What Really Rigs Up RIG-I?

Sailen Barik

Department of Biological, Geological and Environmental Sciences, and Center for Gene Regulation in Health and Disease, Cleveland State University, Cleveland, Ohio, USA

Abstract

RIG-I (retinoic acid-inducible gene 1) is an archetypical member of the cytoplasmic DEAD-box dsRNA helicase family (RLR), the members of which play essential roles in the innate immune response of the metazoan cell. RIG-I functions as a pattern recognition receptor that detects nonself RNA as a pathogen-associated molecular pattern (PAMP). However, the exact molecular nature of the viral RNAs that act as a RIG-I ligand has remained a mystery and a matter of debate. In this article, we offer a critical review of the actual viral RNAs that act as PAMPs to activate RIG-I, as seen from the perspective of a virologist, including a recent report that the viral Leader-read-through transcript is a novel and effective RIG-I ligand.

Introduction

Collectively, the RIG-I-like receptor (or RLR) family consists of pattern recognition receptors (PRRs) that detect cytoplasmic RNA products of a variety of RNA viruses [1–4], such as influenza (an Orthomyxovirus), various members of the Paramyxoviridae family and flaviviruses, such as Dengue and West Nile, to name a few. The major functional domains in RIG-I are two N-terminal CARD domains, a central ATP-dependent RNA helicase domain, and the C-terminal regulatory domain (RD; schematically depicted in fig. 1). Biochemical and structural studies have shown that in its inactive state, RIG-I exists in a closed conformation such that the CARD domains are sequestered by the helicase domain. Upon recognition of the proper pathogen-associated molecular pattern (PAMP) RNA serving as a RIG-I ligand, RIG-I hydrolyzes ATP and undergoes a series of allosteric changes that eventually release the CARD domains free [5–9]. A predominant feature among the majority of RIG-I ligand RNAs is a 5′-triphosphate (5′ppp); recent studies of 5′ppp dsRNA-RIG-I cocrystals have in fact revealed that the α- and β-phosphates of 5′ppp are cradled inside the basic groove of the RD [5–8]. The activated RIG-I is recruited from cytosol to specialized mitochondria-associated membranes via interactions between the opened CARD domains of RIG-I and those of the mitochondrial MAVS/IPS-1 [3–12]. The heterotypic CARD-CARD interaction then initiates a kinase cascade, eventually leading to the activation of NFκB and the IFN regulatory factors, IRF3 and IRF7. The activated NFκB and IRF3/7 translocate to the nucleus and interact with the promoter regions of target genes, including the type I interferons, IFNα and IFNβ, and inflammatory cytokines. The elaborated IFNs bind to their cognate receptors on...
the cell membrane and the resultant signaling cascade induces the expression of hundreds of IFN-stimulated genes, many of which are antiviral factors, leading to the establishment of an antiviral state [13]. The cardinal antiviral role of the RIG-I signaling pathway is also underscored by the fact that many viruses, especially the RNA viruses, encode mechanisms that block RNA detection by RIG-I [14]; reciprocally, loss of the blocking mechanism increases the IFN response, leading to a reduction of viral replication and pathogenicity.

RIG-I was originally discovered as a dsRNA-activated PRR more than a decade ago [1], and its preference for 5′ppp but not 5′cap was unraveled shortly thereafter [15–17]. However, the early studies often used synthetic or in vitro transcribed RNA molecules, some with uncertain features and questionable relevance to viral RNA [18–22].
Many of these RNA were purified, prefolded and transfected into uninfected cells, out of context of virus infection, and thus may not have replicated the nascent folding of the RNA or its association with viral proteins. Nonetheless, accumulated evidence has revealed an amazing diversity in the structures of RNA that can activate RIG-I. The current list of RIG-I-activating RNA PAMPs includes but is not limited to [3, 20, 23]: 5′ppp adjoining the double-stranded stalk of a hairpin; 5′ppp single-stranded RNA; 5′ppp RNA in a large panhandle structure; 5′pp (diphosphate) dsRNA viral genome; internal RNA sequences, such as poly-U/A-rich areas, likely making the 5′ termini less relevant; blunt-ended dsRNA with 5′p (monophosphate) or 5′OH (no phosphate) and 3′OH; various 3′ppp RNA, both single and double stranded, and occasional mRNA, supposedly containing 5′cap. Moreover, RIG-I was shown to prefer RNAs that are less than 300 base pairs long, although this also seems to be variable, perhaps depending on other structural features of the RNA [2, 17, 22–26].

