Published in final edited form as:

Virus Res. 2009 September ; 144(1-2): 1–7. doi:10.1016/j.virusres.2009.02.005.

Characterization of *Ebolavirus* Regulatory Genomic Regions

Gabriele Neumann¹, Shinji Watanabe¹, and Yoshihiro Kawaoka^{1,2,3,4,*}

¹Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA

²Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

³Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama 332-0012, Japan

⁴International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

SUMMARY

For filoviruses, such as *Ebolavirus* and the closely related *Marburgvirus*, transcriptional regulation is poorly understood. The open reading frames (ORFs) that encode the viral proteins are separated by regulatory regions composed of the 3' nontranslated region (NTR) of the upstream gene, highly conserved transcription stop and start signals, and the 5'NTR of the downstream gene. The conserved transcription stop and start signals either overlap, or they are separated by intergenic regions (IGRs) of different lengths. To assess the contribution of the regulatory regions to transcription, we established bicistronic minireplicons in which these regions were flanked by upstream and downstream ORFs, the *Ebolavirus* leader and trailer regions, and by T7 RNA polymerase promoter and ribozyme sequences. We found that the individual viral regulatory regions differ in their ability to direct protein synthesis from the upstream or downstream ORFs. Deletion or modification of the NTRs, IGRs, or transcription stop and start signals affected protein expression levels to various extents; for example, 5'NTRs appear to affect efficient protein expression from the downstream ORF, whereas 3'NTRs seem to attenuate protein expression from the upstream ORF. Overall, our data suggest that the regulation of *Ebolavirus* protein levels is complex.

Keywords

Ebolavirus; transcription; minireplicon; nontranslated region (NTR); intergenic region (IGR)

1. INTRODUCTION

Ebolavirus and the closely related *Marburgvirus* form the family *Filoviridae* in the order *Mononegavirales* (reviewed in Sanchez et al., 2007). Within the genus *Ebolavirus*, there are four species, *Zaire ebolavirus* (ZEBOV), *Sudan ebolavirus*, *Ivory Coast ebolavirus*, and *Reston ebolavirus*. *Ebola-* and *Marburgviruses* cause hemorrhagic fever with mortality rates as high as 80%.

The genome organization and the replication strategy of filoviruses closely resemble those of other nonsegmented, negative-sense RNA viruses (Feldmann et al., 1992, 1993). The genomes

*Corresponding Author: Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, 2015 Linden Drive, Madison, WI 53706, USA. Phone: (608) 265-4925; Fax: (608) 265-5622; kawaokay@svm.vetmed.wisc.edu.

of filoviruses consist of a nonsegmented, single-stranded, negative-sense RNA of about 19 kb in length that encodes seven structural proteins in the following order (in the positive-sense orientation): NP (nucleoprotein) – VP35 (a component of the replication/transcription complex) – VP40 (a matrix protein) – GP (glycoprotein; *Ebolaviruses* also encode a secreted glycoprotein from the GP gene) – VP30 (a component of the replication/transcription complex) – VP24 (a minor matrix protein) – L (an RNA-dependent RNA polymerase) (Fig. 1a). Each gene is flanked by conserved transcription start and stop signals (Fig. 1) (Feldmann et al., 1992; Sanchez et al., 1993). At the ends of the genome, there are extragenic regions, referred to as *leader* and *trailer*, which contain promoters for replication and transcription.

Very little is known about the regulation of filoviral gene transcription. Genes located closer to the 3'-end of the genome are transcribed at higher levels than those located further downstream (Feldmann et al., 1992; Muhlberger et al., 1996). This 'transcriptional gradient' likely results from polymerase complexes that, after transcription terminates, fail to reinitiate at the downstream start signal. However, other factors may also contribute to the regulation of transcription, such as the organization of the transcription stop and start signals, or the composition of noncoding and/or intergenic regions.

For ZEBOV, the transcription start signals comprise 12 nucleotides that differ in only one position among the ZEBOV genes (Fig. 1b). Likewise, the transcription stop signals (11–12 nucleotides in length) are highly conserved among ZEBOV genes; the only difference is the length of the uridine stretch (adenosine stretch in the positive-sense orientation), which functions as a polyadenylation signal, and the transcription stop signal of the L gene, which deviates from the consensus sequences at two positions (Fig. 1b).

The transcription stop and start signals of nonsegmented, negative-sense RNA viruses are separated by intergenic regions (IGRs). For *Ebolavirus*, the IGRs differ in length; for three gene boundaries, VP35-VP40, GP-VP30, and VP24-L, the transcription stop and start signals overlap (Fig. 1a,c) (Sanchez et al., 1993). In this case, the transcription start signal of the downstream gene precedes the transcription stop signal of the upstream gene; i.e., after terminating transcription, the polymerase has to track back to initiate transcription of the downstream gene. The overlap of the two signals is brought about by a conserved pentanucleotide ATTAA that is shared by transcription stop and start signals (Fig. 1). Overlapping transcription stop and start signals also exist for *Marburgviruses*, albeit for different genes. The biological significance of filoviral IGRs and these gene overlaps is not known. The filoviral genes contain gene-specific 5'- and 3'-nontranslated regions (NTRs; Fig. 1a) that are relatively long compared to those of other nonsegmented, negative-sense RNA viruses (Feldmann et al., 1992; Muhlberger et al., 1992; Sanchez et al., 1989, 1992, 1993). Computer analysis has revealed potential stem-loop structures in the 5'NTRs of filoviral mRNAs (Muhlberger et al., 1996; Sanchez et al., 1993), which may, for example, increase mRNA stability. However, the biological significance of these NTRs remains to be established.

