A functional RNA structure in the influenza A virus ribonucleoprotein complex for segment bundling

Naoki Takizawa[a#], Koichi Higashi[b], Risa Karakida Kawaguchi[c], Yasuhiro Gotoh[d], Yutaka Suzuki[e], Tetsuya Hayashi[d], Ken Kurokawa[b]

[a]Laboratory of Virology, Institute of Microbial Chemistry (BIKAKEN), Tokyo, Japan

[b]Genome Evolution Laboratory, National Institute of Genetics, Mishima, Japan

[c]Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Aomi, Koto-ku, Tokyo, Japan

[d]Department of Bacteriology, Faculty of Medical Sciences, Kyushu University, Fukuoka, Japan

[e]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan

#Address correspondence to Naoki Takizawa, takizawan@bikaken.or.jp

Abstract

The influenza A virus genome is segmented into eight viral RNAs (vRNA). Intersegment interactions are necessary for segment bundling, and secondary structures on vRNA are assumed to be involved in the process. However, the RNA structure required for segment bundling remains unidentified because the secondary structure of vRNA in virion was partially unwound by binding viral non-specific RNA binding proteins. Here, we revealed the global intersegment interactions and the secondary structure of the vRNA in virion. We demonstrated that a pseudoknot structure was formed on a segment in the virion and the impairment of replication and packaging of the other specific segment was observed in cells infected with recombinant virus which had mutations in the pseudoknot structure. Moreover, we showed that the intersegment interactions were reconstituted in the recombinant virus. Our data provides the first evidence that the functional RNA structure on the influenza A virus genome affects segment bundling.

Introduction

The influenza A virus (IAV) genome consists of eight single-stranded negative-sense RNA segments (vRNA). One copy of each segment is packaged together into a single virus particle and eight segments are organized in a conserved '7+1' configuration in the virus particle [1,2]. Segment reassortment is one of the driving forces for IAV evolution. Genetic reassortment between the human IAV and the avian/animal IAV can lead to the emergence of a new subtype of IAV which is a candidate for a pandemic influenza strain. The terminal regions of each segment are partially complementary and connected by a short loop at one end [3,4]. Moreover, each genome segment forms the viral ribonucleoprotein (vRNP) with the viral RNA polymerase and nucleoprotein (NP), which is a single-stranded RNA binding protein with no sequence specificity and known to bind viral genome RNAs.

To reveal the mechanism of genome packaging and segment bundling, previous studies have examined motif sequences required for efficient genome packaging and bundling [5]. Identified signal sequences were found to be located in the coding regions at both ends of each segment. Mutations and deletions of some signal sequences resulted in impairment in the bundling eight segments [6–11] and each segment was shown to have a different importance for the viral genome bundling [12], suggesting that the bundling of the segmented genome is a hierarchical process. While the function of these signal sequences remains unknown, it is suggested to be involved in intersegment interactions. Regions

responsible for the interactions between vRNAs have been identified *in vitro*, suggesting the existence of specific intersegment interaction networks important for genome packaging [13–16]. Direct contacts between the vRNPs have also been observed by electron tomographic analyses [17,18]. As such, intersegment interactions are considered to be one of the important key factors to control of precise genome packaging into the virion through the signal sequences at the 5′ and 3′ coding regions.

The importance of the RNA secondary structures has been widely recognized in the formation and regulation of higher-order interactions such as protein binding and intersegment interactions. In many RNA viruses, specific regions of the viral RNA genomes also act as cis-acting regulatory elements that mediate the virus propagation. These cis-acting RNA elements often form highly specialized structural motifs such as stem-loops, and pseudoknots [19]. In IAV, the conserved 5′ and 3′ termini of the vRNAs are annealed to form a hairpin structure necessary for the viral RNA synthesis [20,21]. The comprehensive analysis of specialized structural motifs coded on the IAV genome RNA was carried out by the RNA secondary structure predictions, identifying the conservation and enrichment of stem-loop structures on the IAV genome [22,23]. Furthermore, the mutations that disrupt a predicted stem-loop structure have been shown to result in the reduction of virus propagation [22–25]. Gavazzi et al. identified an *in vitro* direct interaction between segments 2 and 8 of a H5N2 avian IAV strain, and found that these two regions involved in the intersegment interaction may form stem-loop structures that could

potentially initiate the intersegment interaction by forming a kissing-loop complex [14]. These findings resulted in the hypothesis that specialized structural motifs on the IAV genome RNA mediate intersegment interaction networks.

However, it is difficult to identify the functional RNA motif in IAV infected cells by *in silico* prediction because the secondary structure and intersegment interactions of the viral genome RNA can be unwound by binding NP [26]. Recent studies by cross-linking immunoprecipitation (CLIP) analyses have revealed that NP does not bind vRNA uniformly, indicating that the secondary structures of vRNAs are partially unwound by binding NP so that some specific regions of the vRNP can form functional secondary structures that cannot be predicted *in silico* [27,28]. In addition, previous studies have found that mutations of the conserved amino acids in NP impaired viral replication and induced irregular packaging of the genome segments, suggesting that conserved residues on NP are essential for correctly bundling eight segments [29,30]. On the other hand, the actual mechanism how NP is involved in bundling of the segments as well as the impact of NP binding on intersegment interactions remain unknown. For this reason, it is highly demanded to analyze the RNA secondary structures and the intersegment interactions of vRNA simultaneously with the existence of NP binding to reveal the elaborated packaging machinery of the IAV segments.

Recently, the methods to globally map pairwise RNA-RNA interactions were developed [31–33] and the interactions between different segment pairs have been examined

at a genome-wide level by a comprehensive high-throughput sequencing (HTS) approach [34]. Dadonaite et al. showed that a redundant and complex network of intersegment interactions exists in the virion and analyzed the comprehensive secondary structures of the viral genome in virion. While this study suggests that intersegment interactions are essential for bundling of the eight segments, the current RNA-RNA interaction analysis has several problems such as detection biases and comprehensiveness. The comparative analysis of different HTS approaches showed that each HTS approach possesses the own detection bias for base reactivity [35]. Moreover, the positive predictive value of SPLASH, a method used in the previous study to detect RNA-RNA interactions, was around 0.6 for 80S rRNA even with the four replicates, suggesting the contact map relying on the single method represents the partial view of higher-order structures of viral genome in virion. To further understanding the mechanism of segment bundling in IAV, thus, a reliable and high-resolution contact map as well as the comprehensive information on the secondary structures of the IAV genome in virion need to be created.

In this study, we revealed the global intra- and intersegment interactions and the secondary structures of the IAV genome in the virion using HTS technology. We applied multiple high-throughput and massive-scale sequencing techniques not used in the previous study; crosslinking, ligation and sequencing of hybrids (CLASH) [36] and ligation of interacting RNA followed by high-throughput sequencing (LIGR-seq) [33] for identifying intra- and intersegment interactions, and dimethyl sulphate (DMS)-seq [37–39]

and selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE)-seq [40,41] for determination of the RNA secondary structures. We identified a specialized structural motif on the vRNP and showed that unwinding of this structural motif resulted in the impairment of replication and segment bundling. We further demonstrated that a large fraction of intersegment interactions detected between multiple segments were reconstituted in the recombinant viruses that had mutations in the structural motif region. Our results suggest that the structural motif formed on vRNP is required for the bundling of eight segments to preserve the genomic structures of the IAV.

Results

Intersegment structures supported by RNA interactions

To identify global genome RNA interactions in the virion, we developed special protocols for two existing large-scale sequencing methods optimized for the vRNA, which is highly complexed with viral proteins. One is a method called CLASH, which cross-links RNA-protein interactions [36] and another is a LIGR-seq, which cross-links RNAs that form base pairs between each other [33] (Figure 1).

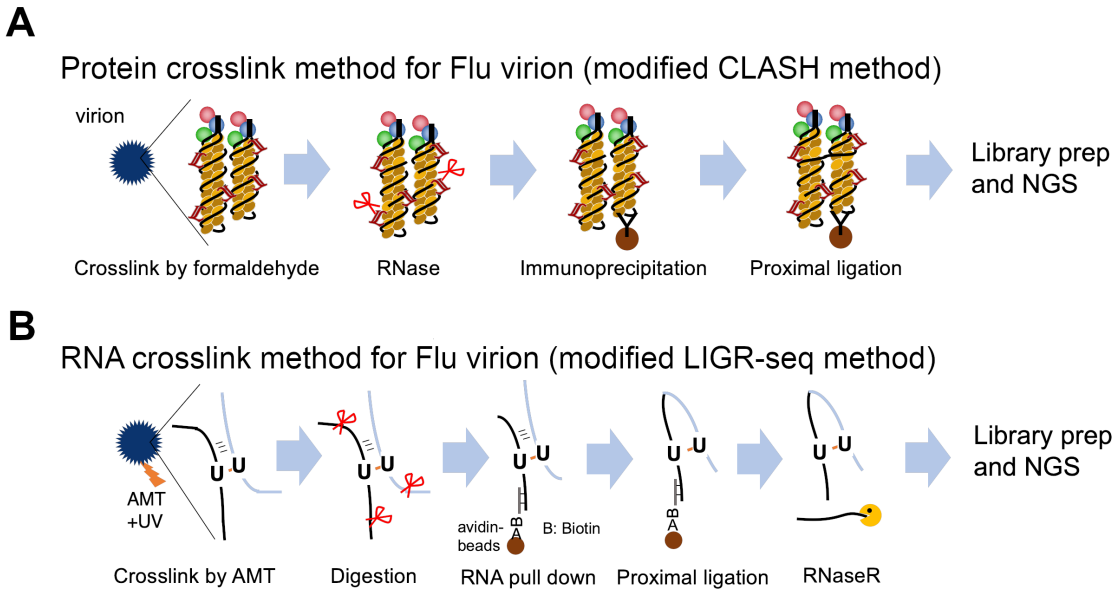


Figure 1. Schematic representation of the methods for the identification of intra- and intersegment interactions. (A) Modified CLASH method. Purified virion was treated with formaldehyde to crosslink between protein and RNA and to crosslink between proteins. After partial RNA digestion by nuclease, vRNP

was purified by immunoprecipitation, and proximal ligation was performed. (B) Modified LIGR-seq method. Purified virion was treated with AMT to crosslink between RNAs. After partial RNA digestion, vRNA was purified by a biotin-conjugated antisense single-strand DNA, and proximal ligation was performed.

