

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Institut de Biologie Moléculaire et Cellulaire

UPR9002 : Architecture et Réactivité de l'ARN

THÈSE présentée par :

Rithu PAUL STANSILAU

soutenue le : 6 Septembre 2024

pour obtenir le grade de : Docteur de l'université de Strasbourg

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

**Impact du réassortiment génétique sur la
structure des ARN des virus Influenza A**

THÈSE dirigée par :

Dr Roland MARQUET

IBMC, UPR9002, Strasbourg

THÈSE codirigée par :

MC Laurence DESPONS

IBMC, UPR9002, Strasbourg

RAPPORTEURS :

Dr Laura GRAF

Institute of Virology, Uniklinik Freiburg

Dr Thibaut CRÉPIN

Institut de Biologie Structurale, Grenoble

EXAMINATEURS :

Dr Catherine ISEL-GRIFFITHS

Institut Pasteur, Paris

Pr David GILMER

IBMP, Strasbourg

Rithu PAUL STANSILAU

Impact du réassortiment génétique sur la structure des ARN des virus Influenza A

Les virus de la grippe A (IAV) sont connus pour leurs épidémies annuelles et leurs pandémies occasionnelles. Le génome segmenté de l'IAV facilite le réassortiment génétique, permettant aux virus d'échanger des segments entiers lors de la co-infection avec des souches génétiquement homologues. Cela accélère l'évolution de l'IAV en générant des souches pandémiques avec de nouvelles combinaisons de segments pouvant échapper aux réponses immunitaires adaptatives existantes. Cependant, le réassortiment n'est pas aléatoire et est contraint par le mécanisme sélectif d'encapsidation du génome de l'IAV. Des signaux d'encapsidation spécifiques à chaque segment interagissent pour garantir qu'au moins une copie de chaque segment est sélectivement incorporée dans un virion. Ces interactions intersegmentaires influencent la fréquence de réassortiment. En étudiant les mécanismes moléculaires de ce processus, nous espérons développer des cibles potentielles pour la recherche antivirale. Nous avons généré quatre virus par génétique inverse 7:1 - H1N1/PR8, H3N2/MO et 2 réassortants à segment unique R1 (PR8_PB2_{MO}) et R6 (PR8_NA_{MO}). La méthodologie SHAPE-MaP a été utilisée pour déterminer les structures secondaires de l'ARN de ces virus dans deux états différents - intacts et désassemblés. Nous avons utilisé vRNAsite pour générer des prédictions d'appariements entre tous les segments des virus. Nous avons couplé nos données de SHAPE-MaP avec les prédictions informatiques de chaque virus pour identifier des régions avec des implications potentielles pour l'encapsidation du génome et le réassortiment génétique de l'IAV.

Mots-clés : Grippe, ARN, SHAPE-MaP, intact, désassemblé, structures, prédictions

Traduit avec l'aide de ChatGPT, OpenAI

Influenza A viruses (IAV) are notorious for annual epidemics and occasional pandemics. The segmented genome of IAV facilitates genetic reassortment, whereby viruses exchange full-length segments during co-infection with genetically homologous strains. This accelerates IAV evolution by generating pandemic strains with novel segment combinations that can evade existing adaptive immune responses. However, reassortment is not random and is constrained by the selective genome packaging mechanism of IAV. Segment-specific packaging signals on each segment interact to ensure that at least one copy of each segment is selectively packaged into a virion. These intersegmental interactions influence the reassortment frequency. By studying the molecular drivers of this process, we aim at developing potential targets for antiviral research. We generated four viruses by 7:1 reverse genetics - H1N1/PR8, H3N2/MO and two single-segment reassortants R1 (PR8_PB2_{MO}) and R6 (PR8_NA_{MO}). SHAPE-MaP was used to determine the RNA secondary structures of these viruses in 2 different states - intact and disassembled. We used vRNAsite to generate pairwise predictions of interactions between all segments of the viruses. We coupled our SHAPE-MaP data with the computational predictions of each virus to generate regions with potential implications for genome packaging and genetic reassortment in IAV.

Keywords : Influenza, RNA, SHAPE-MaP, intact, disassembled, structures, predictions

"I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale."

- Maria Skłodowska Curie

Acknowledgements

This work wouldn't have been possible without some brilliant and good people. I'd like to take a moment to acknowledge their efforts here.

This research was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Actions Innovative Training Networks grant agreement no. 955974 (VIROINF). I am grateful to have been a fellow member of this consortium that successfully brought together many enterprising scientists. I am also thankful to Manja, our VIROINF 'mama' at the FSU, Jena. Her initiative and enthusiasm were fuel for our energy. Her team made sure I had a productive and pleasant time during my secondment at Jena. I was constantly inspired by Gabriel, Nikolas, Óscar, Harshit, Yasas, Lara, Dan, Xue, Lovro, Liuwei, Ali, Mikhail, Swenja, Lygeri, and Berta. We kept each other motivated over the three years to put our best foot forward and become better researchers. I will always cherish the times we explored islands and played laser tag, bowling and darts, all while discussing science. It was truly a once-in-a-lifetime opportunity, and I wish them all the very best, in what I'm sure, is a very bright future ahead for each of them.

It goes without saying that Team Marquet-Paillart at IBMC in Strasbourg played the biggest role in helping me execute this work. Roland was exceptionally supportive of my work, giving me ample freedom to take the onus of this work and execute it how I wanted to. His expertise and guidance on RNA biology helped steer my work towards fruition. Thanks to Roland, I have now begun viewing RNA as an independent entity, unlike when I started out, as I was a pure virologist. I have to thank the rest of UPR 9002 - Jean-Christophe, Anne-Caroline, Antoine, Valerie, Serena, Orian, Cedric, Erwan, Lisa, Aishwarya, Florine, Mathieu, and Cassandre. Anne-Caroline showed me around the lab when I first arrived and played a huge role in helping me understand the lab's day-to-day. JC kept the fun element of the lab high. I could always rely on Antoine to discuss the stats and math of RNA structural biology. Orian, Cedric and Erwan were fellow PhDs who I could rant with. The occasional visits from our exchange students Suresh and others, also helped with the team's morale. That said, I am glad to have been a part of this team for 3 years.

An advantage of the VIROINF consortium was that I got to collaborate with some brilliant minds. Prof. Dr. Martin Schwemmle helmed one such team at the Institute of Virology, Freiburg. Martin, Hardin, and Celia played significant roles in helping me piece together the 'virology part' of my work and the remaining RNA biology. Martin's scientific rigour and enthusiasm were always motivating. Hardin is an exceptionally admirable scientist, and I am grateful for the time he gave me to answer my naïve questions about packaging when it was not a part of his job description. His attention to detail and precision were impeccable, and I don't think I could have asked for a better scientist to learn packaging from. Celia went from being a motivating colleague I could look up to, to a very dear friend. It was nice to learn and

discuss packaging with her, and she kept me grounded with her feisty nature. I had one of the most fun weekends with Celia and Philipp, and I look forward to hanging out with them more often. Thanks to AG Schwemmler, I made friends in Freiburg - Kiki, Precious, Deepti, John, Stella, Julian, Quinn, Lina, Kevin, Peter, and Sebastian. They were extremely supportive of my work, showed me some fun times while in Freiburg, and nurtured me through rough times.

Just like the IAV RNA is held together by the NP protein, the teams at Strasbourg and Freiburg were held together by our very smart technicians Anne and Kerstin respectively. While trying to navigate work simultaneously in Strasbourg and Freiburg, Anne (may her soul rest peacefully) made sure my work in Strasbourg remained above reproach. Likewise, while setting up virus handling in Strasbourg, Kerstin was more than happy to help me with protocols and ensured I always had them right.

The bioinformatics portion of this work wouldn't have been possible without Dr. Daniel Desirò and Beatrice Chane-Woon Ming. They were patient and kind in offering their time while being extremely busy. Their expertise helped me gather data to understand my work better and create cool figures for my thesis. They were always available to support our team and helped me channel my work to better standards by customising the tools to make them biologically coherent.

I would also like to acknowledge the following people for their contributions to my work. A) My thesis committee members – Catherine, David and Laurence, who patiently listened to my challenges during PhD, gave me expert opinions and actionable advice on tackling challenges and improving my work's overall quality. B) Dr Brad Gilbertson at the University of Melbourne, Australia, for providing the pHW2000 plasmids. C) Dr Jonas Fuchs at the Institute of Virology, Freiburg, Germany, for technical assistance with the sequencing of virus stocks. D) Melanie, Lea, and the rest of team Pfeffer at IBMC for their support in helping me establish a Flu-handling setup at IBMC. E) Armelle, Responsable Marchés Pharma & Coating, JRS PHARMA, France for providing Avicel. F) Dr Patrick Irving, University of North Carolina, for providing expert opinion on SHAPE-MaP.

Roberto, Max, Riccardo, Zina, Alessandro, Raquel, Wacym, Enis, Erica, Aishwarya, Suresh, Lea, Gaby, Laura, Lisanne, Guillermo, Marcella, Ambre and Roxanne ensured my social life did not take a backseat during PhD. They welcomed me into their circle and ensured I stayed happy during the last three years. Thanks to them, I have a lifetime of good memories in Europe.

I don't think I would have made it through this experience as gracefully as I did without Nimmy and Bhargavi. They were always available for me, patient with me, supported me, and reminded me of my strengths when I couldn't do it for myself.

Finally, I owe it all to my family—Papa, Amma, and my sisters. Papa and Amma have been my biggest pillars of support. They follow me through my dreams while remaining my compass and anchor. On the other hand, my sisters keep my life entertaining while ensuring my wardrobe and wallet remain empty. This milestone will forever remain a feather in *their* cap.

Table of Contents

PREFACE	13
INTRODUCTION	15
GENETIC REASSORTMENT EXPEDITES EVOLUTION IN SEGMENTED RNA VIRUSES	16
<i>Antigenic shift – A direct consequence of reassortment</i>	17
<i>Determinants of genetic reassortment</i>	19
INFLUENZA VIRUSES	21
<i>General classification and nomenclature</i>	21
<i>Host tropism and transmission patterns</i>	22
<i>Antigenic drift and Flu vaccines</i>	23
INFLUENZA A VIRUSES – STRUCTURE AND BIOLOGY	25
<i>Virion structure</i>	25
<i>Genome structure</i>	26
<i>Viral nucleocapsid protein</i>	28
<i>IAV life cycle – from entry to egress</i>	29
SELECTIVE GENOME PACKAGING IN SEGMENTED VIRUSES	38
<i>Viral nucleocapsid proteins and genomic RNA self-assemble into supramolecular complexes</i>	38
<i>Packaging signals mediate selective genome assembly pathways in segmented RNA viruses</i>	40
REVIEW – THE INFLUENZA A VIRUS GENOME PACKAGING NETWORK – COMPLEX, FLEXIBLE AND YET UNSOLVED	43
KNOWLEDGE GAPS	59
RNA STRUCTURAL BIOLOGY AND CHEMICAL PROBING	61
<i>RNA structure and its implications on virus regulation</i>	61
<i>RNA structural probing strategies</i>	63
<i>Selective 2'-hydroxyl acylation analysed by primer extension - mutational profiling (SHAPE-MaP)</i>	65
<i>Using structural data from SHAPE-MaP to inform functional RNA elements</i>	69
COMPUTATIONAL PREDICTIONS OF RNA SECONDARY STRUCTURES	72
<i>Secondary structure prediction tools</i>	72
<i>RNA Structuromics: Optimising RNA secondary structure prediction with chemical probing data</i>	73
HYPOTHESIS & OBJECTIVES	77
MATERIALS AND METHODS	79
CELLS, PLASMIDS, AND VIRUSES	80
<i>Cell culture</i>	80
<i>Plasmid amplification</i>	80
<i>In vitro transcription to produce naked viral RNA</i>	80
<i>Virus production through 7:1 reverse genetics</i>	81
<i>Virus stock production</i>	82

<i>Growth kinetics</i>	82
<i>Plaque assays for virus titre determination</i>	82
<i>Immunofluorescence assay</i>	82
COMPETITION EXPERIMENTS	83
<i>Mutagenesis of HA_{PR8} and production of PR8_HA_{tag} virus</i>	83
<i>Single cycle replication kinetics</i>	83
<i>Competition co-infections</i>	84
CHEMICAL PROBING AND COMPUTATIONAL BIOLOGY	84
<i>SHAPE-MaP</i>	84
<i>Shapemapper2</i>	86
<i>ΔSHAPE</i>	87
<i>vRNAsite</i>	87
RESULTS & DISCUSSION	89
IN VITRO 7:1 REVERSE GENETICS	90
<i>Production of PR8, MO and eight single-segment reassortants (R1 – R8)</i>	90
<i>Two candidate single-segment reassortants R1 (PR8_PB2_{MO}) and R6 (PR8_NA_{MO}) were chosen for SHAPE-MaP</i>	91
<i>Single cycle replication kinetics showed R1 had slightly better in vitro viral fitness</i>	93
<i>Protein incompatibilities may have caused attenuation of the remaining six reassortants</i>	94
<i>Summary and limitations of the 7:1 reverse genetics system</i>	96
COINFECTION WITH TWO COMPETING VIRUSES	98
<i>Introduction of a mutational tag in the HA segment of PR8</i>	98
<i>Primer optimisation for segment-specific amplification</i>	98
<i>Control coinfection A. PR8_HA_{tag} outperformed PR8</i>	99
<i>Coinfection B. Heterologous coinfection of MDCK.2 with PR8_HA_{tag} and R1</i>	101
<i>Coinfection C. Heterologous coinfection of MDCK.2 with PR8_HA_{tag} and R6</i>	102
SHAPE-MAP	104
<i>Chemical probing on intact and disassembled particles</i>	104
<i>Shapemapper2 produced reactivity profiles for each segment</i>	105
<i>Custom ΔSHAPE parameters were used to compare intact and disassembled particles</i>	113
<i>vRNPs of intact particles showed higher SHAPE reactivities than their disassembled counterparts</i>	113
<i>ΔSHAPE patterns of R1 and R6</i>	115
<i>Caveats of SHAPE-MaP setup</i>	116
VRNASITE	118
CONCLUSIONS & PROSPECTS	123
REFERENCES	127
APPENDICES	147
RÉSUMÉ EN FRANÇAIS	159

Table of Figures

Figure 1: Antigenic shift and drift in Influenza A viruses	17
Figure 2: HA-mediated cellular receptor binding of Influenza A viruses	18
Figure 3: Genetic reassortment in IAV	18
Figure 4: ICTV classification of IAV	21
Figure 5: Structure of Influenza A virus	25
Figure 6: Structure of viral ribonucleoprotein (vRNP)	26
Figure 7: Genome structure of H1N1/PR8	27
Figure 8: Overview of IAV replication	29
Figure 9: Conformational changes of HA during pH-induced membrane fusion	30
Figure 10: Transcription and replication of Influenza genome	32
Figure 11: Schematic of transcription cycle	33
Figure 12: Regulation of IAV RNA polymerase through interactions with molecular cues and proposed model for influenza virus RNA genome replication	35
Figure 13: Model for trafficking of vRNPs across the cytoplasm in an IAV-infected cell	37
Figure 14: Segmented viruses package their genomes randomly or using segment-specific packaging signals	38
Figure 15: RNA secondary structure of PR8_PB2 segment created by RNAstructure	62
Figure 16: Enzymes and chemical probes used to study RNA secondary structures	64
Figure 17: Overview of the SHAPE-MaP assay	66
Figure 18: Immunofluorescence staining of MDCK.2 infected with PR8, MO, R1 and R6	90
Figure 19: Titres of virus supernatants following 7:1 reverse genetics in HEK 293T cells	90
Figure 20: Virus titres of PR8, MO and the eight reassortants	91
Figure 21: Growth kinetics of PR8, MO, and the eight reassortants (R1 - R8)	92
Figure 22: Single-cycle infection of PR8, R1 and R6	93
Figure 23: Competitive coinfection assays	100
Figure 24: Virus titres during SHAPE-MaP protocol	104
Figure 25: Reactivity profile produced by Shapemapper2	105
Figure 26: Circos plot showing SHAPE-MaP reactivities of all segments of PR8/H1N1 virus	106
Figure 27: Circos plot showing SHAPE-MaP reactivities of all segments of R1 virus	107
Figure 28: Circos plot showing SHAPE-MaP reactivities of all segments of R6 virus	108
Figure 29: Median SHAPE-MaP reactivities of vRNP segments one through eight of intact and disassembled PR8	110
Figure 30: Median SHAPE-MaP reactivities of vRNP segments one through eight of intact and disassembled R1	111
Figure 31: Median SHAPE-MaP reactivities of vRNP segments one through eight of intact and disassembled R6	112
Figure 32: Comparison of negative and positive Δ SHAPE regions	114
Figure 33: Nucleotide distribution among negative and positive Δ SHAPE regions	116
Figure 34: Circos plots generated by vRNAsite for PR8 (a), MO (b), R1 (c), R6 (d)	120
Figure 35: Venn diagram showing all the positive Δ SHAPE regions predicted by vRNAsite that are shared by viruses	121
Figure 36: Venn diagram showing all the negative Δ SHAPE regions predicted by vRNAsite that are shared by viruses	122

Table Of Supplementary Figures

<i>Supplementary fig 1: 7:2 reverse genetics and coinfection settings</i>	148
<i>Supplementary fig 2: Competition coinfection strategy</i>	149
<i>Supplementary fig 3: SHAPE reactivity profiles of intact (a) and disassembled (b) PR8 virus</i>	150
<i>Supplementary fig 4: SHAPE reactivity profiles of intact (a) and disassembled (b) R1 virus</i>	151
<i>Supplementary fig 5: SHAPE reactivity profiles of intact (a) and disassembled (b) R6 virus</i>	152
<i>Supplementary fig 6: ΔSHAPE results of segments one through four of PR8</i>	153
<i>Supplementary fig 7: ΔSHAPE results of segments five through eight of PR8</i>	154
<i>Supplementary fig 8: ΔSHAPE results of segments one through four of R1</i>	155
<i>Supplementary fig 9: ΔSHAPE results of segments five through eight of R1</i>	156
<i>Supplementary fig 10: ΔSHAPE results of segments one through four of R6</i>	157
<i>Supplementary fig 11: ΔSHAPE results of segments five through right of R6</i>	158

Abbreviations/Acronyms

1M7	1-Methyl-7-nitroisatoic anhydride
ATCC	American Type Cell culture
BTV	Bluetongue virus
cDNA	Complementary deoxyribonucleic acid
CMCT	1-cyclohexyl (2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate
COVID-19	2019 Coronavirus disease
DIPs	Defective interfering particles
DMEM	Dulbecco's modified minimum essential medium
DMS	Dimethyl sulphate
DNA	Deoxyribonucleic acid
ds	Double stranded
HA	Haemagglutinin
HEK	human embryonic kidney
HPAI	Highly pathogenic avian Influenza
hpi	hours post infection
hpt	hours post transfection
IAV	Influenza A virus
IBVs	Influenza B viruses
ICTV	International Committee on Taxonomy of Viruses
ICVs	Influenza C viruses
IFA	Immunoflourescence assay
IGBMC	Institut de génétique et de biologie moléculaire et cellulaire
IPS	Internal packaging signals
IRIs	intermolecular RNA-RNA long-range interactions
IVTs	<i>In vitro</i> transcripts
LPAI	Low pathogenic avian influenza
M	Matrix
MDCK	Madin Darby canine kidney
MO	Influenza A/Moscow/10/99/H3N2
NA	Neuraminidase
Neu5Ac	5-N-acetylneuraminic acid
NGS	Next generation sequencing
NP	Nucleoprotein
NS	Non-structural
PA	Polymerase acidic subunit
PAGE	polyacrylamide gel electrophoresis

PB1	Polymerase basic subunit 1
PB2	Polymerase basic subunit 2
PCR	Polymerase chain reaction
pdm	pandemic
PR8	Influenza A/Puerto Rico/8/1934/H1N1
PS	Packaging signals
R1	Reassortant 1 (PR8_PB2 _{MO})
R2	Reassortant 2 (PR8_PB1 _{MO})
R3	Reassortant 3 (PR8_PA _{MO})
R4	Reassortant 4 (PR8_HA _{MO})
R5	Reassortant 5 (PR8_NP _{MO})
R6	Reassortant 6 (PR8_NA _{MO})
R7	Reassortant 7 (PR8_M _{MO})
R8	Reassortant 8 (PR8_NS _{MO})
RdRp	RNA dependent RNA polymerase
RNA	Ribonucleic acid
RNase	Ribonuclease
SEC	Size exclusion chromatography
SHAPE-MaP	Selective 2'-hydroxyl acylation analysed by primer extension – mutational profiling
SPLASH	Sequencing of Psoralen crosslinked, ligated, and selected Hybrids
ss	Single stranded
TPCK	N-tosyl-L-phenylalanine chloromethyl ketone
TPS	Terminal packaging signals
UC	ultracentrifugation
UTR	Untranslated regions
vRNP	viral ribonucleoprotein

Preface

Over the last two decades, pandemics and epidemics caused by SARS-CoV-2 (COVID-19), HIV-1, Zika virus, Influenza virus, Monkeypox virus, and Ebola virus have burdened global healthcare infrastructure, the economy and socio-political environment^{1,2,3,4}. The incidence of emerging and re-emerging pathogens and their consequent clinical burden has also increased, necessitating better pandemic preparedness programs. A key aspect of such programs to combat future pandemics is a thorough understanding of patterns of virus evolution. Studying interspecies transmission and host adaptation factors that enable virus evolution is essential to mitigate future pandemics.

RNA viruses are known to evolve at faster rates than DNA viruses. Occasionally, this can lead to the emergence of novel phenotypic variants capable of evading pre-existing adaptive immunity in human populations. As evidenced by the examples stated before, this can result in pandemics with significant morbidity and mortality. Consequently, the molecular mechanisms of acquiring novel mutations that enhance virus fitness and enable cross-species transmission are often the starting point to understanding genetic variability and virus evolution.

RNA viruses with segmented genomes use genetic reassortment as an auxiliary evolutionary mechanism. Multiple subtypes of a virus can infect susceptible natural hosts concurrently. Such infections produce novel progeny virions with genome segments from both parental strains. This accelerates genetic variability in reassortants by facilitating the acquisition of multiple favourable mutations that promote cross-species transmission and the emergence of novel strains over a single reassortment event. Such novel reassortants are antigenically distinct from their parental strains and thus enhance disease severity when introduced into an immunologically naïve population.

Although reassortment has been shown for many segmented viruses such as Bluetongue, Influenza, Rotavirus and even some multipartite viruses, the underlying molecular mechanism remains largely unknown. Sometimes, reassortants produced may be of suboptimal virulence or even non-infectious due to genetic and antigenic mismatches conferred by the novel genome constellation. Consequently, reassortment events that successfully produce novel strains with pandemic potential are rare. This implies reassortment is not random as the nature of interacting genome segments constrains it. As a result, some of these viruses use a selective mechanism to package their genome. This packaging mechanism selectively incorporates compatible genome segments into a budding virion. Therefore, understanding the rules of coordinated genome packaging paves the way for developing pandemic prediction tools, targeted antivirals, and effective vaccine strategies to combat virus evolution.

Influenza A virus (IAV) is an enveloped, negative-sense RNA virus. Its genome is split into eight segments, each coding for an essential viral protein⁵. Therefore, it is crucial to selectively package one copy of each segment to produce an infectious virion during the virus life cycle. Consequently, during genome packaging in the cytosol, the segments assemble progressively into an octameric complex with seven segments surrounding one segment. The synergistic effect of intersegmental RNA-RNA interactions and RNA-protein interactions maintains the conformation of this octameric supramolecular complex. Extensive studies have identified regions that could be involved in forming and maintaining this octameric complex. Other high-throughput studies also suggested that IAV has a redundant genome packaging mechanism with a flexible network of intersegmental RNA-RNA interactions⁶. While this implies an unfailingly rigorous selection mechanism during genome packaging, it also suggests that the network has evolved to withstand any negative pressure that could potentially impair virus fitness.

