Virology 536 (2019) 91–100

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

Priming and realignment by the influenza a virus RdRp is dependent on the length of the host primers and the extent of base pairing to viral RNA

Corey De Vlugt, Dorota Sikora, Lynda Rocheleau, Martin Pelchat*

Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, K1H 8M5, Canada

ARTICLE INFO

Keywords:
Influenza A virus
Viral transcription
Cap-snatching
Capped host primer profiling
Prime-and-realign mechanism

ABSTRACT

Initiation of influenza A virus (IAV) transcription depends on RNA primers derived from host RNAs. During this process, some primers are elongated by a few nucleotides, realigned on the viral RNA templates (vRNA), and then used to initiate another round of transcription. Here, we used information on the host primers used by four IAV strains and four mini-replicons to investigate the characteristics of primer undergoing priming and realignment. We report that primers are biased towards this mechanism on the basis of length and RNA duplex stability with the vRNA templates. Priming and realignment results in primers three nucleotides longer, ending in a nucleotide sequence able to base pair with the 3′ end of the vRNA template. By acting on primers based on length and sequence compatibility with the 3' end of the vRNA, priming and realignment rescues suboptimal primers, converting them into ones that can efficiently initiate transcription.

1. Introduction

The influenza A virus (IAV) is a negative sense single-stranded RNA virus of the family *Orthomyxoviridae*, and causes significant worldwide mortality and morbidity. The IAV genome consists of eight RNA segments (vRNAs), each of which encodes at least one protein. After infection, the viral genome segments are transported into the nucleus, where both viral replication and transcription occur. IAV mRNAs are produced by the viral RNA-dependent RNA polymerase (RdRp) and possess a 5′ cap structure and a poly(A) tail (Caton and Robertson, 1980; Mark et al., 1979; Scholtissek and Rott, 1970; Taylor et al., 1977). The polyadenylation of viral mRNA is accomplished through iterative transcription of a polyuridine tract near the 5′ end of the vRNA template by RdRp stuttering (Poon et al., 1999). To initiate viral transcription, IAV uses a process called "cap-snatching", where the RdRp interacts with the C-terminal domain of host RNA polymerase II (RNAP II), cleaves host RNAs 10–15 nucleotides downstream of their 5' cap structures, and uses them to initiate transcription of the RNA genome segments. This results in chimeric mRNAs with host-derived leader sequences followed by the viral mRNA sequences (Fig. 1; right path) (De Vlugt et al., 2018; Plotch et al., 1981; Te Velthuis and Fodor, 2016).

The IAV RdRp is composed of three subunits: polymerase basic protein 1 (PB1), polymerase basic protein 2 (PB2), and polymerase acidic protein (PA). PB1 contains the functional motifs associated with transcription and is the central subunit of the heterotrimer to which the other subunits connect (Biswas and Nayak, 1994; Braam et al., 1983; Pflug et al., 2014). PB2 binds to the 5′ m^7G cap structure of host RNA (Guilligay et al., 2008; Pflug et al., 2018). PA contains the endonuclease region responsible for cleavage of PB2-bound host RNA (Dias et al., 2009). In addition, the RdRp binds to the 3′ and 5' conserved sequences of each IAV segment (Flick and Hobom, 1999), with the remainder of the vRNA being bound to the viral nucleoprotein (NP)(Labaronne et al., 2016; Ng et al., 2008; Ye et al., 2006). Due to extensive interactions, the functions of each individual subunit are dependent on the other subunits, and on the binding of vRNA segments (Lee et al., 2003; Pflug et al., 2018, 2014; Reich et al., 2017; Thierry et al., 2016).

The RdRp predominantly cap-snatches host RNA based on abundance (Sikora et al., 2017, 2014), and it was suggested that primers with multiple bases complementary to the vRNA templates are preferentially used for transcription (Geerts-Dimitriadou et al., 2011a, 2011b; Te Velthuis and Oymans, 2018). The RdRp cleaves at a phosphodiester bond 10–15 nt downstream of the 5′ cap structure of host transcripts after a purine residue (Beaton and Krug, 1981; Krug et al., 1980; Plotch et al., 1981; Rao et al., 2003). Using purified components, cleavage by the N-terminal region of PA occurs at the 3′ phosphodiester bond of a guanine (G) residue (Datta et al., 2013). A-terminated capped fragments can base pair with the U residue located at the 3′ extremity of the conserved 3′-UCG … -5′ of the vRNA template (3′U1) and be used to initiate RNA synthesis (Beaton and Krug, 1981; Plotch et al., 1981; Rao et al., 2003; Shaw and Lamb, 1984). Transcription can also initiate with

* Corresponding author.
E-mail address: mpelchat@uottawa.ca (M. Pelchat).

https://doi.org/10.1016/j.virol.2019.08.002
Received 21 June 2019; Received in revised form 1 August 2019; Accepted 1 August 2019
Available online 02 August 2019
0042-6822/ © 2019 Elsevier Inc. All rights reserved.

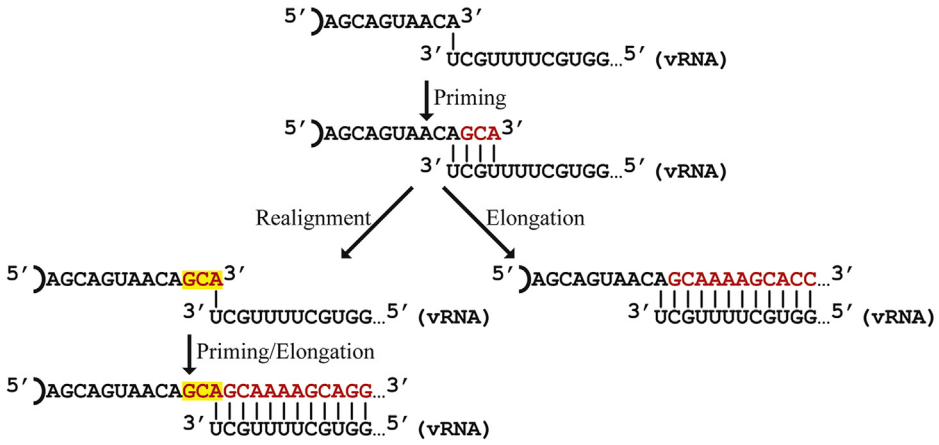


Fig. 1. Priming and transcription of the Influenza A Virus. Host-derived cap-snatched primers are used to prime transcription at the 3′ end of the vRNA. Following the elongation by a few nucleotides (red), primers either undergo PAR (left path) or continue elongation (right path). When PAR occurs, the elongated primers realign to the start of the viral RNA templates (vRNA), and are then used to initiate another round of elongation. The yellow background indicates realigned nucleotides complementary to the 3′ end of the vRNA.

G-terminated primers directed at 3′C2 rather than 3′U1 (Datta et al., 2013; Geerts-Dimitriadou et al., 2011b, 2011a; Sikora et al., 2014; Te Velthuis and Oymans, 2018). *In vitro*, IAV transcription can also be primed by the dinucleotide AG (Deng et al., 2006; McGeoch and Kitron, 1975; Pritlove et al., 1995; Zhang et al., 2010).

Nucleotide sequence repeats complementary to the 3′ end of the vRNA conserved sequence, not present in the cap-snatched primer, have been observed between the 3′ most nucleotide of the host-derived sequence and the viral conserved sequences (Geerts-Dimitriadou et al., 2011a, 2011b; Koppstein et al., 2015; Sikora et al., 2017, 2014). These insertions arise through an iteratively aborted transcription attempt of the first few nucleotides of the viral conserved sequences (Geerts-Dimitriadou et al., 2011b, 2011a; Te Velthuis and Oymans, 2018). In this case, the RdRp transcribes a few nucleotides, realigns the elongated primer to the start of the vRNA, and re-initiates transcription (Fig. 1; left path). This mechanism is called “prime-and-realign” (PAR). PAR is not limited to a single round, and some host-derived sequences have multiple sequence insertions from multiple rounds of PAR. Using high-throughput sequencing and extensive data processing to trim all potential repetitive sequences at the 3’ end of the primers, it was observed that GCA addition were the most common sequence added during PAR on A/WSN/1933 (H1N1) transcripts (Koppstein et al., 2015). Recently, *in vitro* transcription assays using purified recombinant RdRp subunits and selected RNA primers suggested that PAR operates predominantly on short primers or on primers without extensive base complementarity to the vRNA templates (Te Velthuis and Oymans, 2018). However, the determinants of PAR *in vivo*, in infected cells, remain to be elucidated.

