DNA Repair 12 (2013) 672–684

Contents lists available at SciVerse ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

Bidirectional transcription of trinucleotide repeats: Roles for excision repair

Helen Budworth, Cynthia T. McMurray*

Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Available online 11 May 2013

Keywords:
Bidirectional transcription
Trinucleotide
Excision repair
Neurodegenerative
Coding
RNA transcript

ABSTRACT

Genomic instability at repetitive DNA regions in cells of the nervous system leads to a number of neurodegenerative and neuromuscular diseases, including those with an expanded trinucleotide repeat (TNR) tract at or nearby an expressed gene. Expansion causes disease when a particular base sequence is repeated beyond the normal range, interfering with the expression or properties of a gene product. Disease severity and onset depend on the number of repeats. As the length of the repeat tract grows, so does the size of the successive expansions and the likelihood of another unstable event. In fragile X syndrome, for example, CGG repeat instability and pathogenesis are not typically observed below tracts of roughly 50 repeats, but occur frequently at or above 55 repeats, and are virtually certain above 100–300 repeats.

Recent evidence points to bidirectional transcription as a new aspect of TNR instability and pathophysiology. Bidirectional transcription of TNR genes produces novel proteins and/or regulatory RNAs that influence both toxicity and epigenetic changes in TNR promoters. Bidirectional transcription of the TNR tract appears to influence aspects of its stability, gene processing, splicing, gene silencing, and chemical modification of DNAs. Paradoxically, however, some of the same effects are observed on both the expanded TNR gene and on its normal gene counterpart. In this review, we discuss the possible normal and abnormal effects of bidirectional transcription on trinucleotide repeat instability, the role of DNA repair in causing, preventing, or maintaining methylation, and chromatin environment of TNR genes.

Published by Elsevier B.V.

1. Introduction

More than 20 human neurodegenerative diseases are caused by amplification of simple DNA repeat tracts [1–3]. In this review, we will focus primarily on expansion of trinucleotide repeats (TNR). In the past, the position of the TNR in the coding or non-coding region of the gene was identified as a parameter of significant impact on both the mechanism of toxicity and the size of the expansion. TNRs in the coding sequence supported expression of long and "sticky" homopolymeric proteins that disrupted cellular machinery, interfering with cellular processes including trafficking, lipid homeostasis and transcription [4–6]. In a non-coding region of a gene, abnormal heterochromatin formation, gene silencing and hypermethylation frequently accompanied expansion. With the recent discovery of bidirectional transcription of TNR repeat tracts, however, these lines have been blurred.

Historically it was believed that, in eukaryotic genomes, each protein-coding gene has a unique promoter in the 5′-flanking region, and the direction of the promoter is usually unidirectional. Advances in transcriptome analysis, however, have revealed that bidirectional transcription of coding genes is common [7]. Bidirectional transcription across repeat tracts occurs in most, if not all TNR genes [8], and provides previously unknown aspects of disease pathophysiology encompassing both protein- and RNA-mediated toxicity [8–11]. Furthermore, TNR coding transcripts can be alternatively translated in the absence of a starting ATG site, in a process coined as Repeat Associated Non-ATG translation (RAN-translation) [12]. RAN-translation occurs in all reading frames. At CAG tracts, for example, translation of CAG, AGC, and GCA produces toxic homopolymeric proteins of polyglutamine, polyserine, and polyalanine tracts [12,13], respectively. These emerging discoveries provide entirely new disease mechanisms arising from a single gene defect.

The impact of bidirectional transcription on the DNA expansion mutation is unknown. The occurrence of bidirectional transcription, itself, does not require an expanded TNR tract, nor does it depend on the TNR sequence of the disease gene. To name just a few, bidirectional transcription occurs in myotonic dystrophy type 1 (DM1) (CTG) [14], in Huntington's disease-like 2 (HDL2) (CTG) [15], in the *FMR1* gene of fragile X syndrome (FXS) (CGG) [16], in Huntington's disease (HD) (CAG) [17], and in Spinocerebellar ataxia type 7 (SCA7) (CAG) [18], in which CAG, CTG, and CGG repeats are

* Corresponding author. Tel.: +1 510 486 6526; fax: +1 510 486 6880.
E-mail address: ctmcmurray@lbl.gov (C.T. McMurray).

