Evolution of transient RNA structure-RNA polymerase interactions in respiratory RNA virus genomes

Charlotte Rigby[1,2,3], Kimberly Sabsay[1,4], Karishma Bisht[1], Dirk Eggink[5,6], Hamid Jalal[3], Aartjan J.W. te Velthuis[1,*]

[1] Lewis Thomas Laboratory, Department of Molecular Biology, Princeton University, 08544 New Jersey, United States.
[2] University of Cambridge, Department of Pathology, Addenbrooke's Hospital, Cambridge CB2 2QQ, United Kingdom
[3] Public Health England, Addenbrooke's Hospital, Cambridge CB2 2QQ, United Kingdom
[4] Sigler Institute, Princeton University, Princeton, NJ 08544, United States.
[5] Department of Medical Microbiology, Amsterdam UMC, Amsterdam, The Netherlands
[6] Center for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

* Address correspondence to: aj.te.velthuis@princeton.edu

Key words: Coronavirus, RNA structure, influenza A virus, RNA polymerase, t-loop, adaptation.

Short title: Evolution of t-loop stability in RNA virus genomes

Abstract

RNA viruses are important human pathogens that cause seasonal epidemics and occasional pandemics. Examples are influenza A viruses (IAV) and coronaviruses (CoV). When emerging IAV and CoV spill over to humans, they adapt to evade immune responses and optimize their replication and spread in human cells. In IAV, adaptation occurs in all viral proteins, including the viral ribonucleoprotein (RNP) complex. RNPs consists of a copy of the viral RNA polymerase, a double-helical coil of nucleoprotein, and one of the eight segments of the IAV RNA genome. The RNA segments and their transcripts are partially structured to coordinate the packaging of the viral genome and modulate viral mRNA translation. In addition, RNA structures can affect the efficiency of viral RNA synthesis and the activation of host innate immune response. Here, we investigated if RNA structures that modulate IAV replication processivity, so called template loops (t-loops), vary during the adaptation of pandemic and emerging IAV to humans. Using cell culture-based replication assays and *in silico* sequence analyses, we find that the sensitivity of the IAV H3N2 RNA polymerase to t-loops increased between isolates from 1968 and 2017, whereas the total free energy of t-loops in the IAV H3N2 genome was reduced. This reduction is particularly prominent in the PB1 gene. In H1N1 IAV, we find two separate reductions in t-loop free energy, one following the 1918 pandemic and one following the 2009 pandemic. No destabilization of t-loops is observed in the IBV genome, whereas analysis of SARS-CoV-2 isolates reveals destabilization of viral RNA structures. Overall, we propose that a loss of free energy in the RNA genome of emerging respiratory RNA viruses may contribute to the adaption of these viruses to the human population.

Introduction

RNA viruses, such as influenza A viruses (IAV) or coronaviruses (CoV), are important pathogens that cause mild to severe respiratory disease in humans. RNA virus spread among humans occurs in seasonal epidemics, outbreaks of emerging viruses, or occasional pandemics [1]. Seasonal epidemics are caused by viruses that are largely adapted to humans and to which humans have pre-existing immunity, such as the seasonal IAV strains H3N2 and H1N1 or CoV strains NL63 and 229E. In contrast, outbreaks and pandemics are caused by emerging RNA viruses to which humans lack pre-existing immunity [2]. Emerging viruses spill over from an animal reservoir and typically have poor human-to-human transmission. Examples of these are the IAV H5N1 and H7N9 strains. Pandemic viruses can efficiently spread from human-to-human, due to adaptations that they acquired via mutation, or through reassortment or recombination with a human-adapted virus. In recent history, IAV subtypes H1N1, H2N2 and H3N2 have caused pandemics in 1918, 1957, 1968 and 2009, while the CoV severe acute respiratory CoV 2 (SARS-CoV-2) caused the COVID-19 pandemic [3].

Analysis of IAV strains has shown that human-adaptive mutations occur in nearly all viral proteins, but in particular in the IAV surface glycoproteins hemagglutinin (HA) and neuraminidase (NA); the proteins involved in viral replication polymerase basic 1 and 2 (PB1 and PB2), polymerase acidic (PA) and the nucleoprotein (NP); and the protein involved in modulating host-immune responses non-structural 1 (NS1) [4–16]. During the SARS-CoV-2 pandemic, adaptive mutations have been observed primarily in the spike (S) glycoprotein, and potentially important mutations have been reported for the nucleocapsid protein, envelope protein, and non-structural proteins 2, 3, 4, 10, 12 and 13 [17–23]. Research indicates that mutation in the glycoproteins improve host receptor binding, virus transmission, pH stability and host immune evasion, whereas mutations in the viral replication complex improve viral RNA synthesis [17,24–31]. In addition to mutations in the viral proteins, nucleotide changes and deletions have been observed in IAV and SARS-CoV-2 genome RNA structures located in untranslated regions (UTRs), indicating that changes at the RNA structure level may also play a role in the adaptation of RNA viruses [32,33].

Secondary RNA structures in RNA virus genomes play critical roles in IAV and CoV infection cycles and perform functions in packaging, translation, and viral replication. In addition, RNA structures can be the target of therapeutics [34]. Secondary RNA structures can form between neighboring sequences, as well as involve long-range interactions, such as duplex formation between the 5’ and 3’ UTRs of various viruses [35,36]. Disruption of some RNA structures affects RNA virus growth, including packaging signals, however several IAV RNA structures have a large degree of variability among strains [37–39]. Presently, our understanding of the role of RNA secondary structures as well as mutations in these structures during RNA virus replication and adaptation is limited, but it seems likely that RNA structures that negatively impact viral replication in a new host may be modified or removed in time.

In this study, we aimed to gain insight into the evolution of template loops (t-loop), a transient RNA secondary structure that folds around the viral RNA polymerase and affects IAV replication efficiency on short viral RNA templates [40]. Currently, the effect of t-loops in full-length viral genome (segments) cannot be studied due to the lack of appropriate assays. However, we hypothesized that if strong t-loops are present in the viral genome and they affect the efficiency of viral RNA synthesis, they would be destabilized over time to improve viral fitness. These destabilizations would appear as synonymous nucleotide changes in primary sequence data. We expected that these putative t-loop destabilizations would occur

in emerging IAV strains, because they would benefit most from improvements in viral replication. In addition, we hypothesized that after t-loop destabilization, continued adaption of the viral replication complex in a t-loop-less genome would make the viral RNA polymerase sensitive to the re-introduction of t-loops. Hence, we expected human-adapted IAV to be more sensitive to t-loops structures than recently emerged/less human-adapted viruses, and that reassortment events between human and avian-adapted IAV would impose strong selections on strong t-loops, which would be apparent from sequence data.

To test the above hypotheses, we performed cell culture assays as well as analyses of IAV, IBV and SARS-CoV-2 sequence data obtained through surveillance. We observe that the sensitivity of H3N2 IAV RNA polymerases to t-loops increased between 1968 and 2017, but that the total free energy (ΔG) of t-loops in the IAV H3N2 genome was linearly increased, and thus the overall stability reduced, between 1968 and 2008. Reductions in t-loop stability are also observed in H1N1 IAV following the 1918 and 2009 pandemics. No destabilization of t-loops is observed in the IBV genome, which has been circulating in humans for decades, but an analysis of SARS-CoV-2 isolates reveals destabilization of t-loops as well. Overall, we propose that the ΔG of RNA structures in the RNA genome of emerging respiratory RNA viruses may play a role in the adaption of these viruses to humans.

Results

T-loops differentially affect IAV H3N2 RNA polymerase activity and IFN induction.

We previously identified t-loops as RNA elements that can reduce the replication efficiency of H1N1 and H5N1 IAV RNA polymerases on 71-nt long RNA templates called mini viral RNAs (mvRNAs) [40]. In our previous observations, mvRNAs containing a t-loop were more likely to trigger IFN promoter activity than mvRNAs without a t-loop [40]. So far, the sensitivity of the RNA polymerase of emerging and adapted IAV viruses to t-loops has not been directly compared. The IAV H3N2 genome has been continuously circulating in humans for 55 years, making the replication complexes of H3N2 IAVs ideal for performing a direct comparison between an emerged and adapted IAV strain. To perform the comparison, we transfected plasmids expressing the PB1, PB2, PA and NP proteins of the A/WSN/1933 (abbreviated as WSN), A/Netherlands/1968 (abbreviated as H3N2-1968), and A/Netherlands/2017 (abbreviated as H3N2-2017) RNA polymerases into HEK 293T cells along with an IFN-promoter driven luciferase reporter plasmid and a TK-promoter-driven *Renilla* luciferase control plasmid. In addition, we provided a cellular RNA polymerase I-driven plasmid expressing a 71-nt long template without a t-loop (NP71.1) and a 71-nt long template containing a t-loop (NP71.11). These templates were extensively tested previously and used to show that the IFN promoter activity is indicative of poor RNA polymerase activity [40]. Western blot analysis showed efficient expression of the viral proteins (Fig. 1) Measurement of the IFN promoter activity showed no difference in the ability of the WSN and H3N2-1968 RNA polymerase to induce IFN promoter activity. However, the H3N2-2017 RNA polymerase induced an almost 8-fold higher IFN promoter activity (Fig. 1), suggesting that it had a higher sensitivity to the t-loop structure in the template than the WSN and H3N2-1968 RNA polymerase.

