The hidden side of unstable DNA repeats: Mutagenesis at a distance

Kartik A. Shah¹, Sergei M. Mirkin*

Department of Biology, Tufts University, Medford, MA 02155, USA

ABSTRACT

Structure-prone DNA repeats are common components of genomic DNA in all kingdoms of life. In humans, these repeats are linked to genomic instabilities that result in various hereditary disorders, including many cancers. It has long been known that DNA repeats are not only highly polymorphic in length but can also cause chromosomal fragility and stimulate gross chromosomal rearrangements, i.e., deletions, duplications, inversions, translocations and more complex shuffles. More recently, it has become clear that inherently unstable DNA repeats dramatically elevate mutation rates in surrounding DNA segments and that these mutations can occur up to ten kilobases away from the repetitive tract, a phenomenon we call repeat-induced mutagenesis (RIM). This review describes experimental data that led to the discovery and characterization of RIM and discusses the molecular mechanisms that could account for this phenomenon.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

About half of the human genome is made up of repetitive DNA elements i.e., multiple copies of identical sequences, often categorized based on their location and length of the repeating unit (interspersed vs tandem repeats, microsatellites vs minisatellites etc.) [1,2]. While the polymorphic nature of these repeats is believed to contribute to genetic variability, their instability is known to cause various human diseases. One striking example is the expansion of short tandem DNA repeats, a phenomenon responsible for ~40 human hereditary neurological, neurodegenerative and developmental disorders such as Huntington’s disease, myotonic dystrophy type 1, Friedreich’s ataxia, fragile X syndrome, amyotrophic lateral sclerosis and others (reviewed in [3–5]). Molecular mechanisms underlying DNA repeats instability have been extensively studied in various experimental systems, including bacteria, yeast, mice and cultured human cells [6–8]. An unexpected outcome of these studies has been the discovery that besides being inherently unstable, DNA repeats can also induce mutations in flanking DNA sequences, a phenomenon called repeat-induced mutagenesis (RIM) [9]. Here, we review the historical backdrop as well as recent experimental data demonstrating RIM and discuss the molecular pathways through which it compromises genomic integrity.

2. Historical background

The discovery of repeat-induced mutagenesis was totally serendipitous in nature. The story goes back to the 1980s during which many alternative DNA structures, including left-handed Z-DNA, cruciform DNA, three-stranded H-DNA and four-stranded G-quadruplex DNA were discovered (reviewed in [10]). The first of multi-stranded DNA structures to be discovered was H-DNA – an intramolecular DNA triplex formed by homopurine–homopyrimidine mirror repeats under the influence of negative supercoiling [11]. This discovery was almost instantly followed by the realization that intermolecular triplexes could form between a triplex-forming oligonucleotide (TFO) and its homopurine–homopyrimidine target in duplex DNA [12–14]. Researchers found that targeting a homopurine–homopyrimidine sequence within the promoter region of the c-myc proto-oncogene with a TFO repressed its transcription, both in vitro as well as in cultured HeLa cells [15,16]. Because such sequence elements are found commonly in the human genome and often located in the regulatory portions of various genes, it was speculated that TFOS could be used as prospective antigen tools to gain control of gene expression at the transcriptional level [17]. Subsequent demonstrations of TFO-mediated gene modulation by various groups invariably helped antigen technology gain momentum as an attractive therapeutic strategy against viral infections as well as cancer [18].
Investigations into the mechanisms behind TFO-mediated modulation of gene expression revealed two major causes – (1) direct blockage of transcriptional initiation and/or elongation at the site of triplex formation [19–22] and (2) induction of localized mutations by TFOs through site-specific DNA damage (see [23–25] and references therein).

