Published in final edited form as:

Mol Carcinog. 2009 April ; 48(4): 350–361. doi:10.1002/mc.20488.

TRANSCRIPTION DESTABILIZES TRIPLET REPEATS

Yunfu Lin*, Leroy Hubert Jr., and John H. Wilson*

Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030

Abstract

Triplet repeat expansion is the molecular basis for several human diseases. Intensive studies using systems in bacteria, yeast, flies, mammalian cells, and mice have provided important insights into the molecular processes that are responsible for mediating repeat instability. The age-dependent, ongoing repeat instability in somatic tissues, especially in terminally differentiated neurons, strongly suggests a robust role for pathways that are independent of DNA replication. Several genetic studies have indicated that transcription can play a critical role in repeat instability, potentially providing a basis for the instability observed in neurons. Transcription-induced repeat instability can be modulated by several DNA repair proteins, including those involved in mismatch repair (MMR) and transcription-coupled nucleotide excision repair (TC-NER). Though the mechanism is unclear, it is likely that transcription facilitates the formation of repeat-specific secondary structures, which act as intermediates to trigger DNA repair, eventually leading to changes in the length of the repeat tract. In addition, other processes associated with transcription can also modulate repeat instability, as shown in a variety of different systems. Overall, the mechanisms underlying repeat instability in humans are unexpectedly complicated. Because repeat-disease genes are widely expressed, transcription undoubtedly contributes to the repeat instability observed in many diseases, but it may be especially important in nondividing cells. Transcription-induced instability is likely to involve an extensive interplay not only of the core transcription machinery and DNA repair proteins, but also of proteins involved in chromatin remodeling, regulation of supercoiling, and removal of stalled RNA polymerases, as well as local DNA sequence effects.

Keywords

triplet repeats; transcription-induced instability; abnormal DNA structures; mismatch repair; nucleotide excision repair

Short tandem DNA repeats are unstable in all genomes, but at several loci in the human genome repeat instability is associated with disease [1,2]. Expansion of trinucleotide (triplet) repeats is the molecular cause of at least 18 human neurological diseases (repeat diseases), including myotonic dystrophy (DM1), Huntington's disease (HD), and a number of spinocerebellar ataxias (SCAs) [3]. Repeat diseases are characterized by the expansion of a disease-specific triplet repeat tract beyond a threshold of about 25–35 repeats to a length that has pathologic consequences, often involving neuronal death in disease-specific regions of the brain. The discovery of this unique category of human disease raised three important questions: What are the molecular mechanisms underlying triplet repeat instability? How does expansion of repeat sequences cause human diseases? And can these diseases be

*Correspondence to: Yunfu Lin, John H. Wilson, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, Phone: 713-798-5760, Fax: 713-796-9438, yunful@bcm.edu, jwilson@bcm.tmc.edu.

treated or prevented? Here we examine the molecular mechanisms of repeat instability. In the past 15 years, intensive studies using model systems in bacteria, yeast, flies, mammalian cells, and mice have provided important insights into the molecular processes that destabilize repeats, although the exact mechanisms are far from clear. In this review we focus on the emerging role of transcription as an important modulator of triplet repeat instability. Related reviews in this issue discuss replication-independent models of non-B DNA induced genetic instability in mammalian cells (Wang & Vasquez) and the role of non-B DNA conformations in mutagenesis and human disease (Bacolla & Wells).

TRIPLET REPEAT INSTABILITY

Inheritance of repeat diseases typically shows a progressive worsening of the disease phenotype in subsequent generations, as the repeat tract continues to expand. This phenomenon, which physicians labeled anticipation, indicated that a critical period of expansion-biased repeat instability must occur in the germline or at a very early stage of embryonic development. While repeat expansions have been observed in patients and in mice at both stages, it appears that instability during mitotic divisions in the germline is the largest contributor to intergenerational instability [4]. Repeat instability is not confined to the germline, however, but also occurs in characteristic patterns in many somatic tissues of affected individuals [5]. For example, CAG repeats in HD typically are highly unstable in striatum, moderately unstable in liver and kidney, and marginally unstable in heart and muscle [6]. The ongoing repeat instability in critical somatic tissues likely accelerates disease progression. Thus, tissue-specific repeat instability plays two roles in “repeat disease”: germline instability establishes the initial repeat length in an individual, which is the best predictor of disease onset and severity, and instability in key tissues exacerbates the disease symptoms.

The complexity of these tissue-specific patterns of repeat instability challenges any simple understanding of the underlying molecular mechanisms. There is no straightforward correlation with cell division rates (DNA replication), DNA damage (DNA repair and recombination), or disease-gene transcript levels (transcription) [1,5]. The lack of an easy explanation leaves us with a set of basic—and difficult—questions. Why is a repeat tract biased toward expansion during intergenerational transmission, as well as in somatic tissues? What is the molecular basis for the variation in degree of repeat instability from tissue to tissue? Do these patterns of instability arise via one fundamental mechanism, or do distinct mechanisms drive repeat instability in different tissues? Are the same types of repeat at different locations in the genome destabilized by the same mechanism or by different ones?

Repeat instability has been extensively modeled in bacteria, yeast, flies, mammalian cells, and mice. These studies have identified DNA replication, DNA repair, recombination, and transcription as significant potential contributors to repeat instability in patients [1,5]. Each of these processes exposes single-stranded DNA, which allows triplet repeats to form intrastrand secondary structures such as hairpins and slipped-strand DNA duplexes, among other possibilities [7,8]. The combination of secondary structure and repetitive sequence in some way confounds the cell's ability to accurately convert the aberrant structure back to normal length Watson-Crick DNA. The myriad permutations of DNA metabolic processes that expose single strands, different types of secondary structure that repeat tracts can form, and varieties of ways to resolve such structures in cells guarantee a rich complexity of possible pathways by which repeats might change their length. Moreover, these possibilities do not include additional identified contributors to instability such as epigenetic modifications, chromatin structure, and local sequence effects [1]. Almost certainly, multiple pathways will be found to play a role in repeat instability in human patients.

