

Review

Structural and Dynamical Properties of Nucleic Acid Hairpins Implicated in Trinucleotide Repeat Expansion Diseases

Feng Pan ^{1,2} , Pengning Xu ^{1,3} , Christopher Roland ¹, Celeste Sagui ¹ and Keith Weninger ^{1,*}

¹ Department of Physics, North Carolina State University, Raleigh, NC 27695, USA; fpan3.ncsu@gmail.com (F.P.); cmroland@ncsu.edu (C.R.)

² Department of Statistics, Florida State University, Tallahassee, FL 32306, USA

³ Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

* Correspondence: krwening@ncsu.edu; Tel.: +1-(919)-513-3696

Abstract: Dynamic mutations in some human genes containing trinucleotide repeats are associated with severe neurodegenerative and neuromuscular disorders—known as Trinucleotide (or Triplet) Repeat Expansion Diseases (TREDs)—which arise when the repeat number of triplets expands beyond a critical threshold. While the mechanisms causing the DNA triplet expansion are complex and remain largely unknown, it is now recognized that the expandable repeats lead to the formation of nucleotide configurations with atypical structural characteristics that play a crucial role in TREDs. These nonstandard nucleic acid forms include single-stranded hairpins, Z-DNA, triplex structures, G-quartets and slipped-stranded duplexes. Of these, hairpin structures are the most prolific and are associated with the largest number of TREDs and have therefore been the focus of recent single-molecule FRET experiments and molecular dynamics investigations. Here, we review the structural and dynamical properties of nucleic acid hairpins that have emerged from these studies and the implications for repeat expansion mechanisms. The focus will be on CAG, GAC, CTG and GTC hairpins and their stems, their atomistic structures, their stability, and the important role played by structural interrupts.

Keywords: trinucleotide repeats; expansion diseases; hairpin structure; single-molecule FRET; smFRET; molecular dynamics simulations; CAG and GAC; CTG and GTC



Citation: Pan, F.; Xu, P.; Roland, C.; Sagui, C.; Weninger, K. Structural and Dynamical Properties of Nucleic Acid Hairpins Implicated in Trinucleotide Repeat Expansion Diseases. *Biomolecules* **2024**, *14*, 1278. <https://doi.org/10.3390/biom14101278>

Academic Editor: Sofie Lautrup

Received: 28 August 2024

Revised: 26 September 2024

Accepted: 5 October 2024

Published: 10 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The stability of the genome is important to all organisms in order to maintain the complex network of biochemical interactions that underlie life. The survival of any organism is challenged by exposure to agents that damage the genome, be they external agents (such as UV light, ionizing radiation, and toxic chemicals) or harmful products of endogenous metabolic processes (such as reactive oxygen). Such genome-modifying agents can benefit humans, for example when used as anti-microbials (e.g., industrial sterilization of materials by gamma-ray exposure), but they can also result in human diseases with cancer being most prominent.

Beyond these sorts of external stressors, it has been recognized that another mutagenic agent is encoded in the DNA sequence itself: nonstandard or atypical nucleic acid structures that can result from certain DNA sequences [1]. One of the most significant classes of sequence motifs associated with genomic instability is simple sequence repeats (SSRs) consisting of one to six (and even 12) nucleotides that are repeated up to 30 times (and more for pathological cases) [2]. In many cases, these repeated sequences in the genome are surprisingly susceptible to changes in their length.

Some instability of SSRs can be part of the natural processes in cells. It has been estimated that the rate of length mutations in some SSRs is about 10^5 times higher than that of a point mutation [3]. This can lead to frequent polymorphism in the coding regions of the

genes with a correspondingly rapid expansion of the amino acid repeats. This variability aids natural selection by rapidly generating new alleles and facilitating evolution.

In addition to these productive mutations, SSRs may also exhibit “dynamic mutations” that do not follow Mendelian inheritance. About a century ago, in humans, it was observed that the neurological myotonic dystrophy disorder [4,5] was an inherited disease whose age of onset decreased and whose severity increased with successive generations. Many other similar human disorders joined the family of so-called “genetic anticipation” diseases, but it took scientists until the 1990s to discover that these diseases were caused by the intergenerational expansion of SSRs [6–8]. To date, approximately 50 DNA expandable SSR diseases have been identified, and the list is expected to grow [9,10]; see Table 1. Despite being involved in dozens of diseases, the mechanisms underlying the instability of SSRs are generally unconfirmed, and the connections between the change in the repeat sequence and the emergence of disease phenomena are similarly opaque. In this review, we address recent progress toward uncovering mechanisms for a subset of human diseases known as trinucleotide repeat expansion diseases.

Table 1. A list of the locations on genes of the most common TRs along with an abbreviation of the most common diseases that are associated with them [11].

Location on Gene	Repeat	Disease
5'-UTR	CGG	FRAXA, FXTAS
	GCC	FRAXE
	CAG	SCA12
EXON	CAG	HD, HDL2, SBMA
		DRPLA, SCA1, SCA2, SCA3
		SCA6, SCA7, SCA17
	GAC	MSD
	GCG	SPD, HFG, ISSX, CCD
		HPES, OPMD, CCHS, BPES
INTRON	CAA	FRDA
	CCTG	DM2
	ATTCT	SCA10
	TGGAA	SCA31
	GGCCTG	SCA36
	GGGGCC	ALS
3'-UTR	CTG	DM1, HDL2, SCA8

1.1. Trinucleotide Repeat Sequences and Associated Human Diseases

Trinucleotide repeats (TRs) constitute the most common type of SSRs in the exome of all known eukaryotic genomes [12–15]. RNA transcripts containing TRs may greatly outnumber their genomic templates [16]. TRs in DNA may be neutral sequences or have several regulatory roles in gene expression. In RNA, they are believed to play a regulatory role mediated by interactions with TR binding proteins [17–19], regulating splicing, maturation and transport [20–27], while in the cytoplasm, they may regulate mRNA stability and translation [28] or promote repeat-associated non-AUG (RAN) translation [29,30]. The incidence of specific sorts of TRs and their position in genes varies greatly between different genomes, underscoring the vital role of TRs in genome evolution [12,31]. The dynamic changes in the length of TR regions in human genes cause disorders known as Trinucleotide (or Triplet) Repeat Expansion Diseases (TREDs) [32–34]. Triplet expansion diseases are a group of genetic disorders that include diseases such as Huntington’s disease, Myotonic dystrophy, spinocerebellar ataxia, and fragile X syndrome that affect the function of nerve cells, leading to severe neurological problems [8,35].

The genomic instability underlying Huntington’s disease provides an example of trinucleotide repeat expansion phenomena. Huntington’s disease pathology emerges when the CAG repeat sequence in the huntingtin protein gene expands from the typical 17–23 repeat range across a threshold of 36 repeats [36–39]. The expanded CAG region will result

in mRNA containing the expanded region as well as a longer polyglutamine insert within the translated huntingtin protein. Although these expanded polyglutamine regions make the protein more susceptible to aggregation and induce plaques, exactly what processes are toxic and lead to cell death remains under investigation [40–44].

More fundamentally, the processes causing the CAG expansion in the genome are unknown. The expansion is believed to be primarily caused by some sort of slippage during DNA replication, repair, recombination or transcription [1,8–10,32–34,45–53], which involves the transient separation of complementary DNA strands or exposure of a single DNA strand. One important breakthrough has been the recognition that all expandable repeats have atypical structural characteristics [8,10,47,53,54], including single-stranded hairpins, Z-DNA, triplex, G-quartets and slipped-stranded duplexes. In the example of Huntington's disease, models of mechanisms of expansion are focused on hairpins that can result when the CAG repeat-containing single-strand DNA folds on itself (Figure 1) [50,54–57]. Such DNA hairpin structures formed from a trinucleotide repeat have a duplex-type stem with 2/3 correct base pairing and 1/3 mismatched bases and a loop of three or four bases at the turn (Figure 1). These structures are likely recognized by mismatch repair proteins with their subsequent enzymatic processes starting a chain of events that result in the expansion. Cell toxicity and death arise in cells that contain expansions by pathways associated with toxic gain-of-function or loss-of-function for RNA transcripts and translated proteins [9,19,58–67]. The precise connection between TR expansion and pathological processes in cells remains a point of debate that we do not address in this review [41,42]. Rather, our focus here is on processes that relate to the expansion of the TR in the genome.

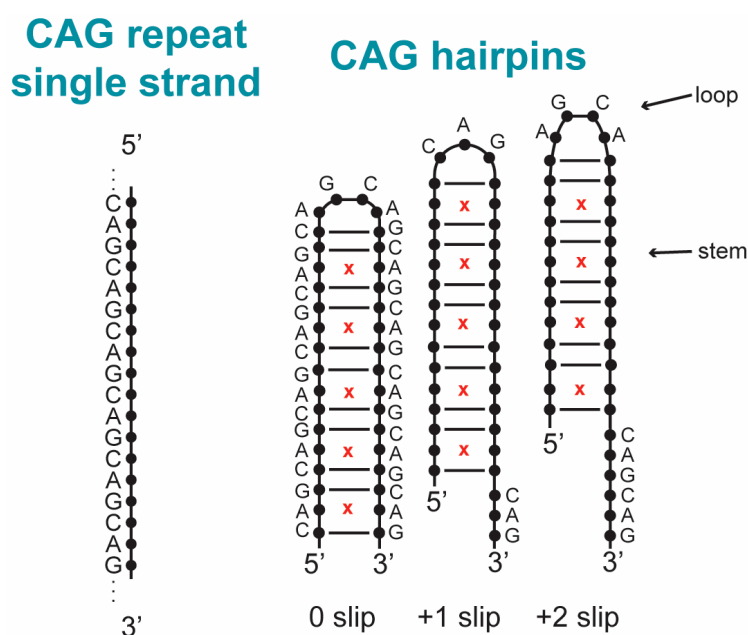


Figure 1. Examples of hairpins formed from CAG-repeat containing DNA. A single strand of CAG repeated DNA (left) can fold on itself to form hairpins (right). The stem is a duplex where a third of the base pairs are A–A mismatches (marked by red x) and the loop at the end includes unpaired bases. Three different states of the hairpin are displayed that are slipped in steps of a CAG unit indicated below the hairpin (leaving a short single strand overhang in this example). Some base labels are omitted for clarity. The stem base pairing pattern is maintained with slipped states. The loops illustrate that even and odd numbers of repeats in the stem result in tetraloops or triloops with 4 or 3 unpaired bases respectively.

1.2. DNA Mismatch Repair Proteins Are Causative for Disease-Related Expansion of Trinucleotide Repeat Sequences

How exactly do atypical structures manage to “gum up the works” and lead to proliferated expansion? After all, do cells not have the capability to repair these kinds of

errors? This is a largely unknown, open-ended question very much at the forefront of current research. The DNA mismatch repair (MMR) system acts after DNA replication and recombination to maintain genomic stability by recognizing and repairing chemically modified bases, mismatched base pairs, or insertion/deletion loops [68–72]. Perplexingly, the action of MMR has been associated with TR expansion [36,52,73–78]. In eukaryotes, MMR is initiated by either one of two heterodimeric MutS homolog complexes, MSH2-MSH6 (MutS α) and MSH2-MSH3 (MutS β), which combine lesion binding with ATP hydrolysis with distinct but sometimes overlying specificities [79,80]. MutS α repairs single-base mismatches and small, one or two-nucleotide loops. MutS β has strong specificity for two- to fifteen-nucleotide loops but can also signal the repair of some single-base mismatches. Mutations in the genes encoding the proteins of the MMR system can lead to further mutations and microsatellite instability throughout the genome [68–70,72], as is the case of hereditary non-polyposis colorectal cancer [74,81–89]. In contrast, MMR machinery that is not defective itself can facilitate different outcomes when encountering some types of DNA damage, where the process of lesion binding can trigger processes leading to cell-cycle checkpoint or apoptosis activation [72,90–94]. In the case of encountering TR duplexes, the MMR system most alarmingly becomes one of the causative factors leading to TR expansion [52,53,73,78]. These instabilities are localized only in the TRED allele, and do not spread throughout the genome as occurs with mutated MMR proteins. Genome-wide association studies (GWAS) correlated variants in MMR proteins to changes in CAG expansion disease severity in Huntington's disease patients [76,95–97], and patient-derived stem cell genetic manipulations further extended these conclusions [98]. In particular, the crucial role of MMR proteins in TR expansion has been shown in transgenic mouse models of Huntington's disease [99–101] and other expansion diseases [102–104]. In crosses of Huntington's disease (HD) transgenic mice with mice lacking MSH2, the CAG expansion was attenuated in both the somatic tissue of their descendants and in their germ cell [105,106]. In transgenic mice, the absence of MSH3 suppresses the expansion of the CTG repeat in the 3' non-coding region of the DM1 transgene as well as the expansion of the CAG repeat in the coding sequence of the HD transgene [52,107]. Instead, the loss of MSH6 has increased somatic repeat instability [108], which suggests that MutS α may try to prevent TR expansion. Findings of high levels of TRs and MSH3 in post-mitotic neurons in mice and humans also supported the notion that high levels of MutS β can cause TR expansion in nondividing cells [109].

While the workings of the MMR system are quite complex and the mechanisms leading to TREDs are still fairly unknown, there are associated features that are robust. In most TREDs, there is a correlation between the number of repeated nucleotides beyond a critical threshold and their increased further expansion and subsequent increased pathology. Thus, the age of disease onset decreases while the disease severity increases with each successive generation as the number of nucleotide repeats continues to increase. Another important breakthrough in the understanding of TREDs has been the recognition that the critical step in all models of repeat instability is the stable or transient formation of atypical non-B DNA stable secondary structures in the expandable repeats [32,46,55,78,109–112] and that these non-B DNA conformations “*not the sequences per se, promote mutagenesis in flanking regions*” [112]. While there are a number of pathogenic structures known to be associated with TREDs, i.e., hairpins, triple helices, Z-DNA, G-quadruplexes and others, it is the hairpin structures based on single-stranded DNA (ssDNA) that are by far the most prolific and commonly associated with TRs. Thus, an important first step in understanding the complex cascade of molecular changes undergone by the MMR system is to probe the structural and dynamical characteristics of nucleic acid hairpins.

In this brief review, we summarize the results of recent experimental and theoretical studies of SSR-based secondary structures implicated in TREDs; we will feature our own work based on large-scale molecular dynamics (MD) simulations [113–122] and single-molecule Fluorescence Resonance Energy Transfer (smFRET) [11,123,124] (for smFRET reviews, see Refs. [125,126]). Primarily, we shall discuss the structure and dynamics of

DNA/RNA hairpins of the most common TRs; these include DNA/RNA CAG, GAC, CTG (CUG for RNA), GTC (GUC for RNA), CCG, and GGC.

2. Results

For TREDs, the most common pathogenic structures are hairpin loops. Until recently, relatively little was known about the structure and dynamics of these loops at the molecular level. However, recent smFRET and nuclear magnetic resonance (NMR) experiments complemented by classical MD simulations are starting to reveal their atomistics, which we review here along with implications for SSR expansion mechanisms. For presentation purposes, it is convenient to discuss the results in terms of the homoduplexes representing the hairpin stem and the actual loop parts separately.

The experimental results discussed below were primarily generated using smFRET measurements. For these studies, DNA containing different numbers of triplet repeat units for several distinct sequences was tethered to a passivated surface by a biotin–streptavidin interaction. The triplet repeat-containing DNA folds on itself into a hairpin configuration. The hairpin includes an acceptor fluorophore on its free end, and a donor is attached at a fixed position on the surface anchoring strand (Figure 2, blue and red dots). The distance between the donor and acceptor changes when the hairpin slips along its axis, allowing the FRET coupling to report the configuration of the hairpin. A prism-type total internal reflection single molecule fluorescence microscope [126] records the signals [123]. An example of data from a GTC hairpin is shown in Figure 2, where anticorrelated intensities from the donor and acceptor result from the spontaneous slipping of the hairpin. Studies of this sort, using different triplet repeat sequences, characterized the sliding dynamics of hairpins in repeat units, and quantitative analysis of the kinetics [127] can reveal the energetics of the process.