The latter was a totally unexpected outcome, nonetheless exploited by researchers, who began deliberately conjugating DNA damaging agents, such as psoralen or bleomycin, to the TFOs in the hope to develop a powerful yet facile method for site-specific genome modification [23,25,26]. In experiments with a supF plasmid reporter system carried out in cultured primate cells, it was found that a 30-nucleotide long TFO increased the rate of localized mutations 10-fold above control. Psoralen-conjugates of this TFO increased the mutation rate up to 100-fold above control upon activation by irradiation [27,28]. Similar TFO-mediated mutagenesis was reported in cultured human cells [29] but was absent however, in xeroderma pigmentosa group A (XPA) cells deficient in nucleotide excision repair or in Cockayne’s syndrome group B (CSB) cells deficient in transcription-coupled repair. These results implied that transcription-coupled and/or nucleotide excision repair pathways are essential for the TFO-induced mutagenesis. Researchers soon discovered that TFOs could also induce point mutations, small insertions and deletions at or around their chromosomal targets in mammalian cells [30,31]. Remarkably, in all cases, these mutations were found to lie either within the TFO target site or less than 50 base pairs upstream of it. Using the same supF plasmid reporter system, it was further demonstrated that the H-DNA-forming region from the human c-myc promoter could induce a greater than 20-fold increase in mutation frequency by itself, i.e., without the presence of any TFO [32]. Additionally, frequent double-strand breaks (DSB) were found to occur around the H-DNA site in this system, implying that naturally occurring triplex-forming DNA sequences could act as an endogenous source of genomic instability [24]. Sequences capable of forming other unusual DNA structures, such as Z-DNA [33] or G-quadruplex DNA [34,35], also appeared to be mutagenic. It was suggested therefore, that unusual DNA structures, which have the ability to stall the DNA replication or transcription machinery, might result in DNA breakage and subsequently lead to mutagenesis (reviewed in [36]).

3. Recent insights into repeat-induced mutagenesis

3.1. Mutations induced by Friedreich’s ataxia GAA repeats

Friedreich’s ataxia (FRDA) is an autosomal recessive neurodegenerative disorder caused by the expansion of (GAA)ₙ repeats in the first intron of the frataxin (FXN) gene [37]. Studies from various labs have revealed that the GAA repeat element is capable of forming an intramolecular triplex and blocking transcriptional elongation in vitro as well as in vivo [38,39]. Expanded (GAA)ₙ repeats also trigger heterochromatinization of surrounding DNA, leading to the inhibition of FXN gene expression (reviewed in [40]). While studying GAA repeat expansions in serially passaged lymphoblastoid cell lines from FRDA patients, researchers inadvertently observed a 3-fold increase in mutagenesis within a 135 bp region immediately upstream of the expanded (GAA)ₙ repeat tract [41]. Our lab recently developed a tractable genetic system to study large-scale repeat expansions in yeast [42]. Various lengths of GAA repeats were cloned into the intron of the artificially split URA3 selectable marker. Large-scale expansions of the repeats would turn off URA3 splicing, allowing for selection of such events on media containing 5-fluoroorotic acid (5-FOA). However, we observed three different types of events among the 5-FOA resistant clones: (1) large-scale expansions of GAA repeats, (2) mutations in the body of the URA3 gene and (3) chromosomal rearrangements. Relative to a control construct without repeats, the rate of mutation was elevated 10-fold in the presence of a (GAA)₁₇₅ run and 100-fold in the presence of (GAA)₁₅₉ run. Thus, doubling the size of the repeat increased mutation rate by an order of magnitude. Interestingly, these mutations consisted largely of point substitutions and were found at significant distances, i.e. up to 1 kb away, both upstream as well as downstream of the repeat tract.

