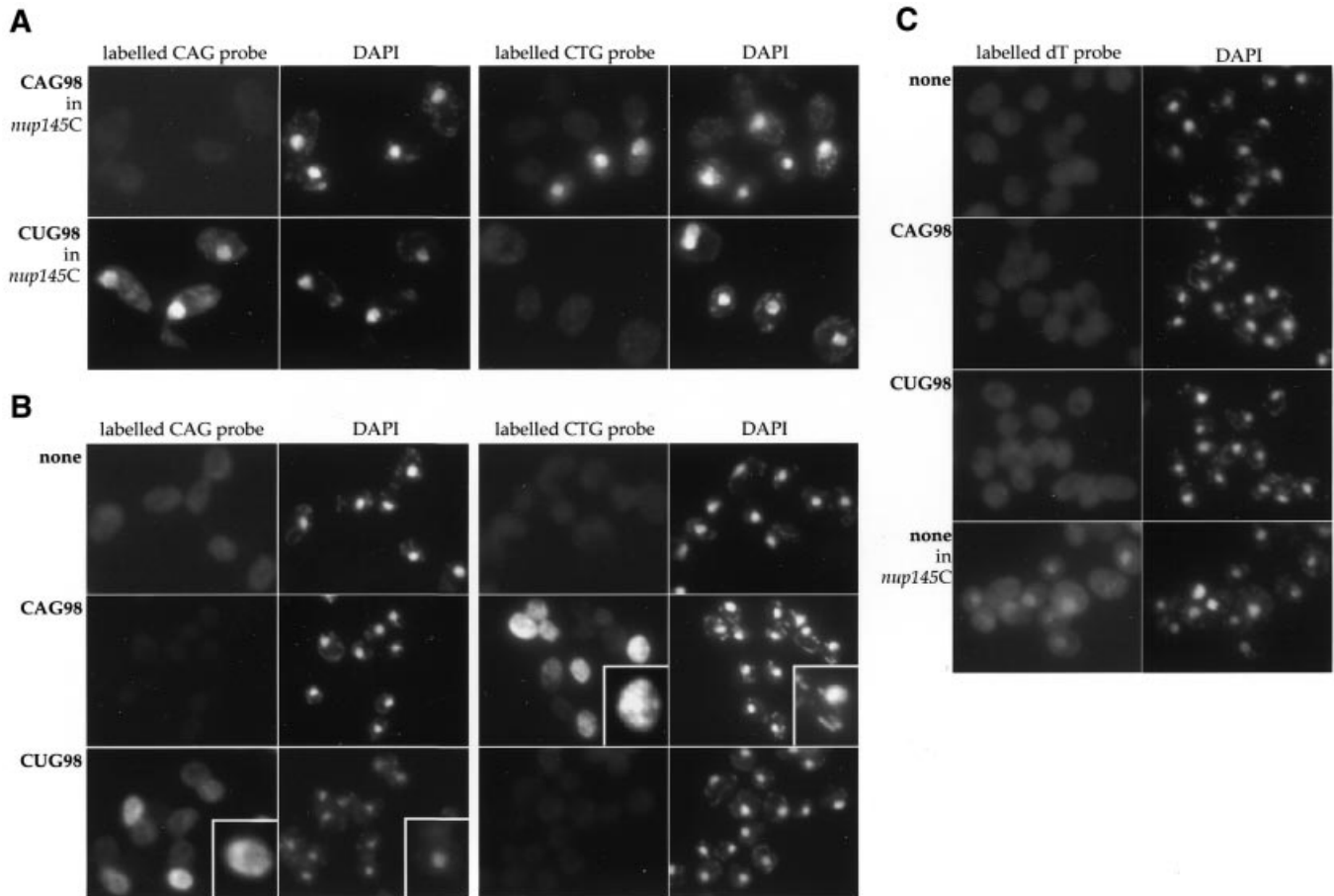
Figure 5. (**A**) FISH labelling in a *nup145CΔ* mutant expressing either *URA3*-CAG$_{98}$ (top) or *URA3*-CUG$_{98}$ (bottom). mRNAs were detected with CAG or CTG labelled oligonucleotide probes, and nuclei were stained with DAPI. (**B**) FISH labelling in strains expressing *URA3* (none), *URA3*-CAG$_{98}$ or *URA3*-CUG$_{98}$. Insets at bottom right of panels show parts of the figure at higher magnification. (**C**) Labelling with an oligo-dT probe.