In our modified CLASH method, the purified virion was treated with formaldehyde to crosslink viral proteins and RNA. Then, the contact map representing intra- and intersegment interactions was constructed as described in materials and methods. To quantify the interaction frequencies, the contact map was normalized using iterative correction and eigenvector decomposition (ICE) normalization method, which has been widely employed for normalization of Hi-C data [42].

We first analyzed the vRNP structure in the virion formed by intrasegment interactions. In our CLASH contact map, we captured the interaction of 3′- and 5′- ends of the vRNA and a loop structure previously observed by electron microscopy [3,4,43], even though the contact signals were broadly distributed (Figure 2A). This interaction of the 3′- and 5′- ends was not observed in the control experiment using the vRNA as a starting source (Figure S1A). These results suggest that the interaction between viral proteins, mainly NPs, supports the formation of the vRNP loop structures observed *in vivo*.

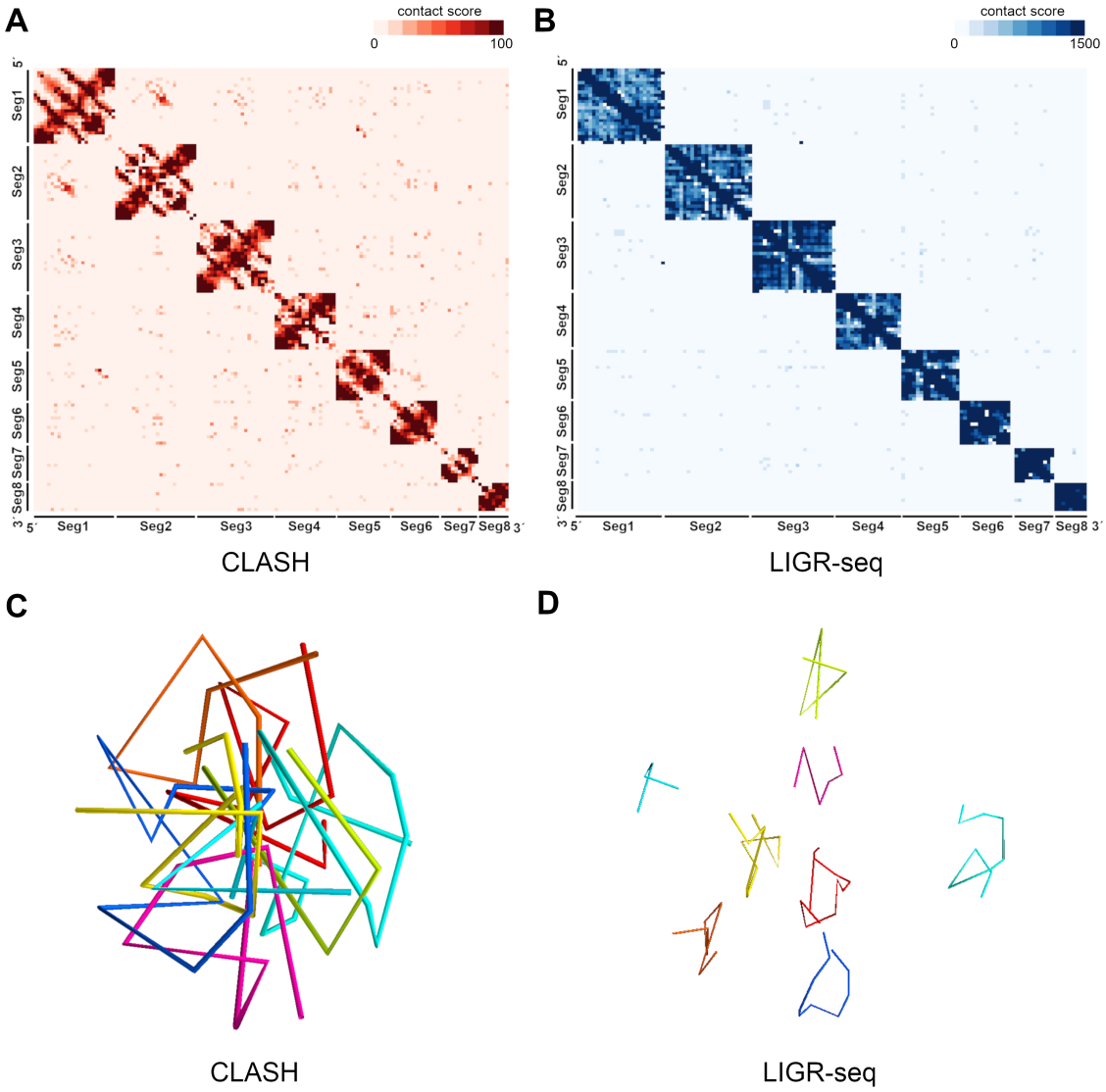


Figure 2. Intra- and intersegment interaction map of segment RNAs. (A and B) Normalized RNA contact score of CLASH and LIGR-seq. Normalized contact scores were calculated from CLASH (A) and LIGR-seq (B) data and contact maps were constructed. The light and shade of colors in each bin represent low and high normalized contact scores. (C and D) Reconstitution of the viral genome arrangement from the contact scores. Segment arrangement was reconstituted

from contact scores of CLASH (C) and LIGR-seq (D) using MiniMDS.

Moreover, the consistency of intrasegment interaction detection is investigated using our modified LIGR-seq. We analyzed the sequence reads following the same workflow of CLASH and obtained the contact score. To assess a bias of selection of the crosslinked RNA, we performed LIGR-seq experiments with or without RNaseR treatment. Background intrasegment signals were reduced in the RNaseR-treated samples and the intersegment signal was enhanced (Figure S1B). We utilized both sequence datasets for the irreproducible discovery rate (IDR) [44] to identify the intersegment interactions. As a result, the interaction of the 3′- and 5′- ends of the vRNA was also captured by LIGR-seq while the strength of signals was substantially weaker (Figure 2B).

Next, we analyzed the intersegment interactions in the virion to reveal the existence of more complicated higher-order structures formed by multiple vRNA segments. Consequently, we identified 290 and 80 intersegment interactions by CLASH and LIGR-seq, respectively. These intersegment interactions formed redundant and complex intersegment interaction networks in both CLASH and LIGR-seq (Figure 2). If intersegment interactions are precisely captured by CLASH or LIGR-seq, the correct segment arrangement can be reconstituted by the estimated distances from the contact scores. Thus, we performed the reconstitution of the segment arrangement using miniMDS [45]. Previous reports suggest that eight vRNPs are packaged in a virion with a

specific ‘1+7’ configuration where one central vRNP segment is surrounded by seven vRNP segments [2,18]. Figure 2C shows the segment arrangement reconstituted from CLASH and all segments were predicted to be entangled (Figure 2C). Conversely, the non-entangled and the ‘1+7’-like vRNP arrangement were reconstituted from LIGR-seq (Figure 2D). These results suggest that overall intersegment interactions were well captured by LIGR-seq compared to CLASH. Moreover, to confirm the reproducibility of the intersegment interactions captured by LIGR-seq, we compared our global map of intersegment interactions with those previously identified by sequencing of psoralen-crosslinked, ligated and selected hybrids (SPLASH) [34]. Intersegment interaction maps from our modified LIGR-seq and two replicates of SPLASH are shown in Figure S2A. Twelve intersegment interactions (12/80, 15% of the total interactions detected by LIGR-seq) were identified in all three results and 22 interactions (22/80, 27.5%) were identified in the LIGR-seq result and either one of the two SPLASH results (Figure S2B), suggesting that suggest that intersegment interactions captured by LIGR-seq are also captured by a different method, partially. We further identified novel 56 interactions in LIGR-seq, showing the high potential of our modified LIGR-seq. These interactions were new candidate interactions that are involved in the viral genome bundling. In addition, the contact score of intersegment interactions in our analysis is expected to be more reliable because the contact scores were normalized by ICE and the information on the reproducibility was included by IDR to filter out false positives in the final normalized

contact scores.

Identification of pseudoknot structure formed in the vRNP

Specific RNA structures, such as kissing-loop complex, have a potential to regulate RNA interactions directly and indirectly. Thus, revealing global vRNA secondary structure is necessary for elucidating the mechanism which defines the intersegment interactions. To reveal the conformational changes of the vRNA secondary structures by binding of the viral proteins, we performed DMS-seq and SHAPE-seq for IAV genome RNA in three different conditions; the vRNA, vRNP, and virion.

These methods are aimed to detect the RNA regions that are more accessible and likely to be attacked by the reagents. Thus, we can infer single-stranded and double-stranded regions at a single nucleotide resolution based on the reactivity scores. However, each method is known to have a large bias in the detectable regions. For example, DMS and NAI can modify different moieties of a single-stranded RNA [46], resulting in a different detectability depending on the sequence content. For this reason, we utilized both DMS-seq and SHAPE-seq to uncover the whole landscape of the secondary structures of vRNA which was highly complex with the viral proteins. The reactivity scores of each nucleotide were calculated from the large-scale sequencing data using reactIDR [47] and BUMHMM [48]. The reactivity score plot of each segment was shown in Figure S3. The overall tendency of RNA structure was assessed for each segment using

the Gini index, which quantifies the evenness of reactivity score distribution [37] (Table 1) and the violin plot of reactivity scores (Figure 3A and S3). As a result, the median of reactivity scores from the vRNP labeled with DMS was higher than that from the DMS-labeled vRNA. The Gini index from the vRNP labeled with DMS was lower than that from the DMS-labeled vRNA except for segment 5 (Figure 3A and Table 1). These results suggest that the overall secondary structure of vRNA is likely to be dissolved by forming in the vRNP. On the other hand, the result of SHAPE-seq shows an opposite tendency. As a result, the median of reactivity scores from the vRNP labeled with NAI was higher than that from the NAI-labeled vRNA in SHAPE-seq. The Gini index from the vRNP labeled with NAI was also higher than that from vRNA when NAI-labeling was applied (Figure S4 and Table 1). These opposite results can be explained by the different specificity of DMS and NAI to modify moieties of a single-stranded RNA. NAI has been shown to modify the 2′-OH group in the ribose backbone, whereas the vRNA has been shown to bind to NP via the phosphodiester backbone [49]. Therefore, the reactivity scores of the vRNP obtained by SHAPE-seq may become lower by NP binding to vRNA backbone, not by forming of the base pairs. For that reason, we utilized the result of DMS-seq to find a differentially structured region with and without NP.

