Positioned to expand

Sergei M. Mirkin & Ekaterina V. Smirnova

Department of Molecular Genetics, University of Illinois at Chicago, Chicago, Illinois 60607, USA. e-mail: mirkin@uic.edu

The position of trinucleotide repeats relative to adjacent origins of DNA replication can drastically affect repeat expansion. This finding offers insight into the cause of several hereditary diseases in humans.

Expansion of trinucleotide repeats (TNRs) is responsible for at least 15 hereditary neurological disorders in humans[1]. They vary in their sequences, are situated in a number of genes and affect gene expression in different ways. Still, there is one underlying similarity: TNRs are stably inherited until the number of elementary units in a repeat exceeds approximately 25. If this threshold is exceeded, TNRs expand during intergenerational transmissions. In a few generations, this results in the addition of as many as several thousand repeats, disruption of gene expression and, ultimately, the onset of disease. Although such large, rapid expansions might suggest that some major flaw in DNA metabolism is involved, familial DNA analyses show that only a single repeat is amplified in every case. Thus, the aberrant DNA metabolism in affected families is triggered by factors positioned in *cis* to the repeat, rather than by mutations in the genes encoding components of the general replication, recombination or repair machinery. On page 37 of this issue, John Cleary and colleagues[2] provide new insight into the nature of the elusive factors governing repeat expansion.

Why do repeats expand?

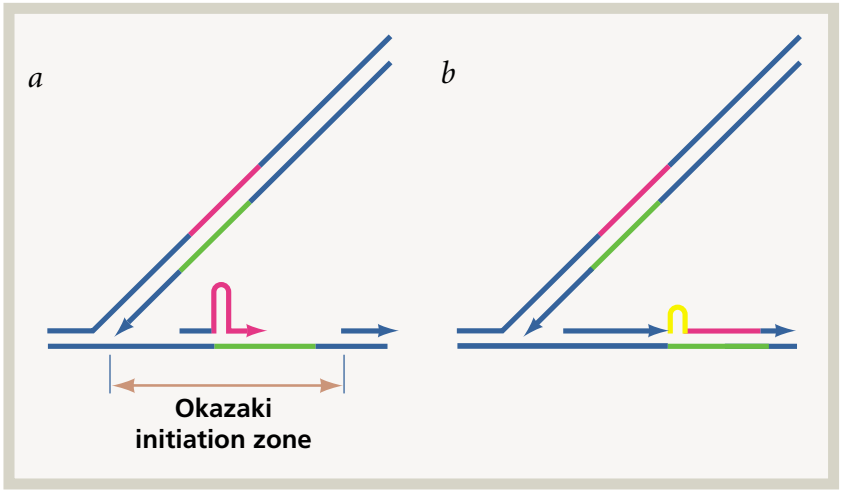
Of all possible TNRs, only three have been shown to expand: $(CAG)_n$•$(CTG)_n$, $(CGG)_n$•$(CCG)_n$ and $(GAA)_n$•$(TTC)_n$. These sequences are distinguished by their ability to form unusual DNA structures: hairpins, G-quartets and triplexes. Formation of these structures by TNRs inhibits the activity of DNA polymerases and other replication proteins *in vitro*[3]. When the TNRs are longer than the threshold lengths, they impede the DNA replication fork progression *in vivo* as well, presumably owing to similar structural problems[4]. It seems, therefore, that irregularities of TNR replication could occasionally result in the addition of extra copies of repeats to newly synthesized DNA strands. In support of this notion, various mutations in genes encoding the replication apparatus of yeast have been shown to increase the rate of TNR expansions[5].

In two commonly discussed models of TNR expansion by replication, it is hypothesized that TNRs expand during lagging strand DNA synthesis (Fig. 1). The first model[6] suggests that replication blockage within a TNR could occasionally lead to partial dissociation of the newly synthesized and template DNA strands and their subsequent misalignment. If the newly synthesized DNA strand forms an unusual structure such as a hairpin, resumption of DNA synthesis leads to the repeat's expansion. Different types of TNR vary in their structural potential; for example, a $(CTG)_n$ strand forms hairpins better than its $(CAG)_n$ partner. Consequently, the frequency of TNR expansions must depend on the direction of replication through these repeats. Consistent with this prediction, $(CAG)_n$•$(CTG)_n$ repeats were shown to expand in model systems when the $(CTG)_n$ run was synthesized as the lagging DNA strand[6].