More recently, we investigated the role of various DNA polymerases in GAA-mediated RIM [9]. In eukaryotes, faithful duplication of the genome requires the action of several DNA polymerases. These include three distinct and specialized replicative DNA polymerases – leading and lagging strand polymerases Pol ε (epsilon) and Pol δ (delta), respectively, as well as polymerase-primase Pol α (alpha) (reviewed in [43,44]). In addition to the replicative DNA polymerases, various translesion (TLS) polymerases also carry out specialized roles, and these include DNA polymerases Pol ζ (zeta) and Rev1 (for a detailed description, see [45] and references therein). Using the above yeast system, we found that defects in the leading or lagging strand polymerase (Pol ε and Pol δ), but not polymerase-primase (Pol α), drastically elevated the rate of RIM relative to a wild-type background. Consistent with prior observations, the repeat-induced mutations in DNA polymerase mutants could be mapped as far as 1 kb away and on either side of the repetitive tract. The increase in rate of RIM in yeast strains containing defective-replicosomes seemed to depend, at least partially, on the presence of TLS polymerases Pol ζ or Rev1. Majority of the point substitutions observed in these cases were C-to-G transversions, which is a characteristic signature of these two TLS polymerases [46,47]. Recent studies have found that TLS polymerases are recruited to the replication fork by tethering to proliferating cell nuclear antigen (PCNA) [48]. The signaling cascade is initiated by post-translational modification of a lysine residue (K164) on PCNA (reviewed in [49]). We observed that blocking modification of this residue through a mutation in PCNA (K164R) also had the same effect as the absence of Pol ζ or Rev1. Note, however, that we did not observe a Pol ζ-dependence of RIM in the wild-type background. Tang et al. studied GAA-mediated RIM in a slightly different yeast system, in which a URA3 reporter was placed more than 1 kb away from a very long (GAA)₂₃₀ repeat [50,51]. Both deletions and mutations were observed in the body of the URA3 reporter, but an overwhelming majority of these events were point substitutions. While the results in this system are largely consistent with our observations described above, there are also some important differences. First, almost all mutational events in [51] were accompanied by changes in the length of the GAA. Second, in many instances, the URA3 gene was found to contain more than one mutation. Third, and the most significant difference was that these mutations were exquisitely Pol ε-dependent. Since deletion of its catalytic subunit (rev3) ablated nearly all of the mutagenesis in the wild-type background. These differences could be explained by the fact that different tract lengths were used in the two systems – (GAA)₁₀₀ in our case and (GAA)₂₃₀ in theirs. However, strain or locus-specific differences cannot be ruled out. Saini et al. also designed a yeast system to study RIM, in which the URA3 reporter was positioned at short (0.4–0.6 kb), medium (8 kb) or long (30 kb) distances on either side of a very long (GAA)₂₃₀ repeat [52]. The presence of a long GAA tract caused substantial chromosomal fragility, resulting in double strand breaks (DSB) at the junction of the repeats. This increase in chromosomal fragility went hand in hand with a significant increase in GAA-mediated mutagenesis and was observed up to 8 kb away from the repeat tract. However, this effect was only seen in strains containing defective DNA Pol δ. Consistent with our observations, GAA-mediated mutagenesis in replisome-defective
strains was found to be partially dependent on TLS polymerase Pol ζ.

3.2. Mutations induced by other repeats

Recent investigations have uncovered a relationship between RIM-associated phenomena and several other classes of repeats, including inverted repeats and telomere-like repeats. Inverted repeats (IR), known to fold into hairpin or cruciform structures, are a potent source of chromosomal rearrangements and genomic instability [53–56]. *Alu* retrotransposons, a type of IR consisting of a ∼320 bp-long repeat element, are found littered throughout the human genome (reviewed in [57]). *Alu*-mediated recombination is known to cause as many as 33 inherited disorders such as insulin-resistant diabetes type II, Tay-Sachs disease, familial hypercholesterolaemia, α-thalassemia and others as well as 16 different types of cancer, including Ewing sarcoma, breast cancer, acute myelogenous leukemia etc. (reviewed in [58]). Saini et al. used their yeast experimental (described above) to assess the mutagenic potential of long inverted repeats, including the 320 bp *Alu* palindrome [52]. Similarly to GAA-mediated RIM, the mutations in IR-mediated RIM were (1) increased up to 30-fold in strains with defective DNA Pol β but not in the wild-type strain, (2) partially dependent on the TLS polymerase Pol ζ and (3) detected as far as 8 kb away from the IR site. Additionally, these inverted repeats were also found to cause substantial chromosomal fragility, resulting in DSBs.

Telomeric DNA repeats help in maintaining chromosomal stability and genomic integrity by protecting the ends of chromosomes from fusion and degradation (reviewed in [59]). While telomeric repeats are usually located at chromosomal termini, interstitial telomeric sequences (ITS) are found at internal sites on the chromosomes of many organisms (reviewed in [60–62]). ITSs co-localize to sites of several human chromosomal aberrations and rearrangements implicated in chromosomal fragility, various cancers, Prader–Willi syndrome etc. (reviewed in [60,61]). While investigating the mechanisms behind ITS instability in our yeast experimental system, we found that even short telomere-like (TGTGTGGG)_n