As a category, transcription-induced pathways gather circumstantial support from two observations on repeat instability in humans. First, even though repeat diseases typically affect only one or a few cell types, usually neurons, most repeat disease-associated genes are ubiquitously expressed in human (and animal) tissues [5]. Widespread expression involving the many tissues that display instability is a prerequisite for any transcription-induced pathway. Thus, transcription has the potential to contribute to repeat instability in multiple tissues, including the germline. Second, transcription provides an ongoing process in terminally differentiated cells such as neurons, which no longer carry out DNA replication, a well-defined initiator of repeat instability that is strongly supported by genetic studies in bacteria, yeast, and mammalian cells. A particularly clear example of age-dependent, expansion-biased repeat instability was observed in the brains of HD transgenic mice [9]. At 3 months of age, only minimal levels of repeat instability were seen in samples from several regions of the brain, including the striatum. Thus, no significant repeat instability had arisen during the many cell divisions that were required to establish the brain. However, by 24 months more than 80% of the cells in the striatum had generated an altered repeat tract length, with a strong bias toward expansion. These observations by no means prove that transcription alters repeat stability in human patients; for example, instability could arise via aberrant repair triggered by periodic DNA damage. Nevertheless, they do set the stage for a serious consideration of that possibility.

TRANSCRIPTION PROMOTES REPEAT INSTABILITY IN MODEL SYSTEMS

Transcription was initially suggested as a possible cause for repeat instability based on observations in four lines of transgenic mice that carried a portion of the huntingtin gene containing 55 CAG repeats [10]. In three lines the transgene was actively expressed and the repeat was unstable, whereas in the fourth the transgene was silent and the repeat was stable. This correlation has been challenged because no simple relationship exists between transcript levels—one measure of transcription—and tissue-specific repeat instability in transgenic mice. In a transgenic DM1 mouse model with a (CTG)₅₅ tract, for example, low levels of DMPK mRNA were present in liver and kidney, but repeat instability was high; by contrast, high levels of DMPK mRNA in gastrocnemius and heart (4-fold and 20-fold, respectively, above liver and kidney) were associated with low levels of repeat instability [11]. These observations do not rule out a transcription-induced pathway of repeat instability for two reasons: stable mRNA levels may not accurately reflect the rates of transcription in different tissues; and transcription may not be the rate-limiting step in the transcription-induced instability pathway, which depends on downstream DNA repair activities in addition to transcription itself. The inability to draw firm conclusions from these sorts of correlations in mice has led to the exploitation of more easily manipulatable model systems.

Transcription Destabilizes Repeat Instability in Bacteria and Yeast

The key feature of all model systems for investigating the role of transcription in repeat instability is the ability to control transcription through the repeat so that the effects of transcription can be isolated from other influences. In bacteria, repeat sequences were cloned into the *lacZ* transcription unit so that transcription could be controlled by the inducer, IPTG (Figure 1) [12,13]. Induction of transcription—up to 10-fold—was shown to increase repeat instability moderately, with effects that depended on the length, orientation, and composition of the repeat sequences. For example, transcription did not detectably alter the stability of CAG repeats with 50 repeat units, but did destabilize repeats tracts of 64 and 175 repeats. GAC repeats displayed a similar sensitivity to transcription, with tract lengths of around 30 being stable and those longer than 49, unstable. Orientation of the repeat tract relative to the direction of transcription also affected the stability of the repeat. In the template strand, (GAC)₄₉, for example, was much more unstable than (GTC)₅₃, but this effect disappeared at longer lengths. Surprisingly, a (CAG)₁₇₅ tract, which carried two

G to A interruptions, displayed a dramatic orientation dependence, with the more susceptible orientation carrying CAG sequence in the template strand. These general effects of length, orientation, and composition of repeat tracts are consistent with transcription-induced secondary structure formation by the repeat.

Transcription-induced repeat stability in bacteria has been interpreted largely as a consequence of interaction between transcription and replication, which occur at the same time [13,14]. DNA polymerase complexes appear to slow down or stall when they encounter an RNA polymerase [15], and they also pause when they encounter long triplet repeats [16]. Thus, it is not hard to imagine that transcription plus long repeats might have synergistic effects on replication-induced destabilization of triplet repeats. The orientation dependence and preponderance of deletions in these studies has led to the speculation that CTG hairpins form on the nontemplate strand during transcription and are then bypassed by the DNA polymerase complex to cause deletions [14].

This simple idea that transcription-induced secondary structures are bypassed by DNA polymerase is unlikely to be the complete story in bacteria because of the demonstrated involvement of nucleotide excision repair (NER). For example, in the absence of transcription, the stability of (CAG)₁₇₅ repeats was unaffected by mutations in the NER repair genes, *uvrA* or *uvrB* [14]. When transcription was induced, however, repeat instability was significantly altered. In one study, mutations in *uvrA* increased instability, while mutations in *uvrB* decreased it [14]. In a second study mutations in *uvrA* and *uvrB* were both shown to reduce repeat instability [17]. Although these two studies differ in their specific conclusions, both indicated that NER can influence triplet repeat stability.

The relationship between triplet repeats and transcription has not been investigated in yeast; however, the stability of GT dinucleotide repeats has been shown to be sensitive to transcription [18]. When transcription was induced by addition of galactose, the rates of change in repeat length increased by up to 9-fold (Figure 1). Although the effect of NER was not examined, another DNA repair process—mismatch repair (MMR)—was shown to interact with transcription. Mutations in MMR genes by themselves increased repeat instability by 100-fold, generally causing 1- to 2-unit changes, consistent with slippage during DNA replication. Turning on transcription boosted this effect another 2- to 3-fold [18]. As in the bacterial studies, these results were interpreted in terms of the ability of transcription to interfere with the fidelity of DNA replication, causing either more slippage by DNA polymerase or inhibiting the normal MMR-mediated repair of such events.

Transcription Destabilizes Repeat Tracts in Human Cells and Flies

Although mammalian cells have been used to examine various aspects of repeat stability [19–21], only recently was a selectable system combined with an inducible promoter to allow the effects of transcription to be sensitively and directly analyzed in human cells (Figure 1). In this assay system, the *HPRT* minigene, driven by a Tet-ON inducible promoter, was modified to carry a (CAG)₉₅ repeat in its intron, oriented so that the CAG repeat appears in the RNA transcript [22]. This arrangement works as a selection assay because long CAG repeats are aberrantly spliced into the mRNA, rendering the encoded protein nonfunctional, whereas short repeat tracts (less than 39 repeats) allow proper splicing, producing functional HPRT [23]. Thus, selection for HPRT⁺ colonies provides a ready measure of CAG-repeat contractions to less than 39 units.