Here, we keep this diversity in the background and review the various viral RNAs that act as PAMPs to bind and activate RIG-I. We focus mainly on nonsegmented (or unsegmented) negative-strand RNA (NNR) viruses of the Paramyxoviridae family, the largest family of RNA viruses studied in reference to RIG-I recognition.

The RNA Landscape of Nonsegmented NNR Viruses

NNR viruses by definition contain NNR genome (antigenome sense), tightly wrapped (encapsidated) by the viral nucleocapsid protein N, generating the N-RNA template [27]. The entire life cycle of these viruses occurs in the cytoplasm. Following infection, the genomic N-RNA is transcribed in the cytoplasm by the viral RNA-dependent RNA polymerase (RdRP) that initiates transcription from the single promoter at the 3′ end of the viral genome (fig. 1). The RdRP proceeds in a stop-start fashion, stopping at each gene end and resuming at the next gene start sequence [28–33]. A certain fraction of transcribing RdRP apparently falls off at each intergenic region, leading to the polarity of expression, such that there is a gradient of expression from the promoter proximal to the distal end [28–34]. In other words, the genes closest to the promoter (3′ end of the genome), such as the Leader RNA (LeRNA), are expressed most abundantly and those that are farther away are expressed in gradually diminishing amounts. Later in the life cycle, the viral RdRP switches from transcription to replication mode, in which all the gene-start and gene-end signals are ignored and a full-length N-wrapped copy of the genome is made, which then serves as a template for replication of more negative-sense genomes. The molecular detail of the transcription-to-replication switch remains foggy, but the N protein is known to play a role [28, 35, 36]. In this mechanism, when de novo translation of the newly synthesized N mRNA generates sufficient N protein, it is available to encapsidate the nascent LeRNA, which somehow suppresses transcription termination. This mechanism guarantees that replication will occur late in the viral life cycle only when enough N is available to encapsidate the product. In other words, the very process of their synthesis guarantees that all full-length genome and antigenome RNA are wrapped with N protein.

Evaluation of NNR Viral RNAs as a Potential RIG-I Ligand

The three well-known major classes of viral RNA are the genome/antigenome RNA, mRNA, and small transcripts, such as the LeRNA; to this list, we will add the newly appreciated Leader-read-through (LeRT) RNA. We will briefly review their candidacy for RIG-I activation, and compare them with the known RNA triggers of RIG-I.

Viral Genome RNA

The two termini of the NNR viral genome and antigenome have partial complementarity because of the necessity of having transcriptional and replicational promoters at the ends, both of which must recruit the viral RdRP. Both the genome and the antigenome RNA are uncapped, with 5′ppp RNA ends, and are not translated. The full-length NNR viral genome and antigenome also exist exclusively as N-RNA ribonucleoprotein [27] and not as naked RNA. X-ray crystallography of actual N-RNA complexes of three NNR viruses (rabies, vesicular stomatitis virus, measles) clearly revealed a consensus RNP (ribonucleoprotein) structure in which several of the bases are completely shielded by the N protein such that Watson-Crick base-pairing may not occur and the rigid ribose-PO₄ backbone would not favor formation of an RNA duplex [37–39]. However, the possibility cannot be ruled out that a small number of complementary bases in the N-RNA genome termini are exposed for base-pairing, or a small fraction of the genome termini may be partially devoid of encapsidation. Even so, it is unclear whether the combination of the imperfect terminal
dsRNA region and 5’ppp will suffice to recruit and activate RIG-I.

Nonetheless, recent studies have shown that NNR viral infection can induce IFN even when the virion is pre-treated with UV light [40] or infection is conducted in the presence of cycloheximide [41], and that the IFN induction is in fact RIG-I dependent. UV light generates cross-linking in the genome RNA and strongly reduces gene expression; cycloheximide inhibits protein synthesis, and, as mentioned before, synthesis of new N protein is required for NNR genome replication. However, these treatments may still allow the synthesis of LeRNA or Leader-read-through RNA (see below). The Leader gene may escape UV damage due to its small target size, and inhibition of translation should not prevent primary transcription of the NNR viral genome. Of note, panhandle structures are well established in segmented RNA viral genomes, such as influenza [42], where it has also been shown to activate RIG-I; however, there are conflicting reports on whether the influenza viral genome RNP can activate RIG-I by itself [41] or requires a transcriptional product [43].

