Expansions of simple tandem repeats are responsible for almost 50 human diseases, the majority of which are severe, degenerative, and not currently treatable or preventable. In this review, we first describe the molecular mechanisms of repeat-induced toxicity, which is the connecting link between repeat expansions and pathology. We then survey alternative DNA structures that are formed by expandable repeats and review the evidence that formation of these structures is at the core of repeat instability. Next, we describe the consequences of the presence of long structure-forming repeats at the molecular level: somatic and intergenerational instability, fragility, and repeat-induced mutagenesis. We discuss the reasons for gender bias in intergenerational repeat instability and the tissue specificity of somatic repeat instability. We also review the known pathways in which DNA replication, transcription, DNA repair, and chromatin state interact and thereby promote repeat instability. We then discuss possible reasons for the persistence of disease-causing DNA repeats in the genome. We describe evidence suggesting that these repeats are a payoff for the advantages of having abundant simple-sequence repeats for eukaryotic genome function and evolvability. Finally, we discuss two unresolved fundamental questions: (i) why does repeat behavior differ between model systems and human pedigrees, and (ii) can we use current knowledge on repeat instability mechanisms to cure repeat expansion diseases?

In 1905, British ophthalmologist Edward Nettleship made a peculiar observation: children that suffered from certain degenerative genetic diseases exhibited pathological symptoms earlier than their parents. Nettleship termed this phenomenon “genetic anticipation” (1). This concept was further substantiated by Bruno Fleischer in his genetic studies of myotonic dystrophy (DM) with cataracts (2).

This work was supported by National Institutes of Health Grant R35GM130322 and by a contribution from the White family (to S. M. M.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

1 To whom correspondence should be addressed. E-mail: sergei.mirkin@tufts.edu.

2 The abbreviations used are: DM, myotonic dystrophy; HD, Huntington’s disease; SBMA, spinal and bulbar muscular atrophy; RED, repeat expansion disease; SCA, spinocerebellar ataxia; DRPLA, dentatorubral-pallidoluysian atrophy; XDP, X-linked dystonia parkinsonism; FAME3, familial adult myoclonic epilepsy 3; FRDA, Friedreich’s ataxia; HFG, hand-foot-genital; polyQ, polyglutamine; RAN, repeat-associated non-ATG; G4, G-quadruplex; DUE, DNA-unwinding element; DSB, double-strand break; RIM, repeat-induced mutagenesis; Pol, polymerase; ssDNA, single-stranded DNA; RPA, replication protein A; BIR, break-induced replication; SSA, single-strand annealing; EJ, end-joining; NHEJ, nonhomologous EJ; MEJ, microhomology-mediated EJ; MMR, mismatch repair; BER, base excision repair; TC-NER, transcription-coupled NER; CSB, Cockayne syndrome group B; IVF, in vitro fertilization; PGD, preimplantation genetic diagnosis.
Today, we know 13 different types of tandem repeats whose expansions cause various human diseases (Table 1). The majority of these repeat expansion diseases (REDS) are neither curable nor preventable at present. The most common cause of REDs is an expansion of CAG repeats (or complementary CTG repeats), which are responsible for 16 conditions, including HD and multiple spinocerebellar ataxias (SCAs). CGG repeat expansions cause six different conditions, including fragile X syndrome. Next, two disorders are caused by GAA repeat expansion. In addition to the expansions of these trinucleotide repeats, expansions of one tetranucleotide (CCTG), five pentanucleotide (ATTCT, TGGAA, TTTTA, TTTCA, and AAGGG), three hexanucleotide (GGCCTG, CTTTCT, and GGGGCC), and one dodecanucleotide (CCCCGCCCGCGG) repeat cause 13 other diseases. A separate class of REDs, so-called polyalanine (poly(A)) diseases, are caused by an in-frame expansion of imperfect GCN repeats. This expansion results in abnormally long stretches of alanine in the corresponding proteins. Altogether, nearly 50 REDs are currently known. Beyond these monogenic diseases, some specific repeat expansions might contribute to the etiology of various complex polygenic psychiatric and brain disorders, such as autism spectrum disorder (13), bipolar spectrum disorders, schizophrenia, and others (reviewed in Ref. 14). The number of known REDs will likely grow, as more than 100 human genes contain DNA repeats that are known to expand in some REDs. These repeats, therefore, can a priori expand, causing a disease (database of genes related to REDs RRID:SCR_018086).

The majority of REDs share two common features. First, the number of inherited repeats typically positively correlates with disease severity and negatively correlates with age of onset. Diseases with strong correlations between the number of repeats and age of onset are SCA7 (16, 17), SCA3 (18), SCA2 (19–21), SCA37 (22), HD (23–26), DM1 (27, 28), dentatorubral-pallidoluysian atrophy (DRPLA) (29, 30), X-linked dystonia parkinsonism (XDP) (31), familial adult myoclonic epilepsy 3 (FAME3) (32), and Friedreich’s ataxia (FRDA) (33, 34). For some disorders, the correlation between the number of repeats and symptom manifestation is less clear, although there typically is a marginally significant trend (35–39). Second, expansion of one disease-causing repeat does not promote expansions of other repeats in the patient’s genome. In other words, each RED patient typically has only a single repeat type expanded (40, 41).

Other characteristics differ among REDs. First, even though repeat expansion is necessary for some RED manifestation, there are REDs in which mutations other than repeat expansions can cause the same disease. Second, the RED mode of inheritance can be either autosomal recessive, dominant, or X-linked. Third, REDs can demonstrate a clear pattern of genetic anticipation (11, 23) or a complete lack thereof (34). Fourth, expandable repeats can be located in the coding part of a gene, in an intron, or in the 5′- or 3′-UTR. Fifth, the size of repeat expansion sufficient to cause pathological symptoms ranges from several repeat units for poly(A) diseases (42–53) to thousands for DM1 (11) and DM2 (54). Sixth, the origin of expanded alleles varies across repeats. For example, single common ancestors had a pathogenic amplification of ATTCT repeats in SCA10 (55–57), CCTG repeats in DM2 (58), and TTTTA/TTTCA repeats in FAME3 (32). Contrastingly, de novo repeat expansions were reported for the CAG repeat in HD (59), the CGG repeat in fragile X syndrome (60), and the GCN repeat in hand-foot-genital (HFG) syndrome (61). Last, the mechanisms of repeat toxicity vary among different repeats and include both a loss and gain of function.

The goal of this review is to depict the molecular mechanisms implicated in REDs. We first describe models of repeat-induced pathogenicity. We then discuss how repeat length instability and the propensity to induce various types of mutations compromises genome integrity. The focal point of the review is the description of the molecular mechanisms responsible for repeat instability and repeat-induced mutagenesis (RIM). Whereas these mechanisms were primarily established in model experimental systems, we discuss their relevance to human REDs whenever it is possible. Then we reflect on the existence of expandable repeats from the standpoint of genome function and evolvability. As for future directions, we discuss possible reasons for the differences in expandable repeats’ behavior in model systems versus human pedigrees as well as how we could use our current knowledge about characteristics and instability of expandable repeats to develop therapies for REDs.

From expanded DNA repeats to disease

Repeat expansions induce changes in cell metabolism that become detrimental to the function of specific tissues. What are these changes? First, an expanded repeat located within a gene may lead to a transcription defect and consequently a loss of function of the carrier gene. Such a loss-of-function mechanism predicts an autosomal recessive or X-linked mode of inheritance. It also predicts that in a subset of patients, an inactivation of the same gene could happen due to a missense, indel, or frameshift mutation, instead of repeat expansion. Both of these predictions are true for autosomal recessive FRDA (62), progressive myoclonus epilepsy of the Unverricht–Lundborg type (EPM1) (63), Congenital insensitivity to pain (44), X-linked Duchenne muscular dystrophy (64), and fragile X syndrome (65, 66).

The majority of REDs exhibit an autosomal dominant mode of inheritance and are caused exclusively by repeat expansions. Moreover, an expansion of a particular repeat may have the same detrimental consequences regardless of the gene in which this expansion happens. For example, expansion of (CAG)ₙ repeats in 13 different genes causes SCAs. Similarly, recently discovered expansions of (TTTCA)ₙ(CTTTA)ₙ in the introns of five different genes are all linked to FAME (32, 67–69). Therefore, researchers have proposed toxic gain-of-function mechanisms, which could be either at the RNA or protein levels.

Loss of function of a gene with an expanded repeat tract

Long repeat tracts can impede transcription of their corresponding gene (70) and therefore lead to decreased production of an essential protein (31, 62, 71). This process could occur in several ways. Formation of RNA-DNA hybrids (R-loops) potentially combined with formation of an unusual secondary
# JBC REVIEWS: Repeat-mediated genome instability