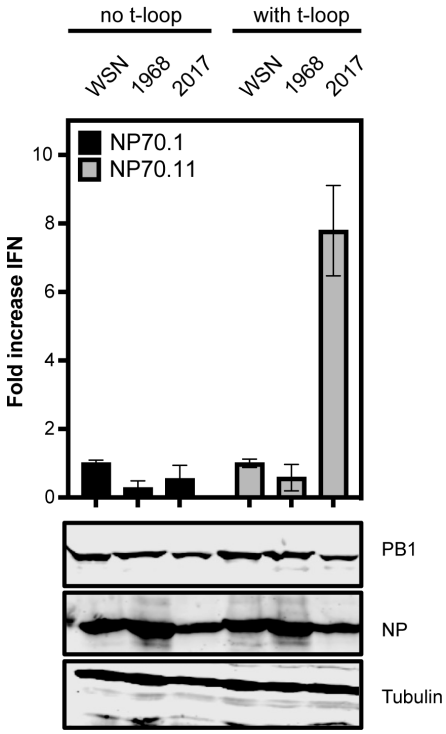

Figure 1. Interferon beta promoter activity in influenza A virus H3N2 RNP reconstitution assay. **A**) Fold change in *IFNB* promoter activation following the replication and transcription of an mvRNA containing no t-loop (NP71.1) and an mvRNA containing a t-loop (NP71.11). Fold change in *IFNB* promoter activation was normalized to the IAV WSN H1N1 control for each template. Bottom panels show western blot analysis of NP and PB1 expression for the analyzed IAV H3N2 replication complexes. Tubulin was used as loading control.

Initial analysis of IAV H3N2 t-loop stability.

The above result suggested that the RNA polymerase of a recent IAV H3N2 isolate is more sensitive to t-loops than an IAV H3N2 isolate from 1968. We hypothesized that any t-loops in the 1968 IAV H3N2 genome would be absent in the 2017 IAV H3N2 genome if they negatively affected viral replication. To explore this, we focussed on an initial set of genome sequences of IAV H3N2 isolates from 1968, 1972, 1982, 1993, 2003, 2008, 2014, and 2017. These isolates capture the full IAV H3N2 tree (Fig. 2A) and did not show a particular trend in the variation of the GC content per genome (Fig. 2B).

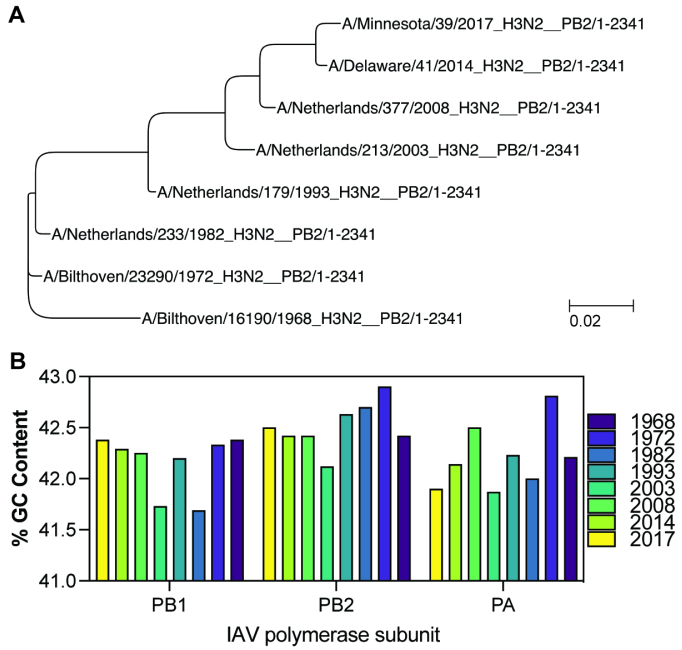

Figure 2. Phylogeny and GC content of IAV H3N2 genomes between 1968 and 2017. A) Phylogenetic analysis of IAV H3N2 based on the PB2 nucleotide sequences of isolates from 1968, 1972, 1982, 1993, 2003, 2008, 2014 and 2017. Bar represents degree of nucleotide change. B) GC content (%) for the genomes in Fig. 2A.

Next, we analyzed the putative RNA structure content in the IAV H3N2 genome. To this end, we modified the t-loop analysis that we described previously [40] to account for NP binding up and downstream of the IAV RNA polymerase during replication and RNP assembly (Fig. 3A and B). Specifically, we used a sliding window analysis, in which a sequence covered by the 20-nt footprint of the IAV RNA polymerase is blocked from participating in secondary RNA structure formation [41,42]. We also assumed that NP has a footprint of 24 nt, and that the separation of the template RNA from the downstream or upstream NP in the RNP makes 24 nt available for base pairing (Fig. 3B) [43]. With both an upstream and downstream NP removed, t-loop as well as any other secondary RNA structure formation can be calculated between the two 24-nt of single-stranded RNA (Fig. 3B). This sliding window was started at the 3' end of the viral genome and moved in 1-nt intervals along the length of the template RNA.

Before discussing our findings, we note a few caveats to our approach. Research has shown that the placement of NP along the viral genome is heterogeneous, and that various secondary RNA structures are present in IAV genome segments *in virio* [44,45]. NP dissociation downstream or upstream of an existing RNA structure may make additional nucleotides available for t-loop formation. However, the stability of such a secondary RNA structure is likely such that it will not spontaneously unwind and contribute to t-loop formation. In addition, if spontaneous unwinding were to occur, any nucleotides placed more further away from the RNA polymerase template channels are unlikely to form a tight t-loop around the RNA polymerase and will thus unlikely contribute to the modulation of RNA polymerase processivity. For simplicity, we have excluded the above two possibilities from our analysis, but point them out because we appreciate that the (transient) RNA structure landscape is likely more complex than we can currently model.

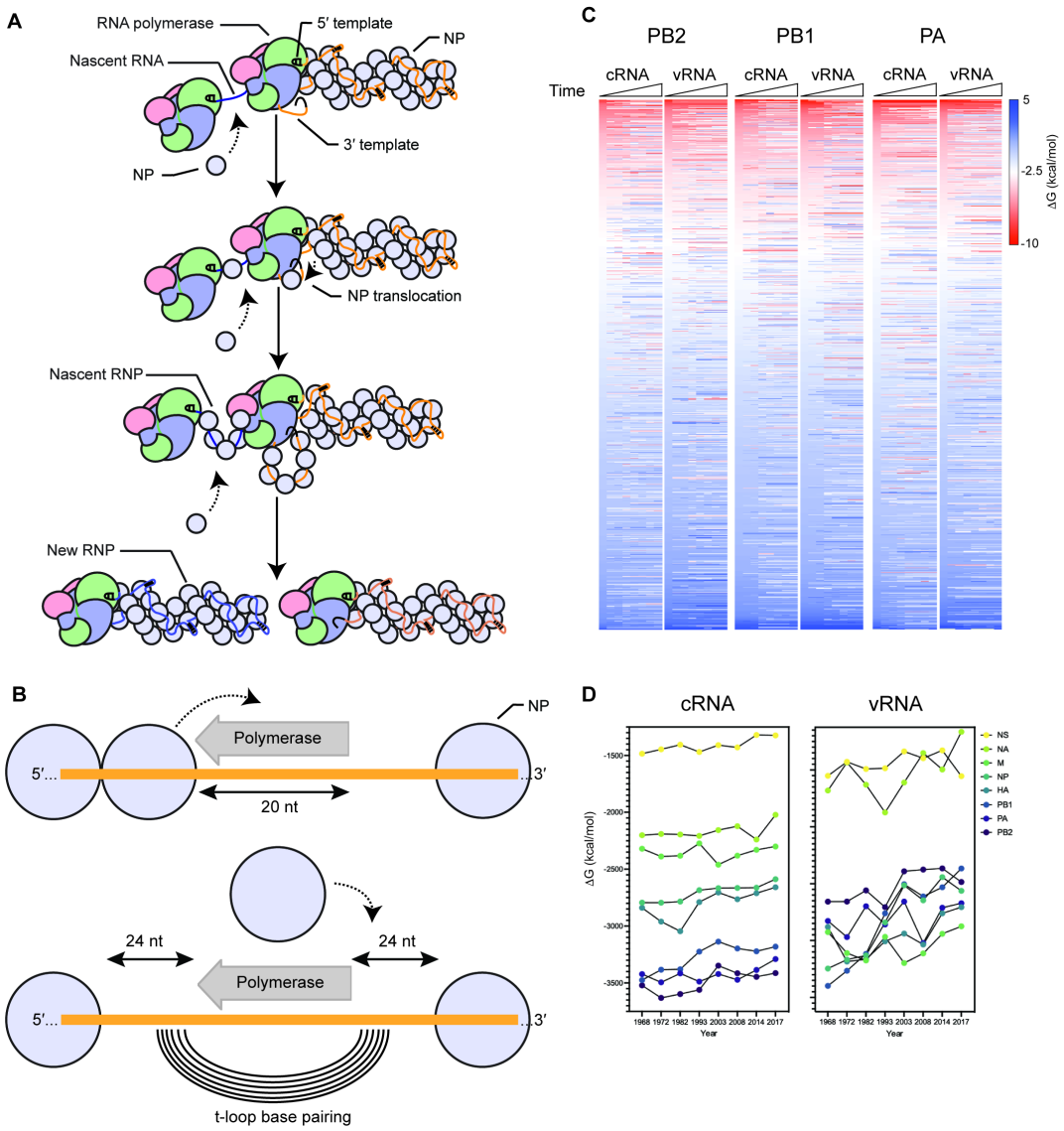

Figure 3. Model for RNP assembly during replication and analysis of influenza A virus H3N2 genomes. A) Model of IAV replication and RNP assembly following the recruitment of a second RNA polymerase and newly synthesized NP molecules. B) Model of t-loop formation that is the basis of this manuscript. The template is shown in green, with the footprint of the polymerase shaded black. Base-pairing between the up- and downstream sequences is illustrated with concentric lines. The footprint of the NP molecules, 24 nt, is also indicated. C) Analysis of the t-loop ΔG in the IAV H3N2 genome segments 1-3. Analysis was performed on the IAV H3N2 genome of Fig. 2 and ΔG sorted by 1968. See Table S1 for unsorted data. Scale bar indicates t-loop ΔG in kcal/mol. D) Total t-loop ΔG in IAV H3N2 genome segments.