In this study, we investigated the population of cap-snatched primers undergoing PAR by the IAV RdRp and used by IAV transcription in infected cells to quantify PAR addition frequencies, to define the characteristics of primers undergoing PAR, and how PAR is affected by the vRNA templates. To this end, we analyzed previously obtained datasets of host-derived sequences located at the 5′ end of viral mRNAs from human lung epithelial (A549) cells infected by A/Puerto Rico/8/1934 (H1N1), A/Hong Kong/1/1968 (H3N2), A/Brisbane/59/2007 (H1N1), and A/WSN/1933 (H1N1)(Gu et al., 2015; Koppstein et al., 2015; Sikora et al., 2017, 2014). To complement these datasets using a non-infectious model, we also profiled the host primers used by four different IAV mini-replicons.

2. Results

2.1. PAR frequencies are conserved among various IAV strains

To investigated the population of the primers undergoing PAR by the IAV RdRp, we extracted the cap-snatched primers using a conservative strategy we previously reported (Sikora et al., 2017). Briefly, reads corresponding to each IAV mRNA were extracted using gene-specific or 5′ UTR sequences. The heterogeneous sequences located at the 5′ end of the IAV mRNAs were obtained by extracting the sequences just upstream of the conserved 5′-GC[A/G]AAAGCAGG-3′ viral motifs, and downstream of either RACE primers (for A/Puerto Rico/8/1934 (H1N1), A/Hong Kong/1/1968 (H3N2) and A/Brisbane/59/2007 (H1N1)) or stretches of at least three consecutive guanine residues (for A/WSN/1933 (H1N1)) (Gu et al., 2015; Koppstein et al., 2015; Sikora et al., 2017, 2014). The extracted sequences were then mapped to windows located from −100 nt to +100 nt around the annotated transcription start sites (TSS) indexed in the human reference genome GRCh38, and the solution closest to the TSS selected. If no match was found, the longest single nucleotide sequence which could be added by PAR (i.e. G, GC, GCA, GCG, GCAA, GCGA, GCAAA, GCGAA, GCAAAA, or GCGAAA) was iteratively trimmed from the 3′ end of the host-derived sequence. The primers able to be trimmed were then remapped to windows around the TSS, as above. The cap-snatched primers were classified into two distinct groups named “NoPAR” or “PAR”, corresponding to cap-snatched primers able to be mapped directly or with at least one round of trimming to the annotated TSS, respectively.

Most cap-snatched primers (85.7 ± 0.6%) do not undergo PAR, indicating that the PAR process is not used during the transcription of the majority of IAV mRNAs (Fig. 2a). The frequency of PAR is similar in all four strains at approximately 14.3 ± 0.6% of all cap-snatched primers (Fig. 2b, S1a-S4a). This indicates that the key determinants of PAR are shared in these strains. Some cap-snatched primers undergo multiple rounds of PAR prior to productive transcription. In our datasets, a maximum of three or four rounds of PAR were observed for most strains. A maximum of nine rounds of PAR was observed in the primers used by A/WSN/1933 (H1N1), which might be due to the high sequencing depth of this library. Of the primers processed by PAR, 95.0 ± 2.9% undergo only one round of PAR (i.e. 13.6% of the 14.3% processed by PAR; Fig. 2c, S1b-S4b). The number of primers processed by PAR decreases exponentially with each round, at around 95% per round, suggesting that this process is used to salvage most of the potentially suboptimal primers at each round and consequently, decreases the probability of a subsequent round of PAR.

2.2. PAR frequencies correlate inversely with the length of the cap-snatched primer

Length was shown to be a key determinant of PAR for other (−) ssRNA viruses that use a cap-snatched primer to prime transcription, with shorter primers more likely to undergo PAR prior to transcription (Liu et al., 2018, 2016; Yao et al., 2012). To determine if the subpopulation of primers which undergo PAR is of a shorter length, we calculated the percentage of primers of a given length in both PAR and NoPAR subpopulations. Also, for the PAR subpopulation, the length of primers before PAR (i.e. Pre-PAR) and after the final round of PAR

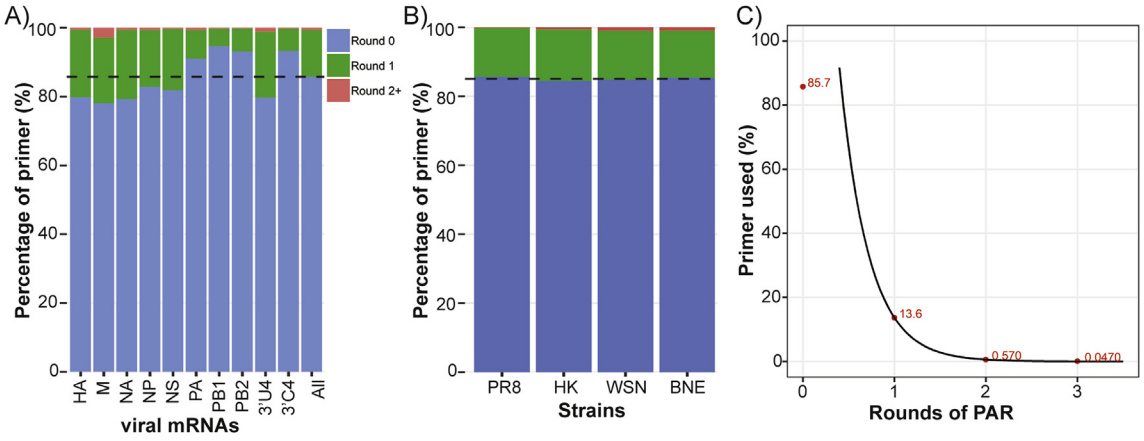


Fig. 2. Relative PAR frequencies of IAV mRNAs from various IAV strains and number of primers processed for each round. (a) Percentage of host-derived primers used directly (Round 0) or that are processed by PAR (Round 1 and above) for each vRNA, vRNA conserved sequence type (3′U4 or 3′C4), and on average (All). Only primers which were matched to within 100 nucleotides of an annotated transcription start site were included in this percentage. Values represent the average of the A/Puerto Rico/8/1934 (H1N1), A/Hong Kong/1/1968 (H3N2), A/Brisbane/59/2007 (H1N1), and A/WSN/1933 (H1N1) strains. The black dashed line indicates the average percentage of primers that do not undergo PAR (85.7%). (b) As in (a) with primers grouped by strain. A/Puerto Rico/8/1934 (H1N1), A/Hong Kong/1/1968 (H3N2), A/WSN/1933 (H1N1) and A/Brisbane/59/2007 (H1N1) are represented by PR8, HK, WSN and BNE, respectively. (c) Percentage of primers used directly for transcription elongation (Round 0) and processed at each round of PAR (Rounds 1 to 3). Black line indicates log-linear trend (R^2 = 0.995). Values in red represent the average of the four strains analyzed.

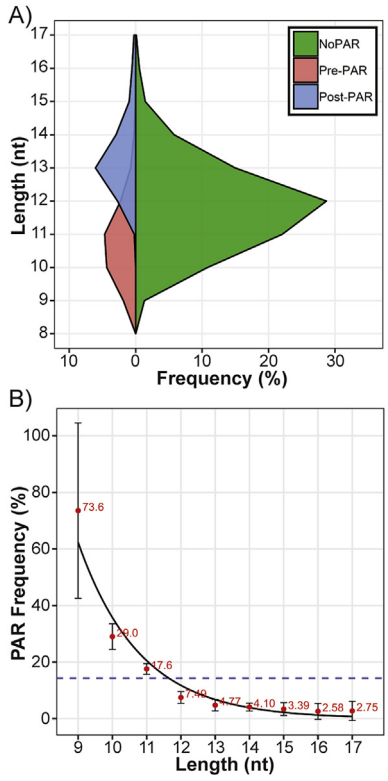


Fig. 3. PAR frequency is dependent on primer length. (a) Length distributions of primers versus the proportion of primers used directly or processed by PAR. (b) PAR frequency inversely correlates with the length of the cap-snatched primers. Values in red and the error bars represent the average of all 4 strains and the corresponding standard deviations, respectively. Black line indicates log-linear trend (R^2 = 0.888 for lengths between 9 and 17). The blue dashed line indicates the average PAR frequency for the four strains (14.3%). Values represent the average of the A/Puerto Rico/8/1934 (H1N1), A/Hong Kong/1/1968 (H3N2), A/Brisbane/59/2007 (H1N1), and A/WSN/1933 (H1N1) strains.