## Table 1
Currently known REDs

<table>
<thead>
<tr>
<th>Repeat unit</th>
<th>Gene and position of the repeat in the gene</th>
<th>Disease</th>
<th>Year of discovery</th>
<th>High-light paper</th>
<th>Inheritance pattern</th>
<th>Healthy range, no. of repeats</th>
<th>Symptomatic range, no. of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTCT</td>
<td>(ATXN10,) intron</td>
<td>Spinocerebellar ataxia 10 (SCA10)</td>
<td>2000</td>
<td>535</td>
<td>Autosomal dominant</td>
<td>10–22</td>
<td>500–4500</td>
</tr>
<tr>
<td></td>
<td>(DA)1, intron</td>
<td>Spinocerebellar ataxia 37 (SCA37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAG</td>
<td>(ATXN1,) coding sequence</td>
<td>Spinocerebellar ataxia1 (SCA1)</td>
<td>1993</td>
<td>536</td>
<td>Autosomal dominant</td>
<td>6–39</td>
<td>41–83</td>
</tr>
<tr>
<td></td>
<td>(ATXN2,) coding sequence</td>
<td>Spinocerebellar ataxia2 (SCA2)</td>
<td>1996</td>
<td>21, 537</td>
<td>Autosomal dominant</td>
<td>&lt;31</td>
<td>33–200</td>
</tr>
<tr>
<td></td>
<td>(ATXN3,) coding sequence</td>
<td>Spinocerebellar ataxia3 (SCA3)</td>
<td>1994</td>
<td>18</td>
<td>Autosomal dominant</td>
<td>&lt;44</td>
<td>52–86</td>
</tr>
<tr>
<td></td>
<td>(CACNA1A,) coding sequence</td>
<td>Spinocerebellar ataxia6 (SCA6)</td>
<td>1997</td>
<td>538</td>
<td>Autosomal dominant</td>
<td>&lt;18</td>
<td>20–33</td>
</tr>
<tr>
<td></td>
<td>(TPB,) coding sequence</td>
<td>Spinocerebellar ataxia 17 (SCA17)</td>
<td>2001</td>
<td>539</td>
<td>Autosomal dominant</td>
<td>25–44</td>
<td>47–63</td>
</tr>
<tr>
<td></td>
<td>(ATN1,) coding sequence</td>
<td>Dentatorubral-pallidolusian atrophy, Naito-Oyanagi disease (DRPLA)</td>
<td>1994</td>
<td>29, 30</td>
<td>Autosomal dominant</td>
<td>6–35</td>
<td>49–88</td>
</tr>
<tr>
<td></td>
<td>(HTT,) coding sequence</td>
<td>Huntington disease (HD)</td>
<td>1993</td>
<td>12</td>
<td>Autosomal dominant</td>
<td>9–29</td>
<td>36–121</td>
</tr>
<tr>
<td></td>
<td>(ATXN8,) coding sequence</td>
<td>Spinocerebellar ataxia8 (SCA8)</td>
<td>1999</td>
<td>114</td>
<td>Autosomal dominant</td>
<td>15–50</td>
<td>71–1300</td>
</tr>
<tr>
<td></td>
<td>(PPPR2RB2, ) 5’ UTR</td>
<td>Spinocerebellar ataxia12 (SCA12)</td>
<td>1999</td>
<td>540</td>
<td>Autosomal dominant</td>
<td>7–32</td>
<td>51–78</td>
</tr>
<tr>
<td></td>
<td>(ATXN7,) coding sequence</td>
<td>Spinocerebellar ataxia7 (SCA7)</td>
<td>1997</td>
<td>17</td>
<td>Autosomal dominant</td>
<td>7–17</td>
<td>38–130</td>
</tr>
<tr>
<td></td>
<td>(AR,) coding sequence</td>
<td>Spinal and bulbar muscular atrophy (SBMA)</td>
<td>1991</td>
<td>10</td>
<td>X-linked recessive</td>
<td>&lt;34</td>
<td>&gt;38</td>
</tr>
<tr>
<td></td>
<td>(GLS, ) 5’ UTR</td>
<td>Glutaminase deficiency (GD)</td>
<td>2019</td>
<td>470</td>
<td>Autosomal recessive</td>
<td>8–16</td>
<td>400–1500</td>
</tr>
<tr>
<td>CCCCGC-</td>
<td>(CSTB, ) 5’ UTR</td>
<td>Progressive myoclonus epilepsy of the Unverricht–Lundborg type (EPM1)</td>
<td>1997</td>
<td>541</td>
<td>Autosomal recessive</td>
<td>2–3</td>
<td>38–77</td>
</tr>
<tr>
<td>CCCCCG-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCCCCTT</td>
<td>(TAF1,) intron</td>
<td>X-linked dystonia parkinsonism (XDP)</td>
<td>2019</td>
<td>31</td>
<td>X-linked recessive</td>
<td>N/A</td>
<td>30–60</td>
</tr>
<tr>
<td>CCTG</td>
<td>(CNBP,) intron</td>
<td>Myotonic myostrophy2 (DM2)</td>
<td>2001</td>
<td>54</td>
<td>Autosomal dominant</td>
<td>&lt;30</td>
<td>75–11,000</td>
</tr>
<tr>
<td>CTG</td>
<td>(DM1, ) 3’ UTR</td>
<td>Myotonic myostrophy1 (DM1)</td>
<td>1992</td>
<td>11</td>
<td>Autosomal dominant</td>
<td>5–37</td>
<td>50–5,000</td>
</tr>
<tr>
<td></td>
<td>(PHH3, ) 3’ UTR</td>
<td>Huntington disease-like 2</td>
<td>2001</td>
<td>542</td>
<td>Autosomal dominant</td>
<td>6–28</td>
<td>&gt;41</td>
</tr>
<tr>
<td></td>
<td>(ATXN8,) coding sequence</td>
<td>Spinocerebellar ataxia8 (SCA8)</td>
<td>1999</td>
<td>114</td>
<td>Autosomal dominant</td>
<td>15–34</td>
<td>90–250</td>
</tr>
<tr>
<td></td>
<td>(TCF4,) intron</td>
<td>Fuchs endothelial corneal dystrophy</td>
<td>2012</td>
<td>543</td>
<td>Autosomal dominant</td>
<td>&lt;40</td>
<td>&gt;50</td>
</tr>
<tr>
<td>GAA</td>
<td>(DMD,) intron</td>
<td>Duchenne muscular dystrophy (DMD)</td>
<td>2016</td>
<td>64</td>
<td>X-linked recessive</td>
<td>11–33</td>
<td>59–82</td>
</tr>
<tr>
<td></td>
<td>(FXN,) intron</td>
<td>Friedreich ataxia 1 (FRDA)</td>
<td>1996</td>
<td>62</td>
<td>Autosomal recessive</td>
<td>5–30</td>
<td>&gt;70</td>
</tr>
<tr>
<td>CGG</td>
<td>(XLYT1, ) 5’ UTR</td>
<td>Baratela–Scott syndrome (BSS)</td>
<td>2019</td>
<td>544</td>
<td>Autosomal recessive</td>
<td>9–20</td>
<td>100–800</td>
</tr>
<tr>
<td></td>
<td>(EMR2, ) 5’ UTR</td>
<td>Mental retardation, X-linked, associated with fragile site FRAXE</td>
<td>1993</td>
<td>545</td>
<td>X-linked recessive</td>
<td>4–39</td>
<td>&gt;200</td>
</tr>
<tr>
<td></td>
<td>(DIP2B, ) 5’ UTR</td>
<td>Mental retardation, associated with fragile site FRA12A</td>
<td>2007</td>
<td>546</td>
<td>Autosomal dominant</td>
<td>12–26</td>
<td>&gt;150</td>
</tr>
<tr>
<td></td>
<td>(FMRI, ) 5’ UTR</td>
<td>Fragile X mental retardation syndrome</td>
<td>1991</td>
<td>5–9</td>
<td>X-linked dominant</td>
<td>6–82</td>
<td>231–2000</td>
</tr>
<tr>
<td></td>
<td>(CBL, ) 5’ UTR</td>
<td>Jacobsen syndrome</td>
<td>1998</td>
<td>547</td>
<td>Not inherited</td>
<td>11</td>
<td>&gt;100</td>
</tr>
<tr>
<td></td>
<td>(NOTCH2NL) coding sequence</td>
<td>Neuronal intranuclear inclusion disease (NIID)</td>
<td>2019</td>
<td>380</td>
<td>Autosomal dominant</td>
<td>13–30</td>
<td>60–959</td>
</tr>
<tr>
<td></td>
<td>(RUNX2,) coding sequence</td>
<td>Cleidocranial dysplasia</td>
<td>1997</td>
<td>42</td>
<td>Autosomal dominant</td>
<td>17</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>(SOX3,) coding sequence</td>
<td>Mental retardation, X-linked Congenital insensitivity to pain (CIP)</td>
<td>2002</td>
<td>43</td>
<td>X-linked recessive</td>
<td>11</td>
<td>15–26</td>
</tr>
<tr>
<td></td>
<td>(PRDM12,) coding sequence</td>
<td>Congenital insensitivity to pain (CIP)</td>
<td>2015</td>
<td>44</td>
<td>Autosomal recessive</td>
<td>12</td>
<td>18–19</td>
</tr>
<tr>
<td></td>
<td>(PARP1,) coding sequence</td>
<td>Oculopharyngeal muscular dystrophy</td>
<td>1998</td>
<td>45</td>
<td>Autosomal dominant</td>
<td>6</td>
<td>7–17</td>
</tr>
<tr>
<td></td>
<td>(HOXD13,) coding sequence</td>
<td>Synpolydactyly 1</td>
<td>1996</td>
<td>46, 47</td>
<td>Autosomal dominant</td>
<td>15</td>
<td>22–29</td>
</tr>
<tr>
<td></td>
<td>(HOXA13,) coding sequence</td>
<td>Hand-foot-genital (HFG) syndrome</td>
<td>2000</td>
<td>48</td>
<td>Autosomal dominant</td>
<td>18</td>
<td>24–26</td>
</tr>
<tr>
<td></td>
<td>(AR,) coding sequence</td>
<td>Epileptic encephalopathy, early infantile, 1</td>
<td>2007</td>
<td>49, 50</td>
<td>X-linked recessive</td>
<td>10–16</td>
<td>17–23</td>
</tr>
<tr>
<td></td>
<td>(PHOX2B,) coding sequence</td>
<td>Central hypoventilation syndrome, congenital</td>
<td>2003</td>
<td>51</td>
<td>Autosomal dominant</td>
<td>20</td>
<td>24–33</td>
</tr>
</tbody>
</table>

---

**Healthy range, no. of repeats** refers to the range of healthy individuals, while **Symptomatic range, no. of repeats** refers to the range of symptomatic individuals.

**Year of discovery** refers to the year the condition was discovered.

**High-light paper** refers to the specific paper where the condition was highlighted.

**Inheritance pattern** indicates whether the condition is inherited in a dominant, recessive, or X-linked manner.

**Repeat unit** indicates the specific repeat sequence associated with the condition.

**Gene and position of the repeat in the gene** specifies the exact location of the repeat within the gene.

**Disease** describes the condition associated with the repeat sequence.
structure within the expanded repeats may physically stall transcription (72). For example, in FRDA, RNA polymerase fails to successfully transcribe through expanded GAA repeats located in the intron of the FXN gene (62, 71), likely because of the formation of R-loop–containing structures (73, 74), such as an H-loop (75). This transcription hindrance leads to the appearance of repressive chromatin marks at and around an expanded repeat, ultimately leading to local heterochromatinization and gene repression (reviewed in Ref. 76) (Fig. 1). Reduced FXN expression causes mitochondrial dysfunction, which eventually leads to cell death (77, 78). Similarly, R-loop formation within expanded CGG (72, 79) repeats in the 5′-UTR of the FXN gene leads to promoter DNA methylation at the repeat and the adjacent promoter (80), followed by histone and DNA methylation, resulting in massive heterochromatinization that can spread up to 1.8 Mb (reviewed in Ref. 81) (Fig. 1). It is, in fact, this constitutive heterochromatinization that slows down DNA replication through this region, leading to the characteristic fragile X phenotype (82).

Expansion of the dodecamer CCCCCGCCCCCGG repeat in the 5′-UTR of the CSTB gene seems to decrease CSTB gene expression by a mechanism different from that discussed above. This repeat is located between two cis-regulatory elements that activate CSTB transcription. Expansion of the repeat increases the distance between the two regulatory elements such that it prevents efficient transcription initiation of the CSTB gene (63).

Toxic gain of function at the level of RNA

Growing evidence suggests that some expanded repeats exhibit cellular toxicity on their own, rather than in a context of their carrier gene. One example is the expansion of CCTG repeats in an intron of the CNBP gene (in DM2). This expansion does not directly affect gene expression: it changes neither the methylation pattern of the gene promoter (83) nor the splicing pattern of CNBP mRNA nor the CNBP protein level (84).

Instead, expanded CCTG repeats are transcribed into toxic repetitive RNA. The toxicity of expanded RNA repeats is likely associated with their ability to fold into unusual RNA secondary structures. The list of known repetitive RNAs with confirmed toxicity includes CUG, CCUG, CGG, CAG, AUUCU, UGGAA, AUUUC, GGGCCUG, and GGGGCCC repeats (22, 85–94).

What are the molecular mechanisms of this toxicity? Expanded repetitive RNAs can engage in multivalent base pairing, which leads to their gelation and aggregation into visible nuclear foci (95). As a consequence, these RNAs remain in the nucleus and might sequester RNA-binding proteins. In the best-studied cases of DM1 and DM2 diseases, expanded RNA repeats sequester Muscleblind (Mbnl) proteins, whose normal function is to regulate splicing of muscle-specific genes (96) (Fig. 2). This mechanism is evidenced by the fact that disruption of the MBNL1 gene leads to the same symptoms as the expression of expanded repeats alone (97). For other repeats, the precise toxicity mechanism is less understood. Expression of r(GGGGCCC)exp causes length-dependent cognitive impairment in mice (98), but the mechanistic connection between RNA expression and symptoms remains unclear. This repetitive RNA folds into G-quadruplexes (99), forms nuclear foci (98, 100), sequesters important RNA-binding proteins (101–104), and impairs mRNA transport (105) likely by damaging the nuclear pore complex (106). However, it is not yet clear which of these features of the r(GGGGCCC)exp repeats is primarily responsible for toxicity.