In this study, we have used Selective 2'-hydroxyl acylation analysed by primer extension-mutational profiling (SHAPE-MaP)⁷ to study the extent of RNA secondary structures on the octameric complex formation in IAV. We used two alternative states of the virus – intact and disassembled to map potential regions that could be involved in intersegmental RNA-RNA interactions. We also used pairwise computational predictions to support the experimental SHAPE-MaP data in mapping potential regions crucial to genome packaging.

Introduction

Genetic reassortment expedites evolution in segmented RNA viruses

Virus evolution is facilitated by acquiring novel beneficial mutations that enable interspecies transmission and enhanced virulence. While mutation rates vary across different viruses, RNA viruses have higher mutation rates ranging between 10^{-6} and 10^{-4} substitutions per nucleotide site⁸. This is due to the lack of an exonuclease proofreading activity of the virus-encoded RdRp, which occasionally results in a gain of function⁹. Such high mutation rates are favoured since they confer a greater adaptive capacity to the viruses. RNA is also a structurally dynamic biomolecule with different effector functions regulated by the nature of its secondary and tertiary structural conformations. For instance, the highly conserved stem loops in the 5' UTR of the HIV-1 genome regulate vital steps such as RNA transcription, polyadenylation, splicing, translation, reverse transcription and packaging¹⁰. Consequently, mutations that alter such structures would impact the corresponding effector functions, either exacerbating the effect or impairing virus propagation.

Usually, random beneficial RNA mutations are acquired incrementally over multiple generations. Occasionally, horizontal transfer of genome elements between viruses of different lineages also facilitates evolution¹¹. This phenomenon is often seen in segmented viruses where RNA genomes exist as several distinct molecules. Concurrent infection of susceptible hosts with two genetically distant subtypes of a segmented virus facilitates genetic reassortment. The resulting progeny virions contain genome segments from both parental strains of infecting viruses. It is a unique evolutionary mechanism of segmented RNA viruses and is essential in novel virus emergence and interspecies transmission^{12,13,14}. Often, reassortment accelerates virus evolution when compared to incremental mutation events. This is due to the acquisition of multiple mutations through a single reassortment event where one or more segments can be swapped into a progeny virus. As a result, segmented RNA viruses exhibit a higher rate of evolution than non-segmented viruses. Therefore, the study of reassortment is vital to detect novel reassortants with increased potency for circulation in an immunologically naïve population and to study reassortants that could escape antibody recognition or are resistant to antivirals¹².

Another concern is the development of antiviral resistance due to the acquisition of resistant genes through genetic reassortment, as described in IAVs. This was observed in adamantane resistance gained by the presence of a non-cognate yet compatible segment of the M1 protein. Similarly, novel drug-resistant mutations may be acquired through reassortment. The emergence of amantadine-resistant H3N2 viruses and oseltamivir-resistant strains are classic examples of reassortment-induced drug-resistance^{15,16}.

Some of the segmented virus families are *Orthomyxoviridae*, *Reoviridae*, *Bromoviridae*, *Cystoviridae*, *Picobirnaviridae*, *Arenaviridae*, *Birnaviridae*, *Bunyaviridae*, *Chrysoviridae*, *Closteroviridae*, and *Partitiviridae*. Many viruses in these families are associated directly or indirectly with human diseases that cause significant morbidity and mortality. While reassortment has been described in many of these viruses, it has been studied extensively in IAVs as a key mechanism for interspecies transmission and the emergence of pandemic strains¹².

Antigenic shift – A direct consequence of reassortment

The phenomenon in which reassortment generates a new epitope-altering virus variant is called antigenic shift (Fig 1). This enables novel variants to circumvent preexisting neutralising antibodies produced actively by natural infection of hosts or obtained passively through

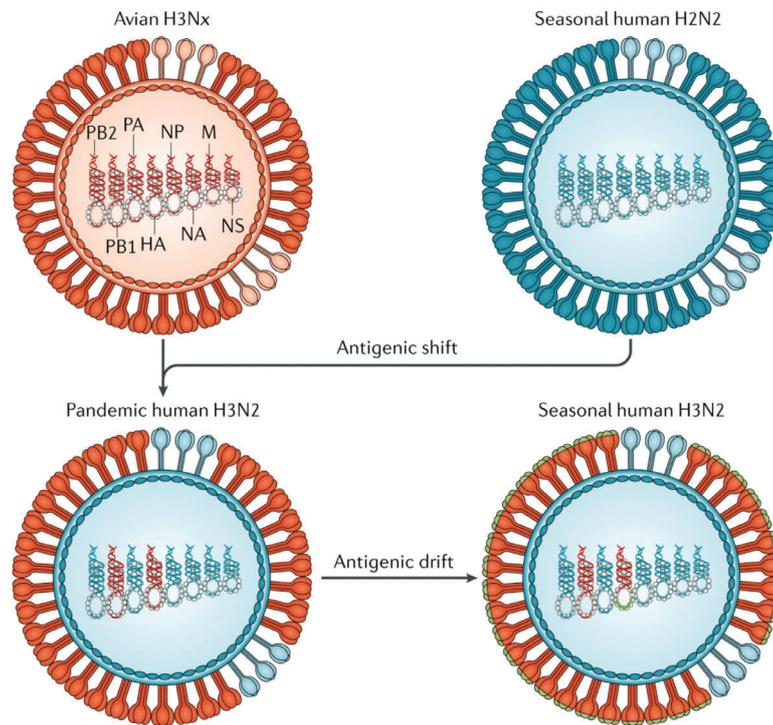


Fig 1 | Antigenic shift and drift in Influenza A viruses

In 1968, co-infections between an avian influenza A H3Nx (where x=1–9) virus and the seasonal human influenza A H2N2 viruses resulted in the exchange of viral segments (reassortment) and the selection of the pandemic human influenza A H3N2 virus, with the RNA polymerase PB1 and haemagglutinin (HA) RNA segments derived from the avian virus and the rest of the segments derived from the human virus. The lack of pre-existing immunity to the antigenically novel H3 HA in humans facilitated human transmission. Similar reassortment processes have taken place during other influenza A virus pandemics. Once H3N2 became established in humans, the virus began to drift, as is the case with all other human seasonal influenza viruses. During drift, small antigenic changes in the HA protein generated by mutation are selected to increase immune evasion, although not as dramatically as during shift. M, RNA encoding M1, matrix protein, and M2, membrane protein; NA, RNA encoding neuraminidase; NP, RNA encoding nucleoprotein; NS1, RNA encoding nonstructural protein; PB2 and PA, RNA encoding RNA polymerases.

Figure and legend are adapted from Krammer et. al., 2018

vaccination. The genomic novelty generated by antigenic shift invalidates all preexisting virus control methods. Therefore, understanding antigenic shift is a significant aspect of mitigating virus evolution.

The ‘Asian flu’ of 1957 was caused by an H2N2 IAV generated by reassortment of the previously circulating human H1N1 virus and an avian H2N2 virus that contributed the PB1, HA, and NA genes to the pandemic strain. Phylogenetic analysis identified three distinct

lineages of the HA protein acquired from humans and two avian species. The global mortality during the pandemic was estimated at 1.1 million; however, this strain disappeared from the human population in 1968¹⁷. This was followed by the ‘Hong Kong influenza’ in 1968, caused by an H3N2 virus containing an HA from an avian IAV and the N2 neuraminidase from the 1957 H2N2 virus¹⁸. It was the largest pandemic since the Asian flu. The HA gene contained two mutations in its receptor binding site from avian viruses, altering its receptor binding specificity from preferential binding for α -2,3-linked sialic acids to α -2,6-linked sialic acids¹⁹ (Fig 2a & 2b).

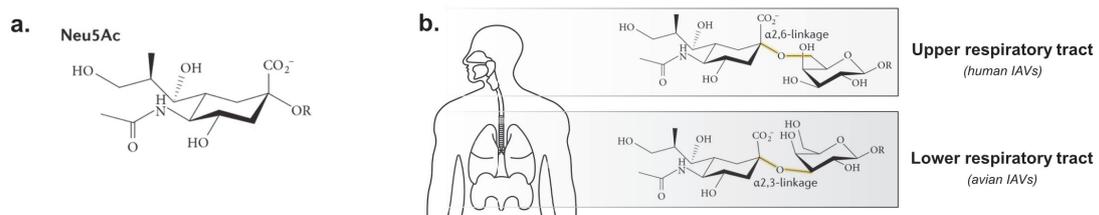


Fig 2 | HA mediated cellular receptor binding of Influenza A viruses

a. Neu5Ac. Sialic acids are derivatives of neuraminic acid, which is a nine-carbon monosaccharide. The C5 carbon is frequently modified with an N-acetyl group to form N-acetylneuraminic acid (Neu5Ac). **b. HA-sialic acid linkage.** Sialic acids are attached to carbohydrate chains on glycoproteins and glycolipids via different glycosidic linkages. The most common linkage types are α -2,3-linkage to a galactose residue, α -2,6-linkage to a galactose moiety or to an N-acetylgalactosamine moiety.

The 2009 pandemic H1N1 virus (H1N1pdm) derives six genes from triple-reassortant North American swine virus lineages and two genes (encoding NA and M proteins) from Eurasian swine virus lineage^{20,21} (Fig 3).

Bluetongue virus (BTV) has a segmented dsRNA genome comprising ten linear genome segments. The outer capsid protein 2 (VP2) is the main serotype-determining antigen encoded by segment 2. There are currently 28 serotypes of VP2 circulating in ruminants²². Due to reassortment, the genomes of BTV strains are highly variable, often showing substantial heterogeneity within the same serotype. Infrequent diagnostic sampling makes it complicated to study reassortment patterns in these viruses. Based on the genetic clustering of the individual segments, there appears to be a bias in segments that reassort, with segments 1, 2 and 7 being preferred combinations in reassortants²³. Frequent reassortment in BTV plays a crucial role in propagating novel resistant

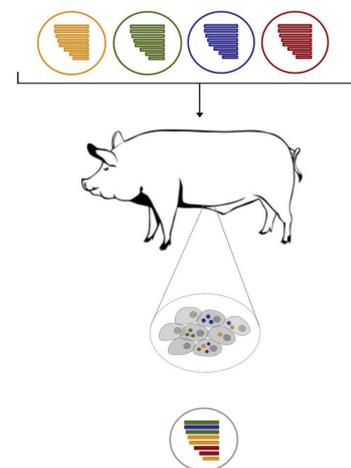


Fig 3 | Genetic reassortment in IAV

The 2009 pandemic H1N1 is a triple reassortant swine virus that underwent reassortment with a Eurasian avian-like swine virus. The NA and M segments were from the Eurasian swine viruses while the remaining six genes came from the triple reassortant.

phenotypes among indigenous sheep breeds, making it difficult to control the incidence of epidemics in these populations¹⁵.

The global mortality rates and healthcare burden caused by these pandemics were staggering, highlighting the need to study the molecular mechanism of antigenic shift to control future pandemics.

Determinants of genetic reassortment

There are various determinants for successful genetic reassortment. Selective and stochastic bottlenecks can negatively impact genetic reassortment and decrease genetic diversity in reassortants. During the coinfection of a host cell by two or more IAVs, some reassortant genotypes are not generated efficiently due to incompatibilities among heterologous RNAs. Low-affinity interactions between the non-cognate RNAs (derived from different parental viruses) are readily outcompeted during coinfection by the optimal, higher-affinity interactions between cognate RNAs (derived from the same parental virus)¹⁴.

Under circumstances where reassortant progeny are successfully produced, they are usually outperformed by one or both parental viruses in viral fitness due to incompatibilities among heterologous viral proteins. Such protein incompatibilities are typically evident only when the reassortant virus infects a new host cell. In this case, the reassortant-encoded proteins reduce transcription and replication levels, eventually decreasing virus growth rates. Mismatches at the protein level among the IAV polymerase complex components have been reported for several heterologous strain combinations^{24,25}. Another common protein incompatibility for IAV is the imbalance between HA avidity and NA activity. Since these two proteins are interdependent in their functions, reassortant genotypes that combine noncognate HA and NA genes often yield viral particles that do not bind efficiently to target cells or do not spread efficiently due to aggregation at the cell surface²⁶.

Therefore, the genetic diversity of parental strains often determines reassortment potential. Similar strains generate a high proportion of reassortant progeny, and divergence between parental viruses generates biased segment combinations²⁷. Other factors that limit or favour reassortment can be the selection of strains used to determine reassortment potential. Moreover, the reassortment potential between strains may change as the virus evolves, reducing the potential for reassortment. Under natural multicycle coinfection conditions, the percentage of reassortants produced, the number of genotypes produced, and the pairwise linkage of segments are essential markers that measure the propensity of reassortment²⁸.

Superinfection exclusion occurs when a pre-existing virus infection within a host prevents a secondary infection with the same or a closely related virus²⁹. Reassortment was induced in host cells superinfected with a virus variant of the same IAV strain within the first 18 hours of

primary infection. However, this was not observed at later time points, possibly due to innate immune responses activated by the primary infection³⁰. Natural infections occur at a low MOI. Therefore, reassortment is often restricted and takes multiple infection cycles before successfully propagating reassortant progeny are produced. This trend was also observed *in vitro* when primary infection was performed under low MOI multicycle conditions³¹. Once multiple infections of the reassortant progeny have been established, it can be an efficient strategy for host adaptation and inter-species transmission. Moreover, for efficient propagation of reassortants at sufficient levels in a population, they must have a genomic composition that confers a modest advantage to viral fitness³¹.

Influenza viruses

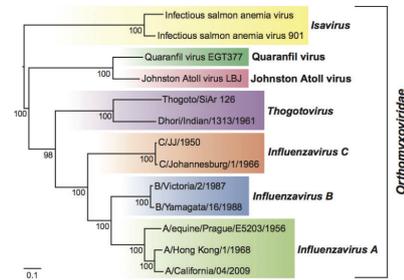
General classification and nomenclature

According to the 2021 ICTV update, Influenza viruses belong to the *Orthornavirae* kingdom and the *Orthomyxoviridae* family (Fig 4). Based on the antigenic differences in the NP and M1 proteins of influenza viruses, they are classified into four types - A, B, C, and D^{32,33}. Influenza is a common human pathogen that causes a contagious respiratory illness often referred to as Flu. It is characterised by mild to severe diseases such as pneumonia and can be associated with high mortality rates.

Influenza A viruses (IAVs) are divided into subtypes based on the two surface proteins HA and NA. HA binds to cellular sialic acids for virus entry, and NA cleaves sialic acids after budding from the host cell to facilitate virion release³⁴. There are 18 and 11 different hemagglutinin and neuraminidase subtypes, respectively. They are named H1 through H18 and N1 through N11 and are indicated in parentheses for all IAVs³⁵. Nomenclature of Influenza A strain mandates the following information in addition to the HA and NA subtypes: the antigenic type (A, B, C, or D), host of origin for non-human viruses (swine, avian); geographical origin or place of isolation, strain number and year of isolation of the strain e.g. A/chicken/Novosibirsk/65/2005 (H5N1)³⁶. IAVs can be subdivided into genetic clades and sub-clades, otherwise called groups and sub-groups, respectively. Genetically different clades and sub-clades may not be antigenically different. Current subtypes of IAVs that routinely circulate in the human population include H1N1 and H3N2³⁷.

Influenza B viruses (IBVs) are classified into two lineages instead of subtypes called B/Yamagata and B/Victoria, respectively. Influenza surveillance data from recent years shows the co-circulation of influenza B viruses from both lineages globally. However, the proportion of IBVs from each lineage that circulates can vary by geographic location and by season. Over the recent years, flu B/Yamagata viruses circulate less frequently than flu B/Victoria viruses globally^{38,39}.

Influenza C virus is detected less frequently in humans and usually causes mild infections; it does not present public health importance compared to IAV or IBV⁴⁰. On the other hand, cattle are the natural hosts for Influenza D viruses, and human cases have not been reported for this virus⁴¹.



Adapted from 9th Report ICTV 2011

Fig 4 | ICTV classification of IAV

Phylogenetic relationships within the family *Orthomyxoviridae*. Nucleotide sequences of PB1 were aligned and their phylogenetic relationships were determined.

Host tropism and transmission patterns

IAV is a zoonotic pathogen that can infect many species, such as birds, swine, and mammals. Diseases caused by zoonotic influenza viruses range from mild conjunctivitis to severe and sometimes fatal pneumonia. The cellular receptors that mediate virus attachment and entry determine the host susceptibility and tissue tropism. HA protein uses the widely abundant sialic acid receptors for virus entry, enabling a broad range of host cells to be infected by IAV. Sialic acids are a family of nine-carbon sugars based around an alpha-keto acid core⁴². The most common sialic acid modification in humans is *N*-acetylation at position 5 to form Neu5Ac (**Fig 2a**). Human-adapted IAV strains usually prefer α -2,6-linked Neu5Ac, displayed by epithelial cells of the upper respiratory tract. Avian-adapted strains typically favour α -2,3-linked Neu5Ac due to its overwhelming presence in the intestines, the primary site for avian IAV infection⁴³ (**Fig 2b**). Therefore, while sialic acids enable broader cellular tropism, the nature of sialic acid-HA interactions allows IAV to discriminate between hosts, restricting host-species jump⁴⁴.

However, pigs are an exception to the cellular infection pattern mentioned above. Pig tracheae contain both α -2,3- and α -2,6-linked receptors that both avian- and human-adapted strains can establish an infection and transmit. This susceptibility enables pigs coinfecting with different strains to serve as 'mixing vessels', which generate and successfully transmit reassortant strains into the human population. Humans susceptible to swine influenza usually come in close contact or inhale aerosols from infected pigs. Reassortant swine influenza viruses can potentially have new antigenic characteristics that enhance replication and transmission, facilitating efficient transfer and propagation within human hosts. Such reassortments have led to pandemics, as seen in 1918 and 2009, when the virus acquired efficient human-to-human transmission capabilities. Three subtypes of swine influenza circulate globally—H3N2, H1N2, and H1N1. The pdmH1N1 influenza gained worldwide attention as "swine flu" during the 2009 pandemic after swine influenza viruses were reassorted with preexisting H1N1 strains⁴⁵ (**Fig 3**).

Waterfowl and shorebirds are considered natural reservoirs of IAVs⁴⁶. Avian influenza spreads typically in birds but can also infect humans by crossing the species barrier. Five subtypes of avian IAVs have caused human infections (H5, H6, H7, H9, and H10 viruses). Avian influenza A viruses are classified into low pathogenicity avian influenza (LPAI) and highly pathogenic avian influenza (HPAI) viruses⁴⁷. The HPAI causes up to 90–100 % mortality in chickens within 48 hours of infection, whereas the LPAI strains cause mild illness or no disease except for ruffled feathers and decreased egg production in infected hosts⁴⁸. Human infections of avian flu are primarily acquired through direct contact with infected poultry or contaminated environments. While human-to-human transmission of avian influenza viruses has not been reported yet, the ongoing avian flu epidemic in poultry and mammals is concerning. If spillage

of these viruses occurs successfully in an immunologically naïve human population, followed by its establishment and transmission, it poses yet another threat of a virus pandemic following COVID-19⁴⁹.

Two novel IAV strains, H17N10 and H18N11, have been isolated from bats. Although they have similar genomes and are close to IAVs, they are distinct from conventional IAVs in aspects such as cellular receptors for entry, virus replication, and NA functions. Thus far, no reassortment has been reported between bats and conventional IAVs; therefore, they do not pose a health risk to humans⁵⁰.

Canine influenza A/H3N2 and A/H3N8 viruses cause a respiratory disease called dog flu. Canine influenza A (H3N2) viruses are different from seasonal influenza A (H3N2) viruses that spread annually in humans. There is no evidence of the spread of canine influenza viruses from dogs to people, and there has not been a single reported case of human infection with a canine influenza virus worldwide⁵¹.

Equine influenza viruses spread initially from birds to horses (horse flu) and belong to two IAV subtypes: H7N7 and H3N8. They affect the nose, throat, and sometimes the lungs of horses and closely related animals, like donkeys and zebras. The EIV H7N7 subtype was last reported in the 1970s and is now considered extinct. The EIV H3N8 subtype, distinct from the canine subtype, was first reported in horses in the 1960s and still spreads globally in horses⁵².

Antigenic drift and Flu vaccines

Seasonal flu epidemics are characterised by cocirculation of three influenza strains (A/H1N1, A/H3N2 and one or two Influenza B lineages). In humans, the continual evolution of these seasonal strains occurs, making some older strains disappear from circulation. One of the major causes of seasonal variations in Influenza is antigenic drift. Drift occurs when an accumulation of mutations, especially in the HA and NA proteins, gives rise to novel antigenic variants with the potential to evade the adaptive immune response. Therefore, such variants impact disease severity during seasonal epidemics. As a result, the human immune system has to evolve constantly to effectively counteract novel antigenic epitopes that try to evade a heterogeneous immune system. Hence, antigenic drift necessitates an annual review of flu vaccine components based on genetic and antigenic characteristics of circulating influenza viruses, their global distribution and prevalence, and the virus inhibition activity of candidate vaccine viruses⁵³.

During the 2019–2020 flu season, predominant A(H1N1)pdm09 viruses belonged to the phylogenetic subclade 6B.1A 183P-5A. As a result, vaccine candidates chosen for that season, including reference strains A/Guangdong-Maonan/SWL1536/2019 selected for egg-based,

and A/Hawaii/70/2019 selected for cell- or recombinant-based vaccines had viruses in subclade 5A with amino acid substitutions D187A and Q189E in HA antigenic site Sb⁵³.

Following the emergence of antigenically distinct B/Victoria viruses in 2017, B/Colorado/06/2017 (clade V1A.1) was selected as the 2018–2019 reference strain. B/Victoria viruses from HA clade V1A.3, with three amino acid deletions at 162–164, represented a small proportion of viruses in early 2019, but this clade increased and replaced V1A.1 viruses which were in circulation then. During the 2019–2020 influenza season, almost all B/Victoria viruses belonged to clade V1A.3. Despite concerns for vaccine mismatch, vaccine efficiency against V1A.3 viruses was similar to estimates for B/Victoria viruses in previous seasons^{53,54}.

The most commonly used influenza vaccines are injected-inactivated influenza vaccines. These come in trivalent (TIV) and quadrivalent (QIV) forms⁵⁵. With the QIV, there are also high-dose formulations designed for people aged 65 or older. Other types of influenza vaccines include: adjuvanted, live-attenuated (LAIV) and, more recently, recombinant influenza vaccines. Additionally, PR8 is the candidate strain for Flu vaccine production. It is used as the backbone to produce 6:2 reassortants with HA and NA segments from the circulating subtypes by genetic reassortment⁵⁶.

Influenza A viruses – structure and biology

Since IAVs adversely impact healthcare and the economy more than other Influenza subtypes, they have been studied extensively. The following section gives a detailed account of the IAV virion features and a review of the virus life cycle. PR8 was used as the prototype to describe the structure.

Virion structure

IAVs are pleiomorphic with either spherical shapes in lab-adapted strains (~100 nm in diameter) or filamentous shapes in clinical isolates (~100 nm wide and up to 20 μm long)^{57,58}. The viral envelope is acquired from the host plasma membrane and contains three transmembrane proteins: HA; NA; and matrix protein 2 (M2) (**Fig 5**). Beneath the viral envelope is the matrix layer, composed of oligomerised M1, which provides structural support⁵⁹. M1 protein has been proposed to be a determinant in forming filamentous virions. The membrane-lining lattice of the M1 protein presumably imposes the shape of elongated or filamentous virions⁶⁰. On the other hand, the low-energy state of membrane vesicles corresponds to spherical morphology. It appears that the disappearance of elongated virions at low pH corresponds to their converting to a spherical shape. In both cases, the spherical morphology likely results from the relaxation of M1-envelope and M1-M1 interactions in response to low pH⁶¹.