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*-TTG$_{87}$ 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*-CTG$_{255}$ 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*-CAG$_{98}$) was 40% of the amount detected in strain GFR109 (*URA3*-CUG$_{98}$) (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 chromatinian 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$_{98}$ in yeast (39). *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 *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 *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 *Drosophila muscleblind* protein or the PKR protein (the last one having yeast homologs), could explain the nuclear retention of mutant *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

REFERENCES

1. Cummings,C.J. and Zoghbi,H.Y. (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.*, **9**, 909–916.
2. Alba,M.M., Santibanez-Koref,M.F. and Hancock,J.M. (1999) Amino acid reiterations in yeast are overrepresented in particular classes of proteins and show evidence of a slippage-like mutational process. *J. Mol. Evol.*, **49**, 789–797.
3. Brook,J.D., McCurrach,M.E., Harley,H.G., Buckler,A.J., Church,D., Aburatani,H., Hunter,K., Stanton,V.P., Thirion,J.P., Hudson,T. *et al.* (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell*, **68**, 799–808.
4. Harley,H.G., Brook,J.D., Rundle,S.A., Crow,S., Reardon,W., Buckler,A.J., Harper,P.S., Housman,D.E. and Shaw,D.J. (1992) Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature*, **355**, 545–546.
5. Koob,M.D., Moseley,M.L., Schut,L.J., Benzow,K.A., Bird,T.D., Day,J.W. and Ranum,L.P.W. (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nature Genet.*, **21**, 379–384.
6. Krahe,R., Ashizawa,T., Abbruzzese,C., Roeder,E., Carango,P., Giacanelli,M., Funanage,V.L. and Siciliano,M.J. (1995) Effect of myotonic dystrophy trinucleotide repeat expansion on *DMPK* transcription and processing. *Genomics*, **28**, 1–14.
7. Taneja,K.L., McCurrach,M., Schalling,M., Housman,D. and Singer,R.H. (1995) Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.*, **128**, 995–1002.
8. Davis,B.M., McCurrach,M.E., Taneja,K.L., Singer,R.H. and Housman,D.E. (1997) Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl Acad. Sci. USA*, **94**, 7388–7393.
9. Mankodi,A., Logigian,E., Callahan,L., McClain,C., White,R., Henderson,D., Krym,M. and Thornton,C.A. (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science*, **289**, 1769–1772.
10. Liquori,C.L., Ricker,K., Moseley,M.L., Jacobsen,J.F., Kress,W., Naylor,S.L., Day,J.W. and Ranum,L.P.W. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of *ZNF9*. *Science*, **293**, 864–867.
11. Miller,J.W., Urbinati,C.R., Teng-umnuay,P., Stenberg,M.G., Byrne,B.J., Thornton,C.A. and Swanson,M.S. (2000) Recruitment of human muscleblind proteins to (CUG)n expansions associated with myotonic dystrophy. *EMBO J.*, **19**, 4439–4448.
12. Parsons,M.A., Sinden,R.R. and Izban,M.G. (1998) Transcriptional properties of RNA polymerase II within triplet repeat-containing DNA from the human myotonic dystrophy and fragile X loci. *J. Biol. Chem.*, **273**, 26998–27008.
13. McDonald,J.F. (1993) Evolution and consequences of transposable elements. *Curr. Opin. Genet. Dev.*, **3**, 855–864.
14. Severinov,K. and Goldfarb,A. (1994) Topology of the product binding site in RNA polymerase revealed by transcript slippage at the phage λ P$_L$ promoter. *J. Biol. Chem.*, **269**, 31701–31705.
15. Jeong,S.W., Lang,W.H. and Reeder,R.H. (1996) The yeast transcription terminator for RNA polymerase I is designed to prevent polymerase slippage. *J. Biol. Chem.*, **271**, 16104–16110.
16. Deng,L. and Shuman,S. (1997) Elongation properties of vaccinia virus RNA polymerase: pausing, slippage, 3′ end addition, and termination site choice. *Biochemistry*, **36**, 15892–15899.
17. Wagner,L.A., Weiss,R.B., Driscoll,R., Dunn,D.S. and Gesteland,R.F. (1990) Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*. *Nucleic Acids Res.*, **18**, 3529–3535.
18. Xiong,X.F. and Reznikoff,W.S. (1993) Transcriptional slippage during the transcription initiation process at a mutant *lac* promoter *in vivo*. *J. Mol. Biol.*, **231**, 569–580.

19. van Leeuwen,F.W., de Kleijn,D.P.V., van den Hurk,H.H., Neubauer,A., Sonnemans,M.A.F., Sluijs,J.A., Köycü,S., Ramdjielal,R.D.J., Salehi,A., Martens,G.J.M. *et al.* (1998) Frameshift mutants of β amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. *Science*, **279**, 242–247.