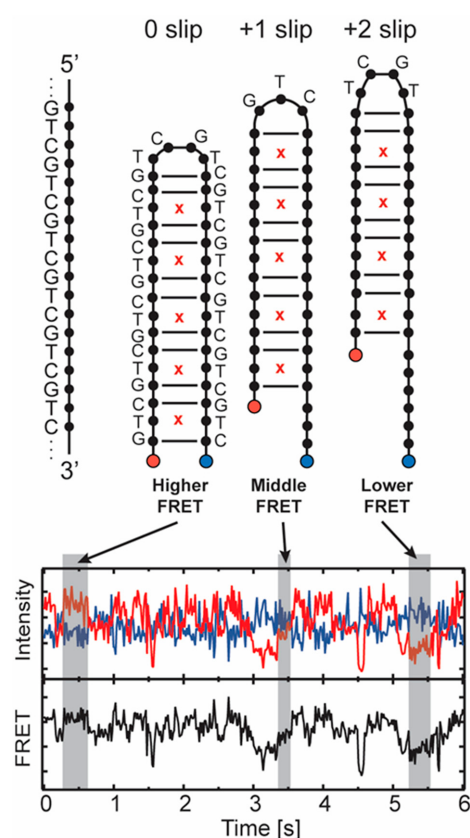


Figure 2. Example of smFRET assay for detecting slipped states of trinucleotide repeat DNA hairpins. On the left, a single-strand GTC repeat DNA is illustrated. To the right, 3 slipped states of the GTC

folded into a hairpin are displayed. The acceptor on the 5' end is shown in red, and the donor is shown in blue. T–T mismatches are shown as red x. The sequence in the loop is shown to illustrate tetraloop and triloop states that associate with the parity of the slipped state, but the stem sequence is not shown for the +1 and +2 slipped states for clarity. The different slipped states result in different donor–acceptor separations and different FRET values as indicated. The example measured data trace (**lower**) shows donor intensity (blue), acceptor intensity (red), and FRET efficiency ratio (black) vs. time. See Refs. [123,124] for more information.

The computational results are based on atomistic Molecular Dynamics (MD) and on our Adaptively Biased Molecular Dynamics (ABMD) method [128,129] (an umbrella sampling approach with a time-dependent potential), which has been combined with generalized Replica Exchange Molecular Dynamics [130] (REMD) along with a Steered Molecular Dynamics [131] procedure (SMD). The resulting codes allow for the efficient evaluation of biomolecular free energies and sampling of conformational space. These procedures ultimately can yield highly accurate free energy curves and equilibrium properties for trinucleotide repeat hairpins. The MD simulations can reveal details of the atomic configurations of the loops of the hairpins and elucidate how these configurations relate to the stability of the structures. The combination of smFRET and molecule dynamics studies help us understand how these hairpin structures contribute to populating states associated with trinucleotide expansions and ultimately connect to the pathogenesis of the related diseases. We explore details of the combination of these investigations below.

2.1. Hairpins Formed from CAG and GAC TRs

CAG TRs give rise to the largest group of neurodegenerative diseases, which may be a reflection of the fact that CAG tracts are overrepresented in the human genome. Thus, studies show [132] that CAG tracts of six or more repeats occur 1055 times in the human genome with 300 such tracts in the exons. CAG tracts of ten or more occur 136 times with 33 tracts in the protein-coding region. These numbers are to be contrasted with those associated with GAC tracts: these occur only sixteen times in the human genome with three in the exon region. The specific TRED expressed depends very much on the location of the CAG tract within a given gene. Thus, CAG repeats in the 5'-UTR region of the PPP2R2V gene give rise to spinocerebellar ataxia type 12, while CAG tracts in the exon are responsible for nine neurodegenerative disorders such as Huntington's disease (HD), bulbar and spinal atrophy (SBMA), dentatorubralpallidolysian atrophy (DRPLA) and several different spinocerebellar ataxias (SCAs). Generically, CAG repeats give rise to what is known as polyglutamine (polyQ) diseases [133], as these tracts all give rise to polyQ expansions. These diseases are associated with a critical TR number for the pathological expansions. For example, in HD, a repeat number less than about 34 is considered normal, while repeats in the 36 to 250 range are pathological, leading to disease expression. Although the pathologies of the CAG-associated TREDS are different, they all give rise to polyQ aggregates with cross-beta conformations ultimately associating with neuronal death [134,135].

It is interesting that GAC TR repeats are also associated with a specific set of diseases that are different from the CAG-associated TREDS. Unlike CAG expansion diseases that often involve the growth of CAG repeat tracks by tens to hundreds of CAG units, GAC diseases are associated with only small changes in the TR repeat number. Multiple epiphyseal dysplasia is caused by a one or two-unit increase in the (GAC)₅ tract within the gene for the human cartilage oligomeric matrix protein to a (GAC)₆ or (GAC)₇ tract [136]. In contrast, within the same gene, contraction of the (GAC)₅ tract by one GAC unit to (GAC)₄ causes pseudoachondroplasia [137,138]. It turns out that the specific structure of the GAC duplexes is dependent on the pH and the ionic strength [139]. As will be discussed later, experimental studies show that the loop dynamics associated with CAG and GAC are very different, which gives rise to their contrasting behavior and may relate to differences in disease phenomena.

2.1.1. CAG and GAC Homoduplexes

The behavior of the A–A mismatches in CAG and GAC homoduplexes determines the homoduplex structures. In CAG TRs, the Watson–Crick base pairs between the mismatches exhibit GpC steps, while in GAC TRs, the corresponding steps are CpG. MD simulations based on the AMBER simulation package [140] and augmented with special free energy methods (ABMD) [129] with carefully chosen collective variables show distinct characteristics of the homoduplexes [117,141]. Figures 3 and 4 display CAG and GAC-based free energy landscapes and corresponding configurations of nucleotides associated with the main minima. The free energy calculations show that the preferred duplex structure is characterized by A–A mismatches stacked inside the helical core with the nucleotide torsion angles in an anti–anti conformation for both RNA and DNA. This corresponds to torsion angles of about 180–200° degrees for RNA and high-anti—about 230–260° degrees—for DNA. Beyond this, the next most favorable conformation corresponds to nucleotides in an anti–syn configuration, which is followed by nucleotides in their syn–syn conformation. Sample free energy landscapes based on the torsion angle (χ) and a collective variable (Ω) probing for the swinging out of nucleotides from their helical core are shown in Figure 3, while Figure 4 shows corresponding sample conformations. It should be noted that for nucleotides in a given torsion angle conformation, several different hydrogen bond structures are possible and have been characterized [117]. These results are consistent with experimental studies on CAG-RNA [141]. We note that the differences between the DNA and RNA anti-conformations are due to the presence of an additional hydroxyl group associated with the RNA sugars. This hydroxyl group interacts with the RNA backbone in such a way as to reduce the range of torsion angles explored by the RNA homoduplex.

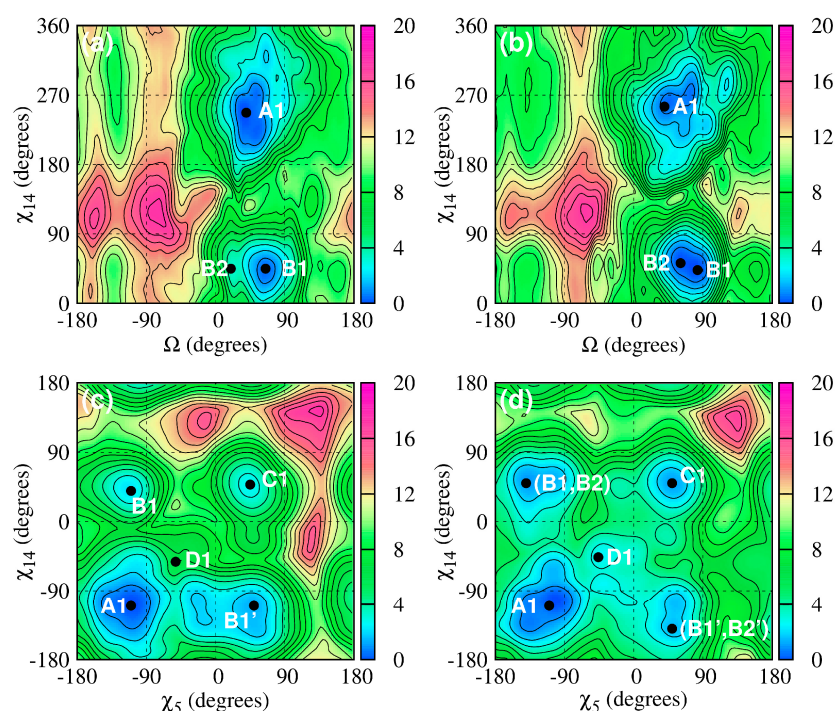


Figure 3. Shown here are sample free energy landscapes for a single A–A mismatch in a DNA-CAG and DNA-GAC. Here, (a) is (χ, Ω) for DNA-CAG; (b) (χ, Ω) for DNA-GAC; (c) (χ, χ) for DNA-CAG and (d) (χ, χ) for DNA-GAC. The letters mark the most important local minima with associated structures shown in Figure 4. The primed letters represent minima that are approximate mirror images of unprimed minima. Here, collective variable χ represents a dihedral angle of a given nucleotide and Ω an angle probing the motion of the nucleotide outside of its helical core. See Ref. [117] for more details.

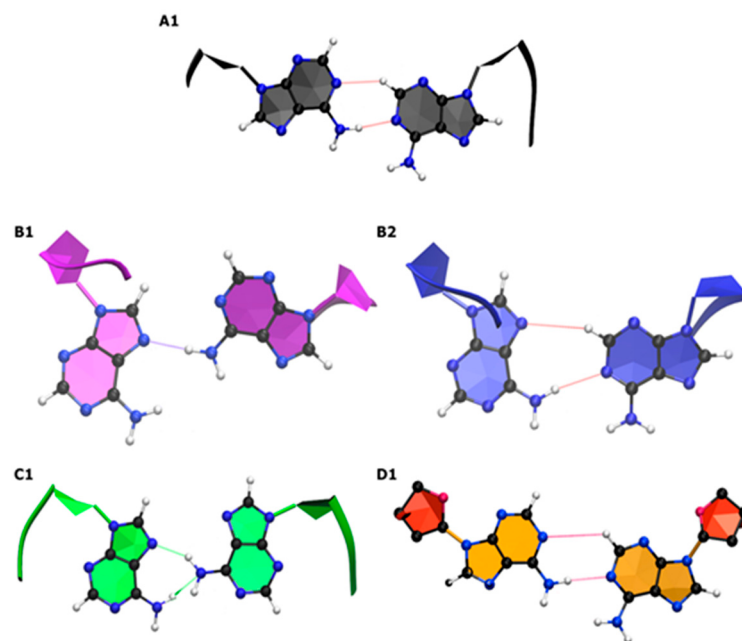


Figure 4. Shown here are sample A–A mismatch conformations associated with the primary minima on the free energy landscapes of Figure 3 with letters denoting different conformations (indicated on Figure 3). Here, (A1) is associated with anti–anti; (B1,B2) with syn–anti; (C1) with syn–syn; and (D1) is a special case in which the χ -angle is syn–syn, but the base conformation appears anti–anti given the sugar ring twisting to be parallel to the bases. Hydrogen bonds associated with the configurations are marked. For more details, please see Ref. [117].

The dynamics of CAG and GAC homoduplexes was characterized via the Principal Component Analysis (PCA) [142]. Large fluctuations have been observed for DNA in its global minimum structure, corresponding primarily to a coupling between the bending and unwinding modes of the helix. Similar fluctuations were found for the A-RNA, but to a lesser degree due to the already noted coupling between the sugar hydroxyl and backbone. Transitions between different nucleotide conformations, e.g., anti–syn to anti–anti, have also been investigated. These kinds of transitions involve local distortions around the A–A mismatches and involve combinations of base flipping, base stacking and base rotation in either the major or minor grooves [117]. Both CAG and GAC DNA homoduplexes experience some degree of unwinding with CAG unwinding occurring at the mismatch and GAC unwinding taking place at the CpG steps. There is, however, no evidence for the formation of transient left-handed helical structures such as Z-DNA.

2.1.2. Structure and Dynamics of CAG and GAC Loops

There is a delicate interplay between the free energies of the loop and stem parts of a TR hairpin which determines its behavior. Recently, it has become possible to probe this behavior by means of combined smFRET experiments and MD simulations [11,123,124]. smFRET was used to probe the parity-dependent slipping behavior in CAG and GAC hairpins, which has important ramifications for the loop structure. In the experiments, a two-stranded system was used involving an anchor and a hairpin strand with acceptor and donor fluorophores placed in suitable, consistent positions such that when the donor and acceptor come near each other, FRET signals of differing intensities are measured (Figure 2 for examples). This in turn signals the opening, closing, and slipping of the hairpin, and its kinetics may be inferred [123,124].

To probe CAG hairpin slippage, two hairpin structures (CAG)₁₄ and (CAG)₁₅ with an even and odd parity were constructed [123,124], and time traces of their FRET signals were analyzed [127]. See Figures 1 and 2 for the schematics associated with the general idea of this parity behavior and sample FRET data. For (CAG)₁₄, three FRET states of

varying efficiencies—0.01, 0.31, and 0.65—were observed. The 0.01 state is rarely visited. It could arise from opening of the hairpin or blinking of the acceptor fluorophore. The populations for the 0.31 and 0.65 states are about 20% and 80%, respectively. Turning to (CAG)₁₅, four different states with FRET efficiencies of 0.01, 0.25, 0.46 and 0.73 were observed. Again, populations for the open state (0.01) were negligible, while for the other states, they were estimated to be 5% for 0.25, 60% for 0.46 and 30% for 0.73, respectively. Subsequent analysis showed that the 0.73 state for (CAG)₁₅ is similar to the 0.65 state for (CAG)₁₄ and corresponds to the donor and acceptor fluorophores moving closer together by a single CAG unit, or a −1 slip. Likewise, the 0.46 state is associated with the hairpin slipping by one CAG unit away from the donor, corresponding to a +1 slip. So clearly, there is back and forth slippage for both strands by a single CAG TR unit. Structurally, for (CAG)₁₅, this allows for the formation of a 5'-AGCA-3' tetraloop with CAG/GAC aligned pairing for the duplex stem. Contrast this with a hairpin with the ends CAG/GAC aligned, i.e., 0 slip. This, in turn, leads to a triloop 5'-CAG-3' structure consisting of a single CAG unit. The smFRET and simulation results indicate that these kinds of loops are less stable than the tetraloops. Analyzing the remaining data in a similar fashion indicates that the 0.31 state of (CAG)₁₄ may be associated with slippage via +2 CAG units, and the 0.25 state of (CAG)₁₅ may be associated with a +2 unit shift [123,124].

This behavior turns out to be robust and forces an even/odd repeat parity onto hairpin slippage. Hairpins with an even number of repeats form a 5'-AGCA-3' tetraloop with a stem fully paired or slipped by two CAG units. By contrast, a hairpin with an odd number of repeats and a fully paired stem is associated with a 5'-CAG-3' triloop. Such conformations, however, are less stable, and the hairpin slips back and forth to form a 5'-AGCA-3' tetraloop with a hanging CAG TR unit in the stem. This differing slippage behavior between hairpins with even and odd repeat numbers illustrates nicely the interplay between loop and stem structure and agrees with the MD simulation results [123,124].

MD simulations have allowed for the direct study of the loop atomistics, which in turn explains their differing stabilities. Figure 5 shows sample loop configurations. The CAG triloop has nucleotides in their anti-conformation with the C-base being flipped out forming a sheared C–G pair characterized by a single hydrogen bond. The loop is then 'held together' by a weak AG/CA step with mismatched A's. The 5'-AGCA-3' tetraloop is more stable as it is held together by favorable stacking energy within the loop. There is also reduced bending and deformation of the backbone with the base held together by a locking CG/GC step. There is also more flexibility in terms of the base conformations of the loop; the AGCA nucleotides may be in either an anti-anti-anti-anti or an anti-anti-anti-syn conformation. In turn, these conformations display populations with a differing number of single and double base stacks within the loop [123]. How does the transition from tri- to tetra-loop take place? The simulations suggest that such transitions are triggered through the disruption of the A–A mismatch closest to the loop. The A base on the 3' strand turns toward the minor groove, forming a transient GACG tetraloop, which presumably is the first step toward the formation of a more stable tetraloop. Importantly, CAG hairpins within a configuration of a three-way junction where the hairpin occupies one arm and the other two arms are CAG-CTG DNA duplexes that permit short migrations of the CAG hairpin [143–145] display similar trends as the simple hairpins described above. The significance of the loop region on the stability of CAG hairpins is further confirmed by smFRET studies of CAG hairpins where the loop region is replaced by poly-A or poly-T linkers. These new sequences did not display the characteristic slipping behavior but rather transitioned between simple fully open and fully closed hairpin conformations [146].