(Post-PAR) was extracted. Our results indicate that the Pre-PAR primers are shorter than those in the NoPAR subpopulation. PAR lengthens these primers by approximately two to three nucleotides (Fig. 3a, S1c-S4c). Additionally, PAR rates decrease exponentially with length (Fig. 3b, S1d-S4d) and PAR rates drop below the strain averages (Fig. 3b, blue dashed line) when primer length reaches 12 nucleotides.

2.3. PAR frequencies correlate inversely with the stability of the RNA duplex

The 3′ end of the IAV segments possesses a conserved region of 12 nucleotides, with a variable pyrimidine at the fourth position: 3′-UCG[U/C]UUUCGUCC-5′. Our results indicate that mRNAs transcribed from 3′C4 templates undergo less PAR than 3′U4 templates (Fig. 2a, S1a-S4a; 3′C4 vs 3′U4), suggesting that RNA duplex stability with the vRNA template affects PAR frequency. To examine the relationship between PAR and RNA duplex stability, we calculated the maximum free energies of the primers used for IAV transcription by taking into account both hydrogen bond formation and stacking interactions (Turner and Mathews, 2010). Our results show an inverse correlation between duplex stability and PAR frequency, mostly due to the identity of the nucleotide annealing with 3′U1 (Fig. 4a, S1e-S4e; in pink), indicating that as the RNA duplex stability increases (i.e. free energy decreases) the PAR frequency decreases.

It was previously reported that the IAV RdRp may prime terminally at 3′U1 with any nucleotide or internally at 3′C2 if the primer ends in a guanine residue (Datta et al., 2013; Reich et al., 2017; Te Velthuis and Oymans, 2018). We repeated the above analysis factoring in either terminal (i.e. 3′U1) or internal (i.e. 3′C2) initiation (Fig. 4a, S1e-S4e; in orange and green). Although the PAR frequencies remain correlated with the free energy of the 3′U1-directed nucleotide, internal initiation at 3′C2 is associated with lower PAR frequencies than priming terminally at 3′U1. Using a linear regression model, priming internally with a complimentary guanine residue results in a reduction of PAR frequency corresponding to an average free energy reduction of 0.13 kcal/mol.

Regarding the contribution of the nucleotide at the fourth position of the vRNA template, we calculated a difference in a maximum free energy of 0.62 kcal/mol between a GCA/CGUU duplex and a GCG/CGCU duplex, due to the additional hydrogen bond on the 3′C4 templates. To determine if PAR rate is affected by this difference in template and free energy, we extracted PAR frequencies for each IAV gene of the four strains, and grouped them based on the nucleotide at 3′Y4. Variability exists in PAR frequencies between the IAV segments

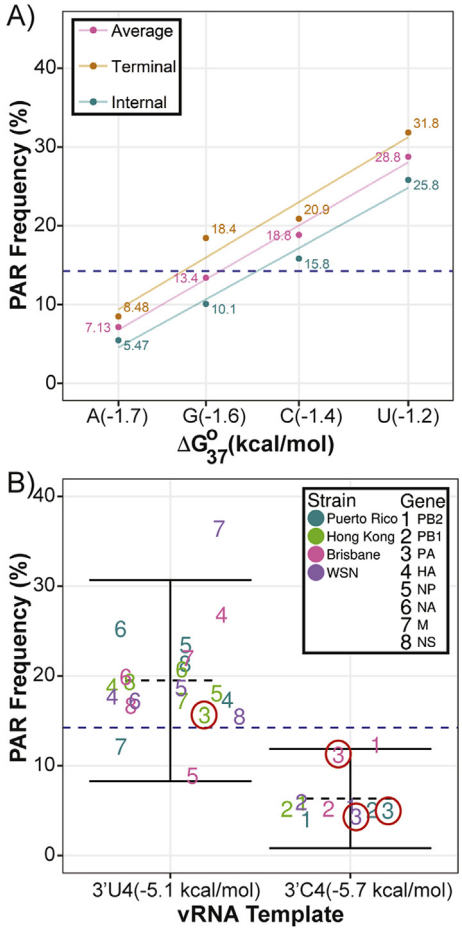


Fig. 4. PAR frequency is dependent on the stability of the RNA duplex between the cap-snatched primer and vRNA. (a) Relationship between PAR frequency and maximum free energies of the RNA duplex between the cap-snatched primer and vRNA. Free energy was calculated using the Turner energy rules for each of the nucleotide annealing with 3′U1. Values represent the average of the A/Puerto Rico/8/1934 (H1N1), A/Hong Kong/1/1968 (H3N2), A/Brisbane/59/2007 (H1N1), and A/WSN/1933 (H1N1) strains. The pink line represents average values (R^2 = 0.992), the orange line represents sequences which primed transcription terminally at 3′U1 (R^2 = 0.958) and the cyan line represents sequences which primed transcription internally at 3′C[2] (R2 = 0.983). Difference between Internal and Terminal groups is about 0.13 kcal/mol (p = 0.0149). (b) The 3′C4 vRNA templates are associated with lower PAR frequency. Gene segments are grouped based on the 3'[U/C]4 variation (p-value = 6.72×10^{-10} between templates). The black dashed line indicates the average PAR frequency for the templates; black bars indicate the 95% confidence interval. Red circles highlight PAR frequency of primers used for transcription of segment 3 (PA); 3′U4 for A/Hong Kong/1/1968 (H3N2) and 3′U4 for the A/Puerto Rico/8/1934 (H1N1), A/Brisbane/59/2007 (H1N1), and A/WSN/1933 (H1N1). The blue dashed line indicates the average PAR frequency for the four strains (14.3%).

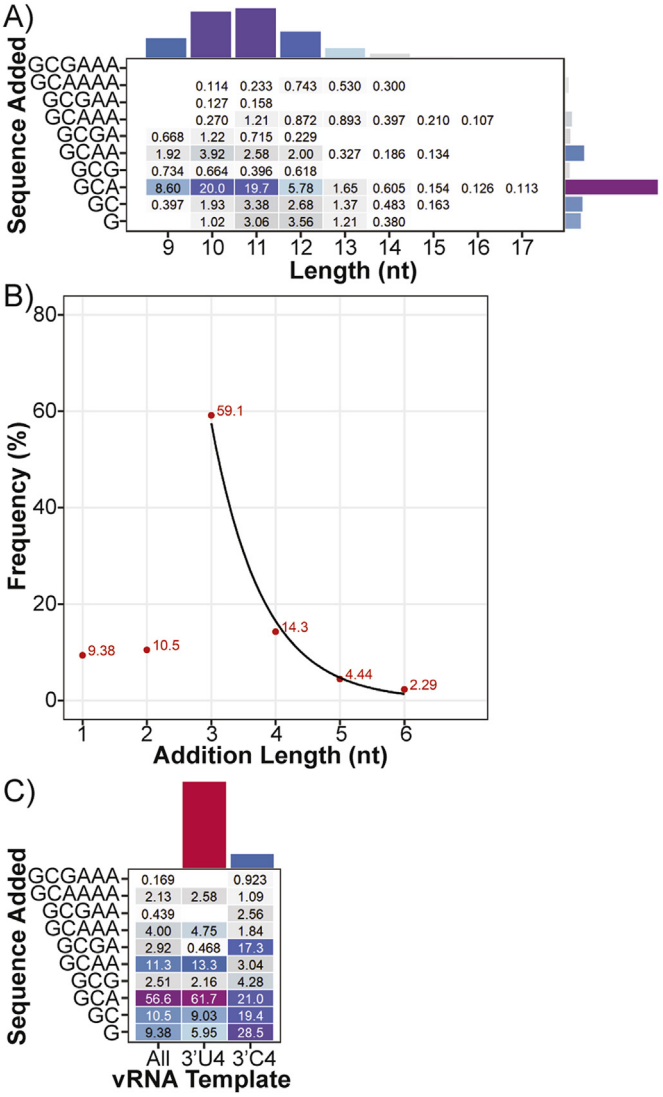


Fig. 5. Nucleotides added during PAR. (a) Nucleotide sequence added by PAR in relation to the length of the primer. The data is represented as a heatmap showing the percentage of primers which underwent PAR, grouped by length and nucleotide sequence added. Vertical bars at the top show the frequency of lengthening events from the indicated primer length; horizontal bars at the right show the frequency of lengthening events with the specific nucleotide sequence addition. (b) Frequency of nucleotide sequences of a given length added by PAR. Log-linear model is calculated between lengths of 3 (GC[A/G]) and six (GC[A/G]AAA) nucleotides (R^2 = 0.976). (c) Nucleotide sequence added by PAR in relation to the 3'[U/C]4 variation present on the vRNA templates. Vertical bars at the top show the frequency of lengthening events for template variants. Values represent the average of the A/Puerto Rico/8/1934 (H1N1), A/Hong Kong/1/1968 (H3N2), A/Brisbane/59/2007 (H1N1), and A/WSN/1933 (H1N1) strains.

transcribed (Fig. 4b, S1f-S4f). However, mRNAs transcribed from gene segments with the 3′C4 template (segments 1 (PB2), 2 (PB1), and 3 (PA) for all strains except A/Hong Kong/1/1968 (H3N2)) undergo PAR less frequently than those transcribed from 3′U4 templates. Significantly, mRNAs transcribed from segment 3 (PA) of A/Hong Kong/1/1968 (H3N2), which contains a 3′U4, undergo PAR at a frequency similar to the other 3′U4 genes, while mRNAs transcribed from the PA segments of the other three strains (3′C4 templates) undergo PAR at a frequency comparable to the other 3′C4-templated genes (Fig. 4b; red circles). Taken together, these results indicate that RNA duplex stability with the vRNA templates affects PAR frequency, and that higher PAR frequencies are associated with lower duplex stability.