The underlying assumption in this system is that by assaying for contractions selectively, insight can be gained into the process of repeat instability in humans, which is generally biased toward expansions. This assumption has been tested in two instances by examining the effects of gene deficiencies in mice. In one case, in which genome-wide demethylation

or DNMT1 knockdown was shown to dramatically increase contractions in cells [21,24], a deficiency of DNMT1, the major maintenance DNA methyltransferase, was shown to increase expansions in the mouse germline, but not in somatic tissues [24]. In the second, knockdown of XPA—a key protein involved in nucleotide excision repair—was shown to decrease contractions in human cells [22]. When tested in mice, *Xpa* nulls were shown to radically reduce expansions in the striatum (the tissue with the highest level of instability), but not in the germline (Hubert and Wilson, unpublished). Thus, to the limited extent it has been tested, this cell-based assay for contractions appears to predict instability in mouse models, and by extension in humans. It is also clear, however, that the cell-based assay does not predict which tissues—somatic or germline—will be affected; that information must be ascertained by direct measure in whole organisms.

The link between transcription and repeat instability was tested in this inducible *HPRT* selection system by addition of doxycycline, which increases transcription through the *HPRT* minigene by 25-fold over a low background. When transcription was induced, the rate of repeat contraction increased with an average of 15-fold to about 6×10^{-6} in four cell lines, which had the *HPRT* minigene integrated at different sites in the genome [22]. These results indicate a strong connection between transcription and repeat instability.

The suggestion from bacterial and yeast studies that transcription might affect repeat instability via an interaction with replication was addressed in this system by comparing transcription-induced contraction frequencies in proliferating and confluent cells, which differ by 10-fold in their rates of cell division. These two cell populations accumulated transcription-induced repeat contractions at the same rate: about 2×10^{-6} contractions per day of doxycycline treatment over a period of a week [22]. The observation that transcription-induced repeat instability, as measured by repeat contractions, was independent of cell division rates suggests strongly that instability does not depend on DNA replication. This conclusion is strengthened by the observation that siRNA knockdown of FEN1, a flap endonuclease required for processing Okazaki fragments, has no effect on repeat contractions in human cells [22], even though mutation of the corresponding gene in yeast (*RAD27*) dramatically increases both repeat expansions and contractions [25]. These studies support the potential role of transcription in inducing repeat instability in terminally differentiated cells such as neurons, which no longer replicate their genomes.

In vivo support for transcription-induced repeat instability was generated by a study in flies that carried a (CAG)₇₈ transgene containing a segment of the *SCA3* gene from humans [26]. By mating, the transgene could be exposed to a germline-specific driver of transcription, so that repeat instability could be compared in parallel lines in the presence and absence of transcription. A striking enhancement of instability—with a 3:1 bias toward repeat expansions—was observed in lines in which the transgene was transcribed [26]. Transgenes at different locations in the genome, which showed the same low level of instability in the absence of transcription, displayed a 5-fold difference in instability when transcription was turned on, even though the levels of transcription were nearly identical. The basis for the difference in instability of transcribed transgenes at different locations is unclear, but may indicate that local context effects also play a role [1]. These studies confirm the importance of transcription in inducing instability *in vivo* and at the same time challenge the naïve expectation that levels of transcription will be directly correlated with levels of repeat instability.

MECHANISMS OF TRANSCRIPTION-INDUCED REPEAT INSTABILITY

Although transcription clearly promotes repeat instability in bacteria, yeast, flies, and mammalian cells, transcription, by itself, cannot alter the length of DNA. Transcription

likely triggers instability indirectly by exposing single strands of DNA, which allows formation of aberrant secondary structures that in turn engage one or more DNA repair processes, which are ultimately responsible for changing the length of the repeat tract. In various model systems, the effects of gene deficiency or gene knockdown on repeat instability have provided important insights in the roles of various proteins on transcription-induced instability (Table 1). The involvement of several repair proteins in transcription-induced repeat instability has been documented in human cells by siRNA knockdowns and in flies by mutation analysis [22,26,27]. Based on genetic studies in human cells, we proposed a speculative model for how the length of a repeat tract might change during transcription (Figure 2). In this model, slipped-strand structures are imagined to form in the wake of a passing RNA polymerase. Depending on the arrangement of CAG loops and CTG hairpins, repair of the structure could lead to contracted, expanded, or unchanged repeat tracts. Right or wrong, this model provides a convenient picture, around which to organize a discussion of potential and identified contributions to transcription-induced repeat instability.

RNA Polymerase Stalls at Repeat Tracks

Demonstrating that secondary structures form during transcription in cells is a challenge, but *in vitro* studies have demonstrated that RNA polymerase alters its behavior at repeat tracks and in the presence of non-B DNA structures. In nuclear extracts of HeLa cells, for example, about 50% of RNA polymerase II (RNAPII) transiently pauses within the first few units of CNG tracts ranging from 17 to 255 repeats [28]. Surprisingly, different lengths of the same type of repeat blocked transcription with similar efficiency. In contrast, the rate of transcription through a repeat tract depended on the sequence of the repeat on the template strand, in the order CTG, CCG, CGG, and CAG, from most inhibitory to least. The transient pausing of RNAPII at repeat tracts could be caused by intrastrand secondary structures, which might act as physical or energetic barriers to transcription.

In other studies, T7 RNA polymerase (RNAP) was shown to pause at sequences capable of forming triple-helical structures, G-rich quadruplex DNA, and Z-DNA [29–31]. For example, elongation by T7 RNAP on a closed-circular plasmid was inhibited at (CG)₁₄ repeats, which can form Z-DNA, and its pausing was significantly enhanced by negative supercoiling, which facilitates formation of Z-DNA [29]. The partial transcription blockage at triplex or quadruplex structures may be directly relevant to the instability of GAA and CCG repeats observed in humans, since they can form triplex and quadruplex structures, respectively [32]. T7 RNAP did not pause at a palindromic sequence, which can form intrastrand hairpins. Palindromic hairpins differ in two key ways from the CNG hairpins that form in slipped-duplex DNA: they are perfectly paired and they are symmetrically arrayed. Hairpins in slipped-duplex DNA, by contrast, contain mismatches at every third base and they project from the parental duplex at asymmetric positions. As discussed below, additional proteins can bind to mismatched CNG hairpins; thus, it may be that protein-DNA complexes formed at a repeat hairpin, rather than the hairpin itself, provides the roadblock for RNAPII.