Analysis of the first three IAV H3N2 genome segments showed a loss of as well as an increase in t-loop stability across the segments in both the cRNA and vRNA sense (Fig. 3C; Table S1). The heatmaps suggested that the more stable t-loops of the 1968 H3N2 genome tended to be destabilized more in the PB1 vRNA segment, relative to the other segments. To get a better sense of the overall t-loop stability across each segment in time, we calculated the total t-loop ΔG per segment, per IAV H3N2 isolate, and observed that in particular the PB1 vRNA segment showed a trend corresponding with t-loop destabilization (Fig. 3D).

The IAV H3N2 shows a progressive loss of t-loop stability in PB1 segment.

To confirm our findings, we next extended our dataset using 50 full-length IAV H3N2 sequences for each of the above years (Table S2). We first analyzed the total minimum free energy (MFE; which attempts to describe large RNA structures and structure forming up and downstream of the RNA polymerase) in segment 1 (PB2), segment 2 (PB1) and segment 3 (PA) as well as the total t-loop stability per IAV genome segment. In line with our initial analysis above, we observed clear segment-specific trends (Fig. 4). The total MFE followed a progressive increase in MFE stability (decrease in ΔG) for the PB2 segment, whereas the PB1 and PA segments demonstrated a loss of MFE stability (Fig. 4). In contrast, the t-loop stability was reduced for all three RNA polymerase segments across the 55 years analyzed. The steepest increase in t-loop ΔG was observed for the PB1 segment between 1968 and 2008. An apparent plateau in PB2 and PB1 t-loop stability followed 2008, but additional years would need to be analyzed to support this. For simplicity, we fit the trends with a linear regression ($p = <0.0001$). In general, both the vRNA and cRNA sense showed similar directions in their increase or decrease in MFE or t-loop ΔG.

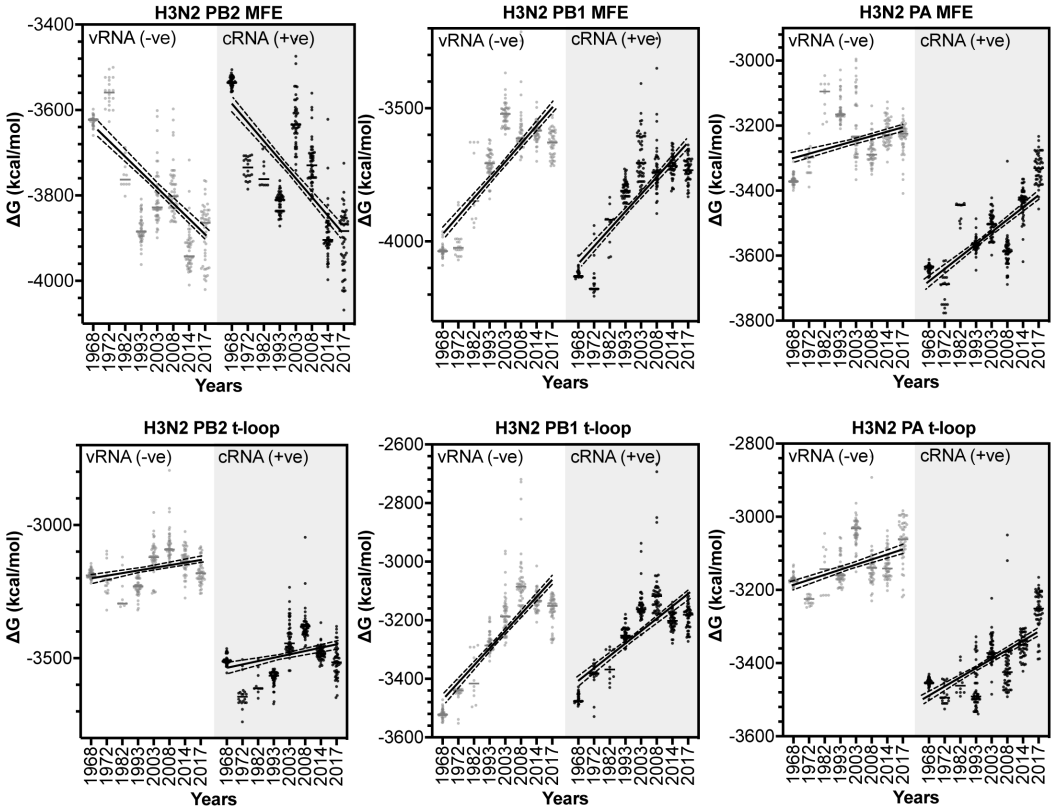

Figure 4. Total MFE and t-loop free energy analysis for IAV H3N2 isolates between 1968 and 2017. Analysis of the cRNA (light gray) and vRNA (dark gray) is shown. When plot points overlap, the points are shaded black. For each year, the first 50 full-length sequences were analyzed and plotted.

Changes in t-loop stability occur independently of amino acid mutation.

To rule out that the observed changes in t-loop ΔG across the IAV H3N2 genome segments was the result of amino acid mutation, we plotted the change in ΔG against the amino acid changes observed in the PB2 (14 changes), PB1 (13 changes) and PA (21 changes) proteins (Fig. 5). Quantification of the number of t-loop ΔG changes that co-occurred with amino acid changes showed that approximately 2 out of 3 t-loop ΔG changes occurred independently of changes at the codon level in segments 1 and 2. For segment 3, 3 out of 4 t-loop ΔG changes

occurred independently of changes at the codon level (Table 1). Overall, this suggests that the majority of t-loop ΔG changes occurs through synonymous mutation.

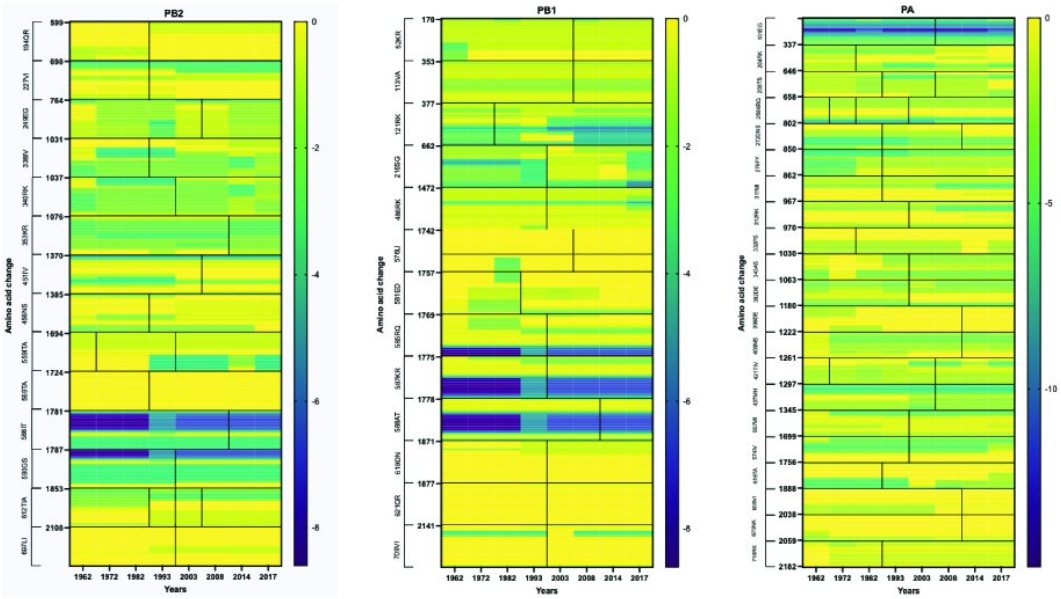

Figure 5. Co-occurrence of t-loop structure stability and amino acids changes. Co-occurrence of changes in t-loop structure and amino acid coding information in the PB2, PB1 and PA IAV H3N2 genome segments. Scale bar indicates t-loop ΔG in kcal/mol. Quantification of co-occurance is shown in Table 1.

Table 1: Coincidence of RNA structure and amino acid mutations.

	Amino acid mutation co-occurrence	
Segment	**Yes**	**No**
Segment 1 (PB2)	33.3 %	66.7 %
Segment 2 (PB1)	30.8 %	69.2 %
Segment 3 (PA)	19 %	81 %

The IAV H1N1 genome shows a loss of secondary structure destabilization.