1568-7864/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.dnarep.2013.04.019

contained within the sense strand. The occurrence of bidirectional transcription also does not depend on the position of the TNR tracts in the gene. The CAG and CTG repeats in HD, HDL2 and SCAs reside in coding regions of the gene, while DM1, FXS and SCA8 contain repeats in the untranslated regions of the gene.

Since TNR sequences potentially encode multiple repeat-containing RNAs, bidirectional transcription increases the repertoire of RNA-mediated regulatory mechanisms that are likely to influence toxicity. However, a major question is how and whether processing of TNR expansion-containing RNA influences TNR instability. Many of the RNA-mediated processing events that arise from TNR expansion genes also occur in their normal counterparts. Thus, a key question is whether TNR expansion has an influence on bidirectional transcription or *vice versa*. TNR expansion leads to aberrant phenotypes such as abnormal heterochromatin spreading. Yet, heterochromatin changes are used in nature to silence and regulate gene expression for specific uses during development. Hypermethylation often accompanies unscheduled gene silencing of TNR genes, yet methylation is required for normal or reversible heterochromatin formation. How a cell knows to regulate the extent of methylation is unknown. Hypermethylation of some TNR genes results in faulty suppression of gene expression, yet there are DNA repair activities present in the cell to maintain or regulate the normal methylated state. Do they fail to work at TNR genes? For example, deamination of 5-methylcytosine to thymine results in formation of mismatch pairs that are removed by either DNA glycosylases or mismatch repair (MMR) [19,20]. The growth arrest and DNA damage 45α protein (Gadd45α) participates with NER factors to maintain active DNA demethylation and maintain the promoter of active genes under a hypomethylated state [21]. While some RNAs arise from loci destined to encode regulatory RNAs, bidirectional transcription of TNR genes gives rise to RNAs that exert abnormal activity. Thus, TNR expansion, genome locus, DNA and histone methylation, and DNA repair are connected, but the links among them are poorly understood. Distinguishing normal from abnormal events is difficult.

In this review, we address whether and how the regulatory RNAs that arise from bidirectional transcription of TNR genes influence TNR instability. Specifically, we will review examples of bidirectional transcription of TNR genes, and discuss models for (1) how sense and antisense regulatory RNA mechanisms that operate at TNR genes are altered for an abnormal outcome, (2) the impact of chromatin modulation on methylation patterns in TNR expansion, and (3) the role of excision repair in regulating expansion, hypermethylation, and unscheduled silencing.

2. Dual sources of toxicity from bidirectionally transcribed TNR expansion genes

In order to understand possible mechanisms for expansion, we must first consider the consequences of bidirectionally transcribed aberrant RNA processing and its effects on toxicity (Fig. 1). Bidirectional transcription across an expanded TNR gene leads to three distinct scenarios for RNA regulation of TNR genes (Table 1): (1) RNAs that code for protein in both the sense and antisense directions, (2) RNAs that code for protein in only one direction, and (3) some non-coding RNAs induce heterochromatin spreading and hypermethylation. Thus, the sense and antisense transcripts serve as the basis for protein-mediated toxicity, RNA-mediated toxicity, or both [9–11].