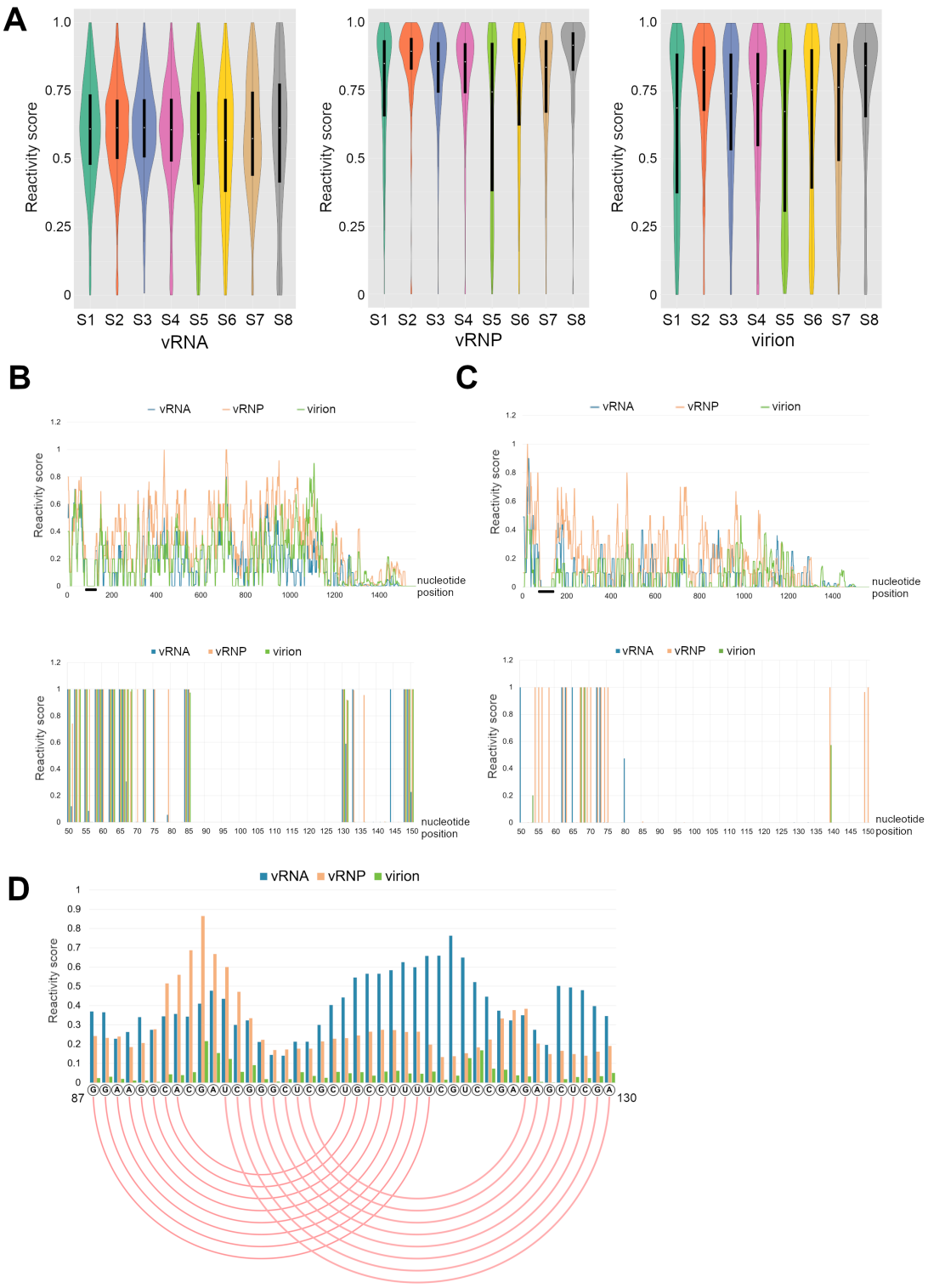


Figure 3. A pseudoknot structure at nucleotide positions 87–130 of segment 5 in the virion. (A) Distribution of reactivity scores from DMS-seq of the vRNA,

vRNP, and virion. Reactivity scores from DMS-seq were calculated by reactIDR and the scores of each segment were shown by violin plot. (B and C) The reactivity scores of segment 5 in the vRNA, vRNP, and virion. Reactivity scores were calculated by BUMHMM and a 10-nt moving average of the reactivity score of segment 5 from DMS-seq (B) and SHAPE-seq (C) were shown (upper panels). The lower panels show the reactivity scores of a specific region indicated by a black line in the upper panels. (D) The reactivity scores and predicted base pairs at nucleotide positions 87 – 130 of segment 5. The reactivity scores of each nucleotide from SHAPE-seq was calculated by reactIDR. Pink lines indicate predicted base pairs in the pseudoknot structure.

Table 1. Gini index of each segment from DMS-seq and SHAPE-seq

DMS-seq	vRNA	95% confidence interval	vRNP	95% confidence interval	virion	95% confidence interval
Seg1	0.187	(0.177-0.198)*	0.155	(0.145-0.166)	0.271	(0.257-0.286)
Seg2	0.163	(0.153-0.173)	0.081	(0.072-0.090)	0.146	(0.135-0.157)
Seg3	0.162	(0.153-0.171)	0.105	(0.097-0.113)	0.199	(0.187-0.211)
Seg4	0.184	(0.171-0.198)	0.116	(0.106-0.128)	0.204	(0.188-0.219)
Seg5	0.243	(0.227-0.259)	0.266	(0.249-0.283)	0.296	(0.277-0.314)
Seg6	0.243	(0.228-0.259)	0.177	(0.162-0.193)	0.264	(0.244-0.284)
Seg7	0.209	(0.193-0.225)	0.151	(0.137-0.170)	0.219	(0.200-0.237)
Seg8	0.265	(0.240-0.291)	0.102	(0.086-0.119)	0.174	(0.150-0.198)
Total	0.199	(0.196-0.204)	0.144	(0.139-0.149)	0.224	(0.218-0.230)

SHAPE-seq	vRNA	95% confidence interval	vRNP	95% confidence interval	virion	95% confidence interval
Seg1	0.197	(0.190-0.201)	0.379	(0.370-0.388)	0.412	(0.404-0.420)
Seg2	0.205	(0.197-0.212)	0.380	(0.370-0.389)	0.373	(0.364-0.381)
Seg3	0.185	(0.178-0.192)	0.432	(0.422-0.440)	0.444	(0.434-0.452)
Seg4	0.198	(0.189-0.207)	0.433	(0.422-0.444)	0.430	(0.420-0.440)
Seg5	0.236	(0.227-0.245)	0.448	(0.437-0.459)	0.406	(0.394-0.417)
Seg6	0.247	(0.235-0.259)	0.436	(0.424-0.448)	0.426	(0.414-0.438)
Seg7	0.229	(0.217-0.243)	0.503	(0.486-0.517)	0.454	(0.439-0.467)
Seg8	0.289	(0.274-0.303)	0.415	(0.400-0.428)	0.348	(0.332-0.363)
Total	0.219	(0.215-0.222)	0.427	(0.424-0.431)	0.417	(0.413-0.421)

* 95% confidence interval was calculated by bootstrap method.

When focusing on the reactivity differences between each segment in DMS-seq, the median of reactivity score of segment 5 vRNP was lower than those of other segments while the Gini index of segment 5 vRNP labeled with DMS was comparable to that of segment 5 vRNA (Figure 3A and Table 1). These results suggest that segment 5 vRNP is highly structured compared with other segments. Because segment 5 is an only segment whose entire vRNA structure was less dissolved by binding viral proteins, we next examined the local secondary structure of segment 5 and found the less reactive region for both NAI and DMS was found in vRNA at nucleotide positions 87 – 130 of segment 5 (Figure 3B and 3C). This less reactive region was consistently observed in the vRNP and virion, suggesting that the secondary structure of this region is not changed by the vRNP formation. Moreover, the intersegment RNA interaction analysis showed the

strong signals between nucleotide positions 1 – 100 of segment 5 and segments 3, 6, and 8 (Figure S5).

Interestingly, this region was categorized in the low-NP-binding regions [28] and predicted to form a pseudoknot structure in the previous studies [22,28]. Figure 3D shows a predicted secondary structure by IPknot [50] for the 87 – 130 nucleotides of segment 5 and the reactivity scores of the NAI-labeled vRNA, vRNP, and virion at each nucleotide position calculated by reactIDR. The reactivity scores of the NAI-labeled vRNP and virion at nucleotide positions 87 – 130 supported the stem and loop structure predicted by IPknot while the reactivity scores in the vRNA only weakly supported the stem and loop structure, suggesting that the pseudoknot structure at nucleotide positions 83 – 130 is more frequently formed in the vRNP and virion. While other additional pseudoknot structure regions were predicted in segments 1 and 8 [28], we did not detect these regions as pseudoknot structures in our DMS-seq or SHAPE-seq results.

To investigate the role of the pseudoknot structure in segment 5 for virus propagation, we constructed a recombinant virus where multiple mutations were introduced to disrupt the pseudoknot structure (referred as 87mut, hereafter). The 87mut virus had three mutations, G96A, C126U, and G129A, that did not induce amino acid changes of NP coded on segment 5 (Figure 4A). Among the three mutations within the pseudoknot region, G96 is located within the loop in our predicted pseudoknot structure, but within the stem region in a slightly different pseudoknot structure predicted in a

previous study [22].