Research on nonsegmented, negative-sense RNA viruses has benefited from the establishment of so-called minireplicon systems. In these systems, virus-like RNAs are synthesized that express quantifiable reporter proteins instead of viral proteins and thus allow one to assess the levels and/or kinetics of gene expression. Minireplicon approaches or reverse genetics approaches (which allow the artificial generation of filoviruses) (Neumann et al., 2002; Volchkov et al., 2001; Watanabe et al., 2004) have been used to identify the filoviral proteins required for replication and transcription (Muhlberger et al., 1999), and to address cross-compatibilities among *Ebola*- and *Marburgvirus* proteins (Boehmann et al., 2005; Theriault et al., 2004). A minireplicon approach was also used to study the role of the *Ebolavirus* leader region in the regulation of transcription (Weik et al., 2005). Here, we use a minireplicon

approach to study the regulation of *Ebolavirus* transcription. Such knowledge will increase our fundamental understanding of the viral life cycle.

2. METHODS

2.1. Generation of bicistronic *Ebolavirus* minireplicons

To generate bicistronic *Ebolavirus* minireplicons, the components described below were PCR-amplified with oligonucleotides that contain recognition sequences for so-called type IIS restriction endonucleases. This strategy allows the nucleotide-specific fusion of any two DNA segments. The respective components were then joined step-wise and inserted into a derivative of pACYC177 (Rose, 1988). In particular, we assembled the following components: the wild-type T7 RNA polymerase promoter, as described (Neumann et al., 2002), the leader region of ZEBOV and the 5' nontranslated region of the NP mRNA (i.e., nucleotides 1–55 and 56–469 of the antigenomic RNA of ZEBOV), the open reading frame (in the antisense orientation) of the luciferase (Luc) gene, type IIS cloning sites for the insertion of regulatory regions (RRs), the open reading frame (in the antisense orientation) of the chloramphenicol-acetyltransferase (CAT) gene, the 3' nontranslated region of the L mRNA and the trailer region of ZEBOV (i.e., nucleotides 18220–18282 and 18283–18959 of the antigenomic RNA of ZEBOV), and a ribozyme sequence, as described (Neumann et al., 2002).

In the next step, the individual RRs of ZEBOV were PCR-amplified and inserted between the Luc and CAT open reading frames, relying on type IIS cloning strategies. The individual RRs encompass nucleotides 2690–3128 (NP/VP35-RR), 4152–4478 (VP35/VP40-RR), 5460–6038 (VP40/GP-RR), 8069–8508 (GP/VP30-RR), 9376–10344 (VP30/VP24-RR), and 11101–11580 (VP24/L-RR); the nucleotide positions refer to the positive-sense orientation. In the resulting constructs, the CAT and Luc open reading frames are thus fused between the individual regulatory regions of the ZEBOV genome. Variants of ZEBOV RRs were generated by site-directed mutagenesis. All DNA segments generated by PCR or mutagenesis were confirmed by sequence analysis.

2.2. Reporter gene assays

To test bicistronic minireplicons for CAT and Luc expression levels, we transfected 10^6 human embryonic kidney (293T) cells with the following plasmids: 1 µg of the respective minireplicon construct, 0.5 µg of pC-T7pol (for the expression of the T7 RNA polymerase), 0.25 µg of pCEZ-NP (expressing ZEBOV NP protein), 0.15 µg of pCEZ-VP30 (expressing ZEBOV VP30 protein), 0.25 µg of pCEZ-VP35 (expressing ZEBOV VP35 protein), and 2 µg of pCEZ-L (expressing ZEBOV L protein). Forty-eight hours later, cells were lysed in 300 µl Glo lysis buffer (Promega) and the total protein content of each sample was determined. Equal amounts of protein were used to determine the Luc expression levels (Steady-Glo luciferase assay system; Promega) and CAT expression levels (CAT ELISA; Roche). All samples were processed in triplicate. Two to three independent experiments were carried out per construct.

3. RESULTS AND DISCUSSION

3.1. Comparison of wild-type regulatory regions

The contribution of the sequences that separate the ZEBOV open reading frames (referred to as regulatory regions, RRs; Fig. 1a) to the regulation of *Ebolavirus* protein expression is unknown. We, therefore, constructed bicistronic minireplicons in which the Luc and CAT ORFs were flanked by the ZEBOV *leader* and *trailer* sequences and separated by ZEBOV RRs (Fig. 2). The viral sequences were flanked by the T7 RNA polymerase promoter and ribozyme sequences.

We first generated a series of bicistronic minireplicons that contained the authentic ZEBOV RRs, which separate the NP/VP35, VP35/VP40, VP40/GP, GP/VP30, VP30/VP24, and VP24/L ORFs (Fig. 2). To assess the levels of reporter gene expression, 293T cells were transfected with the respective minireplicon plasmid, and with plasmids for the expression of the T7 RNA polymerase and the ZEBOV NP, VP30, VP35, and L proteins. Forty-eight hours later, the transfected cells were lysed and subjected to Luc and CAT assays.