---

**Fig. 1.** Molecular mechanisms leading to repeat-induced mutagenesis.  
(A) Short repeats block transcription by forming alternative DNA structures or R-loops. This initiates ‘gratuitous’ transcription-coupled repair (TCR) and generates single strand breaks (SSBs) that can be converted to double strand breaks (DSBs). To repair the break, the ends undergo resection and single-strand annealing followed by gap-filling through TLS DNA polymerases that gives rise to mutations on either side of the repeat tract. (B) Long repeats block replication fork progression and result in reversed forks that are similar to Holliday junctions. Upon isomerization and resolution, these structures generate a one-ended DSB, which can be repaired by break-induced replication (BIR). During BIR, the end undergoes resection followed by invasion and copying of several kilobases of DNA from the donor. Due to the conservative mode of replication in BIR, DNA synthesis is highly error-prone and gives rise to mutations at large distances from the initial break site. Unrepaired SSBs generated by gratuitous TCR can also give rise to a one-ended DSB upon replication. (C) Long structure-prone repeats can form unusual DNA structures even in the G1 phase of the cell cycle. Cleavage of these structures leads to the formation of two-ended DSBs that can anneal ‘out-of-register’ and undergo homologous recombination (HR) repair. Gap-filling of the resected ends by error-prone DNA polymerases can give rise to mutations on either side and at large distances from the break site. In the case of (A) and (C), mutations will be incorporated in the next round of replication (not shown). Repetitive strands are shown in red and green. Dashed lines indicate DNA synthesized during repair and yellow star indicates mutations.
4. Molecular models of repeat-induced mutagenesis

Structure-prone DNA repeats frequently co-localize to breakpoints of various chromosomal aberrations, including deletions, translocations, duplications, inversions and complex chromosomal rearrangements linked to human disease (reviewed in [36,65–68]). It is becoming increasingly clear that formation of some sort of a DNA double-strand break (DSB) intermediate is at the heart of all these aberrations. In several model systems, increased chromosomal fragility (i.e., DSB) goes hand in hand with an increase in RIM [52]. Increased local mutations have also been observed in the vicinity of complex human chromosomal rearrangements [69]. We believe, therefore, that RIM is the hidden and largely unexplored consequence of repeat instability, resulting from the repair of a DSB event. In light of the above findings, we discuss how and when these DSBs could occur and in what way their repair could result in RIM.

4.1. Mutagenesis induced by short structure-prone DNA repeats

The molecular mechanisms of RIM could be somewhat different between short repeats and longer structure-prone repeats. Short repeats are not known to stall replication fork progression in vivo [56,70–72], and unless they are present on strongly supercoiled DNA, formation of alternative DNA structures by short repetitive tracts is not energetically favorable [73]. At the same time, short repetitive tracts, including the triplex-forming GAA repeats and the H-DNA region from the c-myc promoter, are known to stall RNA polymerase in vitro as well as in vivo [74–76]. Could transcriptional stalling by short repeats result in DNA breakage? Transcription-coupled repair (TCR), a sub pathway of nucleotide excision repair (NER), is activated when the transcriptional machinery encounters a damaged DNA template, such as an abasic site or a DNA adduct [77]. An alternative form of TCR, called “gratuitous” TCR, may be invoked when an RNA polymerase pauses at sites of unusual structure formation [78]. A wide body of evidence suggests that various short repeats can indeed stall translocating RNA polymerase in vivo by forming transient structures on the non-template strand or extended DNA–RNA hybrids (R-loops), leading to the initiation of gratuitous TCR [79–82]. Endonuclease components of the TCR machinery can induce single strand breaks (SSBs) in the template strand and it was recently shown that SSBs within R-loops could also be converted to double strand breaks (DSBs) [83], reviewed in [84]).

A DSB formed during the transcription of a repeat locus in the G1 phase of the cell cycle could be repaired by various mechanisms. Of particular interest to us is repair via the single-strand annealing pathway (SSA) (Fig. 1A) where the 5′ ends of the break undergo resection, followed by re-annealing of the 3′ repetitive ends and extension through gap filling [85,86]. DNA synthesis during gap filling can be carried out by various DNA repair polymerases, including TLS polymerase DNA Pol ζ, and give rise to point mutations on either side of the repeat tract. Importantly, re-annealing of repetitive tracts would most likely occur out of register, causing a change in the repeat’s length together with RIM, as was indeed observed in [51].