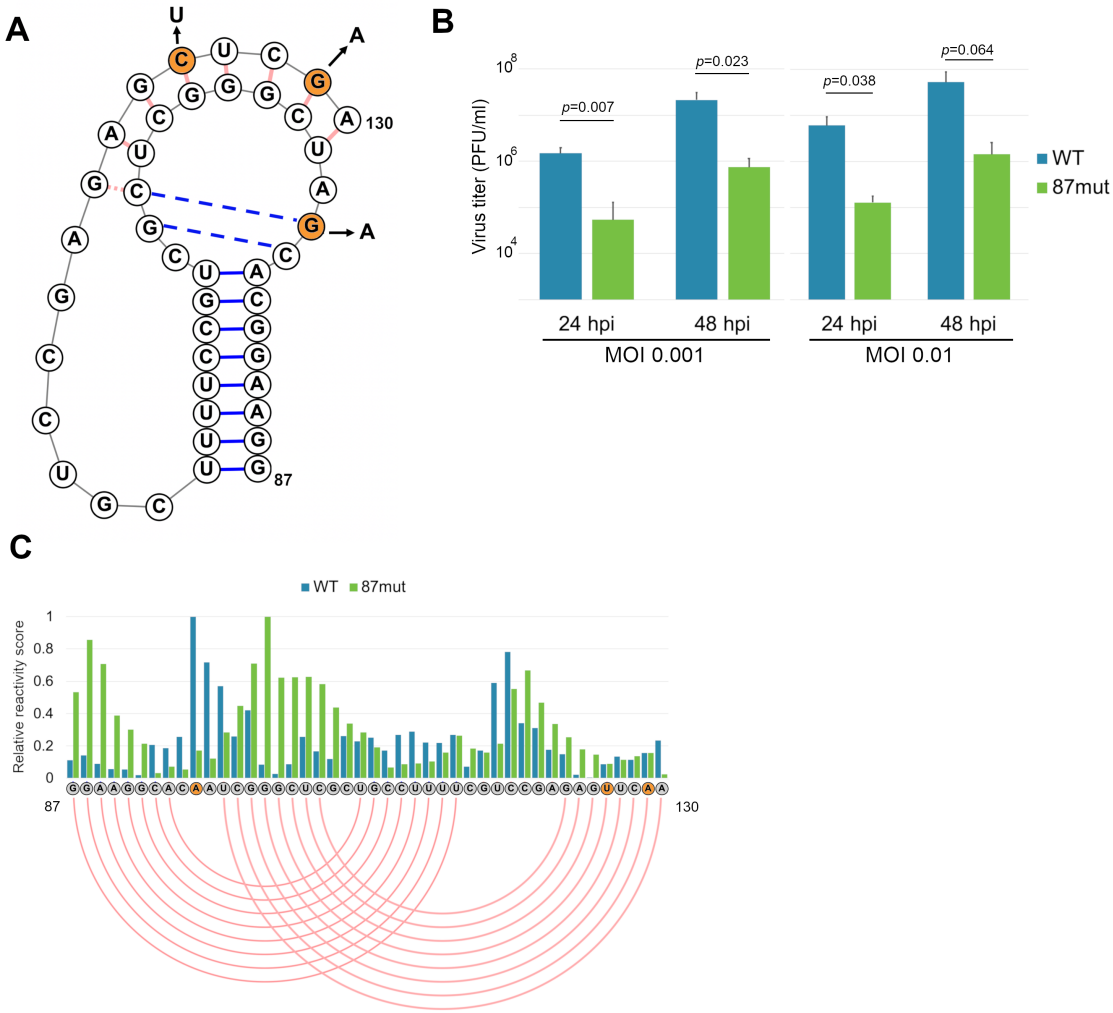


Figure 4. Unwinding the pseudoknot structure in the 87mut virus. (A) The pseudoknot structure of segment 5 and mutations in the 87mut virus. Orange-background nucleotides indicate the mutated nucleotides in the 87mut virus. (B) Virus propagation of the 87mut virus. MDCK cells were infected with the wild type or the 87mut virus at the indicated MOI. The supernatant was collected at indicated hours post infection (hpi) and the virus titer was determined by a

plaque assay. The graph indicates average values with standard deviations from three independent experiments. P-values were calculated by the Student's *t*-test. (C) Relative reactivity scores at nucleotide positions 87 – 130 of segment 5. Reactivity scores from SHAPE-seq using the wild type and the 87mut virus were calculated by reactIDR. The relative reactivity scores were calculated by dividing the reactivity score of each nucleotide by the highest reactivity score in the region. Orange-background nucleotides indicate the mutated nucleotides in the 87mut virus and pink lines indicate the predicted base pairs in the wild type sequence.

We first analyzed whether these mutations affected the virus propagation. As a result, the propagation of the 87mut virus was significantly impaired compared with that of the wild type virus, even though all three mutations did not change any amino acid residues (Figure 4B). We also examined the structural differences using SHAPE-seq for the 87mut virus. The relative reactivity scores at nucleotide positions 87 – 130 of segment 5 for NAI in the 87mut virion did not correspond to the RNA secondary structure predicted from the wild type sequence, indicating that pseudoknot structure became unstable by introduction of these mutations (Figure 4C). These results suggest that the pseudoknot structure in segment 5 plays an important role in virus propagation.

Impairment of segment bundling and viral genome replication by mutations in pseudoknot structure in segment 5

To assess in which step at which virus propagation was impaired in the 87mut virus, we determined the amounts of the vRNA segment in the cells infected with the mutant virus. The relative amount of segment 3 in the cells infected with the 87mut virus was significantly decreased compared with that with the wild type virus (Figure 5A). This result suggests that the mutations in the pseudoknot region have the potential to affect the replication of segment 3.

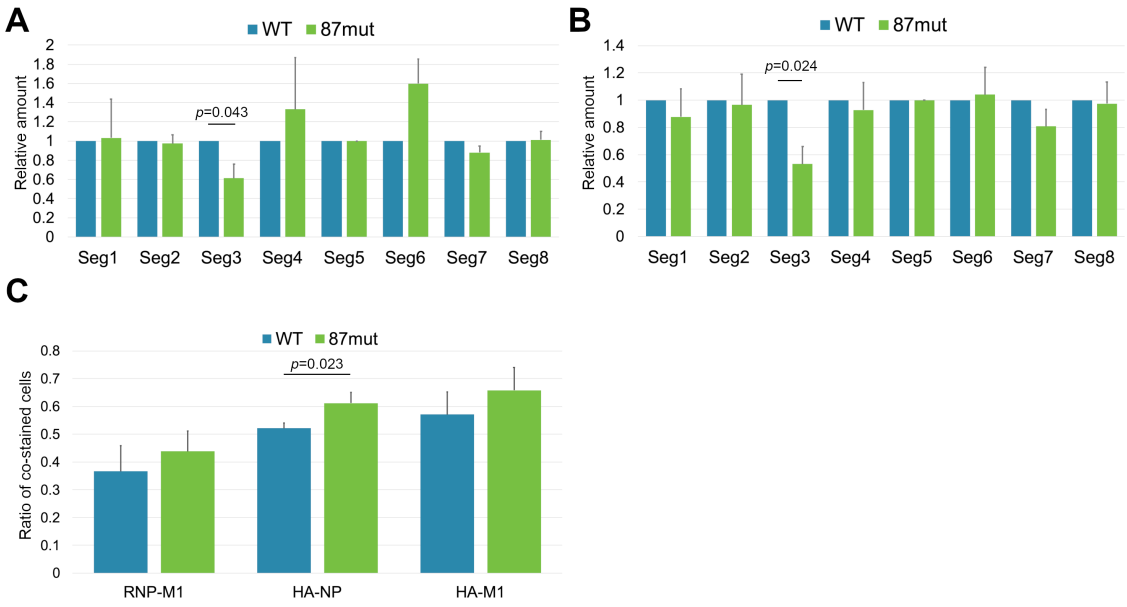


Figure 5. Impairment of replication and packaging of segment 3 and alteration of the segment bundling pattern in the 87mut virus. (A and B) Relative vRNA amount in the infected cells (A) and in the virion (B). The

amount of each segment was determined by RT-qPCR and the relative amount was calculated by double-normalization to the wild type virus and segment 5. The graph indicates average values with standard deviations from three independent experiments. P-values were calculated by the Student's *t*-test. (C) Co-expression rate of two viral proteins in cells infected with the wild type or the 87mut virus. Ratio of the co-stained cells were determined by FACS. The graph indicates average values with standard deviations from three independent experiments. P-values were calculated by the Student's *t*-test.