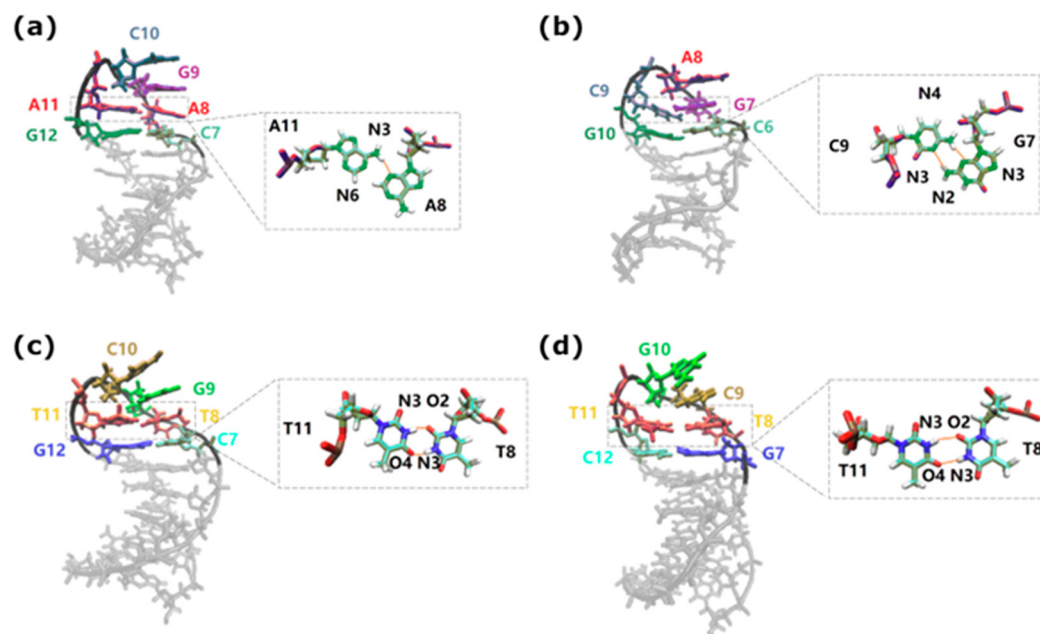


Figure 5. (a) A CAG tetraloop with the hairpin part in A(anti)–G(anti)–C(anti)–A(syn) form, with inset showing the hydrogen bond A8:N3–A11:N6 in the closing A–A mismatch. (b) GAC triloop with the hairpin part of G(anti)–A(anti)–C(syn) form, with inset showing the hydrogen bond G7:N2–C9:N3 and G7:N3–C9:N4 in the closing CG pair. (c) CTG tetraloop with the hairpin part in T(anti)–G(anti)–C(anti)–T(anti) with inset showing the hydrogen bond T8:N3–T11:N4 and T8:O2–T11:N3 in the closing T–T mismatch; (d) a GTC tetraloop with the hairpin part in T(anti)–G(anti)–C(syn)–T(anti) with inset showing the hydrogen bond T8:N3–T11:N4 and T8:O2–T11:N3 in the closing T–T mismatch. For more information, please see Refs. [123,124].

Having understood the basics of CAG slippage, it is interesting to probe how interrupts or mutations change both loop structure and dynamics. This is not just an academic question, as mutations and interrupts strongly influence disease expression. For instance, the point mutation $G \rightarrow A$ which changes CAG to CAA (both coding for the same amino acid glutamine) is known to have a stabilizing effect on HD [147,148], SCA17 [149], and SCA2 [150–152]. Mechanisms of how trinucleotide interrupts lessen TRED phenomena have been suggested to relate to stabilizing the dynamics of hairpins [153]. To explore the effects of CAA interrupts on CAG hairpin dynamics, smFRET experiments and MD simulations have been used to study $(CAG)_6(CAA)(CAG)_8$ and $(CAG)_7(CAA)(CAG)_8$ hairpins [11,123] in which the CAA interrupt is placed at or near the loop position. The experiments show that in the case of the former, the +1 slip is dramatically preferred for the former and the –1 slip is preferred for the latter. Both of these results are consistent with a tetraloop sequence of AACA closed by two G–C Watson–Crick bonds. Overall, the CAA interrupts in CAG TRs dramatically reduce strand slippage in the hairpins. MD simulations results are completely consistent with the experimental results. Essentially, for the $G \rightarrow A$ mutation, which changes the 5'-AGCA-3' tetraloop to 5'-AACA-3', the base stacking is preserved, and the stability considerations are not dramatically altered. However, placing the mutation elsewhere in or around the loop either adds mismatches or destabilizes the stacking, which ultimately gives a less stable hairpin. Similar stabilization has been observed with smFRET for CAA interrupts in CAG hairpins in the context of three-way junctions [143].

Other mutations have also been explored, and the results are briefly summarized here [123,124]. Experimentally, changing the middle CAG $\rightarrow$ AAA in $(CAG)_{15}$ ultimately gives rise to a dominant AAA triloop. MD simulations agree with these results, which may be understood in terms of stronger purine–purine (A/A) stacking over the pyrimidine–purine (C/A) stacking of CAG-based triloops. Other explored variants consider mutations

of CAG → CGG, CAG → CAC, and CAG → CTG. It is interesting to note that the former two mutations are in favor of positioning the mutated nucleotide around the loop, whereas the CAG → CTG mutation favors the configuration that avoids the hairpin loop and puts the thymine closer to the hairpin stem. In terms of CAG → CGG, smFRET experiments on (CAG)₆(CGG)(CAG)₈ are characterized by +1 or −1 slips, which is indicative of GGCA or AGCA tetraloops, respectively. The GGCA is favored over the AGCA loop in the experiment, although comparison is difficult because the alignments of bases in the stem are different, too. MD simulations support the stability of the GGCA tetraloop. Results based on (CAG)₆(CAC)(CAG)₈ hairpins show that an ACCA tetraloop is favored with this mutant. However, shifting the mutation around can lead to more complicated behavior with the possible formation of larger loops [123,124].

smFRET has also been used to study GAC hairpins [124], and its behavior is considerably different from that of the reverse sequence CAG hairpins. Both (GAC)₁₄ and (GAC)₁₅ hairpins were investigated, and the results show that GAC hairpins are most commonly found in a triloop configuration with slips of +1 or −1 being the most common. MD simulations show that the 5'-CAG-3' triloop prefers anti-anti-syn conformations for its hairpin. The hairpin is stabilized by the G/A stacking of the loop bases and a sheared G_{anti}-C_{syn} base pair with two hydrogen bonds. By contrast, the 5'-ACGA-3' tetraloop is characterized by fluctuations between a number of closely related states. The MD simulations indicate a preference for the anti-anti-anti-syn conformation with one persistent hydrogen bond in the A_{anti}-A_{syn} pair. The state has two possible conformations with either no stacking in the loop or one with a two-stack A/C and G/A state. The propensity of GAC to form triloops has also been observed in previous experimental studies based on polymerase extension [154]. The contrasting CAG and GAC hairpin structures likely influence the different expansion propensities of their repeated tracts seen in distinct diseases [41,42,136], but the mechanisms connecting these expansions to disease states remain opaque.

2.2. Hairpins Formed from CTG (CUG for RNA) and GTC (GUG for RNA) Repeats

Turning to the structure and dynamics of CTG expansions (CUG in RNA), we note that these expansions give rise to a number of TREDs, while GTC (GUC in RNA) do not exhibit such pathological expansions. In particular, myotonic dystrophy begins in adulthood [155,156] and is caused by CTG TRs (myotonic dystrophy type 1) or CCTG (myotonic dystrophy type 2) expansions. Thus, CTG expansions in the 3–38 range are considered normal, 39–50 repeats premutation are considered abnormal, and 50+ repeats lead to disease expression. For CTG, the TRs are located in the 3'-UTR of the dystrophin myotonic protein kinase gene; CCTG repeats are associated with the zinc finger (ZNF9) gene [157]. There is considerable evidence that d(CTG) and r(CUG) form hairpin structures with the latter being associated with a toxic mRNA gain-of-function [23]. Experimentally, RNA CUG TRs are known to form homoduplexes in an A-RNA form dominated by their U–U mismatches. These are quite dynamic associated with relatively large fluctuations and are therefore known as “wobble U–U” mismatches [158–161].

2.2.1. CTG (CUG for RNA) and GTC (GUC for RNA) Homoduplexes

The structure and dynamics of d(CTG)/r(CUG) and d(GTC)/r(GUC) homoduplexes have been investigated both experimentally [158–161] and theoretically with MD simulations [119,162]. For d(CTG)/r(CUG), the mismatches are separated with a Watson–Crick base pair GpC step, while CpG steps separate mismatches in d(GTC)/r(GUC). The studies show that the global free energy minima associated with the T–T and U–U mismatches correspond to anti-anti conformations inside the helical core. While both mismatches are very dynamic, the fluctuations associated with the U–U mismatch are larger than the T–T fluctuations. The mismatches themselves are characterized by structures with the hydrogen bond number varying between zero and two. Beyond the global minimum, the next minima correspond to anti-syn or syn-anti conformations followed by syn-syn. As with CAG/GAC TR structures, the anti-anti U–U conformations in r(CUG)/r(GUC)

duplexes are within their regular range, while those corresponding to d(CTG)/d(GTC) are in the “high-anti” range [119].

Dynamical simulations of the homoduplexes show that the T–T and U–U mismatches can flex between a large number of conformations without significantly altering the global helical structure. PCA analysis of the fluctuations of the first eigenvector around the backbone displays a coupling between bending and unwinding modes, which is stronger in the RNA helices. Dominating the mismatched base-pair dynamics are opening, shear and stretching modes. For RNA, there is also a widening of the major groove with under twisting. There is also a decrease in the inclination angle with respect to the standard A-RNA and B-DNA structures [119].

2.2.2. Structure and Dynamics of CTG and GTC Hairpins

CTG and GTC hairpins have been studied with smFRET in a manner similar to the CAG and GAC hairpins as discussed above [124,163]. Additionally, CTG has also been investigated with Nuclear Magnetic Resonance (NMR) [164], and a unified picture of the hairpin structure has emerged. Perhaps not surprising, the structure and behavior of CTG loops resemble those of CAG loops, while GTC loops are more like those with GAC. There are, however, some differences.

A pattern emerges in which CAG, CTG and GTC prefer tetraloop configurations with GTC being more tolerant of triloops; GAC, as already mentioned, prefers a triloop configuration. Both CTG and GTC display parity-dependent behavior based on strand length. Here, we simply give some of the results [124]: (CTG)₁₄ hairpins spent 85% of the time in a 0 slip configuration indicative of a tetraloop, 2% of time in a −2 slip also indicative of a tetraloop and only 4% in a +1 slip consistent with a triloop. Similarly, for (CTG)₁₅, it was observed that hairpins were characterized by slips of +1 (33%) and −1 (57%) consistent with tetraloop structures and only 7% in a triloop configuration. Turning to GTC loops, (GTC)₁₄ displays +1 and −1 slips, but in this case, these are indicative of triloops with populations of 8% and 17%, respectively. However, 60% of the time, the hairpin is in a zero slip or +2 slip (6%) configuration with a tetraloop; for (GTC)₁₅, the data show slips of +1 (26%) and −1 (50%) with tetraloops and 26% with a zero slip with a triloop. Thus, both CAG and CTG hairpins predominantly slip by 2 TR units, jumping from tetraloop to tetraloop. GAC and GTC hairpins, on the other hand, slip primarily by a single unit jumping between tetraloops and triloops, respectively.

Turning to the structural results obtained from the MD simulations [124], it has been noted that for CTG, the 5′-CTG-3′ triloop resembles the 5′-CAG-3′ loop and is characterized by a sheared G–C pair held together by a single hydrogen bond. The triloops are subject to large fluctuations leading to an unraveling of the loop. By contrast, the 5′-TGCT-3′ tetraloop is much more stable and was observed to have either an anti–anti–anti–anti or an anti–syn–anti–anti conformation with differing stacking. The loop is stabilized by two hydrogen bonds between the T_{anti}–T_{anti} mismatches and the flanking GpC step. The NMR studies agree with these results, showing that a structure based on four CTG repeats and clamped with a single GC pair gives rise to a stable hairpin with a TGCT loop [164].

For GTC, the simulations [124] indicate that the triloop 5′-GTC-3′ may exist in anti–anti–anti or syn–anti–anti conformations. However, unlike in the GAC triloop which has a sheared G_{anti}–C_{syn} pair with two hydrogen bonds, there is no hydrogen bond between the G–C pair. The triloop is closed by an TC/GT step and therefore considerably weaker. The 5′-TCGT-3′ tetraloop is stabilized in two configurations: anti–anti–anti–anti and anti–anti–syn–anti. This loop is stabilized by two hydrogen bonds between the T_{anti}–T_{anti} pair, which differs from that of the GAC tetraloop characterized by a single hydrogen bond between the A–A mismatched loop pairs. This feature, combined with the stacking arrangements, stabilizes the GTC tetraloop over the GAC counterpart.

smFRET investigations of CAG hairpins with CAA interrupts showed that there was a dramatic stabilization of slipped states when the G is mutated to A. For CTG, the corresponding experiments involve changing C to T, and so the effects of a TTG interrupt

were examined [124]. This interrupt was placed near the loop, so that (CTG)₆(TTG)(CTG)₈, (CTG)₇(TTG)(CTG)₇ and (CTG)₈(TTG)(CTG)₆ hairpins were studied. For the latter two, switching is very much reduced over the original (CTG)₁₅ hairpin. The (CTG)₇(TTG)(CTG)₇ hairpin is dominated by a single configuration at 0.54 FRET with a +1 slipped state, while (CTG)₈(TTG)(CTG)₆ has a state at 0.75 FRET consistent with both 0 and −1 slip states. These states all correspond to a tetraloop with the interrupting T being placed in the loop. The behavior of (CTG)₆(TTG)(CTG)₈ is somewhat more complex, but it too is consistent with a tetraloop configuration with the additional T placed in the loop [124].

2.3. Hairpins Formed from C and G Exclusive Trinucleotides

Here, we discuss hairpins formed from triplet repeat sequences containing only C and G. The most common examples are CGG and CCG TRs, which are overexpressed in the human genome. CGG TRs may be found in the 5'-UTR of the fragile X mental retardation gene (FMR1) [165]; similarly, TRs of CCG are also to be found in the 5'-UTR of several genes. The normal repeat range for CGG TRs is from about 5 to 54; a higher number of repeats leads to increased disease expression, causing fragile X syndrome (FXS) [166,167]. Specifically, a repeat length in the 55 to 200 range is associated with fragile X tremor ataxia syndrome (FXTAS) in males [168] and premature ovarian failure in females [169]. Longer repeats of 200 or more give rise to inherited fragile X mental retardation syndrome [170]. Likewise, CCG TRs are related to three TREDs, the longest expansion of which in the FRM2 gene gives rise to chromosome X-linked mental retardation (FRAXE) [171]. These repeats have also been implicated in HD and type 1 myotonic dystrophy [172].