2.4. PAR typically adds three nucleotides to the primers

Since both the length of the primer and the vRNA template transcribed are important factors, we analyzed the nucleotide sequence added in relation to both primer length and vRNA template transcribed. The length of the primer does not significantly affect the nucleotide sequence addition, as a GCA is predominantly added irrespective of the initial length (Fig. 5a, S1h-S4h). This addition of three nucleotides (i.e. GCA) is consistent with the length difference between the Pre-PAR and Post-PAR subpopulations illustrated in Fig. 3a. After the addition of three nucleotides, the frequency decreases exponentially (Fig. 5b, S1g-S4g). However, the vRNA template transcribed affects the nucleotide

sequence added by PAR. PAR events on the 3′U4 templates are predominantly associated with GCA additions, whereas on the 3′C4 templates, PAR also frequently occurs following a G or GC addition (Fig. 5c, S1i-S4i). This difference in identities of nucleotide additions on 3′C4 templates is likely due to the lower PAR events due to the stabilisation of the RNA duplex (vertical bars at the top on Fig. 5c, S1h-S4h).

2.5. PAR activity is independent of viral infection

It is possible that PAR activity and the properties of host primers undergoing PAR could be affected by the IAV infection and/or the production of viral proteins unrelated to cap-snatching and transcription. To account for this possibility, and to investigate PAR in a non-infectious model, cap-snatching was profiled from an IAV mini-gene system. Human lung epithelial cells (A549) cells were co-transfected with plasmids expressing the A/Puerto Rico/8/1934 (H1N1) RdRp subunits (i.e. PA, PB1, PB2, and NP) and a mini-gene construct expressing a negative-sense RNA transcript comprising the Firefly luciferase ORF flanked by various UTRs from A/Puerto Rico/8/1934 (H1N1). Then, mRNA was extracted and host primers profiled by high-throughput sequencing, as above. We tested four A/Puerto Rico/8/1934 (H1N1) UTR combinations: $^{5'}$NP-Luc-NP$^{3'}$ (NPhomo), $^{5'}$PB1-Luc-PB1$^{3'}$ (PB1homo), $^{5'}$PB1-Luc-NP$^{3'}$ (NPhetero), and $^{5'}$NP-Luc-PB1$^{3'}$ (PB1hetero). Overall, 31.4 million sequences corresponding to primers used by the luciferase mini-replicons were identified (Fig. 6a, pie chart). Analysis of the length distribution of the host primers indicates that 99.2% are from 9 to 14 nt long, with main peaks at 11–12 nt (Fig. 6a), consistent with the reported length for the primers cap-snatched by the four IAV strains analyzed (Sikora et al., 2017).

Using the same analyses as above, our results indicate that approximately 11.8% of all cap-snatched primers used to transcribe the luciferase mini-replicons undergo PAR (Fig. 6b), that 95.7% of those processed by PAR only use one round of PAR (11.3% of the 11.8% processed by PAR), and that the amount of primers processed by PAR decreases exponentially with each round (Fig. 6c). As observed during the infection by the IAV strains, the length of the Pre-PAR primers is two to three nucleotides shorter than the NoPAR subpopulation (Fig. 6d), and PAR frequencies correlate inversely with the length of the cap-snatched primer (Fig. 6e). Similarly, there is an inverse correlation between duplex stability and PAR frequency (Fig. 6f). Internal initiation at 3′C2 and transcription from 3′C4 templates is associated with lower PAR frequencies (Fig. 6b, f and 6g). PAR on the mini-replicon templates also predominantly adds three nucleotides, and the frequency for longer additions decreases exponentially (Fig. 6h). Finally, as observed for the four IAV strains, PAR events on the 3′U4 templates (i.e. NPhomo and NPhetero) are associated with GCA additions, whereas PAR events on the 3′C4 templates (i.e. PB1homo and PB1hetero) also frequently occur following a G or GC addition (Fig. 5c). Because these results are similar to those observed in IAV-infected cells, PAR appears to be independent of viral infection, and is likely related to the nature of the cap-snatched primers and to their interactions with the vRNA and the RdRp.

2.6. PAR processes primers on the basis of length and RNA duplex stability with the vRNA templates

Given the strong association of PAR with both the length of the primer and the stability of the RNA duplex, we then examined the effects of these variables together on this process by extracting PAR frequencies for all primers on the basis of length and sequence end. Because the identity of the nucleotide at the fourth position of the vRNA template affects PAR frequencies, we also examined PAR separately on the 3′U4 and 3′C4 templates (Fig. 7a-c).

Our analyses indicate that both the variable pyrimidine at the fourth position of the vRNAs and the nucleotide used to prime transcription change the impact of primer length on PAR. Lower PAR rates are associated with longer primers and with the 3′C4 templates (vs. the 3′U4 templates). Overall, a primer length of at least 12 nucleotides is processed less by PAR, except when priming with a uridine. At a length of 12 nucleotides or more, the ability to prime internally is associated with a reduction in PAR rates in comparison to priming with a terminally-directed nucleotide. A similar model was generated using data derived from the IAV mini-replicon system (Fig. 7d-f), indicating that PAR is independent of viral infection, and is related to the combination of the length of the cap-snatched primer and RNA duplex stability with the vRNA templates.

