Cooperative RNA recognition by a viral transcription antiterminator

Ivana G. Molina, Sebastian A. Esperante, Cristina Marino-Buslje, Lucía B. Chemes, Gonzalo de Prat-Gay

PII: S0022-2836(18)30044-5
DOI: https://doi.org/10.1016/j.jmb.2018.01.018
Reference: YJMBI 65600

To appear in:

Received date: 5 December 2017
Revised date: 29 January 2018
Accepted date: 29 January 2018

Please cite this article as: Ivana G. Molina, Sebastian A. Esperante, Cristina Marino-Buslje, Lucía B. Chemes, Gonzalo de Prat-Gay , Cooperative RNA recognition by a viral transcription antiterminator. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Yjmbi(2018), https://doi.org/10.1016/j.jmb.2018.01.018

Cooperative RNA recognition by a viral transcription antiterminator

Ivana G. Molina[1#], Sebastian A. Esperante[1#], Cristina Marino-Buslje[2], Lucía B. Chemes[1*] and Gonzalo de Prat-Gay[1*].

[1] Protein Structure-Function and Engineering Laboratory, Fundación Instituto Leloir and IIBBA-CONICET, Av. Patricias Argentinas 435, 1405 Buenos Aires, Argentina

[2] Structural Bioinformatics Unit, Fundación Instituto Leloir and IIBBA-CONICET, Av. Patricias Argentinas 435, 1405 Buenos Aires, Argentina

# these authors contributed equally

* corresponding authors: gpg@leloir.org.ar; lchemes@leloir.org.ar.

KEYWORDS: $M_{2\text{-}1}$ protein, antiterminator, cooperativity, RNA binding, human respiratory syncytial virus.

ABSTRACT

RNA transcription of mononegavirales decreases gradually from the 3' leader promoter towards the 5' end of the genome, due to a decay in polymerase processivity. In the respiratory syncytial virus (RSV) and metapneumovirus, the $M_{2\text{-}1}$ protein ensures transcription anti-termination. Despite being a homotetramer, RSV $M_{2\text{-}1}$ binds two molecules of RNA of 13mer or longer per tetramer, and temperature sensitive secondary structure in the RNA ligand is unfolded by stoichiometric interaction with $M_{2\text{-}1}$. Fine quantitative analysis shows positive cooperativity, indicative of conformational asymmetry in the tetramer. RNA binds to $M_{2\text{-}1}$ through a fast bimolecular association followed by slow rearrangements corresponding to an induced-fit mechanism, providing a sequential description of the time events of cooperativity. The first binding event of half of the RNA molecule to one of the sites increases the affinity of the second binding event on the adjacent contacting protomer by 15-fold, product of increased effective concentration caused by the entropic link. This mechanism allows high affinity binding with an otherwise relaxed sequence specificity, and instead suggests a yet undefined structural recognition signature in the RNA for modulating gene transcription. This work provides a basis for an essential event for understanding transcription antitermination in pneumoviruses and is counterpart Ebola virus VP30.

INTRODUCTION

RNA is at the center of gene expression and regulation, and is almost always found associated to proteins for either function or protection. Gene expression in RNA viruses requires precise modulation of the relative amounts of its proteins. Paramyxoviruses display a marked transcription polarity from 3' to 5' ends, where the genes at the 3' end of the genome are expressed at a higher level [1]. Such is the case of the human respiratory syncytial virus (HRSV), the leading cause of acute respiratory tract infections in newborns, the elderly and immunocompromised. HRSV comprises a negative single stranded RNA genome of 15 kb, tightly wrapped around the nucleocapsid N protein, the template for transcription and replication by the RNA dependent RNA polymerase (RdRp) L [2]. The polymerase replication complex is composed of L, N and the phosphoprotein polymerase cofactor P. An additional gene product, the $M_{2\text{-}1}$ protein unique to pneumoviruses, is added in the transcription complex. In fact, RSV and its close relative metapneumovirus (MPV) were recently classified into a new family, pneumoviridae, in part because of the singular way they exert control of transcription through the $M_{2\text{-}1}$ protein, being closer to filoviruses (ebola virus) which bear a protein, VP30 with a similar function [3].

The transcriptase mode of RdRp produces the sequential transcription, capping and polyadenylation of ten subgenomic mRNAs, recognizing cis-acting signals flanking each gene. The gene-start (GS) and gene-end (GE) sequences determine the transcription initiation and termination sites for each sequential open reading frame [4]. The $M_{2\text{-}1}$ protein was shown to be essential for transcriptase activity, preventing premature transcriptional termination of long mRNAs acting as intragenic antiterminator and increasing the synthesis of polycistronic read-through mRNAs acting as intergenic antiterminator [4; 5]. It was proposed that $M_{2\text{-}1}$ allows the transcription of promoter-distal genes to counteract the transcriptional attenuation at each gene junction, and down-regulates the efficiency of translation of the second cistron in bicistronic read-

through transcripts [6].

While M_{2-1} is required for transcriptional antitermination, it is dispensable for RNA replication [5; 7]. The overall regulation of transcription and replication of RSV and paramyxoviruses was recently reviewed [8].

Seminal work on M_{2-1} activity showed that: i) the synthesis of long RNAs was much more positively affected than that of short RNAs, ii) processivity enhancement or antitermination by M_{2-1} varied greatly for the different naturally occurring gene junctions, iii) the synthesis of the first two genes (NS1 and NS2, unique to RSV) was independent of M_{2-1}, while all downstream genes were highly dependent [5].

We have previously shown that the RSV M_{2-1} tetramer is highly stable (K_D 10^{-38} M^3) with a strong pH dependent dissociation and unfolding [9]. M_{2-1} interacts with the P tetramer through an intriguing 1:1 tetramer:tetramer interface [10], and its quaternary structure is finely modulated by reversible and pH dependent zinc removal from its Cys_3His_1 motif [11]. The crystallographic structure revealed that RSV M_{2-1} forms a disk-like assembly with tetramerization driven primarily by a long helix forming a four-helix bundle at its center [12] (Figure 1). The tetramerization helix is linked to the core domain responsible for RNA binding activity [13] by a flexible region, where two functionally phosphorylatable serine residues that affect RNA binding and antitermination activity are located [12]. Residues involved in RNA binding have been previously identified by structure guided mutagenesis and RNA binding assays [12]. Positively charged residues within the core surface (K92, K150, R151 and K159) and N-terminal arginine residues (R3 and R4) were shown to participate in RNA binding (Summarized in Figure 1A). Proteins containing CCCH zinc finger motifs have been extensively described to be involved in RNA binding activity. For instance, the eukaryotic poly-A RNA binding protein Nab2 binds specifically to poly-A RNA sequences and participates in polyadenylation regulation and nuclear export of mature mRNPs [14].

Antitermination mechanisms involve the interaction of regulatory proteins with RNA structures, which leads to destabilization of terminator signals, or their bypass by the RNA polymerase. While these mechanisms have been intensely studied in prokaryotic systems [15; 16], far less is known about antitermination in eukaryotic and viral genes [17]. It has been shown that the HRSV $M_{2\text{-}1}$ protein is able to bind to RSV leader RNA sequences [18] and to different gene-start and gene-end sequences with variable affinities and preference for A-rich sequences [12] as well as to long RNAs but with otherwise no clear indication of sequence specificity [18]. It was proposed that $M_{2\text{-}1}$ acts during transcription, recognizing and binding to mRNA-sense gene-end sequences, and preventing premature termination [12; 13]. However, no clear correlation was found between binding affinity to sense gene-end sequences and known sensitivities of each gene to $M_{2\text{-}1}$ antitermination activity, and no unique nucleotide sequence specificity has been identified.

Despite the well described biological activity [5; 19] and structures of $M_{2\text{-}1}$ [12; 20], the biochemical mechanism accounting for the antitermination or processivity enhancement activity of $M_{2\text{-}1}$ is still unknown. In this work, we set out to dissect the initial biochemical events underlying RNA recognition by $M_{2\text{-}1}$ by investigating equilibrium and kinetic determinants that define the reaction. We show an unexpected stoichiometry whereby two molecules of 20mer RNA bind to the tetramer, through a positively cooperative process. In addition, we show that the $M_{2\text{-}1}$ tetramer but not the monomeric RNA-binding $M_{2\text{-}1}$ core domain is able to unfold persistent RNA structure upon binding, which may otherwise slow down or interrupt transcription. We explain the cooperative recognition in kinetic sequential terms in which a fast association reaction is followed by a conformational rearrangement and binding of the second RNA molecule. Our results contribute to the understanding of how high affinity can be attained in the absence of sequence specificity by linkage of RNA stretches binding to spatially contiguous RNA binding sites in a tetramer. We discuss our findings in the light of structural data, in particular that of the

homologous HMPV counterpart and the functionally related VP30 protein from the Ebola virus.

RESULTS

Unexpected stoichiometry for the M_{2-1} antiterminator-RNA complex

An essential step for defining a binding mechanism is the determination of the binding stoichiometry. We started with analyzing the physical interaction of the M_{2-1} tetramer with 6-FAM labeled RNAs by EMSA. Because fluorescently labeled RNAs were used, quenching might limit the detection of the M_{2-1}:RNA complex, and therefore quantification was performed on the free RNA probe (see materials and methods). Since previous work has shown that M_{2-1} is able to bind several RNAs of different sequence compositions [12], we selected a model sequence consisting of repetitions of the pentamer AGUUG referred to as 10mer, 15mer and 20mer RNAs. When titrating the M_{2-1} tetramer into 10mer RNA (AGUUGAGUUG), a large retardation of the bound species was observed with a clear 1:4 M_{2-1} tetramer:RNA stoichiometry, equivalent to one RNA per M_{2-1} monomer (Figure 2A). However, two molecules of the 20mer RNA (AGUUGAGUUGAGUUGAGUUG) bound per M_{2-1} tetramer, equivalent to a 1:2 (M_{2-1} tetramer:RNA) stoichiometry (Figure 2B). The core domain of M_{2-1} consists of a 13.6 KDa monomeric fragment that contains the major RNA binding site (Figure 1A) [13]. The core domain bound to both 10mer and 20mer RNAs with clear 1:1 stoichiometry (Figures 2C and 2D), strongly suggesting that the unexpected stoichiometry of the 20mer binding to the tetramer arises from the RNA contacting two RNA binding sites from adjacent monomers.

Shorter RNAs required higher 6-FAM-RNA probe concentrations (2 to 4 μM for 10mer RNA compared to 200 nM for 20mer RNA, Figure 2) in order to reach saturation in the EMSA experiments, indicating that formation of the 1:2 complex of the M_{2-1} tetramer to 20mer RNA led to tighter binding between the RNA and M_{2-1}. Similarly, 1:1 binding of either 10mers or 20mers to the monomeric M_{2-1} core domain required much higher concentrations, also indicating significantly lower affinities. In order to establish the minimal length that allowed a 1:2 binding mode, we tested binding of the 15mer RNA (AGUUGAGUUGAGUUG). Similar to the 20mer, we

observed a 1:2 association of 15mer RNA to the $M_{2\text{-}1}$ tetramer, and the low RNA concentration (200 nM) required for saturation experiments indicated high affinity (Figure S1A).