Toxic gain of function at the level of protein

A group of REDs are caused by expansions of GCN or CAG repeats located in coding parts of various genes. Translation of these in-frame repeats results in abnormally long poly(A) or polyglutamine poly(Q) tracts in their corresponding proteins. These long tracts could abrogate function of a protein and/or exhibit toxic gain of function.
Poly(A) tracts are common protein motifs that were suggested to promote protein interactions (107) or serve as nuclear export signals (108). At the molecular level, poly(A) peptides undergo a conformational change as their length increases. Whereas short poly(A) tracts preferably form monomeric \( \beta \)-helicies, longer tracts fold into polymeric \( \beta \)-sheets (reviewed in Ref. 109) and into coiled coils (110). These conformational changes can lead to a toxic gain of function by introducing novel types of protein interactions, aggregation, mislocalization, and sequestration of other essential proteins. Commonly, proteins with poly(A) expansions co-aggregate with their WT versions, leading to effective haploinsufficiency, which explains their dominant inheritance patterns (reviewed in Ref. 109).

Similar to expanded poly(A) tracts, long poly(Q) tracts misfold into \( \beta \)-sheet structures that aggregate into toxic inclusion bodies in neurons. These structures might act as so-called “polar zippers,” which exhibit nonspecific affinity to various regulatory proteins in a cell (111). Accumulation of poly(Q) runs is believed to cause neurodegeneration (reviewed in Ref. 112). First discovered in 1997 for CAG repeats in HD (113), poly(Q) accumulation was used to explain symptoms of SCAs and other diseases caused by expansions of translated CAG repeats. However, there was one mysterious exception: SCA8. Unlike other SCAs, SCA8 is caused by an expansion of transcribed CTG rather than CAG repeats. Nonetheless, its symptoms resemble other SCAs (114). This conundrum was partially
resolved after the discovery of bidirectional transcription of the SCA8 locus and subsequent translation of antisense transcripts (115). Yet one question remained unanswered: how can CAG repeats from antisense transcripts be translated without a start codon?

The discovery of an unusual phenomenon called repeat-associated non-ATG (RAN) translation shed some light on this mystery (116). RAN translation requires neither the canonical nor an alternative start codon for initiation and results in production of repetitive proteins in all possible reading frames. Taking into consideration bidirectional transcription, one expanded repeat may produce up to six repetitive polypeptides. To illustrate, RAN translation of the GGGGCC repeat sense transcript, r(GGGGCC)_{exp} results in poly(GA), poly(GP), and poly(AR) peptides, all of which were detected in ALS/FTD patient–derived cells (117). The antisense transcription of the same repeat leads to accumulation of polyGP, polyAP, and polyPR peptides (118, 119). These polypeptides aggregate into high-molecular weight insoluble clusters (120) and seem to be neurotoxic (98, 121), although it is not yet clear to what extent different polypeptides contribute to overall toxicity (122).

Length-dependent RAN translation has been reported for CAG, CGG, GGGGCC, GCUCUG, and TGGAA repeats (reviewed in Refs. 123–125). Nonetheless, we still lack a good understanding of RAN’s mechanistic details. One hypothesis is that formation of unusual secondary RNA structures, such as hairpins (114) or G-quadruplexes (G4s) (126), might somehow recruit ribosomes and trigger this peculiar mode of translation (Fig. 2).

Taken together, mechanisms by which expanded repeats lead to cellular toxicity include loss of function of a protein and a toxic gain of function on the RNA or protein levels. These mechanisms might take place in combination with each other. Unfortunately, for many REDs, there is still no consensus on what mechanism plays a major role in disease progression. For example, accumulation of toxic poly(Q) proteins was historically used to explain HD’s phenotype. However, evidence accumulated in the last decade suggests that gain of function at the RNA level might also contribute to disease manifestation. This is supported by experiments in model systems where CAG repeats interrupted by the CAA codon express decreased cellular toxicity when compared with pure CAG repeat tracts (reviewed in Ref. 127). This is unexpected because both CAA and CAG codons code for glutamine. As such, the length of transcribed poly(Q) tracts does not depend on the presence of a CAA interruption. Additionally, it was recently documented that HD age of onset is better predicted by the length of uninterrupted CAG repeat tracts rather than by the sheer number of consecutive glutamines in the Htt protein (128–130). However, there exists an alternative explanation of the same phenomenon: the disease age of onset might depend on somatic expansions of CAG repeats during the patient’s lifetime, a process that requires repeat integrity (see below).

Overall, due to technical obstacles, it is highly challenging to decipher the precise role of various toxicity mechanisms in a specific RED (125). Therefore, we have a good understanding of possible mechanisms of expanded repeats’ toxicity, even though the precise mechanisms are established for only few REDs.

Dynamic DNA structures as the key to repeat instability

Tandem repeats are abundant in the human genome (131), particularly in centromeres and telomeres, as well as in regulatory regions (132). High variability of tandem repeats between different individuals (133) makes analysis of tandem repeat polymorphisms useful in forensics and paternity testing (134). Nonetheless, most of the known tandem repeats do not expand as dramatically as disease-causing repeats, which may gain up to thousands of repeat units in affected individuals (11, 54).

What are the main features that make certain repeats prone to large-scale expansions resulting in REDs? Numerous studies that we review below attempted to resolve this fundamental question. Initially, the DNA slippage model was proposed to explain instability of tandem repeats. It postulated that repeats may be lost or gained during local misalignment of DNA during replication or repair (135). Indeed, formation of slipped DNA structures followed by repeat unit gains was detected in in vitro replication experiments (136). These structures were also observed in DM1 patient samples (137).

The DNA slippage model is beautiful in its simplicity and sufficient to explain the small-scale instability of any tandem repeat. However, it does not easily address several questions. First, why do some repeats gain thousands of repetitive units and others do not? Second, why does the instability of expandable repeats increase exponentially but not linearly with the size of the repeat? And last, why do repeat interruptions dramatically stabilize expanded repeat tracts?

Relationship between repeat size and instability

Length-dependent instability is one of the most universal properties of expandable repeats. What is the nature of this relationship? The DNA slippage model predicts a linear correlation between a repeat’s length and its propensity to expand or contract: the number of repetitive units should reflect the number of opportunities for local DNA strand misalignments. However, the expandable repeats’ behavior is inconsistent with this prediction.

First, for the majority of expandable repeats, there exists a repeat number threshold length after which the repeat starts to exhibit dramatic intergenerational and somatic instability. Curiously, this threshold length is comparable with the average Okazaki fragment size in eukaryotes (138). Moreover, disease-causing repeats show a nonlinear relationship between the repeat length and instability in experimental systems. Depending on an experimental system, doubling the repeat tract length results in a 6.5-fold increase in instability for GGGGCC repeats (139), up to a 100-fold increase for CGG repeats (140, 141), and up to a 1000-fold increase for GAA repeats (142). Together, data from both patient samples and experimental systems demonstrate that there is a qualitative difference between the instability of long and short repeats. This difference is evident from the differential propensity of repeats to expand or contract, the scale of expansions and contractions, and, likely, mechanisms of expansions and contractions (reviewed in Ref. 143).

Role of repeat interruptions in repeat stability

The DNA slippage model predicts that a small number of repeat interruptions should not dramatically change the overall
[Image 74x585 to 538x733]

Figure 3. Types of dynamic DNA structures and repeat sequences that form them. *, complement of CCTG; **, complement of CCCTCT.

instability of a repeat. However, there appears to be a causal connection between repeat purity and length. Analysis of DNA samples derived from RED patients revealed that the expanded repeat tracts commonly lose repeat interruptions found in their ancestral alleles (also called long normal alleles). For example, (CGG)$_{20–52}$ repeat alleles located in the ancestral alleles (also called long normal alleles). For example, (CGG)$_{20–52}$ repeat alleles located in the FMR1 promoter are typically randomly interrupted by AGG triplets, whereas expanded alleles consist of almost pure CGG repeats (144). The presence of AGG interruptions stabilizes both normal and premutational CGG repeat alleles during intergenerational transmission (145–150) and in experimental systems (151). Overall, it seems like the total length of a pure uninterrupted CGG tract rather than the repeat length overall predicts the level of CGG repeat instability. In other words, a loss of AGG interruption likely predisposes CGG repeat tracts to large-scale expansions (149, 152). Generally, the stabilizing role of interruptions is common for expandable repeats and has been reported for GAA (153, 154), CAG (38, 128, 129, 155–162), CCTG (54, 58, 163–165), and other repeats in somatic tissues, intergenerational transmission, and experimental systems.

Taken together, it appears that DNA slippage cannot be solely responsible for the repeat instability observed in REDs. On the contrary, more complex mechanisms, likely involving formation of alternative DNA structures, should be involved. It is generally believed that formation of such structures allows certain repetitive sequences to escape faithful DNA replication or repair and to undergo large-scale expansion. These structures include DNA triplexes (also known as H-DNA), G4s, and imperfect hairpins. Some AT-rich sequences can also undergo major DNA unwinding; hence, they are called DNA-unwinding elements (DUEs) (Fig. 3).

**H-DNA**

As discovered in 1987, homopurine/homopyrimidine mirror repeats can adopt an intramolecular triple-helical DNA structure called H-DNA. In H-DNA, one of the DNA strands corresponding to one-half of the mirror repeat folds back, forming a triple helix with the remaining duplex part of the mirror repeat. The resulting triple helix constitutes a stack of base triads, which are stabilized by Hoogsteen or reversed Hoogsteen base pairing within the triplets and by $\pi-\pi$ interactions between the stacks (Fig. 3). Two types of triplexes exist (166). First, a pyrimidine triplex, or H-y, consists of TA-T and protonated CG-C$^+$ triads and requires a mild acidic environment to form (167). Second, a purine triplex, or H-r, consists of CG-G and TA-A triads and is stable at a physiological pH in the presence of bivalent cations (168).

H-DNA formation is favored in negatively supercoiled DNA, as it is topologically equivalent to DNA unwinding (169). Triplex-forming motifs are grossly overrepresented in eukaryotic genomes (170), which might point to their evolutionarily conserved role in genome function. On the other hand, formation of triplexes can pose a threat to a genome: triplexes impede DNA replication (171, 172) and transcription (73, 173, 174), promote formation of double-strand breaks (DSBs) (175), and trigger nucleotide excision repair (NER) (176). Three perfect homopurine-homopyrimidine repeats are known in REDs: GAA, AAGGG, and CCCTCT. GAA repeats are the best studied of the three and were reported to form both H-y and H-r triplexes in vitro and in vivo (177–182). Importantly, homopurine/homopyrimidine repeats that lack mirror symmetry (and therefore are unlikely to form stable triplexes) are dramatically more stable than (GAA)$_n$ repeats (183). This indicates that formation of a triplex might be central to repeat instability.

**G4-DNA**

One year after the discovery of H-DNA, a G4 structure was described for a GC-rich sequence from the immunoglobulin switch region (184). G4 is a four-stranded DNA structure, which consists of several G-quartets stacked upon each other and held together by $\pi-\pi$ interactions. Each quartet is formed by four guanines connected by Hoogsteen hydrogen bonds (Fig. 3). Monovalent cations, specifically sodium and potassium, stabilize the G-quartets and their stacking interaction. The consensus motif is typically represented as $G_4N_1G_4N_1G_4N_1G_4N_1G_4$, although G4-DNA with loops longer than 7 bp, with two G-tetrads, or with a bulged nonguanine base inside the three guanine tracts also is possible (reviewed in Ref. 185). G4 can exist in a large variety of isoforms differing in the number of DNA strands (one, two, or four) and relative strand polarity. In the context of this review, an intramolecular G4 structure is of prime relevance and is presented in Fig. 3.