4.2. Mutagenesis induced by long structure-prone DNA repeats

Long structure-prone repeats are known to pose various challenges during DNA replication, recombination and repair. Numerous studies have shown that long repeats are potent blockers of DNA polymerases in vitro and replication fork progression in vivo (reviewed in [88]). Stalled replication forks often lead to fork reversal [89], the resulting structure of which (Fig. 1B) is practically indistinguishable from a Holliday junction [90]. Upon isomerization and enzymatic resolution, a reversed fork could be converted into a one-ended DSB (Fig. 1B), and breaks with only one free end are believed to be repaired by break-induced replication (BIR) [91]. In BIR, the 5′ end of the break undergoes resection, followed by invasion of the 5′ end into a homologous region on the sister chromatid and extension by DNA synthesis. BIR can involve copying of several kilobases of DNA, often times extending all the way to the telomere of the donor chromosome [92]. The nascent DNA is synthesized by DNA Pol δ and is exquisitely dependent on its Pol32 subunit [93]. Despite being independent of error-prone TLS polymerases, BIR appears to be a highly mutagenic form of DSB repair [94–96]. Remarkably, this is due to BIR involving a “conservative” mode of replication [97,98], for which the nascent leading strand serves as a template for the nascent lagging strand (Fig. 1B), thus precluding the role of mismatch repair. Since the mutational signature of BIR can be observed many kilobases away from the site of DSB, it could explain the large-scale effects of RIM and its independence from translesion polymerase observed in our studies [9].

Apart from forming alternative DNA structures during DNA replication in the S phase, long DNA repeats can also do so in the G1 phase of the cell cycle (Fig. 1C). In this scenario, limited DNA supercoiling required for theses structural transitions can easily be generated during transcriptional elongation [99]. Alternative DNA structures are known to be cleaved by various proteins in vivo, resulting in the formation of two-ended DSBs (Fig. 1C). For example, a DNA triplex formed by the long GAA repeats in yeast is converted into a DSB by the endonuclease activity of Mute [100]. The Mre11p-Rad50p-Xrs2p (MRX) complex in yeast processes the ends of two-ended DSBs, giving rise to 3′-overhangs [53]. These overhangs can quickly re-annel, owing to the repetitive nature of their ends, followed by gap-filling carried out by DNA repair polymerases, including translesion DNA Pol ζ (Fig. 1C). This can explain the occurrence of point mutations on either side of the repeat tract. Importantly, re-anneling of repetitive tracts would most likely occur out of register, causing a change in the repeat’s length together with RIM, as was indeed observed in [51].

5. Future directions

While Fig. 1 summarizes current ideas on genetic transactions leading to RIM, detailed molecular mechanisms of this phenomenon still remain to be unraveled. In the short run, genetic and biochemical studies are needed to understand the role of Rad26p in TCR-mediated RIM, Pol32p subunit of DNA Polδ and DNA-helicase Pif1p in BIR-mediated RIM as well as recombination proteins Rad51p and Rad52p in SSA-mediated RIM. Experimental systems designed to study repeat instability in a controlled transcription environment could be used to distinguish the relative contributions of transcription and replication in RIM [101]. One common and important feature of the replication-dependent and replication-independent pathways leading to RIM is the presence of single-stranded DNA (ssDNA) at the DSB site. Sensing of
ssDNA by cell-cycle checkpoint proteins should trigger a DNA damage response by the ATR signaling pathway (Mec1 pathway in S. cerevisiae) [reviewed in [102]]. Furthermore, if the ssDNA is converted to a DSB, the ATM signaling pathway becomes rapidly activated (Tel1 protein in S. cerevisiae). ssDNA sensing and coating by RPA (ssDNA-binding protein) is necessary to activate various checkpoint pathways. However, alternative DNA structures are not known to bind RPA, and could thus escape detection by the checkpoint machinery. Future studies on how the Mec1p and Tel1p kinases of the ATR/ATM pathway affect RIM could help elucidate these mechanisms.

While mutagenesis mediated by short repeats is well documented in mammalian cells [54,66], very little is known about RIM caused by long structure-prone repeats in higher eukaryotes, including humans. Recent studies have observed elevated mutation levels around human chromosomes loci that have undergone complex, repeat-associated genome rearrangements [69] but almost nothing is known about their basis. Thus, developing tractable genetic systems capable of investigating both the mutability (repeat instability) as well as the mutagenicity (RIM) of long repeats is of high priority.

Finally, it would be of great interest to compare mutation rates in DNA segments located adjacent or apart from structure-prone DNA repeats using computational genomics approaches.

Acknowledgements

We thank Tom Petes, Kirill Lobachev, Polina Schcherbakova, Anna Malkova, Anna Aksenenova, Jim Haber, Lorraine Symington and Hanna Klein for many fruitful discussions of the mechanisms for DNA repeats using computational genomics approaches.

References