Quantitative analysis of the $M_{2\text{-}1}$ interaction with RNA

As a complementary approach, we performed solution studies of $M_{2\text{-}1}$-RNA complex formation by equilibrium titrations using different fixed concentration of 6-FAM-labeled RNAs, and monitoring fluorescence changes at increasing concentrations of $M_{2\text{-}1}$ tetramer. Binding of the $M_{2\text{-}1}$ tetramer to 20mer RNA produced a significant decrease in the 6-FAM fluorescence due to quenching of the fluorophore (Figure 3A) and showed saturation at a 1:2 ($M_{2\text{-}1}$tetramer:RNA) stoichiometry for 20 nM or higher 20mer RNA concentrations, confirming the stoichiometry determined by EMSA experiments (Figure 3A vs Figure 2B). Similar results also confirmed a 1:2 $M_{2\text{-}1}$ tetramerRNA binding stoichiometry for the 15mer RNA (Figure S2). Titrations at sub-stoichiometric concentrations of RNA revealed shifts in the binding transitions as a function of RNA length, with transitions occurring at lower $M_{2\text{-}1}$ concentrations for longer RNAs, revealing a gradient of relative affinities for association to the $M_{2\text{-}1}$ tetramer, with 20mer RNA showing the highest affinity (Figure 3B).

We next investigated the complexes where one RNA molecule was bound per $M_{2\text{-}1}$ monomer (equivalent to a 1:1 stoichiometry of the $M_{2\text{-}1}$ Core Domain and a 1:4 stoichiometry in the case of the $M_{2\text{-}1}$ tetramer binding to 10mer RNA) by combining fluorescence and EMSA titrations at different RNA concentrations in order to estimate the equilibrium dissociation constant for each interaction (Table 1, Figures S3 and S4 and Table S1), where averaging of several binding curves allowed the determination dissociation constants. Opposed to what was observed for binding of the $M_{2\text{-}1}$ tetramer to 10mer, 15mer and 20mer RNAs (Figure 3 A-C), binding of RNAs to the $M_{2\text{-}1}$ Core Domain produced small changes in fluorescence intensity, and the best signal to noise ratio for the 20mer RNA interaction with the $M_{2\text{-}1}$ core domain was

obtained by monitoring fluorescence anisotropy (Figure S3A). Where both EMSA and fluorescence techniques could be applied ($M_{2\text{-}1}$ Core Domain binding to 20mer RNA and $M_{2\text{-}1}$ tetramer binding to 10mer RNA), we observed that the K_D values recovered from combining both electrophoretic and spectroscopic techniques were in good agreement (Figure S4 and Table S1). Remarkably, the association of 10mer and 20mer RNAs with the $M_{2\text{-}1}$ core domain monomer had comparable average dissociation constant values (10mer K_D = 236 ± 232 nM and 20mer K_D = 239 ± 110 nM, Table 1). This result led us to conclude that the 10mer RNA saturates binding to the primary binding site located in the $M_{2\text{-}1}$ core domain, with no significant additional interactions being contributed by longer length RNAs. Moreover, these K_D values were in turn in very good agreement with the K_D value of the $M_{2\text{-}1}$ tetramer binding to 10mer RNA (K_D = 286 ± 170 nM, Table 1), indicating that RNA binds in a similar mode to the minimal and independent site, irrespective of whether the RNA binding site is within the context of the $M_{2\text{-}1}$ tetramer or that of the $M_{2\text{-}1}$ Core Domain monomer.

To assess the mechanism of $M_{2\text{-}1}$ interaction with RSV gene regulatory sequences, we investigated association of the $M_{2\text{-}1}$ tetramer to a cognate RNA sequence corresponding to the SH gene-end region (RSV SH-GE) (positive polarity). In order to assess the effects of length on this interaction, we used a previously tested 13mer RNA sequence (13mer RSV-RNA) [12], as well as a longer 20mer RNA (20mer RSV-RNA) spanning additional adjacent sequences from the same gene regulatory region. Both EMSA assays (Figure S1B) and fluorescence titration experiments (Figure 3C) showed that both 13mer RSV-RNA and 20mer RSV-RNA also revealed a 1:2 ($M_{2\text{-}1}$tetramer:RNA) binding stoichiometry. Non-cognate RNA sequences (Figure 3A and 3B) and RSV SH-GE sequences (Figure 3C and S1B), showed transitions occurring at lower $M_{2\text{-}1}$ concentrations for 20mer RNAs, suggesting higher affinity for longer RNAs. In both cases, two RNA molecules

bound per tetramer, where the minimal length, at least for the RSV GE sequence, was now narrowed to 13 nucleotides, but the optimal length was 20 nucleotides.

$M_{2\text{-}1}$ unfolds RNA structure upon binding

We assessed whether $M_{2\text{-}1}$:RNA complex formation involved conformational changes in either or both binding partners using circular dichroism. First, we looked at the Far-UV CD region, where most of the contribution should correspond to the protein backbone, by comparing the 20mer RNA:$M_{2\text{-}1}$ complex to the arithmetic sum of each individual component (RNA and $M_{2\text{-}1}$). Both CD spectra were superimposable, indicating that the overall secondary structure content of $M_{2\text{-}1}$ showed no major changes upon complex formation (Figure S5A). Next, we assessed the Near-UV CD region, where we expected the major contribution to be that of the RNA. Here, we observed that 20mer RNA had a strong positive signal with a maximum centered at around 270 nm (Figure 4A). The $M_{2\text{-}1}$ tetramer lacked signal in this wavelength range, therefore, the signal corresponded to the RNA molecule alone (Figure 4A). This signal arises likely due either to base stacking, base pairing, or both, and was not observed for the shorter (10mer and 15mer) RNA fragments (Figure 4A). This structure was destabilized by temperature and unfolded with little cooperativity (Figure 5B and S5C), indicating that it did not correspond to a stable hairpin-like structure and instead indicated the presence of a significant degree of a folded yet undefined structure. We considered the fact that this structure could be affected by binding to the $M_{2\text{-}1}$ antiterminator. Indeed, addition of the $M_{2\text{-}1}$ tetramer to 20mer RNA showed a gradual decrease in ellipticity at 270 nm down to baseline values, indicative of the RNA structure being unfolded upon complex formation (Figure 4B and S5B). Moreover the stoichiometry observed was 1:2, confirming, by a third and independent method, that two molecules of 20mer RNA bind per $M_{2\text{-}1}$ tetramer (Figure 4B, inset).

Contrary to what we observed for the $M_{2\text{-}1}$ tetramer, the $M_{2\text{-}1}$ core domain monomer was unable to destabilize the RNA structure of 20mer RNA (Figure S5D). The experiment was carried

out at 5 μM, and since we determined the dissociation constant to be 239 ± 110 nM (Table 1 and Figure S4), the protein concentration was in fact 20-fold above the K_D, ensuring complete binding saturation. This led us to conclude that the $M_{2\text{-}1}$ core domain monomer binds to the 20mer RNA, but is unable to unfold it, strongly suggesting a different binding mode.

A similar behavior was observed for the cognate 13 and 20mer RNA-RSV sequences, in fact the molar ellipticity of 20mer RSV-RNA was much more pronounced than non-cognate 20mer RNA suggesting a higher degree of structure or stability (Figure 5A). Furthermore, thermal denaturation scans showed that 20mer RSV-RNA unfolding was cooperative and more stable than the 20mer RNA by a difference of 23 °C in apparent Tm (Figure 5B). Equilibrium titrations with the $M_{2\text{-}1}$ tetramer produced a decrease in the 270 nm band for both RSV 13mer and 20mer RNAs (Figure 5C), confirming the 1:2 $M_{2\text{-}1}$ tetramer:RNA stoichiometry observed for the non-cognate 20mer RNA (Figure 4B, inset). RNA unfolding by $M_{2\text{-}1}$ was fast and occurred within the 10 seconds dead time of the experiment (not shown).

Binding of the anti-terminator to RNAs longer than 13 bases is cooperative

In order to investigate the interaction of the longer 20mer RNAs to $M_{2\text{-}1}$, we performed titrations at sub-stoichiometric concentrations, where it is possible to estimate binding affinities. As can be observed, binding is sub-stoichiometric under 40nM 6-FAM-RNA concentration for non-cognate 20mer RNA and 20mer RSV-RNAs (Figure 3A, C). A close inspection of the raw and normalized binding data for the interactions of $M_{2\text{-}1}$ with non-cognate 20mer RNA and with 20mer RSV RNAs, revealed a clear sigmoidal shape of the binding curves at sub-stoichiometric concentrations with a steep increase in the fraction of bound ligand upon $M_{2\text{-}1}$ addition, indicative of cooperativity (Figure 6A and Figure S6). The distinctive sigmoidal behavior could be observed under equilibrium binding conditions at RNA concentrations ranging from 10 to 40 nM for both non-cognate and cognate RNA ligands (Figure S6). One approach for assessing cooperativity is to

quantify the log units of ligand required to produce a transition from bound fractions of 0.1 to 0.9 [21]. While non-cooperative interactions usually span 2 log units of ligand concentration, negative cooperativity is characterized by shallower transitions that span more than 2 log units, and positive cooperativity by sharper transitions that span less than 2 log units. All binding isotherms for $M_{2\text{-}1}$ tetramer binding to 20mer non-cognate and 20mer RSV RNAs, showed transitions that spanned less than two log units of protein concentration, supporting a positively cooperative behavior (Figure S7). In accordance with this, fitting of the data to the Hill equation at all protein concentrations tested (Figure 6B, Figure S8 and Table S2) allowed us to determine the Hill coefficient (n_H) and apparent binding constant (K_{Dapp}) for all binding curves, which yielded similar values at the different concentrations tested (See materials and methods and Table S2). For both non-cognate and RSV 20mer RNAs, we obtained reproducible fits at different RNA concentrations, with average values of n_H = 2.3 ± 0.5 for non-cognate RNA and n_H = 2.5 ± 0.7 for RSV-RNA, confirming positive cooperativity (Table 1 and Table S2). The average apparent binding constants recovered were K_{Dapp} = 14 ± 6 nM for non-cognate RNA and K_{Dapp} = 16 ± 5 nM for RSV-RNA, revealing a 15-fold increase over the K_D values characteristic of the single-site interactions observed with the core domain monomer and the non-cooperative interactions of the M2-1 tetramer with 10mer RNA (Table 1). The stoichiometry together with these results, suggests that binding to two adjacent sites in the $M_{2\text{-}1}$ core domains within the $M_{2\text{-}1}$ tetramer leads to the observed cooperativity and increase in affinity. As our experiments were performed with a core domain $M_{2\text{-}1}$ monomer that lacks the N-terminal region, we cannot rule out that the observed increase in binding affinity is also partly due to additional interactions of the RNA with the N-terminal region previously shown to contribute to RNA association (R3 and R4) [12], but this energetic contribution is likely to be minor in comparison to that of the main Core Domain RNA binding site.

Kinetic basis for the cooperativity of the $M_{2\text{-}1}$:RNA interaction

To elucidate the different steps involved in cooperative $M_{2\text{-}1}$:RNA association, we performed non-steady state analyses of the binding and dissociation reactions. The formation of the $M_{2\text{-}1}$:RNA complex produced a decrease in the fluorescence intensity and an increase in fluorescence anisotropy signals (Figure 7A and 7B). At 100 nM $M_{2\text{-}1}$ concentration, a major change in the anisotropy signal occurred within the dead time of mixing (5 seconds), followed by a slow decay that was best fit to a mono-exponential process with k_{obs} = 0.020 ± 0.002 s^{-1} (Figure 7A). A similar behavior was observed for all $M_{2\text{-}1}$ concentrations tested indicating the presence of a burst phase (Figure S9 A, C), which corresponds to the fast bimolecular association reaction. The slow decay phase was always mono-exponential (Figure S10), and showed no dependence on $M_{2\text{-}1}$ concentration (Figure 7D).