The second model is based on the observation that flap-endonuclease, which is responsible for removing RNA primers from Okazaki fragments, is inefficient on primers consisting of TNRs[3,7]. Rather then being removed, such a primer could be displaced during the next Okazaki fragment synthesis and then ligated to its 3′ end. This is predicted to result in the addition of extra TNRs, first in the RNA form, and then, after another round of replication, in the DNA form. Indeed, mutations in the yeast flap-nuclease gene (*rad27*) greatly increase the TNR expansion rate[8].

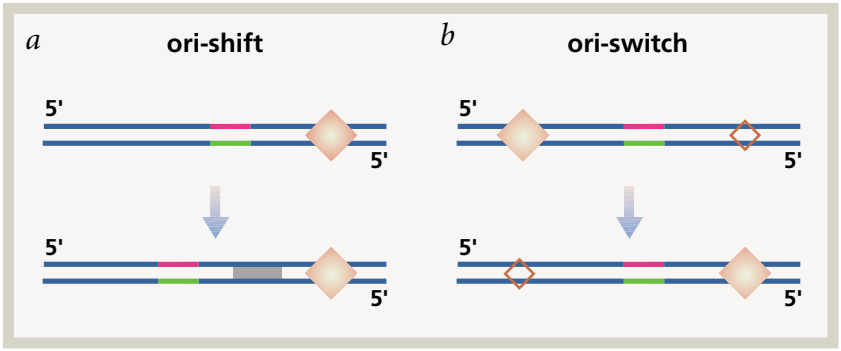
Whereas most data support the replicative model for expansion, alternative mechanisms involving recombination and repair have been proposed[9,10]. However, these alternative mechanisms should also include DNA synthesis through TNRs to account for the large scale of the expansions.

Mechanistic studies of TNR expansion are well underway, but important biological questions remain. What prevents TNRs from expanding in normal individuals and what triggers their expansion in affected families? The first plausible answer to this question came from studies of fragile X syndrome, caused by repeat



BOB CRIMI

Fig. 1 Models of TNR expansion by replication. ***a***, DNA hairpin formation. TNR expansion is caused by formation of a hairpin in the lagging strand and subsequent misalignment of the lagging strand and its template. ***b***, RNA primer displacement. Instead of being removed, RNA primer containing TNR is displaced and ligated to the 3′ end of the next Okazaki fragment, leading to expansion. Blue line, nonrepetitive DNA; red line, structure-prone TNR stand; green line, partner TNR strand; yellow line, TNR-containing RNA primer. Arrows show directionality of DNA synthesis.



BOB CRIMI

Fig. 2 Possible mechanisms triggering TNR-related diseases. ***a***, The 'ori-shift' model. Expansion of a TNR is triggered by an insertion that changes its distance from the ori. ***b***, The 'ori-switch' model. Inactivation of the ori at one side of a TNR and concurrent activation of a cryptic ori on its other side switches directionality of replication and triggers TNR expansion. Line colors: see legend of Fig. 1. Large filled diamond, active replication origin; small empty diamond, cryptic replication origin; gray bar, DNA insertion shifting TNR from the origin.

expansions in *FMR1* (ref. 11), and spinocerebellar ataxia, caused by repeat expansions in *SCA1* (ref. 12). Normal alleles of these disease-related genes seemed to contain TNRs with several interruptions. Those alleles predisposed for expansion lacked some interruptions, and subsequent expansions were limited to the non-interrupted part of the repeat. Thus, interruptions might confer TNR stability, whereas their loss could lead to repeat expansion. It is not clear, however, if this is the first event or if it is caused by earlier genomic alterations. If the loss of interruptions is sufficient for expansion, why are uninterrupted repeats not predisposed to expand in cultured cells? Further, people with fragile X syndrome carry another unstable microsatellite adjacent to the $(CGG)_n$ repeat, suggesting that a *cis*-element outside the expanding repeat is responsible[13].

Location, location, location

The work by Cleary *et al.*[2] helps to explain the nature of the *cis*-factors that are potentially responsible for TNR expansion. In so doing, the authors developed the first experimental system for measuring TNR expansions in mammalian cells. Specifically, they studied the frequency of expansion in $(CAG)_n \bullet (CTG)_n$ tracts depending on their location relative to the origin of DNA replication. The $(CAG)_n \bullet (CTG)_n$ tracts were cloned into a mammalian episomal vector at various distances and in different orientations with respect to the SV40 replication origin. Upon replication in COS cells, episomal DNA was isolated, re-transformed into bacteria and analyzed for changes in repeat length. The TNRs in some of the recombinant constructs seemed to be very unstable, giving rise to expansions and/or deletions at a frequency approaching 10%.