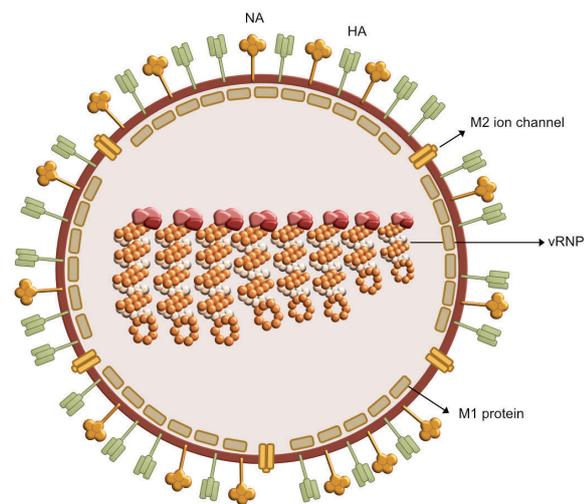


Fig 5 | Structure of Influenza A virus

It is an enveloped virus with eight negative-sense RNA. There are three transmembrane proteins - HA, NA and M2. Beneath the envelope lies a layer of oligomerised M1 proteins which retains the structural integrity of the virus. Tethered to the matrix layer is the viral ribonucleoproteins (vRNPs).

The 13.5kb viral genome is tethered to the matrix layer and split into eight segments associated with the NP proteins to form viral ribonucleoproteins (vRNPs) (Fig 6)⁶².

Genome structure

Eight vRNPs code for ten major proteins essential for viral replication: PB2, PB1, PA, HA, NP, NA, M1, M2, non-structural protein 1 (NS1), and nuclear export protein (NEP/NS2)⁵⁹. IAVs have three

integral membrane proteins: an ion channel protein (M2) and two glycoproteins, haemagglutinin (HA), required for entry into host cells, and neuraminidase (NA), involved in the release of progeny virions from the host cell. The genome also encodes up to 11 non-essential accessory proteins: PB2-S1; PB1-N40; PB1-F2; PA-X; PA-N155; PA-N182; eNP; NA43; M42; NS3; and tNS1⁵⁹ (Fig 7).

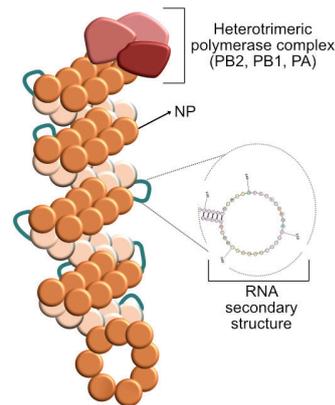


Fig 6 | Structure of viral ribonucleoprotein (vRNP)

Each vRNP has a viral RNA wrapped around a double helical scaffold of oligomers of the NP protein. Each vRNP has a loop at the end. The heterotrimeric polymerase complex comprises PB2, PB1 and PA and binds with the partially complementary vRNA 3' and 5' termini.

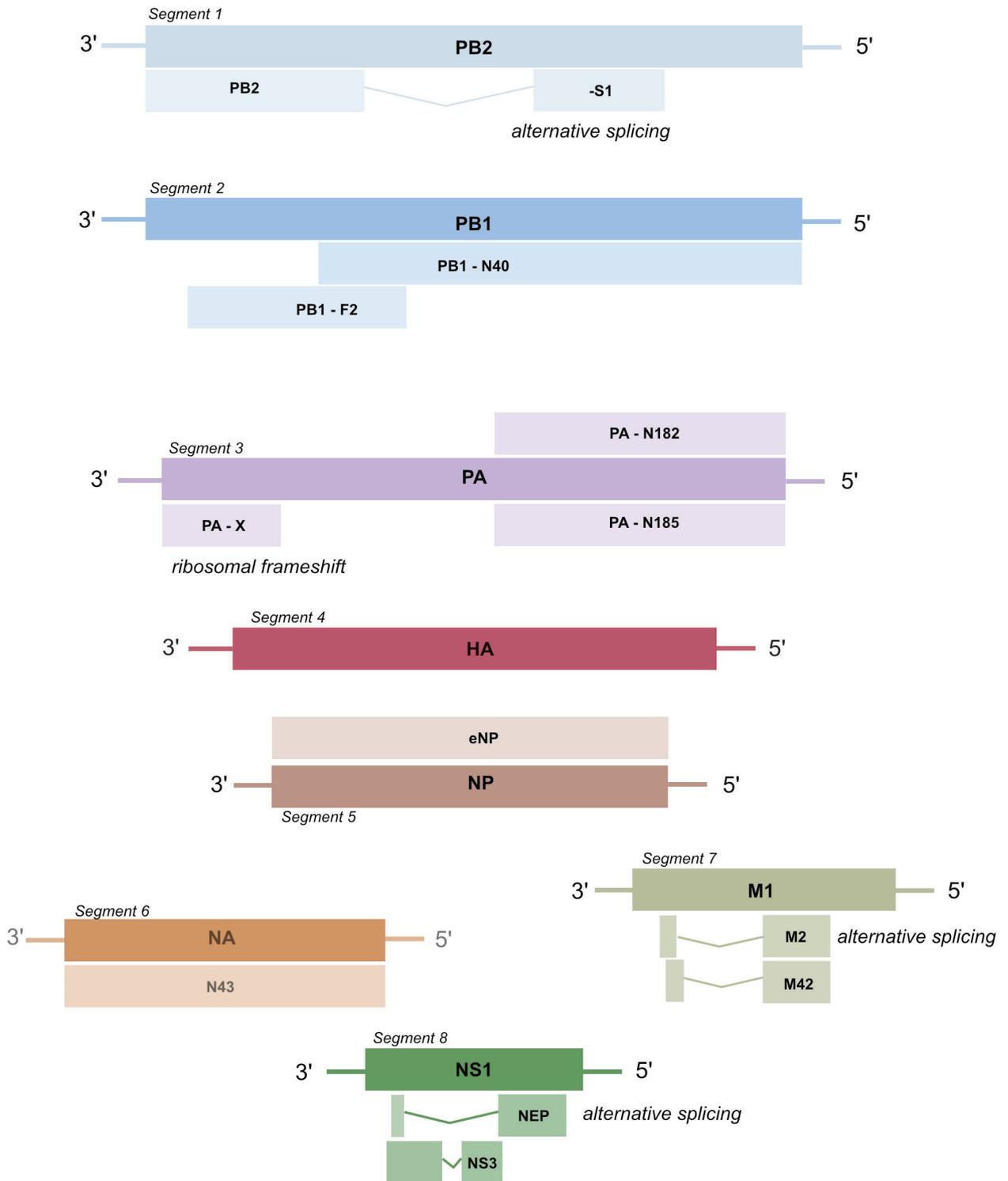


Fig 7 | Genome structure of H1N1/PR8. Dark and lighter boxes represent primary and accessory proteins respectively.

The total genome size is 13.5kb. Each segment ranges in size from 2341 to 890 nucleotides. The genome is multipartite and displayed in anti-sense orientation. The 3' and 5' terminus consists of 12 and 13 highly conserved nucleotides respectively. These partially complementary ends base pair to form the promoter for vRNA replication. PB2 (2341 nt), PB1 (2341 nt), and PA (2233 nt) form the heterotrimeric polymerase complex. Viral RNA polymerase (PB1, PB2 and PA) transcribes one mRNA from each genome segment. Transcription is primed by cap cleaved from cellular mRNAs by cap snatching. mRNA are polyadenylated by the viral polymerase stuttering on a poly U track. MP and NS mRNA can be alternately spliced, giving rise to mRNA coding for M2 and NEP proteins. Some strains can express M42 from the M segment. PB1 mRNA encode three proteins by leaky scanning: PB1, N40 and PB1-F2. PA-X protein is translated by ribosomal frameshift on the PA gene. Leaky scanning may produces N-truncated PA proteins, namely PA-N155 and PA-N182. PB2-S1 would be an alternative splicing from PB2 gene.

Figure and legend is adapted from Carter et. al, 2024 and viralzone.expasy.org respectively

The segments 1–3 encoded PB2, PB1 and PA proteins are estimated to be 84 kDa, 87 kDa, and 83 kDa, respectively. RNA segment 2, the segment that encodes PB1 also encodes a second polypeptide read in an alternative reading frame, PB1-F2, which varies in length between viruses, the full-length protein being around 90 amino acids in length, some strains of virus encode a PB1-F2 of around 55 amino acids, the vast majority of the pandemic A (H1N1) 2009 viruses encode a truncated 11-amino acid PB1-F2. The same RNA segment of some viruses encodes a third polypeptide, PB1 N40, 718 amino acids in length. RNA segment 4 encoded (unglycosylated) HA is 63 kDa (glycosylated HA1 is 48 kDa, HA2 is 29 kDa). The segment 5 encoded NP is 56 kDa. The segment 6 encoded NA is 50 kDa. M1 and M2 proteins encoded by segment 7 are 28 kDa and 11 kDa, respectively. Meanwhile, the segment 8 encoded NS1 and NS2 (NEP) are 27 kDa and 14 kDa, respectively⁶³.

Sequences at the termini of each vRNA sequence are complementary and highly conserved. Twelve of the 15 nucleotides at the 3'ends are shared among all eight segments of the PR8 virus. At the 5' ends, 13 nucleotides out of 16 are conserved. The other three nucleotides at each end of an RNA are segment-specific, and since they are complementary, they may also contribute to the stability of the panhandle. Sequence analyses of different IAV vRNAs suggests that conservation of the panhandle sequence is necessary for viability⁶⁴.

Viral nucleocapsid protein

The crystal structure of NP in a trimeric oligomer identified a mechanism of NP oligomerisation via a flexible C-terminal loop from one molecule that interacted with a neighbouring NP molecule⁶⁵. Each vRNA is wrapped around a double helical scaffold of NP oligomers, forming a right-handed double-stranded helix with a loop at one end and a heterotrimeric viral polymerase at the other end⁶³ (**Fig 6**). This macromolecular structure is called vRNP. The viral polymerase comprises PB2, PB1, and PA, bound to the vRNA 3' and 5' termini⁶⁶. Immunogold labelling of electron micrographs showed that the trimeric viral polymerase is localised to the opposite end of the RNP filament as the loop⁶⁷. The vRNA termini form a base-paired duplex when bound to the polymerase⁶⁸. Thus, the vRNA sequence is left partially exposed as it winds around the NP-scaffolding facilitating small RNA secondary structural elements like to peek out of the NP-scaffold. Insufficient levels of NP prevent the replication machinery of IAV from processing full-length genomes, resulting in aberrant replication products which form pathogen-associated molecular patterns enabling swifter host recognition and antiviral responses⁶⁹.

IAV life cycle – from entry to egress

Fig 8 provides an overview of the virus life cycle. The following section describes each step in detail.

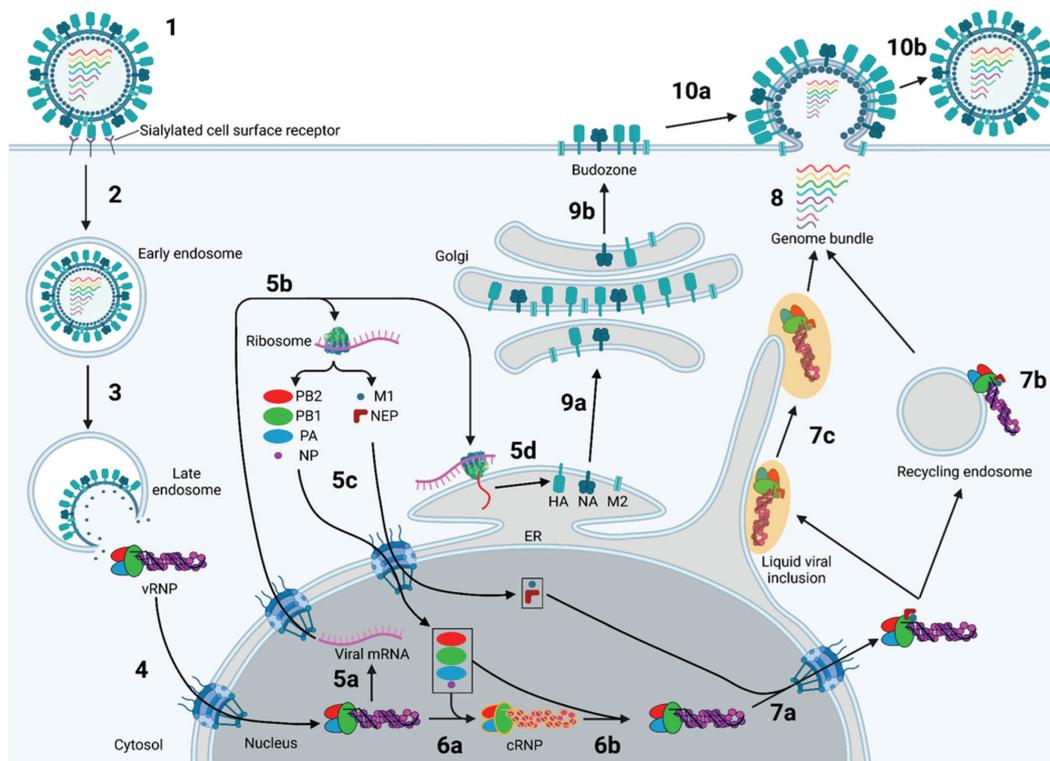


Figure 8 | Overview of IAV replication.

- (1) IAV virions comprise a viral membrane containing HA, NA, and M2 supported by an inner matrix layer of oligomerised M1. The viral genome is split into eight vRNPs, represented as coloured lines for simplicity. IAV virions can also be filamentous, but for clarity, this Figure depicts a spherical virion.
- (2) The replication cycle starts with the binding of HA to sialylated cell surface receptors, causing the virion to enter the cell contained in an early endosome; clathrin-mediated endocytosis is the primary mode of entry, but other mechanisms exist.
- (3) As the early endosome matures to a late endosome, the corresponding drop in luminal pH releases the HA fusion peptide, which buries itself in the endosomal membrane; subsequent refolding of the extended HA intermediate fuses the viral and endosomal membranes, releasing the genome into the cytosol. While only one vRNP is shown for simplicity, in reality, all eight are released as one bundle.
- (4) The vRNP bundle is imported into the nucleus via the classical importin pathway.
- (5a) The heterotrimeric viral polymerase on the vRNP snatches the first 10–13 nt from capped host mRNA and uses it and a vRNA template to transcribe viral mRNA (5b), which is then exported and translated. (5c) Cytosolic viral proteins (PB2, PB1, PA, NP, M1, and NEP) are imported into the nucleus. (5d) Transmembrane viral proteins (HA, NA, M2) are translated into the ER. The accessory proteins are omitted from this Figure for clarity.
- (6a) The vRNP-resident polymerase replicates a vRNA template into cRNA, which binds the nascently imported PB2, PB1, PA, and NP to form a cRNP. (6b) The cRNP is similarly but not identically replicated into progeny vRNPs.
- (7a) The progeny vRNPs are exported from the nucleus with the aid of M1 and NEP before being trafficked to the plasma membrane either in (7b) recycling endosomes returning to the cell surface or (7c) liquid viral inclusions that form and travel along a modified ER.
- (8) The vRNPs form a fully assembled genome bundle en route to the plasma membrane.
- (9a) The viral transmembrane proteins are translated into the ER membrane and trafficked to the Golgi, where HA is activated by proteolytic cleavage, (9b) and then to lipid rafts in the plasma membrane, which coalesce into a viral budzone, from which progeny virions bud.
- (10a) The viral transmembrane proteins induce budding, with the budzone membrane being incorporated into the growing virion. Cytoplasmic M1 oligomerises into a matrix layer beneath the membrane, and the fully assembled genome bundle is recruited into the tip of the budding virion. (10b) Once a sufficient amount of budzone has been incorporated, M2-mediated scission separates the viral and plasma membrane, and the sialidase activity of NA releases the nascent virion from sialylated cell surface receptors.

Figure and legend are adapted from Carter et. al., 2024

IAV uses sialic acid moieties of the host cell glycoprotein receptors for virus adsorption and entry. HA prefers α 2,6-linked Neu5Ac, a common sialic acid present in the epithelial cells of the human upper respiratory tract. However, avian strains prefer α 2,3-linked Neu5Ac displayed by epithelial cells of the lower respiratory tract in humans, thus increasing the disease severity of avian-derived flu in humans⁷⁰ (**Fig 2b**). HA protein is initially translated as a continuous polypeptide (HA0) with an ER signal sequence at the N-terminus cleaved after translation^{59,71}. Post-translation, HA0 homotrimerises before being trafficked to Golgi, where each monomer is proteolytically cleaved into an HA1-HA2 dimer linked by a disulphide bond^{59,72}. The highly conserved residues in the receptor-binding domain of HA1 interact with the glycan base of Neu5Ac-Gal. This binding prompts the glycan to fold back on itself and adopt an umbrella-like volume. Such binding of multiple HA proteins sequesters host transmembrane proteins to induce plasma membrane curvature for virus endocytosis^{59,73} (**Fig 9**).

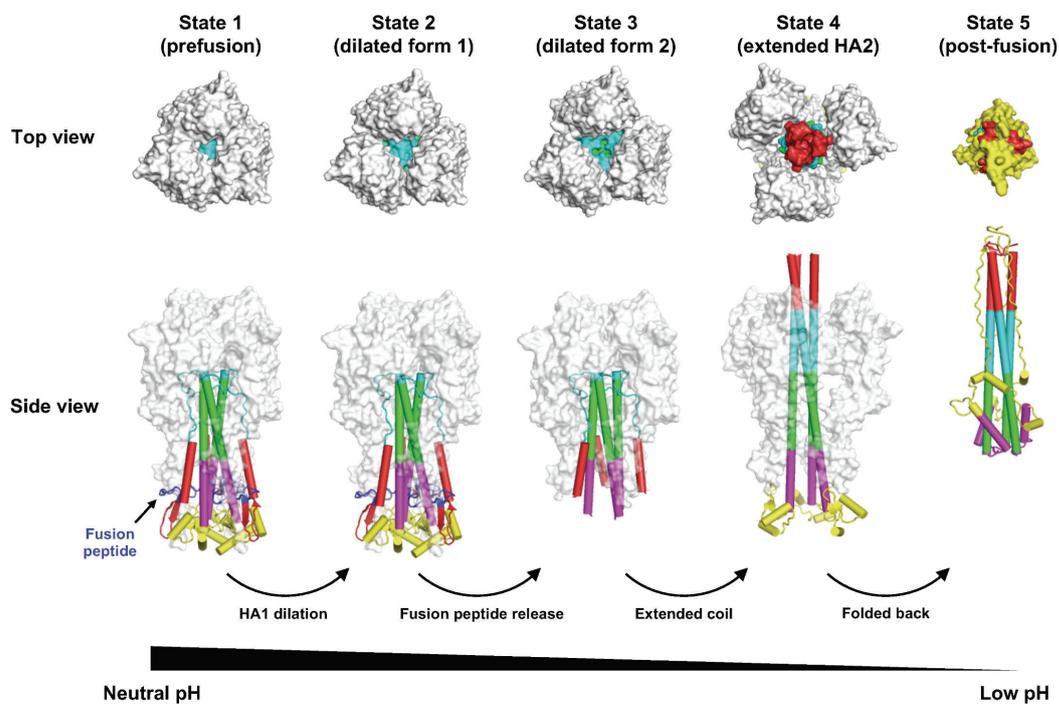


Figure 9 | Conformational changes of HA during pH-induced membrane fusion

Conformational change of HA during pH-induced membrane fusion. Different intermediates states of HA during pH-induced conformational change were identified by cryo-EM. The top and side views of state 1 (prefusion conformation, PDB 6Y5H), state 2 (dilated form 1, PDB 6Y5I), state 3 (dilated form 2, PDB 6Y5J), state 4 (extended HA2, PDB 6Y5K), and state 5 (post-fusion conformation, PDB 1QU1) are shown. Of note, after fusion peptide is released from state 2, the fusion peptide becomes disordered. In state 3, the membrane proximal region (yellow) is also disordered. Different components in the HA2 that are involved in structural rearrangements between pre- and post-fusion structures are in different colors.

Figure and legend are adapted from Wu et. al., 2020

Approximately 70 % of IAV virions enter a host cell via clathrin-mediated endocytosis which involves the formation, scission, and uncoating of clathrin-coated pits (CCPs)^{59,74}. Since filamentous IAV virions are too large for uptake by clathrin-mediated endocytosis, they enter

cells through macropinocytosis. A drop in the pH of the maturing macropinosome induces drastic membrane curvature, which eventually fragments the filamentous virions^{59,75}.

Post-entry into the host cell, the virus is confined to an endocytic vesicle. This vesicle fuses with mildly acidic early endosomes. As the endosomes grow bigger with more virions, they fuse with lysosomes. The M2 ion channel mediates virion acidification by transporting up to 100 protons per second across the viral membrane^{59,76}. This induces a pH drop to 4.5, transforming the HA structure to mediate viral-endosomal membrane fusion releasing the viral genome into the perinuclear space⁶¹. To describe this briefly, as the pH in the endosome begins to drop, an M2-mediated pumping of protons into the virion induces a change in intermolecular interactions of the M1 layer resulting in the weakening of its binding to the viral membrane⁷⁷. The envelope becomes more flexible and prone to fusion following these changes. Concurrently, disruption of M1-vRNP interactions discharges free vRNPs into the host cell cytoplasm. Soon after the M1 layer has been disrupted, HAs complete their conformational changes and induce fusion to prevent coagulation of the vRNPs⁶¹.

The vRNPs are released near the nucleus because unlike most RNA viruses, the IAV genome replicates within the nucleus and import usually occurs through the classical importin pathway^{78,79}. Nuclear import of vRNPs is mediated by the cooperation of importin β 1 with importin α and Ran GTPase. After M1 dissociates from the vRNPs, importin α binds to the nuclear localisation signal on NP^{59,80}. All eight vRNPs are imported as one bundle which separates into individual vRNPs after nuclear import⁷⁹. Phenylalanine-glycine repeats (FG repeats) line the central pore of the nuclear pore complex. Importin β 1 interacts with these FG repeats allowing the vRNP - importin α complex to be released into the nucleus.

Once the vRNPs are trafficked into the nucleus, the heterotrimeric vRdRp engages in transcription and replication of the viral RNA (**Fig 10**).