G4 motifs are highly abundant in eukaryotic and some prokaryotic genomes (186). G4s were detected in vivo using G4-specific antibodies and small molecules. They are found in telomeres and multiple other regions in the genome (reviewed
in Ref. 186) and are enriched in promoters and 5′-UTR sequences of genes. It is believed that G4s are likely to form in nucleosome-depleted, highly transcribed regions (187, 188). G4s interfere with regulation of transcription, genome integrity, and telomere maintenance (reviewed in Refs. 186 and 189–191). Importantly, they can constitute potent barriers for replication (reviewed in Ref. 192) and, in certain circumstances, promote DSB formation (193–195). Because these structures are highly stable at physiological conditions, a number of helicases, including Pif1, Rtel1, FANCJ, Bloom (BLM), and Werner (WRN), have evolved to unfold them (reviewed in Ref. 191). Additionally, various proteins, for example topoisomerase I (196–198) or nucleolin (199), can bind to and/or interact with G4 motifs (reviewed in Ref. 200).

Based on our knowledge of G-quadruplex structures and in vitro data, expandable repeats that should be able to readily adopt G4-DNA are GGGGCC (201–203), CCCCGCCCCGCG, AAGGG, and AGAGGG—the complement of CCCTCT. Other repeats, such as CGG (204), form a peculiar version of G4 composed of G-quartets separated by hydrogen-bonded cytosines. CAGG (the complement of CCTG) and GG CCTG repeats also have a potential to form a weak G4, in theory (Fig. 3).

**Imperfect hairpins**

An inverted DNA repeat may form a hairpin: a DNA secondary structure in which an upstream region anneals to a complementary downstream region via normal Watson–Crick base pairing. When hairpins form on either side of a double helix directly opposite to each other, the overall structure is called DNA cruciform. If two opposite hairpins are shifted relative to each other, the resulting structure is called slipped DNA or S-DNA (Fig. 4A). No expandable repeats can form a perfect hairpin or a cruciform. That said, many expandable repeats, such as CAG (205–207), CGG (204, 207), CCTG (208), GGGGCC (203), CCCCCCCCCGCG (209), and TGGAA (210), form hairpins containing mismatches—imperfect hairpins (Fig. 4A)—as was first demonstrated in a groundbreaking paper from the McMurray laboratory (207). Based on its sequence, another repeat, GGCCTG, also theoretically adopts an imperfect hairpin. Similar to perfect hairpins, complementary imperfect hairpins can also form an S-DNA (Fig. 4A).

Both DNA cruciform and S-DNA are topologically equivalent to completely unwound DNA and, thus, favored by negative DNA supercoiling. There is a principal difference, however, in their formation kinetics. In the case of a DNA cruciform, the first step is slow: it involves the unwinding of a 10-bp DNA segment around the pseudosymmetry axis of an inverted repeat to form a central bubble. Self-annealing of separated DNA strands in this bubble results in cruciform nucleation. The nucleus then quickly extrudes into the full-size cruciform via branch migration. Upon DNA linearization, DNA branch migration would instantly convert the cruciform back to a regular DNA duplex (211) (Fig. 4B).

The kinetics of S-DNA formation is different, owing to the fact that the presence of mismatches completely blocks spontaneous branch migration, thus precluding the extrusion step (212). Consequently, S-DNA formation must be preceded by the unwinding of a long DNA segment followed by the out-of-register self-annealing of individual DNA strands. In vitro, this could be achieved upon denaturing and renaturing of repetitive

---

**Figure 4.** A, DNA cruciform versus S-DNA structures. B, difference in the kinetics of perfect versus imperfect hairpin formation (see “Imperfect hairpins” in the “Dynamic DNA structures as the key to repeat instability” section).
DNA segments. In vivo, this could happen during DNA replication, transcription, or repair, all of which involve unwinding of duplex DNA. After an S-DNA is formed, the same kinetic principles apply in reverse: the S-DNA cannot easily convert into duplex DNA via branch migration. In other words, such S-DNA is kinetically trapped and, thus, relatively stable even though this conformation is thermodynamically unfavorable (Fig. 4B).

**DUEs**

Highly AT-rich repeats ATTCT, TTTTA, and TTTCA are similar in their sequence composition to DUEs, which are involved in DNA unwinding at replication origins. Indeed, the DNA-unwinding capacity of the ATTCT repeat was confirmed experimentally. In supercoiled plasmids, this repeat is highly unwound or even completely unpaired, thus being accessible to small molecules, oligonucleotides, and DNA polymerase (213). In fact, long ATTCT can function as aberrant replication origins in human cells (214), which might contribute to their instability.

Taken together, all expandable DNA repeats can form dynamic DNA structures during biological transactions that involve either extensive DNA unwinding or accumulation of negative supercoiling (169, 215). Surprisingly, mechanisms of repeat instability are not the same for different repeats. These mechanisms depend on the repeat sequence, length, purity, and location within a genome, as well as cell type, developmental stage, and replicative and transcriptional status. Rephrasing Leo Tolstoy: stable DNA sequences are all alike; every unstable DNA sequence is unstable in its own way. In the next section, we review how the presence of expandable repeats affects genome integrity or, in other words, types of repeat-mediated genome instability.

**Types of repeat-mediated genome instability**

**Length instability during intergenerational transmission**

Expandable repeats change their length dramatically while being transmitted between generations. Some of them preferentially expand, which leads to the genetic anticipation phenomenon described above. Contrary to initial beliefs, however, most REDs do not exhibit genetic anticipation, and the repeat length changes in both directions between generations.

Remarkably, parental gender is the most significant determinant of the direction of intergenerational instability. For example, there is a strong bias for expansions of GAA (FRDA), CGG (fragile X), or CCCTCT (XDP) repeats during maternal transmission, whereas paternal transmission typically results in a contraction of the same repeats (31, 34, 140, 216–219). The opposite pattern is observed for the majority of diseases caused by CAG repeat expansions: repeats are stable or slightly biased toward contractions during maternal transmissions, whereas paternal transmissions are biased toward expansions (23, 24, 30, 159, 220–225). Note, however, that CAG repeats in SCA8 show the opposite tendency (114, 226), whereas CAG repeats in SCA2 (20) and SCA7 (16, 17) lack an apparent parental gender bias. Another interesting example is the ATTCT repeat: this repeat expands or contracts dramatically during paternal transmission but stays virtually unchanged during maternal transmission (227, 228).

The second important factor is parental age. Intergenerational instability of GAA (FRDA) (216), CTG (DM1) (229), and CAG (SCA17 (225) and SCA2 (224)) repeats directly correlates with parental age. DRPLA, SCA1, and DM1 mice also show age-dependent intergenerational CAG repeat instability (230–232). Having said that, age dependence is not a universal rule: for example, intergenerational instability of CAG repeats in HD depends neither on the mother’s nor the father’s age both in human (221) and mice (233) pedigrees.

Why does intergenerational repeat instability depend on parental gender and age? The increase of instability with age indicates that repeat expansions and contractions might take place before fertilization. This may simply be because a longer lifetime allows more chances for a repeat to contract or expand. Why repeat instability also depends on parental gender is a more difficult question for which we do not yet have a full answer. Several hypotheses that incorporate the differences between male and female germ cells were proposed to explain this phenomenon. First, the differential bias for expansions and contractions could be due to the differences in duration between spermatogenesis and oogenesis (230, 232, 234, 235). Second, the counterselection for repeat expansions could differ between sperm and oocytes (16, 218). For example, large-scale expansions of CGG repeats in the *FMR1* gene seem to be counterselected in sperm (218, 236), possibly because *FMR1* loss of function reduces sperm mobility, at least in *Drosophila* (237). Third, the difference in the expression level of certain DNA repair proteins such as Msh3 during spermatogenesis and oogenesis might explain different levels of repeat instability (238).

We hypothesize that two other mechanisms might account for the difference in repeat instability during spermatogenesis or oogenesis. First, a differential pattern of origin firing and replication timing might affect repeats’ propensity to expand or contract (as will be discussed below). Second, the fate of chromatin, which is dramatically different between the two gamete types, could differentially affect repeat instability. Mature oocyte chromatin is very loose, to allow rapid production of RNAs and proteins required for successful fertilization and development into an embryo (reviewed in Ref. 239). In contrast, mature sperm chromatin is compact, owing to the temporary replacement of core histones with sperm-specific histone variants, followed by transition proteins and, finally, protamines (reviewed in Refs. 239 and 240).

**Length instability in somatic tissues**

The majority of REDs are not congenital; instead, they have a fairly late age of onset. It was initially thought that this is because repeat-associated toxicity is low and results in a slow, gradual degradation of patients’ tissues that becomes symptomatic later in life (241, 242). Although certainly attractive, this hypothesis fails to explain two groups of data. First, it does not explain why the age of onset of dominant HD is well-predicted by the length of the longer repeat expansion allele in homozgyous patients (243, 244) or why the age of onset of recessive FRDA is determined solely by the shorter GAA expansion allele.
(245–247). Second, it fails to explain why the age of onset of HD is better predicted by the length of noninterrupted CAG repeats in DNA rather than by the length of the polyQ tract in the protein (128–130).

The currently accepted view is that the age of onset and disease progression are governed by a combination of two different mechanisms (243). This hypothesis states that the number of repeats in inherited alleles is typically insufficient to exhibit significant cellular toxicity. However, repeats expand in some somatic tissues (16, 17, 27, 31, 32, 35, 36, 54, 80, 164, 248–251), which causes cell toxicity when a repeat length exceeds a disease-specific threshold. Subsequent disease progression is determined by the mechanism of toxicity, which is specific for each disease (252). This model provides an elegant explanation for the dependence of age of onset on the number of inherited repeats in both heterozygous and homozygous patients.

The level of somatic instability is tissue-specific (253, 254), likely because of differential expression of DNA replication and repair genes across various cell types (255). Somatic instability also depends on the type and length of the repeat (256) and tends to increase with age (54, 257, 258). Somatic instability can be so dramatic that it may preclude precise estimation of the ease-specific threshold. Subsequent disease progression is determined by the mechanism of toxicity, which is specific for each disease (252). This model provides an elegant explanation for the dependence of age of onset on the number of inherited repeats in both heterozygous and homozygous patients.

Role of DNA replication in repeat instability

Duplicating long structure-forming repetitive sequences presents a significant challenge for the replication machinery. This phenomenon is well-established and has been observed in virtually all experimental systems. In this section, we review how replication contributes to repeat instability.

Evidence that replication promotes repeat instability

In experiments with mammalian cells, long GAA (279) or CAG (280) repeats cloned into plasmids expand and contract more frequently if the plasmids are able to replicate upon transfection. This indicated that replication per se promotes repeat instability.

DNA replication is inherently asymmetrical. Whereas the nascent leading strand is synthesized more or less continuously in the direction of replication, the nascent lagging strand is replicated in the opposite direction and in short patches known as Okazaki fragments. Consequently, a floating zone in the lagging strand template called the Okazaki initiation zone stays single-stranded, facilitating the formation of DNA secondary structures. There are also differences in DNA polymerases that synthesize the two DNA strands. The leading strand DNA polymerase ε (Pol ε) is a highly processive and precise enzyme. The lagging strand DNA polymerase δ (Pol δ) is less processive, and its fidelity is an order of magnitude lower than that of Pol ε, which is compensated by a higher activity of mismatch repair (MMR) on the lagging strand. Unlike Pol ε, Pol δ has strand displacement activity, which is needed for Okazaki fragment processing (reviewed in Ref. 281). To additionally complicate the picture, it was recently demonstrated that leading strand...
DNA synthesis is, in fact, initiated by Pol δ followed by a switch to Pol ε (282).