The fluorescence intensity signal decreased upon binding, with no detectable burst phase (Figure 7B and Figure S9 B, D). For all $M_{2\text{-}1}$ concentrations tested, the process was best described by two observed phases (Figure S11). The amplitude of the faster phase (k_{obs2}) became larger for increasing $M_{2\text{-}1}$ concentrations (Figure 7C). Instead, the amplitude of slower phase (k_{obs1}) became smaller for increasing $M_{2\text{-}1}$ concentrations (Figure 7C) and the rate constant showed no dependence on $M_{2\text{-}1}$ concentration (Figure 7D), with a similar value to the decay phase observed in the anisotropy signal.

Taken together, we observed that at least two slower processes were associated to $M_{2\text{-}1}$:RNA complex formation. For the non-cooperative $M_{2\text{-}1}$ core domain:20mer RNA and $M_{2\text{-}1}$ tetramer:10mer RNA complexes, no observable slow phases accompanied binding (not shown). This suggested that the slower phases correspond to conformational rearrangements that follow binding and constitute a signature of the observed cooperativity. We applied a quantitative model to assess the conformational rearrangement. It has been shown that under fast equilibration

conditions for the bimolecular association, an increase in the value of the slow rate constant as a function of ligand concentration is indicative of an induced-fit reaction [22]. While the slower phase showed no dependence on $M_{2\text{-}1}$ protein concentration (Figure 7D, clear circles), the observed rate constant for the faster phase clearly increased as $M_{2\text{-}1}$ concentration was increased, supporting the existence of an induced-fit process. Fitting of the data to this induced fit model [22] (See material and methods, Eq [4]) yielded the rate constants for the forward reaction k_1 = 0.16 ± 0.1 s^{-1} and an equilibrium dissociation constant for the association reaction of K_1 = 89 ± 6 nM (Figure 7D). Since no global changes in secondary structure were observed in $M_{2\text{-}1}$ upon complex formation (Figure S5A), this conformational rearrangement corresponds to a subtle change that is not detected in the CD experiments, as often seen for cooperative or allosteric processes.

The dissociation reaction was probed using displacement experiments, where a stoichiometric $M_{2\text{-}1}$:6-FAM-RNA complex, was mixed with a molar excess of unlabeled RNA. We tested the $M_{2\text{-}1}$:20mer RNA interaction, and observed that dissociation occurred within the dead time of mixing (8 seconds) as reported by both the anisotropy and fluorescence intensity signals and that the binding reaction was fully reversible (Figure S12). Displacement of 20mer RNA from the complex with 15mer RNA required higher molar excess (> 20-fold, not shown). This strongly suggested that the 20mer RNA species establishes additional contacts in the $M_{2\text{-}1}$ tetramer that cannot be displaced by the 15mer RNA molecules, in line with the higher affinity observed for the 20mer species in equilibrium experiments.

DISCUSSION

The single and best described antitermination mechanism in eukaryotic viruses is that of the regulation of HIV transcription [17]. In this case, promoter-proximal pausing in the HIV leader region stalls transcription by RNA polymerase II, allowing the formation of the TAR hairpin in nascent mRNA and a repressor complex with the NELF RNA binding protein. The viral Tat antiterminator factor recruits elongation factors, allowing disruption of repressor complexes and reinitiating transcription of viral genes [23]. A common theme conserved across prokaryotic, eukaryotic and viral genes [24] is the interaction of antiterminator proteins with RNA sequence or structural elements. Antiterminator proteins may bind RNA and disfavor the formation of RNA structures, through a binding that can be sequence specific, as in the case of the hut operon regulator HutP [25] or nonspecific, as in the global response to temperature of cold shock proteins [26].

The mechanism by which the transcription machinery recognizes RSV gene-end (GE) and gene-start (GS) signals is still unclear. The RSV GEs are 12-13 nucleotides long and direct polyadenylation and transcription termination. However, it is unknown whether binding of factors to these regulatory sequences slows down or pauses elongation [27; 28] in order to trigger polyadenylation and release of the nascent mRNA. GEs contain three regions: a conserved penta-nucleotide (3′-UCAAU-5′) followed by a non conserved 3-4 nucleotide AU-rich middle sequence and a 4 to 5 nucleotide poly-U tract. $M_{2\text{-}1}$ was proposed to bind the GE in a sequence specific manner, allowing the transcriptase complex to bypass this functional signal and generate read-through RNA transcripts [12]. $M_{2\text{-}1}$ was shown to increase the levels of read-through transcription at all gene junctions to different extents, but the lack of correlation between the strength of $M_{2\text{-}1}$ binding to different GE regions and the efficiency of antitermination [4; 6], favors the hypothesis that $M_{2\text{-}1}$ potentially binds at several sites in the RNA genome or antigenome beyond the gene end sequences.

Different alternatives for RNA binding sites were proposed in previous work, including the leader region [18], viral mRNAs and A-rich RNA sequences, but the absence of sequence specificity was also discussed for RSV $M_{2\text{-}1}$ [12] and the homologous EBOV VP30 protein [29]. A recent study tested 13mer RNA GEs of a selection of genes, assuming these to be potential specific RNA binding sites for RSV $M_{2\text{-}1}$ [12]. Despite the high sequence conservation inherent to these GEs, the affinities differed 5- to 15-fold for positive and negative sense GEs compared to the SH gene GE taken as reference (46 nM). Moreover, the highest affinity described was for a poly-A tract, which led the authors to propose that $M_{2\text{-}1}$ binds the viral mRNA cotranscriptionally and preferentially at A-rich 3‘ sequences, consistent with the U-rich tracts found in the viral genome GEs. Our results with both cognate and non-cognate 20 mer ligands (K_D=14-16 nM) yield affinities identical to those of the poly-A substrate (K_D=19 nM) [12], which is remarkable considering the experiments were carried out in different labs. These results support the view that $M_{2\text{-}1}$ is able to bind gene end sequences but also other A-rich regions either in the genome or subgenomic mRNAs with comparable affinity, leaving open the question regarding the degree of specificity of $M_{2\text{-}1}$/RNA interactions. In this scenario, $M_{2\text{-}1}$ might associate to multiple sites besides the gene ends, disrupting local RNA structures and facilitating elongation. Among mononegavirales, the EBOV VP30 was shown to display functional [30] and structural [31] similarities to $M_{2\text{-}1}$ proteins, in that it binds RNA, acting as a transcription antitermination or processivity factor. Interestingly, a recent investigation of binding of VP30 to a plethora of RNA substrates also concluded that the binding specificity was rather relaxed, similar to what has been observed for $M_{2\text{-}1}$ [29]. In any case, exclusive RNA binding sites for HRSV, HMPV $M_{2\text{-}1}$ or VP30 are still to be confirmed.

Based on crystallographic studies of the HMPV $M_{2\text{-}1}$ tetramer bound to DNA and AMP [20], Leyrat et al. proposed that the RNA molecule can simultaneously interact with the N-terminal zinc finger region and the C-terminal core domain. The existence of two RNA binding sites was also

supported by the higher affinity of 13mer compared to 8mer poly-A RNA (K_D of 20 nM versus 3.4 μM) [12]. Mutational analyses suggested that the R3 and R4 N-terminal residues participate in RNA binding. However, the R3A/R4A mutant revealed a modest 2-fold change in its K_D relative to WT-RSV-$M_{2\text{-}1}$ [12]. This suggests that the 15-fold increase we observed for 20mer RNAs binding to the $M_{2\text{-}1}$ tetramer originates mainly from tetramerization that brings two core domains in close proximity, allowing longer RNAs to bridge two adjacent RNA binding sites, and that the contribution to the overall binding affinity from additional interactions with R3/R4 is minor.

Another related example is the Nab2 protein, which contains tandem CCCH zinc motifs that recognize poly-A RNA sequences and the length of the RNA molecule also influences RNA binding affinity. Solution structure of Nab2 in complex with RNA showed that three CCCH zinc motifs formed a single ordered structural unit that contacts eight consecutive adenines increasing RNA binding affinity from 16 μM for a 6mer poly-A RNA to 100 nM for a 15 mer poly-A RNA [14].

We have shown that the quaternary arrangement of $M_{2\text{-}1}$ leading to contiguous RNA binding sites is a determinant for tight RNA binding, and our results are summarized in Figure 8. While the $M_{2\text{-}1}$ monomer always binds RNA with a 1:1 stoichiometry, the RNA binding mode of the $M_{2\text{-}1}$ tetramer depends on RNA length. While short RNAs have a 1:4 binding mode (4 RNAs per $M_{2\text{-}1}$ tetramer) and show no cooperativity (Figure 8A), longer RNAs establish a 1:2 ($M_{2\text{-}1}$ tetramer:RNA) binding stoichiometry, where the most likely arrangement involves each RNA molecule bridging two binding equivalent binding sites of adjacent monomers (Figure 8B). Positive cooperativity is observed exclusively in the 1:2 ($M_{2\text{-}1}$ tetramer:RNA) interaction, which enhances the binding affinity over 15-fold, leading to formation of a tight low-nanomolar affinity complex. The RNA sequences used in this work were defined based on previous sequence conservation analysis at gene ends. However, the exact nature of specific $M_{2\text{-}1}$ sites has not been defined, either in sequence or in extension. In a scenario where $M_{2\text{-}1}$ is closely associated with the

polymerase, the two RNA strands bound by the $M_{2\text{-}1}$ tetramer in our model (Figure 8B) may represent the template and transcript RNA. Alternatively, $M_{2\text{-}1}$ could act as an antiterminator by preventing the formation of RNA secondary structures by threading two different regions of a single longer RNA strand. We believe this is an important and imperative issue to address in the future, in particular with longer sequences covering up to full-length transcripts.

Importantly, the bimolecular association of 20mer non-cognate and cognate RNAs with the $M_{2\text{-}1}$ tetramer is accompanied by a fast ($t_{1/2}$ < 4s) and significant conformational rearrangement of the RNA molecule (Figure 8B). Following bimolecular association, subtle rearrangements of the $M_{2\text{-}1}$:RNA complex occur through an induced-fit mechanism ($t_{1/2}$ 100s), to yield the final stable complex.

Several signatures linked to the cooperative behavior can be highlighted. Firstly, $M_{2\text{-}1}$ has the ability to unfold RNA structure, most likely due to disruption of base-stacking interactions, as previously described for ribonucleoprotein interactions within spliceosomal components [32]. The structures formed by these 20mer RNAs are not the highly stable, hairpin type of structure often observed in viral transcription regulatory sites based on intrachain precise base pairing, but might be the result of a locally stable collapsed structure that results from hydrophobic base stacking. The equilibrium between structured and unfolded RNA is very fast, concomitant with the initial collision with the protein that likely selects the unfolded species.