Furthermore, to analyze the vRNA packaging efficiency of the mutant virus, we determined the relative amounts of each vRNA segment in the virion. Consequently, the relative amounts of segment 3 in the 87mut virus was decreased compared with that of the wild type virus (Figure 5B). These results suggest that packaging of a specific segment was impaired by the mutations in the pseudoknot region. Moreover, we performed a FACS analysis of the viral proteins in the cells infected with the mutant virus to quantify the fraction of semi-infectious particles, that represented the segment bundling function [51]. The infected cells were stained with the combination of NP and M1, HA and NP, and HA and M1, respectively. The ratio of the cells expressing HA and NP was increased compared to that of the cells infected with the wild type virus (Figure 5C). This result indicates that the co-packaging efficiency of segments 1, 2, 3, and 5 encoding viral

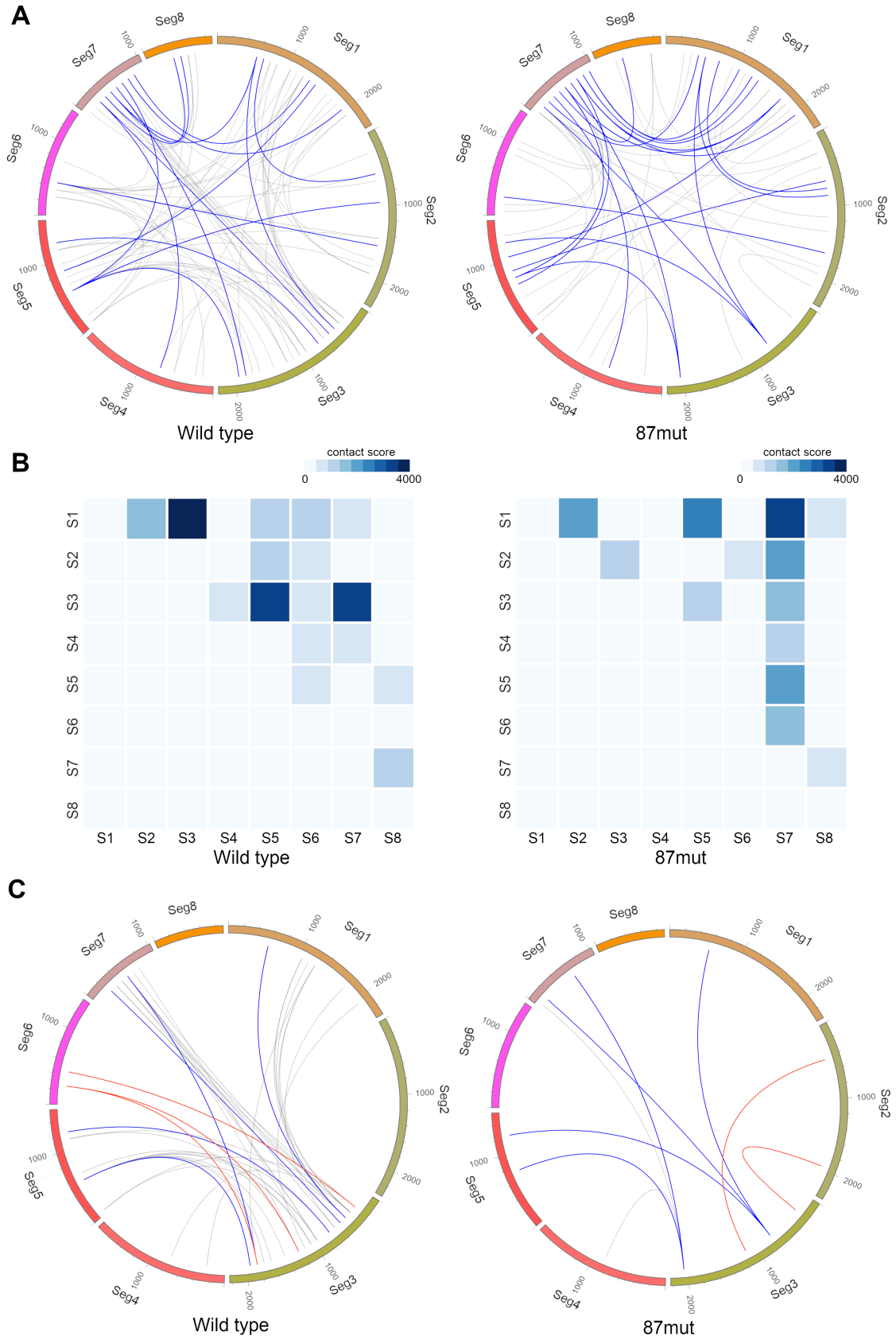
polymerases and NP and segment 4 encoding HA increased in the cells infected with the 87mut virus. If the lack of segment 3 does not affect the bundling of other segments, the co-packaging efficiency of these segments will not be altered. Thus, this result suggests that segment bundling was altered by the mutations in the distant pseudoknot region.

Rearrangement of the intersegment interaction of segment 3 in the 87mut virus

We showed that nucleotide positions 1 – 100 of segment 5 potentially interacted with segment 3 and the mutations that disrupted the pseudoknot structure of nucleotide positions 87 – 130 of segment 5 affected the replication and incorporation of segment 3 into the progeny virion. To analyze whether intersegment interactions were rearranged in the 87mut virus, we further performed LIGR-seq for the 87mut virus and obtained the intersegment interaction map.

First, we analyzed the intersegment interactions identified in both the wild type and the 87mut viruses. We set the allowable limit of error to 300 nt to identify the intersegment interactions in both viruses because vRNA was digested into a fragment that was approximately 300 nt long in the library preparation for large-scale sequencing. As a result, 16 interactions in the wild type virus (16/80, 20% of the total detected interactions) and 21 interactions in the 87mut virus (21/50, 42%) were overlapped with the interactions detected in another (Figure 6A, blue line). The bias to specific segments was not observed in the identified interactions for both the viruses. These results indicate that intersegment

interactions are partially captured both in the wild type and the 87mut viruses but the rearrangement of intersegment interactions occurs in the 87mut virus. Next, we assessed the intensity of the intersegment interactions in the wild type and the 87mut viruses. To analyze the intensity of intersegment interactions, the contact scores for all the interactions in each intersegment combinations were summarized in Figure 6B. The overall contact scores of segments 3 and 6, especially the contact scores between segments 3 and 1, 5, and 7, were reduced in the 87mut virus, while that of segment 7 was enhanced. When we focused on each intersegment interaction between segment 3 and other segments, we successfully identified 5 interactions in both the wild type and the 87mut viruses (Figure 6C). The contact scores of these preserved interactions between segments 3 and 1, 5, and 7 were, however, decreased in the 87mut virus, but the contact scores between these segments were decreased. In the 87mut virus, the interactions between nucleotide positions 1 – 100 of segment 5 and segment 3, segments 2 and 3, and segments 6 and 3 were not maintained, while a novel interaction was observed between segments 2 and 3 (Figure 6C). These results suggest that the intersegment interactions of segment 3 were reduced and partially rearranged by the conformational changes of the pseudoknot structure region in segment 5.


A
Seg1
Seg2
Seg3
Seg4
Seg5
Seg6
Seg7
Seg8
Wild type
87mut
B
contact score
0
4000
S1
S2
S3
S4
S5
S6
S7
S8
Wild type
87mut
C
Wild type
87mut

1

Figure 6. Reconstitution of the intersegment interactions in the 87mut virus. (A) intersegment interactions in the wild type and the 87mut virus. Intersegment interaction maps of the wild type and the 87mut virus were constructed from LIGR-seq. Intersegment interactions identified in both the wild type and the 87mut virus within a limit of error of 300 nt have been indicated by blue lines. (B) Intensity maps of the intersegment interactions in the wild type and the 87mut virus. Total contact scores of all the two-segment combinations from LIGR-seq were calculated. Since the data were symmetric, only the upper-triangular portion has been shown. (C) Intersegment interactions between segment 3 and other segments in the wild type and the 87mut virus. Intersegment interactions from segment 3 identified in both the wild type and the 87mut virus within a limit of error of 300 nt have been indicated by blue lines and interactions between segment 3 and other segments identified only in the wild type or the 87mut virus have been indicated by red lines.

Discussion

To reveal the global intra- and intersegment RNA interactions in the IAV virion, we obtained two RNA contact maps using protein crosslinking and RNA crosslinking methods. In our modified CLASH analysis, the segment arrangement that was previously observed by electron microscopy was not reconstituted (Figure 2). Two opposite-polarity NP strands in a double-helical conformation of the vRNP are connected by NP-NP interactions [3,4]. Thus, formaldehyde crosslinking may largely capture the interstrand NP connections in a double-helical conformation. Previous studies have shown that mutations of the conserved amino acids in NP affected segment bundling [29,30], but the NP interactions between segments have not been observed in our modified CLASH analysis. This result indicates that NP is involved in segment bundling indirectly rather than by direct NP interactions between different segments. Mutations in NP may alter its affinity to RNA and induce structural changes of the vRNP. Such structural changes of the vRNP may affect the intersegment interaction networks. In LIGR-seq analysis, the segment arrangement in the virion was, at least partially, captured (Figure 2). A recent study also revealed a redundant and complex networks of RNA-RNA interactions in the IAV by using other global RNA-RNA interaction detection method [34]. To identify specific intersegment interactions quantitatively, we used ICE normalization and IDR in our analysis. We showed that the intensities of interactions between segments were not equal (Figure 6B). Gao et al. showed that segments 1, 3, 5, and 7 played more important roles

than the other segments for viral genome packaging [12]. In our analysis, total contact sores of segments 1, 3, 5, and 7 were higher compared with those of the other segments (Figure 6B). These findings raise the possibility that the balance of segment interactions generates the hierarchy of segments for viral genome packaging. We also identified a redundant and complex intersegment network (Figures 2B and 6A). Redundancy of the intersegment interaction could provide the plasticity that tolerates the loss of some intersegment interactions. This plasticity may allow the IAV to escape an established immunity by mutations, while excess interactions may prevent the segment reassortment which also generate the diversity of IAV. Thus, the balance of intersegment interactions may be a factor for determining the diversity of IAV.

In the cells infected with the 87mut virus, the packaging of segment 3 were decreased, and bundling of the segments was affected by the mutations (Figure 5B and 5C). Nucleotide positions 1 – 100 of segment 5 interacted with segments 3, 6, and 8 in the wild type virus whereas these interactions were not identified in nucleotide positions 1 - 100 that interacted with segment 7 in the 87mut virus (Figure 6). This result suggests that unwinding of the pseudoknot structure induced a reconstitution of the intersegment interactions of segment 5. The total contact scores of segments 3, 5, and 6 in the 87mut virus were decreased and that of segment 7 was increased (Figure 6B), indicating that unwinding of the pseudoknot structure was able to affect the intersegment interactions of other segments. In particular, the total contact score of segment 3 was strongly decreased

in the 87mut virus. The intersegment interactions of segment 3 with other segments were equally decreased except for those between segments 3 and 2 (Figure 6C). One possible explanation of this finding is that segment 3 is eliminated from the center of the '7+1' vRNP arrangement in the 87mut virus by disruption of the pseudoknot structure of segment 5. Segments 5 and 6 would be located at the periphery, and thus, packaging of these segments is compensated by other intersegment interactions though the total intersegment interactions from these segments were decreased in the 87mut virus. The pseudoknot structure may be indirectly involved in segment bundling rather than involvement in the direct RNA interactions. The secondary structure of this pseudoknot region was considerably different between the wild type and the 87mut viruses (Figure 4C). These secondary structure changes could induce a reconstitution of the intersegment interactions as a result of the NP repositioning. Replication of segment 3 which did not have mutations was also reduced in the cells infected with the 87mut virus (Figure 5A). It is possible that repositioning of NP occurs in segment 3 by the effects of reconstitution of the intersegment interactions in the 87mut virus. Further analysis will be required to clarify the detailed molecular mechanism underlying the replication defect of segment 3 in the 87mut virus.