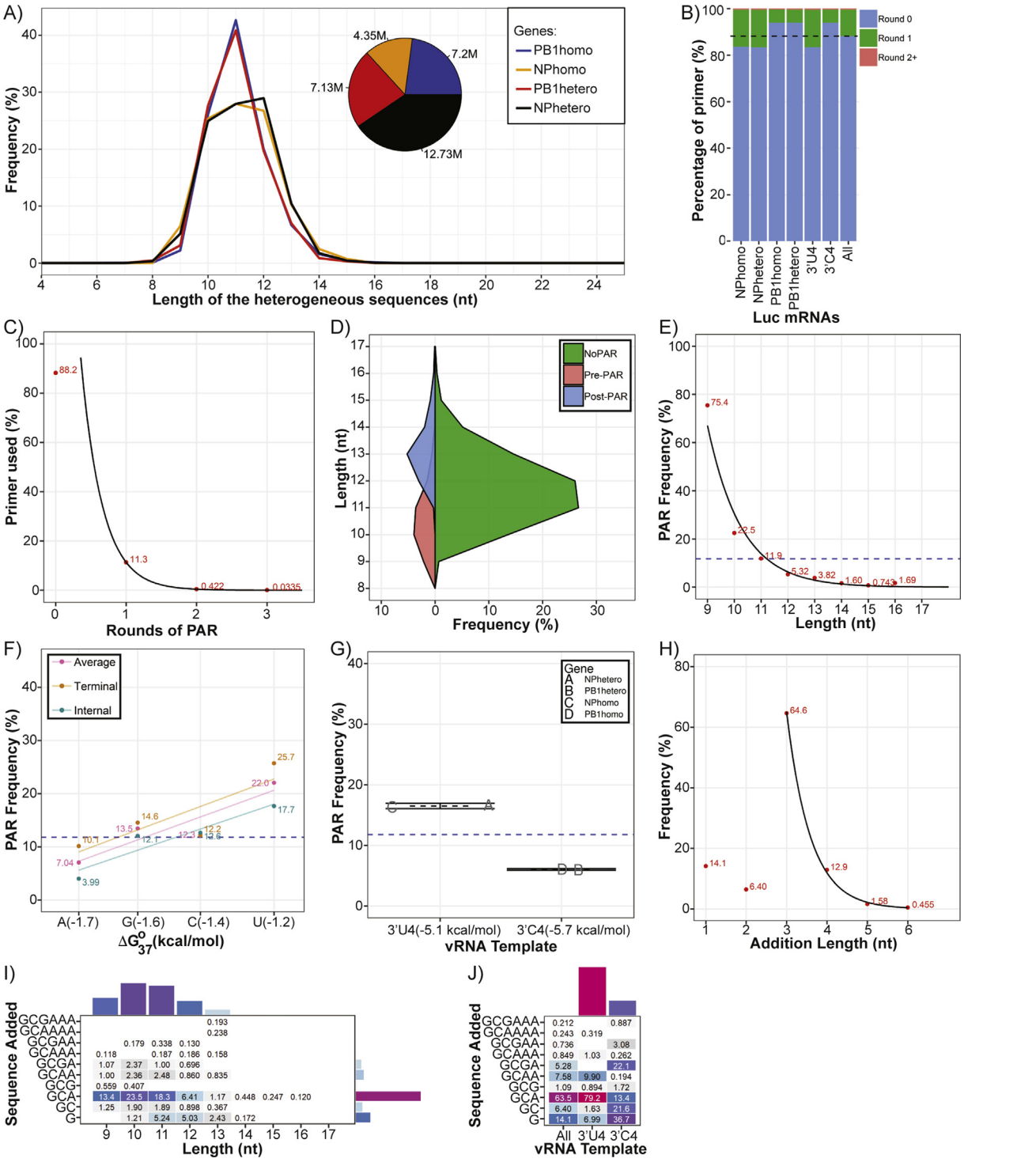
3. Discussion

Previous studies on IAV cap-snatching led to the observation of nucleotide sequence repeats complementary to the 3' end of the vRNA conserved sequence not present in the cap-snatched primers. These sequences are generated by iteratively aborted transcription of the first few nucleotides of the viral conserved sequences (Geerts-Dimitriadou et al., 2011a, 2011b; Sikora et al., 2017, 2014; Te Velthuis and Oymans, 2018). However, the determinants of this process were not elucidated in cells supporting IAV transcription. Here, we used information on sequences located upstream of virus-encoded nucleotides on mRNAs from four IAV strains (A/Puerto Rico/8/1934 (H1N1), A/Hong Kong/1/1968 (H3N2), A/WSN/1933 (H1N1) and A/Brisbane/59/2007 (H1N1)) and four mini-replicons to investigate the physical characteristics of the primers undergoing PAR.

We observed that PAR processes approximately 14.3% of the cap-snatched primers. This PAR frequency is slightly higher than a previous estimate of approximately 10% (Te Velthuis and Oymans, 2018). However, in our dataset, we obtained a comparable PAR frequency of 10.1% for 11-nucleotide long primers ending in AG, which are similar to the primer used in the previous study. Of the primers processed by PAR, approximately 95% undergo only one round. The proportion of primers processed by additional rounds of PAR remains constant at approximately 95% of residual primers per round. Consistent with previous *in vitro* analyses using purified components (Te Velthuis and Oymans, 2018), we found that the key determinants of PAR are length and RNA duplex stability with the vRNA templates. Both of these determinants are addressed by PAR, which typically adds the 3-nucleotide-long sequence GCA. This results in longer primers which are able to base-pair with the terminal uridine of the vRNA templates. Thus, PAR could rescue sub-optimal cap-snatched primers by simultaneously addressing both length and RNA duplex stability.

Length is the major determinant of PAR, and a key element shared among (−)ssRNA viruses that use PAR on cap-snatched primers (Liu et al., 2018, 2016; Yao et al., 2012). Our results suggest that the minimum length to mitigate PAR is related to the distance between the PB2 cap-binding region and PB1 active site. For IAV, this distance is 54 Å, or approximately 12 nucleotides (Reich et al., 2014). At 12 nucleotides, the PAR rate is lower than the strains average PAR rate. In order to prime transcription with a shorter primer, the RdRp would have to compress the template exit tunnel. Such compression would lead to elastic deformity of the enzyme, which would impose an energy penalty in a manner similar to compression of a non-linear spring. This penalty would increase exponentially for every decrease in primer length. This model is consistent with the observation that short primers can substantially inhibit RdRp transcription *in vitro* (Chung et al., 1994). This energy cost is also observed when priming internally at 3′C2, as internal initiation is associated with higher PAR rates when the primer is shorter than 12 nucleotides. However, this penalty does not modify the number of nucleotides transcribed prior to PAR, as length of the nucleotide sequence addition is not correlated with the length of the primer.

In addition to a length of 12 nucleotides, our results indicate that primers ending with AG have lower PAR rates and are used more often directly for transcription. Such primers are not only compatible with the distance between the PB2 cap-binding region and PB1 active site,



(caption on next page)

but also have the lowest free energy priming sequence, and allow for internal initiation. This increase in direct transcription, coupled with the preferential cleavage of host RNAs by PA immediately downstream of a guanine (Datta et al., 2013), may explain the higher observation of primers ending in AG in the cap-snatched populations reported previously (Geerts-Dimitriadou et al., 2011b, 2011a; Sikora et al., 2014). Since these populations are derived from mRNA extracted early after infection, primers from host RNA substrates with faster rates of PA cleavage, and which can be used directly for elongation would accumulate preferentially, leading to the appearance of sequence-specific targeting in a process that appears to be regulated by the abundance of host RNA targets (De Vlugt et al., 2018; Sikora et al., 2017).

PAR is also related to the nucleotide directed at 3′U1 to prime transcription. It was recently reported that mutation of 3′U1 to 3′A1 on a vRNA allows for priming transcription with a primer ending in an uridine, but leads to a large reduction in transcription when priming with a primer ending in an adenine, suggesting the formation of a Watson-Crick base pair between the 3′ ends of the primer and vRNA template (Te Velthuis and Oymans, 2018). In our data sets, we found that primers ending with A, G or C directed at 3′U1 are less associated

Fig. 6. Analysis of PAR determinants on mini-replicons. (a) Length distribution of the heterogeneous sequences successfully mapped to annotated TSS. Pie chart inset: proportion of IAV sequences corresponding to each transcript. (b) Percentage of host-derived primers used directly (Round 0) or that are processed by PAR (Round 1 and above) for each mini-replicon, mini-replicon conserved sequence type (3′U4 or 3′C4), and average (All). Only primers which were matched to within 100 nucleotides of an annotated transcription start site were included. (c) Percentage of primers used directly for transcription elongation (Round 0) and rescued at each round of PAR (Rounds 1 to 3). Black line indicates log-linear trend ($R^2 = 0.994$). (d) Length distributions of primers versus the proportion of primers used directly or processed by PAR. (e) PAR frequency inversely correlates with the length of the cap-snatched primers. The black line indicates log-linear trend ($R^2 = 0.902$). (f) Relationship between PAR frequency and maximum free energies of the RNA duplex between the cap-snatched primer and mini-replicons. Free energy was calculated using the Turner energy rules for each of the nucleotide annealing with 3′U1. The pink line represents average values ($R^2 = 0.850$), the orange line represents sequences which primed transcription terminally at 3′U1 ($R^2 = 0.715$) and the cyan line represents sequences which primed transcription internally at 3′C^2 (R2 = 0.890). Difference between Internal and Terminal groups is about 0.11 kcal/mol (p = 0.123). (g) The mini-replicons having a 3′C4 are associated with lower PAR frequency. Mini-replicons are grouped based on the 3'[U/C]4 variation (p-value = 4.81×10^{-3} between templates). The black bars indicate the 95% confidence interval. (h) Frequency of nucleotide sequences of a given length added by PAR. Log-linear model is calculated between lengths of 3 (GC[A/G]) and six (GC[A/G]AAA) nucleotides ($R^2 = 0.992$). (i) Nucleotide sequence added by PAR in relation to the length of the primer. The data is represented as a heatmap showing the percentage of primers which underwent PAR, grouped by length and nucleotide sequence added. Vertical bars at the top show the frequency of lengthening events from the indicated primer; horizontal bars at the right show the frequency of lengthening events with the specific nucleotide sequence addition. (j) Nucleotide sequence added by PAR in relation to the 3'[U/C]4 variation present on the mini-replicons. Vertical bars at the top show the frequency of lengthening events for template variants. For all panels, the values represent the average of the 4 mini-replicons and the dashed lines indicate the average PAR frequency.

with PAR than those ending with a U (Fig. 7). This suggests that the interaction between these nucleotides may be mediated by the maximum free energy rather than the ability to form a Watson-Crick base pair, and that priming with a 3′U1-directed uridine may result in a less favorable initiation structure. The structure of the putative base pair at 3′U1 has not yet been resolved (Reich et al., 2017); it is possible that the orientation of this nucleotide in the active site is such that it forms a non-canonical base pair. If true, the identity of the 3′ most nucleotide of the vRNA as a uridine residue might be important for the observed flexibility in directly transcribing from primers ending with A, G or C.