The subtle but functionally determinant conformational rearrangements in the complex take place in the protein side and are consistent with an induced-fit mechanism, supported by a quantitative kinetic test [22]. While in the case of $M_{2\text{-}1}$ the evidence indicates that no major changes in secondary structure occur upon binding, loop or hinge motions may reposition the RNA binding domains, as shown for the metapneumovirus $M_{2\text{-}1}$ protein, where transitions from an “open” to a “closed” conformation are observed upon RNA association [20]. Our results highlight the importance

of multi-site binding and conformational transitions in allosteric RNA recognition, such as those taking place in multidomain proteins of different arrangements [33; 34]. In $M_{2\text{-}1}$, the domains bearing the RNA binding sites are tightly connected by the homotetrameric nature, where an otherwise weak binding of a first stretch of the RNA substrate increases the effective concentration resulting in a large increase in affinity, due to an entropic linkage. This also appears to be the case of the functionally related Ebola VP30 factor, which involves an hexameric arrangement and shows an optimum length of 40 bases for binding [29]. VP30 binds to a 3' terminal region of the EBOV genome with an apparent K_D of 0.35 μM [35]. Interestingly, mutations that destroy the capability to form homohexamers or disrupt the zinc finger domain, impair RNA binding and transcriptional activation [30]. A large number of sequences and RNA lengths were tested but the affinity changes were moderate (0.3 to 1.0 μM, corresponding to differences in binding energies of less than 0.7 kcal mol^{-1}) [29]. This further supports that such antitermination activity found so far only in pneumoviruses and filoviruses may not require single high affinity specific binding sites.

Crystallographic, molecular simulations and solution studies revealed that the closely related HMPV $M_{2\text{-}1}$ protein displays a similar folding and tetramerization topology, but an asymmetric configuration with a rigid tetramerization domain and dynamic core domains [12]. Interestingly, SAXS studies revealed a highly dynamic ensemble in solution, in which each core domain in the tetramer exists in equilibrium between open and closed conformations, and treatment with EDTA increases the proportion of open conformers. Moreover, RNA binding was proposed to favor the closed state [12]. It is likely that the highly homologous RSV $M_{2\text{-}1}$ protein will display a similar behavior, but since crystal structure snapshots cannot reveal such dynamic processes occurring in solution, this issue requires further investigation.

Our study provides a first and significant step for understanding antitermination in pneumoviridae and also represents one of the few mechanistic studies of viral antiterminator-

RNA association. Cooperativity is often found in protein-RNA interactions, and in this work we show how association of adjacent protomer binding sites allows high-affinity binding in the absence of sequence specificity. Moreover, we dissect the time events behind cooperativity and show an induced fit event on the protein side and a conformational selection of the RNA ligand. Further studies will hopefully clarify the presence of specific binding sites in the genome or subgenomic mRNAs and the extent of binding to other regions, which should be supported by structural studies that provide a molecular-level detail for specificity through base specific recognition.

The open question remaining is how $M_{2\text{-}1}$ operates on antitermination. Is it through the stabilization of the polymerase-nucleoprotein complex, or through breaking RNA structures that slow down or disrupt the complex? Future directions should aim at structural analysis of the relevant complexes, complemented by solution studies and evaluation of binding of $M_{2\text{-}1}$ in the context of all the components of the RNA polymerase complex, including longer RNA molecules. Hopefully this will also extend to other antiterminator proteins such as Ebola VP30.

MATERIALS AND METHODS

Protein expression and purification. *Full length M_{2-1} protein.* Expression and purification was performed as previously described [9]. Briefly the full length human RSV strain A M_{2-1} protein was expressed in C41(DE3) cells transformed with the pRSETA M_{2-1} plasmid and grown in terrific broth medium. Induction was performed with 0.3 mM Isopropil-β-D-1-tiogalactopiranoside (IPTG) at

DO_{600nm} of 0.6. The soluble protein was precipitated with 40% ammonium sulfate and was subjected to an Heparin Ceramic HyperD affinity column and to a size exclusion chromatography (SEC) on a Superdex 200 gel filtration column (GE Healthcare). *M_{2-1} Core Domain:* Full length M_{2-1} was subjected to limited proteolysis with chymotrypsin enzyme from bovine pancreas (Sigma-Aldrich) in 50 mM Tris.HCl (pH 8) and 2 mM $CaCl_2$ for 90 minutes at 28°C at a ratio 1:60 (protein:chymotrypsin w/w); the reaction was stopped by adding 1 mM phenylmethylsulfonyl-fluoride (PMSF) to obtain a digest product identified by MALDI-TOF MS (Bruker, Daltonics, Billerica, MA) whose sequence corresponds to a fragment of M_{2-1} from Ala73 to Tyr194 (122 amino acids). Molecular filtration of the digestion product was performed in a SEC column (Superdex 75) in 20 mM Tris.HCl (pH 7.0) and 0.2 M NaCl. The M_{2-1} Core domain purity was assessed by SDS-PAGE stained with Coomassie Blue and MALDI-TOF MS. Mass Spectrometry analysis and RP-HPLC revealed a homogeneous sample corresponding to fragment 72-194 of full length M_{2-1}. The sample was concentrated using Amicon centrifugal filter units (Milipore) up to an approximate concentration of 100 μM. Circular dichroism and SEC confirmed secondary structure and monomeric state of M_{2-1} Core domain respectively. The SEC filtration was performed on an analytical Superdex S75 10/300 GL (24 mL) that was calibrated with different standard globular proteins and the void volume (V_0) and total volume (V_t) were determined by loading Blue Dextran and acetone, respectively. Far-UV CD measurements and SEC filtration were performed in 20 mM

Tris.HCl (pH 7), 0.2 M NaCl. The protein concentration was determined spectrophotometrically using molar extinction coefficients (ε_{280}) of 13200 M^{-1} cm^{-1} (full length $M_{2\text{-}1}$) and 4470 M^{-1} cm^{-1} ($M_{2\text{-}1}$ Core domain). The protein concentration is expressed in monomer units for $M_{2\text{-}1}$ Core Domain and moles of tetramer for the $M_{2\text{-}1}$ tetramer.

Labeled and unlabeled RNA oligonucleotides. Unlabeled single chain RNA oligonucleotides were chemically synthesized and purified by Integrated DNA Technologies (IDT). RNAs labeled with 6-Carboxyfluorescein (6-FAM) at their 5' end were subjected to additional desalting and RP-HPLC purification. Oligonucleotides were quantified using corresponding molar extinction coefficients (ε_{260}) by UV spectrophotometry to provide an accurate measure of concentration. The RNA sequences are the following (unlabeled RNAs lacked the 5' 6-FAM moiety):

10-mer: 5'- /6-FAM/AGUUGAGUUG-3'

15-mer: 5'- /6-FAM/AGUUGAGUUGAGUUG-3'

20-mer: 5'- /6-FAM/AGUUGAGUUGAGUUGAGUUG-3'

RSV 13-mer: 5'- /6-FAM/AGUUAAUUAAAAA-3'

RSV 20-mer: 5'- /6-FAM/AUAAAGUAGUUAAUUAAAAA-3'

Electrophoretic mobility shift assay (EMSA). Native gels were prepared with 4% polyacrylamide (29:1, acrylamide/bisacrylamide), 5% glycerol, 0.5X Tris-Glycine buffer (pH 9.0). The gel was run before loading samples for 30 minutes at 60 volts. A binding reaction was prepared in 0.02 M hepes (pH 7.4), 0.08 M NaCl, 5% glycerol, 6-FAM-labeled RNA at the final desired concentration. The samples were incubated 20 min at room temperature to allow for complex formation. The electrophoretic run was performed at 40 volts for 20 minutes followed by 60 volts for 1 hour. All experiments were performed in 1X Tris-Glycine running buffer at 20°C. RNA fluorescence was

detected fluorometrically in a Storm Imager 840 (Molecular Dynamics). Experiments were performed at total RNA concentrations ranging from 200 nM to 4.0 μM. In cases where RNA concentration was higher than 200 nM, a 200 nM 6-FAM-RNA concentration was used and unlabeled RNA was added up to the final RNA concentration. 200 nM 6-FAM-RNA was the lowest probe concentration that yielded a detectable signal in the Molecular Dynamic Storm 840 Imager.

Densitometric analysis of EMSA assays. The digital image of the EMSA gel was used to quantify the fraction of free 6-FAM-RNA, using the ImageJ 1.48V tool. As opposed to EMSA experiments using radioactively labeled oligonucleotides, the use of 6-FAM-RNAs prevented an accurate estimation of the fraction of bound versus free RNA species by densitometry, due to strong quenching of the RNA fluorescence upon binding. Therefore, we estimated the percentage of unbound RNA species alone. The limits of the first band of free RNA (First tube containing no $M_{2\text{-}1}$ tetramer or monomer) were determined and used as a guide to select the rest of the bands. The program generates a value that corresponds to the intensity of the signal within the selected area. The values obtained at increasing $M_{2\text{-}1}$ concentrations were normalized to the intensity value of the first band (corresponding to 100 % free RNA) and expressed as the percentage of unbound RNA (% Unbound RNA). The % Unbound RNA was then plotted as a function of the $M_{2\text{-}1}$:RNA molar ratio, to determine the binding stoichiometry (Figure 2, lower panels) or as a function of $M_{2\text{-}1}$ concentration to estimate the dissociation constant (K_D).

Binding stoichiometry. The binding stoichiometry was determined from titrations performed in stoichiometric conditions, by extrapolation of two linear fits of the initial and final values evaluated.

Circular Dichroism measurements (CD). Far- and Near-UV CD measurements were performed on a Jasco J-810 spectropolarimeter using 5 μM of samples (RNA, protein or protein:RNA complexes) using a cell path of 0.1 or 0.2 cm. The temperature was set using a Peltier temperature controller. Spectra were recorded between 200-260 nm (Far-UV CD) and 260-350 (Near-UV CD) at standard sensitivity, at rate of 100 nm/min, response time of 2 s, a data pitch of 0.2 nm and a bandwidth of 4 nm. All spectra were an average of at least six scans. All experiments were performed in 50 mM Tris-HCl (pH 7.5), 0.1 M NaCl and 1 mM DTT at 5, 20 or 80ºC. Kinetic experiments were performed by following the ellipticity signal at 270 nm after adding 10 μM of $M_{2\text{-}1}$ protein manually (dead time 5 seconds) in a cuvette containing 10 μM 20mer RNA until saturation was reached.

Fluorescence spectroscopy measurements. Fluorescence emission spectra and fluorescence anisotropy were performed on a Jasco FP-6500 spectrofluorometer using an excitation wavelength of 495 nm. The emission spectrum was recorded from 505 to 600 nm. For fluorescence anisotropy measurements, parallel and perpendicular emission components were measured in L-format by excitation at 495

nm and emission at 520 nm. Anisotropy was measured five times using individual G factors and a measurement time of 3 s. For both signals the excitation and emission bandwidths were set to 3 or 5 nm depending the concentration of the 6-FAM-RNA used.

Equilibrium titration measurements. Increasing amounts of protein ($M_{2\text{-}1}$ tetramer or $M_{2\text{-}1}$ Core domain) were added to a cuvette containing a fixed amount of 6-FAM-RNA (10, 15 or 20mer) and were incubated for 5 min at 20 °C to ensure steady state measurements. The volume added

reached less than 10% of total volume in each assay, and thus, the concentration of 6-FAM-RNA can be assumed to have remained constant throughout the titration.

Kinetic measurements. Slow association phases were measured by manual mixing using a Jasco FP-6500 spectrofluorometer. The changes in fluorescence anisotropy and fluorescence emission at 520 nm were measured after the addition of increasing concentrations of $M_{2\text{-}1}$ to a cuvette containing a fixed amount of 6-FAM-RNA (indicated in each figure legend). The individual data traces were fit to a single exponential function to obtain the observed rate constant for the reaction.