In addition to reacting to the presence of aberrant structures in the DNA, RNAPII may also contribute to their formation. For example, G quartets have been observed to form on the nontemplate strand during transcription [33]. Formation of such structures may be favored by extended stretches of RNA-DNA hybrid (R-loops), which can form just behind a passing polymerase in regions where the template strand is C-rich, due to the exceptional stability of rG:dC base pairs [34]. R-loops have been documented in stretches of template that are 34% C [35], which is about the same as for CAG, CTG, or CGG repeats in the template strand. An R-loop in the template strand would favor formation of hairpin structures on the nontemplate strand. Subsequent resolution of the R-loop and re-pairing of the DNA strands

could force the template strand to form a hairpin to compensate for the one in the nontemplate strand, thereby providing a transcription-induced route to slipped duplex formation. If R-loops are involved in repeat instability, then treatments that alter their kinetics of formation or removal should affect repeat stability.

Components of Mismatch Repair Alter Transcription-Induced Repeat Stability

Slipped-strand structures, as illustrated in Figure 2, are reasonably stable, but are unlikely to block progression of RNAPII. Both CAG and CTG hairpins, however, are bound by mismatch repair (MMR) proteins [36,37], which could increase their stability sufficiently to cause problems for RNAPII. The MSH2/MSH3 complex (MutS β), for example, binds to CAG structures with a low-nanomolar K_d [37]. Such a protein-stabilized structure in the template strand might constitute a significant barrier to RNAPII.

The roles of various MMR proteins in repeat instability during transcription were tested in a selective assay in human cells, using siRNA knockdowns [22]. Cells were treated with siRNAs against individual MMR components, transcription through the repeat was then induced, and the frequency of CAG repeat contraction was measured. Knockdowns of MSH2 and MSH3, which form MutS β , reduced the frequencies of transcription-induced contraction, implying that the *normal* activity of MutS β destabilizes repeats. Knockdown of MSH6, which complexes with MSH2 to form MutS α , did not affect repeat stability, suggesting that MutS α is not involved in transcription-induced repeat instability [22].

The destabilizing effect of MutS β mismatch recognition complex during transcription contrasts with the effects of downstream components involved in MMR. Knockdown of MLH1 and PMS2, which form a complex involved in the repair of mismatched nucleotides and small insertions and deletions, increased the frequency of repeat contractions 2- to 3-fold (Lin and Wilson, unpublished). Why mismatch-recognition and downstream-processing complexes should produce opposite effects in transcription-induced repeat instability is unclear, but that question will likely be resolved only by additional biochemical studies. Nevertheless, both sets of results confirm that MMR components play key roles in transcription-induced repeat instability in human cells.

Although no model for transcription-induced repeat instability has yet been reported in mice, the influence of MMR on repeat instability has been extensively investigated. In several different mouse models, both intergenerational and somatic repeat instability is strongly influenced by mutations in MMR genes [38–41]. Transgenic and knock-in models of HD and DM1, with repeat tracts ranging from 84 to 300 repeats, were bred onto backgrounds that were null for various MMR genes and the stabilities of the repeat tracts were analyzed in germline and somatic tissues. Although there is not complete agreement among these studies, perhaps reflecting the differences in the lengths of the tracts or their locations in the genome, it is clear that MMR genes significantly affect repeat stability. On an *MSH2*^{-/-}, *MSH3*^{-/-}, or *PMS2*^{-/-} background, the instability normally evident in sperm and various somatic tissues was substantially reduced [38,42] or shifted toward contractions [43], consistent with the normal role of the corresponding proteins in expansion-biased repeat instability. On an *MSH6*^{-/-} background, triplet repeat instability, especially expansion, was markedly enhanced [38]. This initially surprising result has been interpreted in terms of a competition between MSH3 and MSH6 for binding to MSH2 [38]. In the absence of MSH6, additional MSH2/MSH3 complex is formed, which functions to increase the instability of the repeat tract. These results are similar to the results in human cells, where MMR normally functions to promote transcription-induced repeat instability [22].

Transcription-Coupled Nucleotide Excision Repair Destabilizes Repeat Tracts

If a protein-stabilized CAG or CTG hairpin could block progression of RNAPII, it would likely trigger transcription-coupled NER (TC-NER) since a stalled RNAPII is thought to be the primary signal to initiate that process [44]. siRNA knockdowns were used to assess the roles of various NER proteins in repeat instability during transcription in a selective assay in human cells [27]. Cells were treated with siRNAs against individual NER components, transcription through the repeat was then induced, and the frequency of CAG repeat contraction was measured. Knockdowns of CSB, XPA, ERCC1, and XPG significantly reduced the frequency of contractions, while knockdown of XPC had no effect [22,27]. The involvement of CSB, which is specifically required for TC-NER, and lack of involvement of XPC, which is specific for global genome NER, confirm that repeat instability is linked to the transcription-coupled subpathway of NER [45]. These results, like those with MSH2 and MSH3, indicate that it is the normal activities of TC-NER proteins that cause repeat instability.

Recently, a role for TC-NER in repeat instability has also been suggested based on experiments in flies [26]. The transcription-induced germline instability of a (CAG)₇₈-containing *SCA3* transgene was found to be significantly reduced on a background that carried null mutations in the gene encoding Mus201, the fly homolog of XPG. The reduction in transcription-induced repeat instability argues that the *Mus201* mutations affect TC-NER, rather than global genome NER, in which Mus201 also plays a role. Repeat instability in both the male and female germlines was affected. In the ovaries, the level of *SCA3* mRNA was unaffected by the *Mus201* mutations; however, in the male germline the *Mus201* mutations unexpectedly decreased the level of transcription through the repeat, making it unclear whether the proximate cause of reduced instability was lack of Mus201 or decreased transcription. Nevertheless, these studies support a role for TC-NER in the female germline, and—like the studies in human cells—argue strongly that the normal activity of TC-NER destabilizes repeats.

The specific targeting of TC-NER toward removal of DNA lesions on the template strand for transcription focuses attention on structures that form on the template strand in repeat tracts and their potential stabilization by MutSβ. The possibility that TC-NER and components of MMR might collaborate in the transcription-induced pathway for repeat instability (Figure 2) is supported by experiments in human cells that showed that mixtures of siRNAs against XPA and MSH2 reduced repeat instability to the same extent as either siRNA alone [27]. Although the mechanism for collaboration of MMR components with TC-NER in the pathway for transcription-induced repeat instability is still unclear, interactions between components of MMR and NER have been documented in several other aspects of DNA repair, including repair of UV-induced DNA damage, removal of psoralen crosslinks, and double-strand break repair [5].

MODULATORS OF TRANSCRIPTION-INDUCED REPEAT INSTABILITY

Transcription in eukaryotic cells is a complicated process. In addition to the repetitive linkage of nucleotides into the nascent RNA, transcription involves the remodeling of chromatin structure, promoter identification and initiation of RNA synthesis, unwinding of the DNA helix and resolution of the resultant supercoils, and monitoring for dysfunctional transcription complexes, among other processes. Any—or all—of these elements of the transcription process could influence the pathway for transcription-induced repeat instability in cells and organisms. A number of these transcription-related processes have been examined in model systems for their influence on repeat instability.

