INTERNATIONAL JOURNAL OF APPLIED BIOLOGY AND PHARMACEUTICAL TECHNOLOGY

www.ijabpt.com Volume-5, Issue-2, April-June-2014 Coden : IJABPT Copyrights@2014

ISSN : 0976-4550

Received: 12th Feb-2014 *Revised: 24th March-2014* *Accepted: 26th March-2014*

Review article

R-LOOP HYBRID STRUCTURE FORMATION AND THEIR ROLE IN HEALTH AND DISEASE; POSSIBLE IMPLICATIONS FOR UNUSUAL NUCLEIC ACID STRUCTURE FORMATION?

Manoj G Tyagi

Department of Pharmacology, Christian Medical College, Vellore 632002, Tamilnadu

Author for Correspondence: E-mail: tyagi243@yahoo.co.in

ABSTRACT

The R-loop is a stable RNA–DNA hybrid structure in which the RNA strand is base-paired with one DNA strand of a DNA duplex, leaving the opposite DNA strand single-stranded. This structure can be involved in the hyper-mutation and double stranded DNA breaks in mammalian immunoglobulin (*Ig*) genes, oncogenes and neurodegenerative disease related genes. R-loops have not been studied at the genome scale extensively. It has been shown that many oncogenes and tumour suppressors (e.g. *Tp53, BRCA1, BRCA2 and Ptprd*) and neurodegenerative diseases related genes (e.g. *ATM*, *Park2*, and *GLDC*) could be prone to significant R-loop formation. The recent findings suggest that R-loops provide a novel level of RNA–DNA interactome complexity, playing key roles in gene expression controls, mutagenesis, recombination process, chromosomal rearrangement, alternative splicing, DNA-editing a R-loops have been described *in vivo* at the immunoglobulin class switch sequences and at prokaryotic and mitochondrial origins of replication. However, the biochemical mechanism and determinants of R-loop formation are unclear and how also they can affect epigenetic modifications needs to be elucidated. R-loop hybrid structure could be used as a novel source of prospective therapeutic targets.

INTRODUCTION

The R-loop structure was first characterized about 36 years ago (Roshbash *et al* 1979, Woodward & Roshbash, 1979) as described in Figure 1. Previous studies of R-loop focused on the development of 'R-loop hybridization technique' for visualization of the genetic organization of ribosomal RNA genes in yeast using electron microscopy. The application of this technique also led to the discovery of intron by the observation of splicing of adenovirus 2 late mRNA using a electron microscope (Chow *et al* 1977).Since then many subsequent applications of R-loop hybridization have been developed, which are now widely used for the study of gene structure. In 1995, Drolet and co-workers first demonstrated that R-loop existed *in vivo* in the bacterial cell. In this study, the R-loop formation was shown to be a consequence of transcription process that resulted in hybridization between nascent RNA transcript and DNA template, therefore such a process was called 'co-transcriptional R-loop' formation. R-loops occur *in vivo* within sequences that generate G-rich transcripts at the prokaryotic origins of replication, mitochondria and mammalian immunoglobulin (Ig) class switch sequences (Roy *et al* 2008). The two possible mechanisms of R-loop formation proposed by Lieber and Roy (2009) are 'thread back' and 'extended hybrid' mechanisms (Roy *et al* 2010). According to the thread back mechanism a nascent RNA is single-stranded for a short period of time and later anneals with the template DNA strand. In the extended hybrid mechanism, the nascent RNA that forms upon transcription fails to denature from the template in the transcription bubble, due to the high thermodynamic stability between RNA– DNA hybrids. Until recently, the studies of R-loops have provided various examples of significance of RNA–DNA interactions in a cell. The formation of R-loops during replication process in both prokaryotes and eukaryotes may lead to replication blockage and can be lethal if left unresolved (Camps and Loeb, 2005).