***Data analysis and fitting*.** Fitting was carried out using the software program Profit (Quantum Soft, Uetikon am See, Switzerland) to obtain parameters and their standard deviations.

Equilibrium binding titrations. For complexes with 1:1 stoichiometry, the dissociation constant (K_D) of the complex $M_{2\text{-}1}$:RNA was calculated by fitting the plot of the fluorescence anisotropy signal (r) change as a function of $M_{2\text{-}1}$ concentration to the following equation corresponding to a 1:1 association model,

$$r = r_{free} + \frac{\Delta r_{\text{int}}}{2}\left(\left(x + [RNA] + K_D\right) - \left(\left(x + [RNA] + K_D\right)^2 - 4[RNA]x\right)^{0.5}\right) \qquad (1)$$

where x represents the variable total concentration of the protein ($M_{2\text{-}1}$ or $M_{2\text{-}1}$ Core), [RNA] is the probe concentration, Δr_{int} is the difference in intrinsic fluorescence anisotropy between the free and complexed fragments, r_{free} is the fluorescence anisotropy of the free fragment and K_D is the dissociation constant of the complex $M_{2\text{-}1}$:RNA.

Association Kinetics. The change in fluorescence anisotropy signal was fit to the following mono-exponential equation to obtain the observed rate constant, for each M_{2-1} concentration,

$$\Delta r = A.e^{k.t} + c.t \qquad (2)$$

where A is the amplitude and k_{obs} the observed rate constant. Signal drift was taken into account by a linear time-dependent term (c).

The change in 6-FAM fluorescence signal were fit to a double-exponential equation, to obtain the two observed rate constants, k_{obs1} and k_{obs2} and their respective amplitudes,

$$\Delta r = A_1.e^{k1.t} + A_2.e^{k2.t} + c.t \qquad (3)$$

where A_1 and A_2 are the amplitudes and k_{obs1} and k_{obs2} are the observed rate constants. Signal drift was taken into account by a linear time-dependent term (c).

Induced-Fit model. For the slow phases in the kinetic association experiments, the dependence of the observed rate constant on M_{2-1} concentration was fit to a previously described induced-fit model [22]:

$$k_{obs} = k_{-1} + k_1 \cdot \left(\frac{[M_{2-1}]}{[M_{2-1}] + K_{D2}} \right) \qquad (4)$$

Where k_{obs} is the observed association rate constant at each $M_{2\text{-}1}$ concentration, k_1 and k_{-1} are the forward and reverse rate constants for the post-binding rearrangement reaction, and K_{D2} is the equilibrium dissociation constant for the bimolecular association reaction.

Cooperative interactions. We first tested cooperativity by visual inspection of titration curves, which showed a sigmoidal behavior. Transformation of the x-axis into log scale allowed assessment of the number of log units required for transitioning from 0.1 to 0.9 fraction of saturation [21]. In all cases, we observed transitions, which occurred in less than two log scales, a characteristic feature of positive cooperativity (Figure S7). In second place, we used the Hill formula to assess cooperativity [36]:

$$\Theta = \frac{[M_{2-1}]^{nH}}{[M_{2-1}]^{nH} + [K_D]^{nH}} \qquad (5)$$

Where θ is the degree of saturation, $[M_{2\text{-}1}]$ is the free ligand concentration, K_D is the apparent dissociation constant, and n_H is the Hill coefficient. From the linear transformation of the Hill equation, it is possible to estimate the Hill coefficient n_H from the slope of the linear plot of log θ as a function of log $[M_{2\text{-}1}]$, and the apparent binding constant from the ordinate.

$$\log \frac{\Theta}{1-\Theta} = nH * \log[M_{2-1}] - nH * \log k_D \qquad (6)$$

Complete $M_{2\text{-}1}$ tetramer structure model. Although the $M_{2\text{-}1}$ quaternary complex structure is known (PDB ID: 4C3B) [12], three of the four chains have missing residues. We built a model of the complete quaternary complex based on the complete chain (pdb chain F) with Chimera [37] in order

to attain a realistic view of the surface and to map the RNA interacting residues in the complex (See Figure S13). $M_{2\text{-}1}$ representations in Figure 1 was also created using Chimera.

ACKNOWLEDGEMENTS

We thank Leopoldo L. Gebhard for technical assessment and helpful discussion.

FUNDING

This work was supported by Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) [PICT grant number 2011-0721] and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) [Grant PIP 2012-2014 GI-288] GdPG. LBC, CMB and SAE are CONICET Career Investigators, and IGM is a CONICET Doctoral Fellow.

REFERENCES

1. (2011). *The Biology of Paramyxoviruses* (Samal, S. K., Ed.), Caister Academic Press, Mariland.
2. Collins, P. L., Chanock, R.M., Murphy,. (2001). Respiratory Syncytial Virus. In *Fields Virology* 4th edit. (Knipe DM, H. P., Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE, ed.). Lippincott Williams & Wilkins, Philadelphia.
3. Afonso, C. L., Amarasinghe, G. K., Banyai, K., Bao, Y., Basler, C. F., Bavari, S., Bejerman, N., Blasdell, K. R., Briand, F. X., Briese, T., Bukreyev, A., Calisher, C. H., Chandran, K., Cheng, J., Clawson, A. N., Collins, P. L., Dietzgen, R. G., Dolnik, O., Domier, L. L., Durrwald, R., Dye, J. M., Easton, A. J., Ebihara, H., Farkas, S. L., Freitas-Astua, J., Formenty, P., Fouchier, R. A., Fu, Y., Ghedin, E., Goodin, M. M., Hewson, R., Horie, M., Hyndman, T. H., Jiang, D., Kitajima, E. W., Kobinger, G. P., Kondo, H., Kurath, G., Lamb, R. A., Lenardon, S., Leroy, E. M., Li, C. X., Lin, X. D., Liu, L., Longdon, B., Marton, S., Maisner, A., Muhlberger, E., Netesov, S. V., Nowotny, N., Patterson, J. L., Payne, S. L., Paweska, J. T., Randall, R. E., Rima, B. K., Rota, P., Rubbenstroth, D., Schwemmle, M., Shi, M., Smither, S. J., Stenglein, M. D., Stone, D. M., Takada, A., Terregino, C., Tesh, R. B., Tian, J. H., Tomonaga, K., Tordo, N., Towner, J. S., Vasilakis, N., Verbeek, M., Volchkov, V. E., Wahl-Jensen, V., Walsh, J. A., Walker, P. J., Wang, D., Wang, L. F., Wetzel, T., Whitfield, A. E., Xie, J. T., Yuen, K. Y., Zhang, Y. Z. & Kuhn, J. H. (2016). Taxonomy of the order Mononegavirales: update 2016. *Arch Virol* **161**, 2351-60.
4. Sutherland, K. A., Collins, P. L. & Peeples, M. E. (2001). Synergistic effects of gene-end signal mutations and the M2-1 protein on transcription termination by respiratory syncytial virus. *Virology* **288**, 295-307.
5. Fearns, R. & Collins, P. L. (1999). Role of the M2-1 transcription antitermination protein of respiratory syncytial virus in sequential transcription. *J Virol* **73**, 5852-64.
6. Hardy, R. W., Harmon, S. B. & Wertz, G. W. (1999). Diverse gene junctions of respiratory syncytial virus modulate the efficiency of transcription termination and respond differently to M2-mediated antitermination. *J Virol* **73**, 170-6.
7. Collins, P. L., Hill, M. G., Cristina, J. & Grosfeld, H. (1996). Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus. *Proc Natl Acad Sci U S A* **93**, 81-5.
8. Noton, S. L. & Fearns, R. (2015). Initiation and regulation of paramyxovirus transcription and replication. *Virology* **479-480**, 545-54.
9. Esperante, S. A., Chemes, L. B., Sanchez, I. E. & de Prat-Gay, G. (2011). The respiratory syncytial virus transcription antiterminator M(2-1) is a highly stable, zinc binding tetramer with strong pH-dependent dissociation and a monomeric unfolding intermediate. *Biochemistry* **50**, 8529-39.
10. Esperante, S. A., Paris, G. & de Prat-Gay, G. (2012). Modular unfolding and dissociation of the human respiratory syncytial virus phosphoprotein p and its interaction with the m(2-1) antiterminator: a singular tetramer-tetramer interface arrangement. *Biochemistry* **51**, 8100-10.
11. Esperante, S. A., Noval, M. G., Altieri, T. A., de Oliveira, G. A., Silva, J. L. & de Prat-Gay, G. (2013). Fine modulation of the respiratory syncytial virus M2-1 protein quaternary structure by reversible zinc removal from its Cys(3)-His(1) motif. *Biochemistry* **52**, 6779-89.
12. Tanner, S. J., Ariza, A., Richard, C. A., Kyle, H. F., Dods, R. L., Blondot, M. L., Wu, W., Trincao, J., Trinh, C. H., Hiscox, J. A., Carroll, M. W., Silman, N. J., Eleouet, J. F., Edwards, T. A. & Barr, J. N. (2014). Crystal structure of the essential transcription antiterminator M2-1 protein of

human respiratory syncytial virus and implications of its phosphorylation. *Proc Natl Acad Sci U S A* **111**, 1580-5.

13. Blondot, M. L., Dubosclard, V., Fix, J., Lassoued, S., Aumont-Nicaise, M., Bontems, F., Eleouet, J. F. & Sizun, C. (2012). Structure and functional analysis of the RNA- and viral phosphoprotein-binding domain of respiratory syncytial virus M2-1 protein. *PLoS Pathog* **8**, e1002734.
14. Brockmann, C., Soucek, S., Kuhlmann, S. I., Mills-Lujan, K., Kelly, S. M., Yang, J. C., Iglesias, N., Stutz, F., Corbett, A. H., Neuhaus, D. & Stewart, M. (2012). Structural basis for polyadenosine-RNA binding by Nab2 Zn fingers and its function in mRNA nuclear export. *Structure* **20**, 1007-18.
15. Santangelo, T. J. & Artsimovitch, I. (2011). Termination and antitermination: RNA polymerase runs a stop sign. *Nat Rev Microbiol* **9**, 319-29.
16. Weisberg, R. A. & Gottesman, M. E. (1999). Processive antitermination. *J Bacteriol* **181**, 359-67.
17. Zhang, J. & Landick, R. (2016). A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem Sci* **41**, 293-310.
18. Cuesta, I., Geng, X., Asenjo, A. & Villanueva, N. (2000). Structural phosphoprotein M2-1 of the human respiratory syncytial virus is an RNA binding protein. *J Virol* **74**, 9858-67.
19. Hardy, R. W. & Wertz, G. W. (2000). The Cys(3)-His(1) motif of the respiratory syncytial virus M2-1 protein is essential for protein function. *J Virol* **74**, 5880-5.
20. Leyrat, C., Renner, M., Harlos, K., Huiskonen, J. T. & Grimes, J. M. (2014). Drastic changes in conformational dynamics of the antiterminator M2-1 regulate transcription efficiency in Pneumovirinae. *Elife* **3**, e02674.
21. Cattoni, D. I., Chara, O., Kaufman, S. B. & Gonzalez Flecha, F. L. (2015). Cooperativity in Binding Processes: New Insights from Phenomenological Modeling. *PLoS One* **10**, e0146043.
22. Vogt, A. D. & Di Cera, E. (2012). Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. *Biochemistry* **51**, 5894-902.
23. Nilson, K. A. & Price, D. H. (2011). The Role of RNA Polymerase II Elongation Control in HIV-1 Gene Expression, Replication, and Latency. *Genet Res Int* **2011**, 726901.
24. Greenblatt, J., Nodwell, J. R. & Mason, S. W. (1993). Transcriptional antitermination. *Nature* **364**, 401-6.
25. Kumar, P. K., Kumarevel, T. & Mizuno, H. (2006). Structural basis of HutP-mediated transcription anti-termination. *Curr Opin Struct Biol* **16**, 18-26.
26. Phadtare, S. & Severinov, K. (2010). RNA remodeling and gene regulation by cold shock proteins. *RNA Biol* **7**, 788-95.
27. Gusarov, I. & Nudler, E. (1999). The mechanism of intrinsic transcription termination. *Mol Cell* **3**, 495-504.
28. Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* **34**, 1062-6.
29. Schlereth, J., Grunweller, A., Biedenkopf, N., Becker, S. & Hartmann, R. K. (2016). RNA binding specificity of Ebola virus transcription factor VP30. *RNA Biol* **13**, 783-98.
30. Biedenkopf, N., Schlereth, J., Grunweller, A., Becker, S. & Hartmann, R. K. (2016). RNA Binding of Ebola Virus VP30 Is Essential for Activating Viral Transcription. *J Virol* **90**, 7481-96.
31. Hartlieb, B., Muziol, T., Weissenhorn, W. & Becker, S. (2007). Crystal structure of the C-terminal domain of Ebola virus VP30 reveals a role in transcription and nucleocapsid association. *Proc Natl Acad Sci U S A* **104**, 624-9.