PAR typically occurs after the formation of four base pairs. Nucleotide sequence additions longer than three nucleotides were observed; however, the frequency of longer additions decreases exponentially between 3 and 6 nucleotides. This suggests that the IAV

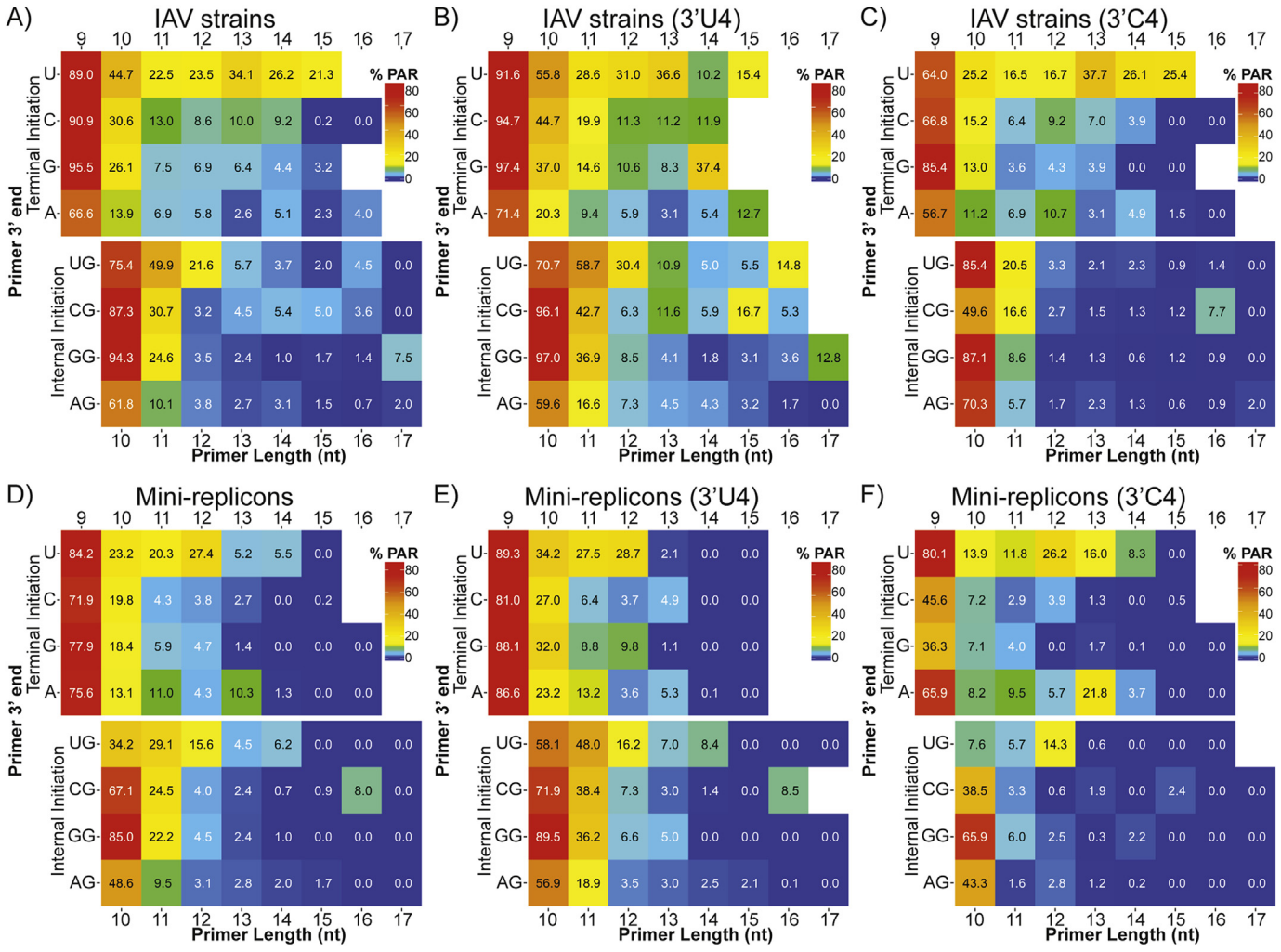


Fig. 7. PAR is dependent on the combined effects of both length and RNA duplex stability. Heatmap representation PAR frequencies based on length and nucleotide sequences used to prime transcription on all IAV templates (a), 3′U4 IAV templates (b), 3′C4 IAV templates (c), all mini-replicons (d), 3′U4 mini-replicons (e), and 3′C4 mini-replicons (f). Values represent the average of either the 4 IAV strains or the 4 mini-replicons analyzed.

RdRp could accommodate four base pairs during transcription. This number of base pairs is consistent with the number of nucleotides on the vRNA templates used during IAV polyadenylation, where polymerase stuttering requires at least four consecutive uridines (Li and Palese, 1994; Zheng et al., 1999). Once these four base pairs are formed during initiation, the product and template would be trafficked out of the respective exit tunnels. If the relative stability of the RNA duplex (hydrogen bonds, stacking interactions, RNA protein interactions, elastic deformity, etc.) is too low, the product may realign without net movement of the template, leading to PAR.

4. Conclusions

The cleavage of host mRNA occurs between 10 to 13 nucleotides downstream of the 5′ cap and peaks at 11 and 12 nucleotides (Gu et al., 2015; Koppstein et al., 2015; Plotch et al., 1981; Sikora et al., 2017, 2014). This length range is consistent with the estimated distance between the PB2 cap-binding and PA endonuclease regions (~50 Å, estimated to be at least 10 nucleotides), but is shorter than the estimated distance between the PB2 cap-binding region and the PB1 active site (54 Å, estimated to be at least 12 nucleotides) (Reich et al., 2014). This indicates that RdRp may cleave primers at a length shorter than optimal for priming transcription. Our analysis supports a model where PAR is evaluated when four base pairs are within the RdRp active site. At this point, the free energy is evaluated, and if the RNA duplex is not sufficiently stable, PAR occurs more frequently. Priming transcription with a short primer likely causes elastic deformity of RdRp leading to an energy penalty and a loss of duplex stability. The free energy of the RNA duplex is modified by the nucleotide directed at 3′U1 and the vRNA template transcribed. If the combination of these energies is insufficient, the probability of PAR increases. In this case, a three-nucleotide long sequence ending in an adenine residue (i.e. GCA) is typically produced, and rescues the primer on the basis of both length and RNA duplex stability. For example, a 9-nucleotide long primer ending in a uridine has an 89% chance to undergo PAR. After one round of PAR (and the addition of GCA), this elongated primer (12 nt long ending in A) has a 94.2% chance of not undergoing a subsequent round of PAR, and being used directly for the elongation of IAV mRNA. Thus, our results suggest that PAR could rescue suboptimal primers based on the combined effects of both length and RNA duplex stability. More importantly, the rescue of sub-optimal primers by PAR is independent of viral infection, and is mainly related to the nature of the cap-snatched primers and to their interactions with the vRNA and the RdRp.