Positive and Negative Supercoiling

As RNA polymerase transcribes the template strand, it changes the local topology of the DNA, introducing positive supercoils ahead of and negative supercoils behind the moving polymerase. Statistical mechanical calculations show that CNG repeat tracts supercoil more readily than random B-DNA sequences [46], raising the possibility that supercoiling pressures could play a role in repeat instability. In bacteria, where the levels of supercoiling can be manipulated by inhibition or mutation of specific topoisomerases, GAA and CGG repeat tracts (but not CAG tracts) were found to be destabilized by higher negative supercoil density, and these effects were enhanced by transcription through the repeat [47]. Thus, transcription and supercoiling each cause repeat instability in bacteria, and their effects may be tied together.

Using a selectable system in human cells, we showed that inhibition of topoisomerase I activity by camptothecin, or by knockdown with siRNAs, increased the instability of a (CAG)₉₅ repeat tract several fold, but only when transcription through the repeat was induced (Hubert, Lin, and Wilson, unpublished). The link between topoisomerase I activity, transcription, and CAG repeat instability in human cells is not yet clear. One reasonable possibility is that when topoisomerase I activity is compromised, transcription through the CAG repeat tract generates a higher-than-normal level of local supercoiling, eliciting more efficient formation of secondary structures in the repeat tract, which in turn triggers downstream repair events that change tract length. Alternatively, direct repair of stuck topoisomerase-I complexes—especially in the case of camptothecin inhibition—in transcribed areas may enhance exposure of single strands in the repeat region, triggering structure formation, repair, and tract-length change. Regardless of mechanism, these studies clearly tie DNA topoisomerase I—and by extension, DNA topology—to transcription-induced triplet repeat instability.

Unwinding DNA Secondary Structures

Since the proximate cause of repeat instability is thought to be the presence of abnormal secondary structures, it would seem natural that DNA helicases might provide a route for their resolution that would not entail a change in tract length. By unwinding a slipped-strand structure, for example, a helicase would give the DNA strands a second chance to pair correctly. The possibility that DNA helicases are involved in repeat stability has been examined only in yeast [48]. There, the Srs2 helicase was identified in a blind screen for mutations that stimulate repeat expansion in a selective system in which a triplet repeat tract separated a distance-sensitive promoter from the start site for transcription. In this system, which can be used to measure either expansion or contraction, a short repeat allows transcription, but a long repeat prevents it [48]. Mutations affecting Srs2 stimulated CTG, CAG, and CGG repeat expansions several fold, but had no effect on repeat contractions. Thus, as might be expected, the normal activity of Srs2 protects against repeat expansion.

Epistasis analysis in this system suggests that Srs2 functions in a pathway with DNA polymerase delta, implying that expansions are prevented as a consequence of Srs2 function during DNA replication [48]. The selective effect of Srs2 on expansions raises the possibility of a link to transcription, as well. In the expansion assay, a short repeat allows transcription, and the location of the repeat immediately upstream of the transcription start site means that the repeat will be exposed to negative supercoiling pressure. By contrast, in the contraction assay, a long repeat prevents transcription, eliminating transcription-induced supercoiling pressure on the repeat. Thus, in principle, transcription-induced supercoiling could be responsible for generating the secondary structures that are subsequently resolved by Srs2 during replication. Clearly, the involvement of DNA helicases in repeat stability is a topic deserving of additional study in yeast and other model systems.

Fate of Stalled RNA Polymerase

Transcription-induced repeat instability is linked to TC-NER in human cells [27]. The classic signal to engage TC-NER is a stalled RNAPII complex at a site of DNA damage [49]. In Figure 2, we suggest that RNAPII might also stall at protein-stabilized secondary structures in a repeat tract. Regardless of the nature of the block, it is likely that RNAPII will need to be displaced to permit access by NER proteins. Proposed mechanisms for moving RNAPII out of the way in eukaryotic cells include RNAPII backtracking and proteasomal degradation of RNAPII, among other possibilities. In human cells, CAG repeat instability was reduced by siRNA knockdown of TFIIS [27], a transcription factor required for backtracking of a stalled RNAPII, as well as during transcription initiation and elongation [50,51]. Knockdown of TFIIS did not affect transcription through the *HPRT* minigene [27], ruling out the possibility that repeat stabilization was caused by reduced transcription through the repeat. This result suggests that RNAPII backtracking may be involved in transcription-induced repeat instability. The same system also provides support for the idea of proteasomal degradation of a stalled RNAPII [27]. Treatment with the proteasome inhibitor, MG132, stabilized repeat tracts, as expected if TC-NER was prevented. In addition, the knockdown of either component of the BRCA1/BARD1 E3 ubiquitin ligase, which can transfer ubiquitin to a stalled RNAPII, also stabilized repeat tracts. Although the BRCA1/BARD1 ligase has multiple roles in the cell, these results are consistent with the notion that RNAPII degradation is a part of the transcription-induced pathway for repeat instability.

The potential involvement of both RNAPII backtracking (TFIIS) and RNAPII degradation (BRCA1/BARD1) in transcription-induced CAG repeat contraction raises the question of how these apparent alternatives might be employed in the *same* pathway, as they seem to be, since the effect of combined knockdown of TFIIS and BRCA1 is equal to either knockdown alone [27]. One possibility is that RNAPII backtracking is required to set the stage for ubiquitination of RNAPII by BRCA1/BARD1. Another possibility is that the two processes are linked because of the specific properties of the repeats themselves. Backtracking from a hairpin in the middle of a repeat tract may not allow RNAPII to escape; instead, the hairpin may grow in size as RNAPII attempts to back off, maintaining contact with RNAPII, preventing access by repair factors, and ultimately triggering ubiquitination and degradation. The true relevance of these processes to transcription-induced repeat instability will need to be clarified in biochemical studies.