32. Williams, S. G. & Hall, K. B. (2014). Linkage and allostery in snRNP protein/RNA complexes. *Biochemistry* **53**, 3529-39.
33. Mackereth, C. D. & Sattler, M. (2012). Dynamics in multi-domain protein recognition of RNA. *Curr Opin Struct Biol* **22**, 287-96.
34. Recht, M. I. & Williamson, J. R. (2001). Central domain assembly: thermodynamics and kinetics of S6 and S18 binding to an S15-RNA complex. *J Mol Biol* **313**, 35-48.
35. John, S. P., Wang, T., Steffen, S., Longhi, S., Schmaljohn, C. S. & Jonsson, C. B. (2007). Ebola virus VP30 is an RNA binding protein. *J Virol* **81**, 8967-76.
36. Ames, J. B., Porumb, T., Tanaka, T., Ikura, M. & Stryer, L. (1995). Amino-terminal myristoylation induces cooperative calcium binding to recoverin. *J Biol Chem* **270**, 4526-33.
37. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12.

FIGURE LEGENDS

Figure 1. Schematic and surface structure representation of RSV $M_{2\text{-}1}$ modules and RNA binding residues. A) Schematic representation of $M_{2\text{-}1}$. The residues corresponding to the Zinc binding region, oligomerization region and the core domain obtained by chymotripsin digestion, are indicated above the graph. The alpha-helices and RNA binding residues reported, are indicated below the graph and colored in cyan [12] and red [13]. B) Surface representation of the $M_{2\text{-}1}$ tetramer structure (PDB: 4C3D completed as explained in Materials and Methods and Figure S13) showing a bottom view (left), side view (center) and N-terminal top view (right). RNA binding residues: K92-V97, L149-L152 and D155-K159 are highlighted in red and R3-R4 in cyan. One monomer is differentially colored in dark gray.

Figure 2. Stoichiometry of the $M_{2\text{-}1}$:RNA interaction probed by Electrophoretic mobility shift assay (EMSA). Upper Panel: Native acrylamide gel electrophoresis of complexes. (A-B) $M_{2\text{-}1}$ tetramer:RNA interactions. A fixed concentration of 4 μM 10mer (A) or 200 nM 20mer (B) 6-FAM labeled RNA was incubated 20 minutes with increasing amounts of $M_{2\text{-}1}$ tetramer up to a 1.25 to 5 molar excess. (C-D) $M_{2\text{-}1}$ Core domain:RNA interactions. A fixed concentration of 4 μM 10mer (C) or 20mer (D) 6-FAM labeled RNA was incubated at room temperature with increasing amounts of $M_{2\text{-}1}$ core domain up to a 5-fold molar excess. Lower Panel: Stoichiometry for each tested interaction (A-D) was determined by densitometry of free RNA bands. Plots represent the fraction of free RNA as a function of $M_{2\text{-}1}$:RNA molar ratio. Experiments were performed in duplicate (filled and open circles).

Figure 3. Equilibrium titration of the $M_{2\text{-}1}$ tetamer:RNA interaction. (A) Fluorescence titration of 10 nM, 20 nM and 40 nM 20mer 6-FAM labeled RNA with increasing amounts of $M_{2\text{-}1}$ tetramer. Fluorescence intensity at 525 nm is represented as a function of the $M_{2\text{-}1}$ tetramer:RNA molar

ratio. Stoichiometric titration at a 1:2 (M_{2-1}tetramer:RNA) ratio is observed as 6-FAM-RNA probe concentration is increased. (B) Fluorescence titration of 10mer (filled triangles), 15mer (open circles) and 20mer (filled circles) RNAs with the M_{2-1} tetramer. Fixed concentrations of 50 nM (10mer RNA) or 25 nM (15mer and 20mer RNA) were used. (C) Fluorescence titration of 10 nM, 20 nM and 40 nM 20mer 6-FAM labeled RSV-RNA with increasing amounts of M_{2-1} tetramer. Fluorescence intensity at 525 nm is represented as a function of the M_{2-1}tetramer:RNA molar ratio. Experiments were performed in 50 mM Tris-HCl (pH 7.5), 0.1 M NaCl and 1 mM DTT at 20ºC. Stoichiometric titration at a 1:2 (M_{2-1} tetramer:RNA) ratio is observed as 6-FAM-RNA probe concentration is increased.

Figure 4. Changes in RNA structure induced upon M_{2-1}:RNA interaction followed by circular dichroism. (A) Near-UV CD spectrum of 5 μM M_{2-1} tetramer (dashed line), 5 μM 20mer RNA (black solid line), 5 μM 15mer RNA (gray solid line) and 5 μM 10mer RNA (dotted line). (B) Titration of the M_{2-1} tetramer:20mer RNA complex formation followed by circular dichroism. Near-UV CD spectra of 5 μM 20mer RNA incubated with increasing amounts of M_{2-1} tetramer up to 10 μM. Inset: Changes in the molar ellipticity signal for the 20mer RNA band followed at 270 nm as a function of the M_{2-1} tetramer:RNA molar ratio. Stoichiometric titration at a 1:2 M_{2-1}tetramer:RNA molar ratio is observed. Spectra were acquired in 50 mM Tris-HCl (pH 7.5), 0.1 M NaCl and 1 mM DTT at 20ºC.

Figure 5. RNA structure and M_{2-1}:RSV-RNA interaction followed by circular dichroism. (A) Near-UV CD spectra of 5 μM of: 20mer RNA (black solid line), 20mer RSV-RNA (dashed line) and 13mer RSV-RNA (dotted line). (B) Thermal denaturation of 5 μM of: 20mer RSV-RNA (black dots) and non-cognate 20mer-RNA (gray dots), from 5ºC to 80ºC followed at 270 nm. (C) Titration of 13mer

RSV-RNA and 20mer RSV RNA with increasing amounts of $M_{2\text{-}1}$ tetramer followed by circular dichroism. Changes in the molar ellipticity signal for the RNA band followed at 270 nm as a function of the $M_{2\text{-}1}$ tetramer:RNA molar ratio. Stoichiometric titration at a 1:2 ($M_{2\text{-}1}$tetramer:RNA) molar ratio is observed in both cases. Spectra were acquired in 50 mM Tris-HCl (pH 7.5), 0.1 M NaCl and 1 mM DTT at 20ºC.

Figure 6. Cooperative association of non-cognate and cognate RNAs to the $M_{2\text{-}1}$ tetramer. A) Equilibrium titration of 20mer 6-FAM-labeled RNA with the $M_{2\text{-}1}$ tetramer. A sigmoidal shape is clearly recognizable in the binding curves. Shown is a fit to the Hill equation (see materials and methods. Eq [5]) which yielded K_{Dapp} = 12.0 ± 3.2 nM and n_H = 2.2 ± 0.2. Inset: Detail of the first segment of the curve where the sigmoidal shape is observed. B) Linear transformation of the binding data shown in panel A (see materials and methods, Eq [6]) used to estimate the Hill coefficient for three different concentrations of each RNA species (Figure S8 and Table S1).

Figure 7. $M_{2\text{-}1}$tetramer:20mer RNA binding kinetics and analysis of slow association phases in the $M_{2\text{-}1}$:RNA interaction. Time traces of the fluorescence anisotropy (A) and 525 nm fluorescence intensity (B) signals were obtained using 5 nM 20mer 6-FAM labeled RNA. After 100 seconds, 100 nM $M_{2\text{-}1}$ tetramer was added by manual mixing (deadtime 5 seconds) and traces followed until equilibrium was reached. Experiments were performed in 50 mM Tris-HCl (pH 7.5), 0.1 M NaCl and 1 mM DTT at 20 ºC. The anisotropy signal was fitted to a monoexponential decay (see materials and methods), with k_{obs}= 0.020 ± 0.002 s^{-1} (A) and the fluorescence signal was fitted to a monoexponential decay (black solid line) or to a biexponential decay, with k_{obs1}= 0.011 ± 0.001 s^{-1}, and k_{obs2}= 0.097 ± 0.004 s^{-1} (B). Inset: residuals of the curve fittings. (C) Analysis of the amplitude of slow (clear circles) and fast (filled circles) kinetic association phases obtained from

fluorescence intensity data as a function of $M_{2\text{-}1}$ concentration following $M_{2\text{-}1}$:RNA binding. 20mer 6-FAM labeled RNA concentration was 5 nM. (D) Dependence of the observed kinetic rate constant on $M_{2\text{-}1}$ concentration and fit to an induced-fit model (see materials and methods, Eq [4]). Filled circles: fast fluorescence intensity phase, Clear circles: slow fluorescence intensity phase, clear triangles: fluorescence anisotropy phase. The parameters obtained for the fit were: Fast fluorescence phase k_1 = 0.16 ± 0.1 s^{-1}, K_{-1} = 89 ± 6 nM (k_{-1} was fixed to 0.01 s^{-1}).

Figure 8. Model of cooperative interaction between the $M_{2\text{-}1}$ tetramer and RNA. A) The $M_{2\text{-}1}$ tetramer associates in a non-cooperative mode to short RNAs, which each cover a single binding site from each monomer. The resulting complex has 4 RNA molecules bound per $M_{2\text{-}1}$ tetramer, with a K_D = 250 nM. B) The model depicts sequential binding of two long RNA molecules to the $M_{2\text{-}1}$ tetramer. The fast bimolecular association steps proceed in a stepwise manner, and involve disruption of base stacking interactions in the RNA. Positive cooperativity is observed, with an apparent binding constant of K_{Dapp} = 16 nM, and the Hill coefficient is 2.5 for cognate RNA. Following binding, slow conformational rearrangements of the complex occur, yielding a final complex formed by two RNA molecules per $M_{2\text{-}1}$ tetramer.