5. Methods

5.1. Library preparation from cells transfected with the mini-replicons

Plasmids encoding the PB1, PB2, PA, and NP proteins from A/Puerto Rico/8/1934 (H1N1), and a reporter plasmid expressing a negative-sense RNA transcript consisting of the firefly luciferase ORF flanked by A/Puerto Rico/8/1934 (H1N1) NP UTRs, under the control of an RNA polymerase I promoter and terminator, were obtained from E.G. Brown (University of Ottawa, Ontario, Canada) (Ping et al., 2010). PCR was performed to replace the NP UTRs with the A/Puerto Rico/8/1934 (H1N1) PB1 UTRs on the reporter plasmid. Specifically, the complete NP or PB1 UTR sequences were added to forward and reverse luciferase primers and cloned into the *Bsm*B1 site of pHH21 (Neumann et al., 1999). Four firefly luciferase-expressing minigene plasmids were used in this study: $^{5'}$NP-Luc-NP$^{3'}$ (NPhomo), $^{5'}$PB1-Luc-PB1$^{3'}$ (PB1homo), $^{5'}$PB1-Luc-NP$^{3'}$ (NPhetero), and $^{5'}$NP-Luc-PB1$^{3'}$ (PB1hetero). A549 cells were co-transfected with plasmids encoding the PB1, PB2, PA, and NP proteins, together with one of the luciferase-expressing plasmids. Transfections were performed using Lipofectamine 2000 (Invitrogen) following manufacturer's instructions. Two days post-transfection, the cells were collected, washed twice with 1X PBS, and total RNA was extracted using the TRIzol Reagent (Invitrogen) according to manufacturer's instructions. The ExactSTART Eukaryotic mRNA 5' & 3' -RACE Kit (Epicentre Biotechnologies) was used to amplify the 5′-end of the purified RNA, according to manufacturer's instructions. Following 5′ RACE, the products were used for 25 cycles of PCR using the kit-provided primer and a primer just upstream of the start codon of the firefly luciferase gene. Sequence identifier tags for multiplexing and adapter sequences compatible with the Illumina Sequencing platform were then added by PCR. The PCR products were gel-purified using the QIAquick Gel Extraction Kit (Qiagen) and the sequences verified by Sanger Sequencing (StemCore Laboratories, Ottawa, Canada). Products were then mixed and sent for high-throughput sequencing using the Illumina HiSeq 2000 System (McGill University and Genome Quebec Innovation Centre, Montreal, Canada). Raw sequencing data have been deposited on the Sequence Read Archive of NCBI [SRA accession PRJNA542626].

5.2. Extraction and mapping of the cap-snatched host leaders

In addition to sequencing data from the mini-replicons, high-throughput sequencing data sets from the 5′ ends of IAV mRNA were retrieved from the Sequence Read Archive of NSCI for A/Puerto Rico/8/1934 (H1N1; SRR5017127), A/Hong Kong/1/1968 (H3N2;; SRR1301043), A/WSN/1933 (H1N1; GSE65101), A/Brisbane/59/2007 (H1N1; GSE67493) (Gu et al., 2015; Koppstein et al., 2015; Sikora et al., 2017, 2014). The raw datasets from the mini-replicons, A/Puerto Rico/8/1934 (H1N1), A/Hong Kong/1/1968 (H3N2) and A/Brisbane/59/2007 strains were de-multiplexed using cutadapt v1.8.1 using the gene-specific and RACE primer sequences. For the A/WSN/1933 dataset, sequences downstream of a stretch of at least 3 guanines were extracted using a custom R script. For all strains, the non-viral sequence upstream of the vRNA conserved sequences (5′-GC[AG]AAAGCAGG-3′) were then extracted and mapped as previously (Sikora et al., 2017). Briefly, the host-derived primers 9 nucleotides and longer were matched to a ± 100 nucleotide window around transcription start sites annotated in the human reference genome GRCh38 using Bowtie v.1.0.0 (Langmead et al., 2009). If no match was found, the longest single nucleotide sequence which could be added by PAR was trimmed from the 3′ end of the host-derived sequence. This iterative matching and trimming procedure was performed until a match was found. Sequences which could not be matched were discarded. In the event of multiple matches, the match closest to the annotated transcription start site was selected; if two or more matches were found, and these matches were equidistant from the transcription start site, the sequences were randomly assigned among all matches. Additionally, since priming can occur at 3′C2 rather than 3′U1 (Datta et al., 2013; Geerts-Dimitriadou et al., 2011b, 2011a; Sikora et al., 2014; Te Velthuis and Oymans, 2018), a 3′C2-directed guanine, when present in the reference genome immediately downstream of the cap-snatched primer, was added to the primers.

5.3. Analysis of PAR events

Primers matching without any rounds of trimming were categorized as “NoPAR”. Primers which were matched after at least one round of trimming were classified as “PAR”. To reduce noise in this dataset, only cap-snatched primers with at least 10 reads matched to a given transcription start site were included. Data was further limited to primers between 9 and 16 nucleotides in length. Free energy was calculated using the Turner Energy Rules Nearest Neighbour method (Turner and Mathews, 2010). The maximum free energy values were used for these calculations, and an initiation penalty was omitted. All data analysis was performed using custom R scripts (supplied as Supplementary data).

Acknowledgements

This work was funded by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC Canada; #04888) awarded to M. Pelchat. The authors wish to acknowledge the genotyping platform of the McGill University and Genome Quebec Innovation Centre for technical assistance. The authors would like to thank Dr Earl G. Brown for providing the plasmids to produce the PB1, PB2, PA, and NP proteins from A/Puerto Rico/8/1934 (H1N1), the mini-replicon constructs and for fruitful discussions. M.P. designed the project. D.S. performed the library preparation for high-throughput sequencing. L.R., C.D.V. and M.P. wrote the required in-house Perl and R scripts. M.P. and C.D.V. conducted data analyses. C.D.V., D.S. and M.P. contributed to writing the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2019.08.002.

References

Poon, L.L., Pritlove, D.C., Fodor, E., Brownlee, G.G., 1999. Direct evidence that the poly (A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. J. Virol. 73 3473–6. doi:10074205.

Beaton, A.R., Krug, R.M., 1981. Selected host cell capped RNA fragments prime influenza viral RNA transcription in vivo. Nucleic Acids Res. 9, 4423–4436. https://doi.org/10.1093/nar/9.17.4423.

Biswas, S.K., Nayak, D.P., 1994. Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. J. Virol. 68, 1819–1826.

Braam, J., Ulmanen, I., Krug, R.M., 1983. Molecular model of a eucaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription. Cell 34, 609–618. https://doi.org/10.1016/0092-8674(83)90393-8.

Caton, A.J., Robertson, J.S., 1980. Structure of the host-derived sequences present at the 5' ends of influenza virus mRNA. Nucleic Acids Res. 8, 2591–2603. https://doi.org/10.1093/nar/8.12.2591.

Chung, T.D.Y., Cianci, C., Hagen, M., Terry, B., Matthews, J.T., Krystal, M., Colonno, R.J., 1994. Biochemical studies on capped RNA primers identify a class of oligonucleotide inhibitors of the influenza virus RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 91, 2372–2376. https://doi.org/10.1073/pnas.91.6.2372.

Datta, K., Wolkerstorfer, A., Szolar, O.H.J., Cusack, S., Klumpp, K., 2013. Characterization of PA-N terminal domain of Influenza A polymerase reveals sequence specific RNA cleavage. Nucleic Acids Res. 41, 8289–8299. https://doi.org/10.1093/nar/gkt603.

De Vlugt, C., Sikora, D., Pelchat, M., 2018. Insight into influenza: a virus cap-snatching. Viruses 10, 641. https://doi.org/10.3390/v10110641.

Deng, T., Vreede, F.T., Brownlee, G.G., 2006. Different de novo initiation strategies are used by influenza virus RNA polymerase on its cRNA and viral RNA promoters during viral RNA replication. J. Virol. 80, 2337–2348. https://doi.org/10.1128/JVI.80.5.2337-2348.2006.

Dias, A., Bouvier, D., Crépin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S., Ruigrok, R.W.H., 2009. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature 458, 914–918. https://doi.org/10.1038/nature07745.

Flick, R., Hobom, G., 1999. Transient bicistronic vRNA segments for indirect selection of recombinant influenza viruses. Virology. https://doi.org/10.1006/viro.1999.9895.

Geerts-Dimitriadou, C., Goldbach, R., Kormelink, R., 2011a. Preferential use of RNA leader sequences during influenza A transcription initiation in vivo. Virology 409, 27–32. https://doi.org/10.1016/j.virol.2010.09.006.

Geerts-Dimitriadou, C., Zwart, M.P., Goldbach, R., Kormelink, R., 2011b. Base-pairing promotes leader selection to prime in vitro influenza genome transcription. Virology 409, 17–26. https://doi.org/10.1016/j.virol.2010.09.003.

Gu, W., Gallagher, G.R., Dai, W., Liu, P., Li, R., Trombly, M.I., Gammon, D.B., Mello, C.C., Wang, J.P., Finberg, R.W., 2015. Influenza A virus preferentially snatches noncoding RNA caps. RNA 21, 2067–2075. https://doi.org/10.1261/rna.054221.115.

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W.H.H., Ortin, J., Hart, D.J., Cusack, S., 2008. The structural basis for cap binding by influenza virus polymerase subunit PB2. Nat. Struct. Mol. Biol. 15, 500–506. https://doi.org/10.1038/nsmb.1421.