The above data showed that the putative t-loops in the H3N2 PB1 segment were destabilized more quickly than the t-loops in the H3N2 PA and PB2 segments. We wondered if the different origin of the PB1 segment, which had been introduced into IAV H3N2 through a reassortment event with an IAV from the avian reservoir prior to 1968, was driving this increased rate in transient RNA structure destabilization [46]. To explore the possibility that adaptation of an avian origin genome segment to the human host was contributing to the change in t-loop stability, we analyzed the genome sequences of IAV H1N1 isolates (Fig. 6).

In recent memory, H1N1 IAV has been associated with two pandemics in which one or more genome segments were of avian IAV origin [47,48]. In 1918, an avian IAV spilled over into humans and gradually adapted, until it stopped circulating in humans at the start of the 1957 H2N2 pandemic. The same partially adapted H1N1 strain then remerged in 1977 and continued to circulate as seasonal IAV. In 2009, a triple-reassorted pandemic virus emerged, which contained PB2 and PA segments from an avian IAV and a PB1 segment from a human-adapted H3N2 IAV [49,50].

Analysis of the t-loop stability in the three longest H1N1 IAV genome segments revealed a gradual loss of overall t-loop stability in all three segments between 1918 and

1957. Between 1977 and 2009, the t-loop stability remained relatively unchanged for the PB1 segment, whereas the PA segment continued to show some destabilization. The t-loop stability in the PB2 segment is more diverse, and particularly wide-spread from 1998 to 2009. In 2009, a shift in t-loop stability occurred in the PA segment, in line with our expectations for the introduction of an avian-adapted gene segment through reassortment. At the same time, the H3N2-derived PB1 segment displayed a rapid reduction in t-loop stability, whereas the avian-adapted PB2 segment showed only a modest reduction in t-loop stability. Analysis of the overall MFE ΔG in the H1N1 genome segments showed that introduction of a new segment also caused a shift in MFE ΔG. However, in contrast to the t-loop ΔG, no trend in MFE ΔG change was observed following the reassortment event. This suggests that t-loop adaptation occurs independently of changes in other RNA structures in the genome. Overall, the above analysis suggests that IAV reassortment is followed by changes in t-loop stability, and that these changes can occur in both avian-adapted as well as human-adapted IAV genome segments.

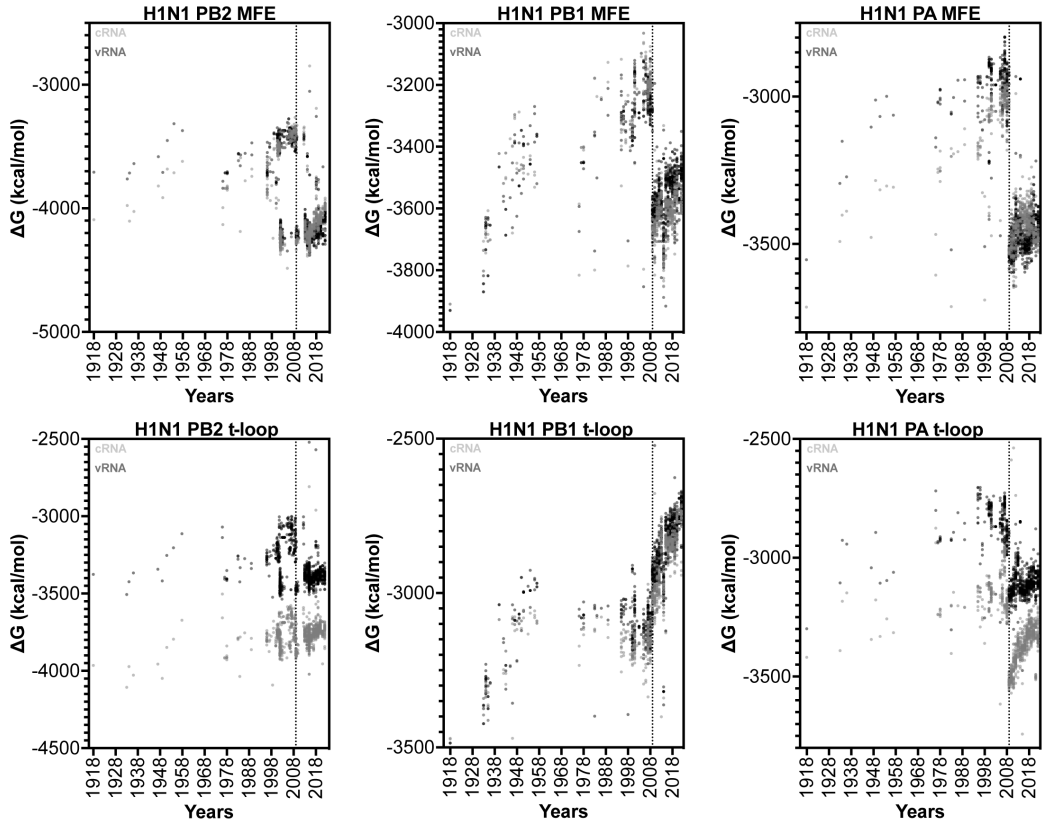

Figure 6. Total MFE and t-loop free energy analysis for IAV H1N1 isolates between 1918 and 2021. Analysis of the cRNA (light gray) and vRNA (dark gray) is shown. When plot points overlap, the points are shaded black for clarity. For each year, the first 50 full-length sequences were analyzed and plotted.

The influenza B virus genome does not show secondary structure destabilization.

Having observed the relatively rapid change in IAV H3N2 and H1N1 secondary structure stability following IAV reassortment, we analyzed the genome of the influenza B virus (IBV) (Fig. 7). Since current IBV strains have been co-circulating among humans and not been associated with spillover events from a non-mammalian animal reservoir, we expected no (rapid) change in t-loop ΔG over time. Analysis of the t-loop stability showed indeed no trend

that was indicative of stabilization or destabilization. The lack of a trend was both evident for both the vRNA as well as the cRNA sense of the three segments investigated (Fig. 7).

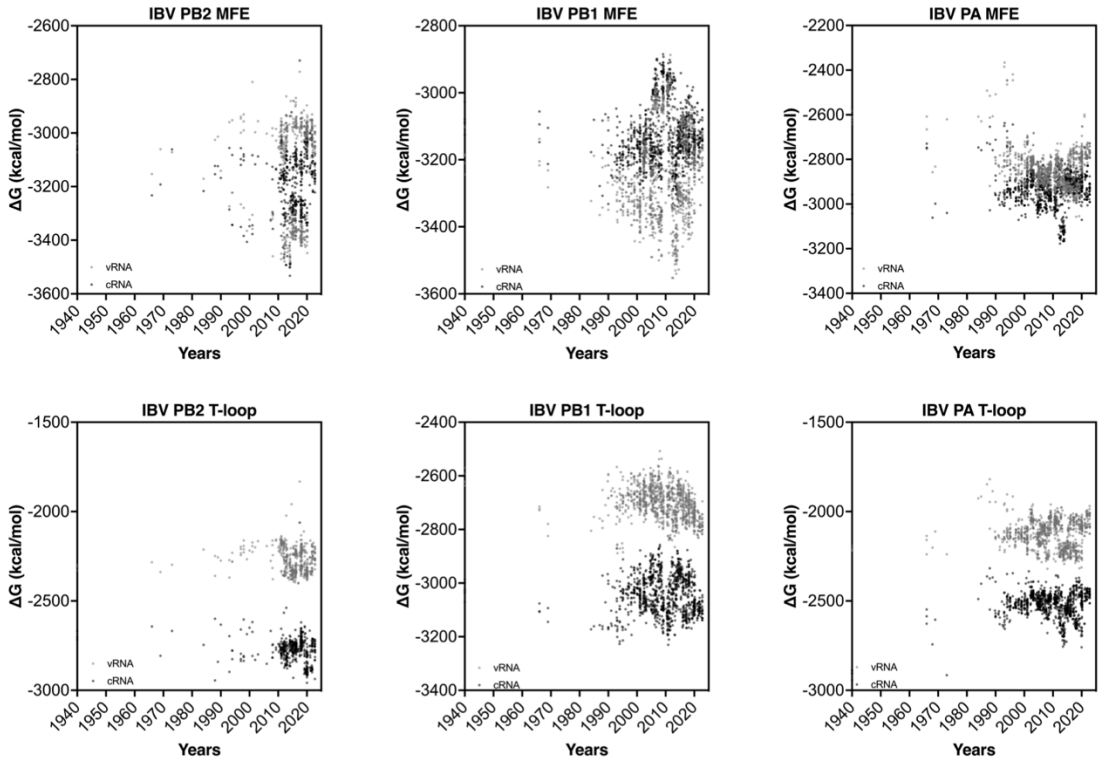

Figure 7. Total MFE and t-loop free energy analysis for IBV. Analysis of the cRNA (light gray) and vRNA (dark gray) is shown. When plot points overlap, the points are shaded black for clarity. For each year, the first 50 full-length sequences were analyzed and plotted.

In contrast to the t-loop analysis, analysis of the MFE showed a more complicated picture for the first three segments of the IBV genome (Fig. 7). For both the PB2 and PB1 segments, we noted a bifurcation in the MFE stability around 1990. This bifurcation, which reduced and increased stability, was particularly prominent for the vRNA sense, and likely linked to the two IBV lineages that have been co-circulating since the 1980s. The MFE of the PA segment was more broadly distributed and did not show a pattern. Overall, these analyses indicate that the t-loop stability in the IBV genome segments is different from the IAV genome segments. It is tempting to speculate that this difference between IAV and IBV is caused by the absence of spillover of IBV strains from a non-mammalian animal reservoir.