Overall, our study presents the global secondary structure and intersegment interactions of the IAV genome in the virion. We showed the formation of redundant and complex intersegment interactions and a pseudoknot structure in the vRNP. These

findings will help us to better understand not only the role(s) of RNA structures in segment bundling but also the molecular mechanisms underlying the emergence of potential pandemic IAV that is generated by segment reassortment. A redundant intersegment interaction is necessary for permitting mutations that disrupt one intersegment interaction, while too much intersegment interaction redundancy will prevent segment reassortment with different virus strains. Further exploration is required to understand the regulation of the formation of intersegment interactions. We also showed a functional pseudoknot structure on the vRNP. The finding further highlights the role(s) of the functional RNA structure in the IAV propagation and will contribute to develop a new class of anti-influenza drugs that bind and unwind the specific RNA structure in the IAV genome.

Materials and methods

Cells

MDCK cells were maintained in a minimal essential medium (MEM) (Sigma-Aldrich, ST. Louis, MO) containing 10% fetal bovine serum and penicillin/streptomycin (Nacalai Tesque, Kyoto, Japan). HEK293T cells (kindly provided by Dr. Y. Kawaoka) were maintained in a Dulbecco's modified Eagle's medium (DMEM) with high glucose concentration (Sigma-Aldrich) containing 10% fetal bovine serum and penicillin/streptomycin.

Viruses

Influenza virus A/PR/8/34 (H1N1) (PR8) was grown in the allantoic sacs of 11 days-old chick embryos at 35.5˚C for 48 h. Allantoic fluids of the infected eggs were collected and the virions were precipitated after stirring for 3 h at 4˚C in the presence of 8% (w/v) PEG6000 (Wako Pure Chemical Industries, Osaka, Japan). The precipitated virion was suspended in PBS(+) and purified by centrifugation on 30% to 60% sucrose gradients in PBS(+) at 100k × *g* for 1.5 h in an SW28 rotor at 4˚C. Viral bands were pooled and re-precipitated by centrifugation in PBS(+) at 120k × *g* for 1.5 h in an SW28 rotor at 4˚C. The precipitated virion was suspended in a Hepes buffer (10 mM Hepes-NaOH [pH 7.4], 100 mM NaCl, 1 mM DTT, 20% glycerol) and stored at -80˚C until use.

MDCK cells were infected with the recombinant virus at a multiplicity of

infection (MOI) of 0.1. At 48 h post infection (hpi) the supernatants were collected, and cell debris were removed by low-speed centrifugation (3k × *g*, 5 min). Virions in the pre-cleared supernatant were precipitated by centrifugation at 100k × *g* for 1.5 h in an SW32 rotor at 4˚C. The precipitated virion was suspended in PBS(-) and purified by centrifugation with 30% to 60% sucrose gradients in PBS(-) at 100k × *g* for 1.5 h in an SW55 rotor at 4˚C. Viral bands were pooled and re-precipitated by centrifugation in PBS(-) at 120k × *g* for 1.5 h in an SW55 rotor at 4˚C. The precipitated virion was suspended in a DMS buffer (40 mM Hepes-NaOH [pH 7.4], 100 mM NaCl, 0.5 mM $MgCl_2$) and stored at -80˚C until use.

Formaldehyde crosslinking and RNA ligation experimental method

The purified virion was treated with or without formaldehyde (final 0.1%) in PBS(-)-10% fetal bovine serum for 2 min at 25˚C. The crosslink was stopped by adding 1 M glycine and 4 volumes of Tris buffer (50 mM Tris-HCl [pH 7.9] and 100 mM NaCl). The crosslinked virion was purified by centrifugation at 130k × *g* for 1.5 h in an SW55 rotor at 4˚C through 30% sucrose. The virus pellet was suspended in a nuclease buffer (30 mM Sodium acetate [pH 4.6], 280 mM NaCl, 1 mM $ZnSO_4$, 0.25% Triton X-100, and 10 U RNase inhibitor [Toyobo, Osaka, Japan]). The viral RNA was digested with 180 U of the S1 Nuclease (Takara Bio, Otsu, Japan) for 10 min at 30˚C. The reaction was stopped by adding EDTA (final concentration 10 mM). Dynabeads-ProteinA (Thermo Fisher

Scientific, Waltham, MA) were washed and suspended in TNE buffer (20 mM Tris-HCl [pH 7.9], 100 mM NaCl, 1 mM EDTA, and 0.25% Triton X-100). One µg of the anti-NP monoclonal antibody mAb61A5 [52] was added to the Dynabeads-ProteinA solution and incubated under rotating conditions for 1 h at 4˚C. The beads were washed and resuspended in the TNE buffer, and the nuclease-treated virus solution was added to the beads. After incubation under rotating conditions for 1 h at 4˚C, the beads were washed three times with the TNE buffer and suspended in an RNA ligation buffer (50 mM Hepes-NaOH [pH 7.4], 10 mM $MgCl_2$, 1 mM DTT, 1 mM ATP, and 8 U RNase inhibitor [Toyobo]). Proximity ligation was performed with 40 U of the T4 RNA ligase I (Takara Bio) at 16˚C for 16 h and the beads were mixed in a Thermomixer (Eppendorf, Hamburg, Germany) at 1.5k rpm for 15 s every 5 min. The vRNA was eluted by adding 10 µg of Proteinase K at 37˚C for 15 min. After reverse-crosslink at 65˚C for 45 min, the vRNA was extracted with phenol/chloroform. The cDNA was synthesized with a random hexamer using the SuperScript III (Thermo Fisher Scientific) and the NEBNext mRNA Second Strand Synthesis Module (New England Biolabs, Ipswich, MA). Sequencing libraries were constructed using the KAPA Hyper Prep Kit (Kapa Biosystems, Wilmington, MA). Sequencing was performed using a HiSeq2500 sequencer (Illumina, San Diego, CA) (2 × 100-bp PE). The sequence data have been deposited in DDBJ Sequence Read Archive (DRA Accession: DRA005635)

AMT crosslinking and RNA ligation experimental method

The purified virion was treated with or without AMT (final concentration 100 μg/ml) and crosslinked using 365 nm UV for 20 min (1.56 J/cm^2). AMT was added again at 10 min after UV irradiation. The viral RNA was extracted with phenol/chloroform and digested with the NEBNext Magnesium RNA Fragmentation Module (New England Biolabs) at 94˚C for 2.5 min. The fragmented RNA was treated with CIAP (Takara Bio) at 37˚C for 30 min, and phosphorylation was carried out by the T4 polynucleotide kinase (Toyobo) at 37˚C for 1 h.

RNA pull-down procedure was performed using modification of a previously described method [53]. The biotinylated cDNA probes for the RNA pull-down were synthesized by PCR. cDNA of the PR8 strain was synthesized with the Uni12 primer using ReverTra Ace (Toyobo). The fragments of each segment were amplified with a linker sequence containing specific primers (5′- GTCAGTCAGTCAGTCAGTCA-(specific sequence)-3′ and 5′- GGATCCAATTGGCCAACTCG-(specific sequence)-3′) using Taq polymerase (Roche, Basel, Switzerland). Specific primers with a linker sequence were designed to amplify every 120 bp of the coding region with a 15-bp overlapped region. To synthesize the biotinylated cDNA probe, the amplified fragments of each segment were mixed and a second procedure of PCR was performed with the biotinylated linker primer (5′ biotin-GTCAGTCAGTCAGTCAGTCA-3′ and 5′-GGATCCAATTGGCCAACTCG-3′) using Taq polymerase. The biotinylated cDNA

probe for each segment was mixed with the Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific) and incubated on a rotating wheel at 37˚C for 1 h in a LiCl hybridization buffer (10 mM Tris-HCl [pH 7.9], 500 mM LiCl, 1 mM EDTA, and 0.1 % NP-40). The beads were washed with the LiCl hybridization buffer and 1x SSPE, and were suspended and incubated in 0.15 M NaOH for 10 min. After incubation, the mixture was neutralized with 100 mM Tris-HCl (pH 7.9) and 1.25 M AcOH. The beads were washed with 0.1 M NaOH and the LiCl hybridization buffer and were suspended in the LiCl hybridization buffer. The crosslinked RNA was boiled at 85˚C for 3 min and was added to the beads. The beads were incubated at 55˚C in a Thermomixer at 1,500 rpm for 2 h. The beads were washed with 1x and 0.1x SSPE at 55˚C. Proximity ligation was performed with 40 U of the T4 RNA ligase I at 16˚C for 16 h in an RNA ligase buffer and the beads were mixed in a Thermomixer at 1,000 rpm for 15 sec every 15 min. The vRNA was eluted with 5 U of DNase I (Takara Bio) at 37˚C for 30 min in a DNase buffer (40 mM Tris-HCl [pH 7.5], 8 mM $MgCl_2$, and 5 mM DTT) and was extracted with phenol/chloroform. The eluted vRNA was treated with RNaseR (Lucigen, Madison, WI) in an RNaseR buffer (20 mM Tris-HCl [pH 8.0], 100 mM KCl, and 0.1 mM $MgCl_2$) at 37 ˚C for 30 min and was extracted with phenol/chloroform. The cDNA was synthesized with a random hexamer using the SuperScript III and the NEBNext mRNA Second Strand Synthesis Module. Sequencing libraries were constructed using the KAPA Hyper Prep Kit. Sequencing was performed using a HiSeq2500 (2 × 100-bp PE). The sequence data have been deposited

in DDBJ Sequence Read Archive (DRA Accession: DRA005778, DRA009492, and DRA009493)

Data analysis for segment interactions

Raw reads were cleaned and trimmed into the first 25 bases with Trimmomatic v0.36 [54]. The cleaned reads were aligned to the A/PR/8/34 genome using bowtie2 with default parameters [55]. Obtained sequencing reads were classified into two categories: inter- and intrasegment interaction. We detected the former interaction by the paired end reads that were mapped at two different segments and the latter ones that were mapped to the same segment at an inverted direction and a long insert length (more than 500 nt in LIGR-seq and more than 750 nt in CLASH). The start positions of the selected pair-reads were counted in every 100 nt, and the contact map was constructed for the intra- and intersegment interactions. To normalize the biases, we adapted ICE normalization method developed for the Hi-C analysis [42]. The normalized count in each 100 nt bin was referred to as the contact score. To discriminate accurately between the true and false signals obtained in the duplicated experiments, we utilized IDR with a threshold value of 0.1 [44]. The segment arrangement was reconstituted with miniMDS [45].