Koppstein, D., Ashour, J., Bartel, D.P., 2015. Sequencing the cap-snatching repertoire of H1N1 influenza provides insight into the mechanism of viral transcription initiation. Nucleic Acids Res. 43, 5052–5064. https://doi.org/10.1093/nar/gkv333.

Krug, R.M., Broni, B. a, LaFiandra, a J., Morgan, M. a, Shatkin, a J., 1980. Priming and inhibitory activities of RNAs for the influenza viral transcriptase do not require base pairing with the virion template RNA. Proc. Natl. Acad. Sci. U.S.A. 77, 5874–5878. https://doi.org/10.1073/pnas.77.10.5874.

Labaronne, A., Swale, C., Monod, A., Schoehn, G., Crépin, T., Ruigrok, R.W.H., 2016. Binding of RNA by the nucleoproteins of influenza viruses a and B. Viruses 8. https://doi.org/10.3390/v8090247.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Lee, M.-K., Bae, S.-H., Park, C.-J., Cheong, H.-K., Cheong, C., Choi, B.-S., 2003. A single-nucleotide natural variation (U4 to C4) in an influenza A virus promoter exhibits a large structural change: implications for differential viral RNA synthesis by RNA-dependent RNA polymerase. Nucleic Acids Res. 31, 1216–1223. https://doi.org/10.1093/nar/gkg214.

Li, X., Palese, P., 1994. Characterization of the polyadenylation signal of influenza virus RNA. J. Virol. 68, 1245–1249.

Liu, X., Xiong, G., Qiu, P., Du, Z., Kormelink, R., Zheng, L., Zhang, J., Ding, X., Yang, L., Zhang, S., Wu, Z., 2016. Inherent properties not conserved in other tenuiviruses increase priming and realignment cycles during transcription of Rice stripe virus. Virology 496, 287–298. https://doi.org/10.1016/j.virol.2016.06.018.

Liu, X., Jin, J., Qiu, P., Gao, F., Lin, W., Xie, G., He, S., Liu, S., Du, Z., Wu, Z., 2018. Rice stripe tenuivirus has a greater tendency to use the prime-and-realign mechanism in transcription of genomic than in transcription of antigenomic template RNAs. J. Virol. 92, e01414–e01417. https://doi.org/10.1128/JVI.01414-17.

Mark, G.E., Taylor, J.M., Broni, B., Krug, R.M., 1979. Nuclear accumulation of influenza viral RNA transcripts and the effects of cycloheximide, actinomycin D, and alpha-amanitin. J. Virol. 29, 744–752.

McGeoch, D., Kitron, N., 1975. Influenza virion RNA-dependent RNA polymerase: stimulation by guanosine and related compounds. J. Virol. 15, 686–695.

Neumann, G., Watanabe, T., Ito, H., Watanabe, S., Goto, H., Gao, P., Hughes, M., Perez, D.R., Donis, R., Hoffmann, E., Hobom, G., Kawaoka, Y., 1999. Generation of influenza A viruses entirely from cloned cDNAs. Proc. Natl. Acad. Sci. U.S.A. 96 (16), 9345–9350.

Ng, A.K.-L., Zhang, H., Tan, K., Li, Z., Liu, J. -h., Chan, P.K.-S., Li, S.-M., Chan, W.-Y., Au, S.W.-N., Joachimiak, A., Walz, T., Wang, J.-H., Shaw, P.-C., 2008. Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. FASEB J. https://doi.org/10.1096/fj.08-112110.

Pflug, A., Guilligay, D., Reich, S., Cusack, S., 2014. Structure of influenza A polymerase bound to the viral RNA promoter. Nature 516, 355–360. https://doi.org/10.1038/nature14008.

Pflug, A., Gaudon, S., Resa-Infante, P., Lethier, M., Reich, S., Schulze, W.M., Cusack, S., 2018. Capped RNA primer binding to influenza polymerase and implications for the mechanism of cap-binding inhibitors. Nucleic Acids Res. 46, 956–971. https://doi.org/10.1093/nar/gkx1210.

Ping, J., Dankar, S.K., Forbes, N.E., Keleta, L., Zhou, Y., Tyler, S., Brown, E.G., 2010. PB2 and hemagglutinin mutations are major determinants of host range and virulence in mouse-adapted influenza A virus. J. Virol. 84, 10606–10618. https://doi.org/10.1128/JVI.01187-10.

Plotch, S.J., Bouloy, M., Ulmanen, I., Krug, R.M., 1981. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell 23, 847–858. https://doi.org/10.1016/0092-8674(81)90449-9.

Pritlove, D.C., Fodor, E., Seong, B.L., Brownlee, G.G., 1995. In vitro transcription and polymerase binding studies of the termini of influenza A virus cRNA: evidence for a cRNA panhandle. J. Gen. Virol. 76 (Pt 9), 2205–2213. https://doi.org/10.1099/0022-1317-76-9-2205.

Rao, P., Yuan, W., Krug, R.M., 2003. Crucial role of CA cleavage sites in the cap-snatching mechanism for initiating viral mRNA synthesis. EMBO J. 22, 1188–1198. https://doi.org/10.1093/emboj/cdg109.

Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crépin, T., Hart, D., Lunardi, T., Nanao, M., Ruigrok, R.W.H., Cusack, S., 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. Nature 516, 361–366. https://doi.org/10.1038/nature14009.

Reich, S., Guilligay, D., Cusack, S., 2017. An in vitro fluorescence based study of initiation of RNA synthesis by influenza B polymerase. Nucleic Acids Res. 45, 3353–3368. https://doi.org/10.1093/nar/gkx043.

Scholtissek, C., Rott, R., 1970. Synthesis in vivo of influenza virus plus and minus strand RNA and its preferential inhibition by antibiotics. Virology 40, 989–996. https://doi.org/10.1016/0042-6822(70)90145-5.

Shaw, M.W., Lamb, R.A., 1984. A specific sub-set of host-cell mRNAs prime influenza virus mRNA synthesis. Virus Res. 1, 455–467. https://doi.org/10.1016/0168-1702(84)90003-0.

Sikora, D., Rocheleau, L., Brown, E.G., Pelchat, M., 2014. Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. Sci. Rep. 4, 6181. https://doi.org/10.1038/srep06181.

Sikora, D., Rocheleau, L., Brown, E.G., Pelchat, M., 2017. Influenza A virus cap-snatches host RNAs based on their abundance early after infection. Virology 509, 167–177. https://doi.org/10.1016/j.virol.2017.06.020.

Taylor, J.M., Illmensee, R., Litwin, S., Herring, L., Broni, B., Krug, R.M., 1977. Use of specific radioactive probes to study transcription and replication of the influenza virus genome. J. Virol. 21, 530–540.

Te Velthuis, A.J.W., Fodor, E., 2016. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. Nat. Rev. Microbiol. 14, 479–493. https://doi.org/10.1038/nrmicro.2016.87.

Te Velthuis, A.J.W., Oymans, J., 2018. Initiation, elongation, and realignment during influenza virus mRNA synthesis. J. Virol. 92https://doi.org/10.1128/JVI.01775-17. e01775-17.

Thierry, E., Guilligay, D., Kosinski, J., Bock, T., Gaudon, S., Round, A., Pflug, A., Hengrung, N., El Omari, K., Baudin, F., Hart, D.J., Beck, M., Cusack, S., 2016. Influenza polymerase can adopt an alternative configuration involving a radical repacking of PB2 domains. Mol. Cell 61, 125–137. https://doi.org/10.1016/j.molcel.2015.11.016.

Turner, D.H., Mathews, D.H., 2010. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res. 38, D280–D282. https://doi.org/10.1093/nar/gkp892.

Yao, M., Zhang, T., Zhou, T., Zhou, Y., Zhou, X., Tao, X., 2012. Repetitive prime-and-realign mechanism converts short capped RNA leaders into longer ones that may be more suitable for elongation during rice stripe virus transcription initiation. J. Gen. Virol. 93, 194–202. https://doi.org/10.1099/vir.0.033902-0.

Ye, Q., Krug, R.M., Tao, Y.J., 2006. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. Nature. https://doi.org/10.1038/nature05379.

Zhang, S., Wang, J., Wang, Q., Toyoda, T., 2010. Internal initiation of influenza virus replication of viral RNA and complementary RNA in vitro. J. Biol. Chem. 285, 41194–41201. https://doi.org/10.1074/jbc.M110.130062.

Zheng, H., Lee, H.A., Palese, P., García-Sastre, A., 1999. Influenza A virus RNA polymerase has the ability to stutter at the polyadenylation site of a viral RNA template during RNA replication. J. Virol. 73, 5240–5243.