The SARS-CoV-2 genome shows a loss of secondary structure destabilization.

In 2019, SARS-CoV-2 emerged as the causative agent of the COVID-19 pandemic (Wu et al., 2020). Over the course of the pandemic, point mutations and deletions occurred and accumulated in the SARS-CoV-2 genome. Some of these mutations have been linked to amino acid changes that improved host receptor binding, while others caused deletions of known secondary structures. However, many synonymous mutations and less obvious deletions have not yet been linked to improved viral functions, although it has been observed that some variants display increased replication in tissue culture, suggesting that these mutations and deletions may contribute to viral replication efficiency [51]. Based on our observations with IAV, we hypothesized that t-loop-like structures may be changing in the SARS-CoV-2 genome.

To investigate the above question, we first estimated the footprint of the SARS-CoV-2 replication and transcription complex (RTC). To this end, we aligned motif C of the cryo-EM structure of the SARS-CoV-2 RTC (PDB 7CYQ) with motif C of the cryo-EM structure of the IAV RNA polymerase (PDB 6T0V) in Pymol, and estimated the route that template RNA would take through the polymerase and that 20-22 nt of the template would be covered by the RTC during RNA synthesis [41,52]. Our initial analyses showed that the above footprint parameter range did not affect the trends for the calculated MFE ΔG and the t-loop ΔG. We therefore set the footprint to the lower of these values for the data presented below, in order to collect more t-loop datapoints per genome.

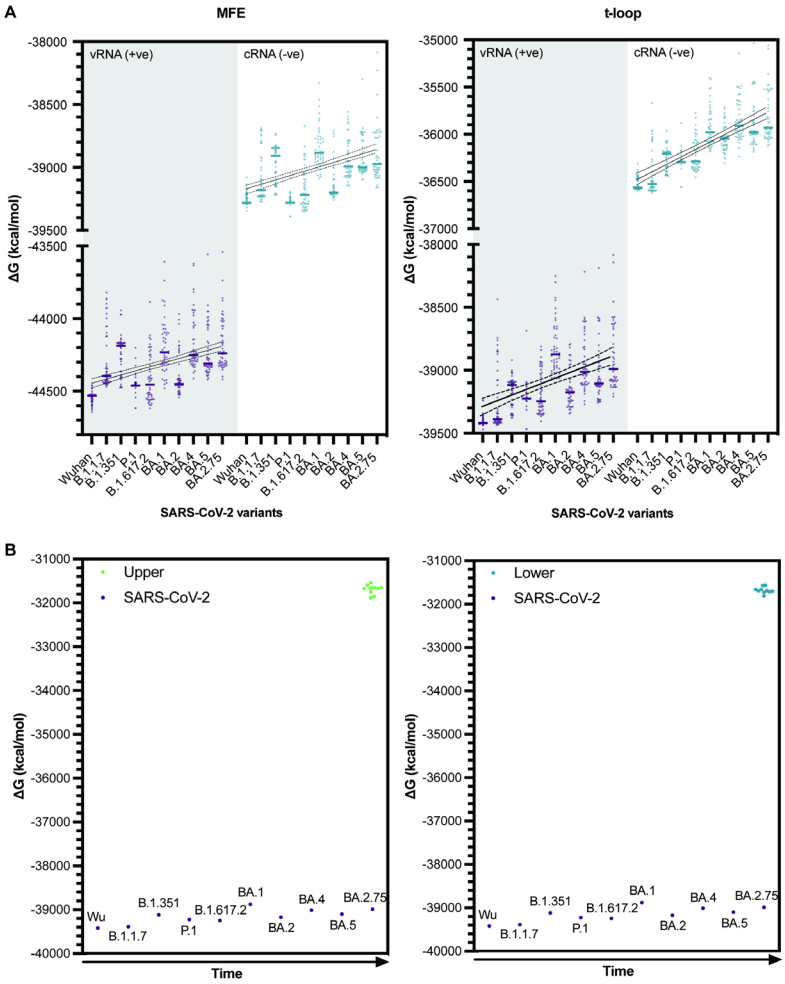

Figure 8. Total MFE and t-loop free energy analysis for SARS-CoV-2. A) Analysis of the positive sense RNA (shaded light gray; purple plot points) and negative sense RNA (gray-blue plot points) is shown. When plot points overlap, the points are shaded black for clarity. For each year, the first 50 full-length sequences were analyzed and plotted. B) Comparison of the t-loop free energy in the SARS-CoV-2 genome following random mutation of the Wuhan strain relative to the mean observed t-loop free energy in SARS-CoV-2 isolates.

We next analyzed the original SARS-CoV-2 isolate and several SARS-CoV-2 variants of concern using our computational model. For each variant we analysed the first 50, full-length, high-quality sequences published. We found that both in the positive and negative sense of the full-length genome, the MFE ΔG and t-loop ΔG trended towards a less negative ΔG, and thus less stable RNA structure content, over the course of the COVID-19 pandemic (Fig. 8A). The trends could be fit with linear regression ($p \leq 0.0001$).

Unlike IAV strains, SARS-CoV-2 has only been circulating among humans for a short time. To investigate that the observed trends were generated by chance, we simulated mutations in the reference SARS-CoV-2 genome, with frequencies of 1.8×10^{-3} and 3.95×10^{-4} per nucleotide per year as the upper and lower bounds of the whole genome mutation rate [53,54]. We performed the mutation simulations for 3 consecutive years, while maintaining the codon sequence, and next calculated the MFE and t-loop ΔG for the *in silico* variants. As shown in Fig. 8B, the t-loop ΔG for the *in silico* variants was less negative than the actual SARS-CoV-2 variants, suggesting that the observed changes in the analyzed SARS-CoV-2 variants had not likely arisen by chance. At present, this mutagenesis analysis is limited and does not take into account codon conservation.

Discussion

Respiratory RNA viruses evolve as they circulate among humans. It is well understood how mutations in viral proteins affect enzyme function and protein-protein interactions. In addition, mutations in long-range RNA secondary structures may affect RNA genome packaging or genome circularization during positive sense RNA virus replication. Currently little is still understood about the role of transient RNA secondary structures, or t-loops, which were proposed to only form when the viral replication complex copies the viral genome and modulates the local RNA secondary structure [40].

We here considered the role of the IAV RNA polymerase and NP footprint to compute putative t-loops in the IAV genome segments PB2, PB1 and PA. We observed that the transient RNA structure landscape starts to evolve when emerged IAV strains circulate in humans. Interestingly, we also observed that the RNA polymerase of a 2017 H3N2 IAV isolate was more sensitive to t-loops than the RNA polymerase of a 1968 H3N2 IAV isolate, which suggests that the transient RNA structure landscape and t-loop sensitivity of the RNA polymerase are matched. We observed a similar trend in the evolution of SARS-CoV-2, which recently emerged as a zoonotic virus from an unknown animal reservoir. Such a trend was not evident in the IBV genome, which has circulated in humans without the introduction of non-mammalian adapted sequences for decades. Possible drivers for the observed trends in IAV and SARS-CoV-2 include differences in cellular environment between hosts, including differences in the temperature at the site of replication [55].

A bias in mutations towards bases not involved in base pair formation has been previously observed [56]. It is possible that the change in the transient RNA structure landscape provides a fitness advantage to emerging respiratory RNA viruses, either by increasing the efficiency of viral replication or modulating the volume of aberrant viral products formed during viral RNA synthesis. In turn, the increased efficiency of viral replication could limit the detection of viral RNA by host pathogen receptors, such as RIG-I and MDA5, and contribute less to the severity to clinical disease. In addition, a reduction in RNA secondary structure (more equivalent to the MFE data in our analysis) has been proposed to reduce recruitment of human proteins, especially those found in stress granules [57].

Our observations were almost unilaterally based on *in silico* models that were derived from experiments performed on short IAV templates [40]. Cell-based tools will need to be developed to confirm the existence of t-loops in IAV and SARS-CoV-2 genomes and test our findings. Our *in silico* model also has putative limitations, because it does not account for secondary RNA structures that reside between NP molecules or structures that may be stabilized by RNA-protein interactions in a cellular environment. Such RNA-protein

interactions may modulate viral replication processivity in addition to t-loops. Our model also fails to account for the possibility of multiple RNA conformations, instead relying on prediction of the most thermodynamically stable structure. Future work will need to be performed, if possible, to map such potential elements and their variation among viral strains.

Even with the above caveats in mind, our results suggest that changes in transient secondary structure stability are occurring as emerging respiratory RNA virus adapt to humans. This implies that these elements affect viral replication and that a reduction in stability provides a fitness benefit for replication in human hosts. Removing elements from the viral genome that affect processivity and result in early termination or recombination may increase the efficiency of viral replication. Along the same lines, an increase in stability should provide a fitness advantage in non-mammalian hosts, should a human-adapted virus spill back into a non-mammalian host. Together, these results advance our insight into respiratory RNA virus evolution, and we believe that they yield new testable hypotheses that will help strengthen our framework for studying respiratory RNA virus replication and adaptation.