Convergent Transcription

Recent reports indicate that a surprisingly high fraction of human genes are associated with antisense transcripts, suggesting an unexpected level of convergent transcription in the human genome [52]. Intriguingly, antisense transcription has been found at several repeat disease gene loci, including DM1 [53], SCA8 [54], and, fragile X syndrome [55]. The possibility that antisense transcription might play a role in repeat instability could help to explain a set of observations that is otherwise difficult to reconcile with the simple concept of transcription-induced repeat instability. Transgenic mouse models for HD, SCA1, SCA7, and SBMA, with a range of repeat lengths, have been developed using expressed cDNAs instead of genomic fragments [5]. The repeat tracts in these models are typically much more stable than in models built with genomic fragments. In one comprehensive study, transgenic mice with a full-length *SCA7* cDNA were compared with those carrying a 13.5-kb *SCA7* genomic fragment, both of which contained a (CAG)₉₂ repeat tract [56]. Strikingly, mice with the cDNA construct displayed minimal repeat instability, while mice with the genomic fragment showed high instability [56]. Repeats in the genomic fragment could be partially stabilized by deleting most of the 3' flanking sequences, suggesting a role for a *cis*-element

in generating repeat instability. One possibility—among many—is that this element is needed for antisense transcription.

As one approach for testing whether convergent transcription could destabilize repeats, we modified our selective assay in human cells so that transcription could be independently induced from either end of the gene (Lin and Wilson, unpublished). Inducing transcription from one end or the other stimulated repeat instability to about the same extent; however, simultaneous transcription from both ends produced a synergistic increase in repeat instability that was more than the sum of either induction alone (Lin and Wilson, unpublished). These results suggest that convergent transcription may be especially destabilizing for triplet repeat tracts.

Chromatin Structure

Transcription is regulated in part by chromatin structure, which is linked to DNA and histone modification. The potential involvement of these epigenetic marks in repeat instability has not been thoroughly investigated, although a few observations support that possibility. At the fragile X syndrome 1 locus in humans, for example, CGG repeats, which are subject to CpG methylation, become highly methylated when the repeat tract exceeds about 200 units in length [1]. Methylation prevents transcription and the methylated repeat tracts are stable [4]. Similarly, CAG repeats in CHO cells and human patient cells were shown to be exquisitely sensitive to 5-aza-cytidine-induced genome-wide demethylation [21]. In the SCA1 mouse model, deficiency of the maintenance DNA methyltransferase, DNMT1, was found to promote repeat expansion [24]. Aberrant DNA and histone methylation was also found at a conserved CpG island adjacent to the repeat tract [24]. The connection, if any, between these epigenetic modifications and transcription is unclear, but the general role of chromatin structure in repeat instability is an area that is ripe for investigation.

SUMMARY AND PERSPECTIVES

One certainty in the field of triplet repeat instability is the central importance of the ability of repeat tracts to form aberrant secondary structures such as hairpins and slipped-strand duplexes. These secondary structures are thought to arise when the strands of DNA are separated, which permits intra-strand structures to form. Such structures evidently make it difficult for the cell's normal DNA repair machinery to return the sequence to its initial length, leading to expansion-biased instability in the germline and somatic tissues of humans. For human patients the identities of the critical processes that expose single strands to allow aberrant structure formation and of those that mishandle the problem of structure removal are not known for sure. It is not even clear whether repeat instability results from one major pathway or from multiple distinct or overlapping pathways.

In this review, we have focused on the transcription-induced pathway for repeat instability. There is ample evidence from studies in model systems that transcription destabilizes repeats, and in human cells genetic data indicate that the normal functions of MMR and TC-NER stimulate repeat instability. At this point, however, the relevance of transcription to the triplet repeat instability observed in human germline and somatic tissues is unclear. The only thing that is certain is that the disease genes are widely transcribed, a pre-requisite for such a pathway. This pathway is also attractive because it could account naturally for the age-dependent accumulation of repeat-length changes in terminally differentiated cells such as neurons, which no longer replicate their DNA. What is needed is a system in mice in which transcription can be induced in germline or specific tissues. Such an inducible system would allow the influence of transcription to be assessed and to be tested in combination with genetic alterations in likely DNA repair genes.

Acknowledgments

Work on triplet repeat instability is supported by grants from the NIH to J.H.W. (GM38219) and to L.H. (1F31HG004918-01).

Abbreviations

APEX	apurinic/apyrimidinic endonuclease 1
BARD1	BRCA1-associated RING domain 1
BRCA1	breast cancer type 1 susceptibility protein
CREB	cAMP response element binding protein
CSB	Cockayne's syndrome type B
DM1	myotonic dystrophy 1
DMPK	myotonic dystrophy protein kinase
DNMT1	DNA methyltransferase 1
ERCC1	excision repair cross complementation group 1
FEN1	flap endonuclease 1
HAT	hypoxanthine aminopterin thymine
HD	Huntington's disease
HPRT	hypoxanthin phosphoribosyltransferase
IPTG	isopropyl β -D-1-thiogalactopyranoside
MLH	<i>mutL</i> homologue
MMR	mismatch repair
MSH	<i>mutS</i> homologue
MutSa	MSH2/MSH6 complex
MutSβ	MSH2/MSH3 complex
NER	nucleotide excision repair
OGG1	8-oxoguanine DNA glycosylase
PMS2	postmeiotic segregation increased 2
RNAP	RNA polymerase
RNAPII	RNA polymerase II
SBMA	spinal and bulbar muscular atrophy
SCA	spinocerebellar ataxia
siRNA	small interfering RNA
TC-NER	transcription-coupled nucleotide excision repair
TFIIS	transcription factor IIS
TOP1	topoisomerase 1
UV	ultraviolet
XP(A	C, F, G), xeroderma pigmentosum group (A, C, F, G)-complementing protein