Not surprisingly therefore, the asymmetry of the replication fork leads to orientation-dependent repeat instability. Indeed, CAG/CTG repeats tend to contract if the CTG tract is in the lagging strand template and tend to expand if the CAG tract is in the lagging strand template in bacterial (35, 283), yeast (161, 162, 284, 285), and mammalian cells (206, 280, 286). Typically, this orientation dependence is explained by the fact that an imperfect hairpin formed by a (CTG) tract is more stable than the one formed by a (CAG) tract. Thus, when a (CTG)n tract is located in the lagging strand template, DNA polymerase might somehow jump through a hairpin formed by this repeat, resulting in a contraction. Alternatively, when the (CTG)n tract is in the nascent DNA strand during lagging strand replication, it can form slipped hairpins that, if unrepaired, would lead to expansions (reviewed in Ref. 287). Similarly to CAG repeats, the CGG and GGGGCC repeats form imperfect hairpins, and the hairpins formed by the G-rich strands are more stable (287). In addition, the G-rich strand of these two repeats may fold into a G4 (201–204). Consistent with the formation of a hairpin or a G4 during lagging strand synthesis, several studies have found that when located in the lagging strand template, CGG (141, 151, 288, 289) and CCGGGG (139) tracts are prone to repeat contraction. GAA repeats do not fold into a hairpin or a G4 and instead form an H-DNA. They also undergo orientation-dependent instability. In a majority of studies, GAA repeats are more unstable when (GAA)n tracts are located in the lagging strand template (73, 142, 153, 290), although we recently found that there is almost no difference in the repeat contraction rate among the two orientations (183).

Altogether, the dependence of repeat instability on the orientation relative to the origin of replication led us to propose the “ori-switch” model. It postulates that inactivation of a replication origin on one side of a repeat and the subsequent switch to the origin located on its opposite side changes the pattern of repeat instability, predisposing them to either expansions or contractions (292).

Repeat instability also strongly depends on the distance between the repeat and the nearest origin of replication (206, 208, 214, 284, 286, 293). The mechanisms of this dependence are not well-understood. Our “ori-shift” model (292) posits that this may be caused by the differences in repeat position within the Okazaki initiation zone. Given recent observations (282), this may also be caused by the repeat position relative to the location where Pol δ switches to Pol ε during leading strand synthesis.

Overall, it seems highly likely that a repeat’s position within the replication fork plays a big role in repeat-induced pathogenesis in humans. First, the “ori-switch” and “ori-shift” models explain why the same repeat tract exhibits dramatically different patterns of instability in two different locations in the genome (250, 279). Second, it could explain time- and tissue-specific differences in repeat instability. For example, the fragile X CGG repeat is believed to expand and contract during gametogenesis and early embryogenesis but stays relatively stable later in a person’s life (60, 272, 294–297). Tissue-specific differences in origin-firing patterns might explain why CGG instability is limited to such a narrow developmental window. It is possible that during embryonal development, the CGG repeat is in an unstable position relative to the active origin of replication. During later cell division, a different origin might be utilized, thus fixing repeat length (298, 299). Additionally, expanded CGG repeats may themselves influence activation of the near origins and, as such, put themselves in an unstable position (82). This hypothesis explains the existence of a length threshold required for CGG repeat instability.

Role of Okazaki fragment processing in repeat instability

Another piece of evidence for the role of replication in repeat instability comes from the observation that impaired Okazaki fragment processing destabilizes various types of repeats, at least in yeast. During lagging strand replication, Pol δ partially displaces the 5′-end of a preceding Okazaki fragment while replicating the succeeding fragment. Specific 5′ flap endonucleases then process the flap, which is followed by the ligation of two fragments. In yeast, two endonucleases do the bulk of the job: Rad27, which only cuts short flaps (300–302), and Dna2, which cuts longer flaps (301–306). Mutations in either of these nucleases lead to dramatic increases in the expansions of various repeats, including CAG, GAA, CGG, G, GT, CAGT, CAACG, CAATCGGT, yeast telomeric repeats, and several minisatellites (142, 289, 307–318). In other words, the effect of rad27 or dna2 mutations is virtually independent of the repeat sequence. The classic interpretation of these data involves the incorporation of a long repetitive flap into the nascent strand during DNA replication, resulting in an elongated repeat tract (319) (Fig. 5A). However, the same mutations also increase repeat contraction rates (183, 289, 307, 309, 310, 312–314, 316–318), a phenomenon that cannot be explained by the flap incorporation model. A more general model suggests that in the absence of fully functional Rad27 or Dna2, the genome-wide accumulation of long ssDNA species (314, 320, 321), such as flaps or gaps, titrates ssDNA-binding protein replication protein A (RPA) from the replication regions (322). Due to the lack of RPA, transiently formed ssDNA within the repetitive tracts is likely to equilibrate into DNA secondary structures, which trigger repeat instability (Fig. 5B) (183).

Are these models applicable to humans? The yeast Rad27 and human Fen1 proteins are highly conserved, such that Fen1 overexpression rescues the rad27Δ phenotype in yeast (323). However, despite these similarities, attempts to find mutations in the FEN1 gene that modify the phenotype of CAG repeat expansion in HD patients were fruitless (324–327). Surely, failure to identify such modifiers could potentially be explained by the high toxicity or lethality of FEN1 mutations. However, experiments with siRNA knockdown of Fen1 in a cell line system (328–330) and with Fen1-deficient mice generally do not detect a role for Fen1 in CAG repeat stability (331), except for one study where the nuclease activity of Fen1 was found to counteract CAG repeat instability, although the effect was not as dramatic as in yeast (332). Additionally, knockdown of Fen1 in mammalian cells does not promote GAA repeat instability.

Therefore, there could exist fundamental differences between

3 J. Cebrian, unpublished results.
yeast and humans, either in the Okazaki fragment processing pathways or in the regulation of RPA homeostasis.

Direct evidence that expandable repeats are hard to replicate

Contemporary techniques, such as quantification of nascent strand abundance, two-dimensional gel electrophoresis, EM, and single-molecule visualization of replicating DNA allow researchers to directly measure replication fork progression. These techniques unconditionally proved that expanded tracts of CAG, GAA, CGG, and GGGGCC repeats physically stall the replication fork in a length-dependent manner in every experimental system studied, including bacterial, yeast, and mammalian cells. Notably, the effect is most pronounced for the CGG and GGGGCC repeats (73, 139, 290, 333–341). How do expanded repeats stall the replication fork? First, replication through a long repetitive sequence might exhaust the local dNTP pool, which slows down the replisome (342). Second, and even more important, DNA sequences prone to form alternative DNA structures block replication both in vivo and in vitro, particularly when their purine-rich strands are in the lagging template (171, 343). Third, these alternative DNA structures may recruit various proteins, which could create “protein bumps” that impede replisome progression (335).

What happens after the replication fork has stalled depends on the repeat sequence, replication mode, and other cellular factors. Consequently, the amount of fork stall may (290) or may not reflect the amount of repeat instability (139, 335, 340). In the best-case scenario, the replication fork can temporarily slow down but quickly recover and continue DNA replication as if nothing had happened. However, a longer pause may lead to the formation of single-stranded gaps, which might result in the accumulation of mutations around the repeat tract (Fig. 6). A prolonged stall can, in turn, facilitate DNA slippage or template switching, both of which can lead to repeat length instability. During template switching, a stalled replicative polymerase, either Pol δ or Pol ε, temporarily switches its template and replicates from the nascent strand to bypass a lesion. If a polymerase stalls at a repetitive template, it might invade the opposite strand out of register and thus end up synthesizing the wrong number of repeats (reviewed in Ref. 344) (Fig. 6). Data from our laboratory revealed that this process could be responsible for GAA repeat expansion in yeast (142).

A dramatic consequence resulting from the replication fork colliding with a potent barrier is fork reversal (345). Fork reversal is a process in which two nascent strands anneal to each other, whereas their template strands reanneal back (Fig. 6), resulting in the formation of the so-called “chicken foot structure” containing a four-way junction. The biological role of fork reversal is to allow a bypass of a DNA lesion in a template strand using an unperturbed nascent strand as a temporal template (reviewed in Ref. 345). Formation of dynamic DNA structures could serve as such a lesion, as reversed forks were detected in vivo for CAG (346) and GAA repeats (338). Furthermore, formation of a reversed fork could promote repeat instability. Indeed, transition of a normal replication fork into a chicken

Figure 5. Two models depicting the potential role of Rad27 flap endonuclease in the instability of structure-forming repeats. A, flap ligation model. In the absence of Rad27, long flaps that formed during Okazaki maturation might become incorporated into the nascent DNA during lagging strand synthesis, resulting in repeat expansion. B, RPA depletion model. In the absence of Rad27, long flaps that accumulate genome-wide during Okazaki fragment maturation titrate RPA out from the repetitive DNA. This promotes formation of DNA secondary structures (the hairpin is shown) within the repeat and can result in repeat contraction or expansion.
foot structure involves several strand-reannealing steps and as such provides multiple opportunities for potential misalignment of tandem repeats. Furthermore, chicken foot structures are substrates for structure-specific nucleases, which can transform reversed forks into one-ended DSBs. These DSBs must be repaired through break-induced replication (BIR), a conservative mode of DNA replication, which leaves its mutagenic trace as much as several kilobases away from the initial break site (reviewed in Ref. 347) (Fig. 6).

If the replication fork cannot recover from a stall, it leaves unreplicated ssDNA regions that can be broken into a two-ended DSB (Fig. 6), which would need to be repaired via homologous replication (HR). However, mitotic cell division may start before a DSB is formed. This is common for late replicating fragile sites, like the CGG repeat in the FMR1 promoter (348). In this scenario, ssDNA regions transform into anaphase bridges, which may lead a permanent loss of a part of or even a whole chromosome (261, 341).

How to maintain hard-to-replicate expandable repeats?

Cells have evolved a number of mechanisms to minimize fork stalling and ensure, as much as possible, smooth progression of the replication fork through repetitive DNA. Replication fork-stabilizing factors, Tof1 and Mrc1, prevent CAG repeat instability and fragility of CAG, ATTCT, and GAA repeats in yeast (142, 263, 308, 349). In addition, specific helicases assist replication forks by unwinding unwanted DNA secondary structures. Yeast helicases Srs2 and Sgs1 unwind hairpins formed by CTG and CGG repeats in vitro, with Srs2 being more efficient (350). In vivo, knocking out SRS2, but not SGS1, promotes instability and fragility of CAG and CGG repeats (351–353), although some reports found that Sgs1 also maintains CAG (354) and CGG repeat stability (289). Despite this disagreement between studies, the amounts of CAG repeat instability in srs2Δ or sgs1Δ mutants reflect the amount of fork stalling. Furthermore, human helicase Rtel1, which unwinds hairpins formed by CAG repeats and complements the yeast srs2Δ mutant, prevents CAG repeat expansions in human cells (355). These observations support the model in which Srs2/Rtel1 and, to a lesser extent, Sgs1 helicases unwind DNA hairpins during DNA replication to prevent fork stalling. Interestingly, the secondary structure–unwinding activity of Srs2 and Sgs1 seems to be specific to hairpin unwinding. Indeed, the instability and fragility of non-hairpin-forming repeats, such as telomeric or GAA repeats, does not depend on Srs2 (142, 183, 351) and is mildly affected by Sgs1, if at all (142). It is not yet known whether there are other helicases that unwind DNA hairpins. Recently, for example, human DNA helicase B was reported to localize to replication stalls at CGG repeats (356), although direct evidence that this helicase unwinds hairpins is so far absent.