We detected negligible levels of CAT and Luc expression in nontransfected cells or cells transfected with a bicistronic minireplicon only (i.e., in the absence of support proteins; data not shown). By contrast, significant levels of CAT and Luc expression were detected in cells that received a minireplicon in combination with plasmids expressing the T7 RNA polymerase and the components of the ZEBOV replication/transcription complex (Fig. 2). For data analysis, we defined the Luc and CAT expression levels detected for the VP40/GP RR as 100%. First, we compared Luc expression levels. Transcription of the Luc gene is initiated in the leader region, which is conserved among the constructs tested. Differences in Luc expression levels should therefore reflect differences in mRNA stability (which may be affected by the 3' nontranslated region), and/or differences in the efficiency of transcription termination. In fact, we detected significant differences among the ZEBOV RRs in their ability to direct expression of the upstream ORF (Fig. 2). The VP40/GP-RR resulted in the highest levels of Luc expression, suggesting efficient termination of transcription and/or high mRNA stability. The VP35/VP40-, GP/VP30-, and VP24/L-RRs directed intermediate levels of Luc expression, whereas the NP/VP35- and VP30/VP24-RRs directed significantly lower levels of Luc expression. These data suggest that the sequences in the latter two RRs may down-regulate the expression levels of the NP and VP30 proteins, respectively. RRs that contain identical transcription stop signals (such as the VP35/VP40-, GP/VP30-, and VP30/VP24-RRs; see Fig. 1) directed different levels of Luc expression, suggesting that the nontranslated regions and/or the arrangement of the transcription stop/start signals (non-overlapping vs. overlapping) contribute to efficient protein expression.

Next, we assessed the levels of CAT expression. In our experimental system, CAT transcription initiates at the transcription start signal in the RR and is terminated at the transcription stop signal in the trailer region, which is identical among all of our test constructs; CAT expression is hence controlled by the 5' nontranslated region under investigation and the 3' nontranslated region of the L gene. Differences in CAT expression levels should therefore reflect differences in the efficiency of transcription initiation, differences in mRNA stability, and/or differences in the level of readthrough transcripts, which may prevent re-initiation of transcription. We did not detect significant differences in CAT expression levels among the NP/VP35-, VP35/VP40-, VP40/GP-, and VP30/VP24-RRs, all of which directed efficient CAT expression.

Two RRs – the GP/VP30- and VP24/L-RRs – were slightly less efficient in directing CAT expression ($p=0.015$ and $p=0.022$, respectively). Hence, sequences in these RRs may down-regulate the expression levels of the VP30 and L proteins. VP30 protein expression may thus be down-regulated in *Ebolavirus*-infected cells via at least two mechanisms: sequences in the VP30 transcription start signal or 5'NTR (as suggested by the low CAT expression levels obtained with the GP/VP30-RR), and sequences in the VP30 transcription stop signal or 3'NTR (as suggested by the low Luc expression levels obtained with the VP30/VP24-RR). VP30 is a transcription activator (Weik et al., 2002) and its levels may have to be controlled tightly to balance transcription vs. replication.

3.2. Significance of noncoding regions for ZEBOV protein expression

The *Ebolavirus* genes contain relatively long 5' and 3'NTRs of unknown function, which may, for example, form secondary structures that affect the stability of the respective mRNAs (Muhlberger et al., 1996; Sanchez et al., 1993). To address this question, we tested several

deletion constructs, based on the NP/VP35-RR (Fig. 3). In one construct (NP/VP35-RR Δ 3'NTR), we deleted the 3'NTR, that is, the sequences between the stop codon of the NP ORF (i.e., the Luc ORF in the minireplicon) and the transcription stop signal (Fig. 3). In another construct (NP/VP35-RR Δ 5'NTR), we deleted the 5'NTR, that is, the sequence between the transcription start signal and the start codon of the VP35 ORF (i.e., the CAT ORF in the minireplicon). A third construct (NP/VP35-RR Δ NTRs) combined these two deletions, so that the Luc stop codon directly bordered the transcription stop signal and the transcription start signal directly bordered the CAT start codon.

Interestingly, deletion of the 3'NTR of the NP gene increased Luc expression (Fig. 3). Hence, the 3'-noncoding region of the NP gene is not only dispensable for efficient expression of the upstream gene, but seems to attenuate this process. Deletion of this region had only a minor effect on CAT expression.

Deletion of the 5'NTR of the VP35 gene reduced the CAT expression levels relative to the wild-type RR (Fig. 3), possibly by affecting mRNA stability. Computational analysis suggested that the transcription start signals and the downstream 5'NTRs fold into stem-loop structures (Muhlberger et al., 1996; Sanchez et al., 1993), which would be destroyed on deletion of the 5'NTRs. Collectively, our findings suggest that the 5'NTR is important for efficient expression from the downstream ORF, although its deletion does not completely abrogate CAT expression.