Defective Interfering Particle Genome
Most NNR viruses generate defective interfering (DI) particles as a result of template switching by the viral RdRP, resulting in a ‘copy-back’ product with high self-complementarity [44]. However, like its full-length counterpart, the DI genome is also considered wrapped with the N protein. Although DI particles have been implicated in IFN induction for a long time [44–46], its exact PRR has received attention only recently. In one of the first direct in vivo studies [47], endogenous RNA-RIG-I complexes were isolated from cells infected with influenza virus and Sendai virus (an NNR virus of the Paramyxoviridae family), and the RNAs were characterized by deep sequencing. RIG-I was indeed found to be specifically and preferentially associated with the copy-back Sendai virus DI particle RNA and not with the full-length viral genome or mRNAs. In the influenza virus-infected cells, RIG-I again preferentially associated with DI RNAs, but also with shorter genomic segments, since the influenza viral genome, like DI particles of NNR viruses, contains a panhandle of near-perfect terminal complementarity. It was, however, not known whether the RIG-associated DI genome RNA was encapsidated or naked.

Viral mRNAs
All NNR viral mRNAs possess standard features of eukaryotic mRNA, including 5’ cap and 3’ poly(A), and therefore cannot be differentiated as a PAMP and should not activate RIG-I. In fact, in direct tests, 5’ capped mRNA failed to bind RIG-I [16, 17]. In measles, which is an NNR virus, methodical screening of each gene mRNA clearly showed their inability to activate IFNβ [15]; although a subsequent study [47] showed that internal AU-rich sequences in the large L gene mRNA can activate RIG-I in vivo, its molecular mechanism remains undetermined. Lastly, unlike cellular mRNA, the NNR viral mRNAs are transcribed in the cytoplasm and should be covered by translating ribosomes, perhaps even as being transcribed (nascent). Thus, cytoplasmic viral mRNA transcription and translation are ‘coupled’, as in the prokaryotes. For these reasons, viral mRNAs by and large are not major triggers of RIG-I.

Viral LeRNA
Interestingly, the most abundant NNR viral transcript is a short noncoding RNA, generated from the most promoter-proximal transcription unit, known as Leader [48, 49]. The LeRNA is relatively short, roughly 45–50 nt long in different viruses, and is generated by multiple rounds of transcription initiation (fig. 1). Being neither 5’ capped nor polyadenylated, but containing 5’ppp, the LeRNA is currently of unknown function (fig. 1), and as a noncoding RNA it is neither translated nor shielded by ribosomes. The LeRNA, such as that of measles virus and Ebola viruses, received initial attention as a candidate RIG-I trigger because naked LeRNA with its 5’ppp, when introduced into uninfected cells, activated RIG-I and induced IFN [15–18]. Another study also showed that transfected LeRNA of rabies, measles and Ebola viruses induced IFN, and that the poly-U/A-rich regions in these leaders are important for this induction [50]. However, the short length of the LeRNA and lack of significant self-complementarity makes it unlikely that it will have a double-stranded structure that is required for efficient RIG-I recruitment [18]. We have shown that native respiratory syncytial virus (RSV) LeRNA, naturally produced in the RSV-infected cell, is not accessible to RIG-I because it is promptly wrapped and shielded by cellular La antigen (protein) [51]. Later, when enough viral N protein is made, it competes with La to bind LeRNA. Thus, the LeRNA in the RSV-infected cell exists either as La-RNA or N-RNA ribonucleoprotein complex, neither of which activates RIG-I or IFN (fig. 1) [51, 52].