2.3.1. CGG, GGC, CCG and GCC Homoduplexes

The structural and dynamical characteristics of the C-rich and G-rich TR homoduplexes for DNA and RNA have been investigated, as these are characteristic of hairpin stems [118]. As with other TR-based duplexes, the properties of CGG (GGC) and CCG (GCC) TR hairpins are dominated by the behavior of G–G and C–C mismatches. Here, it is important to consider the Watson–Crick pairs surrounding the mismatches. Homoduplexes made from CGG and CCG are characterized by GpC steps between the mismatches, while CpG steps surround the mismatches in the GGC and GCC alignments. Free energy studies [118] show that the global minimum associated with the C–C mismatches in both DNA and RNA CCG or GCC correspond to the anti–anti configuration. Next, in terms of increasing free energy are anti–syn conformations followed by syn–syn. In this regard, the dihedral structure of the C–C mismatches is similar to the A–A and T–T (U–U for RNA mismatches in CAG/GAC and CTG/GTC TRs). However, the global free energy minimum associated with G–G mismatches within GGC and CGG TRs corresponds to the anti–syn conformation. In terms of the free energy, the next higher minimum corresponds to anti–anti, which is followed by syn–syn. We note that for RNA-CCG, it has been difficult to unambiguously resolve the actual ground state structure, as the anti–syn and anti–anti structures appear to degenerate within the limits of the calculations [118]. This difference appears to be due to the presence of a significant triple G-base stacking, which is not present in the other structures. Experimentally, there are crystallographic data for RNA CGG and CCG TRs—which have GpC steps between the mismatches—and these agree with the free energy results [155,173,174].

Direct MD simulations of the different homoduplexes in different conformations allow for the exploration of the transitions to the global minimum conformations [118]. On average, duplexes with C–C mismatches tend to transition to the anti–anti state faster than those with G–G mismatches. An examination of the MD runs shows that the additional stacking of the G-bases is what tends to slow down the transition. Another important feature is that for DNA GCC homoduplexes, the mismatched Cs may be extruded from the helical core to form a so-called “e-motif” discussed in the next section. Finally, a qualitative comparison of the relative stability of the mismatched homoduplexes via a fast laser melting

technique [175] indicated that DNA ordering stability is as follows: $G_{\text{rich}}(\text{CpG}) > C_{\text{rich}}(\text{GpC}) > G_{\text{rich}}(\text{GpC})$, which is in agreement with the currently available experimental data [118].

2.3.2. E-Motif Formed by C- and G-Rich DNA Homoduplexes

In the last section, the structure and dynamical properties of homoduplexes with C–C and G–G mismatches are briefly summarized. It turns out, however, that the DNA homoduplexes can form structures that are considerably more complex by extruding the mismatches from their helical core in a well-defined manner to form the so-called “e-motif” [176], as shown in Figure 6. E-motifs were first discovered experimentally in an antiparallel DNA 5'-(CCG)₂-3' duplex via NMR (PDB 1NOQ). In this duplex, slippage conspired to form a central C–C mismatch surrounded by CpG Watson–Crick pairs [176]. These mismatched Cs flip out in the minor groove such that their base moieties point in the 5' direction on each of the strands. See Figure 6 for sample e-motif configurations as well as contrasting nucleotides extruded from the helical core not in such a configuration. Since then, e-motif like structures have been observed in other C-rich DNA strands as well as other more complex structures such as an i-motif [177–179].

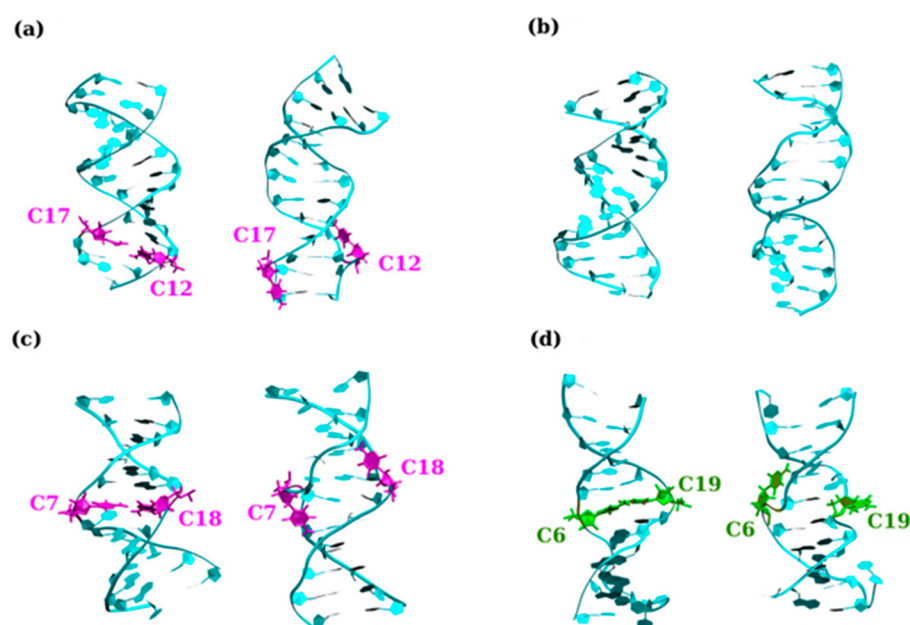


Figure 6. Sample configurations illustrating the e-motif as obtained with MD simulations. Initial (left) and final (right) structures are shown for (a) GCC4; (b) CCG4; (c) DC-1; and (d) DC-2. The C–C bases that ultimately form the e-motif are shown in purple. Bases shown in green are bases that are flipped out of the inner DNA helix but ultimately do not form an e-motif. See Ref. [122] for details.

Theoretically, e-motif structures were observed to form during MD simulations of DNA d(CCCCGG) hexanucleotide repeats [114]. Motivated by these observations, other C-rich duplexes were subsequently examined in order to determine which sequences give rise to e-motifs either in an isolated or an extended form [122]. Specifically, the nonequivalent reading frames (GCC)_n and (CCG)_n for TRs and (CCCGGC)_n, (CGGCCC)_n and (CCCGGG)_n for hexanucleotide repeats were examined. The results show for the basic e-motif, the C-bases of the *i*th residue in a mismatch, flip out in a symmetrical manner toward the *i*-2 residue (in the 5' direction of each strand). The e-motif is stabilized by hydrogen bonds between the flipped out base and the *i*-2 base along the strand. This may be either a C-base (GCC) or a G-base for (CCCGGC). E-motif creation is made favorable through the creation of pseudo-GpC steps between non-adjacent base pairs when the C–C mismatches are extruded. As a result, the e-motif is stable in homoduplexes with GCC and CCCGGC but not in other reading frames [114]. It is interesting that an extended e-motif with all the mismatched Cs may also be formed. Such an arrangement is stabilized by the

favorable stacking provided by the pseudo-GpC steps and the hydrogen bonds between the extruded Cs and the other nucleotides as well as a stacking between the Cs themselves. Ultimately, this results in a very stable but somewhat peculiar secondary structure.

While the e-motif structure formation is a feature of DNA GCC duplexes, it is not a characteristic of RNA-based duplexes. In our simulations, the C-bases in RNA duplexes were observed to flip out in a transient manner, but an e-motif was never formed. This is due to the presence of the additional hydroxyl group on RNA which tends to form hydrogen bonds with nearby sugars and the backbone. Additionally, the canonical A-form of RNA prevents good stacking for either GpC or CpG steps [118,122].

The e-motif is primarily associated with the extrusion of C-bases in B-DNA; it has recently been shown that e-like motifs are also associated with the extrusion of G-bases in left-handed Z-DNA, leading to the formation of a novel eGZ-motif [115]. Experimental CD spectroscopy experiments demonstrate that under high salt conditions, CGG may adopt a left-handed conformation [180,181]. MD simulations were used to examine all possible Z-DNA helices that could form after such a B- to Z-DNA transition [115]. It turns out that for such structures, the GG mismatch conformations may be arranged either intra- or extra-helical conformations, leading to the formation of BZ or ZZ junctions, or in an alternating extruded conformation. The most stable conformation is associated with alternately extruded Gs followed by a helix with symmetrically extruded Gs forming a ZZ junction. The extruded Gs in the left-handed helices are primarily in a syn conformation with favorable hydrogen bonds and stacking interactions, forming novel eGZ-motif structures [115].

2.3.3. Structure and Dynamics of CCG and GGC Hairpins

CGG hairpins, both with and without AGG interrupts, have recently been studied with smFRET [182]. As with the hairpin structures already discussed above, the CGG hairpins exhibit even/odd parity behavior. It was found that CGG repeats predominantly form a so-called “blunt-end” structure consistent with a tetraloop structure. Thus, $(\text{CGG})_n$ with n -even is believed to form a 5'-GGCG-3' tetraloop with no overhangs and a matched stem, while n -odd forms tetraloop 5'-GGCG-3' with a single G overhang in the stem. There is also evidence for the formation of a small quadruplex population [182].

It is important to note that AGG interrupts in CGG TRs play an important role in suppressing FXS gene expression. AGG interrupts are found in most healthy individuals, occurring roughly every 9-10 CGG repeats [183], while those with a family history of FXS tend not to have such interrupts [184]. The recent mapping of CGG alleles shows that there is a strong effect of AGG interruptions on CGG TR expansion [185,186].

Given the importance of AGG in suppressing FXS disease expression, its effect was investigated by placing such interrupts in various location of a CGG-based hairpin. Placing the interrupt at the loop—i.e., $(\text{CGG})_9\text{AGG}(\text{CGG})_9$ —leads to a dynamic reconfiguration of the loop from a 5'-GAGG-3' loop with a single G overhang in the stem to a 5'-GGAG-3' loop with three nucleotide GGC (i.e., one TR unit, from end) overhang. Since AGG interruptions in healthy individuals occur rather frequently, there are additional strands such as $(\text{CGG})_4\text{AGG}(\text{CGG})_n\text{AGG}(\text{CGG})_4$ with n standing for the number of CGG repeats separating the interrupts. For these kinds of strands, a somewhat modified parity-dependent behavior is observed. For odd- n , data are consistent with a tetraloop hairpin structure 5'-GGCG-3' with a three-nucleotide overhang at one terminus. However, there now is an GGAG/GGAG bulge formed by the two AGG interrupts in the CGG hairpin stem. For n -even, a 5'-CGG-3' triloop appears to be formed with a similar bulge at the two interrupts as well as a GGC overhang in the stem end. The exception is when $n = 8$ for which the flanking $(\text{CGG})_4$ participates to form a “dumbbell”-type structure with two GAGG tetraloops at both ends [187]. Thus, the presence of AGG interrupts favors the formation of the more stable 5'-GGAG-3' tetraloop and/or the formation of stable loops within the CGG-based homoduplexes. These in turn are believed to restrict hairpin reconfiguration dynamics and inhibit TR repeat expansion.

3. Discussion

The structure and dynamics of trinucleotide repeat DNA hairpins are determined by a subtle balance between stem and loop interaction free energies. Single base changes arising from interrupts in the sequence pattern can have major impacts on these behaviors. The fact that these interruptions also have a substantial biological influence on disease-related expansion phenomena strongly suggests that the dynamics of trinucleotide repeat hairpins play a key role in these disease mechanisms [188].

Here, we have described recent progress using atomistic molecular dynamics simulations and smFRET experiments to characterize these atypical DNA structures and dynamics. Parity-dependent effects and strongly preferred slipping in one or two trinucleotide repeat units in different systems have been characterized. These reflect the preference for certain configurations of the loop that restrict stem alignment freedom. Different behaviors among distinct sequence variants highlight the impact of stacking energies for adjacent bases [11,117–124].

These observations motivate models of how the dynamics of three-way junctions and cruciforms might contribute to trinucleotide expansions. See Figure 7 for a schematic. Genomic expansions likely involve the opening of the complementary duplex at some point during replication, transcription or repair-related processes. This duplex opening provides opportunities for hairpin-like structures to form from the single-stranded repeat regions folding on themselves. These hairpins would initially form opposite from each other in the opened duplex and have a high likelihood to collapse into each other, reforming the duplex. In some cases, the different slipping dynamics of the complementary sequences could allow them to migrate apart with different preferred directions or at different speeds [124,145]. Once they are isolated, they could persist sufficiently long to be engaged by the proteins that have been identified as involved in expansion phenomena. In these sorts of models, the dynamics of the hairpins are fundamental in determining the details of disease phenomena.

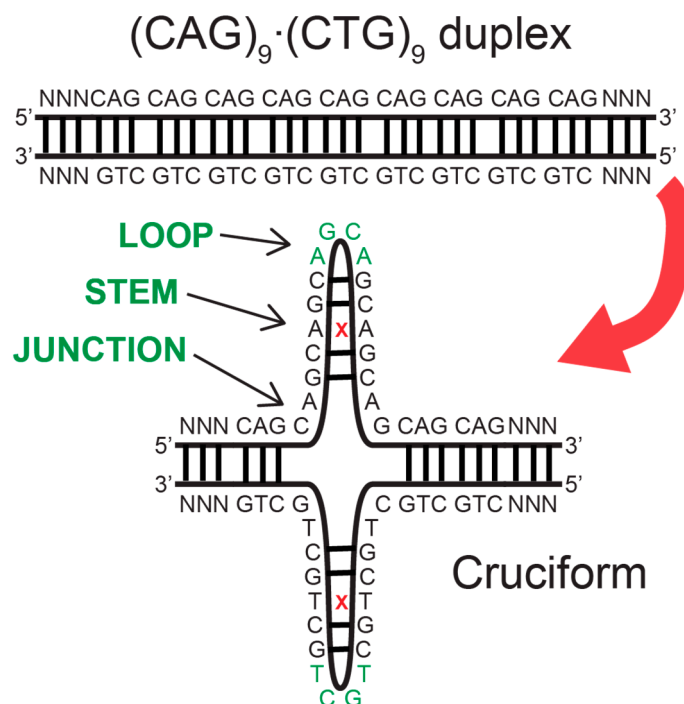


Figure 7. Schematic of a trinucleotide repeat containing duplex DNA (**upper**) opening into a cruciform arrangement (**lower**). The loop, stem, and junction region of the hairpin loop in the cruciform are indicated. The mismatched base pair in the stem is indicated by the red x. See Ref. [124] for more information.

Structures formed from human disease-associated triplet repeats are dynamic with common themes being hairpins slipping in steps of multiples of the repeat unit. Strong evidence that the dynamics of these structures are significant for disease processes was recently provided by the discovery that a small molecule, naphthyridine-azaquinolone (NA), that binds to CAG hairpins, suppresses their slipping and also is curative for Huntington's disease in mouse models [57]. Developing methods may provide windows into the formation of these structures inside living cells [189,190]. GWAS studies, transgenic mouse models, and other methods implicate several proteins, especially Fan1 [191–196] and mismatch repair (MMR) proteins in the mechanisms that lead to the disease-associated expansion of TNR in the human genome [76,197]. A critical frontier is discovering the interplay between the conformational dynamics of these atypical DNA structures and the proteins that are causative for expansion. Despite our lack of mechanistic understanding of how these proteins are involved, therapeutics for many of these diseases are already in development that pursue strategies to alter these protein–DNA interactions [198,199].

The largest family of proteins implicated in promoting TNR expansion is MMR proteins. DNA MMR's primary function is to repair rare single base mismatches or short insertion/deletion loops that escape polymerase proofreading during replication [69,70]. The MutS α variant (MSH2-MSH6) targets the repair of mismatched bases, whereas the MutS β variant (MSH2-MSH3) preferentially initiates the repair of loops [69,70]. The knock-out and GWAS data (above) implicate DNA binding activity of several proteins with most attention focusing on MutS β and the DNA backbone nicking function of MutL γ , but the actual events resulting in expansion are unknown. Recent cryo-EM studies of MutS β revealed multiple classes of conformations on short (CAG)₂ loops and longer CAG hairpins [200,201]. These states inspired models of possible event sequences but have not yet incorporated any interaction with MutL variants.

Several experimental threads have suggested a spectrum of ideas about detailed mechanisms promoting expansions that include MutS β -induced DNA bending [202] and the recruitment of distinct MutL variants [203]. A common theme in these models is a dynamic CAG hairpin that can change size and migrate [57,124,163,204]. Hairpin dynamics can generate asymmetric structures on the complementary strands of duplex DNA that can be differentially modified by MMR proteins. Precisely when and where the MMR proteins nick the DNA leads to distinct outcomes.