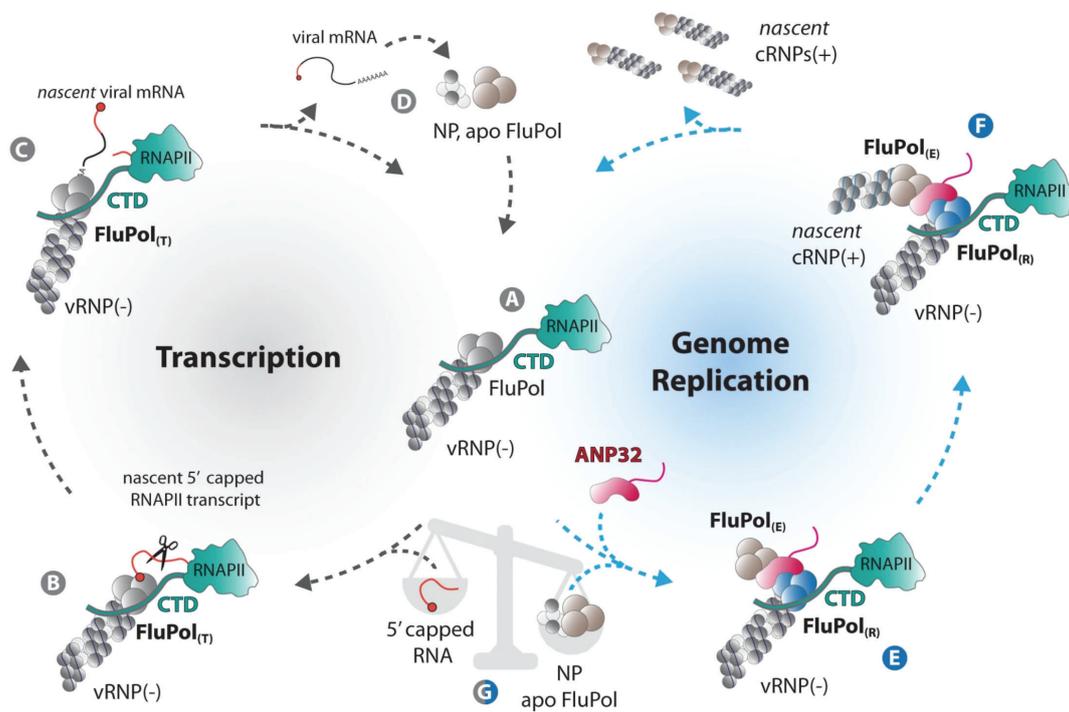


Fig 10 | Transcription and replication of Influenza genome

Upon influenza virus infection, incoming vRNPs are imported into the nucleus and bind to the host RNAP II CTD through bipartite interaction sites on the FluPol (**A**). This intimate association allows FluPol, in the transcriptase conformation (FluPol(T)), to cleave short capped oligomers derived from nascent RNAP II transcripts in a process referred to as 'cap-snatching' to initiate primary transcription of viral mRNAs (**B**). Polyadenylation is achieved by a non-canonical mechanism involving stuttering of the viral polymerase at a polyadenylation signal (**C**). The 5' and 3' vRNA extremities always remain bound to the polymerase which allows efficient recycling from the termination to the initiation state (C–A). Upon translation of viral mRNAs, de novo synthesised FluPols in an apo state (not viral RNA-bound) and NPs are imported into the nucleus (**D**). The apo FluPol, in conjunction with the host factor ANP32, associates with the parental CTD-associated FluPol(T) and triggers its conformational transition into a replicating FluPol(R), to form an asymmetric FluPol(R)-FluPol(E) dimer (**E**) where FluPol (**E**) is encapsidating the newly synthesised cRNA in conjunction with NP (**F**). The FluPol(R)-ANP32-FluPol(E) replication complex remains associated to the RNAP II through direct binding of the CTD to FluPol(R). Anchoring of the parental vRNP to the CTD allows it to engage into successive cycles of either viral genome replication or mRNA transcription, depending on the availability of NP, apo FluPol and/or nascent capped oligomers derived from actively transcribing RNAP II (**G**). Such switching between both activities allows efficient adaptation to waving levels of de novo synthesised vRNP components in the nucleus of an infected cell and is key to ensure a correct balance between genome replication and mRNA transcription.

Figure and legend are adapted from Krischuns et. al., 2024

Transcription occurs first since replication can begin and proceed only after producing adequate amounts of viral proteins (**Fig 11**).

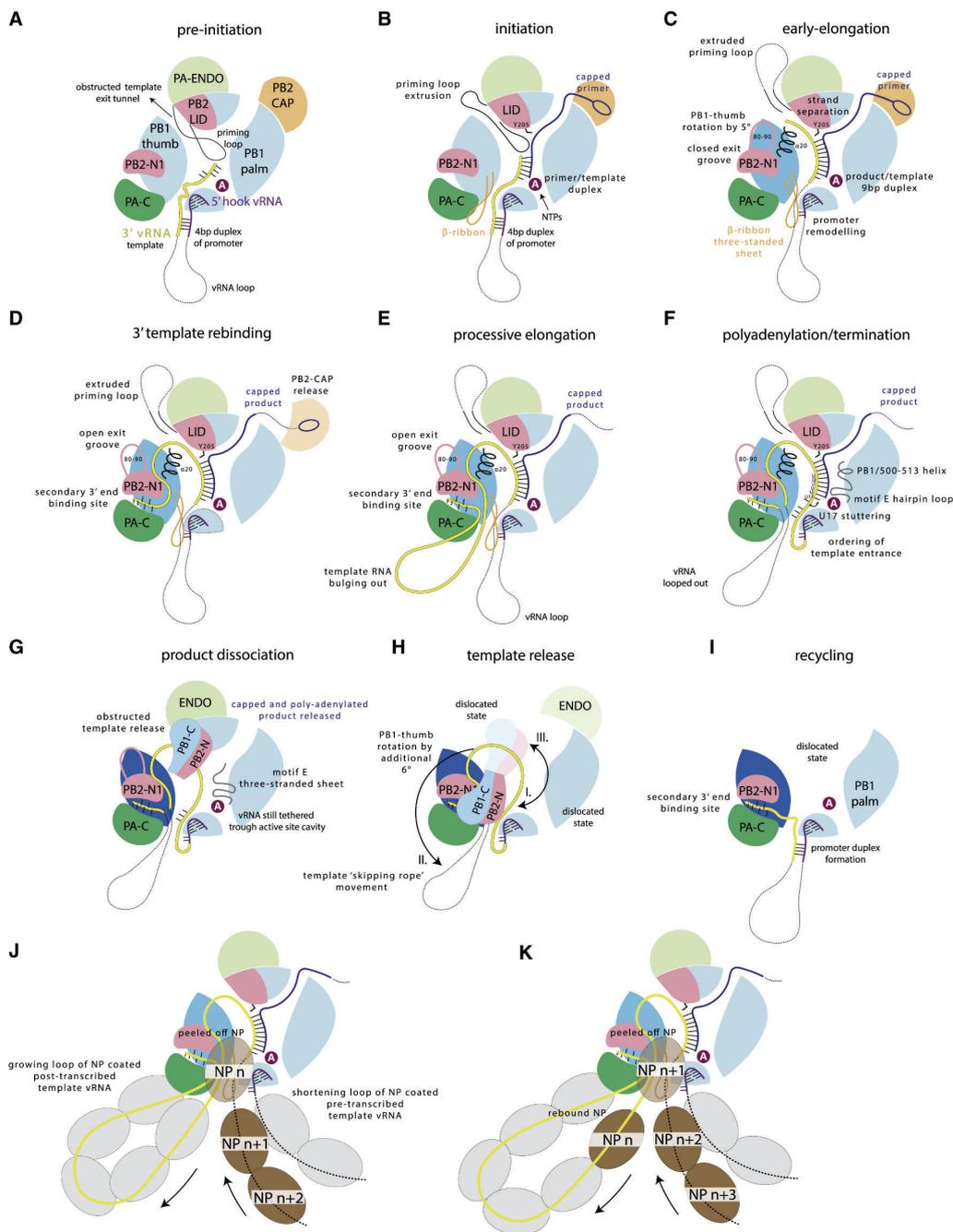


Fig 11 | Schematic of the transcription cycle

(A) In the promoter-bound pre-initiation state the template 3' end takes an indirect route into the active site (white A in magenta sphere).
 (B) A stochastic displacement of the priming loop tip allows primer-template hybridization and incorporation of the first few nucleotides. The growing product/template duplex gradually displaces the priming loop out of the exit channel.
 (C) Template translocation initially occurs by straightening of the template entrance pathway, but eventually the promoter melts and the β -ribbon collapses. Thumb rotation widens the active site cavity allowing growth of the product/template duplex to the steady-state nine base pairs, before abutting against the PB2 helical lid, forcing strand separation.
 (D) The exiting template is guided along the positively charged exit groove and after translocating ~24 nucleotides, its extremity docks into the secondary 3' end site and remains bound there until the end of the transcription cycle.
 (E) Further translocation forces the exiting template RNA to loop out.
 (F) The polymerase processively extrudes the template until the incoming vRNA loop fully tightens up, whereupon the polymerase flips template base U17 in and out of the active site, producing the poly(A) tail.
 (G) Due to mismatches in the A-U rich product-template duplex, the polyadenylated viral mRNA is eventually released, concomitant with a switch of the polymerase to the open, dislocated state. However, the template RNA is still threaded through the polymerase, trapped inside a tunnel.
 (H) The template can be released, for instance, by an outward flip of the PB1-C/PB2-N helical bundle that allows it to swish out of the active site cavity like a skipping rope held at each end.
 (I) Once the vRNA is completely outside, the promoter can reform with the template 3' end still bound in the secondary site (recycling state). However, for robust promoter formation, the polymerase must transition back to the non-dislocated, pre-initiation configuration, which forces release of the 3' end from the secondary site and preferential binding in the active site. Another round of transcription can then start.
 (J) In the vRNP context, NP molecules must be successively stripped off the incoming template (dotted) as it translocates into the polymerase.
 (K) Because the transcribing polymerase sequesters roughly the same length of RNA as binds to a single NP (~25 nucleotides), at least one NP must remain free of RNA (light brown NP(n)) before re-docking onto the growing loop of the exiting template.

Figure and legend are adapted from Wandzik et. al., 2020

The binding of the IAV polymerase to the vRNA promoter opens the polymerase conformation enabling transcription^{59,81,82}. For transcription initiation, mRNA synthesis is primed by cap-snatching 5' 7-methylguanylated (m7G) host mRNA fragments ranging in 10 - 13 nucleotide length from an actively transcribing host RNA polymerase II (Pol II)^{83–85}. The C-terminal domain of Pol II contains 52 YSPTSPS heptad repeats that bind the IAV vRdRp through the C-terminal domain of PA^{59,85}. The PB2-cap binding domain is utilised by the 5' vRNA terminus to bind to the m7G host mRNA cap⁸⁶. In the inactive conformation, the PB2 cap-binding domain is obstructed⁸⁷ and becomes accessible in the pre-initiation conformation^{59,88}. The PA endonuclease domain cleaves the host mRNA at a 5'-G[^]C-3' site 10-13 nucleotides downstream of the cap after it is captured by the PB2 cap-binding domain^{59,89,90}. The PB2 cap-binding domain positions the capped primer into the catalytic center of PB1 where extension of the vRNA occurs⁸⁹. At the PB1-PA interface, nucleotides 1-10 of the 5' vRNA terminus remains bound to the vRdRp to form an intramolecular structure called the 5' hook^{59,91}. The cleaved mRNA base pairs with the 3' terminus of vRNA, and is elongated by two catalytic divalent metal cations (Mg^{2+} *in vivo*, Mn^{2+} *in vitro*) coordinated by three conserved aspartate residues^{92,59,93}.

Transcription continues until interrupted by the tight binding of the vRNA 5' hook to its binding pocket^{59, 81,94}. mRNA is polyadenylated via reiterative stuttering at the oligo (U) motif of vRNA 5' nucleotides 17–22⁹⁵. A strain is introduced when vRNA 5' U17 enters the active site⁹⁴. An incoming ATP molecule stabilises U17 within the active site to allow its incorporation into the mRNA. This is followed by the ejection of the strained U17 out of the active site^{59,94}. U17 then re-enters the free active site to be stabilised by another free ATP, after which it slips out of the active site again. This process is repeated until 30–180 adenosines have been newly added, after which the vRNA–mRNA duplex dissociates^{59,96}. The PB1 C and PB2 N1 domains rotate to open the template exit tunnel, allowing the release of the product strand, and they rotate back to seal the tunnel^{59,94}. The polymerase and the vRNA promoter returns to its pre-initiation conformation for a new round of transcription and the vRNA exits the polymerase catalytic core^{59,94}.

Following transcription termination, the viral mRNA undergoes post-transcriptional modifications and nuclear export. The mRNA from segments 7 and 8 undergo splicing to form the mRNA for M2 and NEP, respectively^{59,97,98,99}. Post-transcription, the m7G cap, poly(A) tail, and UTRs of the viral mRNA associated with various RNA binding proteins recruit the mRNAs to the nuclear pore complex. The mRNA is released into the cytosol for ribosomal translation^{59,100}. The soluble viral proteins (PB2, PB1, PA, NP, M1, NEP) are trafficked directly into the cytosol, whereas the transmembrane proteins (HA, NA, and M2) to the endoplasmic reticulum^{59,101}.

Following this, the vRNP undergoes primary genomic replication (**Fig 12**). IAV genome replication occurs in two phases: initial transcription of vRNA to complementary positive-

sense RNA (cRNA). The nascent proteins convert it into a complementary RNP (cRNP) intermediate. The cRNPs are used as templates for secondary replication to form progeny vRNPs^{59,102,103}. Although the mRNA and cRNA are positive-sense copies of the vRNA, they have differences such as the lack of a 5' m7G cap and poly (A) tail in cRNA^{104,105}. Nuclear export and trafficking of the progeny vRNPs occur via a combination of recycling endosomes and a modified endoplasmic reticulum forming a bundle with all eight segments while en route^{59,106,107,108,77,109}.

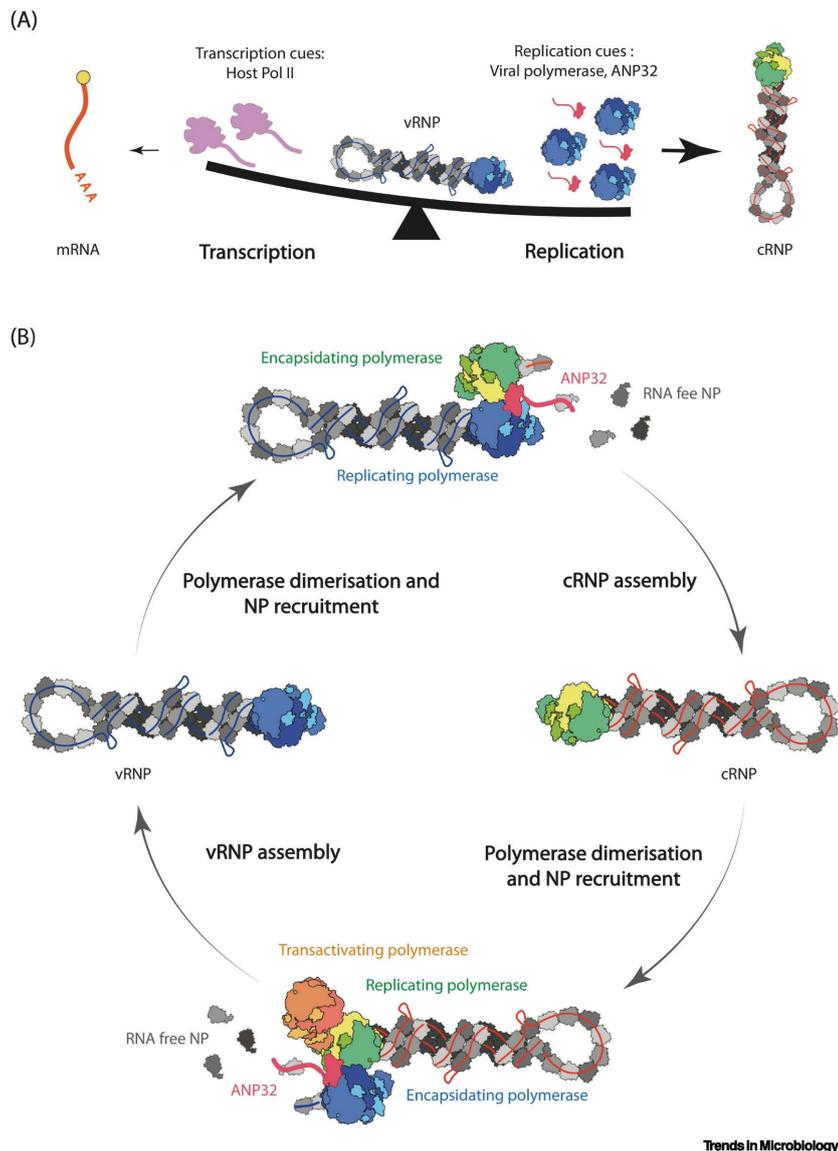


Fig 12 | Regulation of the influenza virus RNA polymerase through interactions with molecular cues and proposed model for influenza virus RNA genome replication.

(A) Interactions with host RNA polymerase II drive incoming viral ribonucleoproteins (vRNPs) towards transcription to produce viral mRNA while interactions with host acidic nuclear phosphoprotein 32 (ANP32) proteins newly synthesised viral polymerase drive vRNPs towards replication producing RNP.

(B) Replication cycle the viral polymerase showing the roles of replicating, encapsidating, and transactivating polymerases.

Abbreviations: cRNA - complementary RNA, cRNP - complementary ribonucleoprotein

Figure and legend are adapted from Zhu et. al., 2023

Two polymerases bridged by host acidic nuclear phosphoprotein 32 (ANP32A) mediate primary IAV genome replication¹¹⁰. In this dimer conformation, one polymerase remains bound to the terminal promoter region of a 47-nucleotide vRNA, while the other remains RNA-free¹¹¹. The RNA-bound polymerase is the replicating polymerase and the RNA-free polymerase is the encapsidating polymerase¹¹¹. Primary replication occurs through *de novo* initiation, using the PB1 priming loop to form a pppApG primer base-paired to the 3' end of a vRNA^{59,112,113}. The pppApG primer stabilises the vRNA 3' terminus in the active site and destabilises the 5'-3' vRNA duplex to favour the initiation of genome replication^{59,114}. The 5'-3' base-pairing in the distal promoter region positions the template RNA at the polymerase active site. Nucleotide extension occurs in the same mechanism as transcription⁸⁷. cRNA is more prone to degradation by host factors since it lacks the necessary viral proteins^{59,104}. Adding an encapsidating polymerase and the co-transcriptional addition of NP via homo-oligomerisation makes cRNP stable^{59,115}.

Secondary replication also occurs like primary replication and requires pppApG primer^{59,116}. However, in this case, the pppApG primer is synthesised internally¹¹⁷. To copy the vRNA into a cRNA, RdRp initiates *de novo* at positions 1 and 2 of the 3' end of the vRNA template (terminal initiation)¹¹⁴. When the RdRp has to copy the cRNA back into vRNA, it initiates *de novo* at positions 4 and 5 of the 3' terminus of the cRNA template (internal initiation)¹¹⁴. Upon polymerase dimerisation, the priming loop and PB2 N1 domain undergo conformational changes, facilitating the backtracking of the cRNA 3' terminus. The elongation mechanism is the same as primary replication. The energy from new base pairs between the cRNA and cRNA strands breaks the cRNA pairs, and the emerging vRNA is captured by the other polymerase and encapsidated to form progeny vRNPs^{59,118}.

CRM1, a nuclear export receptor that binds cargo with a nuclear export signal (NES) mediates the nuclear export of progeny vRNPs¹¹⁹. NES recognition and cargo transport across the nuclear pore complex depend on CRM1 binding to RanGTP⁶². The vRNP, M1, and NEP are exported together by CRM1, NEP bridges the vRNP-M1¹²⁰ and CRM1-RanGTP complexes have been shown to be involved nuclear export of vRNPs¹²¹.

At the end of the infectious cycle, all virus components are trafficked to the apical plasma membrane of polarised epithelial cells for assembly and budding (**Fig 13**). Rab11 is a host GTPase that trafficks vesicles along actin and microtubule networks^{122,123}. Rab11 selectively interacts with vRNPs and not cRNPs through the PB2 627 domain^{59,124}. IAV infection impairs normal Rab11 function to induce the formation of a modified tubular ER within an infected cell. Liquid viral inclusions (LVIs) adjacent to ER exit sites transfer the vRNPs to the plasma membrane^{59,125,126}. The assembled 7+1 complex is incorporated into the budding virion with the transmembrane viral proteins¹²⁰. Once vRNPs are incorporated into a budding virion, M2 induces scission of the viral and plasma membranes, and the sialidase activity of NA releases progeny virions from the cell surface¹²⁷. Assembly of viral components at the budding site

causes asymmetry of the lipid bilayer and outward membrane bending, facilitating budding initiation. Although poorly understood, bud scission involves cellular and viral factors such as disassembly of cortical actin microfilaments, discontinuity in the M1 layer underneath the lipid bilayer, absence of lipid rafts and outer membrane spikes, and presence of M2 to pinch the membrane off, all facilitating bud fission¹²⁸.

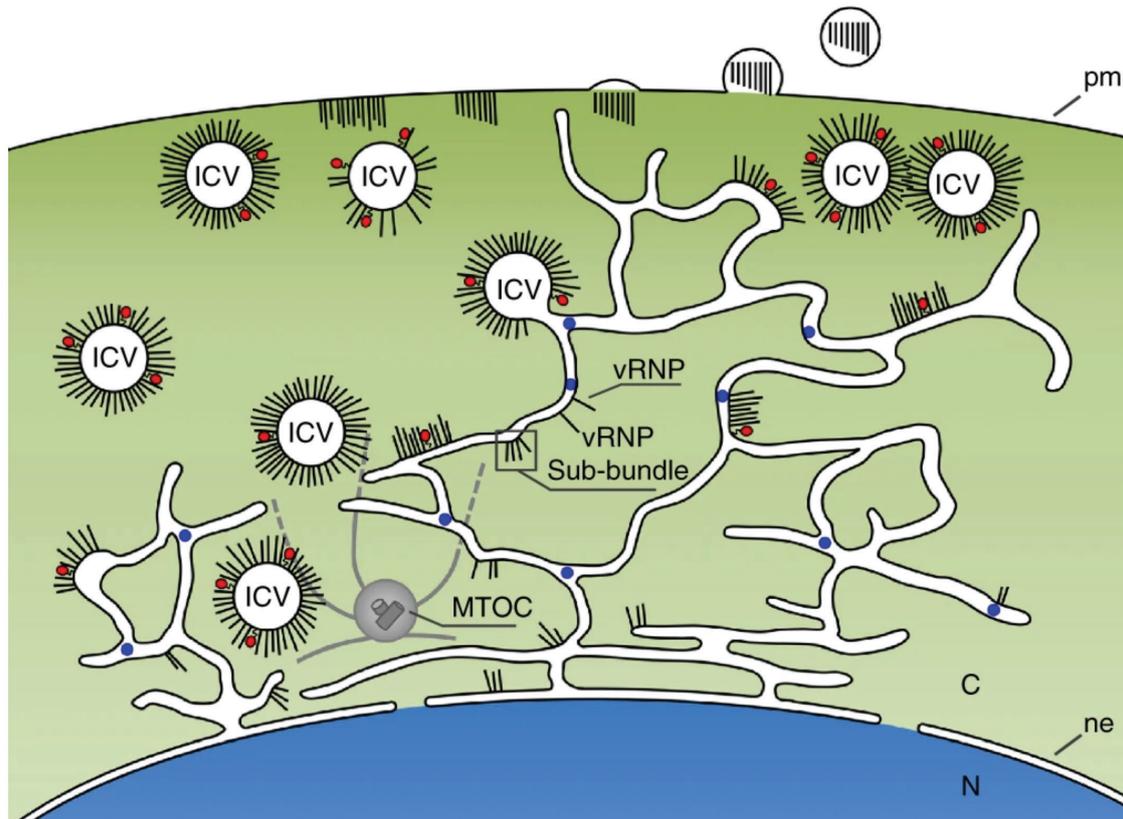


Fig 13 | Model for the trafficking of vRNPs across the cytoplasm in an IAV-infected cell.

The remodeled tubulo-vesicular ER (in blue) extends around the microtubule organizing center (MTOC in gray) and from the nuclear envelope (ne) to the plasma membrane (pm). After their exit from the nucleus, individual vRNPs and/or sub-bundles of vRNPs are targeted to the modified ER. Irregularly coated vesicles (ICVs) loaded with vRNPs and with the Rab11 molecule (in red) might bud from the ER and ensure the transport of vRNPs to the plasma membrane. The frequently observed pairing of ICVs could favor RNA–RNA interactions among vRNPs and the progressive assembly of sets of 8 distinct vRNPs. vRNPs are released from ICVs and possibly transferred to the plasma membrane in a touch-and-go process.

Figure and legend are adapted from Martin et. al., 2017

Selective genome packaging in segmented viruses

Viral nucleocapsid proteins and genomic RNA self-assemble into supramolecular complexes

Monopartite genomes can easily be packaged into single virions. On the contrary, viruses with segmented genomes have to package a unique copy of each cognate segment to ensure the production of an infectious virion. While some viruses, such as IAV and Rotavirus, package the entire genome into a single virion, others like Bromoviridae are multi-particulate viruses since they encapsidate a part of their genome into one virus particle and the remaining into one or more sub-virus particles^{129,130}. Understandably, efficient genome packaging in these viruses involves several cooperative steps of the viral lifecycle, such as synchronised vRNA/s trafficking through the cytosol, capsid self-assembly, membrane budding and cessation (**Fig 14**).