Viral LeRT RNA
The most recent hunt for RIG-I ligands have implicated the NNR viral LeRT transcript [53]. Specifically, a
fraction of the LeRNA that reads through its transcription termination site into the next gene, generating what we will call ‘Leader-read-through’ or LeRT RNA, was shown to be a major inducer of IFN via RIG-I [53]. These authors used Newcastle disease virus, an NNR virus, in which the leader sequence is followed by the nucleocapsid protein gene, N. They identified the Newcastle disease virus LeRT RNA as the 1.8-kb-long Le-N RNA (GenBank No. GQ849007.1), and found that it contains 5’ppp and 3’poly(A’). This RNA and the viral replication complex was shown to actually colocalize in cytoplasmic antiviral stress granules along with cellular RIG-I [53]. Similar IFN-inducing read-through RNA species were also confirmed in two other NNR viruses, namely vesicular stomatitis virus (Le-N) and RSV (Le-NS1). In what follows, we discuss the implications of these findings in light of the biogenesis and unique properties of LeRT.

Transcription termination sites have an intrinsic efficiency of termination that may be regulated by multiple factors, such as template sequence, nascent RNA structure, and proteins that regulate termination. Extensive studies in prokaryotes have shown that the transcription termination efficiency determines the ratio of termination and read-through [54]. We note that the RdRP was first discovered in an NNR virus (vesicular stomatitis virus) nearly half a century ago [55] and its penchant for intergenic read-through has since been observed whenever an appropriate probe was used. In RSV, for example, read-through transcription was estimated to be as high as 12% [56, 57]. Read-through was observed essentially in all gene junctions, such as Le-Gene1, Gene1-Gene2, etc., although the absolute amount of read-through transcript is progressively lower for promoter-distal junctions due to the polarity effect that reduces the overall transcription of distal genes. Thus, in NNR viruses, Le-Gene1 RNA is the most abundant read-through transcript. Regardless of their abundance, the read-through transcripts in biology in general have garnered little attention. This is in part due to the general notion that 100% efficiency is the de facto goal of transcription termination and that any inefficiency of this process is an unintended failure. The LeRT RNA has suffered from the additional perception that it is a useless extension of the LeRNA, itself of unknown function.

What qualifies the LeRT RNA as a RIG-I agonist? Note that the LeRT RNA is essentially a chimera of LeRNA and the Gene1 transcript. We believe that the LeRT RNA uniquely combines the following features that make it well suited for RIG-I activation. First, transcription of LeRT is initiated in the same manner as the LeRNA, and thus, like the LeRNA, it too contains 5’ppp, an important signal for RIG-I recruitment. Second, because of its lack of 5’cap, it is not recruited by the translational machinery, and thus not covered by ribosomes. Third, unlike LeRNA [51], it should not bind the L protein. This is because L preferentially binds to RNA with a free 3’OH end [58], a feature possessed by LeRNA [51]. The LeRT, in contrast, contains a 3’poly(A) tail since it terminates at the standard transcription termination site of Gene1 [53, 56, 59]. Fourth, the LeRT is also not encapsidated by the viral N protein as it is produced by the RdRP in the transcription mode [59]. Finally, since it is much longer than the LeRNA [53, 56, 59], it may form a certain degree of double-stranded structure which, together with the 5’ppp, should facilitate optimal RIG-I recruitment [18]. This may even occur on the nascent LeRT before termination. The exact secondary structure of LeRT will vary with the LeRNA and Gene1 sequences, which may regulate the efficiency of RIG-binding among individual NNR viruses. In summary, the LeRT is a long noncoding RNA that has all the features of optimal RIG-I recruitment, a combination not found in any other NNR viral RNA species. As implicated before, a fraction of Le-Gene1 RT transcript may also read through the second termination site to generate Le-Gene1-Gene2 RNA, but in substantially lower amounts compared to Le-Gene1 RNA [56, 57]. It is, however, entirely possible that the Gene1-Gene2 intergenic sequence is less termination proficient in specific NNR viruses, and that in these viruses the Le-Gene1-Gene2 RNA may make significant contributions to RIG-I activation. In contrast, the Gene1-Gene2 read-through RNA, if produced, will contain 5’cap (from the Gene1 mRNA initiation) and will not activate RIG-I.