Material and methods

Cells and plasmids

Human embryonic kidney 293T (HEK293T) cells were originally sourced from ITCC and routinely checked for mycoplasma infection. HEK293T cells were maintained in DMEM containing 10 % FBS, glutamate, pyruvate, and high glucose (Gibco). WSN pcDNA3-based protein expression plasmids and pPol-based viral template RNA expression plasmids were described previously [58–60]. The H3N2 pPPI4-based protein expression was described by [66]. The firefly luciferase reporter plasmid under the control of the *IFNB* promoter [pIFΔ(−116)lucter] and the transfection control plasmids constitutively expressing *Renilla* luciferase (pTK-*Renilla*) were described previously [60]*.*

Transfections and IFN reporter assays

Transfections to perform RNP reconstitutions and IFN-β reporter assays in HEK293T cells were essentially performed as described previously [60]. RNP reconstitutions were carried out in a 24-well format by transfecting 250 ng of the pcDNA or pPPI4 plasmids encoding PB1, PB2, PA, NP and 250 ng of pPolI plasmid encoding a mvRNA template under the control of the pPolI promoter. HEK293T cells were additionally transfected with 100 ng of plasmid expressing firefly luciferase from the IFN-beta promoter and 10 ng of plasmid expressing *Renilla* luciferase. Transfections were performed with lipofectamine 2000 (Invitrogen) to the manufacturer's specifications. Twenty-four hours later, the medium was aspirated, cells were washed with 1 ml PBS and split into two fractions. Cells were pelleted at 2,500 rpm using a benchtop centrifuge. The pellets from one fraction were resuspended in 30 µl of 1x Laemmli buffer for western blotting and the other fraction was resuspended in 50 µl PBS for the IFN-β promoter activity assay. The assay was done in duplicate using 25 µl of cell suspension in PBS per well in a white 96-well plate format. Next, 25 µl of DualGlo reagent (Promega) was added per well, samples were incubated at RT for 10 min (in dark), and the firefly luciferase readings were taken using a Synergy LX Multimode Microplate Reader (Biotek). Twenty-five µl of Stop-Glo reagent/well was added next, plate was incubated for 10 min at RT (in dark), and the Renilla luciferase readings were taken. Firefly luciferase values were normalised by the *Renilla* luciferase values. Data analysis was performed in Prism 9.5.0 (Graphpad).

Antibodies and Western blotting

IAV proteins were detected using rabbit polyclonal antibodies anti-PB1 (GTX125923, GeneTex), and anti-NP (GTX125989, GeneTex) diluted 1:2000 in blocking buffer (PBS, 5% bovine serum albumin (RPI), 0.1% Tween-20 (RPI)). Cellular proteins were detected using the rat monoclonal antibody anti-tubulin (MCA77G, Bio-Rad) diluted 1:3000 in blocking buffer. Secondary antibodies IRDye 800 donkey anti-rabbit (926-32213, LI-COR) and IRDye 680 goat anti-rat (926-68076, LI-COR) were used to detect Western signals with a LI-COR Odyssey scanner.

IAV H3N2 and H1N1, and IBV sequence acquisition and processing

Initially, our analysis was performed on a limited, representative set of H3N2 sequences selected by Vigeveno et al [66]. To expand this set, 50 full-length IAV H3N2 sequences per year were acquired from EpiFlu (gisaid.org) for 1968 – 2014, and 2017 was downloaded from NCBI (see Table S2). Fifty full-length IAV H1N1 per year were acquired from NCBI for 1918 – 2021. The acquired sequences had >90% full gene segment length and <1% ambiguous nucleotide residues in all gene segments. The 11-12 nt-long conserved IAV promoter sequences were added manually if they had been trimmed and/or were missing in the downloaded sequences. Incomplete sequences that could not be manually updated were removed.

SARS-CoV-2 variant sequence acquisition and processing

Sequences were acquired from EpiCoV (gisaid.org) are listed in Table S2. Briefly, sequences were required to fit the following criteria: high coverage (<1 % N and <0.05 % unique amino acid mutations), human isolate, complete date record, and considered a complete sequence. Sequences were then sorted by sample collection date and the first 50 complete sequences selected for downstream analysis. This approach allowed us to only include the first 50 sequences of a new variant of concern, limiting putative sequence drift, incorrect variant calling, and recombination between variants. A limitation of our data selection procedure was that it overrepresented the geographic regions in which variants of concern emerged.

After selection, sequences were aligned using EMBL-EBLI Clustal Omega [61]. We noted substantial sequence variation at the 3’ and 5’ ends of the SARS-CoV-2 genome, which was likely caused by differences in amplicon design or genome assembly among labs. Prior to further analysis, we trimmed the genome sequences to nucleotide positions 602 and 27,552 of the SARS-CoV-2 reference sequence (NC_045512.2). The resulting sequences were approximately 26935 nt in length, with variation stemming from insertions or deletions.

***In silico* mutation of the SARS-CoV genome**

The SARS-CoV-2 reference genome (NC_045512.2) was mutated at frequencies of 1.8×10^{-3} and 3.95×10^{-4} per nucleotide per year [53,54]. These frequencies were used as upper and lower estimations for SARS-CoV-2s whole genome mutation rate.

In silico analysis of t-loop and other secondary RNA structures.

Our t-loop analysis was essentially performed as described previously [40], but with minor modifications to account for NP binding up and downstream of the IAV RNA polymerase during replication and RNP assembly (Fig. 2). Specifically, we used a sliding window analysis, in which a sequence covered by the 20-nt footprint of the IAV RNA polymerase is blocked

from participating in secondary RNA structure formation. Our *in silico* analysis starts with the RNA polymerase binding to the 3' end of a viral genome (segment), which places residue 16 within the 20-nt footprint in the active site [41,42]. We next assume that NP has a footprint of 24 nt, and that the separation of the template RNA from the downstream or upstream NP in the RNP makes 24 nt available for base pairing (Fig. 2B) [43]. With both an upstream and downstream NP are removed, t-loop as well as any other secondary RNA structure formation can be calculated between the two 24-nt stretches.

The above sliding window was started at the 3' end of the viral genome and moved in 1-nt intervals along the length of the template RNA. For each window, the RNA sequence upstream of the RNA polymerase footprint, downstream of the RNA polymerase footprint, and between the upstream and downstream sequences was analysed and the stability ΔG of the RNA structures calculated using the ViennaRNA package v.2.5 within Python 3.9 [62]. Locations were recorded based on the position of the active site (i.e., starting at position 16 with ΔG>0, because no template had emerged from the RNA polymerase yet). Next, the total ΔG that the RNA polymerase encounters while copying the template was calculated.

Codon analysis

To assess if changes in t-loop ΔG were caused by codon changes or emerging by chance, computational models for codon shifting and random mutations were utilized [63,64]. Specifically, the SARS-CoV-2 reference sequence was randomly mutated 10 times to the upper and lower bounds of its mutation rate [53,54]. Next RNA structures were analyzed using the sliding window approach described above. To establish how codon shuffling may impact SARS-CoV-2s RNA structures, the SARS-CoV-2 reference sequence underwent mutation for each permutation (n3, dn23, dn31 and dn231) after which the RNA structure stability was analyzed.

Consensus sequence generation and GC content analysis

H3N2 consensus sequences for PB2, PB1 and PA were generated using EMBOSS Cons and GC content was calculated using Bio.SeqUtils package from BioPython [61,65]. The python script is provided in the Supplemental Methods.

Acknowledgments

The authors would like to thank Dr. Michael Oade and Sarah van Leeuwen for discussions and reagents. Portions of the work reported in this paper were performed using the Princeton Research Computing resources at Princeton University, which is a consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology's Research Computing. CVR was supported by a studentship from Public Health England. KRS was supported by NIH grants R01 GM140032 and R01 AI170520. KB was supported by NIH grant DP2 AI175474. ATV was supported by NIH grants DP2 AI175474 and R01 AI170520, and Wellcome Trust and Royal Society grant 206579/17/Z.