We also found that deletion of the 5'NTR of the VP35 gene increased Luc expression levels (Fig. 3), although the deleted region was not part of the luciferase (NP) gene. This finding may suggest cooperative effects such as secondary structure formation between the 5' and 3'NTRs, providing another level of regulation of gene expression. Deletion of both NTRs resulted in significantly higher expression levels of the upstream gene, and reduced expression levels of the downstream gene (Fig. 3).

We then asked whether the observed effects were specific for the NP/VP35-RR, or extend to other RRs by generating a similar deletion construct based on the VP35/VP40-RR (Fig. 3, lower panel). Deletion of both NTRs (VP35/VP40-RR Δ NTR) again resulted in the upregulation of Luc and downregulation of CAT expression, although the relative contributions of the 5' and 3'NTRs to the Luc and CAT expression levels differed from those observed with the NP/VP35-RR. Together, these findings indicate that the 5' and 3'NTRs play opposite roles in the regulation of ZEBOV protein expression: 5'NTRs appear critical for efficient expression of the downstream ORF, whereas 3'NTRs seem to attenuate expression of the upstream ORF.

3.3. Significance of overlapping transcription start/stop signals for ZEBOV protein expression

The ZEBOV RRs VP35/VP40- and GP/VP30-RR are characterized by overlapping transcription start/stop signals (Fig. 1c). In addition, the VP24/L-RR is unusual in that it contains two transcription stop signals, the second of which overlaps with the transcription start signal (Fig. 1c). The significance of these overlapping stop/start signals for the regulation of ZEBOV transcription is unknown. We, therefore, separated the overlapping signals of the VP35/VP40-RR by repositioning either the transcription stop signal or the transcription start signal first (Fig. 4).

When the transcription stop signal was in front of the start signal (VP35/VP40-RRSTOP-START), Luc expression increased relative to the wild-type RR. A stop signal in the first position may thus be more efficient than when situated downstream of a start signal. Compared to the authentic VP35/VP40 regulatory region, VP35/VP40-RR-STOP-START directed

slightly reduced levels of CAT expression, a finding that may be attributed to the lack of an intergenic region in this construct (this issue is addressed in more detail in section 3.5).

When the transcription stop signal was shifted downstream by duplicating the conserved pentanucleotide sequence (VP35/VP40-RR-START-STOP; Fig. 4), Luc expression levels were comparable to those obtained for the wild-type VP35/VP40 RR. VP35/VP40-RR-START-STOP also directed efficient CAT expression, suggesting that after transcription termination, the polymerase complex can slide backwards along the full length of the transcription start signal. It is not clear how the polymerase complex then ‘ignores’ the transcription stop signal that it encounters immediately downstream. Collectively, our findings suggest that overlapping transcription stop and start signals are a means to fine-tune the expression levels of filoviral proteins.

To address the significance of the dual transcription stop signal in the VP24/L RR, we disabled the first or the second signal by replacing the polyadenylation site, that is, in the positive-sense orientation, the adenosine stretch (Fig. 4). For both mutant constructs (VP24/L-RRΔSTOP#1 and VP24/L-RRΔSTOP#2), we detected higher Luc but lower CAT expression levels relative to the wild-type construct, VP24/L-RR. Among the constructs tested, the wild-type sequence resulted in the lowest levels of Luc expression (equivalent to VP24 expression in the *Ebolavirus* genome), suggesting that two closely spaced transcription stop signals may attenuate termination, for example, by interference between neighboring polymerase complexes. The two closely spaced transcription stop signals may thus serve as a means to attenuate VP24 expression. On the other hand, the wild-type sequence resulted in the highest levels of CAT expression (equivalent to L expression in the *Ebolavirus* genome). Surprisingly, disruption of either of the two transcription stop signals significantly affected CAT expression levels (i.e., the downstream transcription unit), although the functional transcription start signal remained. These findings suggest complex fine-tuning of protein expression at the level of transcription start and stop signals.

3.4. Significance of intergenic regions (IGRs) for ZEBOV protein expression

The filoviral transcription stop and start signals either overlap (see previous section) or are separated by non-transcribed intergenic regions (IGRs). The IGRs are comprised of 4 or 5 nucleotides (as for the NP/VP35-, VP40/GP-, and VP24/L-RRs), or they are considerably longer (e.g., 144 nucleotides for the VP30/VP24-RR). The significance of the IGRs for transcriptional regulation remains unknown. We, therefore, deleted the IGR that separates the NP and VP35 transcription units (Fig. 5a). The resulting construct (NP/VP35-RRΔIGR) directed efficient Luc expression, but moderately reduced levels of CAT relative to the authentic NP/VP35-RR. CAT expression was also reduced for VP35/VP40-RR-STOP-START (Fig. 5a, see also Fig. 4) in which the overlapping transcription stop/start signals were separated. These findings demonstrate that although the ZEBOV IGRs are not required for gene expression, they may contribute to efficient protein expression from the downstream ORF.