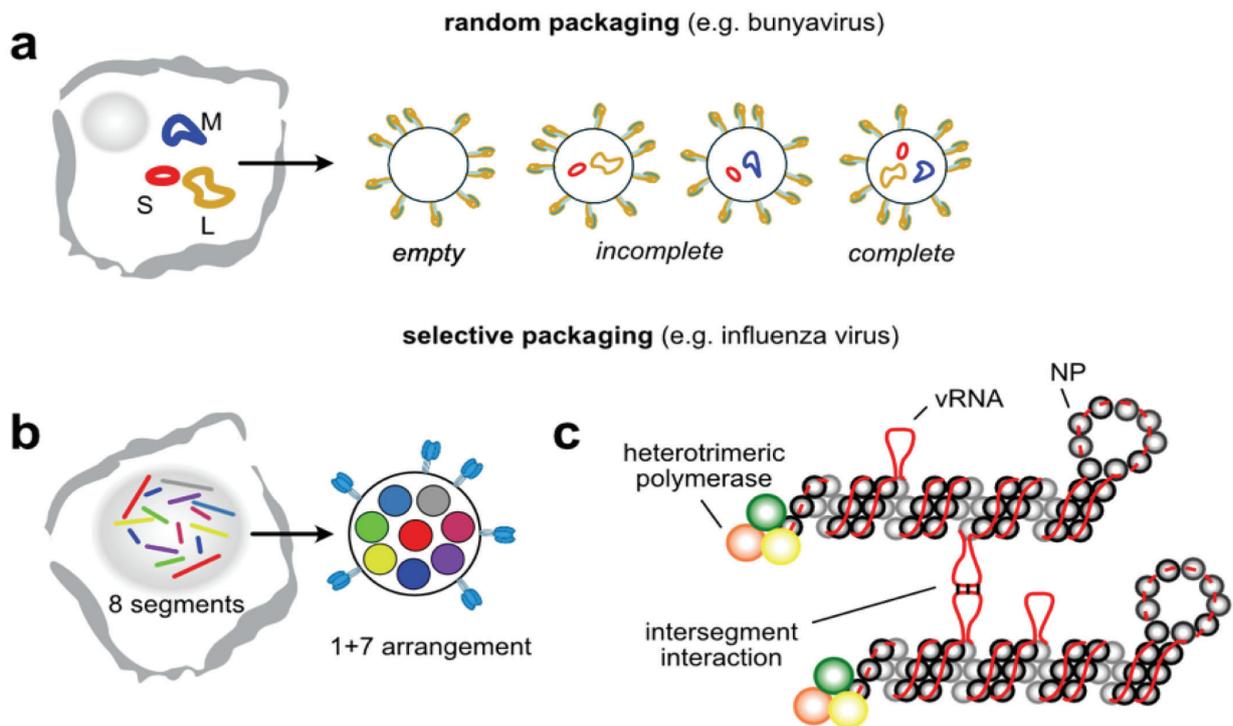


Fig 14 | Segmented viruses package their genomes randomly or using segment specific packaging signals.

(a) Bunyavirus randomly package three genome segments, small (S), medium (M) and large (L), such that many progeny virions are empty or incomplete; **(b)** Influenza viruses selectively package eight genome segments. Genome segments within budding virions are organised into a 7+1 arrangement with one central segment surrounded by seven others; **(c)** Influenza vRNAs are packaged as viral ribonucleoproteins (vRNPs). vRNAs are bound by nucleoprotein (NP) and a heterotrimeric polymerase. vRNA is incompletely coated by NP allowing for intersegment vRNA-RNA interactions to occur as a possible mechanism underlying the selective packaging process.

Figure and legend are adapted from Ye et. al., 2021

The biophysical basis for genome packaging in segmented single-stranded viruses was accepted to be thermodynamically driven by electrostatic interactions between negatively charged phosphate groups in vRNA and basic amino acids located in flexible tails called arginine-rich motifs of capsid proteins^{131,132}. The charge ratio of negative charge on RNA to positive charge on proteins is usually 2:1 and correlates with the segment length^{131,133}. However, these electrostatic interactions were proven nonspecific. On the contrary, in segmented viruses, since cellular RNAs are rarely found in progeny virions, it is evident that vRNAs are selectively recognised from a pool of host cellular RNAs by specific interactions of viral capsid proteins with vRNA segments. This recognition process determines the precision of segmented genomes' packaging to produce infectious virions. Therefore, biophysics and electrostatics alone do not explain the precision with which viruses distinguish their own genome from host RNA; instead, they raise more questions on specific molecular mechanisms involving coordinated viral protein and vRNA interactions¹³⁴.

Viral self-assembly is where capsid proteins associate with each other and co-assemble with vRNA to form a highly ordered supramolecular structure¹³⁴. Co-assembly was first studied in TMV, suggesting that ssRNA genome molecules can act as assembly templates¹³⁵. Negatively charged ssRNA molecules act as non-specific entities that link the positively charged NP proteins, with the 'stem-loop' side branches of the RNA molecules having specific affinity for the capsid proteins¹³⁶. In IAV, multiple and dispersed yet specific interactions exist between the genome segments that always remain associated with multiple NP copies, forming a helical nucleocapsid structure or vRNPs. The NP interacts with RNA via a positively charged cleft¹³⁷. Nucleocapsid proteins are aligned parallelly along the vRNA segment, sandwiched between two domains composed of conserved helix motifs¹³⁸. Multiple copies of vRNPs aggregate into a 7+1 complex before being packaged into a budding virion¹³⁹. Maintaining the integrity of this supramolecular 7+1 complex, most likely through packaging signals, during genome assembly is essential to produce infectious virions during the viral lifecycle.

Packaging signals in most viruses are short, evolutionarily conserved RNA sequences with a specific affinity for nucleocapsids¹³⁵. They are dispersed across the entirety of the viral genome with multiple secondary structural elements capable of nucleocapsid recognition. The packaging signal-mediated genome assembly in spherical viruses was modelled using the Gillespie algorithm. This model states that i) capsomers interact with different packaging signals at different rates determined by capsomer-packaging signal affinity, ii) the free energy of the capsomer-capsomer bonds determines the strength of their interaction rates¹³⁴. The specificity of selective genome packaging was later attributed to the affinity and number of packaging signals dispersed on a vRNA. Perlmutter *et al.*, showed that there were both high-affinity and low-affinity packaging signals and genome assembly was higher in instances where a combination of high and low-affinity packaging signals occurred simultaneously¹⁴⁰.

Although the self-assembly of virus capsomers and genome occurs prominently in nature, the underlying mechanisms remain elusive. Therefore, identifying the molecular mechanism that enables this selective formation of supramolecular complexes can pave the way for developing new antiviral candidates. However, direct measurements of such interactions are challenging because the interactions between protein subunits are generally weak and depend on the solution conditions^{141,142}. Additionally, the role of vRNA in assembly and whether a small cluster of RNA-bound proteins initiates the progressive assembly of the genome all remain avenues that require investigation. Under these circumstances, models that suggest the collective roles of specific NP-NP interactions coupled with high-affinity NP-RNA interactions that allow for specific and deterministic genome assembly pathways are more promising for studying selective and coordinated genome packaging in segmented viruses.

Packaging signals mediate selective genome assembly pathways in segmented RNA viruses

When protein-protein or sequence-independent protein-RNA interactions are too weak to nucleate assembly, packaging signals that can enhance protein-RNA interactions and RNA-mediated protein-protein interactions induce nucleation and facilitate subsequent assembly¹⁴⁰. Assembly and specificity are also sensitive to the affinity and number of these packaging signals¹⁴⁰. Capsid assembly is impaired when the packaging signal-mediated interactions are disrupted¹⁴⁰. Multiple RNA conformations may anneal during assembly through reversible interactions and/or cooperative RNA-protein rearrangements that can account for an ensemble of assembly pathways¹⁴⁰. Simulations demonstrate that packaging signals can alter assembly pathways significantly compared to non-cognate RNAs¹⁴⁰.

The following section briefly summarises selective genome packaging in some viruses where packaging signals have been identified. For most viruses, these packaging signals overlap the 5' and 3' UTRs and coding regions of the viral RNA, ensuring assembly occurs in competition with other genome functions, such as replication and transcription. While comparing this data, it is evident that all segmented viruses share some key determinants of genome assembly and packaging. This knowledge can be extrapolated to broaden experimental design and future studies of genome packaging in other lesser-known viruses.

The *Reoviridae* family of segmented dsRNA viruses includes clinically and economically significant human, animal and plant pathogens, such as rotaviruses, bluetongue viruses and rice dwarf viruses³¹. The rotavirus A virion is a non-enveloped, triple-layered particle comprising a dsRNA genome of 11 segments (S1 – S11). They range in size from 0.5–3.3 kb and each segment has a central ORF flanked by 5' and 3' UTRs. The rotavirus reassortment and packaging process is understood poorly because the field lacks *in vitro* packaging assays and efficient reverse genetics techniques. It has been hypothesised that rotavirus A genome reassortment is similar to IAV genetic reassortment³¹. The 11 distinct RNAs are speculated to

engage with each other through cis-acting RNA elements. These RNAs form a supramolecular complex encapsidated by the core-shell protein during early virion assembly. The segment-specific packaging signals are predicted to be present on the 5' and 3' termini. *In silico* analyses of nucleotide sequences from strains of rotavirus A have identified several putative RNA structural elements in these terminal regions that may represent packaging signals³¹. They reside in the 5' and 3' UTRs and are dispersed in the internal coding sequences³¹. During or immediately following their packaging into a core assembly intermediate, viral polymerases convert rotavirus A vRNAs into dsRNAs by viral polymerases. The nascent core assembly intermediate then undergoes additional morphogenesis to become an infectious triple-layered, non-enveloped particle³¹. Moreover, RNA-RNA SELEX experiments against S11 revealed multiple areas of the RNA underwent conformational rearrangements when non-structural protein 2 binds to S11, while the most stable S11 intramolecular helices H1-H3 remain largely inaccessible¹⁴³. This is consistent with the high affinity of NSP2 shown for ssRNA but not dsRNA. Substitution of NSP2 with a mutant DC-NSP2 which had a significantly lower affinity for ssRNA reduced the formation of intersegmental interactions, supporting the proposed model of NSP2-mediated remodelling of ssRNAs. Such a mechanism would account for the NSP2-facilitated selection of thermodynamically favourable inter-segment interactions that may not always follow strict pairing rules¹⁴⁴.

The bluetongue virus (BTV) genome is organised into ten discrete double-stranded RNA molecules named S1-S10. Genome packaging occurs via the formation of supramolecular complexes by segments interacting with specific sequences in the 3' UTRs. The segments follow a sequential packaging pathway from smallest to largest segment during virus capsid assembly^{31,145}. Putative packaging signals in the 3' UTRs of BTV segments were targeted by several nuclease-resistant oligoribonucleotides (ORNs), and their effects on virus replication in cell culture were assessed. ORNs complementary to the 3' UTR of BTV RNAs significantly inhibited virus replication without affecting protein synthesis^{31,146}. When the same ORNs were added to an RNA-RNA interaction assay that measured the formation of supramolecular complexes between and among different RNA segments, it was observed that the complex formation by the segments was inhibited. Deletions or substitution mutations of these targeted sequences not only reduced the formation of RNA complexes but also prevented the rescue of infectious virions by reverse genetics. When the 3'UTRs were exchanged between segments, segment-specific RNA recognition was impaired, implying that these regions most likely are involved in forming RNA secondary structures that enable segment-specific recognition for genome packaging. Additionally, inhibition of *in-trans* packaging with ORNs suggested that these interactions could also be used as potential targets for antiviral assays^{31,146,147,148}.

Bromoviridae (BMV) encapsidates three genomic RNAs and a subgenomic RNA (sgRNA) into three individual virions of identical size and morphology by a single coat protein¹⁴⁹. Genomic RNAs 1 and 2 are packaged individually into separate particles, whereas genomic RNA3 and

subgenomic RNA4 are co-packaged into a single particle¹⁵⁰. A highly conserved tRNA-like structure at the 3' end has been hypothesised as the nucleating element (NE) for capsid protein subunits and a cis-acting, position-dependent packaging element (PE) of 187 nucleotides present in the nonstructural movement protein gene has been identified as the integral components of the packaging core¹⁵⁰. Selective recognition of a domain specific to each BMV RNA ensures precise packaging of each segment.

The *Cystoviridae* family comprises segmented dsRNA viruses that infect Gram-negative bacteria^{31,20}. The prototype is Pseudomonas phage $\phi 6$ and it has three dsRNA segments averaging a total length of more than 13 kb encoding 13 viral proteins. Each segment contains several ORFs flanked by 5' and 3' UTRs, and are named small (S; 2.9 kb), medium (M; 4.1 kb) and large (L; 6.4 kb) according to their sizes. Using an *in vitro* packaging system, it was shown that $\phi 6$ (+) RNAs are inserted into a pre-formed procapsid core individually and sequentially through an entry portal at one five-fold icosahedral axis³¹. *In vitro* assays also demonstrated that the cis-acting RNA sequence and structural elements crucial for packaging are located in the 5' UTRs^{31,151}. A 5'-terminal 18 bp conserved sequence shared by the S, M and L segments, allows $\phi 6$ to distinguish between viral and host RNAs³¹. The gene-specific packaging signals differentiating S, M and L segments during packaging are located ~200 bp downstream of this 18 bp conserved sequence³¹.

Although there have been extensive studies on the packaging mechanism of the viruses mentioned above, the molecular mechanism of IAVs remains the most elusive. The global burden inflicted by seasonal and pandemic IAVs also necessitates a better understanding of this process. Therefore, genome packaging in segmented viruses has been studied the most in IAVs, and segment-specific packaging signals have also been identified. The section below describes a detailed review of this current knowledge on the genome packaging of IAVs. It details the virus structure, packaging signals mapped thus far on the IAV genome, and potential mechanisms of genome assembly that eventually culminate in the successful packaging of the eight-segmented RNA genome. It also gives an overview of the possible caveats in the current techniques used to study genome packaging in IAVs.

Critical Reviews and Perspectives

The influenza A virus genome packaging network — complex, flexible and yet unsolved

Celia Jakob^{1,2,†}, Rithu Paul-Stansilaus^{3,†}, Martin Schwemmle^{1,2,*}, Roland Marquet^{3,*} and Hardin Bolte^{1,2}

¹Institute of Virology, Medical Center – University of Freiburg, 79104 Freiburg, Germany, ²Faculty of Medicine, University of Freiburg, 79110 Freiburg, Germany and ³Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, IBMC, 67000 Strasbourg, France

Received June 02, 2022; Revised July 19, 2022; Editorial Decision July 22, 2022

ABSTRACT

The genome of influenza A virus (IAV) consists of eight unique viral RNA segments. This genome organization allows genetic reassortment between co-infecting IAV strains, whereby new IAVs with altered genome segment compositions emerge. While it is known that reassortment events can create pandemic IAVs, it remains impossible to anticipate reassortment outcomes with pandemic prospects. Recent research indicates that reassortment is promoted by a viral genome packaging mechanism that delivers the eight genome segments as a supramolecular complex into the virus particle. This finding holds promise of predicting pandemic IAVs by understanding the intermolecular interactions governing this genome packaging mechanism. Here, we critically review the prevailing mechanistic model postulating that IAV genome packaging is orchestrated by a network of intersegmental RNA–RNA interactions. Although we find supporting evidence, including segment-specific packaging signals and experimentally proposed RNA–RNA interaction networks, this mechanistic model remains debatable due to a current shortage of functionally validated intersegmental RNA–RNA interactions. We speculate that identifying such functional intersegmental RNA–RNA contacts might be hampered by limitations of the utilized probing techniques and the inherent complexity of the genome packaging mechanism. Nevertheless, we anticipate that improved probing strate-

gies combined with a mutagenesis-based validation could facilitate their discovery.

INTRODUCTION

The natural reservoir of influenza A viruses (IAVs) are aquatic birds, yet spill-over events have led to the introduction and subsequent establishment of numerous IAV lineages in other species, including swine, poultry, bats and humans (1,2). The ability of IAVs to cross inter-species barriers and adapt to new hosts is enabled by an enormous genetic variability which is brought about by reassortment events of the eight genome segments between co-infecting IAVs. Genetic reassortment has a great impact on IAV evolution and can create devastating pandemic IAVs, as exemplified by the 1957 (Asian flu), 1968 (Hong Kong flu) and 2009 (Swine flu) pandemic IAVs that originated via reassortment between avian, swine and human virus strains (3–5). As the frequency of reassortment events between avian and mammalian IAVs increases in porcine facilities (6,7), there is growing concern that novel reassortants could invade the human population and elicit the next flu pandemic (2). Thus, a profound understanding of the mechanisms underlying reassortment is urgently needed to predict and combat future pandemic IAVs.

Recent research findings suggest that reassortment is driven by a selective genome packaging mechanism that assembles the eight viral genome segments as a supramolecular complex into the virus particle, thereby allowing the exchange of cognate genome segments between co-infecting IAVs (8–10). The currently favoured mechanistic model postulates that this genome complex forms by means of a network of intersegmental RNA–RNA interactions. In this review, we will scrutinize this mechanistic model by criti-

[†]To whom correspondence should be addressed. Tel: +49 761 203 6526; Fax: +49 761 203 6626; Email: martin.schwemmle@uniklinik-freiburg.de
Correspondence may also be addressed to Roland Marquet. Tel: +33 3 88 41 70 54; Fax: +33 3 88 60 22 18; Email: r.marquet@ibmc-cnrs.unistra.fr

[‡]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

complex, flexible and yet unsolved

2 *Nucleic Acids Research*, 2022

cally reviewing the available body of evidence. We will describe the genome segment loci known to coordinate IAV genome packaging and discuss the significance of recently proposed RNA–RNA interaction networks obtained by high-throughput crosslinking experiments. Although we reveal ample evidence to support the prevailing mechanistic model, we also realize that the identification of functionally relevant RNA–RNA interactions between genome segments remains a major challenge. We would therefore like to stimulate critical rethinking of the experimental approaches used to study intersegmental RNA–RNA interactions and suggest potential avenues to identify functional intermolecular contacts involved in genome packaging and reassortment.

THE IAV GENOME PACKAGING MODEL AND ITS CHALLENGES

From single vRNPs to a supramolecular complex

The IAV genome comprises eight negative-sense, single-stranded viral RNA segments (vRNAs) that vary in length from 890 to 2341 nucleotides (11). All genome segments share the same structure, characterized by a broad central coding region flanked by two shorter non-coding regions (NCRs). The NCRs consist of conserved terminal regions (spanning 12 nucleotides at the 3' end and 13 nucleotides at the 5' end) and segment-specific portions that vary in length between 5 and 45 nucleotides (11–13) (Figure 1A). The conserved NCR termini and two adjacent nucleotides together form a panhandle structure that is bound by the heterotrimeric viral polymerase and serves as a promoter during genome replication (14–18). The remaining portions of the NCRs and the coding regions of the vRNAs are non-uniformly associated with multiple copies of the viral nucleoprotein (NP) (19–21), forming rod-shaped, helical viral ribonucleoproteins (vRNPs) (Figure 1B) (22–26).

Following viral infection, these vRNPs are released from the infecting virus particles into the cytoplasm and imported into the nucleus to be transcribed and replicated. Newly formed vRNPs are then exported out of the nucleus and transported to viral budding sites at the plasma membrane. The prevailing model assumes that during this transport, individual vRNPs are progressively assembled into an octameric supramolecular genome complex (Figure 1C) (27–33). While the specific cellular compartment hosting this assembly process remains unknown, (7 + 1) genome complexes in which seven vRNPs surround a central one are incorporated into viral particles at the plasma membrane (34–37). Throughout this review, these processes of genome complex formation and its subsequent incorporation into virions will be collectively referred to as genome packaging.

Past research suggests that IAV genome packaging is coordinated by a network of specific intersegmental RNA–RNA interactions that is formed by discrete genomic loci known as packaging signals. While initial studies have mapped these packaging signals towards the segment termini, more recent studies suggest that they are also present in internal vRNA regions lying beyond the vRNA termini. In the following sections, we review the current knowledge on packaging signals and revisit their presumed role in forming intersegmental RNA–RNA interactions.

A network of terminal packaging signals coordinates genome packaging

Terminal packaging signals were first proposed in studies analysing the genome content of defective interfering (DI) IAV particles. Unlike 'standard' virus particles which package eight full-length vRNAs, DI particles commonly package one genome segment that harbours a large internal deletion but retains the NCRs plus short portions of the adjacent coding regions (38,39). Such DI-RNAs have been described for all genome segments (40–43), yet often derive from vRNAs 1, 2 and 3 encoding viral polymerase subunits. In these cases, the truncated genomes do not give rise to a functional viral polymerase, which renders the DI particle replication-incompetent. However, during co-infection with infectious 'standard' virions when all viral proteins are expressed, these DI-RNAs are replicated and packaged into viral particles. Since DI-RNAs interfere with their full-length counterparts for replication and packaging under co-infection conditions, they can reduce the production of replication-competent 'standard' particles (44,45), a finding which coined their nomenclature. These early observations suggested that the retained terminal ends in DI-RNAs mediate vRNA incorporation into virions and thus contain packaging signals.

To map these proposed terminal packaging signals of each vRNA in detail, artificial genome segments were created in which a GFP reporter gene was flanked by the NCRs plus varying portions of the adjacent coding region of the studied vRNA (Figure 2A). In plasmid-based co-transfection assays, such artificial vRNAs were propagated by the IAV replication machinery and subsequently packaged into virus-like particles (VLPs) in the presence of the seven remaining wild-type vRNAs. Cells were subsequently infected with the released VLPs and a helper virus, and successful packaging events of these artificial vRNAs were detected by counting the number of GFP-positive cells. Such (7 + 1) VLP assays were performed with artificial vRNAs derived from all genome segments, allowing the systematic probing of the minimal terminal sequences required for vRNA packaging (46–55). While the exact nucleotide sequences varied depending on the genome segment under study, the NCRs and a minimum of 9–222 nucleotides of the adjacent 3' and 5' coding regions were necessary to efficiently package the reporter vRNAs into VLPs (8,56) (Figure 1A). These sequences were similar to those found in DI-RNAs (9), supporting the idea that segment-specific terminal packaging signals drive the incorporation of vRNAs.