Two previous observations made in bacteria and yeast were confirmed: only repeats of premutational lengths were unstable, and expansions were only observed in one orientation of a repeat relative to the replication origin. Cleary *et al.*[2] suggest that the formation of hairpins within the $(CTG)_n$-containing Okazaki fragment is responsible for the repeat's expansion (Fig. 1).

Surprisingly, however, the very same repeat in the same orientation could be switched from expansion to deletion simply by being moved a mere 130 bp further from the origin. This could change the position of the repeat within the so-called 'Okazaki initiation zone'—a single-stranded portion of approximately 290 nucleotides of the lagging strand template[14]. Cleary *et al.*[2] speculate that TNRs expand when positioned at the 3′ end and contract when at the 5′ end of the Okazaki initiation zone.

Shift or switch

Is this model of any relevance to a human disease? In families with myotonic dystrophy, all chromosomes carrying expanded $(CTG)_n$ repeats also contain an *Alu* element inserted 5 kb from the TNR[15]. Cleary *et al.*[2] suggest that this insertion might change the distance between the TNR and its replication origin, poising it to expand. As such, their hypothesis might be called 'ori-shift' (Fig. 2*a*). What about other TNR-mediated diseases for which no data on such insertions have been reported? An alternative hypothesis could be that the first event triggering expansion is inactivation of the regular replication origin situated on one side of the repeat and concurrent activation of a cryptic origin on its other side. This hypothesis might be called 'ori-switch' (Fig. 2*b*). Switching the directionality of the TNR's replication has been shown to be a crucial factor for repeat expansion in many studies, including Cleary *et. al.*

Taken together, the 'ori-shift' and 'ori-switch' hypotheses could explain the underlying mechanisms of TNR-mediated diseases. They consider a change in relative orientation or distance between a TNR and its replication origin as the primary event that puts a TNR into an expansion-prone position. These models could also explain why TNR expansions preferentially occur during intergenerational transmissions, as different origins might be used for repeat replication in differentiated, as opposed to germline or embryonic cells. Progress in this field could come from the fine-mapping of replication origins adjacent to TNRs under different developmental programs.□

1. Wells, R.D. & Warren, S.T. (eds) *Genetic Instabilities and Hereditary Neurological Disorders* (Academic Press, San Diego, 1998).
2. Cleary, J.D., Nichol, K., Wang, Y.-H. & Pearson, C.E. *Nature Genet.* **31**, 37–46 (2002).
3. McMurray, C.T. *Proc. Natl Acad. Sci. USA* **96**, 1823–1825 (1999).
4. Samadashwily, G.M., Raca, G. & Mirkin, S.M. *Nature Genet.* **17**, 298–304 (1997).
5. Ireland, M.J., Reinke, S.S. & Livingston, D.M. *Genetics* **155**, 1657–1665 (2000).
6. Kang, S., Jaworski, A., Ohshima, K. & Wells, R.D. *Nature Genet.* **10**, 213–218 (1995).
7. Gordenin, D.A., Kunkel, T.A. & Resnick, M.A. *Nature Genet.* **16**, 116–118 (1997).
8. Freudenreich, C.H., Kantrow, S.M. & Zakian, V.A. *Science* **279**, 853–856 (1998).
9. Jakupciak, J.P. & Wells, R.D. *J. Biol. Chem.* **274**, 23468–23479 (1999).
10. Kovtun, I.V. & McMurray, C.T. *Nature Genet.* **27**, 407–411 (2001).
11. Chung, M.Y. *et al. Nature Genet.* **5**, 254–258 (1993).
12. Kunst, C.B. & Warren, S.T. *Cell* **77**, 853–861 (1994).
13. Zhong, N., Dobkin, C. & Brown, W.T. *Nature Genet.* **5**, 248–253 (1993).
14. DePamphilis, M.L. & Wassarman, P.M. *Annu. Rev. Biochem.* **49**, 627–666 (1980).
15. Neville, C.E., Mahadevan, M.S., Barcelo, J.M. & Korneluk, R.G. *Hum. Mol. Genet.* **3**, 45–51 (1994).