Based on secondary structure predictions, the long VP30/VP24-IGR may adopt a pronounced stem-loop structure with two ‘arms’ (Fig. 5b), prompting us to delete one or both of these arms. Deletion of either arm increased the expression levels of the upstream gene (i.e., Luc), even though the deleted regions are not part of the upstream transcription unit. These findings again suggest that ZEBOV protein expression levels are regulated by complex structural interactions in the noncoding and/or nontranscribed regions. CAT expression levels were unaffected by deletion of the 5’-arm but reduced by deletion of the 3’-arm (Fig. 5b). Deletion of both arms nearly abrogated protein expression from both the upstream and downstream ORFs (Fig. 5b). Hence, for this regulatory region, juxtaposed transcription stop/start signals do not function efficiently, in contrast to our findings with the NP/VP35- and VP35/VP40-RRs (Fig. 5a). These

data suggest that sequences in the VP30/VP24-IGR, and/or cooperative effects between the IGR and the 5' and/or 3'NTRs, are critical for this regulatory region.

3.5. Significance of divergent polyadenylation and transcription start signals for ZEBOV protein expression

The ZEBOV transcription stop and start signals are highly conserved among the genes; however, minor differences do exist (Fig. 1b). The polyadenylation signal of ZEBOV genes comprises five or six uridine residues (adenosine residues in the positive-sense orientation; Fig. 1b). To assess the significance of these differences, we deleted an adenosine residue from the NP polyadenylation site (which is composed of six adenosine residues; Fig. 6) or inserted an additional adenosine nucleotide into the VP35 polyadenylation site (which comprises five adenosine residues), resulting in NP/VP35-RRΔA and VP35/VP40-RR+A, respectively. Deletion of an adenosine residue from the NP polyadenylation signal had no significant effects on Luc or CAT expression levels (Fig. 6). By contrast, insertion of an adenosine residue at the VP35 polyadenylation signal significantly increased Luc expression. Thus, in the context of this overlapping transcription stop/start signal, a longer polyadenylation signal may result in more efficient polyadenylation and/or mRNA release, resulting in higher expression levels from the upstream ORF. It is interesting to note that all three ZEBOV gene overlaps are characterized by short polyadenylation signals (Fig. 1c), which may be important to balance expression from the upstream and downstream ORFs.

The NP and L genes contain a 'G' at the third position of the transcription start signal, in contrast to all of the other genes, which are characterized by a 'T' at this position (Fig. 1b). Replacement of the 'G' with a 'T' in the L gene transcription start signal (VP24/L-RR-T) resulted in increased CAT expression levels relative to the authentic signal (Fig. 6). The 'G' residue in the transcription start signals of the NP and L genes may therefore serve to reduce the expression levels of these two proteins. The NP gene is in the first position of the genome and therefore transcribed at the highest levels; the divergent transcription start signal may thus serve to attenuate NP expression. The L gene sits in the last position of the genome and hence is transcribed at the lowest levels. Despite this fact, a weak transcription start signal may be necessary to further down-regulate L expression. On the other hand, the presence of two transcription stop signals may increase L expression levels relative to constructs that possess only one stop signal (see Fig. 4, lower panel). Tight regulation of L protein expression has been reported for other nonsegmented, negative-sense RNA viruses: VSV (New Jersey strain) and rabies virus regulate L protein expression through sequences in the intergenic regions (Finke et al., 2000; Stillman & Whitt, 1998), and human RSV regulates L protein expression through an overlapping transcription stop/start signal (Collins et al., 1987; Fearn & Collins, 1999). Thus, fine-tuning of L protein expression may be a common feature of nonsegmented, negative-sense RNA viruses, although different mechanisms may be in play.

In conclusion, our findings suggest that the expression levels of *Ebolavirus* proteins are not only determined by their position in the genome, but also by additional features, such as the sequence and composition of the transcription stop and start signals, and the sequence and/or secondary structure of the intergenic region and/or the noncoding regions. Such knowledge improves our understanding of the *Ebolavirus* life cycle.

Acknowledgments

We thank Susan Watson for editing the manuscript. This work was sponsored by National Institute of Allergy and Infectious Disease Public Health Service research grants, by grants-in-aid and by the Special Coordination Funds for Strategic Cooperation to Control Emerging and Reemerging Infections from the Ministries of Education, Culture, Sports, Science, Japan, and by the NIH/NIAID Regional Center of Excellence for Bio-defense and Emerging Infectious Diseases Research (RCE) Program. The authors wish to acknowledge membership within and support from the Region V 'Great Lakes' RCE (NIH award 1-U54-AI-057153).

