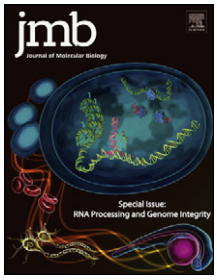
Review



R Loops and Links to Human Disease

Patricia Richard and James L. Manley

Department of Biological Sciences, *Columbia University, New York, NY 10027, USA*

Correspondence to Patricia Richard: *pr2232@columbia.edu*
http://dx.doi.org/10.1016/j.jmb.2016.08.031
Edited by Aguilera Andrés

Abstract

Aberrant R-loop structures are increasingly being realized as an important contributor to human disease. R loops, which are mainly co-transcriptional, abundant RNA/DNA hybrids, form naturally and can indeed be beneficial for transcription regulation at certain loci. However, their unwanted persistence elsewhere or in particular situations can lead to DNA double-strand breaks, chromosome rearrangements, and hypermutation, which are all sources of genomic instability. Mutations in genes involved in R-loop resolution or mutations leading to R-loop formation at specific genes affect the normal physiology of the cell. We discuss here the examples of diseases for which a link with R loops has been described, as well as how disease-causing mutations might participate in the development and/or progression of diseases that include repeat-associated conditions, other neurological disorders, and cancers.

Introduction

R loops are nucleic acid structures in which a nascent G-rich transcript hybridizes with the DNA template strand, leaving the non-template DNA single stranded [1]. R loops were first described in prokaryotes more than 20 years ago [2]. For a long time, they were considered largely as a by-product of transcription that did not have significant consequences and did not generate much interest. However, during the past decade or so, an increasing number of studies have revealed very important functions of R loops in transcription that when disturbed can be linked to a variety of diseases. Genome-wide mapping studies of R loops in human, mouse, and yeast cells by DNA/RNA immunoprecipitation followed by high-throughput sequencing (DRIP-seq) have tremendously helped to understand how and where R loops form [3–7]. Recent high-resolution DRIP-seq experiments suggest that R loops are abundant structures and are estimated to occupy up to 5% of mammalian genomes and 8% of the budding yeast genome [6,7].

R loops occur naturally during transcription and can have important functions. For example, they are important for class switch recombination of immunoglobulin (Ig) genes in activated B cells [8] and for mitochondrial replication [9,10]. R loops are also found frequently at GC-rich regions such as typify many promoters and 3′end regions, where they appear to play significant roles in transcription [3,4,7,11,12]. However, in some cases, persistence of R loops can have deleterious effects. For example, they can result in the accumulation of DNA double-strand breaks (DSBs) [13], leading to DNA rearrangements and genome instability [1,14]. It is thus not surprising that R loops have been linked to disease, including some cancers and several neurodegenerative disorders.

Several reviews describing how R loops form and how they can be resolved or lead to genome instability have appeared in the last few years [1,15–18]. The role of R loops in disease has also been recently reviewed [16,17,19]. However, examples of R loops in pathological conditions have continued to bloom, and here, we highlight recent advances in understanding how R loops are connected to disease.

R Loops and Genome Stability

Before delving into the interconnections between R loops and disease, we briefly highlight the many