Can other DNA secondary structures such as G4 or triplex also be unwound by specialized helicases? Because G4 is so stable under physiological conditions, there is a team of helicases that unwind this structure. This team includes FANCJ, Rtel1, BLM, and Pif1, to name just a few (reviewed in Refs. 191, 192, and 357). Triplex-unwinding helicases, on the other hand, are still largely unknown. Reports have demonstrated in vitro triplex-unwinding activity for several proteins, namely human Dhx9 (358), Ddx11 (359), and BLM and WRN helicases (360) as well as the yeast Stm1 protein (361). However, in vivo evidence for the involvement of these helicases in triplex unwinding is scarce. In particular, data from our laboratory did not show evidence for Stm1 and Chl1 (Saccharomyces cerevisiae Ddx11) triplex-unwinding activity in yeast in vivo (183).

Role of transcription in repeat instability

Remarkably, expandable repeats can exhibit instability in the absence of replication (257, 362), and, in some circumstances,
transcription through expandable repeats is sufficient to destabilize them. As a matter of fact, transcription through expanded repeats promotes their fragility (257, 363), instability (154, 330, 364–367), and RIM (368). On the face of it, this phenomenon seems unexpected, considering that there is no DNA synthesis involved in transcription. Nonetheless, transcription can affect repeat instability via at least three different mechanisms. First, transcription can change the chromatin landscape and, as such, alter repeat stability indirectly (see below). Second, the Smith laboratory recently showed that DNA replication often initiates at transcription start sites in human cells (369). This means that transcription determines origin-firing patterns, which could directly control repeat instability (see above). Third, formation of transcription-dependent R-loops directly influences repeat instability.

An R-loop is a DNA–RNA hybrid that forms during transcription when a nascent RNA strand invades the DNA duplex and reaneals to its template DNA via normal Watson–Crick base pairing. In a eukaryotic cell, R-loops act as a double-edged sword. On the one hand, they play an important role in the regulation of gene expression and transcription termination. On the other hand, they can stall replication and transcription, promote replication-transcription collisions, and, as such, lead to DSB formation and genome instability (reviewed in Refs. 370 and 371).

In vitro, transcription through CAG, CGG, and GAA repeats induces single and double R-loops. The amount and size of these R-loops correlate with an increase in negative supercoiling upstream of an elongating RNA polymerase (365, 372). R-loops are also detected at expanded GAA, CGG, GGGGCC, and CAG (72, 373) repeats in patient samples as well as in several experimental systems (79, 365). Most importantly, R-loop formation within long repeats likely promotes their instability. This is evidenced by the fact that RNase H, which degrades R-loops, counteracts repeat instability (75, 365, 374).

How can formation of R-loops lead to repeat destabilization? When an R-loop forms on a template strand, it leaves the complementary part of the DNA duplex single-stranded, thus promoting the formation of dynamic DNA structures. The result is a composite DNA structure consisting of an R-loop and an alternative DNA structure (Fig. 7) that facilitates repeat instability. For example, formation of a DNA hairpin opposite to an R-loop strand (S-loop) was suggested to explain the R-loop–induced instability of CAG and CGG repeats (79, 365, 374). In a G4-forming repeat, a structure called a G-loop is possible. It can come in two flavors: either G4-DNA is formed in the non-template strand, or a hybrid G4-DNA–RNA structure is formed between the non-template DNA and a nascent RNA strand. Such structures might form in human cells with expanded CGG and GGGGCC repeats (373, 375). Likewise, a triplex-forming repeat might form an H-loop; in this structure, an R-loop is formed between an RNA transcript and a single-stranded portion of H-DNA (74, 75) (Fig. 7). Collision between a replication fork and an H-loop formed by a GAA repeat might lead to the formation of a DSB, followed by repeat expansion in the course of BIR (75).

Another level of complexity is added by the fact that most of the mammalian genome is bidirectionally transcribed (376, 377), and repetitive regions are no exception (115, 378–380). Bidirectional transcription was detected for CAG (381–384), CGG (380, 385), GGGGCC (379), and other repeats (reviewed in Ref. 378) and can synergistically destabilize them (386), likely due to formation of double R-loops (387).

Role of DNA repair in repeat instability

Replication and/or transcription through expanded repeats fosters the formation of DNA secondary structures, which might lead to the formation of a DNA nick or a DSB. This DNA damage then serves as a substrate for various pathways of DNA repair. Consequently, DNA repair plays an important role in repeat instability.

HR

As discussed above, structure-forming repeats promote DSBs either during replication or transcription through the repeat. Thus, it is not surprising that these repeats serve as hotspots for homologous recombination in Escherichia coli (388–393) and yeast (394). These recombination events are often accompanied by a repeat length change (388, 390, 392, 395–398), likely because of an out-of-register strand invasion during the classical HR. However, in yeast, the majority of unstable repeats are not hotspots for recombination during meiosis (289, 394, 399), and
their barely detectable meiotic instability depends on the Spo11 nuclease (267, 400). Data from human (401, 402) and mouse (403, 404) pedigrees also agree that repeat instability generally does not arise from unequal crossing over. A notable exception to this rule is poly(A) diseases, which are caused by expansions of GCN repeats (46, 47, 50, 52, 53). In this case, expanded alleles likely originate via unequal crossing over between parental repeats (44, 51, 52, 109, 398) (Fig. 8A). This explains the lack of detectable somatic instability of GCN repeats as well as their small scale of expansion.

Besides canonical HR, two other pathways, single-strand annealing (SSA) and BIR, were implicated in repeat instability. The SSA pathway involves the resection of DSB ends followed by the annealing of flanking direct repeats, which normally results in a deletion. SSA does not require strand invasion performed by Rad51, but it does require strand annealing by Rad52 or Rad59. As such, one would expect this pathway to give rise to repeat contractions, rather than expansions (Fig. 8B). Indeed, an artificial induction of a DSB within a repetitive tract drives repeat contractions via SSA (396, 405, 406).

The BIR pathway is a highly error-prone, conservative mode of DNA replication that specifically repairs one-ended DSBs. BIR is identified by its mutational signature and dependence on Rad52 recombinase and a processivity subunit of DNA polymerase δ, Pol32 (PolD3 and PolD4 in mammals) (reviewed in Ref. 347). We have recently shown that in yeast, large-scale expansions of CAG repeats (407) and transcription-induced expansions of GAA repeats (75) happen through BIR. Moreover, BIR drives instability of carrier-size CGG repeats in a mammalian cell culture system (276). In the latter case, small-scale expansions and contractions of CGG repeats are accompanied by point mutations and complex rearrangements in the reporter gene. In all of these scenarios, BIR is likely triggered by a secondary structure formation at the repeat locus, and mutations accumulate during error-prone synthesis by Pol δ (Fig. 8C). Aside from BIR, RIM spanning to kilobases away from the repeat can also arise via a different mechanism. Replication through a repeat can leave behind a long single-stranded gap to be filled in by an error-prone translesion polymerase (264, 268, 389).

Importantly, structure-forming repeats not only promote recombination, but can also interfere with it. Any recombination event has to start with resection and typically involves ssDNA end invasion. If this end readily folds into a secondary structure, it could impede the recombination process (408). Unusual DNA structures may inhibit later stages of HR as well. For example, triplexes inhibit branch migration of Holliday junctions in vitro (409).

End-joining (EJ) pathways

In addition to various HR pathways, cells have evolved several EJ pathways to repair DSBs. As follows from the name, EJ pathways involve the direct fusion of two broken DNA ends.
The classic pathway, named nonhomologous EJ (NHEJ), involves threading the DNA into Ku70/80 protein rings followed by the direct ligation of the two DNA ends with DNA ligase IV. This pathway leads to a virtually error-free mode of repair. However, when ends cannot be directly ligated, they need to be processed. This can result in small (up to 4-bp) deletions. In a situation when NHEJ cannot be executed, cells employ more error-prone, noncanonical variations of EJ pathways, such as alternative NHEJ, microhomology-mediated EJ (MMEJ), or synthesis-mediated MMEJ (reviewed in Ref. 270). We do not yet know all of the details and differences between these mechanisms, but what they have in common is that they (i) are highly error-prone and typically result in indels up to 20 bp in size, (ii) do not require classic NHEJ mediators such as Ku70/80 proteins or DNA ligase IV, and (iii) typically act when classic NHEJ is compromised (reviewed in Refs. 270 and 410).

The classic NHEJ seems to be protective against repeat instability, as evidenced by experiments with CAG repeats in yeast (395, 411) and CGG repeats in mice (412). Probably, NHEJ directly ligates a subset of DNA ends originated from a DSB within a repeat, without an alteration in repeat length. Otherwise, these ends would have been repaired via repeat-stabilizing HR or noncanonical EJ pathways.

How can noncanonical EJ pathways destabilize repetitive sequences? It was found that Pol β, which might participate in MMEJ (413), promotes CAG repeat expansions in vitro (414). Knocking down key alternative NHEJ modulators, namely XRCC1, LIG3, and PARP1, suppresses CAG repeat contraction induced by environmental stresses, such as cold, heat, and hypoxia, in human cell cultures (329).

At first glance, it might seem that HR plays a much bigger role in repeat instability than EJ pathways. However, this may simply reflect the fact that most mechanistic studies regarding repeat instability were primarily conducted in replicating yeast or mammalian cells. In replicating cells, DSBs typically arise during DNA replication and are predominantly repaired via HR pathways. Independent of replication, specifically in G₀ and G₁ cell cycle stages, DSBs originate in the course of transcription and result from oxidative damage to DNA. These breaks are typically repaired via EJ pathways (reviewed in Refs. 415 and 416). To fully elucidate the relative contributions of HR and EJ in repeat instability, one needs to carefully compare repeat instability between isogenic replicating and nonreplicating cells.

**MMR**

MMR recognizes and fixes mismatches as well as small loop-outs that distort the double-helix structure. In eukaryotes, mismatch repair starts by the binding of MutS complexes that recognize a lesion. Mismatches and 1–3-bp loop-outs are recognized by the MutSα (Msh2 and Msh6), and larger loop-outs are recognized by MutSβ (Msh2 and Msh3). Upon mismatch recognition, one of several MutL complexes excises the lesion to allow Pol δ to fill the resulting gap (reviewed in Ref. 417). Mutations in MMR give rise to Lynch syndrome, an autosomal dominant disease associated with a high risk of colon cancer. On the molecular level, Lynch syndrome is characterized by high microsatellite instability: the patients have a higher rate of length instability among various short tandem repeats (reviewed in Ref. 418).

Because MMR protects from microsatellite instability in colon cancer, does it also suppress repeat expansion in REDs? Surprisingly, both somatic and intergenerational expansion of CAG and CGG repeats are virtually absent in mice lacking Mlh1, Mlh3, Msh2, or Msh3 (257, 419–427). In other words, MMR promotes rather than suppresses repeat expansions. The profound effect of MMR on repeat instability could explain the strain-specific differences in the amount of repeat instability: differences in the expression level of Mlh1 and/or Msh3 in different mice strains correlate with the amount of repeat somatic instability in those mice (419, 428, 429). Interestingly, the effect of knocking out Msh6 or Exo1 has either a much smaller (420, 423) or an opposite effect (422, 430) on repeat instability in mice. In agreement with mouse data, knocking out components of mismatch repair in other systems decreases CAG repeat instability rates (335, 389, 431, 432), although some studies have failed to detect an effect (161, 280, 433). Instability of other repeats, such as GAA, also seems to be promoted by components of MMR, but to a lesser extent, and not specifically by MutSβ (434–439).