Currently, many therapeutic approaches are focused on affecting MSH3, which impacts MutS β but not MutS α with the intent of maintaining most MMR activity [198,199,205]. Gene therapy and editing, antibodies and other biologics, as well as small molecules are all being developed and tested for impacts on TNR expansion disorders in cell and mouse models as well as in a few human trials [98,199,206,207]. Increased understanding of the dynamics of TNR hairpins and the impact on hairpin dynamics on the interactions they have with MMR proteins will enhance the targeting of these sorts of efforts to develop approaches to treat and prevent these diseases.

Determining how proteins promote somatic TNR expansion is challenging because (i) the precise TNR-containing DNA structure that recruits MMR proteins during expansion remains unconfirmed (with exciting new participants emerging occasionally [208]), (ii) the DNA structures that form and the DNA–protein interactions are dynamic and transient with a myriad of possible complexes possible, and (iii) these protein interactions are regulated by coordinated ATP dependencies that are difficult to synchronize in biochemical studies. Despite the difficulties, major advances will follow uncovering these DNA configurations, protein complexes, interactions, and ATPase regulatory mechanisms. Such efforts will be foundational to guide searches for TNR expansion disease therapeutics that seek to modify them to treat or prevent the disease.

Author Contributions: Writing—original draft preparation, C.R.; writing—review and editing, F.P., P.X., C.R., C.S. and K.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation (NSF) grant number MCB-2409309.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created for this review.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Wells, R.D.; Ashizawa, T. *Genetic Instabilities and Neurological Diseases*; Elsevier: Boston, MA, USA, 2006.
- Ellegren, H. Microsatellites: Simple Sequences with Complex Evolution. *Nat. Rev. Genet.* **2004**, *5*, 435–445. [[CrossRef](#)] [[PubMed](#)]
- Kashi, Y.; King, D.; Soller, M. Simple Sequence Repeats as a Source of Quantitative Genetic Variation. *Trends Genet.* **1997**, *13*, 74–78. [[CrossRef](#)] [[PubMed](#)]
- Caburet, S.; Cocquet, J.; Vaiman, D.; Veitia, R. Coding Repeats and Evolutionary “Agility”. *BioEssays* **2005**, *27*, 581–587. [[CrossRef](#)] [[PubMed](#)]
- Fleischer, B. Über myototische Dystrophie mit Katarakt. *Albrecht Von Graefes Arch. Klin. Exp. Optalmol.* **1918**, *96*, 905–914.
- Oberle, I.; Rouseau, F.; Heit, D.; Devy, D.; Zengerling, S.; Mandel, J. Molecular-Basis of the Fragile-X Syndrome and Diagnostic Applications. *Am. J. Hum. Genet.* **1991**, *49*, 76.
- Campuzano, V.; Montermini, L.; Molt, M.; Pianese, L.; Cossee, M.; Cavalcanti, F.; Monros, E.; Rodius, F.; Duclos, F.; Monticelli, A. Friedreich’s Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion. *Science* **1996**, *271*, 1423–1427. [[CrossRef](#)]
- Mirkin, S.M. DNA Structures, Repeat Expansions and Human Hereditary Disorders. *Curr. Opin. Struct. Biol.* **2006**, *16*, 351–358. [[CrossRef](#)]
- Mirkin, S.M. Expandable DNA Repeats and Human Disease. *Nature* **2007**, *447*, 932–940. [[CrossRef](#)]
- Pearson, C.E.; Edamura, K.N.; Cleary, J.D. Repeat Instability: Mechanisms of Dynamic Mutations. *Nat. Rev. Genet.* **2005**, *6*, 729–742. [[CrossRef](#)]
- Pan, F.; Zhang, Y.; Xu, P.; Man, V.H.; Roland, C.; Weninger, K.; Sagui, C. Molecular Conformations and Dynamics of Nucleotide Repeats Associated with Neurodegenerative Diseases: Double Helices and CAG Hairpin Loops. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2819–2832. [[CrossRef](#)]
- Toth, G.; Gaspari, Z.; Jurka, J. Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Res.* **2000**, *10*, 967–981. [[CrossRef](#)] [[PubMed](#)]
- Borstnik, B.; Pumpernik, D. Tandem repeats in protein coding regions of primate genes. *Genom. Res.* **2002**, *12*, 909–915. [[CrossRef](#)]
- Bacolla, A.; Larson, J.; Collins, J.; Li, J.; Milosavljevic, A.; Stenson, P.; Cooper, D.; Wells, R. Abundance and Length of Simple Repeats in Vertebrate Genomes Are Determined by Their Structural Properties. *Gen. Res.* **2008**, *18*, 1545–1553. [[CrossRef](#)]
- Clark, R.; Bhaskar, S.; Miyahar, M.; Dalgliesh, G.; Bidichandani, S. Expansion of GAA Trinucleotide Repeats in Mammals. *Genomics* **2006**, *87*, 57–67. [[CrossRef](#)] [[PubMed](#)]
- Galka-Marciniak, P.; Urbanek, M.O.; Krzyzosiak, W.J. Triplet Repeats in Transcripts: Structural Insights into RNA Toxicity. *Biol. Chem.* **2012**, *393*, 1299–1315. [[CrossRef](#)] [[PubMed](#)]
- McLaughlin, B.; Spencer, C.; Eberwine, J. CAG Trinucleotide RNA Repeats Interact with RNA- Binding Proteins. *Am. J. Hum. Genet.* **1996**, *59*, 561–569.
- Timchenko, L.; Timchenko, N.; Caskey, C.; Roberts, R. Novel Proteins with Binding Specificity for DNA CTG Repeats and RNA CUG Repeats: Implications for Myotonic Dystrophy. *Hum. Mol. Genet.* **1996**, *5*, 115–121. [[CrossRef](#)]
- Sobczak, K.; Mezer, M.; Michlewski, G.; Krol, J.; Krzyzosiak, W. RNA Structure of Trinucleotide Repeats Associated with Human Neurological Diseases. *Nucleic Acids Res.* **2003**, *31*, 5469–5482. [[CrossRef](#)]
- Timchenko, L. Myotonic Dystrophy: The Role of RNA CUG Triplet Repeats. *Am. J. Hum. Genet.* **1999**, *64*, 360–364. [[CrossRef](#)]
- Mankodi, A.; Takahashi, M.; Jiang, H.; Beck, C.; Bowers, W.; Moxley, R.; Cannon, S.; Thornton, C. Expanded CUG Repeats Trigger Aberrant Splicing of CIC-1 Chloride Channel Pre-mRNA and Hyperexcitability of Skeletal Muscle in Myotonic Dystrophy. *Mol. Cell* **2002**, *10*, 35–44. [[CrossRef](#)]
- Lu, X.; Timchenko, N.; Timchenko, L. Cardiac Delav-Type RNA-Binding Protein (ETR-3) Binds to RNA CUG Repeats Expanded in Myotonic Dystrophy. *Hum. Mol. Genet.* **1999**, *8*, 53–60. [[CrossRef](#)] [[PubMed](#)]
- Philips, A.; Timchenko, L.; Cooper, T. Disruption of Splicing Regulated by a CUG-Binding Protein in Myotonic Dystrophy. *Science* **1998**, *280*, 737–741. [[CrossRef](#)] [[PubMed](#)]
- Tapscott, S.; Thornton, C. Biomedicine—Reconstructing myotonic dystrophy. *Science* **2001**, *293*, 816–817. [[CrossRef](#)] [[PubMed](#)]
- Charlet, B.N.; Savkur, R.; Singh, G.; Philips, A.; Grice, E.; Cooper, T. Loss of the Chloride Channel in DM1 Skeletal Muscle Due to Misregulated Alternative Splicing: A Likely Cause of Myotonia. *Eur. J. Hum. Genet.* **2002**, *10*, 257–258.
- Amack, J.; Mahadevan, M. The Myotonic Dystrophy Expanded CUG Repeat Tract Is Necessary but Not Sufficient to Disrupt C2C12 Myoblast Differentiation. *Hum. Mol. Genet.* **2021**, *10*, 1879–1887. [[CrossRef](#)]
- Amack, J.; Reagan, S.; Mahadevan, M. Mutant DMPK 3’-UTR Transcripts Disrupt C2C12 Myogenic Differentiation by Compromising MyoD. *J. Cell Biol.* **2002**, *159*, 419–429. [[CrossRef](#)]

28. Savkur, R.; Philips, A.; Cooper, T. Aberrant Regulation of Insulin Receptor Alternative Splicing Is Associated with Insulin Resistance in Myotonic Dystrophy. *Nat. Genet.* **2001**, *29*, 40–47. [[CrossRef](#)]
29. Pearson, C. Repeat Associated Non-ATG Translation Initiation: One DNA, Two Transcripts, Seven Reading Frames, Potentially Nine Toxic Entities. *PLoS Genet.* **2011**, *7*, e1002018. [[CrossRef](#)]
30. Castelli, L.; Huang, W.-P.; Chang, K.-Y.; Hautbergue, G. Mechanism of Repeat Associated Non-AUG Translation in Neurological Microsatellite Expansion Disorders. *Biochem. Soc. Trans.* **2021**, *49*, 775–792. [[CrossRef](#)]
31. Timchenko, N.; Welm, A.; Lu, X.; Timchenko, L. CUG Repeat Binding Protein (CUGBP1) Interacts with the 5' Region of C/EBP Beta mRNA and Regulates Translation of C/EBP Beta Isoforms. *Nucleic Acids Res.* **1999**, *27*, 4517–4525. [[CrossRef](#)]
32. Wells, R.D.; Warren, S. *Genetic Instabilities and Neurological Diseases*; Academic Press: San Diego, CA, USA, 1998.
33. Orr, H.; Zoghbi, H. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* **2007**, *30*, 575. [[CrossRef](#)] [[PubMed](#)]
34. Pearson, C.; Sinden, R. Slipped Strand DNA (S-DNA and SI-DNA), Trinucleotide Repeat Instability and Mismatch Repair: A Short Review. In Proceedings of the Structure, Motion, Interaction and Expression of Biological Macromolecules, Albany, NY, USA, 17–21 June 1998; Sarma, R., Sarma, M., Eds.; US NIH, SUNY: Albany, NY, USA, 1998; Volume 2, pp. 191–207.
35. Paulson, H. Repeat Expansion Diseases. *Handb. Clin. Neurol.* **2018**, *147*, 105–123. [[PubMed](#)]
36. Iyer, R.R.; Pluciennik, A. DNA Mismatch Repair and Its Role in Huntington's Disease. *J. Huntingtons Dis.* **2021**, *10*, 75–94. [[CrossRef](#)] [[PubMed](#)]
37. Lee, J.M.; Correia, K.; Loupe, J.; Kim, K.H.; Barker, D.; Hong, E.P.; Chao, M.J.; Long, J.D.; Lucente, D.; Vonsattel, J.P.G.; et al. CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell* **2019**, *178*, 887–900. [[CrossRef](#)] [[PubMed](#)]
38. Goldberg, Y.P.; Telenius, H.; Hayden, M.R. The Molecular Genetics of Huntington's Disease. *Curr. Opin. Neurol.* **1994**, *7*, 325–332. [[CrossRef](#)] [[PubMed](#)]
39. MacDonald, M.E.; Ambrose, C.M.; Duyao, M.P.; Myers, R.H.; Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S.A.; James, M.; Groot, N.; et al. A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **1993**, *72*, 971–983. [[CrossRef](#)]
40. Greco, T.M.; Secker, C.; Ramos, E.S.; Federspiel, J.D.; Liu, J.P.; Perez, A.M.; Al-Ramahi, I.; Cantle, J.P.; Carroll, J.B.; Botas, J.; et al. Dynamics of Huntingtin Protein Interactions in the Striatum Identifies Candidate Modifiers of Huntington Disease. *Cell Syst.* **2022**, *13*, 304–320. [[CrossRef](#)]
41. Malik, I.; Kelley, C.P.; Wang, E.T.; Todd, P.K. Molecular Mechanisms Underlying Nucleotide Repeat Expansion Disorders. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 589–607. [[CrossRef](#)]
42. Bunting, E.L.; Hamilton, J.; Tabrizi, S.J. Polyglutamine Diseases. *Curr. Opin. Neurobiol.* **2022**, *72*, 39–47. [[CrossRef](#)]
43. Tenchov, R.; Sasso, J.M.; Zhou, Q.A. Polyglutamine (PolyQ) Diseases: Navigating the Landscape of Neurodegeneration. *ACS Chem. Neurosci.* **2024**, *15*, 2665–2694. [[CrossRef](#)]
44. Bonsor, M.; Ammar, O.; Schnoegl, S.; Wanker, E.E.; Silva Ramos, E. Polyglutamine Disease Proteins: Commonalities and Differences in Interaction Profiles and Pathological Effects. *Proteomics* **2024**, *24*, 2300114. [[CrossRef](#)] [[PubMed](#)]
45. Kim, J.; Mirkin, S. The Balancing Act of DNA Repeat Expansions. *Curr. Opin. Genet. Devel.* **2013**, *23*, 280–288. [[CrossRef](#)]
46. Khristich, A.N.; Mirkin, S.M. On the Wrong DNA Track: Molecular Mechanisms of Repeat-Mediated Genome Instability. *J. Biol. Chem.* **2020**, *295*, 4134–4170. [[CrossRef](#)] [[PubMed](#)]
47. Pearson, C.E.; Sinden, R.R. Trinucleotide Repeat DNA Structures: Dynamic Mutations from Dynamic DNA. *Curr. Opin. Struct. Biol.* **1998**, *8*, 321–330. [[CrossRef](#)] [[PubMed](#)]
48. Wells, R.D.; Dere, R.; Hebert, M.L.; Napierala, M.; Son, L.S. Advances in Mechanisms of Genetic Instability Related to Hereditary Neurological Diseases. *Nucleic Acids Res.* **2005**, *33*, 3785–3798. [[CrossRef](#)]
49. Cleary, J.; Walsh, J.; Hofmeister, J.; Shanka, G.; Kuskowski, M.; Selkoe, D.; Ashe, K. Natural Oligomers of the Amyloid-Protein Specifically Disrupt Cognitive Function. *Nat. Neurosci.* **2005**, *8*, 79–84. [[CrossRef](#)]
50. López Castel, A.; Cleary, J.D.; Pearson, C.E. Repeat Instability as the Basis for Human Diseases and as a Potential Target for Therapy. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 165–170. [[CrossRef](#)]
51. Dion, V.; Wilson, J. Instability and Chromatin Structure of Expanded Trinucleotide Repeats. *Trends Genet.* **2009**, *25*, 288–297. [[CrossRef](#)]
52. McMurray, C.T. Hijacking of the Mismatch Repair System to Cause CAG Expansion and Cell Death in Neurodegenerative Disease. *DNA Repair* **2008**, *7*, 1121–1134. [[CrossRef](#)]
53. McMurray, C.T. DNA Secondary Structure: A Common and Causative Factor for Expansion in Human Disease. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1823–1825. [[CrossRef](#)]
54. Pearson, C.E.; Sinden, R.R. Alternative Structures in Duplex DNA Formed within the Trinucleotide Repeats of the Myotonic Dystrophy and Fragile X Loci. *Biochemistry* **1996**, *35*, 5041–5053. [[CrossRef](#)]
55. Schmidt, M.H.M.; Pearson, C.E. Disease-Associated Repeat Instability and Mismatch Repair. *DNA Repair* **2016**, *38*, 117–126. [[CrossRef](#)] [[PubMed](#)]
56. Panigrahi, G.B.; Lau, R.; Montgomery, S.E.; Leonard, M.R.; Pearson, C.E. Slipped (CTG) $\bullet$ (CAG) Repeats Can Be Correctly Repaired, Escape Repair or Undergo Error-Prone Repair. *Nat. Struct. Mol. Biol.* **2005**, *12*, 654–662. [[CrossRef](#)]