connections between these structures and genome instability, as this is central to their role in many pathological conditions. The first strong links to genome instability were discovered over 10 years ago, in yeast and chicken cells. In yeast, deletion of *hpr1*, which encodes a component of the THO complex that couples transcription with mRNA export, was found to induce the inhibition of transcription elongation, DNA damage, and hyper-recombination through the formation of R loops [20]. The same connection was shown later with the mammalian THO complex [21]. In chicken DT40 cells, depletion of the splicing factor SRSF1 was shown to cause DSBs and DNA hypermutation, again due to the formation of R loops during transcription [22,23]. In fact, it is now known that many factors involved in RNA metabolism and processing play roles in preventing R-loop formation by packaging nascent RNAs into ribonucleoproteins (RNPs) that prevent its hybridization to the DNA template [20,23–26]. In the presence of R loops, the displaced ssDNA becomes sensitive to nucleases such as activation-induced cytidine deaminase, which leads to DSBs necessary to create hypermutations and diversity at the *Ig* locus in B cells [27]. XPF and XPG, two endonucleases of the nucleotide excision repair machinery, have also been found to process R loops into DSBs [13]. Such DSB-inducing mechanisms could participate in R-loop-dependent genome instability. Loss of topoisomerase I, which plays roles in replication and transcription, favors R-loop formation through increased levels of negative supercoiling. This can drive the stalling of both RNA polymerase (RNAP) and replication forks, leading to chromosome breaks [2,28,29]. Stalled replication forks are unstable structures prone to recombination and genome instability [30]. R loops have been found to cause replication impairments in various conditions and organisms, from bacteria to humans [31–34]. Collision between the transcription and replication machineries at common fragile sites in higher eukaryotes, the hot spot of chromosomal rearrangements, has been found responsible for R-loop formation and genomic instability [35]. However, it is unclear what comes first. Does the collision between RNAP and the replisome cause RNAP stalling and lead to R-loop formation? Or does R-loop formation cause RNAP stalling and then collision that arrests the replication fork? While evidence point to both possibilities, genomic instability appears to be the common outcome in the development of several diseases.

R-loop formation at trinucleotide repeat-associated diseases

More than 40 genetic disorders are caused by gene-specific repeat expansions. These include Huntington's disease [CAG repeats in *huntingtin* (*HTT*)], myotonic dystrophy type 1 [CTG repeats in *dystrophia myotonica protein kinase* (*DMPK*)], spinocerebellar ataxia type 1 [CAG repeats in *ataxin 1* (*ATXN1*)], fragile X mental retardation or fragile X syndrome (FXS) [CGG repeats in *fragile X mental retardation 1* (*FMR1*)], and Friedreich ataxia [GAA repeats in *frataxin* (*FXN*)] [36,37]. Those extended repeats can lead to various outcomes, such as transcription inhibition and expression of toxic RNAs or toxic polyglutamine proteins. A significant number of these have been linked to R loops (Table 1). For example, such repeats often have high GC content [36] and can potentially form R loops [38]. Indeed, *in vitro* transcription of CAG, CTG, GAA (not so high GC content), CGG, and CCG repeats can form R loops in *Escherichia coli* [39]. R loops formed *in vitro* and in bacteria at GAA repeats have been linked to RNA polymerase arrest [40]. Using a similar system, CTG repeats lead to the formation of R loops, promoting repeat instability in bacteria and human cells carrying transgenic repeats [41]. Repeat instability leading to expansion or deletion of the repeats has been shown in other disease models in bacteria [42], flies [43], and human cells [44]. Very often, the repeats are transcribed in both directions, which is believed to stimulate repeat instability [39,44–47]. This increased instability has been proposed to result from the formation of R loops on both strands, referred to as double R loops [39,48]. Interestingly, R-loop composition affects instability by leading to repeat expansion or deletion [48].

R loops have been well documented in the *FXN* and *FMR1* genes. Using transformed lymphoblastoid cell lines derived from FRDA and FXS patient cells, Groh *et al.* confirmed that highly stable R loops form at the expanded repeats of both genes [49]. FRDA is the most common inherited ataxia, caused by the GAA expansion in the first intron of *FXN*, leading to transcriptional silencing of *FXN* [50]. Importantly, R-loop enrichment has been found to be associated with the repressive H3K9me2 chromatin mark at *FXN* in Friedreich ataxia cells [49]. Interestingly, H3K9me2 is also found to be associated with R loops forming at gene 3′ ends, perhaps contributing to efficient transcription termination by promoting RNAPII pausing and release from the DNA template [11,12]. While a decrease in H3K9me2 does not affect R-loop or mRNA levels, increased R-loop formation, by the addition of the topoisomerase I inhibitor camptothecin, leads to increased H3K9me2 and transcription inhibition, suggesting that histone modification is a consequence of R-loop formation [49]. Expanded repeats at *FXN* appear to be another example of R-loop-dependent transcription elongation inhibition [20,51].

The scenario at *FMR1* appears quite different. The *FMR1* promoter and repeat region (in the 5′UTR) are hypermethylated in FXS patient cells containing more than 200 repeats (full mutation), resulting in the repression of *FMR1* transcription [52,53]. The