REFERENCES

- Boehmann Y, Enterlein S, Randolph A, Muhlberger E. A reconstituted replication and transcription system for Ebola virus Reston and comparison with Ebola virus Zaire. *Virology* 2005;332(1):406–417. [PubMed: 15661171]
- Collins PL, Olmsted RA, Spriggs MK, Johnson PR, Buckler-White AJ. Gene overlap and site-specific attenuation of transcription of the viral polymerase L gene of human respiratory syncytial virus. *Proc Natl Acad Sci U S A* 1987;84(15):5134–5138. [PubMed: 2440043]
- Fearn R, Collins PL. Model for polymerase access to the overlapped L gene of respiratory syncytial virus. *J Virol* 1999;73(1):388–397. [PubMed: 9847343]
- Feldmann H, Klenk HD, Sanchez A. Molecular biology and evolution of filoviruses. *Arch Virol Suppl* 1993;7:81–100. [PubMed: 8219816]
- Feldmann H, Muhlberger E, Randolph A, Will C, Kiley MP, Sanchez A, Klenk HD. Marburg virus, a filovirus: messenger RNAs, gene order, and regulatory elements of the replication cycle. *Virus Res* 1992;24(1):1–19. [PubMed: 1626422]
- Finke S, Cox JH, Conzelmann KK. Differential transcription attenuation of rabies virus genes by intergenic regions: generation of recombinant viruses overexpressing the polymerase gene. *J Virol* 2000;74(16):7261–7269. [PubMed: 10906180]
- Muhlberger E, Sanchez A, Randolph A, Will C, Kiley MP, Klenk HD, Feldmann H. The nucleotide sequence of the L gene of Marburg virus, a filovirus: homologies with paramyxoviruses and rhabdoviruses. *Virology* 1992;187(2):534–547. [PubMed: 1546452]
- Muhlberger E, Trommer S, Funke C, Volchkov V, Klenk HD, Becker S. Termini of all mRNA species of Marburg virus: sequence and secondary structure. *Virology* 1996;223(2):376–380. [PubMed: 8806574]
- Muhlberger E, Weik M, Volchkov VE, Klenk HD, Becker S. Comparison of the transcription and replication strategies of marburg virus and Ebola virus by using artificial replication systems. *J Virol* 1999;73(3):2333–2342. [PubMed: 9971816]
- Neumann G, Feldmann H, Watanabe S, Lukashevich I, Kawaoka Y. Reverse genetics demonstrates that proteolytic processing of the Ebola virus glycoprotein is not essential for replication in cell culture. *J Virol* 2002;76(1):406–410. [PubMed: 11739705]
- Rose RE. The nucleotide sequence of pACYC177. *Nucleic Acids Res* 1988;16(1):356. [PubMed: 3340534]
- Sanchez, A.; Geisbert, TW.; Feldmann, H. Filoviridae: Marburg and Ebola Viruses. In: Knipe, DM.; Howley, PM.; Griffin, DE.; Lamb, RA.; Martin, MA.; Roizman, B.; Straus, SE., editors. *Fields Virology*, 5 edn. Philadelphia: Lippincott Williams & Wilkins; 2007. p. 1409–1448.
- Sanchez A, Kiley MP, Holloway BP, Auperin DD. Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. *Virus Res* 1993;29(3):215–240. [PubMed: 8237108]
- Sanchez A, Kiley MP, Holloway BP, McCormick JB, Auperin DD. The nucleoprotein gene of Ebola virus: cloning, sequencing, and in vitro expression. *Virology* 1989;170(1):81–91. [PubMed: 2718390]
- Sanchez A, Kiley MP, Klenk HD, Feldmann H. Sequence analysis of the Marburg virus nucleoprotein gene: comparison to Ebola virus and other non-segmented negative-strand RNA viruses. *J Gen Virol* 1992;73(2):347–357. [PubMed: 1538192]
- Stillman EA, Whitt MA. The length and sequence composition of vesicular stomatitis virus intergenic regions affect mRNA levels and the site of transcript initiation. *J Virol* 1998;72(7):5565–5572. [PubMed: 9621014]
- Theriault S, Groseth A, Neumann G, Kawaoka Y, Feldmann H. Rescue of Ebola virus from cDNA using heterologous support proteins. *Virus Res* 2004;106(1):43–50. [PubMed: 15522446]
- Volchkov VE, Volchkova VA, Muhlberger E, Kolesnikova LV, Weik M, Dolnik O, Klenk HD. Recovery of infectious Ebola virus from complementary DNA: RNA editing of the GP gene and viral cytotoxicity. *Science* 2001;291(5510):1965–1969. [PubMed: 11239157]

- Watanabe S, Watanabe T, Noda T, Takada A, Feldmann H, Jasenosky LD, Kawaoka Y. Production of novel ebola virus-like particles from cDNAs: an alternative to ebola virus generation by reverse genetics. *J Virol* 2004;78(2):999–1005. [PubMed: 14694131]
- Weik M, Enterlein S, Schlenz K, Muhlberger E. The Ebola virus genomic replication promoter is bipartite and follows the rule of six. *J Virol* 2005;79(16):10660–10671. [PubMed: 16051858]
- Weik M, Modrof J, Klenk HD, Becker S, Muhlberger E. Ebola Virus VP30-Mediated Transcription Is Regulated by RNA Secondary Structure Formation. *J Virol* 2002;76(17):8532–8539. [PubMed: 12163572]

(b)

Gene	Transcription START signal	Transcription STOP signal
NP	GAGGAAGATTAA <u> </u>	ATTAAGAAAAAA
VP35	GATGAAGATTAA	ATTAAGAAAAAA
VP40	GATGAAGATTAA	ATTAAGAAAAAA
GP	GATGAAGATTAA	ATTAAGAAAAAA
VP30	GATGAAGATTAA	ATTAAGAAAAAA
VP24	GATGAAGATTAA	ATTAAGAAAAAA
L	GAGGAAGATTAA <u> </u>	ATTAT .AAAAAA

(c)

Gene Border	Arrangement of Transcription Stop and Start Signals
NP/VP35	<u>ATTAAGAAAAA</u> CTAAT <u>GATGAAGATTAA</u>
VP35/VP40	<u>GATGAAGATTAA</u> <u>GAAAAA</u>
VP40/GP	<u>ATTAAGAAAAA</u> TCGGA <u>GATGAAGATTAA</u>
GP/VP30	<u>GATGAAGATTAA</u> <u>GAAAAA</u>
VP30/VP24	<u>ATTAAGAAAAA</u> 144nts <u>GATGAAGATTAA</u>
VP24/L	<u>ATTAAGAAAAA</u> GCCT <u>GAGGAAGATTAA</u> <u>GAAAAA</u>

Figure 1.