Does MMR play as big of a role in mediating repeat expansions in RED patients as it does in mice and other experimental systems? Given the lack of experimental data, analysis of big data sets can address this question. Numerous genome-wide association studies have found that SNPs in the MLH1 and MSH3 genes act as genetic modifiers of HD's age of onset (128, 130, 325–327, 440). Likewise, an SNP in PMS2 modifies age of onset in several other poly(Q) diseases (441). Furthermore, an SNP in MSH3 affects the CTG repeat somatic instability level in DM1 patients (28, 440).

What molecular mechanism underlies MMR’s propensity to promote repeat expansion? Based on in vitro data, it seems like MutSβ is very efficient at repairing short slip-outs of 1–3 CAG/CTG repeats, but it cannot handle long or clustered slip-outs (442). Other studies have found that the MutSβ (420, 424) and, possibly, MutSα (430) complexes bind to a CAG or CGG hairpin DNA efficiently but unproductively, meaning that cleavage does not follow binding (420, 424). Overall, it seems that secondary structures formed by expanded repeats, such as hairpins, can escape repair by MMR. Consequently, they might become incorporated into the DNA, resulting in a repeat expansion.

There are still plenty of unanswered questions in regard to the interaction between MMR and structure-forming repeats. What are the direct consequences of MutSβ’s unproductive binding to a hairpin? What is the precise contribution of MutSα and MutL complexes in this process? What are the mechanistic details of MMR’s interaction with non-hairpin-forming repeats, such as GAA? Future studies focusing on the role of MMR in repeat instability will shed light on these mysteries.

**Base excision repair (BER)**

BER is a conserved DNA repair pathway that corrects modified or damaged DNA bases. In the course of BER, a specialized DNA glycosylase removes the damaged DNA base to create an abasic site, after which AP endonuclease 1 makes a nick. Then
DNA Pol β substitutes the abasic site for the correct nucleotide. In a case in which the abasic site cannot be easily removed, long patch BER comes into play. In this case, other DNA polymerases, including Pol δ, perform strand displacement synthesis, followed by flap removal by Fen1 and DNA ligation (reviewed in Ref. 443).

Substantial evidence suggests that BER promotes somatic CAG repeat instability. First, components of BER machinery, namely Neil1 (444), Ogg1 (445), and Pol β (446, 447), promote CAG repeat instability in mouse tissues or cell extracts. In yeast, BER promotes R-loop–mediated contractions of CAG repeats (374). Finally, Pol β is enriched at CAG repeats in a tissue-specific manner, with tissues prone to higher CAG repeat somatic instability showing a larger accumulation of the protein (448).

What is the molecular mechanism of BER-promoted repeat instability? At least in vitro, Pol β is inherently slippery when synthesizing over CAG repeats and often adds (449) or skips over (450, 451) a portion of the repeat. Importantly, purified MutSβ complex additionally promotes Polβ-mediated repeat instability (449). Furthermore, in vivo, MutSβ and Pol β colocalize as a response to DNA damage (451) and together accumulate at CAG repeats during DNA replication (449). Overall, it appears that BER and MMR cooperate together to promote CAG repeat instability (449, 451). It is possible that during BER, CAG repeats may fold into a hairpin or a loop-out structure, which, if incorporated into the nascent strand, would lead to a repeat expansion. MutSβ likely stabilizes this structure, thus enhancing the chances for expansion (449, 452) (Fig. 9A). Notably, the oxidative damage that accumulates in neurons with age creates more opportunities for repeat-destabilizing BER repair, establishing a so-called “toxic cycle” of oxidative damage and repeat elongation. This model explains age-dependent CAG repeat instability in post-mitotic tissues (445).

NER

NER is another conserved DNA repair pathway. It removes DNA lesions that distort the DNA double helix: bulky DNA adducts, abasic sites, nicks, gaps, and DNA secondary structures. In the course of NER, the recognition of a lesion is followed by local DNA unwinding. Then the lesion-containing oligonucleotide is excised from the DNA, the remaining gap is

Figure 9. Two hypothetical hybrid DNA repair pathways that could promote repeat instability. A, an oxidized nucleotide within a repetitive tract is being removed by the BER pathway to create a nick. In the course of strand displacement synthesis, the displaced flap might form a DNA secondary structure (e.g. a hairpin). The structure is stabilized by MutSβ and gets incorporated into the DNA, resulting in a repeat expansion. B, when a repeat is located in an actively transcribed gene, RNA polymerase might promote formation of a DNA secondary structure, such as S-DNA. These structures, when additionally stabilized by MutSβ, stall the next RNA polymerase, triggering NER. The NER repair might lead to either excision or incorporation of a secondary structure, leading to repeat contraction (this scenario is shown) or expansion (not shown), respectively.
filled, and the DNA ends are ligated back together. Two major NER pathways exist. First, global genomic NER removes DNA lesions genome-wide. Second, transcription-coupled NER (TC-NER) specializes in removing lesions that occur during transcription and impede RNA polymerase progression (453).

Remarkably, TC-NER appears to be responsible for transcription-induced repeat instability in dividing yeast (454, 455), human cell cultures (330, 456), and mouse neurons (457). The latter might reflect a role for this pathway in somatic expansions of CAG repeats (Fig. 9B). It remains unclear, however, whether NER also plays a role in intergenerational repeat instability. On the one hand, NER indeed promotes germline CAG repeat instability in flies (366). Additionally, the Cockayne syndrome group B (CSB) protein, which is required for TC-NER, promotes CGG repeat expansion during maternal, but not paternal, transmission in mice (458). On the other hand, intergenerational CAG repeat expansion is inhibited by CSB (459). We do not yet know how to explain this contradiction.

In sum, an emerging consensus in the field is that several hybrid pathways that combine components of MMR, BER, and NER could be responsible for repeat instability (Fig. 9). The length and type of a repetitive sequence, the replicational and transcriptional state of a cell, and the relative concentration of various repair proteins likely determine the choice of pathway and, ultimately, the repair outcome: repeat expansion, contraction, or no change (459).

Other DNA repair pathways

Genome-wide association studies have also identified the FAN1 gene, which encodes a flap endonuclease with a poorly understood biological role, as an age-of-onset modifier in several polyQ diseases (128, 130, 325–327, 441, 460). Moreover, the study of prediagnostic carriers of expanded CAG alleles in HD revealed that FAN1 SNPs can reliably predict age of onset (461). These SNPs likely affect the FAN1 transcription level. This is supported by the fact that an increased Fan1 expression level negatively correlates with HD age of onset and prevents CAG repeat expansion in human cell cultures (462). Thus, high expression of the FAN1 gene probably prevents CAG repeat somatic instability, thus delaying age of onset. This effect might not be exclusive to CAG repeats, as Fan1 also inhibits somatic expansions of CGG repeats in mice (463).

We do not yet know how Fan1 protects from repeat expansions mechanistically, as its normal biological role is not clear. Fan1 is a conserved flap endonuclease (464) that participates in interstrand cross-link repair either independently or as a part of the Falconi anemia repair pathway (reviewed in Ref. 465). Importantly, Fan1 might interact with MMR (466–468) and play a role in the processing of stalled replication forks (reviewed in Ref. 465). One proposed hypothesis is that Fan1 either prevents MMR proteins from binding to CAG repeat loop-outs or sequesters MMR proteins, thus counteracting MMR’s promotion of CAG repeat expansion (462). Another hypothesis is that Fan1 may substitute for Exo1 in the course of an MMR-like process to promote an error-free repair of a repeat loop-out (463). It would be intriguing to investigate these possibilities in further detail.

Role of the chromatin environment in repeat instability

Expanded tracts of structure-forming repeats may dramatically alter the surrounding chromatin. They can lead to the formation of repressive histone marks, such as histone H3 Lys-9 and histone H4 Lys-20 trimethylation, decreased histone acetylation, and changes in DNA methylation. These epigenetic changes often contribute to disease phenotypes, as they can alter gene expression patterns (469, 470). The chromatin environment, in turn, can promote or inhibit repeat instability. For example, decreased nucleosome occupancy likely promotes large-scale GAA repeat expansion in yeast (367). Consistently, bioinformatic analysis revealed that, in human cells, expandable CAG repeats are located in open chromatin regions, unlike stable CAG repeats (471).

How can an open chromatin region facilitate repeat instability? First, the loss of nucleosomes undoubtedly facilitates formation of dynamic DNA structures, due to the release of negative DNA supercoils previously stored at nucleosomes (472). Second, large nucleosome-free regions might provide more opportunities for template switching during DNA replication (367). Third, open chromatin regions are more accessible to various DNA repair factors, particularly MMR (reviewed in Refs. 417 and 473). As such, open chromatin might facilitate error-prone DNA repair within repeats (417).

Furthermore, patterns of nucleosome assembly might be an important factor controlling repeat expansion (455, 474). The fact that ATTCT (475) and CAG (474, 476, 477) repeats are potent nucleosome-positioning elements in vitro likely contributes to their instability. Types of nucleosomes also can affect repeat instability, as evidenced by the recent finding that the presence of a specific histone variant contributes to stable maintenance of the CAG repeat (478). Importantly, other DNA-binding proteins also might alter repeat stability. For example, binding of the CCCTC-binding factor to the flanking regions of several CAG repeat diseases might stabilize this repeat (479).

A recent finding suggested a more general model for the role of chromatin structure on repeat expansions. Based on Hi-C data, it was suggested that the position of a repeat relative to the 3D genome architecture influences its propensity to expand. This was based on the observation that most disease-causing repeats are located at 3D chromatin boundaries. Moreover, CGG repeat expansion in the FMR1 promoter shifts the gene’s position relative to the chromatin boundaries, which results in its complete silencing and ultimately causes the disease (480). This discovery might explain why some human tandem repeats are relatively stable and others are prone to dramatic expansions.

Why do disease-causing DNA repeats persist in the genome?

The majority of REDs are severe, debilitating disorders, like fragile X, FRDA, ALS, Huntington’s disease, etc. Why are expandable DNA repeats not eliminated from the human genome if they are so detrimental to human health? Several hypotheses address this conundrum.

The most intuitive explanation is based on the fact that the majority of REDs have a late onset and typically exhibit symp-
In summary, when a person already has passed reproductive age. Thus, long repeat alleles are not subject to purifying selection. In support of this hypothesis, repeats in coding areas never expand as much as repeats in noncoding areas, likely because long repeats within a coding region are toxic and quickly counterselected. However, this hypothesis does not consider the existence of REDs with childhood or even congenital onset.

An alternative explanation could be that the existence of REDs is the price that eukaryotes pay for the use of tandem repeats in genome function and evolution. Numerous tandem repeats appear to be inherent to the eukaryotic genome: in humans, tandem repeats constitute nearly 3% of the genome, a similar proportion as protein-coding sequences (131). All tandem repeats are unstable in length, and the rate of their length instability exceeds the rate of point mutations by several orders of magnitude (132, 481).

What could be the advantages of having so many tandem repeats in a genome?

One well-known advantage is that tandem repeats serve as principal genomic structural elements, such as telomeres (482), centromeres (483), and regions of constitutive heterochromatin (484). Importantly, the same mechanisms that lead to repeat expansions in human disease are involved in the maintenance of those structural elements (reviewed in Ref. 344).

Another popular hypothesis is that the abundance of tandem repeats increases genome evolvability (132). Several pieces of evidence support this idea. First, tandem repeat loci (but not the lengths of the repeats) are conserved in related species (471, 485). At least some of these repeats seem to be actively maintained (486), which indicates that the existence of tandem repeats in these loci could be advantageous from an evolutionary standpoint. Sometimes, tandem repeat sites even evolve independently in homologous genes (132).