FIGURES

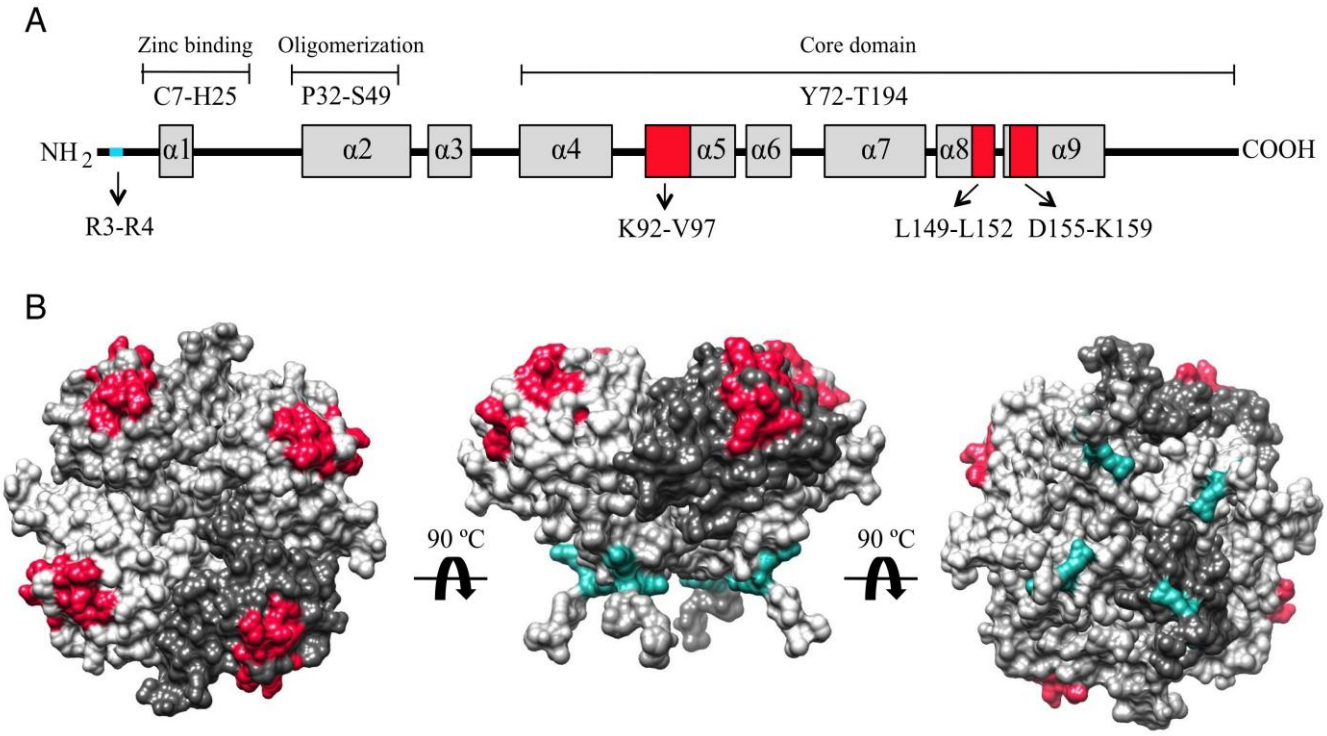


Figure 1.

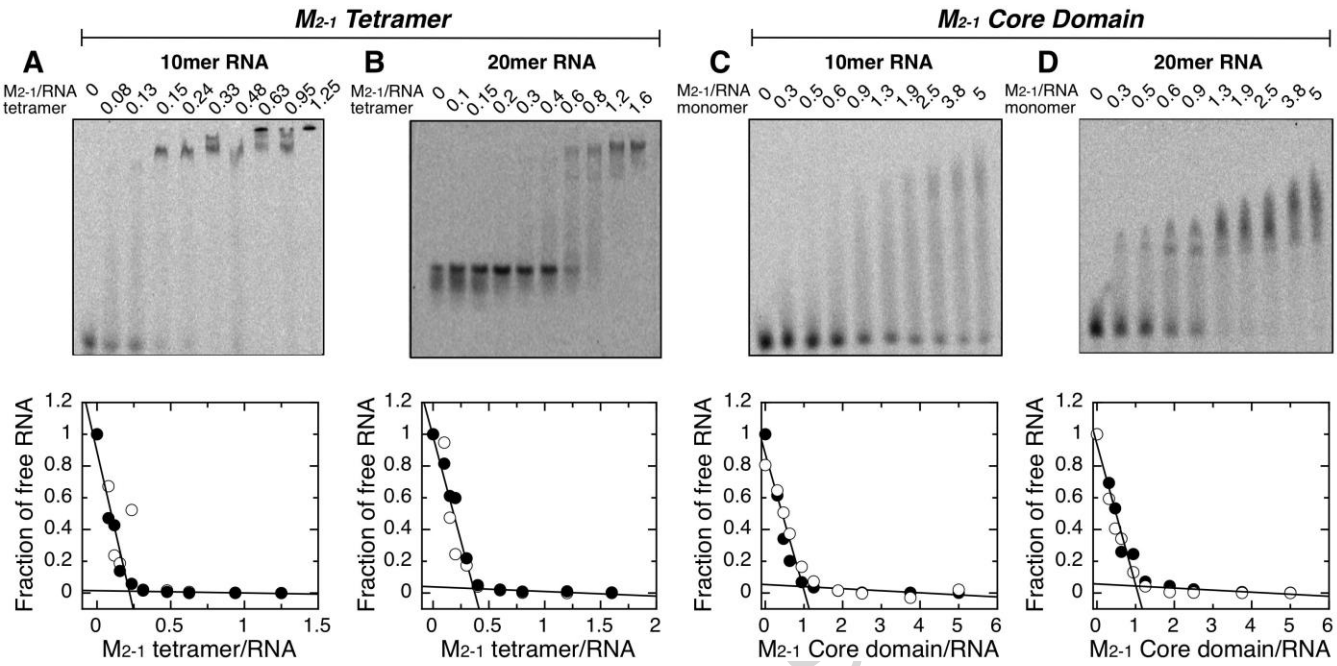


Figure 2.

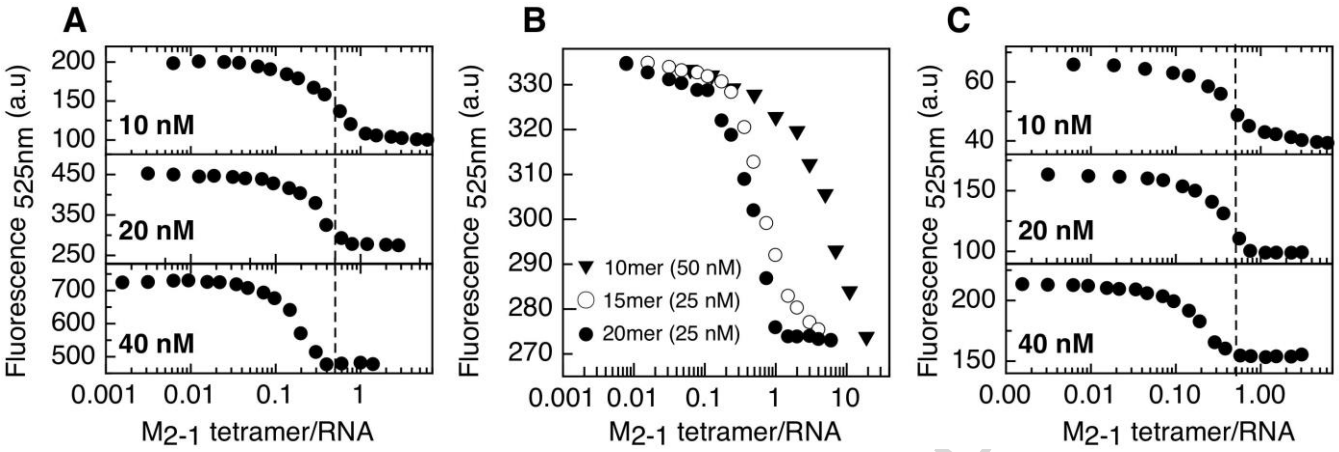


Figure 3.

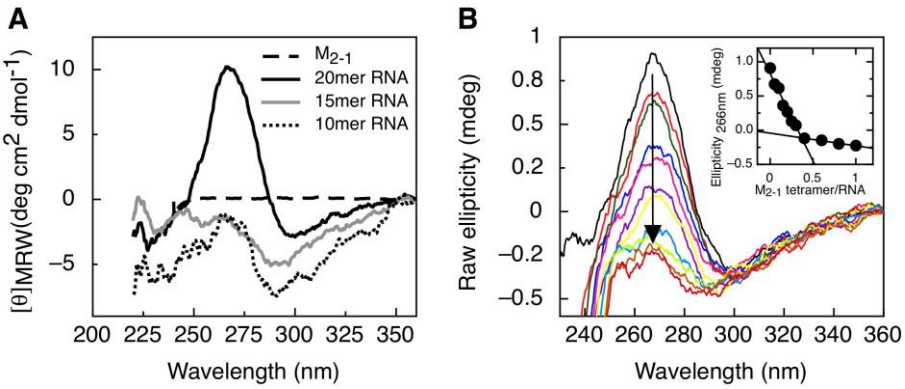


Figure 4.

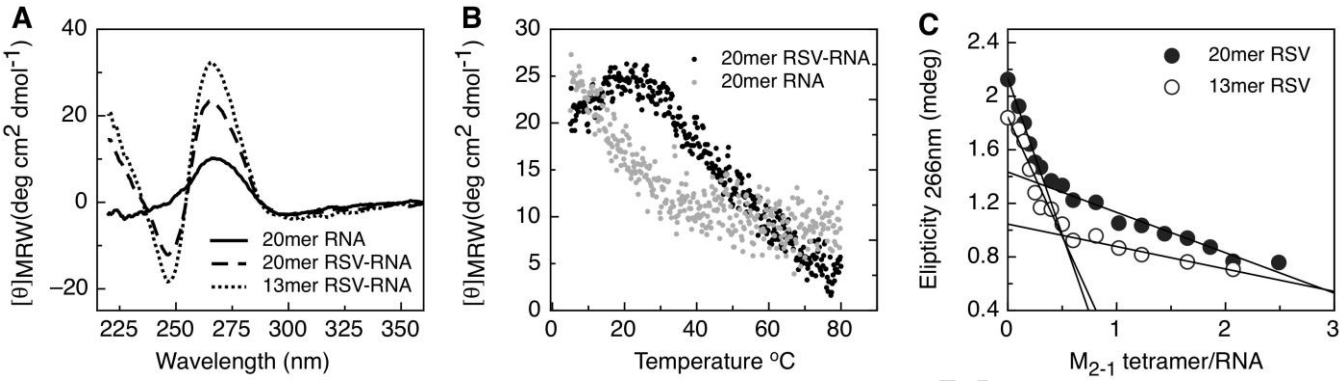


Figure 5.