20. van den Hurk,W.H., Willems,H.J.J., Bloemen,M. and Martens,G.J.M. (2001) Novel frameshift mutations near short simple repeats. *J. Biol. Chem.*, **276**, 11496–11498.

21. Gari,E., Piedrafita,L., Aldea,M. and Herrero,E. (1997) A set of vectors with a tetracycline-regulatable promoter system for modulating gene expression in *Saccharomyces cerevisiae*. *Yeast*, **13**, 837–848.

22. Kang,S., Jaworski,A., Ohshima,K. and Wells,R.D. (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nature Genet.*, **10**, 213–217.

23. Richard,G.-F., Goellner,G.M., McMurray,C.T. and Haber,J.E. (2000) Recombination-induced CAG trinucleotide repeat expansions in yeast involve the *MRE11/RAD50/XRS2* complex. *EMBO J.*, **19**, 2381–2390.

24. Fairhead,C., Llorente,B., Denis,F., Soler,M. and Dujon,B. (1996) New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using 'split-marker' recombination. *Yeast*, **12**, 1439–1457.

25. Chen,D.-C., Yang,B.-C. and Kuo,T.-T. (1992) One-step transformation of yeast in stationary phase. *Curr. Genet.*, **21**, 83–84.

26. Teixeira,M.T., Siniossoglou,S., Podtelejnikov,S., Benichou,J.-C., Mann,M., Dujon,B., Hurt,E. and Fabre,E. (1997) Two functionally distinct domains generated by *in vivo* cleavage of Nup145p: a novel biogenesis pathway for nucleoporins. *EMBO J.*, **16**, 5086–5097.

27. Gietz,R.D., Schiestl,R.H., Willems,A.R. and Woods,R.A. (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, **11**, 355–360.

28. Richard,G.-F., Fairhead,C. and Dujon,B. (1997) Complete transcriptional map of yeast chromosome XI in different life conditions. *J. Mol. Biol.*, **268**, 303–321.

29. Jensen,T.H., Patricio,K., McCarthy,T. and Rosbash,M. (2001) A block to mRNA nuclear export in *S. cerevisiae* leads to hyperadenylation of transcripts that accumulate at the site of transcription. *Mol. Cell*, **7**, 887–898.

30. Fairhead,C. and Dujon,B. (1994) Transcript map of two regions from chromosome XI of *Saccharomyces cerevisiae* for interpretation of systematic sequencing results. *Yeast*, **10**, 1403–1413.

31. Gacy,A.M., Goellner,G., Juranic,N., Macura,S. and McMurray,C.T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell*, **81**, 533–540.

32. Yu,A. and Mitas,M. (1995) The purine-rich trinucleotide repeat sequences d(CAG)15 and d(GAC)15 form hairpins. *Nucleic Acids Res.*, **23**, 4055–4057.

33. Mitas,M., Yu,A., Dill,J., Kamp,T.J., Chambers,E.J. and Haworth,I.S. (1995) Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)15. *Nucleic Acids Res.*, **23**, 1050–1059.

34. Mitas,M. (1997) Trinucleotide repeats associated with human diseases. *Nucleic Acids Res.*, **25**, 2245–2253.

35. Tian,B., White,R.J., Xia,T., Welle,S., Turner,D.H., Mathews,M.B. and Thornton,C.A. (2000) Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA*, **6**, 79–87.

36. Macdonald,L.E., Zhou,Y. and McAllister,W.T. (1993) Termination and slippage by bacteriophage T7 RNA polymerase. *J. Mol. Biol.*, **232**, 1030–1047.

37. Jacques,J.-P. and Kolakofsky,D. (1991) Pseudo-templated transcription in prokaryotic and eucaryotic organisms. *Genes Dev.*, **5**, 707–713.

38. Samadashwily,G., Raca,G. and Mirkin,S.M. (1997) Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.*, **17**, 298–304.

39. Richard,G.-F. and Dujon,B. (1997) Trinucleotide repeats in yeast. *Res. Microbiol.*, **148**, 731–744.

40. Kumazawa,Y., Yokogawa,T., Tsurui,H., Miura,K.-I. and Watanabe,K. (1992) Effect of the higher-order structure of tRNAs on the stability of hybrids with oligodeoxyribonucleotides: separation of tRNA by an efficient solution hybridization *Nucleic Acids Res.*, **20**, 2223–2232.