57. Nakamori, M.; Panigrahi, G.B.; Lanni, S.; Gall-Duncan, T.; Hayakawa, H.; Tanaka, H.; Luo, J.; Otabe, T.; Li, J.; Sakata, A.; et al. A Slipped-CAG DNA-Binding Small Molecule Induces Trinucleotide-Repeat Contractions in Vivo. *Nat. Genet.* **2020**, *52*, 146–159. [[CrossRef](#)] [[PubMed](#)]
58. Ranum, L.P.W.; Cooper, T.A. RNA-mediated neuromuscular disorders. *Ann. Rev. Neurosci.* **2006**, *6*, 259–277. [[CrossRef](#)]
59. Daughters, R.; Tuttle, D.; Gao, W.; Ikeda, Y.; Moseley, M.; Ebner, T.; Swanson, M.; Ranum, L. RNA Gain-of-Function in Spinocerebellar Ataxia Type 8. *PLoS Genet.* **2009**, *5*, e1000600. [[CrossRef](#)] [[PubMed](#)]
60. Li, B.; Bonini, N. Roles of Trinucleotide-Repeat RNA in Neurological Disease and Degeneration. *Trends Neurosci.* **2010**, *33*, 292–298. [[CrossRef](#)] [[PubMed](#)]
61. Jin, P.; Zarnescu, D.; Zhang, F.; Pearson, C.; Lucchesi, J.; Moses, K.; Warren, S. RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Neuron* **2003**, *39*, 739–747. [[CrossRef](#)]
62. Jiang, H.; Mankodi, A.; Swanson, M.; Moxley, R.; Thornton, C. Myotonic Dystrophy Type 1 Is Associated with Nuclear Foci of Mutant RNA, Sequestration of Muscle Blind Proteins and Deregulated Alternative Splicing in Neurons. *Hum. Mol. Genet.* **2004**, *13*, 3079–3088. [[CrossRef](#)]
63. Krzyzosiak, W.J.; Sobczak, K.; Wojciechowska, M.; Fiszer, A.; Mykowska, A.; Kozłowski, P. Triplet Repeat RNA Structure and Its Role as Pathogenic Agent and Therapeutic Target. *Nucleic Acids Res.* **2012**, *40*, 11–26. [[CrossRef](#)]
64. Campuzano, V.; Montermini, L.; Lutz, Y.; Cova, L.; Hindelan, C.; Jiralerspong, S.; Trottier, Y.; Kish, S.; Faucheux, B.; Trouillas, P.; et al. Frataxin Is Reduced in Friedreich Ataxia Patients and Is Associated with Mitochondrial Membranes. *Hum. Mol. Genet.* **1997**, *6*, 1771–1780. [[CrossRef](#)] [[PubMed](#)]
65. Kim, E.; Napierala, M.; Dent, S. Hyperexpansion of GAA Repeats Affects Post-Initiation Steps of FXN Transcription in Friedreich's Ataxia. *Nucl. Acids Res.* **2011**, *39*, 8366–8377. [[CrossRef](#)] [[PubMed](#)]
66. Kumari, D.; Biacsi, R.; Usdin, K. Repeat Expansion in Intron 1 of the Frataxin Gene Reduces Transcription Initiation in Friedreich Ataxia. *FASEB J.* **2011**, *25*, 895.2. [[CrossRef](#)]
67. Punga, T.; Buehler, M. Long Intronic GAA Repeats Causing Friedreich Ataxia Impede Transcription Elongation. *EMBO Mol. Med.* **2010**, *2*, 120–129. [[CrossRef](#)]
68. Iyer, R.R.; Pluciennik, A.; Burdett, V.; Modrich, P.L. DNA Mismatch Repair: Functions and Mechanisms. *Chem. Rev.* **2006**, *106*, 302–323. [[CrossRef](#)] [[PubMed](#)]
69. Kunkel, T.A.; Erie, D.A. DNA Mismatch Repair. *Annu. Rev. Biochem.* **2005**, *74*, 681–710. [[CrossRef](#)] [[PubMed](#)]
70. Kunkel, T.A.; Erie, D.A. Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu. Rev. Genet.* **2015**, *49*, 291–313. [[CrossRef](#)] [[PubMed](#)]
71. Modrich, P. Mechanisms in Eukaryotic Mismatch Repair. *J. Biol. Chem.* **2006**, *281*, 30305–30309. [[CrossRef](#)]
72. Schofield, M.J.; Hsieh, P. DNA Mismatch Repair: Molecular Mechanisms and Biological Function. *Annu. Rev. Microbiol.* **2003**, *57*, 579–608. [[CrossRef](#)]
73. McMurray, C.T. Mechanisms of Trinucleotide Repeat Instability during Human Development. *Nat. Rev. Genet.* **2010**, *11*, 786–799. [[CrossRef](#)]
74. Goellner, G.; Tester, D.; Thiboddeau, S.; Almqvist, E.; Goldberg, Y.; Hayden, M.; McMurray, C. Different Mechanisms Underlie DNA Instability in Huntington Disease and Colorectal Cancer. *Am. J. Hum. Genet.* **1997**, *60*, 879. [[PubMed](#)]
75. Iyer, R.R.; Pluciennik, A.; Napierala, M.; Wells, R.D. DNA Triplet Repeat Expansion and Mismatch Repair. *Annu. Rev. Biochem.* **2015**, *84*, 199–226. [[CrossRef](#)] [[PubMed](#)]
76. Rajagopal, S.; Donaldson, J.; Flower, M.; Hensman Moss, D.J.; Tabrizi, S.J. Genetic Modifiers of Repeat Expansion Disorders. *Emerg. Top. Life Sci.* **2023**, *7*, 325–337. [[PubMed](#)]
77. Jones, L.; Houlden, H.; Tabrizi, S.J. DNA Repair in the Trinucleotide Repeat Disorders. *Lancet Neurol.* **2017**, *16*, 88–96. [[CrossRef](#)] [[PubMed](#)]
78. Owen, B.A.L.; Yang, Z.; Lai, M.; Gajek, M.; Badger, J.D.; Hayes, J.J.; Edelmann, W.; Kucherlapati, R.; Wilson, T.M.; McMurray, C.T. (CAG)_n-Hairpin DNA Binds to Msh2–Msh3 and Changes Properties of Mismatch Recognition. *Nat. Struct. Mol. Biol.* **2005**, *12*, 663–670. [[CrossRef](#)]
79. Warren, J.J.; Pohlhaus, T.J.; Changela, A.; Iyer, R.R.; Modrich, P.L.; Beese, L.S. Structure of the Human MutS α DNA Lesion Recognition Complex. *Mol. Cell* **2007**, *26*, 579–592. [[CrossRef](#)]
80. Gupta, S.; Gellert, M.; Yang, W. Mechanism of Mismatch Recognition Revealed by Human MutS β Bound to Unpaired DNA Loops. *Nat. Struct. Mol. Biol.* **2011**, *19*, 72–79. [[CrossRef](#)]
81. Hsieh, P.; Yamane, K. DNA Mismatch Repair: Molecular Mechanism, Cancer, and Ageing. *Mech. Ageing Dev.* **2008**, *129*, 391–407. [[CrossRef](#)]
82. Karran, P.; Offman, J.; Bignami, M. Human Mismatch Repair, Drug-Induced DNA Damage, and Secondary Cancer. *Biochimie* **2003**, *85*, 1149–1160. [[CrossRef](#)]
83. Meier, B.; Volkova, N.V.; Hong, Y.; Schofield, P.; Campbell, P.J.; Gerstung, M.; Gartner, A. Mutational Signatures of DNA Mismatch Repair Deficiency in C. Elegans and Human Cancers. *Genome Res.* **2018**, *28*, 666–675. [[CrossRef](#)]
84. Fishel, R. The Selection for Mismatch Repair Defects in Hereditary Nonpolyposis Colorectal Cancer: Revising the Mutator Hypothesis. *Cancer Res.* **2001**, *61*, 7369–7374. [[PubMed](#)]
85. Li, G.M. DNA Mismatch Repair and Cancer. *Front. Biosci.* **2003**, *8*, d997–d1017. [[CrossRef](#)] [[PubMed](#)]

86. Modrich, P.; Lahue, R. Mismatch Repair in Replication Fidelity, Genetic Recombination, and Cancer Biology. *Annu. Rev. Biochem.* **1996**, *65*, 101–133. [[CrossRef](#)] [[PubMed](#)]
87. Muller, A.; Fishel, R. Mismatch Repair and the Hereditary Non-Polyposis Colorectal Cancer Syndrome (HNPCC). *Cancer Investig.* **2002**, *20*, 102–109. [[CrossRef](#)] [[PubMed](#)]
88. Heinen, C.D. Mismatch Repair Defects and Lynch Syndrome: The Role of the Basic Scientist in the Battle against Cancer. *DNA Repair* **2016**, *38*, 127–134. [[CrossRef](#)]
89. Martín-López, J.V.; Fishel, R. The Mechanism of Mismatch Repair and the Functional Analysis of Mismatch Repair Defects in Lynch Syndrome. *Fam. Cancer* **2013**, *12*, 159–168. [[CrossRef](#)]
90. Wyatt, M.D.; Pittman, D.L. Methylating Agents and DNA Repair Responses: Methylated Bases and Sources of Strand Break. *Chem. Res. Toxicol.* **2006**, *19*, 1580. [[CrossRef](#)]
91. Roos, W.P.; Kaina, B. DNA Damage-Induced Cell Death by Apoptosis. *Trends Mol. Med.* **2006**, *12*, 440. [[CrossRef](#)]
92. Kaina, B. DNA Damage-Triggered Apoptosis: Critical Role of DNA Repair, Double-Strand Breaks, Cell Proliferation and Signaling. *Biochem. Pharmacol.* **2003**, *66*, 1547–1554. [[CrossRef](#)]
93. Li, Z.; Pearlman, A.; Hsieh, P. DNA Mismatch Repair and the DNA Damage Response. *DNA Repair* **2016**, *38*, 94–101. [[CrossRef](#)]
94. Gupta, D.; Heinen, C. The Mismatch Repair-Dependent DNA Damage Response: Mechanism and Implication. *DNA Repair* **2019**, *78*, 60–69. [[CrossRef](#)] [[PubMed](#)]
95. Bettencourt, C.; Hensman-Moss, D.; Flower, M.; Wiethoff, S.; Brice, A.; Goizet, C.; Stevanin, G.; Koutsis, G.; Karadima, G.; Panas, M.; et al. DNA Repair Pathways Underlie a Common Genetic Mechanism Modulating Onset in Polyglutamine Diseases. *Ann. Neurol.* **2016**, *79*, 983–990. [[CrossRef](#)] [[PubMed](#)]
96. Ciosi, M.; Maxwell, A.; Cumming, S.A.; Hensman Moss, D.J.; Alshammari, A.M.; Flower, M.D.; Durr, A.; Leavitt, B.R.; Roos, R.A.C.; Holmans, P.; et al. A Genetic Association Study of Glutamine-Encoding DNA Sequence Structures, Somatic CAG Expansion, and DNA Repair Gene Variants, with Huntington Disease Clinical Outcomes. *EBioMedicine* **2019**, *48*, 568–580. [[CrossRef](#)]
97. Lee, J.M.; Wheeler, V.C.; Chao, M.J.; Vonsattel, J.P.G.; Pinto, R.M.; Lucente, D.; Abu-Elneel, K.; Ramos, E.M.; Mysore, J.S.; Gillis, T.; et al. Identification of Genetic Factors That Modify Clinical Onset of Huntington’s Disease. *Cell* **2015**, *162*, 516–526. [[CrossRef](#)] [[PubMed](#)]
98. Ferguson, R.; Goold, R.; Coupland, L.; Flower, M.; Tabrizi, S.J. Therapeutic Validation of MMR-Associated Genetic Modifiers in a Human Ex Vivo Model of Huntington Disease. *Am. J. Hum. Genet.* **2024**, *111*, 1165–1183. [[CrossRef](#)] [[PubMed](#)]
99. Pinto, R.M.; Dragileva, E.; Kirby, A.; Lloret, A.; Lopez, E.; St. Claire, J.; Panigrahi, G.B.; Hou, C.; Holloway, K.; Gillis, T.; et al. Mismatch Repair Genes Mlh1 and Mlh3 Modify CAG Instability in Huntington’s Disease Mice: Genome-Wide and Candidate Approaches. *PLoS Genet.* **2013**, *9*, 1003930. [[CrossRef](#)]
100. Dragileva, E.; Hendricks, A.; Teed, A.; Gillis, T.; Lopez, E.T.; Friedberg, E.C.; Kucherlapati, R.; Edelmann, W.; Lunetta, K.L.; MacDonald, M.E.; et al. Intergenerational and Striatal CAG Repeat Instability in Huntington’s Disease Knock-in Mice Involve Different DNA Repair Genes. *Neurobiol. Dis.* **2009**, *33*, 37–47. [[CrossRef](#)]
101. Tomé, S.; Manley, K.; Simard, J.P.; Clark, G.W.; Slean, M.M.; Swami, M.; Shelbourne, P.F.; Tillier, E.R.M.; Monckton, D.G.; Messer, A.; et al. MSH3 Polymorphisms and Protein Levels Affect CAG Repeat Instability in Huntington’s Disease Mice. *PLoS Genet.* **2013**, *9*, e1003280. [[CrossRef](#)]
102. Nakatani, R.; Nakamori, M.; Fujimura, H.; Mochizuki, H.; Takahashi, M.P. Large Expansion of CTG•CAG Repeats Is Exacerbated by MutS β in Human Cells. *Sci. Rep.* **2015**, *5*, 11020. [[CrossRef](#)]
103. Bourn, R.L.; de Biase, I.; Pinto, R.M.; Sandi, C.; Al-Mahdawi, S.; Pook, M.A.; Bidichandani, S.I. Pms2 Suppresses Large Expansions of the (GAA•TTC) $_n$ Sequence in Neuronal Tissues. *PLoS ONE* **2012**, *7*, e47085. [[CrossRef](#)]
104. Zhao, X.N.; Kumari, D.; Gupta, S.; Wu, D.; Evanitsky, M.; Yang, W.; Usdin, K. Muts β Generates Both Expansions and Contractions in a Mouse Model of the Fragile X-Associated Disorders. *Hum. Mol. Genet.* **2015**, *24*, 7087–7096. [[CrossRef](#)] [[PubMed](#)]
105. Manley, K.; Shirley, T.L.; Flaherty, L.; Messer, A. MSH2 Deficiency Prevents in Vivo Somatic Instability of the CAG Repeat in Huntington Disease Transgenic Mice. *Nat. Genet.* **1999**, *23*, 471. [[CrossRef](#)] [[PubMed](#)]
106. Kovtun, I.V.; McMurray, C.T. Trinucleotide Expansion in Haploid Germ Cells by Gap Repair. *Nat. Genet.* **2001**, *27*, 407–411. [[CrossRef](#)] [[PubMed](#)]
107. van den Broek, W.J.; Nelen, M.R.; Wansink, D.G.; Coerwinkel, M.M.; te Riele, H.; Groenen, P.J.; Wieringa, B. Somatic Expansion Behaviour of the (CTG) $_n$ Repeat in Myotonic Dystrophy Knock-in Mice Is Differentially Affected by Msh3 and Msh6 Mismatch-Repair Proteins. *Hum. Mol. Genet.* **2002**, *11*, 191–198. [[CrossRef](#)] [[PubMed](#)]
108. Gonitell, R.; Moffitt, H.; Sathasivam, K.; Woodman, B.; Detloff, P.J.; Faull, R.L.M.; Bates, G.P. DNA Instability in Postmitotic Neurons. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3467–3472. [[CrossRef](#)]
109. Moore, H.; Greenwell, P.W.; Liu, C.P.; Arnheim, N.; Petes, T.D. Triplet Repeats Form Secondary Structures That Escape DNA Repair in Yeast. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1504–1509. [[CrossRef](#)]
110. Polak, U.; Mcivor, E.; Dent, S.Y.R.; Wells, R.D.; Napierala, M. Expanded Complexity of Unstable Repeat Diseases. *BioFactors* **2013**, *39*, 164–175. [[CrossRef](#)]
111. Usdin, K.; House, N.C.M.; Freudenreich, C.H. Repeat Instability during DNA Repair: Insights from Model Systems. *Crit. Rev. Biochem. Mol. Biol.* **2015**, *50*, 142–167. [[CrossRef](#)]