Second, about 15% of all human genes contain tandem repeats in their promoters or coding regions (reviewed in Ref. 132). Importantly, length variation of tandem repeats located in regulatory regions correlates with variance in the expression of these genes (487, 488). On top of that, tandem repeats located in gene promoters were shown to drive evolution of gene expression in yeast (489). Moreover, genes that contain tandem repeats in their coding or promoter regions are typically regulatory genes or transcription factors (reviewed in Ref. 132). Furthermore, a recent study found that DNA secondary structures, which commonly form within tandem repeats, are abundant in vivo and may control gene regulation (490). Taken together, these observations point to a potential role of tandem repeats in regulatory evolution.

Third, tandem repeats located in coding regions, typically in-frame trinucleotide repeats, could drive protein evolution. There are well-documented examples of tandem repeat tracts within protein-coding sequences whose length variation permits quick adaptation to a particular environmental change (132, 491, 492). One such example is the tandem repeats located in coding regions of circadian clock genes. The length of these repeats directly determines the circadian rhythms of various species, including fungi, flies, and, possibly, birds. In other words, these repeats act as a tool used when a species has to quickly change its regimen in response to habitat change (reviewed in Ref. 132). Another example is repetitive protein motifs located in genes that define vertebrate body morphology. Length variation of these repetitive sequences likely allows rapid evolution of a body shape without dramatically altering overall body topology (493). For example, the ratio of poly(A)/polyQ repeat tract lengths in the canine RUNX-2 gene defines head morphology in dogs. Along the same line, deletion of a portion of a polyQ tandem repeat in the canine ALX4 gene causes dogs to grow an extra claw (493). In humans, on the other hand, an expansion of poly(A) repeats in the coding part of the HOXD13 and HOXA13 genes causes REDs, characterized by abnormal limb morphology, synpolydactyly 1 (46, 47), and HFG syndrome, respectively (48).

An additional proof of the tandem repeats’ role in genome evolvability is their contribution to cancer progression. For example, it was recently found that tandem repeat length variation allows cells to quickly evolve during the early stages of malignant transformation (494). Furthermore, as discussed above, translocation junctions in cancer genomes often coincide with structure-forming tandem repeats (269).

Finally, we cannot resist mentioning a radical hypothesis suggested by a titan in the field of molecular evolution stipulating that expansions of unstable triplet repeats were at the heart of early life evolution (495).

Unanswered questions

To what extent do model experimental systems reflect repeat instability seen in humans?

The elephant in the room of the RED field is the substantial difference in how expandable repeats behave in humans as compared to model experimental systems. In humans, intergenerational transmissions of premutational or full mutation repeat alleles lead to a dramatic change in length with nearly 100% probability. As discussed above, the direction of instability (contraction or expansion) and its scale depend on multiple factors, including parental gender and age. There are likely different mechanisms that are responsible for various types of instability: contractions or expansions, small-scale or large-scale. Each of these types of instability seem to predominantly happen in a given germ cell type and within a given developmental window (496). That being said, expansions are prevalent over contractions in multiple circumstances (see above for examples).

Unlike human pedigrees, repeats studied in commonly used simple experimental systems, such as yeast or mammalian cell cultures, exhibit instability with a probability on the order of $10^{-3}$ to $10^{-6}$ (263) and predominantly contract rather than expand (142, 183, 497), with rare exceptions (154, 498). Furthermore, the expansions that are observed in these systems are smaller in scale compared with those observed in human pedigrees, even under selection for expansions (142, 161, 263). Mouse models, at first glance, better recapitulate the patterns of repeat instability in humans. Yet even mice rarely, if at all, exhibit the addition of hundreds of repeats characteristic of human pedigrees (230, 233, 499, 500). This is particularly surprising given the fact that mice generally tolerate longer tan-
dem repeats better than humans, as illustrated by extremely long mouse telomeres (501). There is no coherent explanation for this difference between humans and model experimental systems. Understanding the reasons for this difference is an important goal for future studies.

Despite this difference, we are convinced that investigating mechanisms of repeat instability in model systems provides critical insights. One example is the role of MMR in somatic repeat instability, which became evident from experiments in yeast and mice and was then confirmed by genome-wide association studies in RED patients (see above). Other examples are the “ori-shift” and “ori-switch” models (292), discussed above. The inspiration for these models came from the observation that repeat instability in yeast, bacteria, and mammalian episodes uniformly depends on the repeat orientation and distance relative to the nearest origin of replication (see above). A decade later, an SNP in the closest origin of replication to the CGG repeat in fragile X locus was found to correlate with this repeat’s replication program and propensity to expand in human pedigrees (502).

Overall, simple model systems provide the advantage of conducting detailed molecular genetic analysis of repeat instability mechanisms, which is impossible to do in humans and difficult in mice. The remaining challenge, however, is to relate a mechanism unraveled in a model system with a mode of repeat instability observed in a particular cell type and/or developmental window in humans. Overcoming this challenge would shed light on the origin of REDs and, in the long run, help with developing therapies.

Can REDs be cured?

The most important and urgent question facing the field of REDs is whether we can start to use the extensive knowledge about expandable repeats to fight these debilitating and currently incurable diseases. In attempts to develop therapies for REDs, researchers try to obviate repeats’ toxicity on different levels. Typically, potential therapies aim at the downstream effects arising from repeat expansion to reduce their toxic consequences. The goal is to either reverse gene silencing (70, 81, 469, 503–505) or alleviate adverse effects of toxic RNA (98, 101, 203, 506–514) and/or toxic proteins (98, 101, 203, 506, 515). These methods, while worthy and promising, warrant several caveats. First, undergoing a therapy that targets downstream repeat-mediated toxicity would likely be a lifelong commitment for a patient. Second, the precise toxicity mechanisms are unestablished for the majority of REDs. Thus, it is unclear which toxic factor is the best therapeutic target. Third, targeting downstream effects of repeat expansions could alleviate disease symptoms post-onset but is unlikely to postpone disease onset.

Another direction could be inhibiting cellular pathways that promote somatic repeat expansion (252). For example, inhibiting MMR or BER could reduce somatic expansions of CAG repeats (426, 439). An obvious drawback of this approach is that inhibition of DNA repair pathways could dramatically increase cancer risk, as indicated by Lynch syndrome and other cancers linked to mutations in MMR (reviewed in Ref. 418).

Yet another promising therapeutic approach could be to prevent the formation of dynamic DNA structures that are at the heart of repeat expansions. One way to achieve this is to introduce endogenous oligonucleotides. The binding of an oligonucleotide to a single-stranded repetitive strand would compete with the formation of DNA secondary structure and therefore shift the equilibrium away from secondary structure formation. Indeed, the addition of (CTG)n or (CAG)n oligonucleotides complementary to the lagging strand template rescues replication fork stalling within the repeats in HeLa cells (336). The downside of this approach is that the half-life of DNA oligonucleotides is short in the human body. To address this problem, chemically modified oligonucleotides that do not easily degrade could be used. Alternatively, one could use peptide nucleic acids or polyamides that recognize expandable DNA repeats and preclude them from forming alternative DNA conformation. As a proof of principle, GAA-specific polyamides were recently shown to rescue fork stalling at GAA repeats in patient-derived samples (339). Additionally, another recent report showed that a small molecule binding to an S-DNA structure promotes CAG repeat contractions in vivo (549).

Finally, directly promoting repeat contractions in somatic cells might present a powerful therapeutic approach. Several teams have achieved progress in this direction using transcription activator-like effector nucleases, zinc finger nucleases, or CRISPR-Cas9 to induce a DSB within or adjacent to a repeat. This approach promoted repeat contractions in bacterial (406), yeast (433, 516), primate (517), and human (206, 397, 518–524) cells (also reviewed in Ref. 525). However, at this stage of research, the safety of such technologies remains unclear. As with any other genome-editing techniques, the absence of off-target effects should be thoroughly verified. An additional concern is that besides repeat contractions, DSBs might also drive repeat expansions, which could counteract their therapeutic effects. Instead of inducing a DSB within a repetitive DNA, it might be safer to create a nick using Cas9 nickases. Indeed, the Dion laboratory has recently found that creating a nick within an expanded CAG repeat facilitates its contraction but not expansion (519).

Several methods can be utilized to drive repeat contractions in somatic tissues of RED patients. One possibility could be viral delivery of a Cas9 or related nuclease that make a targeted DSB or a nick at a repeat. Recently, for example, delivery of Cas9 targeted to the mutated HTT gene via adeno-associated virus reduced HD symptoms in mice (526). Alternatively, repeats may be shortened in situ in patient-derived iPSCs, followed by their differentiation. This approach allows one to scrupulously select for cells in which the repeat contracted successfully and to ensure the absence of off-target effects. The resultant cells can be transplanted back into the patient. A challenge to all of these approaches is that it probably would be very hard to “correct” the repeat’s length in all affected cells. Thus, additional studies are needed to evaluate whether contracting a repeat in a subset of affected tissues could prove sufficient to reverse disease progression.

As of today, there is still no effective way to manipulate repeat length in RED patients. However, contemporary diagnostic methods allow one to precisely estimate the length of all repeats (4153).
known expandable repeats in a given person. As such, future parents might choose to know their carrier status for the most common REDs, such as HD, fragile X, or DM1. This type of genetic testing is especially encouraged for people with a family history of REDs. For example, screening programs that aim to identify carriers of DM1 have existed in Canada since 1988 (527). Aspiring parents who appear to be carriers of an RED can then choose to perform in vitro fertilization (IVF) with preimplantation genetic diagnosis (PGD) of the repeat length to ensure that their offspring do not inherit a disease-size repeat allele (528–532). Note, however, that the procedure of IVF/PGD comes with serious ethical and legal challenges (533). For example, a mother-to-be might want to perform prenatal testing of a fetus, because her male partner has a family history of HD. However, that might reveal his carrier status, even if he would prefer not to know it (534).

Concluding remarks

Almost 30 years have passed from the day when the first RED was discovered (6). Since then, scientists have unraveled the basic principles that underlie mechanisms of repeat instability. First, the repetitive tract forms a DNA secondary structure—typically transiently—during DNA unwinding in the course of replication, transcription, or repair. Then, after a biological transaction that often involves some complex interplay between various modes of DNA repair, the DNA secondary structure gets excised or incorporated into the DNA. As a result, the repetitive tract either contracts or expands, respectively. This paradigm applies to the majority of repeat instabilities, although the detailed mechanisms depend on the repeat sequence, length, and location within a genome as well as the organism or cell type examined.

As of now, the majority of REDs are incurable and nonpreventable. Because structure-forming DNA repeats are central to the development of these diseases, direct targeting of those repeats could be a promising therapeutic strategy. In the last few years, researchers have undertaken the first attempts aimed at developing such therapies (206, 339, 397, 518–522, 524, 525, 526). Hopefully, one day, these approaches will be used to combat REDs.

Acknowledgments—We thank Jorge Cebrian for sharing unpublished results and Julia Hisey and Ethan Hoffman for proofreading the manuscript.

References


dependent structure formation in Friedreich ataxia (GAA), (TTC), repeats at neutral pH. \textit{Nucleic Acids Res.} \textbf{32}, 1224–1231 \textsc{CrossRef Medline}


307. Xie, Y., Counter, C., and Alani, E. (1999) Characterization of the repeat tract instability and mutator phenotypes conferred by a Tn3 insertion in RFC1, the large subunit of the yeast clamp loader. Genetics 151, 499–509 Medline
local depletion of specific dNTPs? Bioessays 35, 306–313 CrossRef Medline


JMLH1, PMS1, and PMS2. J. Biol. Chem. 282, 2976–2986 CrossRef Medline


JBC REVIEWS: Repeat-mediated genome instability


JBC REVIEWS: Repeat-mediated genome instability