DMS-seq and SHAPE-seq

NAI was synthesized using a previously described method [40]. One µl of DMS (Wako Pure

Chemical Industries) or 5 µl of NAI was added to the purified virion in 100 µl of DMS buffer. After incubation for 5 min (DMS) or for 15 min (NAI) at 25˚C, 10 µl of 1 M DTT was added to stop the reaction. Then, the vRNA was extracted with phenol/chloroform. Sequencing libraries were prepared using a previously described method [56]. Briefly, cDNA was synthesized with random hexamers containing the Illumina adapters at their 5′- ends using ReverTra Ace. The ssDNA linker containing a 5′ phosphate and 3′ C3 spacer was ligated to the synthesized cDNA using the Circligase I (Lucigen). The resultant cDNA was amplified by an adapter-based PCR using the KAPA HiFi DNA polymerase (Kapa Biosystems) and subjected to high-throughput sequencing using a MiSeq (Illumina) (2 × 75-bp PE). The sequence data have been deposited in DDBJ Sequence Read Archive (DRA Accession: DRA009494). Raw reads were cleaned and trimmed with Trimmomatic v0.36, and the cleaned reads were aligned to the A/PR/8/34 genome using bowtie2 with default parameters. The reactivity of each nucleotide was calculated using reactIDR [47] and BUMHMM [48] with a default setting. Computational prediction of the RNA secondary structure with pseudoknot structures was performed by IPknot [50].

Recombinant viruses

Viral protein expression vectors were generously provided by Dr. Y. Kawaoka [57] and the viral RNA expression vectors derived from the strain A/PR/8 were generously provided

by Dr. F. Momose [58]. To construct the pPolI-PR8 mutant vector, an inverted PCR was performed using the pPolI-PR8 segment 5 vector as a template with specific primer sets (5′-GCAGCGAGCCCGATTGTGCCTTCC-3′ and 5′-CTTTTCGTCCGAGAGTTCAAAGAC-3′). After DpnI treatment, phosphorylation, ligation, and transformation into an *Escherichia coli* Mach1 (Thermo Fisher Scientific) was performed. Recombinant viruses were generated using a reverse genetics approach [57]. Briefly, HEK293T cells were transfected with the viral protein and the viral RNA expression vectors using polyethylenimine (Polysciences, Warrington, PA). After 24 h of transfection, the culture medium was changed to the OPTI-MEM I (Thermo Fisher Scientific) containing 0.6 μg/ml TPCK-trypsin (Sigma-Aldrich). After incubation for 24 h, the cell culture supernatant was collected. The virus titer was determined by a plaque assay.

RT-qPCR

MDCK cells were infected with the virus at an MOI of 1. Total RNA was extracted using the ISOGEN reagent (Nippon Gene, Tokyo, Japan) at 8 hpi. For preparation of the vRNA in the supernatant from the infected cells, MDCK cells were infected with the virus at an MOI of 0.1, and the cells were suspended in MEM containing 0.6 μg/ml TPCK-trypsin. At 48 hpi, the supernatant was collected, and cell debris was removed by low speed centrifugation (500 × *g*, 5 min) and filtration through a 0.45-μm filter (EMD Millipore,

Billerica, MA). The RNA in the supernatant was extracted by phenol/chloroform. For RT-qPCR, the cDNA was synthesized with the Uni12 primer using ReverTra Ace (Toyobo). The synthesized cDNA was mixed with the Thunderbird SYBR qPCR mix (Toyobo) and a specific primer set for each segment. The qPCR reactions were performed using a Thermal Cycler Dice Real Time System TP800 (Takara Bio), and the relative amounts of each segment were calculated.

FACS analysis

Rabbit polyclonal antibodies against NP and M1 were generously provided by Dr. K. Nagata [59] and Dr. N. Kobayashi [60], respectively. Mouse monoclonal antibody RA5-22 against HA was provided by BEI Resources, NIAID, NIH (Bethesda, MD). Alexa Fluor 488-conjugated anti-mouse IgG and Alexa Fluor 647-conjugated anti-rabbit IgG were purchased from Thermo Fisher Scientific and BioLegend (San Diego, CA), respectively.

MDCK cells were infected with the virus at an MOI of 0.01. At 14 hpi, the infected cells were collected by trypsin and fixed with 4% paraformaldehyde at 25˚C for 10 min. The fixed cells were permeabilized with 0.2% NP-40 in PBS(-) at 25˚C for 15min. The cells were immersed in 0.2% BSA in PBS(-) at 25˚C for 1 h, and incubated with primary antibodies at 25˚C for 1 h. After washing with PBS(-), the cells were incubated with secondary antibodies at 25˚C for 1 h. The cells were suspended in PBS(-), and analyzed by the FACS Lyric Flow Cytometer (BD, Franklin Lakes, NJ)

Acknowledgements

We thank Dr. Yoshihiro Kawaoka (University of Tokyo) for kindly providing plasmids for reverse genetics system, Dr. Kyosuke Nagata (University of Tsukuba) for kindly providing polyclonal anti-NP antibody, Dr. Fumitaka Momose (Kitasato University) for kindly providing monoclonal anti-NP antibody and plasmids, Dr. Nobuyuki Kobayashi (Nagasaki University) for kindly providing anti-M1 antibody, and Ms. Yukiko Iwata for technical support of experiments.

References

1. Noda, T. *et al.* Importance of the 1+7 configuration of ribonucleoprotein complexes for influenza A virus genome packaging. *Nat. Commun.* **9**, 1–10 (2018).
2. Noda, T. *et al.* Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* **439**, 490–492 (2006).
3. Moeller, A., Kirchdoerfer, R. N., Potter, C. S., Carragher, B. & Wilson, I. A. Organization of the Influenza Virus Replication Machinery. *Science (80-.).* **338**, 1631–1634 (2012).
4. Arranz, R. *et al.* The Structure of Native Influenza Virion Ribonucleoproteins. *Science (80-.).* **338**, 1634–1637 (2012).
5. Gerber, M., Isel, C., Moules, V. & Marquet, R. Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol.* **22**, 446–455 (2014).
6. Hutchinson, E. C., Curran, M. D., Read, E. K., Gog, J. R. & Digard, P. Mutational analysis of cis-acting RNA signals in segment 7 of influenza A virus. *J. Virol.* **82**, 11869–11879 (2008).
7. Hutchinson, E. C., Wise, H. M., Kudryavtseva, K., Curran, M. D. & Digard, P. Characterisation of influenza A viruses with mutations in segment 5 packaging signals. *Vaccine* **27**, 6270–6275 (2009).
8. Marsh, G. A. *et al.* Highly Conserved Regions of Influenza A Virus Polymerase Gene Segments Are Critical for Efficient Viral RNA Packaging. *J. Virol.* **82**, 2295–2304 (2008).
9. Marsh, G. A., Hatami, R. & Palese, P. Specific residues of the influenza A virus hemagglutinin viral RNA are important for efficient packaging into budding virions. *J. Virol.* **81**, 9727–9736 (2007).
10. Muramoto, Y. *et al.* Hierarchy among viral RNA (vRNA) segments in their role in vRNA incorporation into influenza A virions. *J. Virol.* **80**, 2318–2325 (2006).
11. Zhao, L. *et al.* New Insights into the Nonconserved Noncoding Region of the Subtype-Determinant Hemagglutinin and Neuraminidase Segments of Influenza A Viruses. *J. Virol.* **88**, 11493–11503 (2014).
12. Gao, Q. *et al.* The influenza A virus PB2, PA, NP and M segments play a pivotal role during genome packaging. *J. Virol.* **86**, 7043–7051 (2012).
13. Gavazzi, C. *et al.* An in vitro network of intermolecular interactions between viral RNA segments of an avian H5N2 influenza A virus: comparison with a human H3N2 virus. *Nucleic Acids Res.* **41**, 1241–1254 (2013).
14. Gavazzi, C. *et al.* A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16604–16609 (2013).