R-loop occurrence in neurodegenerative diseases:

A number of studies here proposed and revealed that R-loop formation structure is involved in transcription related mutation (TAM) (Lin *et al* 2010). Recent studies demonstrated a correlation between R-loop formation and activation-induced deaminase (AID) activity, the enzyme which is involved in generation of mutations and recombination events in oncogenes, such as Bcl6 and Myc (Duquette *et al* 2005), and also may affect genome instability. Interestingly, R-loops are often associated with neurodegenerative diseases, including spinocerebellar ataxia type 1 (SCA1), myotonic dystrophy (DM1) and fragile X type A (FRAXA). R-loop forming structures can be found in the Fmr1 and Fxn genes that are responsible for causing neurodegenerative disease. It was demonstrated that R-loops could co-localize with some classes of trinucleotide repeat tracks that occur in these genes (McIvor *et al* 2010). R-loop structures are found when Fmr1 and Fxn genes are transcribed. The RNA–DNA hybridization via R-loop mechanism can produce genetic instability that may be associated with the expansion of the trinucleotide repeats within the disease related genes (Naik *et al* 2010).

G4 DNA structures and the influence of R-loop formation in Genome:

G-quadruplexes form *in vitro* in guanine-rich sequences that contain four tracts of at least three guanines separated by other bases and are stabilized by G-quartets (Burge *et al* 2006). The G-quartets arise from the association of four guanines into a cyclic Hoogsten hydrogen-bonding arrangement in which each guanine base makes two hydrogen bonds with its neighbor using different hydrogen-bonding positions to the canonical Watson-Crick base pairing (Watson and Crick, 1953). The planar G-quartets stack on top of each other, giving rise to four-stranded helical structures. These structures, called G-tetraplex, G-quadruplex, or G4 DNA, may involve intramolecular or intermolecular interactions, and the phosphodiester backbones of the four participating strands may be in parallel or antiparallel orientation.

It is suggested that G-quadruplexes impose a structural barrier to DNA replication and various nucleic acid processing enzymes and are a potential source of genetic instability if not resolved. Identification of several DNA helicases that efficiently unwind and disrupt G4 DNA indicates that eukaryotic cells possess the mechanism for resolution of G4 DNA structures (Sharma, 2011). The R-loops only form when in vitro transcription occurs in the direction that results in a G-rich transcript. There are not many studies of how G-rich or how long the regions must be, nor has there been any sequence modification to assess any aspect of these G-rich regions for their propensity to form R-loops.

R-Loop Formation at Immunoglobulin Class Switch Sequences:

Ig class switch recombination (CSR) sequences have been the focus of most of these *in vitro* studies (Reaben *et al* 1994), although studies on mitochondrial and prokaryotic replication origins have also been done (Daniels and Lieber,1995). Ig CSR occurs at switch regions. In mammals, recombination occurs between the Sμ region, which is located upstream of the constant exons encoding Igμ heavy chain, and any one of the downstream switch regions, Sα, Sγ, or Sε, which are located upstream of the constant exons encoding the Igα, Igγ, and Igε heavy chains, respectively . In mammals, the Ig switch regions are usually several kilobases in length, G-rich on the nontemplate strand (thereby generating a G-rich RNA transcript), and repetitive with a repeat length of approximately 25 and 80 bp. Many of the G's are in clusters of 2 to 5 nucleotides (nt). This suggests a physiologically important role of G clustering at these sequences.

R loops and its role in transcription and genomic instability:

The R loops can accumulate naturally at the G-rich 5′-UTR regions immediately downstream of the CpG-non-methylated promoters in humans (Ginno *et al*., 2012). It has been proposed that the displaced ssDNA in the R loop acts as a signal to recruit either the protective H3K4 trimethyl mark or the DNA demethylases complex (Ginno et al., 2012). Thus, R loops do not only play roles at the 3′-UTR of the genes but may also form at the promoter region leading potentially to the activation/inactivation of genes. This observation is also particularly relevant because it is the first case in which an R loop is shown to form as a natural event to control an epigenetic mark (Reddy *et al* 2010). On the other hand, co-transcriptional R loops are linked to different forms of genome instability including mutations, recombination, and chromosome rearrangements as well as chromosome loss. R loops provide the appropriate substrate for the mutagenic action of specific DNA-modifying enzymes such as AID (Aguilera and Garcia-Muse, 2012). However, apart from AID-mediated specific events, high spontaneous mutagenicity is one of the consequences of R loops, as shown after ASF/SF2 depletion in DT40 cells. The single stranded DNA displaced by the RNA:DNA hybrid may be critical for such mutagenicity, provided that ssDNA is more susceptible to mutagenic DNA damage than double stranded DNA (Xu and Clayton, 1996).