Ebolavirus genome organization. (a) *Ebolavirus* genome organization in the positive-sense orientation. The open reading frames (ORFs) are flanked by leader (*le*) and trailer (*tr*) sequences and are separated by regulatory regions (RRs). The RRs are composed of 3' nontranslated region (NTRs), highly conserved transcription stop and start signals, and 5'NTRs. The transcription stop and start signals overlap or are separated by intergenic regions (IGRs). RRs with non-overlapping or overlapping transcription start/stop signals are shown in diagonal bars or brick bars, respectively. Transcription start and stop signals are indicated by lines above or below the nucleotide sequence, respectively. (b) Comparison of transcription stop and start signals. Nucleotides that deviate from the consensus sequences are underlined (for transcription start signals) or non-underlined (for transcription stop signals). (c) Arrangements of transcription stop and start signals for the *Ebolavirus* gene borders. For the NP/VP35, VP40/GP, VP30/VP24, and VP24/L regulatory regions, the transcription stop and start signals are separated by IGRs. By contrast, the VP35/VP40 and GP/VP30 transcription stop and start signals overlap. The VP24/L regulatory region contains two transcription stop signals, the second of which overlaps with the transcription start signal. Transcription start and stop signals are indicated by lines above or below the nucleotide sequence. All sequences shown are in the positive-sense orientation.

Minireplicon Construct						Relative Activity (%)	
						Luc	CAT
<i>le</i>	Luc	NP/VP35-RR	CAT	<i>tr</i>		60.9±3.6	98.6±1.2
<i>le</i>	Luc	VP35/VP40-RR	CAT	<i>tr</i>		84.8±8.2	99.7±0.9
<i>le</i>	Luc	VP40/GP-RR	CAT	<i>tr</i>		100±0	100±0
<i>le</i>	Luc	GP/VP30-RR	CAT	<i>tr</i>		82.9±4.7	77.0±9.7
<i>le</i>	Luc	VP30/VP24-RR	CAT	<i>tr</i>		47.6±3.8	101.6±0.8
<i>le</i>	Luc	VP24/L-RR	CAT	<i>tr</i>		83.4±2.1	87.3±6.1

Figure 2.

Effects of wild-type regulatory regions on *Ebolavirus* protein expression. Minireplicons were constructed that contain all six wild-type *Ebolavirus* RRs. Shown are the relative Luc and CAT expression levels ($\pm$ standard deviation) relative to those of the VP40/GP-RR, which was arbitrarily set as 100%.

	Minireplicon Construct	Relative Activity (%)	
		Luc	CAT
	<div> <div>STOP</div> <div>START</div> </div>		
NP/VP35-RR	<div> <div>le</div> <div>Luc</div> <div>3'NTR</div> <div>IGR</div> <div>5'NTR</div> <div>CAT</div> <div>tr</div> </div>	100±0	100±0
NP/VP35-RRΔ3'NTR	<div> <div>le</div> <div>Luc</div> <div>Δ3'NTR</div> <div>IGR</div> <div>5'NTR</div> <div>CAT</div> <div>tr</div> </div>	185.1±3.0	109.1±1.6
NP/VP35-RRΔ5'NTR	<div> <div>le</div> <div>Luc</div> <div>3'NTR</div> <div>IGR</div> <div>Δ5'NTR</div> <div>CAT</div> <div>tr</div> </div>	147.3±6.7	41.6±25.4
NP/VP35-RRΔNTR	<div> <div>le</div> <div>Luc</div> <div>Δ3'NTR</div> <div>IGR</div> <div>Δ5'NTR</div> <div>CAT</div> <div>tr</div> </div>	286.8±21.4	67.0±31.5
	<div> <div>START/STOP</div> </div>		
VP35/VP40-RR	<div> <div>le</div> <div>Luc</div> <div>3'NTR</div> <div>5'NTR</div> <div>CAT</div> <div>tr</div> </div>	100±0	100±0
VP35/VP40-RRΔNTR	<div> <div>le</div> <div>Luc</div> <div>Δ3'NTR</div> <div>Δ5'NTR</div> <div>CAT</div> <div>tr</div> </div>	170.9±2.2	14.2±8.5

Figure 3.

Effects of nontranslated regions on *Ebolavirus* protein expression. Based on the NP/VP35 regulatory region, the 5'NTR, 3'NTR, or both NTRs were deleted and the relative levels of Luc and CAT expression determined. Both NTRs were also deleted from the VP35/VP40 regulatory region. Deleted regions are shown in gray. The position of the transcription stop and start signals is indicated by bars and triangles, respectively.