15. Fournier, E. *et al.* Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine* **30**, 7359–7367 (2012).
16. Gilbertson, B. *et al.* Influenza NA and PB1 Gene Segments Interact during the Formation of Viral Progeny: Localization of the Binding Region within the PB1 Gene. *Viruses* **8**, 238 (2016).
17. Fournier, E. *et al.* A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res.* **40**, 2197–2209 (2012).
18. Noda, T. *et al.* Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat. Commun.* **3**, 639 (2012).
19. Rausch, J. W., Sztuba-Solinska, J. & Le Grice, S. F. J. Probing the structures of viral RNA regulatory elements with SHAPE and related methodologies. *Front. Microbiol.* **8**, 1–15 (2018).
20. Fodor, E., Pritlove, D. C. & Brownlee, G. G. The influenza virus panhandle is involved in the initiation of transcription. *J. Virol.* **68**, 4092–4096 (1994).
21. Hsu, M. T., Parvin, J. D., Gupta, S., Krystal, M. & Palese, P. Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. *Proc. Natl. Acad. Sci.* **84**, 8140–8144 (1987).
22. Gultyaev, A. P. *et al.* RNA structural constraints in the evolution of the influenza A virus genome NP segment. *RNA Biol.* **11**, 942–952 (2014).
23. Kobayashi, Y., Dadonaite, B., Doremalen, N. Van, Barclay, W. S. & Pybus, O. G. Computational and molecular analysis of conserved influenza A virus RNA secondary structures involved in infectious virion production. *RNA Biol.* **13**, 883–894 (2016).
24. Gultyaev, A. P. *et al.* Subtype-specific structural constraints in the evolution of influenza A virus hemagglutinin genes. *Sci. Rep.* **6**, 1–15 (2016).
25. Takizawa, N. *et al.* Local structural changes of the influenza A virus ribonucleoprotein complex by single mutations in the specific residues involved in efficient genome packaging. *Virology* **531**, 126–140 (2019).
26. Baudin, F., Bach, C., Cusack, S. & Ruigrok, R. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J.* **13**, 3158–3165 (1994).
27. Lee, N. *et al.* Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic Acids Res.* **45**, 8968–8977 (2017).
28. Williams, G. D. *et al.* Nucleotide resolution mapping of influenza A virus nucleoprotein-RNA interactions reveals RNA features required for replication. *Nat. Commun.* **9**, 465 (2018).
29. Bolte, H., Rosu, M. E., Hagelauer, E., García-Sastre, A. & Schwemmle, M. Packaging of the Influenza Virus Genome Is Governed by a Plastic Network of RNA- and Nucleoprotein-Mediated Interactions. *J. Virol.* **93**, 1–11 (2018).
30. Moreira, É. A. *et al.* A conserved influenza A virus nucleoprotein code controls

specific viral genome packaging. *Nat. Commun.* **7**, 12861 (2016).
31. Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267–1279 (2016).
32. Aw, J. G. A. *et al.* In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Mol. Cell* **62**, 603–617 (2016).
33. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B. J. Global Mapping of Human RNA-RNA Interactions. *Mol. Cell* **62**, 618–626 (2016).
34. Dadonaite, B. *et al.* The structure of the influenza A virus genome. *Nat. Microbiol.* 236620 (2019). doi:10.1038/s41564-019-0513-7
35. Sexton, A. N., Wang, P. Y., Rutenberg-Schoenberg, M. & Simon, M. D. Interpreting Reverse Transcriptase Termination and Mutation Events for Greater Insight into the Chemical Probing of RNA. *Biochemistry* **56**, 4713–4721 (2017).
36. Kudla, G., Granneman, S., Hahn, D., Beggs, J. D. & Tollervey, D. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc. Natl. Acad. Sci.* **108**, 10010–5 (2011).
37. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2013).
38. Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2013).
39. Wan, Y. *et al.* Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505**, 706–709 (2014).
40. Spitale, R. C. *et al.* RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9**, 18–20 (2013).
41. Lucks, J. B. *et al.* Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci.* **108**, 11063–11068 (2011).
42. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
43. Compans, R. W., Content, J. & Duesberg, P. H. Structure of the ribonucleoprotein of influenza virus. *J. Virol.* **10**, 795–800 (1972).
44. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
45. Rieber, L. & Mahony, S. miniMDS: 3D structural inference from high-resolution Hi-C data. *Bioinformatics* **33**, i261–i266 (2017).
46. Mitchell, D., Assmann, S. M. & Bevilacqua, P. C. Probing RNA structure in vivo. *Curr. Opin. Struct. Biol.* **59**, 151–158 (2019).
47. Kawaguchi, R., Kiryu, H., Iwakiri, J. & Sese, J. reactIDR: evaluation of the statistical reproducibility of high-throughput structural analyses towards a robust

RNA structure prediction. *BMC Bioinformatics* **20**, 130 (2019).
48. Selega, A., Sirocchi, C., Iosub, I., Granneman, S. & Sanguinetti, G. Robust statistical modeling improves sensitivity of high-throughput RNA structure probing experiments. *Nat. Methods* **14**, 83–89 (2017).
49. Ye, Q., Krug, R. M. & Tao, Y. J. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444**, 1078–1082 (2006).
50. Sato, K., Kato, Y., Hamada, M., Akutsu, T. & Asai, K. IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics* **27**, i85–i93 (2011).
51. Brooke, C. B. *et al.* Most Influenza A Virions Fail To Express at Least One Essential Viral Protein. *J. Virol.* **87**, 3155–3162 (2013).
52. Momose, F., Kikuchi, Y., Komase, K. & Morikawa, Y. Visualization of microtubule-mediated transport of influenza viral progeny ribonucleoprotein. *Microbes Infect.* **9**, 1422–1433 (2007).
53. Engreitz, J. M. *et al.* The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science (80-.).* **341**, 1237973–1237973 (2013).
54. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
55. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
56. Ding, Y., Kwok, C. K., Tang, Y., Bevilacqua, P. C. & Assmann, S. M. Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat. Protoc.* **10**, 1050–1066 (2015).
57. Neumann, G. *et al.* Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl. Acad. Sci.* **96**, 9345–9350 (1999).
58. Ohkura, T., Momose, F., Ichikawa, R., Takeuchi, K. & Morikawa, Y. Influenza A virus hemagglutinin and neuraminidase mutually accelerate their apical targeting through clustering of lipid rafts. *J. Virol.* **88**, 10039–10055 (2014).
59. Kawaguchi, A., Momose, F. & Nagata, K. Replication-coupled and host factor-mediated encapsidation of the influenza virus genome by viral nucleoprotein. *J. Virol.* **85**, 6197–6204 (2011).
60. Takizawa, N., Watanabe, K., Nouno, K., Kobayashi, N. & Nagata, K. Association of functional influenza viral proteins and RNAs with nuclear chromatin and sub-chromatin structure. *Microbes Infect.* **8**, 823–833 (2006).

Figure legends

Figure 1. Schematic representation of the methods for the identification of intra- and intersegment interactions. (A) Modified CLASH method. Purified virion was treated with formaldehyde to crosslink between protein and RNA and to crosslink between proteins. After partial RNA digestion by nuclease, vRNP was purified by immunoprecipitation, and proximal ligation was performed. (B) Modified LIGR-seq method. Purified virion was treated with AMT to crosslink between RNAs. After partial RNA digestion, vRNA was purified by a biotin-conjugated antisense single-strand DNA, and proximal ligation was performed.

Figure 2. Intra- and intersegment interaction map of segment RNAs. (A and B) Normalized RNA contact score of CLASH and LIGR-seq. Normalized contact scores were calculated from CLASH (A) and LIGR-seq (B) data and contact maps were constructed. The light and shade of colors in each bin represent low and high normalized contact scores. (C and D) Reconstitution of the viral genome arrangement from the contact scores. Segment arrangement was reconstituted from contact scores of CLASH (C) and LIGR-seq (D) using MiniMDS.

Figure 3. A pseudoknot structure at nucleotide positions 87–130 of segment 5 in the virion. (A) Distribution of reactivity scores from DMS-seq of the vRNA, vRNP, and

virion. Reactivity scores from DMS-seq were calculated by reactIDR and the scores of each segment were shown by violin plot. (B and C) The reactivity scores of segment 5 in the vRNA, vRNP, and virion. Reactivity scores were calculated by BUMHMM and a 10-nt moving average of the reactivity score of segment 5 from DMS-seq (B) and SHAPE-seq (C) were shown (upper panels). The lower panels show the reactivity scores of a specific region indicated by a black line in the upper panels. (D) The reactivity scores and predicted base pairs at nucleotide positions 87 – 130 of segment 5. The reactivity scores of each nucleotide from SHAPE-seq was calculated by reactIDR. Pink lines indicate predicted base pairs in the pseudoknot structure.

Figure 4. Unwinding the pseudoknot structure in the 87mut virus. (A) The pseudoknot structure of segment 5 and mutations in the 87mut virus. Orange-background nucleotides indicate the mutated nucleotides in the 87mut virus. (B) Virus propagation of the 87mut virus. MDCK cells were infected with the wild type or the 87mut virus at the indicated MOI. The supernatant was collected at indicated hours post infection (hpi) and the virus titer was determined by a plaque assay. The graph indicates average values with standard deviations from three independent experiments. P-values were calculated by the Student's *t*-test. (C) Relative reactivity scores at nucleotide positions 87 – 130 of segment 5. Reactivity scores from SHAPE-seq using the wild type and the 87mut virus were calculated by reactIDR. The relative reactivity scores were calculated by dividing the

reactivity score of each nucleotide by the highest reactivity score in the region. Orange-background nucleotides indicate the mutated nucleotides in the 87mut virus and pink lines indicate the predicted base pairs in the wild type sequence.

Figure 5. Impairment of replication and packaging of segment 3 and alteration of the segment bundling pattern in the 87mut virus. (A and B) Relative vRNA amount in the infected cells (A) and in the virion (B). The amount of each segment was determined by RT-qPCR and the relative amount was calculated by double-normalization to the wild type virus and segment 5. The graph indicates average values with standard deviations from three independent experiments. P-values were calculated by the Student's *t*-test. (C) Co-expression rate of two viral proteins in cells infected with the wild type or the 87mut virus. Ratio of the co-stained cells were determined by FACS. The graph indicates average values with standard deviations from three independent experiments. P-values were calculated by the Student's *t*-test.

Figure 6. Reconstitution of the intersegment interactions in the 87mut virus. (A) intersegment interactions in the wild type and the 87mut virus. Intersegment interaction maps of the wild type and the 87mut virus were constructed from LIGR-seq. Intersegment interactions identified in both the wild type and the 87mut virus within a limit of error of 300 nt have been indicated by blue lines. (B) Intensity maps of the intersegment

interactions in the wild type and the 87mut virus. Total contact scores of all the two-segment combinations from LIGR-seq were calculated. Since the data were symmetric, only the upper-triangular portion has been shown. (C) Intersegment interactions between segment 3 and other segments in the wild type and the 87mut virus. Intersegment interactions from segment 3 identified in both the wild type and the 87mut virus within a limit of error of 300 nt have been indicated by blue lines and interactions between segment 3 and other segments identified only in the wild type or the 87mut virus have been indicated by red lines.