RNAs containing CAG repeats in coding transcripts express polyglutamine in the sense strand (Fig. 1), but by RAN-translation can also produce polyalanine and polyserine, exacerbating the potential toxicity [12,13]. At the same time, expanded CUG non-coding RNAs form ribonuclear foci and mislocalize or misregulate

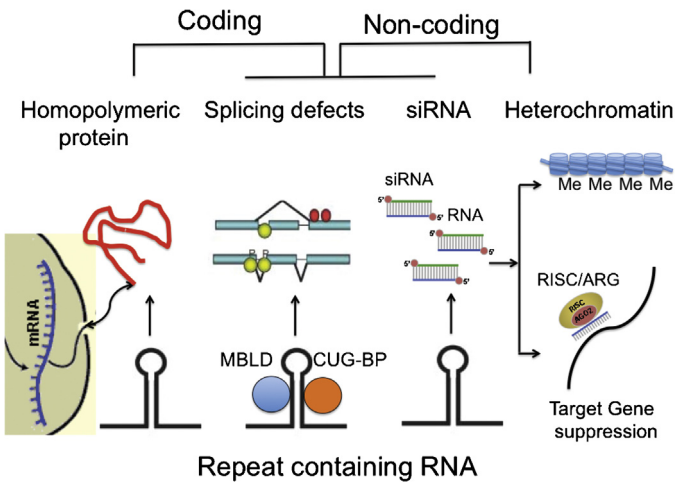

Fig. 1. Possible fates of bidirectional RNA transcripts at trinucleotide repeats. Coding and non-coding RNAs from TNR repeats form secondary structure with several possible fates. (Homopolymeric proteins) The RNA could serve as a coding template for homopolymeric toxic proteins (red squiggle). (Splicing defects) TNR RNA binds to RNA splicing factors Muscleblind (MBLD) (blue circle) and CUG binding proteins (CUG-BP) (orange circle) to alter RNA splicing and binding of other splicing factors (yellow and red balls). (SiRNA) A non-coding RNA hairpin is cleaved by DICER into short 21 nt double stranded RNA. The siRNA can induce silencing in cis, by formation of heterochromatin formation, or in trans by targeting complementary sequences of other genes for degradation *via* an RNA-I silencing complex (RISC) and Argonaute (AGO) complex. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

RNA binding proteins, notably splicing factors (Fig. 1) [9–12]. Two key splicing factors exhibit altered affinities for RNA as a consequence of CUG repeats [22,23]. CUG-BP1 (CUG binding protein 1) is a member of the CUG-BP/ETR-3-like (CELF) family of proteins involved in RNA splicing, RNA editing, and translation [22]. The second splicing regulatory protein is muscleblind-like (MBNL) [24]. Although discovered first in conjunction with DM1, MBNL is implicated in multiple TNR expansion diseases [24]. Both CUG-BP1 and MBNL bind to CUG repeats, but the two proteins antagonistically regulate alternative splicing [23]. In the case of the skeletal muscle chloride channel *CLCN1/ClC-1*, for example, it is alternatively spliced in DM1 to include exon 7A [25]. Elevation of this splice form leads to a reduction in *CLCN1* function, which contributes to myotonia [25]. CUG RNA binding to MBNL represses inclusion of exon 7A, while binding of CELF4 proteins promotes inclusion [25]. The two opposing effects occur through binding of the respective proteins to distinct regions of the RNA. The length of the repeat affects only MBNL binding [26]. Thus, as the CTG region expands, there is a gain-of-function in CELF4 and a loss-of-function in MBNL, conditions that tend to favor generation of the alternatively spliced forms of DM1 [27,28].

In general, the RNA-dependent mechanisms for toxicity generate multiple toxic sources. Expression of misfolded proteins, generation of siRNA, and aberrant splicing appear to be common features (Fig. 1). To illustrate them, we discuss a few examples of the bidirectional transcription of TNR disease genes, and consider how RNA processing plays a role in the toxicity of TNR genes.

2.1. RNAs that code for protein in both the sense and antisense directions

A natural antisense transcript exists at the HD repeat locus that contains the repeat tract [17]. The sense strand of the normal Huntington's disease gene (*htt*) harbors a CAG tract that codes for a polyglutamine region, the antisense HD transcript (*httas*) is 5′ capped, poly (A) tailed and contains three exons, alternatively spliced into two products: htt AS-1 (exons 1 and 3) and htt AS-2