References

1. Cleary JD, Pearson CE. The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet Genome Res.* 2003; 100(1–4):25–55. [PubMed: 14526163]
2. Lenzmeier BA, Freudenreich CH. Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair. *Cytogenet Genome Res.* 2003; 100(1–4):7–24. [PubMed: 14526162]
3. Orr HT, Zoghbi HY. Trinucleotide repeat disorders. *Annu Rev Neurosci.* 2007; 30:575–621. [PubMed: 17417937]
4. Pearson CE, Edamura KN, Cleary JD. Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet.* 2005; 6(10):729–742. [PubMed: 16205713]
5. Lin, Y.; Dion, V.; Wilson, JH. Transcription and triplet repeat instability. In: Wells, R.; Ashizawa, T., editors. *Genetic Instability and Neurological Diseases*. Elsevier; Amsterdam: 2006. p. 691-704.
6. Telenius H, Kremer B, Goldberg YP, et al. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat Genet.* 1994; 6(4):409–414. [PubMed: 8054984]
7. Pearson CE, Sinden RR. Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry.* 1996; 35(15):5041–5053. [PubMed: 8664297]
8. Gacy AM, Goellner G, Juranic N, Macura S, McMurray CT. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell.* 1995; 81(4):533–540. [PubMed: 7758107]
9. Kennedy L, Evans E, Chen CM, et al. Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum Mol Genet.* 2003; 12(24):3359–3367. [PubMed: 14570710]
10. Mangiarini L, Sathasivam K, Mahal A, Mott R, Seller M, Bates GP. Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat Genet.* 1997; 15(2):197–200. [PubMed: 9020849]
11. Lia AS, Seznec H, Hofmann-Radvanyi H, et al. Somatic instability of the CTG repeat in mice transgenic for the myotonic dystrophy region is age dependent but not correlated to the relative intertissue transcription levels and proliferative capacities. *Hum Mol Genet.* 1998; 7(8):1285–1291. [PubMed: 9668171]
12. Mochmann LH, Wells RD. Transcription influences the types of deletion and expansion products in an orientation-dependent manner from GAC*GTC repeats. *Nucleic Acids Res.* 2004; 32(15):4469–4479. [PubMed: 15317871]
13. Bowater RP, Jaworski A, Larson JE, Parniewski P, Wells RD. Transcription increases the deletion frequency of long CTG. CAG triplet repeats from plasmids in *Escherichia coli*. *Nucleic Acids Res.* 1997; 25(14):2861–2868. [PubMed: 9207036]
14. Parniewski P, Bacolla A, Jaworski A, Wells RD. Nucleotide excision repair affects the stability of long transcribed (CTG*CAG) tracts in an orientation-dependent manner in *Escherichia coli*. *Nucleic Acids Res.* 1999; 27(2):616–623. [PubMed: 9862988]
15. Liu B, Alberts BM. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science.* 1995; 267(5201):1131–1137. [PubMed: 7855590]
16. Kang S, Jaworski A, Ohshima K, Wells RD. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nat Genet.* 1995; 10(2):213–218. [PubMed: 7663518]
17. Oussatcheva EA, Hashem VI, Zou Y, Sinden RR, Potaman VN. Involvement of the nucleotide excision repair protein UvrA in instability of CAG*CTG repeat sequences in *Escherichia coli*. *J Biol Chem.* 2001; 276(33):30878–30884. [PubMed: 11413147]
18. Wierdl M, Greene CN, Datta A, Jinks-Robertson S, Petes TD. Destabilization of simple repetitive DNA sequences by transcription in yeast. *Genetics.* 1996; 143(2):713–721. [PubMed: 8725221]
19. Lin Y, Dion V, Wilson JH. A novel selectable system for detecting expansion of CAG. CTG repeats in mammalian cells. *Mutat Res.* 2005; 572(1–2):123–131. [PubMed: 15790495]
20. Claassen DA, Lahue RS. Expansions of CAG{middle dot}CTG repeats in immortalized human astrocytes. *Hum Mol Genet.* 2007; 16(24):3088–3096. [PubMed: 17881653]

21. Gorbunova V, Seluanov A, Mittelman D, Wilson JH. Genome-wide demethylation destabilizes CTG{middle dot}CAG trinucleotide repeats in mammalian cells. *Hum Mol Genet.* 2004; 13(23): 2979–2989. [PubMed: 15459182]
22. Lin Y, Dion V, Wilson JH. Transcription promotes contraction of CAG repeat tracts in human cells. *Nat Struct Mol Biol.* 2006; 13(2):179–180. [PubMed: 16388310]
23. Gorbunova V, Seluanov A, Dion V, Sandor Z, Meservy JL, Wilson JH. Selectable system for monitoring the instability of CTG/CAG triplet repeats in mammalian cells. *Mol Cell Biol.* 2003; 23(13):4485–4493. [PubMed: 12808091]
24. Dion V, Lin Y, Hubert L Jr, Waterland RA, Wilson JH. Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Hum Mol Genet.* 2008; 17(9):1306–1317. [PubMed: 18252747]
25. Freudenreich CH, Kantrow SM, Zakian VA. Expansion and length-dependent fragility of CTG repeats in yeast. *Science.* 1998; 279(5352):853–856. [PubMed: 9452383]
26. Jung J, Bonini N. CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. *Science.* 2007; 315(5820):1857–1859. [PubMed: 17332375]
27. Lin Y, Wilson JH. Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. *Mol Cell Biol.* 2007; 27(17):6209–6217. [PubMed: 17591697]
28. Parsons MA, Sinden RR, Izban MG. Transcriptional properties of RNA polymerase II within triplet repeat-containing DNA from the human myotonic dystrophy and fragile X loci. *J Biol Chem.* 1998; 273(41):26998–27008. [PubMed: 9756950]
29. Ditlevson JV, Tornaletti S, Belotserkovskii BP, et al. Inhibitory effect of a short Z-DNA forming sequence on transcription elongation by T7 RNA polymerase. *Nucleic Acids Res.* 2008
30. Belotserkovskii BP, De Silva E, Tornaletti S, Wang G, Vasquez KM, Hanawalt PC. A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J Biol Chem.* 2007; 282(44):32433–32441. [PubMed: 17785457]
31. Tornaletti S, Park-Snyder S, Hanawalt PC. G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. *J Biol Chem.* 2008
32. Wells RD. Non-B DNA conformations, mutagenesis and disease. *Trends Biochem Sci.* 2007; 32(6):271–278. [PubMed: 17493823]
33. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* 2004; 18(13):1618–1629. [PubMed: 15231739]
34. Sugimoto N, Nakano S, Katoh M, et al. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry.* 1995; 34(35):11211–11216. [PubMed: 7545436]
35. Li X, Manley JL. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell.* 2005; 122(3):365–378. [PubMed: 16096057]
36. Pearson CE, Ewel A, Acharya S, Fishel RA, Sinden RR. Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum Mol Genet.* 1997; 6(7): 1117–1123. [PubMed: 9215683]
37. Owen BA, Yang Z, Lai M, et al. (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat Struct Mol Biol.* 2005; 12(8):663–670. [PubMed: 16025128]
38. van den Broek WJ, Nelen MR, Wansink DG, et al. Somatic expansion behaviour of the (CTG)_n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum Mol Genet.* 2002; 11(2):191–198. [PubMed: 11809728]
39. Savouret C, Brisson E, Essers J, et al. CTG repeat instability and size variation timing in DNA repair-deficient mice. *Embo J.* 2003; 22(9):2264–2273. [PubMed: 12727892]
40. Savouret C, Garcia-Cordier C, Megret J, te Riele H, Junien C, Gourdon G. MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Mol Cell Biol.* 2004; 24(2):629–637. [PubMed: 14701736]
41. Kovtun IV, McMurray CT. Trinucleotide expansion in haploid germ cells by gap repair. *Nat Genet.* 2001; 27(4):407–411. [PubMed: 11279522]