References

1. Alvarez-Munoz S, Upegui-Porras N, Gomez AP, Ramirez-Nieto G. Key Factors That Enable the Pandemic Potential of RNA Viruses and Inter-Species Transmission: A Systematic Review. Viruses. 2021;13: 537. doi:10.3390/v13040537

2. Taubenberger JK, Morens DM. Influenza: The Once and Future Pandemic. Public Health Rep. 2010;125: 16–26.

3. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 2020;382: 727–733. doi:10.1056/nejmoa2001017

4. Bouvier NM, Palese P. The biology of influenza viruses. Vaccine. 2008;26 Suppl 4: D49-53. doi:10.1016/j.vaccine.2008.07.039

5. Selman M, Dankar SK, Forbes NE, Jia J-J, Brown EG. Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. Emerg Microbes Infect. 2012;1: e42. doi:10.1038/emi.2012.38

6. Forbes N, Selman M, Pelchat M, Jia JJ, Stintzi A, Brown EG. Identification of Adaptive Mutations in the Influenza A Virus Non-Structural 1 Gene That Increase Cytoplasmic Localization and Differentially Regulate Host Gene Expression. PLOS ONE. 2013;8: e84673. doi:10.1371/journal.pone.0084673

7. Czudai-Matwich V, Otte A, Matrosovich M, Gabriel G, Klenk H-D. PB2 Mutations D701N and S714R Promote Adaptation of an Influenza H5N1 Virus to a Mammalian Host. J Virol. 2014;88: 8735–8742. doi:10.1128/JVI.00422-14

8. Chang P, Sealy JE, Sadeyen J-R, Bhat S, Lukosaityte D, Sun Y, et al. Immune Escape Adaptive Mutations in the H7N9 Avian Influenza Hemagglutinin Protein Increase Virus Replication Fitness and Decrease Pandemic Potential. J Virol. 2020;94: e00216-20. doi:10.1128/JVI.00216-20

9. Wen L, Chu H, Wong BH-Y, Wang D, Li C, Zhao X, et al. Large-scale sequence analysis reveals novel human-adaptive markers in PB2 segment of seasonal influenza A viruses. Emerg Microbes Infect. 2018;7: 1–12. doi:10.1038/s41426-018-0050-0

10. Lin R-W, Chen G-W, Sung H-H, Lin R-J, Yen L-C, Tseng Y-L, et al. Naturally occurring mutations in PB1 affect influenza A virus replication fidelity, virulence, and adaptability. J Biomed Sci. 2019;26: 55. doi:10.1186/s12929-019-0547-4

11. Mänz B, Schwemmle M, Brunotte L. Adaptation of Avian Influenza A Virus Polymerase in Mammals To Overcome the Host Species Barrier. J Virol. 2013;87: 7200–7209. doi:10.1128/JVI.00980-13

12. Elderfield RA, Watson SJ, Godlee A, Adamson WE, Thompson CI, Dunning J, et al. Accumulation of Human-Adapting Mutations during Circulation of A(H1N1)pdm09 Influenza Virus in Humans in the United Kingdom. J Virol. 2014;88: 13269–13283. doi:10.1128/JVI.01636-14

13. Miotto O, Heiny AT, Albrecht R, García-Sastre A, Tan TW, August JT, et al. Complete-Proteome Mapping of Human Influenza A Adaptive Mutations: Implications for Human Transmissibility of Zoonotic Strains. PLOS ONE. 2010;5: e9025. doi:10.1371/journal.pone.0009025

14. Di Lella S, Herrmann A, Mair CM. Modulation of the pH Stability of Influenza Virus Hemagglutinin: A Host Cell Adaptation Strategy. Biophys J. 2016;110: 2293–2301. doi:10.1016/j.bpj.2016.04.035

15. Jiménez-Alberto A, Alvarado-Facundo E, Ribas-Aparicio RM, Castelán-Vega JA. Analysis of Adaptation Mutants in the Hemagglutinin of the Influenza A(H1N1)pdm09 Virus. PLOS ONE. 2013;8: e70005. doi:10.1371/journal.pone.0070005

16. Thompson AJ, Paulson JC. Adaptation of influenza viruses to human airway receptors. J Biol Chem. 2021;296. doi:10.1074/jbc.REV120.013309

17. Kistler KE, Huddleston J, Bedford T. Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. Cell Host Microbe. 2022;30: 545-555.e4. doi:10.1016/j.chom.2022.03.018

18. Sonnleitner ST, Prelog M, Sonnleitner S, Hinterbichler E, Halbfurter H, Kopecky DBC, et al. Cumulative SARS-CoV-2 mutations and corresponding changes in immunity in an immunocompromised patient indicate viral evolution within the host. Nat Commun. 2022;13: 2560. doi:10.1038/s41467-022-30163-4

19. Abbasian MH, Mahmanzar M, Rahimian K, Mahdavi B, Tokhanbigli S, Moradi B, et al. Global landscape of SARS-CoV-2 mutations and conserved regions. J Transl Med. 2023;21: 152. doi:10.1186/s12967-023-03996-w

20. Harari S, Tahor M, Rutsinsky N, Meijer S, Miller D, Henig O, et al. Drivers of adaptive evolution during chronic SARS-CoV-2 infections. Nat Med. 2022;28: 1501–1508. doi:10.1038/s41591-022-01882-4

21. Rochman ND, Wolf YI, Faure G, Mutz P, Zhang F, Koonin EV. Ongoing global and regional adaptive evolution of SARS-CoV-2. Proc Natl Acad Sci. 2021;118: e2104241118. doi:10.1073/pnas.2104241118

22. Johnson BA, Zhou Y, Lokugamage KG, Vu MN, Bopp N, Crocquet-Valdes PA, et al. Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis. PLoS Pathog. 2022;18: e1010627. doi:10.1371/journal.ppat.1010627

23. Mou K, Abdalla M, Wei DQ, Khan MT, Lodhi MS, Darwish DB, et al. Emerging mutations in envelope protein of SARS-CoV-2 and their effect on thermodynamic properties. Inform Med Unlocked. 2021;25: 100675. doi:10.1016/j.imu.2021.100675

24. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. eLife. 2020;9: e61312. doi:10.7554/eLife.61312

25. Verkhivker GM, Agajanian S, Oztas DY, Gupta G. Comparative Perturbation-Based Modeling of the SARS-CoV-2 Spike Protein Binding with Host Receptor and Neutralizing Antibodies: Structurally Adaptable Allosteric Communication Hotspots Define Spike Sites Targeted by Global Circulating Mutations. Biochemistry. 2021;60: 1459–1484. doi:10.1021/acs.biochem.1c00139

26. Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, et al. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. Nature. 2022;602: 294–299. doi:10.1038/s41586-021-04245-0

27. Laha S, Chakraborty J, Das S, Manna SK, Biswas S, Chatterjee R. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. Infect Genet Evol. 2020;85: 104445. doi:10.1016/j.meegid.2020.104445

28. Kudriavtsev AV, Vakhrusheva AV, Novoseletsky VN, Bozdaganyan ME, Shaitan KV, Kirpichnikov MP, et al. Immune Escape Associated with RBD Omicron Mutations and SARS-CoV-2 Evolution Dynamics. Viruses. 2022;14: 1603. doi:10.3390/v14081603

29. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science. 2021;371: 1139–1142. doi:10.1126/science.abf6950

30. Gupta D, Sharma P, Singh M, Kumar M, Ethayathulla AS, Kaur P. Structural and functional insights into the spike protein mutations of emerging SARS-CoV-2 variants. Cell Mol Life Sci CMLS. 2021;78: 7967–7989. doi:10.1007/s00018-021-04008-0

31. Yang Z, Han Y, Ding S, Shi W, Zhou T, Finzi A, et al. SARS-CoV-2 Variants Increase Kinetic Stability of Open Spike Conformations as an Evolutionary Strategy. mBio. 2021;13: e0322721. doi:10.1128/mbio.03227-21

32. Priore SF, Kauffmann AD, Baman JR, Turner DH. The Influenza A PB1-F2 and N40 Start Codons Are Contained within an RNA Pseudoknot. Biochemistry. 2015;54: 3413–3415. doi:10.1021/bi501564d

33. Ryder SP, Morgan BR, Coskun P, Antkowiak K, Massi F. Analysis of Emerging Variants in Structured Regions of the SARS-CoV-2 Genome. Evol Bioinforma. 2021;17: 11769343211014168. doi:10.1177/11769343211014167

34. Piasecka J, Lenartowicz E, Soszynska-Jozwiak M, Szutkowska B, Kierzek R, Kierzek E. RNA Secondary Structure Motifs of the Influenza A Virus as Targets for siRNA-Mediated RNA Interference. Mol Ther - Nucleic Acids. 2020;19: 627–642. doi:10.1016/j.omtn.2019.12.018

35. Baranovskaya I, Sergeeva M, Fadeev A, Kadirova R, Ivanova A, Ramsay E, et al. Changes in RNA secondary structure affect NS1 protein expression during early stage influenza virus infection. Virol J. 2019;16. doi:10.1186/s12985-019-1271-0

36. Furuse Y, Oshitani H. Evolution of the influenza A virus untranslated regions. Infect Genet Evol. 2011;11: 1150–1154. doi:10.1016/j.meegid.2011.04.006

37. Huston NC, Wan H, Strine MS, de Cesaris Araujo Tavares R, Wilen CB, Pyle AM. Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. Mol Cell. 2021;81: 584-598.e5. doi:10.1016/j.molcel.2020.12.041

38. Fujii K, Ozawa M, Iwatsuki-Horimoto K, Horimoto T, Kawaoka Y. Incorporation of influenza A virus genome segments does not absolutely require wild-type sequences. J Gen Virol. 2009;90: 1734–1740. doi:10.1099/vir.0.010355-0

39. Hutchinson EC, Wise HM, Kudryavtseva K, Curran MD, Digard P. Characterisation of influenza A viruses with mutations in segment 5 packaging signals. Vaccine. 2009;27: 6270–6275. doi:10.1016/j.vaccine.2009.05.053

40. French H, Pitré E, Oade MS, Elshina E, Bisht K, King A, et al. Transient RNA structures cause aberrant influenza virus replication and innate immune activation. Sci Adv. 2022;8: 8655–8655. doi:10.1126/sciadv.abp8655

41. Wandzik JM, Kouba T, Karuppasamy M, Pflug A, Drncova P, Provaznik J, et al. A Structure-Based Model for the Complete Transcription Cycle of Influenza Polymerase. Cell. 2020;181: 877-893.e21. doi:10.1016/J.CELL.2020.03.061

42. Kouba T, Drncová P, Cusack S. Structural snapshots of actively transcribing influenza polymerase. Nat Struct Mol Biol. 2019;26: 460–470. doi:10.1038/s41594-019-0232-z