After the discovery of terminal packaging signals, research was intensified to understand their mechanism of action. Since terminal packaging signals contain conserved nucleotide stretches, a series of WSN/H1N1 and PR8/H1N1 mutants were created, by altering either a single 3' or 5' terminal packaging signal of a given genome segment with synonymous nucleotide substitutions (48,57–65). These mutant viruses were then propagated in cell culture to assess the production of infectious particles and packaging of the eight vRNAs (Figure 2B). Intriguingly, many mutant viruses formed more non-infectious particles than a wild-type control virus in compensation for less infectious virions (58,60,62,63,65). While some mutants ineffi-

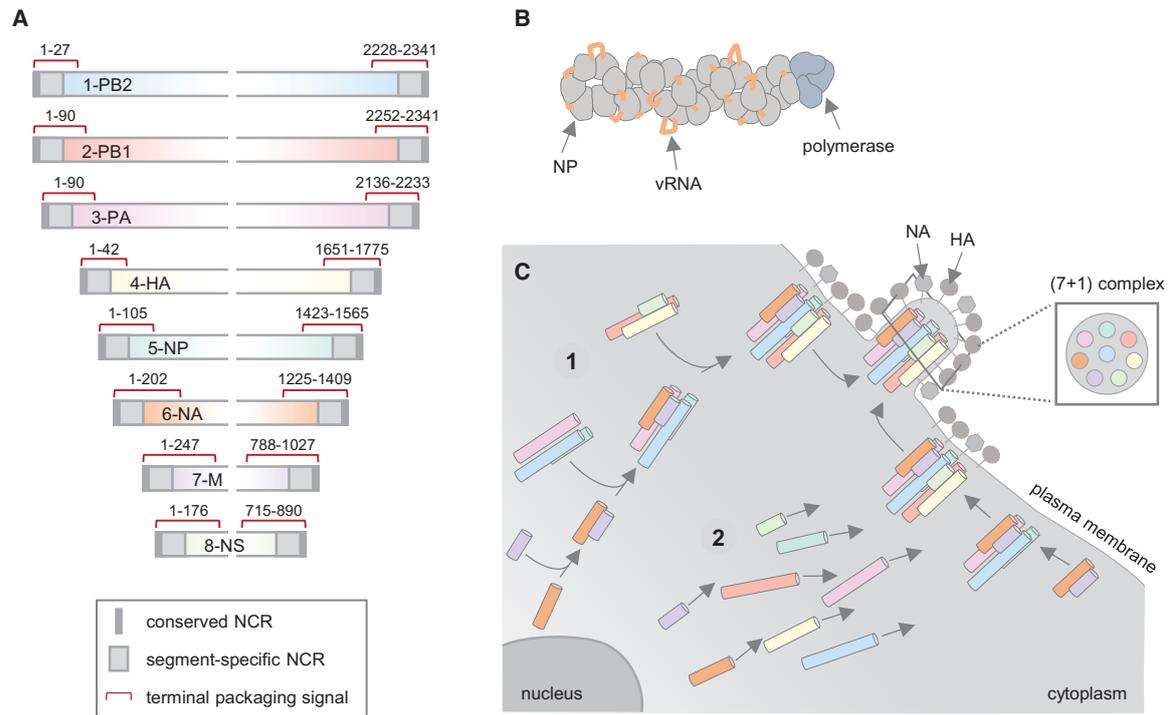


Figure 1. IAV genome structure and genome packaging model. (A) An illustration of the IAV genome based on the WSN/H1N1 strain. The genome segments are shown in negative-sense orientation from 3' to 5'. The identified terminal packaging signals are indicated with red lines. A detailed summary of all characterized packaging signals is presented in (8,56). The conserved 5' and 3' segment termini of the non-coding regions (NCRs) are highlighted in dark grey, whereas the segment-specific parts of the NCRs are indicated in light grey. PB, polymerase basic subunit; PA, polymerase acidic subunit; HA, hemagglutinin; NP, nucleoprotein; NA, neuraminidase; M, matrix; NS, non-structural. (B) An illustration of a viral ribonucleoprotein (vRNP). (C) The current genome packaging model proposes that the eight genome segments sequentially assemble into a (7 + 1) genome complex. This process might either occur (1) *en route* to the plasma membrane or (2) directly at the viral budding site, while the first scenario is currently preferred. The inset depicts a schematic cross-section through a budding virion showing the (7 + 1) arrangement of the vRNPs as observed by electron microscopy (34,35,36).

ciently packaged the mutated vRNA (57,59,61,63) (Figure 3A), many other mutants showed impaired packaging of multiple genome segments, which often, but not always, included the mutated one (57,59,60,62,63,64,65) (Figure 3B). In rare cases, mutants produced increased amounts of empty virions (65) (Figure 3C) or showed a 'random' packaging phenotype characterized by inefficient packaging of all eight vRNAs and the production of vast amounts of non-infectious virions (58). This range of different genome packaging defects suggested that the terminal packaging signals are involved in intricate vRNP–vRNP interactions that coordinate packaging of a genome complex.

Since 3' and 5' terminal packaging signals are present in all genome segments, it was envisioned that they collectively participate in a network of vRNP–vRNP interactions involving all vRNPs. Recently, Bolte and colleagues provided compelling evidence for this idea (63). Intrigued by the finding that single mutated terminal packaging signals in vRNAs 1, 2 or 3 provoked none or only minor genome packaging defects in SC35M/H7N7 (as opposed to the same mutations in WSN/H1N1 (59)), they combined these seemingly 'silent' mutations to create SC35M mutants with up to three mutated terminal packaging signals (Figure 3D). This approach revealed that the combination of two or three mutated terminal packaging signals caused the formation of non-infectious virions and reduced packaging of multiple

vRNAs unlike the single mutations, suggesting that terminal packaging signals are involved in a redundant network of vRNP–vRNP interactions that tolerates the loss of critical vRNP–vRNP contacts to some extent. Interestingly, the packaging phenotypes resulting from different combinations of mutated terminal packaging signals were generally unpredictable, albeit following certain patterns (63). This finding hinted at plastic rearrangements in the network of vRNP–vRNP interactions in response to the loss of certain interactions, potentially mediated by functionally redundant terminal packaging signals.

This apparent flexibility of the vRNP–vRNP interaction network is further supported by findings that mutated terminal packaging signals have varying effects on genome packaging depending on the analysed IAV strain (59,63). Thus, it is plausible that different IAVs use strain-specific vRNP–vRNP interaction networks which respond differently to the same mutation. Although speculative, a flexible rewiring of the interaction network might be achieved through specific combinations of conserved terminal packaging signals in vRNAs 1, 2, 3, 5 and 7 and subtype- or even strain-specific terminal packaging signals in vRNAs 4, 6 and 8 (48,51,66,67).

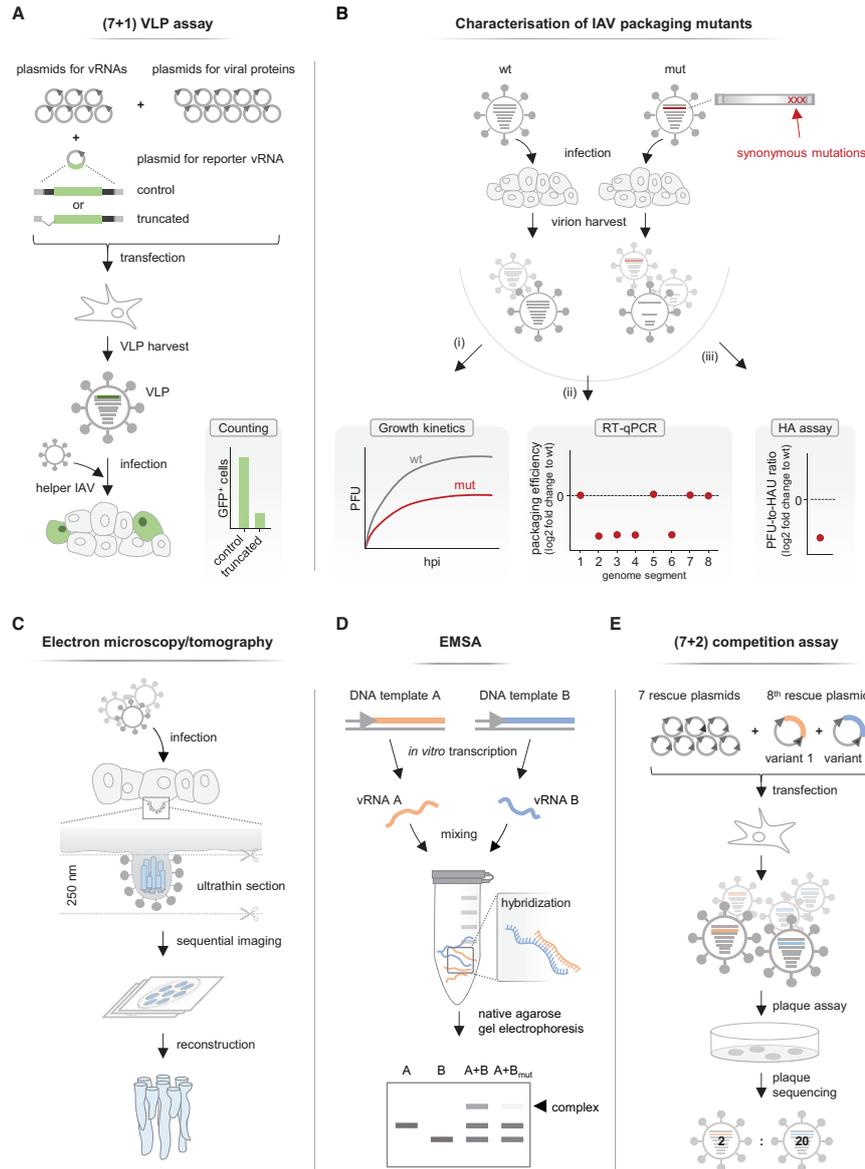


Figure 2. Established methods to study IAV genome packaging. **(A)** (7 + 1) virus like particle (VLP) assay. To generate VLPs, cells are co-transfected with ten plasmids encoding viral proteins, seven plasmids encoding full-length vRNAs and one plasmid encoding an artificial eighth vRNA which comprises a GFP-reporter gene flanked by the 5' and 3' ends including the non-coding regions (shown in grey) and parts of the coding region (shown in black) of the vRNA under study. VLPs released from transfected cells are subsequently used to infect new cells. Replication of the reporter vRNA is facilitated by superinfection with a helper IAV. The number of cells expressing GFP reflects the packaging efficiency of the reporter vRNA. Terminal packaging signals were mapped by shortening the coding region as indicated here by a truncated artificial vRNA. Further mappings were performed by introducing synonymous mutations into the terminal packaging signals (not shown). Variations of this (7 + 1) VLP assay are extensively described in (111). **(B)** Characterization of IAV packaging mutants. Cell cultures are infected either with wild-type virus (wt) or a mutant virus with synonymous mutations in a terminal packaging signal (mut). Newly formed viral particles are collected at various time points post-infection (hpi). Viral growth is monitored by determining the number of plaque-forming units (PFU) (i). A decrease of infectious particles can indicate impaired genome packaging. The vRNA amounts packaged into viral particles are measured using RT-qPCR and used to calculate relative packaging efficiencies (ii). A genome packaging defect of the mutant virus is often characterized by the relative loss of certain vRNAs in the virion population. The relative number of total particles can be determined using HA assay (iii). A decreased ratio of infectious particles (PFU) to total particles (measured as hemagglutination units [HAU]), is characteristic for a genome packaging defect. **(C)** Analysis of budding viral particles by electron microscopy and electron tomography. Infected cells are fixed, stained and embedded, and a tilt series of an ultrathin section is recorded using an electron microscope. These images can be used to count the number of vRNPs within single viral particles and to reconstruct a 3D presentation of the packaged genome complex. **(D)** Electrophoretic mobility shift assay (EMSA). Pairs of *in vitro* transcribed vRNAs are synthesized, mixed and analysed for vRNA-vRNA complex formation by native agarose gel electrophoresis. A size shift compared to single-vRNA controls indicates complex formation of the two vRNAs. Interaction sites can be mapped by mutating a putative interaction site in one partner vRNA (e.g. B_{mut}). Disruption of the vRNA-vRNA complex due to the mutations can be visualized as a loss of the size shift. **(E)** (7 + 2) competition assay. Cells are co-transfected with seven rescue plasmids encoding different vRNAs and two rescue plasmid variants coding for the missing eighth vRNA. These variants can either be a wild-type and a mutated vRNA or vRNA variants of different IAV strains. The released virus particles are subsequently plaque-purified and subjected to genotyping by sequencing. The preferentially packaged vRNA variant is found in the majority of viral plaques.

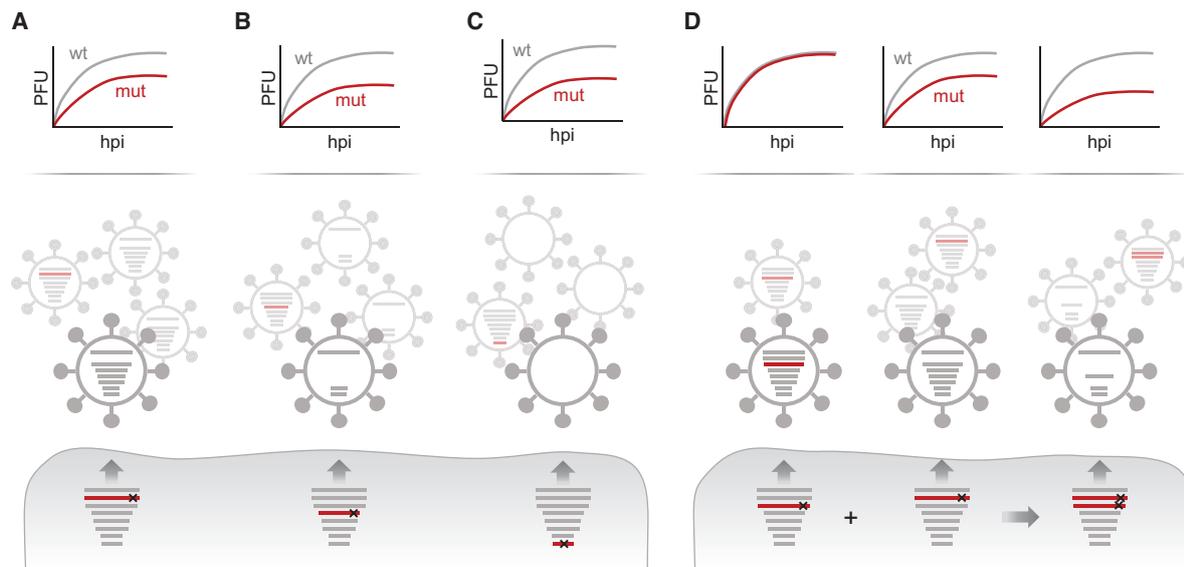


Figure 3. Packaging phenotypes of various IAV packaging mutants. Indicated are schematic growth kinetics (upper panel) and packaging phenotypes (middle panel) of IAVs with mutated packaging signals (mut) compared to a wild-type virus (wt). (A–C) Mutations within a packaging signal of one genome segment may cause reduced packaging of the respective vRNA (A), or multiple vRNAs (B), or all vRNAs (C) into the viral particle population resulting in reduced viral growth. (D) In some cases, mutations within a packaging signal of one genome segment do not obviously affect genome packaging and viral growth when compared to the wild-type virus. However, combination with another mutated (and thus packaging-defective) vRNA can provoke reduced packaging of several genome segments accompanied by impaired viral growth. PFU, plaque-forming units; hpi, hours post infection.

The mechanism of action of the terminal packaging signals remains unknown

Although numerous terminal packaging signals were identified, their mechanism of action has remained under investigation ever since. As they consist of RNA nucleotides, it was speculated that they form intersegmental RNA–RNA interactions. Early hints supporting this idea came from electron-tomography experiments that visualized electron-dense structures between neighbouring vRNPs in the (7 + 1) genome complex of budding viral particles (Figure 2C). Fournier and colleagues observed a ‘platform’ located at the top of this complex where the viral polymerases of the vRNPs are presumably located (36). Its size was sufficient to accommodate potential interactions between terminal packaging signals in the vicinity of the viral polymerases. In addition, Noda and colleagues found string-like structures that connected adjacent vRNPs all along their surface, indicating intersegmental contacts mediated by terminal packaging signals and potential internal vRNA regions (35). While these electron-dense structures might indeed represent RNA–RNA interactions between adjacent vRNPs, it has been a challenge to distinguish true RNA–RNA contacts from background noise due to the limited resolution of electron tomography.

The currently favoured mechanistic model proposes that the terminal packaging signals adopt local RNA secondary structure that loops out of the vRNPs and form sequence-specific intermolecular interactions. Such kissing-loop interactions have been previously observed for other viruses where they regulate various processes (68–73). In agreement with this mechanistic model, structural probing of *in vitro* transcribed vRNAs and computational predictions have shown that the terminal packaging signals of several

genome segments adopt defined RNA secondary structures (74–80). Moreover, SHAPE-MaP analysis (Figure 4A) of viral particles suggested that some local RNA structures are also present in vRNPs and that the 5′ terminal packaging signals tend to be more structured compared to adjacent internal vRNA regions (81). Finally, CLIP experiments (Figure 4B) indicated that certain parts of the vRNAs, including some terminal packaging signals, are relatively free of NP and thus able to fold into structural elements (19–21).

Some of these identified RNA structure elements were proven crucial for IAV genome packaging. For example, mutational studies confirmed the role of a pseudoknot residing in the 5′ terminal packaging signal of genome segment 5. Disruption of this structural element by mutagenesis caused attenuated viral growth and reduced packaging of multiple genome segments (21,62,74). In contrast, no such genome packaging defect was observed when mutations were designed to preserve or repair the pseudoknot. Similarly, Hagey *et al.* showed the biological significance of a conserved stem-loop within the 5′ terminal packaging signal of genome segment 1 (80). Disruption of this stem-loop either by mutagenesis or treatment with antisense-oligos led to reduced viral infectivity and a drop in the packaging efficiencies of segment 1 and multiple other genome segments. This genome packaging defect could be restored by repairing the stem-loop with compensatory mutations, proving the importance of this structural element. Despite these insights, secondary structures of other terminal packaging signals, especially in vRNPs, and their relevance in genome packaging remain poorly understood.

To identify intersegmental RNA–RNA interactions crucial for IAV genome packaging, Fournier and colleagues transcribed the eight vRNAs of Moscow/H3N2 *in vitro* and

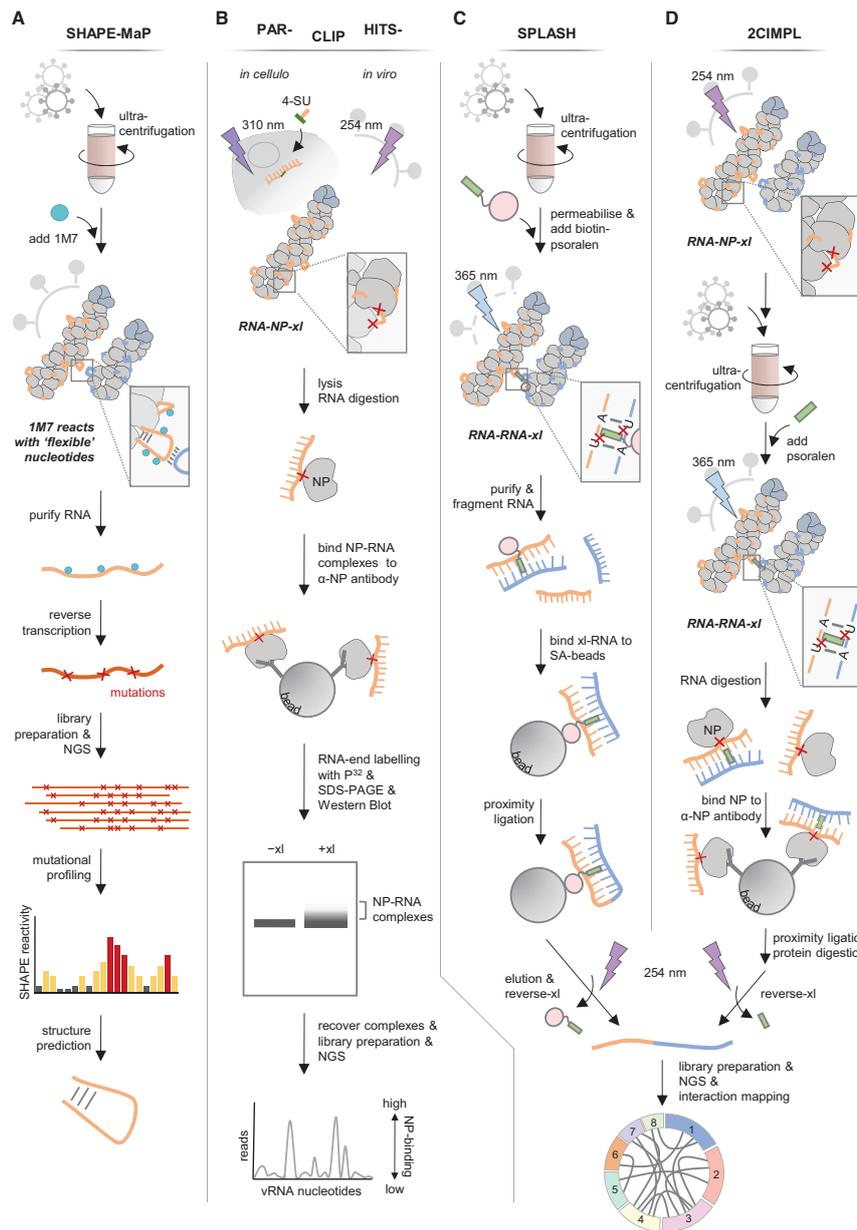


Figure 4. Novel high-throughput methods to study IAV genome packaging. **(A)** Selective 2'-hydroxyl acylation analysed by primer extension and mutational profiling (SHAPE-MaP). Purified viral particles are treated with the SHAPE reagent 1M7 that reacts with the sugar moiety of 'flexible' nucleotides. During reverse transcription of the purified RNA, point mutations are introduced at sites modified with 1M7. The resulting cDNA is subjected to library preparation and next generation sequencing (NGS). Mutation frequencies compared to a DMSO-treated control are used to measure SHAPE reactivities and to predict RNA secondary structures. **(B)** Photoactivatable ribonucleoside enhanced-crosslinking immunoprecipitation (PAR-CLIP) and high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP). NP-RNA complexes can be crosslinked in infected cells ('*in cellulo*') or virus particles ('*in viro*') by using either 4-thiouridine (4-SU) incorporation into newly synthesized RNA followed by UV irradiation at 310 nm or by directly using UV irradiation at 254 nm, respectively. Upon cell/virion lysis, RNA is partially digested, and NP-RNA complexes are enriched using anti-NP antibody coated beads. Radioactive end-labelling of the RNA with P^{32} facilitates visualization of NP-RNA complexes upon SDS-PAGE and Western Blot. NP-RNA complexes are recovered, and the RNA is isolated and subjected to library preparation and NGS. An increase in the normalized read coverage compared to a conventional RNA-seq library of the virus suggests NP-binding sites in the vRNA. **(C)** Sequencing of psoralen crosslinked, ligated, and selected hybrids (SPLASH). Viral particles are ultracentrifuged, permeabilized and treated with biotin-psoralen (biotin, pink circle; psoralen, green rectangle). Upon UV irradiation at 365 nm, psoralen crosslinks (xl), in particular, interstacked pyrimidines. Whole RNA is isolated, fragmented and enriched for biotinylated RNA using streptavidin (SA) beads. Proximity ligation of hybridized and psoralen-bound RNAs leads to the formation of chimeric RNAs. After crosslink-reversal at 254 nm, the RNA is subjected to library preparation followed by NGS. Chimeric RNAs are split-mapped to the IAV genome, and intersegmental RNA-RNA interactions are visualized in circos plots. **(D)** Dual crosslinking, immunoprecipitation, and proximity ligation (2CIMPL). Cell culture supernatant with virus particles is irradiated with 254-nm UV light to crosslink NP-RNA complexes. Virus particles are ultracentrifuged and treated with psoralen and 365-nm UV light to crosslink RNA-RNA interactions. RNA is partially digested, and NP-bound RNA is enriched using anti-NP antibody-coated beads. Proximity ligation of hybridized and NP-bound RNAs leads to the formation of chimeric RNAs. Subsequently, NP is digested and psoralen is removed by UV irradiation at 254 nm. The recovered RNA is subjected to library preparation and NGS. Chimeric RNAs are split-mapped to the IAV genome, and intersegmental RNA-RNA interactions are visualized in circos plots.

analysed their pair-wise interactions in an electrophoretic mobility shift assay (EMSA) (Figure 2D) (36,82). This approach identified numerous *in vitro* RNA–RNA interactions that could be combined into a network comprising all eight vRNAs. Subsequent experiments revealed that some of these *in vitro* interactions were formed by previously defined terminal packaging signals. For example, mutating the 5′ terminal packaging signal of segment 7 largely prevented an *in vitro* interaction between vRNAs 7 and 6. Importantly, these mutations also impaired packaging of the mutated vRNA 7 compared with a wild-type vRNA 7 in a (7 + 2) rescue assay (Figure 2E), suggesting a crucial role of this interaction site. However, the interaction site of vRNA 6 could not be pinpointed, as neither deleting its 3′ nor 5′ packaging signal disrupted the *in vitro* interaction, leaving the intermolecular interaction and its functional relevance elusive.