42. Manley K, Shirley TL, Flaherty L, Messer A. Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet.* 1999; 23(4):471–473. [PubMed: 10581038]
43. Gomes-Pereira M, Fortune MT, Ingram L, McAbney JP, Monckton DG. Pms2 is a genetic enhancer of trinucleotide CAG. CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum Mol Genet.* 2004; 13(16):1815–1825. [PubMed: 15198993]
44. Svejstrup JQ. Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol.* 2002; 3(1):21–29. [PubMed: 11823795]
45. Hanawalt PC. Subpathways of nucleotide excision repair and their regulation. *Oncogene.* 2002; 21(58):8949–8956. [PubMed: 12483511]
46. Gellibolian R, Bacolla A, Wells RD. Triplet repeat instability and DNA topology: an expansion model based on statistical mechanics. *J Biol Chem.* 1997; 272(27):16793–16797. [PubMed: 9201984]
47. Napierala M, Bacolla A, Wells RD. Increased negative superhelical density in vivo enhances the genetic instability of triplet repeat sequences. *J Biol Chem.* 2005; 280(45):37366–37376. [PubMed: 16166072]
48. Bhattacharyya S, Lahue RS. *Saccharomyces cerevisiae* Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. *Mol Cell Biol.* 2004; 24(17):7324–7330. [PubMed: 15314145]
49. Tornaletti S, Patrick SM, Turchi JJ, Hanawalt PC. Behavior of T7 RNA polymerase and mammalian RNA polymerase II at site-specific cisplatin adducts in the template DNA. *J Biol Chem.* 2003; 278(37):35791–35797. [PubMed: 12829693]
50. Guglielmi B, Soutourina J, Esnault C, Werner M. TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proc Natl Acad Sci U S A.* 2007; 104(41):16062–16067. [PubMed: 17901206]
51. Kim B, Nesvizhskii AI, Rani PG, Hahn S, Aebersold R, Ranish JA. The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc Natl Acad Sci U S A.* 2007; 104(41):16068–16073. [PubMed: 17913884]
52. Katayama S, Tomaru Y, Kasukawa T, et al. Antisense transcription in the mammalian transcriptome. *Science.* 2005; 309(5740):1564–1566. [PubMed: 16141073]
53. Cho DH, Thienes CP, Mahoney SE, Analau E, Filippova GN, Tapscott SJ. Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol Cell.* 2005; 20(3):483–489. [PubMed: 16285929]
54. Moseley ML, Zu T, Ikeda Y, et al. Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat Genet.* 2006; 38(7):758–769. [PubMed: 16804541]
55. Ladd PD, Smith LE, Rabaia NA, et al. An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Hum Mol Genet.* 2007; 16(24):3174–3187. [PubMed: 17921506]
56. Libby RT, Monckton DG, Fu YH, et al. Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. *Hum Mol Genet.* 2003; 12(1):41–50. [PubMed: 12490531]

Figure 1.

Substrates used for studying transcription-induced repeat instability in bacteria, yeast, flies, and human cells. In bacteria, transcription is turned on by addition of the inducer, IPTG. Changes in tract length are followed unselectively by gel analysis. In yeast and flies, transcription is turned on by addition of galactose. For yeast, the use of a dinucleotide repeat in the coding sequence shifts the reading frame so that the *Ura3* composed of the fusion gene cannot be made. Expansions or contractions of the repeat restore the reading frame and permit cell survival under selection. In flies, changes in tract length are followed unselectively by PCR. In human cells, the transcription is turned on by addition of doxycycline. CAG contractions are followed by HAT selection, which kills *HPRT⁻* cells.

Figure 2.

Speculative model for transcription-induced repeat instability. In this model, transcription is from left to right and the repeat is oriented so that CTG is on the template (bottom) strand and CAG is on the nontemplate (top) strand. The passage of RNAPII induces secondary structures to form in the repeat tract. Because of their lower thermal stability, CAG structures are shown as loops, while the more stable CTG structures are shown as hairpins. The higher single-strand character of CAG loops would allow them to branch-migrate, permitting multiple smaller loops to coalesce into a larger loop. Stabilization of a CTG hairpin by MutS β stalls the next RNAPII, which triggers TC-NER. Displacement of RNAPII allows NER proteins to gain access to the hairpin and initiate repair. As shown by the three different pathways, the particular arrangement of the CAG loops versus the CTG hairpins could give rise to three different outcomes: contraction, no change, and expansion of the bottom strand. Remaining structures on the template strand may be repaired by this mechanism, as well. Remaining structures on the nontemplate strand may involve other mechanisms.

Table 1
Effect of deficiency of repair proteins on transcription-induced repeat instability in model systems.

Pathway	Gene	Bacteria	Yeast ^a	Flies	Human Cells	Mice
MMR	<i>MSH2</i>		Destabilized [18]		Stabilize [22]	
	<i>MSH3</i>				Stabilize [22]	
	<i>MSH6</i>				No effect [22]	
	<i>MLH1</i>				Destabilize ^c	
	<i>PMS1</i>		Destabilized [18]			
	<i>PMS2</i>				Destabilize ^c	
	<i>MLH3</i>				No effect ^c	
NER	<i>DNMT1</i>				Destabilize [54]	
	<i>XPA</i>				Stabilize [22]	
	<i>CSB</i>				Stabilize [26]	
	<i>XPC</i>				No effect [22]	
	<i>XPF</i>				Stabilize [26]	
	<i>XPG</i>			Stabilize [25]	Stabilize [26]	
	<i>ERCC1</i>				Stabilize [26]	
	<i>uvrA</i>	Stab/Destab ^b [14,17]				
	<i>uvrB</i>	Stabilized [17]				
	<i>TFIIIS</i>				Stabilize [26]	
Others	<i>BRCA1</i>				Stabilize [26]	
	<i>BARD1</i>				Stabilize [26]	
	<i>TOP1</i>				Destabilize ^d	
	<i>FEN1</i>				No effect [22]	
	<i>APEX1</i>				No effect [26]	
	<i>OGG1</i>				No effect [26]	
	<i>CREB</i>			Destabilize [25]		

^a Only dinucleotide repeat instability was studied in yeast.

^b One study showed stabilization, while a second study showed destabilization.

^c Lin and Wilson, unpublished.

^d Hubert, Lin, and Wilson, unpublished.