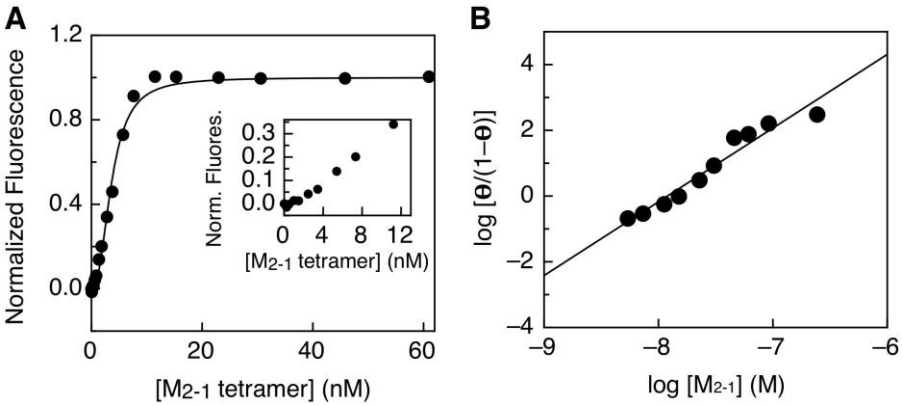


Figure 6.

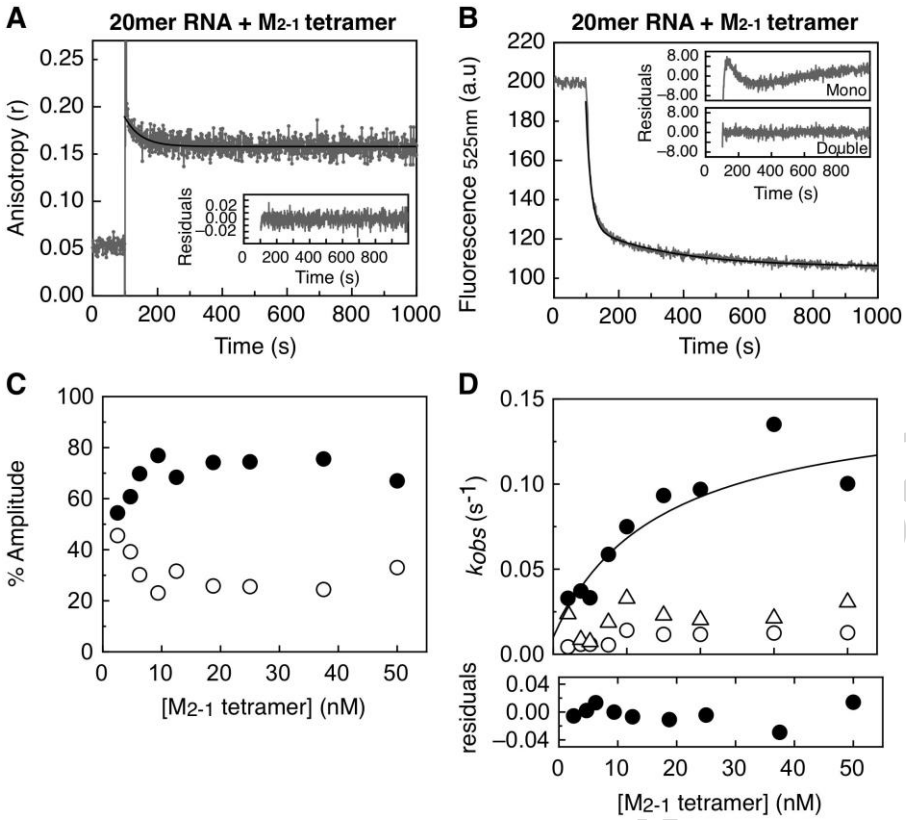


Figure 7.

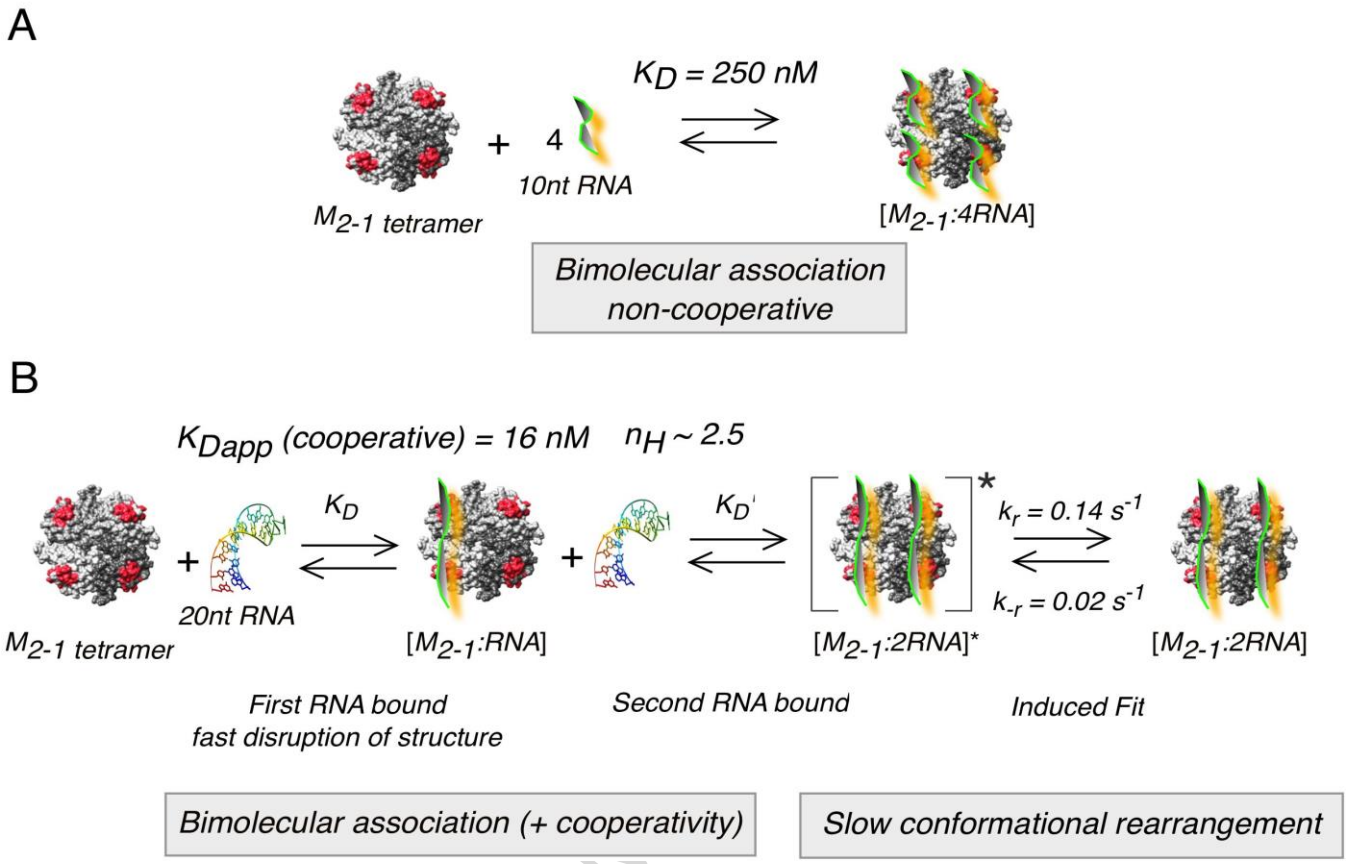


Figure 8

TABLE

Table 1. Binding of the $M_{2\text{-}1}$ Core Domain monomer and $M_{2\text{-}1}$ tetramer to RNA

$M_{2\text{-}1}$ Species	RNA length	Stoich [a] EMSA **$M_{2\text{-}1}$:RNA**	Stoich [b] Fluo **$M_{2\text{-}1}$:RNA**	Stoich [c] CD **$M_{2\text{-}1}$:RNA**	K_D (nM) [d]	Fitted Signal	Hill Coef n_H
model RNAs binding to $M_{2\text{-}1}$ Core Domain monomer							
Core Domain	10	1:1	ND	NS	236 ± 232	EMSA[$]	ND
Core Domain	20	1:1	1:1	NC	239 ± 110	Anisotropy/ EMSA	ND
model RNAs binding to $M_{2\text{-}1}$ tetramer							
Tetramer	10	**1:4**	**1:4**	NS	286 ± 171	Anisotropy/ EMSA	ND
Tetramer	15	**1:2**	**1:2**	NS	ND	ND	ND
Tetramer	20	**1:2**	**1:2**	**1:2**	14 ± 6 **	Fluorescence	2.3 ± 0.5**
RSV cognate RNA binding to $M_{2\text{-}1}$ tetramer							
Tetramer	13	2:1	ND	2:1	ND	ND	ND
Tetramer	20	ND	2:1	2:1	16 ± 5 **	Fluorescence	2.5 ± 0.7**

ND = not determined; NS = no signal in RNA; NC = no change in RNA upon binding

[a] Stoichiometry was determined by EMSA performed in saturating conditions (4µM FITC-RNA). Data is from 2 or more independent experiments.

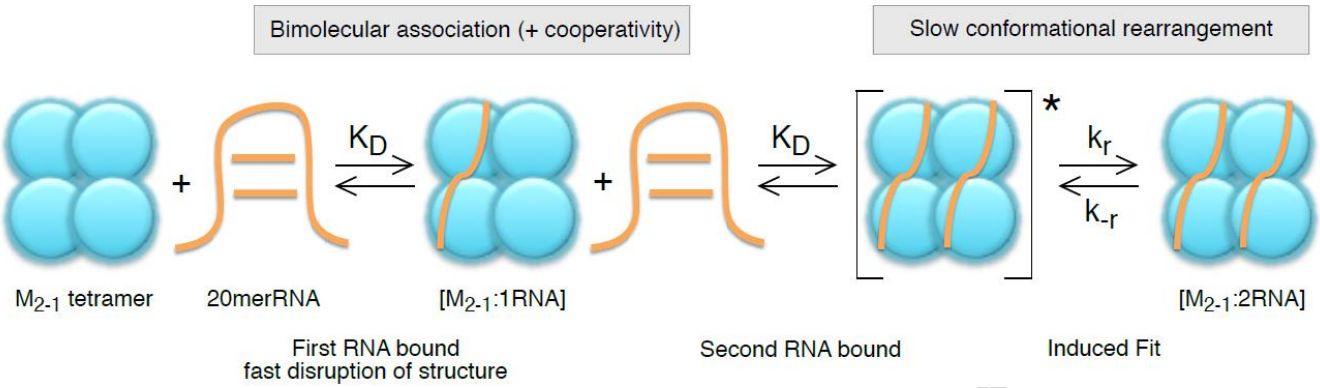
[b] Stoichiometry was determined by equilibrium titration with FITC-RNA in saturating conditions (cc > 1µM). Data is from 3 or more independent experiments.

[c] Stoichiometry was determined by CD measurements of RNA-$M_{2\text{-}1}$ titrations at 5µM 20mer-RNA concentration. Data is from 1 experiment.

[d] Data is from 2 to 7 independent experiments

** Apparent KD and nH were obtained from fitting to the Hill equation (See materials and methods, Eq. [6]). Data represent average values from 6 to 7 experiments at different concentrations (Table S2)

[$] Fitting quality for fluorescence/anisotropy data was poor due to sample oligomerization or aggregation and was therefore excluded fromthe reported K_D



Graphical abstract

Highlights

Transcription antitermination finely tunes protein levels in mononegavirales.

RSV antiterminator $M_{2\text{-}1}$ binds two 20mer RNAs cooperatively and 10mers independently.

Cooperative binding unfolds collapsed type of structure in cognate and non-cognate RNA.

Successive binding and slow rearrangement events link cooperativity to induced fit.

Insights into primary events of transcription antitermination of mononegavirales.