Further evidence that terminal packaging signals form intersegmental RNA–RNA interactions was provided by Essere and colleagues while studying reassortment events between Moscow/H3N2 and Finch/H5N2 viruses using a co-transfection-based rescue assay (83). They observed that vRNA 4 of the H5N2 virus was incorporated into the H3N2 genetic background only when combined with the H5N2-vRNA 7. The authors speculated that this co-segregation event was enabled by an RNA–RNA interaction between the H5N2-vRNAs 4 and 7 mediated by terminal packaging signals. Indeed, replacing the 3′ terminal packaging signal of the H3N2-vRNA 7 with that of the H5N2-vRNA 7 was sufficient to facilitate the co-segregation event. However, while the nucleotides of vRNA 7 involved in this putative interaction were mapped in detail, the partner region in vRNA 4 was not identified, which left the exact intersegmental RNA–RNA contact obscure.

Recently, Miyamoto and colleagues described a functional interplay between the terminal packaging signals of vRNAs 1 and 4 in WSN/H1N1 (65). By introducing synonymous mutations into the 5′ terminal packaging signal of vRNA 4 they created a mutant virus that failed to efficiently package vRNAs 4 and 6. Passaging of this packaging mutant in cell culture, however, selected a virus revertant with a wild-type-like packaging phenotype, as it acquired single point mutations in the 5′ terminal packaging signals of genome segments 4 and 1. Although this finding suggested that these modified terminal packaging signals re-established an intersegmental RNA–RNA interaction to alleviate the genome packaging defect, this hypothesis was not unambiguously supported by computational predictions and EMSAs.

In conclusion, the experimental data reviewed thus far suggest a crucial role of the terminal packaging signals in the formation of an octameric genome complex and its incorporation into the viral particle. For some genome segments, the terminal packaging signals are known to adopt RNA secondary structures responsible for coordinated genome packaging. Although a few studies using virus mutants suggest that terminal packaging signals establish RNA–RNA interactions between genome segments, this mechanistic concept remains to be conclusively proven.

Functional RNA–RNA interactions between internal vRNA regions are known but rare

While many studies have characterized terminal packaging signals, relatively little research has explored a potential role of internal vRNA regions in IAV genome packaging. Nevertheless, a few studies assessed the impact of internal vRNA deletions on virus replication and genome packaging (45,84,85,86). Recombinant viruses harbouring previously defined DI-vRNAs were generated and propagated in cells *trans*-complemented for the missing viral protein. Subsequent analyses of viral particle preparations by plaque assay (on the *trans*-complemented cells), HA-assay and RT-qPCR (Figure 2B) evaluated whether these clonal DI-viruses had growth and genome packaging defects. Characterization of a virus with a DI-vRNA 1 revealed that this truncated genome segment was inefficiently packaged into viral particles compared to vRNAs 5 and 8 (the other vRNAs were not tested) (45). Moreover, this DI-virus showed a reduced PFU-to-HAU ratio compared to the wild-type control virus, indicating that it formed non-infectious particles which lacked one or multiple full-length vRNAs (85). Together, these findings suggested that this DI-vRNA 1 lacks internal packaging signals crucial to coordinate genome packaging. Furthermore, a recent study found that DI-RNAs derived from genome segments 2, 3 and 4 are also inefficiently packaged into viral particles when compared to their full-length counterparts (43), suggesting the presence of internal packaging signals in these vRNAs. However, whether these DI-RNAs lower the packaging efficiencies of other full-length vRNAs like the tested DI-vRNA 1 remains to be investigated.

Although these deletion studies implied the existence of internal packaging signals, they could not dismiss the possibility that the internal vRNA deletions disturbed the correct folding and functioning of adjacent terminal packaging signals, thereby indirectly causing the observed genome packaging defects. Bolte and colleagues ruled out this ambiguity and identified a putative internal packaging signal by introducing synonymous mutations into a short conserved internal vRNA region of genome segment 3 in SC35M/H7N7 (63). While the exclusive mutation of this region did not provoke a genome packaging defect, combining it with mutations in the 5′ terminal packaging signal of vRNA 2 reduced packaging of four genome segments. Importantly, the virus harbouring only the latter mutations failed to exclusively package vRNA 2. This finding suggested that internal packaging signals indeed exist and contribute together with terminal packaging signals to the vRNP–vRNP interaction network.

While the molecular role of terminal packaging signals in IAV genome packaging remains disputable, a few internal vRNA regions have been proven to form intersegmental RNA–RNA interactions. Gavazzi and colleagues obtained first indications for this concept while investigating pair-wise interactions between *in vitro* transcribed vRNAs of Finch/H5N2 using an EMSA (87,88). They revealed that the eight vRNAs formed a complete intersegmental network that was reminiscent of the network obtained with Moscow/H3N2-derived vRNAs (82). However, in contrast to the *in vitro* interactions of the H3N2-vRNAs that mainly

involved terminal packaging signals, multiple *in vitro* interactions of the H5N2-vRNAs formed between internal vRNA regions. Deletion studies and computational predictions allowed the researchers to localize potential interaction sites in the vRNAs that could be later confirmed by disrupting the *in vitro* interactions using antisense-oligos. Importantly, two of these *in vitro* interactions were further validated using *trans*-complementary mutagenesis. In this approach, mutations introduced into either interaction partner disrupted the *in vitro* RNA–RNA interaction, whereas combining both mutated vRNAs restored it, proving that these vRNA regions establish specific intermolecular base-pairings. Despite these findings, none of the *in vitro* RNA–RNA interactions were validated to play a role in IAV genome packaging.

However, Gavazzi and colleagues identified in a subsequent study an additional *in vitro* RNA–RNA interaction between internal regions of vRNAs 2 and 8 of Finch/H5N2 (88) that was crucial to IAV genome packaging. By generating *trans*-complementary virus mutants (Figure 5) they could show that mutation of either interaction site decreased the PFU-to-HAU ratio compared to that of the wild-type virus, indicating formation of more non-infectious particles. RT-qPCR experiments revealed that the mutant viruses poorly packaged four vRNAs (the other four vRNAs were not tested), while EM analysis of budding virus particles showed large amounts of empty virions, suggesting that loss of the intersegmental RNA–RNA interaction reduced packaging of all eight vRNAs. Finally, the virus harbouring both mutated genome segments showed a restored PFU-to-HAU ratio and a reduced number of empty virions compared to the single-vRNA mutant viruses, proving that an intermolecular kissing interaction between the two genome segments is crucial for genome packaging. While this intersegmental RNA–RNA contact is important for genome packaging in Finch/H5N2, it may not be relevant to many other IAV strains as sequence analyses suggest that it is only partially conserved in other H5N2 strains and not conserved in other subtypes.

More evidence of intersegmental RNA–RNA interactions involving internal packaging signals was provided by studies investigating reassortment events during IAV vaccine production. IAV vaccines are usually produced in eggs by co-infection of an egg-adapted parental virus (e.g. PR8/H1N1) and a human isolate (e.g. Udorn/H3N2). The resulting reassortant viruses ideally replicate well in eggs and possess vRNAs 4 and 6 of the seasonal virus to elicit an immune response against the surface glycoproteins hemagglutinin and neuraminidase. While analysing reassortant viruses produced from co-infections between PR8 and Udorn, Cobbin and colleagues observed that the Udorn-vRNA 6 often co-segregated with the Udorn-vRNA 2 into the PR8 genetic background but rarely with the PR8-vRNA 2 (89). This imbalance, however, was surprising since both recombinant reassortants replicated efficiently in eggs. Subsequent (7 + 2) competition assays confirmed the preferential co-segregation event of the Udorn-vRNAs 6 and 2 in cell culture and revealed that it depended on a 300-nucleotide spanning region in Udorn-vRNA 2, which lies beyond the terminal packaging signals (90). As described in detail in the following section, a subsequent analysis could

pinpoint the exact nucleotides that establish this functional interaction between Udorn-vRNAs 2 and 6 (81).

In conclusion, two intersegmental RNA–RNA interactions between internal vRNA regions could be precisely mapped and validated to coordinate genome packaging. However, the short list of internal packaging signals, in contrast to the extensive list of terminal packaging signals, remains a challenge in evaluating their general role in genome packaging.

High-throughput probing of vRNA–vRNA interaction networks

Techniques coupling RNA–RNA crosslinking to next generation sequencing have recently enabled the high-throughput identification of intersegmental RNA–RNA interactions. These techniques commonly use psoralen derivatives which intercalate into double-stranded RNA regions and crosslink them upon UV irradiation. Ligation of the crosslinked RNA regions creates chimeric RNAs, which are subsequently reverse-transcribed and sequenced. Computational analysis of the chimeric reads recovers the initially crosslinked RNA–RNA interactions, which can be used to build up an interaction network with precise intermolecular base-pairings.

Dadonaite *et al.* performed sequencing of psoralen crosslinked, ligated, and selected hybrids (SPLASH) (Figure 4C) on purified viral particles of WSN/H1N1 (81). They identified an extensive, complex and redundant intersegmental RNA–RNA interaction network comprising hundreds of interactions connecting all eight vRNAs. Importantly, the interaction sites were not restricted to the terminal vRNA regions but distributed along the entire length of the genome segments. In that way, the contacts were either formed between terminal packaging signals, or between internal vRNA regions, or between both. However, the previously described terminal packaging signals showed varying detection frequencies, and many of them were even absent from the 50 most frequent RNA–RNA contacts of the network. Comparative analyses revealed that the SPLASH interaction networks of the closely related WSN/H1N1 and PR8/H1N1 strains were similar, sharing many interactions, albeit the detected frequencies of many overlapping interactions varied. In contrast, the SPLASH network of the distantly related Udorn/H3N2 strain was largely different from these H1N1 networks, sharing only very few contacts, suggesting that specific nucleotide stretches in the vRNAs determine the architecture of the SPLASH networks. In a similar approach, Le Sage *et al.* performed dual crosslinking, immunoprecipitation and proximity ligation (2CIMPL) (Figure 4D) using viral particles of WSN/H1N1 (91). While the 2CIMPL workflow also used psoralen, it implemented some changes compared to SPLASH, one of which was that it mapped RNA–RNA interactions forming between vRNA regions crosslinked to NP. The network identified by 2CIMPL also showed a complex and redundant architecture like the SPLASH network; however, despite using the same virus, only 10% of the identified intersegmental RNA–RNA interactions overlapped. This discrepancy could be due to the different workflows.

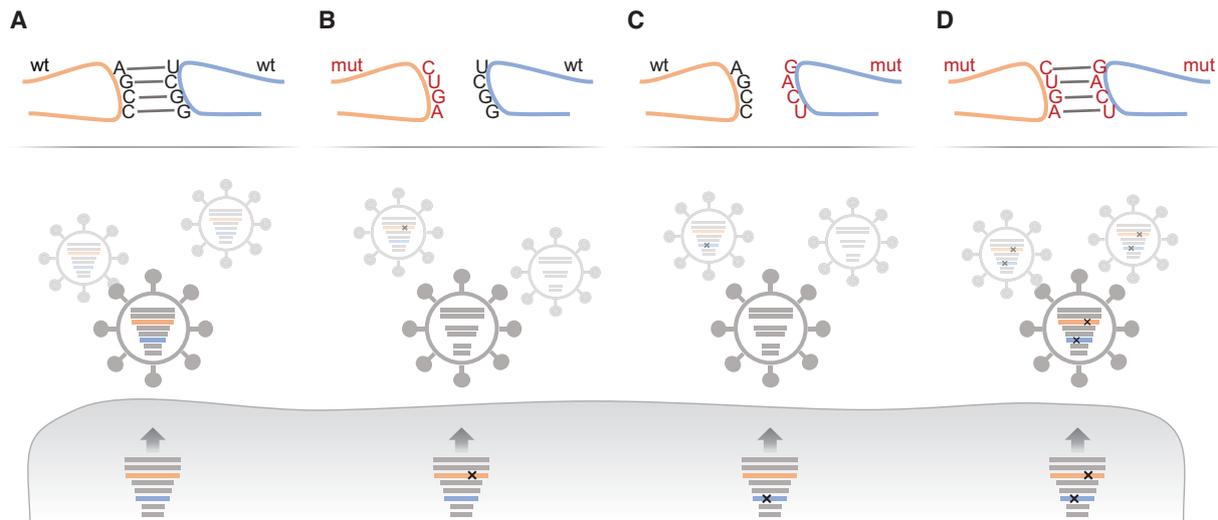


Figure 5. *Trans*-complementary virus mutants as tools to validate vRNA–vRNA interactions important for genome packaging. (A) In a wild-type virus, two vRNAs (in orange and blue) form an intersegmental RNA–RNA interaction that is required for genome packaging. (B, C) Introduction of synonymous mutations into either of the interaction partners abrogates this vRNA–vRNA interaction leading to impaired packaging of the mutated genome segments and possibly other vRNAs. (D) Combining the two mutated vRNAs from panels B and C repairs the intersegmental RNA–RNA interaction and restores genome packaging.

The SPLASH and 2CIMPL workflows identified hundreds of novel potential intersegmental RNA–RNA interactions; however, only few of these were assessed for their relevance in IAV genome packaging. Nevertheless, Dadonaite and colleagues could show in a (7 + 2) competition assay that a preferential co-segregation of Udorn-vRNAs 2 and 6 is mediated by specific base pairings between these two genome segments (81). They also confirmed that this interaction occurs in some other H3N2 viruses but is absent in Wyoming/H3N2 due to four nucleotide changes in the interacting site of vRNA 6. Changing these Wyoming-specific nucleotides to the Udorn-specific ones restored the interaction between the Wyoming-vRNAs 2 and 6 and allowed their preferential co-packaging. Importantly, SPLASH analysis of the respective reassortant viruses confirmed the absence or presence of this interaction. In addition, Le Sage *et al.* focused on a ‘hotspot’ region in vRNA 5 that interacted with multiple partner sites on different genome segments (91). Its mutation caused a genome-wide rearrangement of the intersegmental 2CIMPL network. Though this rearrangement was not accompanied by a detectable genome packaging defect, a potential compensatory function of newly established interactions was not addressed. Thus, albeit hundreds of novel intersegmental RNA–RNA interactions were discovered by SPLASH and 2CIMPL, their significance for genome packaging remains largely unknown.

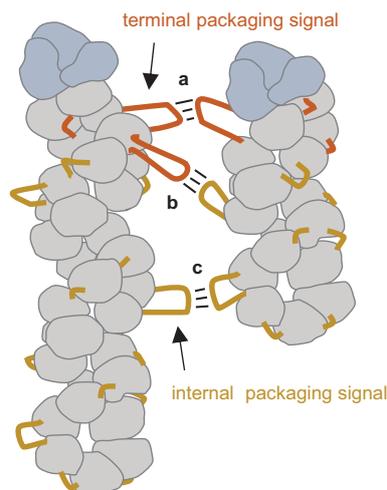
The prevailing genome packaging model currently fails the stress test

The eight IAV genome segments are known to be selectively packaged into viral particles as an octameric genome complex. Several lines of evidence support that this process is facilitated by an extensive and partially flexible network of intersegmental RNA–RNA interactions formed by terminal and internal packaging signals. However, two key as-

pects of this mechanistic model still lack conclusive evidence. Firstly, it has not yet been proven that the terminal packaging signals coordinate genome packaging by forming intersegmental RNA–RNA interactions (Figure 6). Although mutational studies and the SPLASH and 2CIMPL networks suggest that intersegmental RNA–RNA interactions involving terminal packaging signals exist, their relevance in genome packaging remains to be functionally validated. Likewise, only two intersegmental RNA–RNA interactions formed by internal vRNA regions were functionally proven so far (81,88,90). Since these RNA–RNA contacts are virus-strain specific, it is questionable whether interactions involving internal vRNA regions play a major role in genome packaging. Secondly, the relative contribution of terminal and internal packaging signals in genome packaging is currently unclear. While there is an extensive list of packaging mutants harbouring dysfunctional terminal packaging signals, relatively few packaging mutants with mutated internal vRNA regions are known. Nevertheless, it is possible that many more internal packaging signals exist in the IAV genome, and it is tempting to speculate that previous studies have overlooked them by focusing on the terminal vRNA regions due to their presence in DIs and their conservation across IAV strains. Recent studies showing that DI-RNAs are packaged less efficiently than their full-length counterparts indeed suggest the presence of internal packaging signals in many genome segments and pave the way for future studies to identify them.

TOWARDS A ROBUST GENOME PACKAGING MODEL: PITFALLS AND PROSPECTS

Despite recent progress, our current mechanistic understanding of the genome packaging process is not sufficient to fully accept the prevailing genome packaging model. To develop a robust understanding of the genome packaging mechanism we (i) propose avenues to test the biological sig-



(a) Interaction between terminal packaging signals

- (+) terminal packaging signals have been found in all eight vRNAs
- (-) however, there is little evidence for direct vRNA-vRNA interactions

(b) Interaction between terminal and internal packaging signals

- (-) vRNA-vRNA interactions are indicated by crosslinking studies; however, their biological relevance is unclear

(c) Interaction between internal packaging signals

- (+) two vRNA-vRNA interactions have been identified and validated
- (-) more vRNA-vRNA interactions are suggested by crosslinking studies; however, their biological relevance is unclear

Figure 6. Intersegmental RNA–RNA interactions in IAV genome packaging: an appealing model with open questions. For details see main text.

nificance of the experimentally postulated intersegmental RNA–RNA interaction networks, (ii) suggest to evaluate and improve the accuracy of the existent RNA–RNA interaction probing strategies and (iii) encourage the field to explore the potential role of the IAV nucleoprotein (NP) in modulating intersegmental RNA–RNA contacts.

Validation of proposed vRNA–vRNA interactions

One of the current challenges in IAV genome packaging is the shortage of validated intersegmental RNA–RNA interactions. An attractive avenue out of this problem might be to extend functional testing of interactions identified by SPLASH and 2CIMPL using *trans*-complementary virus mutants (Figure 5). Thus, the significance of a proposed intersegmental RNA–RNA interaction could be confirmed by showing that its disruption has negative impact on viral growth and genome packaging. Conversely, repairing the targeted interaction through *trans*-complementation would alleviate these defects and prove the base-pairing mechanism.

On the downside, this approach may be challenging. One issue could be that the disruption of a proposed RNA–RNA contact does not lead to a detectable genome packaging defect due to compensatory mechanisms such as mutation-induced global network rearrangements (91) or the presence of functionally redundant interactions (63). In these cases, it might be difficult to prove that the disrupted interaction is nonetheless crucial for genome packaging. While (7 + 2) competition assays may help reveal the impact of the disrupted interaction (81), combinatorial mutagenesis to disrupt the networks at multiple parts could also provide a solution to prove functional importance (63). Another inherent problem of *trans*-complementary mutagenesis is the limited range of available mutations. The nucleotide substitutions must not only be chosen to disrupt the interaction from both sites but also complement each other. In addition, the mutations should ideally be synonymous to preserve the function of the encoded viral proteins. Consequently, some interactions might not be readily confirmable as has been already previously noted (87). Nevertheless, pre-

diction programs can help in the design of suitable *trans*-complementary mutants (92). It is conceivable that this rigorous validation process will prove to be a Sisyphean task and yet, it is a promising option to substantiate the prevailing genome packaging model.

Finding accurate probing strategies

Apart from the complications discussed above, another problem in the validation process could be false-positive and false-negative RNA–RNA interactions. Mapping of RNA–RNA contacts using EMSAs with *in vitro* transcribed vRNAs were mostly performed in the absence of NP. However, inside viral particles and infected cells, vRNA is bound by NP which influences RNA secondary structure (81,93) and thus possibly also the formation of intersegmental RNA–RNA interactions. Consequently, EMSAs neglecting NP may miss crucial RNA–RNA contacts or identify non-functional ones. Furthermore, EMSAs have only analysed RNA–RNA interactions between pairs of genome segments so far. This artificial situation does not necessarily recapitulate RNA–RNA contacts between eight vRNPs and thus might allow RNA–RNA interactions that are precluded in the genome complex due to specific positioning of the vRNPs (35,36). These methodological problems could be reasons why only one of the many *in vitro* RNA–RNA interactions identified by EMSAs was found crucial in genome packaging.

The discovery of the SPLASH and 2CIMPL networks holds promise of identifying functionally relevant RNA–RNA interactions on a global scale. However, researchers should not assume that these networks represent the true RNA–RNA interaction networks coordinating genome packaging until proven. Indeed, psoralen-based identification workflows tend to introduce specific biases which possibly affect the finally recovered networks. For example, psoralen largely prefers to crosslink RNA–RNA interactions comprising staggered pyrimidines (94). Thus, other RNA–RNA contacts lacking this specific nucleotide composition and geometry are probably absent from the identified networks. Moreover, heavily crosslinked RNA–RNA

interactions tend to be lost during the RNA purification procedures used by SPLASH and 2CIMPL (95). This bias could result in a spurious underrepresentation of heavily crosslinked interactions in the obtained networks compared to sparsely crosslinked ones. Another problem might be ‘pseudo-interactions’ that form after the initial crosslinking step at later stages of the workflow through hybridization of single-stranded RNA regions, followed by ligation and detection. Though such ‘pseudo-interactions’ have not been demonstrated so far, the current workflows are not designed to exclude them or control for them. Ultimately, a combination of these and other biases (96) might skew the identified networks far away from the real ones. This might explain why only 10% of the SPLASH and 2CIMPL networks overlap and why these networks lack many of the previously characterized terminal packaging signals.

Such skewed networks would impose a mammoth task on researchers trying to validate interaction candidates by mutagenesis. Determined by the specific workflow, many packaging-relevant RNA–RNA interactions might be masked by a collection of ‘pseudo-interactions’. Likewise, the redundancy and thus mutational robustness of the true interaction network could be underestimated if many packaging-relevant interactions are missed because they are not crosslinked by psoralen. Together, these obstacles may complicate the identification of functional intersegmental RNA–RNA interactions.

Thus, to identify packaging-relevant RNA–RNA interactions, improved or even new strategies might have to be envisioned. While some biases of the SPLASH and 2CIMPL workflows are potentially eliminable (95), others such as the crosslinking preference of psoralen are not. Other probing techniques such as vRIC-seq (97) could be alternatives for the identification of RNA–RNA interactions; however, it is important to note that all currently available probing techniques probably have inherent biases, and therefore cannot draw an accurate picture of the real interaction network on their own. Nevertheless, comparing datasets obtained by multiple probing techniques across related virus strains using suitable statistical frameworks might help identify an overlapping set of candidate interactions that could play a conserved role in genome packaging. In addition, comparative analyses between wild-type viruses and IAVs with mutated terminal packaging signals may offer a shortcut for identifying packaging-relevant RNA–RNA contacts involving terminal vRNA regions.

Besides the technical limitations stated above, the biggest hurdle yet might be to discover the optimal probing material allowing the identification of packaging-relevant RNA–RNA interactions. In the current SPLASH and 2CIMPL workflows, RNA–RNA contacts are probed inside viral particles that have been released from infected cells and subsequently concentrated by ultracentrifugation. This strategy assumes that the intersegmental RNA–RNA interactions crucial for genome packaging are preserved under these conditions. However, it is documented that ultracentrifugation deforms viral particles (98) and possibly rearranges the genome complex. These structural rearrangements might be accompanied by the disruption of essential RNA–RNA interactions or the formation of artificial contacts which would contribute to a skewed interaction net-

work. A better probing material could be budding virus particles because they contain well-organized (7 + 1) genome complexes wherein adjacent vRNPs interact via string-like structures that potentially represent packaging-relevant RNA–RNA interactions (35,36). However, if these contacts are preserved in released viral particles remains disputable as virions shrink after being released from cells (34,35,99), which may induce ‘bending’ of the longest vRNPs and subsequent rearrangements of the genome complex (100). Finally, infected cells could be used to probe intersegmental RNA–RNA interactions. Ideal probing environments could be liquid organelles that form in the cytoplasm during the late phase of infection and probably host IAV genome assembly (32). Though attractive, probing of RNA–RNA interactions inside confined environments such as liquid organelles or budding viral particles would require new sophisticated techniques.