Nevertheless, it is possible that R loops could be a source of unscheduled mutagenic DNA replication. G-rich sequences immediately downstream of the poly (A) signal are common in mammalian genes and potential G-quadruplex-forming sequences are enriched at the 3′-UTR regions of genes. Given the potential of G-rich sequences to stabilize R loops, it might be possible that R loops are intrinsic elements of termination pause sites. Consistently, in T7 RNAP *in vitro* systems transcription has a tendency to stop at G-rich sequences, in which unusually stable R loops are formed (Aguilera and Gomez-Gonzalez, 2008). As with other structures formed by trinucleotide repeat DNAs, such as slipped DNAs, cruciforms and triplexes, etc., the aberrant processing of these R-loops could give rise to repeat length changes. Such aberrant processing may involve transcription-coupled nucleotide excision repair factors that recognize, bind and attempt to repair the R-loops during transcription similar to the base excision repair (Subathra devi *et al* 2014). Nucleotide excision repair proteins have been shown to affect transcription-enhanced genetic instability of CAG/CTG repeats in bacterial, fly and mammalian cell models (Panigrahi et al 2005, Lin *et al* 2006).

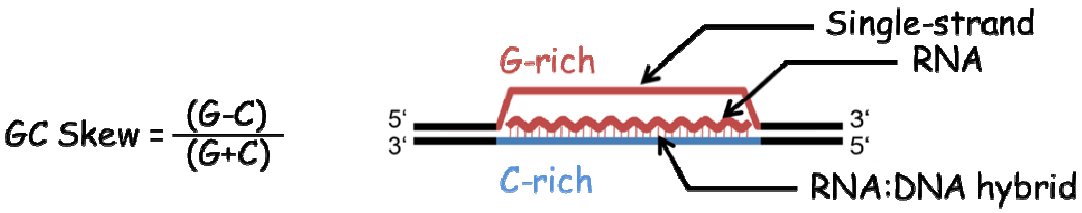

Figure 1: R Loop a three stranded nucleic acid structure

Fig. Courtesy Dr.Frederic Chedin, Department of Molecular and Cell Biology, Univ.of California at Davis, USA

CONCLUSION

The prediction of R-loop formation in over half of the human genes reveals a novel level of RNA–DNA interactome complexity that perhaps will lead to a better understanding of the role of R-loop forming structure in gene expression controls and epigenetic modifications. The specific conformation of RNA–DNA hybrid formation also provides a unique target for controlling the transfer of genetic information through binding by small molecules. I expect that R-loop database will help researchers in the R-loop analysis and design of the experiments aimed to discover mutated sites and epigenetic modifications in R loop formation in identified genes. It is now possible to suggest that R-loop database will be useful for drug discovery and identification of new classes of therapeutic targets.

REFERENCES

Aguilera A, Gómez-González B (2008). Genome instability: a mechanistic view of its causes and consequences. Nat Rev Genet. Mar; 9(3):204-17.

Andrés Aguilera,Tatiana García-Muse (2012). R Loops: From Transcription Byproducts to Threats to Genome Stability. Molecular Cell, Volume 46, Issue 2, 115-124, 27

Burge S, G. N. Parkinson, P. Hazel, A. K. Todd, and S. Neidle (2006). "Quadruplex DNA: sequence, topology and structure," Nucleic Acids Research, vol. 34, no. 19, pp. 5402–5415.

Camps, M. and Loeb, L.A. (2005) Critical role of R-loops in processing replication blocks. Front Biosci, 10, 689–698.