43. Ortega J, Martín-Benito J, Zürcher T, Valpuesta JM, Carrascosa JL, Ortín J. Ultrastructural and Functional Analyses of Recombinant Influenza Virus Ribonucleoproteins Suggest Dimerization of Nucleoprotein during Virus Amplification. J Virol. 2000;74: 156–163. doi:10.1128/JVI.74.1.156-163.2000

44. Dadonaite B, Gilbertson B, Knight ML, Trifkovic S, Rockman S, Laederach A, et al. The structure of the influenza A virus genome. Nat Microbiol. 2019;4: 1781–1789. doi:10.1038/s41564-019-0513-7

45. Le Sage V, Nanni AV, Bhagwat AR, Snyder DJ, Cooper VS, Lakdawala SS, et al. Non-Uniform and Non-Random Binding of Nucleoprotein to Influenza A and B Viral RNA. Viruses. 2018;10: 522. doi:10.3390/v10100522

46. Wendel I, Rubbenstroth D, Doedt J, Kochs G, Wilhelm J, Staeheli P, et al. The Avian-Origin PB1 Gene Segment Facilitated Replication and Transmissibility of the H3N2/1968 Pandemic Influenza Virus. J Virol. 2015;89: 4170–4179. doi:10.1128/JVI.03194-14

47. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. Evolution and ecology of influenza A viruses. Microbiol Rev. 1992;56: 152–179. doi:10.1128/mr.56.1.152-179.1992

48. Scholtissek C, Rohde W, Von Hoyningen V, Rott R. On the origin of the human influenza virus subtypes H2N2 and H3N2. Virology. 1978;87: 13–20. doi:10.1016/0042-6822(78)90153-8

675 49. Taubenberger JK, Morens DM. Pandemic influenza--including a risk assessment of
676 H5N1. Rev Sci Tech Int Off Epizoot. 2009;28: 187–202. doi:10.20506/rst.28.1.1879

677 50. Zeng H, Pappas C, Katz JM, Tumpey TM. The 2009 Pandemic H1N1 and Triple-
678 Reassortant Swine H1N1 Influenza Viruses Replicate Efficiently but Elicit an Attenuated
679 Inflammatory Response in Polarized Human Bronchial Epithelial Cells. J Virol. 2011;85:
680 686–696. doi:10.1128/JVI.01568-10

681 51. Pyke AT, Nair N, van den Hurk AF, Burtonclay P, Nguyen S, Barcelon J, et al. Replication
682 Kinetics of B.1.351 and B.1.1.7 SARS-CoV-2 Variants of Concern Including Assessment
683 of a B.1.1.7 Mutant Carrying a Defective ORF7a Gene. Viruses. 2021;13: 1087–1087.
684 doi:10.3390/v13061087

685 52. Yan L, Ge J, Zheng L, Zhang Y, Gao Y, Wang T, et al. Cryo-EM Structure of an Extended
686 SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in
687 Cap Synthesis. Cell. 2021;184: 184-193.e10. doi:10.1016/j.cell.2020.11.016

688 53. Li X, Wang W, Zhao X, Zai J, Zhao Q, Li Y, et al. Transmission dynamics and evolutionary
689 history of 2019-nCoV. J Med Virol. 2020;92: 501–511. doi:10.1002/JMV.25701

690 54. Shen S, Zhang Z, He F. The phylogenetic relationship within SARS-CoV-2s: An expanding
691 basal clade. Mol Phylogenet Evol. 2021;157: 107017–107017.
692 doi:10.1016/J.YMPEV.2020.107017

693 55. Priore SF, Moss WN, Turner DH. Influenza A Virus Coding Regions Exhibit Host-Specific
694 Global Ordered RNA Structure. PLOS ONE. 2012;7: e35989.
695 doi:10.1371/journal.pone.0035989

696 56. Simmonds P. Pervasive RNA secondary structure in the genomes of SARS-CoV-2 and
697 other coronaviruses. mBio. 2020;11: 1–15. doi:10.1128/mBio.01661-20

698 57. Vandelli A, Monti M, Milanetti E, Armaos A, Rupert J, Zacco E, et al. Structural analysis
699 of SARS-CoV-2 genome and predictions of the human interactome. Nucleic Acids Res.
700 2020;48: 11270–11283. doi:10.1093/nar/gkaa864

701 58. Deng T, Sharps J, Fodor E, Brownlee GG. In Vitro Assembly of PB2 with a PB1-PA Dimer
702 Supports a New Model of Assembly of Influenza A Virus Polymerase Subunits into a
703 Functional Trimeric Complex. J Virol. 2005;79: 8669–8674. doi:10.1128/JVI.79.13.8669-
704 8674.2005

705 59. Vreede FT, Jung TE, Brownlee GG. Model Suggesting that Replication of Influenza Virus
706 Is Regulated by Stabilization of Replicative Intermediates. J Virol. 2004;78: 9568–9572.
707 doi:10.1128/JVI.78.17.9568-9572.2004

708 60. Te Velthuis AJW, Long JC, Bauer DLV, Fan RLY, Yen H-L, Sharps J, et al. Mini viral RNAs
709 act as innate immune agonists during influenza virus infection. Nat Microbiol. 2018;3:
710 1234–1242. doi:10.1038/s41564-018-0240-5

61. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022; gkac240–gkac240. doi:10.1093/nar/gkac240

62. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6: 26–26. doi:10.1186/1748-7188-6-26

63. Jorge DM de M, Mills RE, Lauring AS. CodonShuffle: a tool for generating and analyzing synonymously mutated sequences. Virus Evol. 2015;1: vev012–vev012. doi:10.1093/ve/vev012

64. Kühl MA, Stich B, Ries DC. Mutation-Simulator: fine-grained simulation of random mutations in any genome. Bioinformatics. 2021;37: 568–569. doi:10.1093/bioinformatics/btaa716

65. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25: 1422–1423. doi:10.1093/bioinformatics/btp163

66. Vigeveno R, Han AX, de Vries RP, Parker E, de Haan K, et al. Long term evolution of human seasonal influenza virus A(H3N2) is associated with an increase in polymerase complex activity. BioRxiv.2022;

Figure legends

Figure 1. Interferon beta promoter activity in RNP reconstitution assay. **A**) Fold change in *IFNB* promoter activation following the replication and transcription of an mvRNA containing no t-loop (NP71.1) and an mvRNA containing a t-loop (NP71.11). Fold change in *IFNB* promoter activation was normalized to the WSN control for each template. Bottom panels show western blot analysis of NP and PB1 expression for the analyzed IAV replication complexes. Tubulin was used as loading control.

Figure 2. Phylogeny and GC content of IAV H3N2 genomes between 1968 and 2017. A) Phylogenetic analysis of IAV H3N2 based on the PB2 nucleotide sequences of isolates from 1968, 1972, 1982, 1993, 2003, 2008, 2014 and 2017. Bar represents degree of nucleotide change. B) GC content (%) for the genomes in Fig. 2A.

Figure 3. Model for RNP assembly during replication and t-loop formation. A) Model of IAV replication and RNP assembly following the recruitment of a second RNA polymerase and newly synthesized NP molecules. B) Model of t-loop formation that is the basis of this manuscript. The template is shown in green, with the footprint of the polymerase shaded black. Base-pairing between the up- and downstream sequences is illustrated with concentric lines. The footprint of the NP molecules, 24 nt, is also indicated. C) Analysis of the t-loop ΔG in the IAV H3N2 genome segments 1-3. Analysis was performed on the IAV H3N2 genome of Fig. 2. Scale bar indicates t-loop ΔG in kcal/mol. D) Total t-loop ΔG in IAV H3N2 genome segments.

Figure 4. Total MFE and t-loop free energy analysis for IAV H3N2 isolates between 1968 and 2017. Analysis of the cRNA (light gray) and vRNA (dark gray) is shown. When plot points overlap, the points are shaded black. For each year, the first 50 full-length sequences were analyzed and plotted.

Figure 5. Co-occurrence of t-loop structure stability and amino acids changes. Co-occurrence of changes in t-loop structure and amino acid coding information in the PB2, PB1 and PA IAV H3N2 genome segments. Scale bar indicates t-loop ΔG in kcal/mol. Quantification of co-occurance is shown in Table 1.

Figure 6. Total MFE and t-loop free energy analysis for IAV H1N1 isolates between 1918 and 2021. Analysis of the cRNA (light gray) and vRNA (dark gray) is shown. When plot points overlap, the points are shaded black for clarity. For each year, the first 50 full-length sequences were analyzed and plotted.

Figure 7. Total MFE and t-loop free energy analysis for IBV. Analysis of the cRNA (light gray) and vRNA (dark gray) is shown. When plot points overlap, the points are shaded black for clarity. For each year, the first 50 full-length sequences were analyzed and plotted.

Figure 8. Total MFE and t-loop free energy analysis for SARS-CoV-2. A) Analysis of the positive sense RNA (shaded light gray; purple plot points) and negative sense RNA (gray-blue plot points) is shown. When plot points overlap, the points are shaded black for clarity. For each year, the first 50 full-length sequences were analyzed and plotted. B) Comparison of the

777 free energy in the SARS-CoV-2 genome following random mutation of the Wuhan strain
778 relative to the observed free energy in SARS-CoV-2 isolates.