112. Wojciechowska, M.; Napierala, M.; Larson, J.E.; Wells, R.D. Non-B DNA Conformations Formed by Long Repeating Tracts of Myotonic Dystrophy Type 1, Myotonic Dystrophy Type 2, and Friedreich's Ataxia Genes, Not the Sequences per Se, Promote Mutagenesis in Flanking Regions. *J. Biol. Chem.* **2006**, *281*, 24531–24543. [[CrossRef](#)]
113. Zhang, Y.; Roland, C.; Sagui, C. Structural and Dynamical Characterization of DNA and RNA Quadruplexes Obtained from the GGGGCC and GGGCCT Hexanucleotide Repeats Associated with C9FTD/ALS and SCA36 Diseases. *ACS Chem. Neurosci.* **2018**, *9*, 1104–1117. [[CrossRef](#)]
114. Zhang, Y.; Roland, C.; Sagui, C. Structure and Dynamics of DNA and RNA Double Helices Obtained from the GGGGCC and CCCCCG Hexanucleotide Repeats That Are the Hallmark of C9FTD/ALS Diseases. *ACS Chem. Neurosci.* **2017**, *8*, 578–591. [[CrossRef](#)]
115. Fakharzadeh, A.; Zhang, J.; Roland, C.; Sagui, C. Novel EGZ-Motif Formed by Regularly Extruded Guanine Bases in a Left-Handed Z-DNA Helix as a Major Motif behind CGG Trinucleotide Repeats. *Nucleic Acids Res.* **2022**, *50*, 4860–4876. [[CrossRef](#)] [[PubMed](#)]
116. Zhang, J.; Fakharzadeh, A.; Roland, C.; Sagui, C. RNA as a Major-Groove Ligand: RNA-RNA and RNA-DNA Triplexes Formed by GAA and UUC or TTC Sequences. *ACS Omega* **2022**, *7*, 38728–38743. [[CrossRef](#)] [[PubMed](#)]
117. Pan, F.; Man, V.H.; Roland, C.; Sagui, C. Structure and Dynamics of DNA and RNA Double Helices of CAG and GAC Trinucleotide Repeats. *Biophys. J.* **2017**, *113*, 19–36. [[CrossRef](#)] [[PubMed](#)]
118. Pan, F.; Man, V.H.; Roland, C.; Sagui, C. Structure and Dynamics of DNA and RNA Double Helices Obtained from the CCG and GGC Trinucleotide Repeats. *J. Phys. Chem. B* **2018**, *122*, 4491–4512. [[CrossRef](#)]
119. Fakharzadeh, A.; Qu, J.; Pan, F.; Sagui, C.; Roland, C. Structure and Dynamics of DNA and RNA Double Helices Formed by d(CTG), d(GTC), r(CUG) and r(GUC) Trinucleotide Repeats and Associated DNA-RNA Hybrids. *J. Phys. Chem. B* **2023**, *127*, 7907. [[CrossRef](#)]
120. Zhang, J.; Fakharzadeh, A.; Pan, F.; Roland, C.; Sagui, C. Atypical Structures of GAA/TTC Trinucleotide Repeats Underlying Friedreich's Ataxia: DNA Triplexes and RNA/DNA Hybrids. *Nucleic Acids Res.* **2020**, *48*, 9899–9917. [[CrossRef](#)]
121. Zhang, J.; Fakharzadeh, A.; Pan, F.; Roland, C.; Sagui, C. Construction of DNA/RNA Triplex Helices Based on GAA/TTC Trinucleotide Repeats. *Bio Protoc.* **2021**, *11*, e4155. [[CrossRef](#)]
122. Pan, F.; Zhang, Y.; Man, V.H.; Roland, C.; Sagui, C. E-Motif Formed by Extrahelical Cytosine Bases in DNA Homoduplexes of Trinucleotide and Hexanucleotide Repeats. *Nucleic Acids Res.* **2018**, *46*, 942–955. [[CrossRef](#)]
123. Xu, P.; Pan, F.; Roland, C.; Sagui, C.; Weninger, K. Dynamics of Strand Slippage in DNA Hairpins Formed by CAG Repeats: Roles of Sequence Parity and Trinucleotide Interrupts. *Nucleic Acids Res.* **2020**, *48*, 2232–2245. [[CrossRef](#)]
124. Xu, P.; Zhang, J.; Pan, F.; Mahn, C.; Roland, C.; Sagui, C.; Weninger, K. Frustration Between Preferred States of Complementary Trinucleotide Repeat DNA Hairpins Anticorrelates with Expansion Disease Propensity. *J. Mol. Biol.* **2023**, *435*, 168086. [[CrossRef](#)]
125. Weiss, S. Measuring Conformational Dynamics of Biomolecules by Single Molecule Fluorescence Spectroscopy. *Nat. Struct. Biol.* **2000**, *7*, 724–729. [[CrossRef](#)]
126. Roy, R.; Hohng, S.; Ha, T. A Practical Guide to Single-Molecule FRET. *Nat. Methods* **2008**, *5*, 507–516. [[CrossRef](#)]
127. Götz, M.; Barth, A.; Bohr, S.S.-R.; Börner, R.; Chen, J.; Cordes, T.; Erie, D.A.; Gebhardt, C.; Hadzic, M.C.A.S.; Hamilton, G.L.; et al. A Blind Benchmark of Analysis Tools to Infer Kinetic Rate Constants from Single-Molecule FRET Trajectories. *Nat. Commun.* **2022**, *13*, 5402. [[CrossRef](#)]
128. Moradi, M.; Babin, V.; Sagui, C.; Roland, C. Recipes for Free Energy Calculations in Biomolecular Systems. *Methods Mol. Biol.* **2013**, *924*, 313–337.
129. Babin, V.; Roland, C.; Sagui, C. Adaptively Biased Molecular Dynamics for Free Energy Calculations. *J. Chem. Phys.* **2008**, *128*, 134101. [[CrossRef](#)]
130. Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional Replica-Exchange Method for Free-Energy Calculations. *J. Chem. Phys.* **2000**, *113*, 6042–6051. [[CrossRef](#)]
131. Izrailev, S.; Stepaniants, S.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. Steered Molecular Dynamics. In Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling, Berlin, Germany, 21–24 May 1997; pp. 39–65.
132. Kozlowski, P.; Mezer, M.; Krzyzisiak, W. Trinucleotide Repeats in Human Genome and Exome. *Nucl. Acids Res.* **2010**, *38*, 4039. [[CrossRef](#)]
133. Zoghbi, H.Y.; Orr, H.T. Glutamine Repeats and Neurodegeneration. *Annu. Rev. Neurosci.* **2000**, *23*, 217–247. [[CrossRef](#)]
134. Sikorski, P.; Atkins, E. New Model for Crystalline Polyglutamine Assemblies and Their Connection with Amyloid Fibrils. *Biomacromolecules* **2005**, *6*, 425–432. [[CrossRef](#)]
135. Man, V.H.; Roland, C.; Sagui, C. Structural Determinants of Polyglutamine Protofibrils and Crystallites. *ACS Chem. Neurosci.* **2015**, *6*, 632–645. [[CrossRef](#)] [[PubMed](#)]
136. Délot, E.; King, L.M.; Briggs, M.D.; Wilcox, W.R.; Cohn, D.H. Trinucleotide Expansion Mutations in the Cartilage Oligomeric Matrix Protein (COMP) Gene. *Hum. Mol. Genet.* **1999**, *8*, 123–128. [[CrossRef](#)] [[PubMed](#)]
137. Hecht, J.T.; Nelson, L.D.; Crowder, E.; Wang, Y.; Elder, F.F.B.; Harrison, W.R.; Francomano, C.A.; Prange, C.K.; Lennon, G.G.; Deere, M.; et al. Mutations in Exon 17B of Cartilage Oligomeric Matrix Protein (COMP) Cause Pseudoachondroplasia. *Nat. Genet.* **1995**, *10*, 325–329. [[CrossRef](#)] [[PubMed](#)]

138. Briggs, M.D.; Mortier, G.R.; Cole, W.G.; King, L.M.; Golik, S.S.; Bonaventure, J.; Nuytinck, L.; De Paepe, A.; Leroy, J.G.; Biesecker, L.; et al. Diverse Mutations in the Gene for Cartilage Oligomeric Matrix Protein in the Pseudoachondroplasia-Multiple Epiphyseal Dysplasia Disease Spectrum. *Am. J. Hum. Genet.* **1998**, *62*, 311–319. [\[CrossRef\]](#) [\[PubMed\]](#)
139. Vorlíková, M.; Kejnovská, I.; Tumorová, M.; Kypr, J. Conformational Properties of DNA Fragments Containing GAC Trinucleotide Repeats Associated with Skeletal Dysplasias. *Eur. Biophys. J.* **2001**, *30*, 179–185. [\[CrossRef\]](#)
140. Case, D.A.; Belfon, K.; Ben-Shalom, I.Y.; Brozell, S.R.; Cerutti, D.S.; Cheatham, T.E., III; Cruzeiro, V.W.D.; Darden, T.A.; Duke, R.E.; Giambasu, G.; et al. *AMBER 20*; University of California: San Francisco, CA, USA, 2000.
141. Yildirim, I.; Park, H.; Disney, M.D.; Schatz, G.C.A. Dynamic Structural Model of Expanded RNA CAG Repeats: A Refined X-Ray Structure and Computational Investigations Using Molecular Dynamics and Umbrella Sampling Simulations. *J. Am. Chem. Soc.* **2012**, *40*, 4273–4287. [\[CrossRef\]](#)
142. Amadei, A.; Linssen, A.B.; Berendsen, H.J.C. Essential Dynamics of Proteins. *Proteins Struct. Funct. Genet.* **1993**, *17*, 412–425. [\[CrossRef\]](#)
143. Cadden, G.; Wilkens, S.; Magennis, S. A Single CAA Interrupt in a DNA Three-Way Junction Containing a CAG Repeat Hairpin Results in Parity-Dependent Trapping. *Nucl. Acids Res.* **2024**, *52*, 9317–9327. [\[CrossRef\]](#)
144. Bianco, S.; Hu, T.; Henrich, O.; Magennis, S. Heterogeneous migration routes of DNA triplet repeat slip-outs. *Biophys. Rep.* **2022**, *2*, 100070. [\[CrossRef\]](#)
145. Hu, T.; Morten, M.J.; Magennis, S.W. Conformational and Migrational Dynamics of Slipped-Strand DNA Three-Way Junctions Containing Trinucleotide Repeats. *Nat. Commun.* **2021**, *12*, 204. [\[CrossRef\]](#)
146. Mitchell, M.; Leveille, M.; Solecki, R.; Tran, T.; Cannon, B. Sequence-Dependent Effect of Monovalent Cations on the Structural Dynamics of Trinucleotide-Repeat DNA Hairpins. *J. Phys. Chem. B* **2018**, *122*, 11841–11851. [\[CrossRef\]](#) [\[PubMed\]](#)
147. Wright, G.E.B.; Black, H.F.; Collins, J.A.; Gall-Duncan, T.; Caron, N.S.; Pearson, C.E.; Hayden, M.R. Interrupting Sequence Variants and Age of Onset in Huntington’s Disease: Clinical Implications and Emerging Therapies. *Lancet Neurol.* **2020**, *19*, 930–939. [\[CrossRef\]](#) [\[PubMed\]](#)
148. Wright, G.E.B.; Collins, J.A.; Kay, C.; McDonald, C.; Dolzhenko, E.; Xia, Q.; Bećanović, K.; Drögemöller, B.I.; Semaka, A.; Nguyen, C.M.; et al. Length of Uninterrupted CAG, Independent of Polyglutamine Size, Results in Increased Somatic Instability, Hastening Onset of Huntington Disease. *Am. J. Hum. Genet.* **2019**, *104*, 1116–1126. [\[CrossRef\]](#) [\[PubMed\]](#)
149. Gao, R.; Matsuura, T.; Coolbaugh, M.; Zuhlke, C.; Nakamura, K.; Rasmussen, A.; Siciliano, M.; Ashizawa, T.; Lin, X. Instability of Expanded CAG/CAA Repeats in Spinocerebellar Ataxia Type 17. *Eur. J. Hum. Genet.* **2008**, *16*, 215. [\[CrossRef\]](#)
150. Charles, P.; Camuzat, A.; Benammar, N.; Sellal, F.; Destee, A.; Bonnet, A.; Lesage, S.; Le Ber, I.; Stevanin, G.; Durr, A. Are Interrupted SCA2 CAG Repeat Expansions Responsible for Parkinsonism? *Neurology* **2007**, *69*, 1970–1975. [\[CrossRef\]](#)
151. Elden, A.C.; Kim, H.-J.; Hart, M.P.; Chen-Plotkin, A.S.; Johnson, B.S.; Fang, X.; Armarkola, M.; Geser, F.; Greene, R.; Lu, M.M. Ataxin-2 Intermediate-Length Polyglutamine Expansions Are Associated with Increased Risk for ALS. *Nature* **2010**, *466*, 1069. [\[CrossRef\]](#)
152. Choudhry, S.; Mukerji, M.; Srivastava, A.K.; Jain, S.; Brahmachari, S.K. CAG Repeat Instability at SCA2 Locus: Anchoring CAA Interruptions and Linked Single Nucleotide Polymorphisms. *Hum. Mol. Genet.* **2001**, *10*, 2437–2446. [\[CrossRef\]](#)
153. Volker, J.; Breslauer, K. How Sequence Alterations Enhance the Stability and Delay Expansion of DNA Triplet Repeat Domains. *QRB Discov.* **2023**, *4*, e8. [\[CrossRef\]](#)
154. Hartenstine, M.J.; Goodman, M.F.; Petruska, J. Base Stacking and Even/Odd Behavior of Hairpin Loops in DNA Triplet Repeat Slippage and Expansion with DNA Polymerase. *J. Biol. Chem.* **2000**, *275*, 18382–18390. [\[CrossRef\]](#)
155. Brook, J.D.; McCurrach, M.E.; Harley, H.G.; Buckler, A.J.; Church, D.; Aburatani, H.; Hunter, K.; Stanton, V.P.; Thirion, J.-P.; Hudson, T.; et al. Molecular Basis of Myotonic Dystrophy: Expansion of a Trinucleotide (CTG) Repeat at the 3’ End of a Trans-Script Encoding a Protein Kinase Family Member. *Cell* **1992**, *68*, 799. [\[CrossRef\]](#)
156. Mahadevan, M.; Tsilfidi, C.; Sabourin, L.; Shutler, G.; Amemiya, C.; Jansen, G.; Neville, C.; Narang, M.; Barcelo, J.; O’Hoy, K. Myotonic Dystrophy Mutation: An Unstable CTG Repeat in the Untranslated Region of the Gene. *Science* **1992**, *255*, 1253. [\[CrossRef\]](#) [\[PubMed\]](#)
157. Davis, B.; McCurrach, M.; Taneja, K.; Singer, R.; Housman, D. Expansion of a CUG Trinucleotide Repeat in the 3’ Untranslated Region of Myotonic Dystrophy Protein Kinase Transcripts Results in Nuclear Retention of Transcripts. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 7388. [\[CrossRef\]](#) [\[PubMed\]](#)
158. Kiliszek, A.; Kierzek, R.; Krzyzosiak, W.J.; Rypniewski, W. Structural Insights into CUG Repeats Containing the ‘Stretched U-U Wobble’: Implications for Myotonic Dystrophy. *Nucleic Acids Res.* **2009**, *37*, 4149–4156. [\[CrossRef\]](#) [\[PubMed\]](#)
159. Coonrod, L.; Lohman, J.R.; Berglund, J.A. Utilizing the GAAA Tetra Loop/Receptor to Facilitate Crystal Packing and the Determination of the Structure of a CUG RNA Helix. *Biochemistry* **2012**, *42*, 8330.
160. Tamjar, J.; Katorcha, E.; Popov, A.; Malinina, L. Structural Dynamics of Double- Helical RNAs Composed of CUG/CUG and CUG/CGG Repeats. *J. Biomol. Struct. Dyn.* **2012**, *30*, 505. [\[CrossRef\]](#)
161. Napierala, M.; Krzyzosiak, W.J. CUG repeats present in myotonin kinase RNA form metastable “slippery” hairpins. *J. Biol. Chem.* **1997**, *272*, 31079. [\[CrossRef\]](#)
162. Yildirim, I.; Chakraborty, D.; Disney, M.; Wales, D.; Schatz, G. Computational Investigation of RNA CUG Repeats Responsible for Myotonic Dystrophy 1. *J. Chem. Theory Comput.* **2015**, *11*, 4943. [\[CrossRef\]](#)