Chow,L.T, Gelinas,R.E, Broker,T.R. and Roberts, R.J. (1977). An amazing sequence arrangement at the 50 ends of adenovirus 2 messenger RNA. Cell, 12, 1–8.

Daniels GA, Lieber MR. RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. Nucleic Acids Res. 1995;23:5006-5011.

Drolet,M., Phoenix,P, Menzel,R, Masse,E, Liu,L.F. and Crouch,R.J. (1995) Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. Proc. Natl Acad. Sci. USA, 92, 3526–3530.

Duquette,M.L., Pham,P., Goodman,M.F. and Maizels,N. (2005) AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hyper mutation. Oncogene. 1; 24(38):5791-8.

Ginno P A., P L. Lott, H C. Christensen, I Korf, and Frederic Chedin (2012). R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. Mol. Cell. 2012 Mar 30;45(6):814-25

Lin,Y., Dent,S.Y, Wilson,J.H., Wells,R.D. and Napierala,M. (2010) R loops stimulate genetic instability of CTG.CAG repeats. Proc. Natl Acad. Sci. USA, 107, 692–697.

Lin Y, Dion V, Wilson JH (2006).Transcription promotes contraction of CAG repeat tracts in human cells. Nat. Struct. Mol. Biol.13:179-180.

McIvor,E.I, Polak,U. and Napierala,M. (2010) New insights into repeat instability: Role of RNA.DNA hybrids. RNA Biol., 7, 551–558.

Naik, A.K, Lieber,M.R. and Raghavan,S.C. (2010) Cytosines, but not purines, determine recombination activating gene (RAG)- induced breaks on heteroduplex DNA structures: implications for genomic instability. J. Biol. Chem., 285, 7587–7597.

Panigrahi GB, Lau R, Montgomery SE, Leonard MR, Pearson CE (2005). Slipped (CTG)*(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. Nat. Struct. Mol. Biol. 12:654-662.

Reaban, M. E, J. Lebowitz, and J. A. Griffin (1994). Transcription induces the formation of a stable RNA: DNA hybrid in the immunoglobulin alpha switch region. J. Biol. Chem. 269:21850–21857.

Reddy,K, Tam,M, Bowater,R.P, Barber,M, Tomlinson,M., Nichol Edamura,K, Wang,Y.H. and Pearson,C.E. (2010) Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. Nucleic Acids Res., 39, 1749–1762.

Rosbash,M, Blank,D, Fahrner,K., Hereford,L, Ricciardi,R, Roberts,B, Ruby,S. and Woolford,J. (1979) R-looping and structural gene identification of recombinant DNA. Methods Enzymol, 68, 454–469.

Roy,D. and Lieber,M.R. (2009) G clustering is important for the initiation of transcription-induced R-loops *in vitro*, whereas high G density without clustering is sufficient thereafter. Mol. Cell. Biol., 29, 3124–3133.

Roy,D., Yu,K. and Lieber,M.R. (2008) Mechanism of R-loop formation at immunoglobulin class switch sequences. Mol. Cell. Biol., 28, 50–60

Roy,D., Zhang,Z, Lu,Z, Hsieh,C.L and Lieber, M.R. (2010) Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. Mol. Cell Biol., 30,146-159

Sharma S. (2011). Non-B DNA Secondary Structures and Their Resolution by RecQ Helicases. Journal of Nucleic Acids, 1-15

Subthara Devi C, Kritika Kedarinath, Payal Choudhary, VishakhaTyagi, Mohanasrinivasan. V (2014).Liposome mediated drug delivery for leukocyte adhesion deficieny I (LAD I): Targeting the mutated gene *ITGB2* and expression of CD18 protein. Frontiers in Biology. Volume 9, Issue 1, 1-4

Xu, B, and D. A. Clayton (1996). RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving at primers. EMBO J. 15:3135–3143.

Watson JD, Crick FH (1953). A Structure for Deoxyribose Nucleic Acid. Nature 171, 737-738

Woolford, J.L. Jr and Rosbash, M. (1979). The use of R-looping for structural gene identification and mRNA purification. Nucleic Acids Res., 6, 2483–2497.