The NNR viral intergenic sequences are highly divergent, not only among different viruses but also within the same viral genome. Moreover, although the primary structure of the L protein, the major subunit of the viral RdRP, is highly similar in all NNR viruses [60], the accessory transcription factors such as the phosphoprotein P are highly dissimilar. Overall, it is imperative that the exact frequency of the read-through will vary among NNR viruses. As all NNR viral mRNAs are capped, it is reasonable to assume that internal cap-independent translation initiation of Gene1 does not occur from the LeRT RNA and, thus, the LeRT adds no obvious value to viral gene expression. In fact, as there is a single RdRP-loading site at the 3’ end of the viral genome, any RdRP that reads through Gene1 should cause a correspondingly lower level of the Gene1 mRNA for not initiating at the normal Gene1 start site. It is thus tempting to speculate that the
NNR viral termination frequency was regulated over the course of host-virus coevolution, such that the ratio between Le-Gene1 RT and Gene1 mRNA is fine-tuned to achieve the right balance between host antiviral defense and viral replication. A mutation in RSV L (Asn1049 → Asp) has been described that is termination deficient and produces elevated levels of read-through RNA [61]; it would be worthwhile to test whether infection by this virus elicits higher amounts of IFN. As the RdRP is absent in human cells, another tantalizing possibility is that the host cell may have evolved specific antitermination factor(s) to aid in the suppression of termination, thereby stimulating more read-through and enhancing the innate antiviral defense. Finally, reverse genetic mutation of the termination signal in recombinant viruses should overproduce the read-through RNA, leading to enhanced innate immunity and highly attenuated vaccine strains.

**Conclusion**

RIG-I, a proximal trigger of the IFN pathway, seems to be a versatile PRR in its ability to interact with a variety of viral RNA, serving as a PAMP. Much of our knowledge of the structural changes in RIG-I draws on the crystal structure of uncomplexed RIG-I (apo-RIG-I) compared with RIG-I bound to various immunostimulatory ligands, such as blunt-ended 5′ppp or unphosphorylated dsRNA. However, RIG-I is clearly activated by alternative RNA structures, likely due to the plasticity of its RNA-binding site. Multiple features of the PAMP RNA, such as 5′ppp and double-strandedness, may act synergistically to bind RIG-I, but it is possible that the lack of the less important one, such as the double-strandedness, can be compensated by internal poly-U/A-rich motifs in some sequence contexts. Even with the 5′ppp RNA ligand, the γ-phosphate makes no contact with RIG-I, explaining why 5′pp RNA can also serve as a ligand. The flexibility of RNA molecules may allow compensatory binding, even if at a lower affinity, if the concentration of the RNA is high enough. The molar levels of various NNR gene RNA species change over the virus replication cycle. The Leader and LeRT RNA, for example, are the most abundant early in infection, whereas genome-length RNA predominates at the late stage. Evidence is coming to light that the RNA viral genome generates a large number of short transcripts that were hitherto unknown [62–66]. While each transcript may activate RIG-I only weakly, all of them may collectively generate a physiologically relevant level of RIG-I activation. In interrogating an RNA’s ability to activate RIG-I, some of the confusion may have arisen from the fact the binding of RIG-I to the RNA is independent of ATP, but the subsequent translocation of RIG-I to the interior of the RNA, essential for RIG-I activation, requires ATP hydrolysis [67–69]. Thus, many RNA species may score positive in RNA-binding studies in vitro but may not actually activate RIG-I and, hence, score negative in downstream signaling assay, such as IFN induction. Recent studies have also revealed that the CARD domains of RIG-I, opened after RNA-binding, undergo several posttranslational modifications, which is important for downstream RIG-I signaling, mainly its docking to MAVS. These modifications include: ubiquitination by TRIM25 and the RNA-binding MEXC3 E3 ligases [70–74], phosphorylation by protein kinase C and casein kinase 2, and dephosphorylation by protein phosphatase 1 [75–78]. Regulation of any of these modifications by specific viruses, cell types or culture conditions may abrogate functional RIG-I signaling downstream of RNA binding, which will make the RNA appear ineffective as a RIG-I ligand, perhaps explaining some of the observed variance and diversity.

**Acknowledgements**

I sincerely thank Dr. Ganes Sen (Cleveland Clinic, Cleveland, Ohio, USA) and Dr. Malathi Krishnamurthy (University of Toledo, USA) for critical reading of the manuscript.

**References**

Viral Ligands of the RNA Helicase, RIG-I