163. Ni, C.W.; Wei, Y.J.; Shen, Y.I.; Lee, I.R. Long-Range Hairpin Slippage Reconfiguration Dynamics in Trinucleotide Repeat Sequences. *J. Phys. Chem. Lett.* **2019**, *10*, 3985–3990. [\[CrossRef\]](#)
164. Wan, L.; He, A.; Li, J.; Guo, P.; Han, D. High-Resolution NMR Structures of Intra Strand Hairpins Formed by CTG Trinucleotide Repeats. *ACS Chem. Neurosci.* **2024**, *15*, 868–876. [\[CrossRef\]](#)
165. Fu, Y.H.; Kuhl, D.P.; Pizzuti, A.; Pieretti, M.; Sutcliffe, J.S.; Richards, S.; Verkert, A.J.; Holden, J.J., Jr.; RGF, W.; ST, O.; et al. Variation of the CGG Repeat at the Fragile X Site Results in Genetic Instability: Resolution of the Sherman Paradox. *Cell* **1991**, *67*, 1047–1058. [\[CrossRef\]](#)
166. Zhong, N.; Ju, W.; Pietrofesa, J.; Wang, D.; Dobkin, C.; Brown, W.T. Fragile X “Gray Zone” Alleles: AGG Patterns, Expansion Risks, and Associated Haplotypes. *Am. J. Med. Genet.* **1996**, *64*, 261–265. [\[CrossRef\]](#)
167. Dombrowski, C.; Lévesque, S.; Morel, M.L.; Rouillard, P.; Morgan, K.; Rousseau, F. Pre-Mutation and Intermediate-Size FMR1 Alleles in 10572 Males from the General Population: Loss of an AGG Interruption Is a Late Event in the Generation of Fragile X Syndrome Alleles. *Hum. Mol. Genet.* **2002**, *11*, 371–378. [\[CrossRef\]](#) [\[PubMed\]](#)
168. Hagerman, R.; Leehey, M.; Heinrichs, W.; Tassone, F.; Wilson, R.; Hills, J.; Grigsby, J.; Gage, B.; Hagerman, P. Intention Tremor, Parkinsonism, and Generalized Brain Atrophy in Male Carriers of Fragile X. *Neurology* **2001**, *57*, 127–130. [\[CrossRef\]](#) [\[PubMed\]](#)
169. Sherman, S.L. Premature Ovarian Failure among Fragile X Premutation Carriers: Parent-of-Origin Effect? *Am. J. Hum. Genet.* **2000**, *67*, 11–13. [\[CrossRef\]](#) [\[PubMed\]](#)
170. Glass, I. X linked mental retardation. *J. Med. Genet.* **1991**, *28*, 361–371. [\[CrossRef\]](#)
171. Gu, Y.; Shen, Y.; Gibbs, R.A.; Nelson, D.L. Identification of FMR2, A Novel Gene Associated with the FRAXE CCG Repeat and CpG Island. *Nat. Genet.* **1996**, *13*, 109–113. [\[CrossRef\]](#)
172. Braida, C.; Stefanatos, R.K.; Adam, B.; Mahajan, N.; Smeets, H.J.; Niel, F.; Goizet, C.; Arveiler, B.; Koenig, M.; Lagier-Tourenne, C.; et al. Variant CCG and GGC Repeats within the CTG Ex-Pansion Dramatically Modify Mutational Dynamics and Likely Contribute toward Un- Usual Symptoms in Some Myotonic Dystrophy Type 1 Patients. *Hum. Mol. Genet.* **2010**, *19*, 1399–1412. [\[CrossRef\]](#)
173. Kuma, A.; Fang, P.; Park, H.; Guo, M.; Nettles, K.W.; Disney, M. A Crystal Structure of a Model of the Repeating r(CGG) Transcript Found in Fragile X Syndrome. *ChemBioChem* **2011**, *12*, 2140–2142. [\[CrossRef\]](#)
174. Kiliszek, A.; Kierzek, R.; Krzyzosiak, W.J.; Rypniewski, W. Crystallographic Characterization of CCG Repeats. *Nucl. Acids Res.* **2012**, *40*, 8155–8162. [\[CrossRef\]](#)
175. Man, V.H.; Pan, F.; Sagui, C.; Roland, C. Comparative Melting and Healing of B-DNA and Z-DNA by an Infrared Laser Pulse. *J. Chem. Phys.* **2016**, *144*, 145101. [\[CrossRef\]](#)
176. Gao, X.; Huang, X.; Smith, G.; Zheng, M.; Liu, H. New Antiparallel Duplex Motif of DNA CCG Repeats That Is Stabilized by Extrahelical Basis Symmetrically Located in the Minor-Groove. *J. Am. Chem. Soc.* **1995**, *117*, 8883–8884. [\[CrossRef\]](#)
177. Assi, H.; Garavis, M.; Gonzalez, C.; Damha, M. I-Motif DNA: Structural Features and Significance to Cell Biology. *Nucl. Acids Res.* **2018**, *46*, 8038. [\[PubMed\]](#)
178. Chen, Y.W.; Jhan, C.R.; Neidle, S.; Hou, M.H. Structural Basis for the Identification of an I-Motif Tetraplex Core with a Parallel-Duplex Junction as a Structural Motif in CCG Triplet Repeats. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 10682. [\[CrossRef\]](#) [\[PubMed\]](#)
179. Zeraati, M.; Langley, D.; Schofield, P.; Moye, A.; Rouet, R.; Hughes, W.; Bryan, T.; Dinger, M.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nat. Chem.* **2018**, *10*, 631. [\[CrossRef\]](#) [\[PubMed\]](#)
180. Fojtik, P.; Kejnovska, I.; Vorlickova, M. The Guanine-Rich Fragile X-Chromosome Repeats Are Reluctant to Form Tetramers. *Nucl. Acids Res.* **2004**, *32*, 298. [\[CrossRef\]](#)
181. Renciuik, D.; Kypr, J.; Vorlickova, M. CCG Repeats Associated with Fragile X Chromosome Form Left-Handed Z-DNA Structure. *Biopolymers* **2011**, *95*, 174. [\[CrossRef\]](#)
182. Shen, Y.-I.; Cheng, K.-C.; Wei, Y.-J.; Lee, I.-R. Structural Dynamics Role of AGG Interruptions in Inhibition CGG Repeat Expansions Associated with Fragile X Syndrome. *ACS Chem. Neurosci.* **2024**, *15*, 230–235. [\[CrossRef\]](#)
183. Eichler, E.E.; Holden, J.J.; Popovich, B.W.; Reiss, A.L.; Snow, K.; Thibodeau, S.N.; Richards, C.S.; Ward, P.A.; Nelson, D.L. Length of Uninterrupted CGG Repeats Determines Instability in the FMR1 Gene. *Nat. Genet.* **1994**, *8*, 88–94. [\[CrossRef\]](#)
184. Falik-Zaccari, T.C.; Shachek, E.; Yalon, M.; Lis, Z.; Borochowitz, Z.; Macpherson, J.N.; Nelson, D.L.; Eichler, E.E. Predisposition to the Fragile X-Syndrome in Jews of Tunisian Descent Is Due to the Absence of AGG Interruptions on a Rare Mediterranean Haplotype. *Am. J. Hum. Genet.* **1997**, *60*, 103–112.
185. Yrigollen, C.M.; Durbin-Johnson, B.; Gane, L.; Nelson, D.L.; Hagerman, R.; Hagerman, P.J.; Tassone, F. AGG Interruptions with the Maternal FMR1 Gene Reduces the Irish Offspring with Fragile X Syndrome. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **2012**, *14*, 729–736.
186. Latham, G.J.; Coppinger, J.; Hadd, A.G.; Nolin, S.L. The Role of AGG Interruptions in Fragile X Repeat Expansions: A Twenty Year Perspective. *Front. Genet.* **2014**, *5*, 244. [\[CrossRef\]](#) [\[PubMed\]](#)
187. Jarem, D.A.; Huckaby, L.V.; Delaney, S. AGG Interruptions in (CGG)_n DNA Repeat Tracts Modulate the Structure and Thermodynamics of Non-B Conformations in Vitro. *Biochemistry* **2010**, *49*, 6826–6837. [\[CrossRef\]](#) [\[PubMed\]](#)
188. Rajan-Babu, I.-S.; Dolzenko, E.; Eberle, M.; Friendman, J. Sequence Composition Changes in Short Tandem Repeats: Heterogeneity Detection, Mechanisms and Clinical Implications. *Nat. Rev. Genet.* **2024**, *25*, 476–499. [\[CrossRef\]](#) [\[PubMed\]](#)

189. Matos-Rodrigues, G.; Hisey, J.; Nussenzweig, A.; Mirkin, S. Detection of Alternative DNA Structures and Its Implications for Human Disease. *Mol. Cell* **2023**, *83*, 3622–3641. [\[CrossRef\]](#) [\[PubMed\]](#)
190. Darlow, J.M.; Leach, D.R.F. The Effects of Trinucleotide Repeats Found in Human Inherited Disorders on Palindrome Inviability in Escherichia Coli Suggest Hairpin Folding Preferences in Vivo. *Genetics* **1995**, *141*, 825–832. [\[CrossRef\]](#)
191. Zhao, X.N.; Usdin, K. FAN1 Protects against Repeat Expansions in a Fragile X Mouse Model. *DNA Repair* **2018**, *69*, 1–5. [\[CrossRef\]](#)
192. Goold, R.; Flower, M.; Moss, D.M.; Medway, C.; Wood-Kaczmar, A.; Andre, R.; Farshim, P.; Bates, G.P.; Holmans, P.; Jones, L.; et al. FAN1 Modifies Huntington's Progression by Stabilizing the Expanded HTT CAG Repeat. *Hum. Mol. Genet.* **2019**, *28*, 650–661. [\[CrossRef\]](#)
193. Goold, R.; Hamilton, J.; Menneteau, T.; Flower, M.; Bunting, E.L.; Aldous, S.G.; Porro, A.; Vicente, J.R.; Allen, N.D.; Wilkinson, H.; et al. FAN1 Controls Mismatch Repair Complex Assembly via MLH1 Retention to Stabilize CAG Repeat Expansion in Huntington's Disease. *Cell Rep.* **2021**, *36*, 109649. [\[CrossRef\]](#)
194. Loupe, J.M.; Pinto, R.M.; Kim, K.H.; Gillis, T.; Mysore, J.S.; Andrew, M.A.; Kolvalenko, M.; Murtha, R.; Seong, I.; Gusella, J.F.; et al. Promotion and Somatic CAG Repeat Expansion by FAN1 Knock-out in Huntington's Disease Knock-in Mice Is Blocked by MLH1 Knock-Out. *Hum. Mol. Genet.* **2020**, *29*, 3044–3053. [\[CrossRef\]](#)
195. Desmukh, A.L.; Porro, A.; Mohiuddun, M.; Lanni, S.; Panigrahi, G.B.; Caron, M.C.; Masson, J.Y.; Sartori, A.A.; Pearson, C.E. FAN1, a DNA Repair Nuclease as a Modifier of Repeat Expansion Disorders. *Hungtons Dis.* **2021**, *10*, 95–122.
196. Phadte, A.S.; Bhatia, M.; Ebert, H.; Abdulla, H.; Elrazaq, E.A.; Komolov, K.E.; Pluciennik, A. FAN1 Removes Triplet Repeat Extrusions via a PCNA- and RFC-Dependent Mechanism. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2302103120. [\[CrossRef\]](#) [\[PubMed\]](#)
197. Kadyrova, L.Y.; Mieczkowski, P.A.; Kadyrov, F.A. Genome-Wide Contributions of the MutS α and MutS β -Dependent DNA Mismatch Repair Pathways to the Maintenance of Genetic Stability in Saccharomyces Cerevisiae. *J. Biol. Chem.* **2023**, *299*, 104705. [\[CrossRef\]](#) [\[PubMed\]](#)
198. Tabrizi, S.J.; Flower, M.D.; Ross, C.A.; Wild, E.J. Huntington Disease: New Insights into Molecular Pathogenesis and Therapeutic Opportunities. *Nat Rev. Neurol.* **2020**, *16*, 529–546. [\[CrossRef\]](#) [\[PubMed\]](#)
199. Tabrizi, S.J.; Estevez-Fraga, C.; van Roon-Mom, W.M.C.; Flower, M.D.; Scahill, R.I.; Wild, E.J.; Muñoz-Sanjuan, I.; Sampaio, C.; Rosser, A.E.; Leavitt, B.R. Potential Disease-Modifying Therapies for Huntington's Disease: Lessons Learned and Future Opportunities. *Lancet Neurol.* **2022**, *21*, 645–658. [\[CrossRef\]](#) [\[PubMed\]](#)
200. Li, J.; Wang, H.; Yang, W. Tandem MutS β Binding to Long Extruded DNA Trinucleotide Repeats Underpins Pathogenic Expansions. *bioRxiv*, 2023; preprint.
201. Lee, J.-H.; Thomsen, M.; Daub, H.; Steinbacher, S.C.; Szttyler, A.; Thieulin-Pardo, G.; Neudegger, T.; Plotnikov, N.V.; Iyer, R.R.; Wilkinson, H.A.; et al. Conformational Dynamics and DNA Recognition by Human MutS β . *bioRxiv*, 2023; preprint.
202. Lang, W.H.; Coats, J.E.; Majka, J.; Hura, G.L.; Lin, Y.; Rasnik, I.; McMurray, C.T. Conformational Trapping of Mismatch Recognition Complex MSH2/MSH3 on Repair-Resistant DNA Loops. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, E837–E844. [\[CrossRef\]](#)
203. Kadyrova, L.Y.; Gujar, V.; Burdett, V.; Modrich, P.L.; Kadyrov, F.A. Human MutL γ , the MLH1–MLH3 Heterodimer, Is an Endonuclease That Promotes DNA Expansion. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 3535–3542. [\[CrossRef\]](#)
204. Huang, T.Y.; Chang, C.K.; Kao, Y.F.; Chin, C.H.; Ni, C.W.; Hsu, H.Y.; Hu, N.J.; Hsieh, L.C.; Chou, S.H.; Lee, I.R.; et al. Parity-Dependent Hairpin Configurations of Repetitive DNA Sequence Promote Slippage Associated with DNA Expansion. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 9535–9540. [\[CrossRef\]](#)
205. O'Reilly, D.; Belgrad, J.; Ferguson, C.; Summers, A.; Sapp, E.; McHugh, C.; Mathews, E.; Boudi, A.; Buchwald, J.; Ly, S.; et al. Di-Valent SiRNA-Mediated Silencing of MSH3 Blocks Somatic Repeat Expansion in Mouse Models of Huntington's Disease. *Mol. Ther.* **2023**, *31*, 1661–1674. [\[CrossRef\]](#) [\[PubMed\]](#)
206. Aldous, S.G.; Smith, E.J.; Landles, C.; Osborne, G.F.; Cañibano-Pico, M.; Nita, I.M.; Phillips, J.; Zhang, Y.; Jin, B.; Hirst, M.B.; et al. A CAG Repeat Threshold for Therapeutics Targeting Somatic Instability in Huntington's Disease. *Brain* **2024**, *147*, 1784–1798. [\[CrossRef\]](#) [\[PubMed\]](#)
207. Choi, D.E.; Shin, J.W.; Zeng, S.; Hong, E.P.; Jang, J.H.; Loupe, J.M.; Wheeler, V.C.; Stutzman, H.E.; Kleinstiver, B.; Lee, J.M. Base Editing Strategies to Convert CAG to CAA Diminish the Disease-Causing Mutation in Huntington's Disease. *Elife* **2024**, *12*, RP89782. [\[CrossRef\]](#) [\[PubMed\]](#)
208. Gall-Duncan, T.; Luo, J.; Jurkovic, C.M.; Fisher, L.A.; Fujita, K.; Desmukh, A.L.; Harding, R.J.; Tran, S.; Mehkary, M.; Li, V.; et al. Antagonistic Roles of Canonical and Alternative-RPA in Disease-Associated Tandem CAG Repeat Instability. *Cell* **2023**, *186*, 4898–4919. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.