	Minireplicon Construct	Relative Activity (%)	
		Luc	CAT
VP35/VP40-RR	<u>GATGAAGATTAAGAAAAA</u>	100±0	100±0
VP35/VP40-RR – STOP-START	<u>ATTAAGAAAAAGATGAAGATTAA</u>	128.2±0.8	84.2±11.9
VP35/VP40-RR – START-STOP	<u>GATGAAGATTAAATTAAGAAAAA</u>	116.1±11.7	103.5±1.7
VP24/L-RR	<u>ATTAAGAAAAAA</u> GCCT <u>GAGGAAGATTAAGAAAAA</u>	100±0	100±0
VP24/L-RRΔSTOP#1	<u>ATTAA</u> <u>CTCTCTC</u> GCCT <u>GAGGAAGATTAAGAAAAA</u>	132.1±6.0	37.0±3.4
VP24/L-RRΔSTOP#2	<u>ATTAAGAAAAAA</u> GCCT <u>GAGGAAGATTAA</u> <u>CTCTCT</u>	114.2±10.9	75.4±1.4

Figure 4.

Effects of overlapping transcription stop/start signals on *Ebolavirus* protein expression. The overlapping signals in the VP35/VP40-RR were separated by placing the stop or start signal first. The VP24/L-RR contains two transcription stop signals, the second of which overlaps with the transcription start signal. To assess the significance of the dual transcription stop signals, we disabled each of the two polyadenylation signals. Altered nucleotides are boxed. Transcription start and stop signals are indicated by lines above or below the nucleotide sequence.

(a)	Minireplicon Construct		Relative Activity (%)	
			Luc	CAT
NP/VP35-RR	<u>ATTAAGAAAAAA</u>	CTAAT <u>GATGAAGATTAA</u>	100±0	100±0
NP/VP35-RRΔIGR	<u>ATTAAGAAAAAA</u>	..Δ.. <u>GATGAAGATTAA</u>	106.2±10.9	82.2±17.5
VP35/VP40-RR	<u>GATGAAGATTAA</u>	<u>GAAAAA</u>	100±0	100±0
VP35/VP40-RR – STOP-START	<u>ATTAAGAAAAA</u>	<u>GATGAATATTAA</u>	128.2±0.8	84.2±11.9

(b)

	5'	3'	Relative Activity (%)	
			Luc	CAT
VP30/VP24-RR	<u>ATTAAGAAAAA</u>	CTG...TGGTAT...TAA <u>GATGAAGATTAA</u>	100±0	100±0
VP30/VP24-RRΔ5'IGR	<u>ATTAAGAAAAA</u>	Δ ₁₋₅₀ TAT...TAA <u>GATGAAGATTAA</u>	121.8±9.8	100.8±0.9
VP30/VP24-RRΔ3'IGR	<u>ATTAAGAAAAA</u>	CTG...TGG Δ ₅₁₋₁₄₄ <u>GATGAAGATTAA</u>	129.2±2.8	55.5±10.4
VP30/VP24-RRΔIGR	<u>ATTAAGAAAAA</u>	Δ ₁₋₁₄₄ <u>GATGAAGATTAA</u>	5.0±2.3	3.3±1.3

Figure 5.

Effects of intergenic regions (IGRs) on *Ebolavirus* protein expression. (a) Significance of the NP/VP35 IGR. To assess the significance of the NP/VP35 IGR for the regulation of protein expression levels, the entire IGR was deleted, resulting in juxtaposed transcription stop and start signals. For comparison, another construct with juxtaposed transcription stop and start signals is shown (VP35/VP40-RR-STOP-START). This construct resulted from the separation of the overlapping transcription stop and start signals in the VP35/VP40-RR (see Fig. 5). (b) Significance of the VP30/VP24 IGR. The VP30/VP24-RR contains a long IGR with a predicted ‘two-armed’ stem-loop structure. To assess the significance of this structural feature for *Ebolavirus* protein expression levels, one or both arms were deleted.

	Minireplicon Construct	Relative Activity (%)	
		Luc	CAT
NP/VP35-RR	<u>ATTAAGAAAAAA</u> CTAAT <u>GATGAAGATTAA</u>	100±0	100±0
NP/VP35-RRΔA	<u>ATTAAGAAAAA</u> A CTAAT <u>GATGAAGATTAA</u>	104.6±8.4	102.4±2.1
VP35/VP40-RR	<u>GATGAAGATTAAGAAAAA</u>	100±0	100±0
VP35/VP40-RR+A	<u>GATGAAGATTAAGAAAAA</u> A	147.7±7.4	109.8±1.7
VP24/L-RR	<u>ATTAAGAAAAAA</u> GCCT <u>GAGGAAGATTAAGAAAAA</u>	100±0	100±0
VP24/L-RR-T	<u>ATTAAGAAAAAA</u> GCCT <u>GAT</u> T <u>GAAGATTAAGAAAAA</u>	124.8±3.0	144.9±9.7

Figure 6.

Effects of mutations in the transcription stop or start signals on *Ebolavirus* protein expression levels. To assess the significance of the length of the polyadenylation signal (i.e., the adenosine stretch in the positive-sense orientation), we shortened the polyadenylation signal in the NP/VP35-RR, or elongated the signal in the VP35/VP40-RR. The VP24/L-RR deviates at one position from the consensus transcription start signal. To assess the significance of this atypical transcription start signal, it was converted to the consensus sequence. Altered nucleotides are boxed. Transcription start and stop signals are indicated by lines above or below the nucleotide sequence.