In conclusion, the identification of functional intersegmental RNA–RNA contacts will greatly depend on the accuracy of the applied probing strategies, and mutational analyses will be an important tool to benchmark them. Improvements of the existent probing techniques and development of novel strategies could finally help paint a clear picture of the intersegmental RNA–RNA networks that control genome packaging and reassortment.

Exploring the potential role of NP in modulating vRNA–vRNA interactions

Only recently, it was recognized that in addition to terminal and internal packaging signals, NP also serves a critical role in genome packaging. NP is the main protein component of vRNPs and consists of a head domain, a body domain, and a flexible tail loop (101). During genome replication, multiple NP molecules oligomerize on the nascent vRNA by inserting the tail loop into an insertion pocket in the body domain of another NP. This NP–vRNA complex folds back and twists around itself to form a helical vRNP together with the viral polymerase (Figure 1B) (23,24,26). Although the details of the vRNP structure are poorly understood, NP likely binds the negatively charged sugar-phosphate backbone of the vRNA through a positively charged RNA-binding groove located between the NP head and body domains, thereby presenting the bases of the bound vRNA outward of the vRNP (93,101–103).

By mutagenesis, Moreira and colleagues identified conserved amino acid residues in the NP head and body domains crucial for genome packaging (104). In their approach, they generated SC35M/H7N7 viruses, in which either seven NP-head domain residues (rNP7) or 18 NP-body domain residues (rCH2) were replaced with the corresponding ones of a distantly related bat-born IAV of the H17N10 subtype. Viral growth and RT-qPCR analyses revealed that these viruses with NP amino acid substitutions produced many non-infectious virions due to reduced packaging of multiple vRNAs. In this way, these NP mutant viruses were reminiscent of SC35M viruses with multiple mutated terminal packaging signals (63), suggesting that both types of alterations impaired the same underlying mechanism. Importantly, the poorly packaged vRNA subsets varied between the NP mutant viruses, indicating that each set of

amino acid substitutions disrupted a distinct set of vRNP–vRNP interactions. Additionally, Moreira and colleagues discovered an rNP7-R31G virus revertant with an extra amino acid substitution in the NP-body domain which showed wild-type-like genome packaging (104), suggesting repaired vRNP–vRNP contacts. However, Bolte and colleagues found that adding single mutated terminal packaging signals to the rNP7-R31G genetic background reduced packaging of multiple vRNAs, whereas adding them to the wild-type SC35M genetic background had little or no effect on genome packaging (63), indicating that the rNP7-R31G revertant virus established a distinct vRNP–vRNP interaction network.

These findings have established a crucial role of NP in selective genome packaging; however, the underlying molecular mechanism remains speculative. An attractive scenario is that binding of NP to the vRNAs helps them to adopt their native structure which is crucial to expose packaging signals and establish intersegmental RNA–RNA interactions. This modulatory role of NP is supported by two observations: firstly, SHAPE analyses suggested that although the vRNA structure is mainly determined by its sequence, NP can induce some local structural changes upon binding to the vRNA (81). Secondly, CLIP studies found that the eight vRNAs are non-uniformly bound by NP and retain unbound regions (19–21). Taking these findings together, it is plausible that NP binds to specific vRNA regions and thereby allows neighbouring regions such as packaging signals to remain free and adopt local secondary structures to participate in intersegmental RNA–RNA interactions.

While there is only very limited structural information available on NP–RNA interactions (26,103), multiple amino acid residues within the putative RNA-binding groove of NP have been functionally mapped (101,105–109). Interestingly, some of the rNP7 amino acid residues identified by Moreira and colleagues are identical with or located close to these putative RNA-binding residues (104). Consequently, it is possible that their replacement alters the affinity of NP towards specific vRNA regions and induces vRNA structural changes that impede crucial intersegmental RNA–RNA interactions. This is further supported by the observation that alanine substitutions of basic amino acid residues in the putative RNA-binding groove impair genome packaging (110). Likewise, some of the amino acid residues altered in the rCH2 mutant are located inside or in proximity to an accessory RNA-binding region of NP (104,108) and thus might similarly alter specific vRNA structures and intersegmental RNA–RNA contacts.

Some amino acid residues identified to be important for genome packaging do not lie within known RNA-binding regions of NP (104), suggesting that they are not directly involved in RNA binding. It is possible that these NP amino acid residues influence vRNA structuring through NP–NP interactions that control the relative positioning of NP molecules and their RNA-binding regions within vRNPs. Since RNA-binding is distributed across multiple NP molecules in vRNPs, the overall configuration of the NP backbone might be involved in structuring the bound vRNAs in their entirety. While early cryo-EM studies revealed a rigid configuration of the NP backbone showing a regular helical structure (23,24), recent cryo-EM studies identi-

fied the NP backbone to be structurally flexible and contain NP molecules with distinct orientations (25,26). These flexible NP orientations may place RNA-binding regions at specific positions in the NP backbone, thereby guiding which vRNA regions are bound by NP and helping the encapsidated vRNA to find its native structure that exposes packaging signals for intersegmental RNA–RNA interactions. Consequently, certain NP amino acid substitutions could disrupt essential intersegmental RNA–RNA contacts by changing the NP backbone configuration through altered NP–NP interactions.

These mechanistic possibilities have yet to be explored, and it will be critical in the future to test whether NP mutant viruses, such as those found by Moreira and colleagues, display alterations in vRNA structure and intersegmental RNA–RNA interactions responsible for the observed genome packaging defects. The success of these studies will depend on accurate techniques to probe intersegmental RNA–RNA interactions as discussed in the preceding sections. However, understanding how NP possibly modulates vRNA–vRNA interactions will require a broader panel of sophisticated and accurate techniques that is suitable to decipher additional changes in RNA–NP and NP–NP interactions as well as changes in the vRNP configuration between wild-type and NP mutant viruses.

CONCLUDING REMARKS

Here, we provided a comprehensive description of known packaging signals in the IAV genome segments and challenged the prevailing mechanistic model that they establish a specific, yet flexible network of intersegmental RNA–RNA interactions. This mechanistic model emerged with the discovery of the terminal packaging signals that provided an intuitive explanation for how mutually interacting genome segments could be packaged into virus particles in the form of a supramolecular complex. Eventually, the discovery of two functional RNA–RNA interactions between internal vRNA regions has provided a proof of this mechanistic concept. However, the lack of functional vRNA–vRNA interactions formed by terminal packaging signals remains a major weak point that needs to be addressed in the future.

While crosslinking-based RNA–RNA interaction probing techniques hold promise of identifying additional functional vRNA–vRNA contacts, it becomes evident that they might suffer from biases that portray a distorted image of the vRNA–vRNA interaction networks coordinating IAV genome packaging. Nevertheless, careful consideration of these biases and improved experimental designs coupled with *trans*-complementary mutagenesis may eventually expand the limited set of validated intersegmental RNA–RNA contacts and clarify the roles played by terminal packaging signals and internal regions in IAV genome packaging.

The recent discovery that NP is involved in IAV genome packaging suggests an additional level of mechanistic complexity that awaits future investigation. While the precise role of NP is currently unclear, NP mutant viruses with genome packaging defects could serve as valuable tools

to identify functional RNA–RNA interactions involved in IAV genome packaging.

Finally, we would like to mention that (i) the roles played by other viral proteins and host cell factors in IAV genome packaging and (ii) a systematic analysis of the current bioinformatics approaches to predict and study the involved intersegmental RNA–RNA interactions lie beyond the scope of this review and therefore have not been reviewed here. Understandably, insights obtained on these aspects could contribute significantly to our current knowledge of IAV genome packaging and genetic reassortment.

FUNDING

FluCode Project [ProjetANR-19-CE11-0027] funded by the French National Research Agency (ANR) and German Research Foundation (DFG) [to C.J., H.B., M.S., R.M.]; European Union's Horizon 2020 research and innovation program, under the Marie Skłodowska-Curie Actions Innovative Training Networks [955974 (VIROINF) to R.P.S.]. The contents of this manuscript solely reflect the authors' personal views on the field and do not reflect the views of the agencies involved. Funding for open access charge: FluCode Project [ProjetANR-19-CE11-0027].

Conflict of interest statement. None declared.

REFERENCES

- Taubenberger, J.K. and Kash, J.C. (2010) Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe*, **7**, 440–451.
- Kessler, S., Harder, T.C., Schwemmler, M. and Ciminski, K. (2021) Influenza A viruses and zoonotic events—are we creating our own reservoirs? *Viruses*, **13**, 2250.
- Smith, G.J., Vijaykrishna, D., Bahl, J., Lycett, S.J., Worobey, M., Pybus, O.G., Ma, S.K., Cheung, C.L., Raghvani, J., Bhatt, S. *et al.* (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*, **459**, 1122–1125.
- Viboud, C., Simonsen, L., Fuentes, R., Flores, J., Miller, M.A. and Chowell, G. (2016) Global mortality impact of the 1957–1959 influenza pandemic. *J. Infect. Dis.*, **213**, 738–745.
- Honigsbaum, M. (2020) Revisiting the 1957 and 1968 influenza pandemics. *Lancet*, **395**, 1824–1826.
- Henritzi, D., Petric, P.P., Lewis, N.S., Graaf, A., Pessia, A., Starick, E., Breithaupt, A., Strebel, G., Luttermann, C., Parker, L.M.K. *et al.* (2020) Surveillance of European domestic pig populations identifies an emerging reservoir of potentially zoonotic swine influenza A viruses. *Cell Host Microbe*, **28**, 614–627.
- Sun, H., Xiao, Y., Liu, J., Wang, D., Li, F., Wang, C., Li, C., Zhu, J., Song, J., Sun, H. *et al.* (2020) Prevalent Eurasian avian-like H1N1 swine influenza virus with 2009 pandemic viral genes facilitating human infection. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 17204–17210.
- Gerber, M., Isel, C., Moules, V. and Marquet, R. (2014) Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol.*, **22**, 446–455.
- Hutchinson, E.C., von Kirchbach, J.C., Gog, J.R. and Digard, P. (2010) Genome packaging in influenza A virus. *J. Gen. Virol.*, **91**, 313–328.
- Lowen, A.C. (2017) Constraints, drivers, and implications of influenza A virus reassortment. *Annu Rev Virol.*, **4**, 105–121.
- Ferhadian, D., Contrant, M., Printz-Schweigert, A., Smyth, R.P., Paillart, J.C. and Marquet, R. (2018) Structural and functional motifs in influenza virus RNAs. *Front Microbiol.*, **9**, 559.
- Zhao, L., Peng, Y., Zhou, K., Cao, M., Wang, J., Wang, X., Jiang, T. and Deng, T. (2014) New insights into the nonconserved noncoding region of the subtype-determinant hemagglutinin and neuraminidase segments of influenza A viruses. *J. Virol.*, **88**, 11493–11503.
- Benkaroun, J., Robertson, G.J., Whitney, H. and Lang, A.S. (2018) Analysis of the variability in the non-coding regions of influenza A viruses. *Vet Sci*, **5**, 76.
- Bae, S.H., Cheong, H.K., Lee, J.H., Cheong, C., Kainosho, M. and Choi, B.S. (2001) Structural features of an influenza virus promoter and their implications for viral RNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10602–10607.
- Park, C.J., Bae, S.H., Lee, M.K., Varani, G. and Choi, B.S. (2003) Solution structure of the influenza A virus cRNA promoter: implications for differential recognition of viral promoter structures by RNA-dependent RNA polymerase. *Nucleic Acids Res.*, **31**, 2824–2832.
- Pflug, A., Guilligay, D., Reich, S. and Cusack, S. (2014) Structure of influenza A polymerase bound to the viral RNA promoter. *Nature*, **516**, 355–360.
- Fan, H., Walker, A.P., Carrique, L., Keown, J.R., Serna Martin, I., Karia, D., Sharps, J., Hengrung, N., Pardon, E., Steyaert, J. *et al.* (2019) Structures of influenza A virus RNA polymerase offer insight into viral genome replication. *Nature*, **573**, 287–290.
- Fodor, E. and Te Velthuis, A.J.W. (2020) Structure and function of the influenza virus transcription and replication machinery. *Cold Spring Harb. Perspect. Med.*, **10**, a038398.
- Lee, N., Le Sage, V., Nanni, A.V., Snyder, D.J., Cooper, V.S. and Lakdawala, S.S. (2017) Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic Acids Res.*, **45**, 8968–8977.
- Le Sage, V., Nanni, A.V., Bhagwat, A.R., Snyder, D.J., Cooper, V.S., Lakdawala, S.S. and Lee, N. (2018) Non-uniform and non-random binding of nucleoprotein to influenza A and B viral RNA. *Viruses*, **10**, 522.
- Williams, G.D., Townsend, D., Wylie, K.M., Kim, P.J., Amarasinghe, G.K., Kutluay, S.B. and Boon, A.C.M. (2018) Nucleotide resolution mapping of influenza A virus nucleoprotein-RNA interactions reveals RNA features required for replication. *Nat. Commun.*, **9**, 465.
- Compans, R.W., Content, J. and Duesberg, P.H. (1972) Structure of the ribonucleoprotein of influenza virus. *J. Virol.*, **10**, 795–800.
- Arranz, R., Coloma, R., Chichon, F.J., Conesa, J.J., Carrascosa, J.L., Valpuesta, J.M., Ortin, J. and Martin-Benito, J. (2012) The structure of native influenza virion ribonucleoproteins. *Science*, **338**, 1634–1637.
- Moeller, A., Kirchoefer, R.N., Potter, C.S., Carragher, B. and Wilson, I.A. (2012) Organization of the influenza virus replication machinery. *Science*, **338**, 1631–1634.
- Gallagher, J.R., Torian, U., McCraw, D.M. and Harris, A.K. (2017) Structural studies of influenza virus RNPs by electron microscopy indicate molecular conortions within NP supra-structures. *J. Struct. Biol.*, **197**, 294–307.
- Coloma, R., Arranz, R., de la Rosa-Trevin, J.M., Sorzano, C.O.S., Munier, S., Carlero, D., Naffakh, N., Ortin, J. and Martin-Benito, J. (2020) Structural insights into influenza A virus ribonucleoproteins reveal a processive helical track as transcription mechanism. *Nat. Microbiol.*, **5**, 727–734.
- Chou, Y.Y., Heaton, N.S., Gao, Q., Palese, P., Singer, R.H. and Lionnet, T. (2013) Colocalization of different influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS Pathog.*, **9**, e1003358.
- Lakdawala, S.S., Wu, Y., Wawrzusin, P., Kabat, J., Broadbent, A.J., Lamirande, E.W., Fodor, E., Altan-Bonnet, N., Shroff, H. and Subbarao, K. (2014) Influenza A virus assembly intermediates fuse in the cytoplasm. *PLoS Pathog.*, **10**, e1003971.
- Giese, S., Bolte, H. and Schwemmler, M. (2016) The feat of packaging eight unique genome segments. *Viruses*, **8**, 165.
- de Castro Martin, I.F., Fournier, G., Sachse, M., Pizarro-Cerda, J., Risco, C. and Naffakh, N. (2017) Influenza virus genome reaches the plasma membrane via a modified endoplasmic reticulum and Rab11-dependent vesicles. *Nat. Commun.*, **8**, 1396.
- Vale-Costa, S. and Amorim, M.J. (2017) Clustering of rab11 vesicles in influenza A virus infected cells creates hotspots containing the 8 viral ribonucleoproteins. *Small GTPases*, **8**, 71–77.
- Alenquer, M., Vale-Costa, S., Etibor, T.A., Ferreira, F., Sousa, A.L. and Amorim, M.J. (2019) Influenza A virus ribonucleoproteins form

- liquid organelles at endoplasmic reticulum exit sites. *Nat. Commun.*, **10**, 1629.
33. Haralampiev, I., Prisner, S., Nitzan, M., Schade, M., Jolmes, F., Schreiber, M., Loidolt-Krüger, M., Jongen, K., Chamiolo, J., Nilson, N. *et al.* (2020) Selective flexible packaging pathways of the segmented genome of influenza a virus. *Nat. Commun.*, **11**, 4355.
 34. Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R.H. and Kawaoka, Y. (2006) Architecture of ribonucleoprotein complexes in influenza a virus particles. *Nature*, **439**, 490–492.
 35. Noda, T., Sugita, Y., Aoyama, K., Hirase, A., Kawakami, E., Miyazawa, A., Sagara, H. and Kawaoka, Y. (2012) Three-dimensional analysis of ribonucleoprotein complexes in influenza a virus. *Nat. Commun.*, **3**, 639.
 36. Fournier, E., Moules, V., Essere, B., Paillart, J.C., Sirbat, J.D., Isel, C., Cavalier, A., Rolland, J.P., Thomas, D., Lina, B. *et al.* (2012) A supramolecular assembly formed by influenza a virus genomic RNA segments. *Nucleic Acids Res.*, **40**, 2197–2209.
 37. Nakatsu, S., Sagara, H., Sakai-Tagawa, Y., Sugaya, N., Noda, T. and Kawaoka, Y. (2016) Complete and incomplete genome packaging of influenza a and b viruses. *MBio*, **7**, e01248-16.
 38. Jennings, P.A., Finch, J.T., Winter, G. and Robertson, J.S. (1983) Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell*, **34**, 619–627.
 39. Dimmock, N.J. and Easton, A.J. (2014) Defective interfering influenza virus RNAs: time to reevaluate their clinical potential as broad-spectrum antivirals? *J. Virol.*, **88**, 5217–5227.
 40. Nayak, D.P., Chambers, T.M. and Akkina, R.K. (1985) Defective-interfering (DI) RNAs of influenza viruses: origin, structure, expression, and interference. *Curr. Top. Microbiol. Immunol.*, **114**, 103–151.
 41. Pelz, L., Rüdiger, D., Dogra, T., Alnaji, F.G., Genzel, Y., Brooke, C.B., Kupke, S.Y. and Reichl, U. (2021) Semi-continuous propagation of influenza a virus and its defective interfering particles: analyzing the dynamic competition to select candidates for antiviral therapy. *J. Virol.*, **95**, e0117421.
 42. Mendes, M. and Russell, A.B. (2021) Library-based analysis reveals segment and length dependent characteristics of defective influenza genomes. *PLoS Pathog.*, **17**, e1010125.
 43. Alnaji, F.G., Reiser, W.K., Rivera-Cardona, J., Te Velthuis, A.J.W. and Brooke, C.B. (2021) Influenza a virus defective viral genomes are inefficiently packaged into virions relative to wild-type genomic RNAs. *Mbio*, **12**, e0295921.
 44. Huang, A.S. and Baltimore, D. (1970) Defective viral particles and viral disease processes. *Nature*, **226**, 325–327.
 45. Hein, M.D., Arora, P., Marichal-Gallardo, P., Winkler, M., Genzel, Y., Pöhlmann, S., Schughart, K., Kupke, S.Y. and Reichl, U. (2021) Cell culture-based production and in vivo characterization of purely clonal defective interfering influenza virus particles. *BMC Biol.*, **19**, 91.
 46. Fujii, Y., Goto, H., Watanabe, T., Yoshida, T. and Kawaoka, Y. (2003) Selective incorporation of influenza virus RNA segments into virions. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 2002–2007.
 47. Watanabe, T., Watanabe, S., Noda, T., Fujii, Y. and Kawaoka, Y. (2003) Exploitation of nucleic acid packaging signals to generate a novel influenza virus-based vector stably expressing two foreign genes. *J. Virol.*, **77**, 10575–10583.
 48. Fujii, K., Fujii, Y., Noda, T., Muramoto, Y., Watanabe, T., Takada, A., Goto, H., Horimoto, T. and Kawaoka, Y. (2005) Importance of both the coding and the segment-specific noncoding regions of the influenza a virus NS segment for its efficient incorporation into virions. *J. Virol.*, **79**, 3766–3774.
 49. Liang, Y., Hong, Y. and Parslow, T.G. (2005) cis-Acting packaging signals in the influenza virus PB1, PB2, and PA genomic RNA segments. *J. Virol.*, **79**, 10348–10355.
 50. Muramoto, Y., Takada, A., Fujii, K., Noda, T., Iwatsuki-Horimoto, K., Watanabe, S., Horimoto, T., Kida, H. and Kawaoka, Y. (2006) Hierarchy among viral RNA (vRNA) segments in their role in vRNA incorporation into influenza a virions. *J. Virol.*, **80**, 2318–2325.
 51. Gog, J.R., Afonso Edos, S., Dalton, R.M., Leclercq, I., Tiley, L., Elton, D., von Kirchbach, J.C., Naffakh, N., Escriou, N. and Digard, P. (2007) Codon conservation in the influenza a virus genome defines RNA packaging signals. *Nucleic Acids Res.*, **35**, 1897–1907.
 52. Ozawa, M., Fujii, K., Muramoto, Y., Yamada, S., Yamayoshi, S., Takada, A., Goto, H., Horimoto, T. and Kawaoka, Y. (2007) Contributions of two nuclear localization signals of influenza a virus nucleoprotein to viral replication. *J. Virol.*, **81**, 30–41.
 53. Liang, Y., Huang, T., Ly, H., Parslow, T.G. and Liang, Y. (2008) Mutational analyses of packaging signals in influenza virus PA, PB1, and PB2 genomic RNA segments. *J. Virol.*, **82**, 229–236.
 54. Ozawa, M., Maeda, J., Iwatsuki-Horimoto, K., Watanabe, S., Goto, H., Horimoto, T. and Kawaoka, Y. (2009) Nucleotide sequence requirements at the 5' end of the influenza a virus m RNA segment for efficient virus replication. *J. Virol.*, **83**, 3384–3388.
 55. Goto, H., Muramoto, Y., Noda, T. and Kawaoka, Y. (2013) The genome-packaging signal of the influenza a virus genome comprises a genome incorporation signal and a genome-bundling signal. *J. Virol.*, **87**, 11316–11322.
 56. Li, X., Gu, M., Zheng, Q., Gao, R. and Liu, X. (2021) Packaging signal of influenza a virus. *Virol J*, **18**, 36.
 57. Marsh, G.A., Hatami, R. and Palese, P. (2007) Specific residues of the influenza a virus hemagglutinin viral RNA are important for efficient packaging into budding virions. *J. Virol.*, **81**, 9727–9736.
 58. Hutchinson, E.C., Curran, M.D., Read, E.K., Gog, J.R. and Digard, P. (2008) Mutational analysis of cis-acting RNA signals in segment 7 of influenza a virus. *J. Virol.*, **82**, 11869–11879.
 59. Marsh, G.A., Rabadan, R., Levine, A.J. and Palese, P. (2008) Highly conserved regions of influenza a virus polymerase gene segments are critical for efficient viral RNA packaging. *J. Virol.*, **82**, 2295–2304.
 60. Hutchinson, E.C., Wise, H.M., Kudryavtseva, K., Curran, M.D. and Digard, P. (2009) Characterisation of influenza a viruses with mutations in segment 5 packaging signals. *Vaccine*, **27**, 6270–6275.
 61. Wise, H.M., Barbezange, C., Jagger, B.W., Dalton, R.M., Gog, J.R., Curran, M.D., Taubenberger, J.K., Anderson, E.C. and Digard, P. (2011) Overlapping signals for translational regulation and packaging of influenza a virus segment 2. *Nucleic Acids Res.*, **39**, 7775–7790.
 62. Anhlan, D., Hrinčius, E.R., Scholtissek, C. and Ludwig, S. (2012) Introduction of silent mutations into the NP gene of influenza a viruses as a possible strategy for the creation of a live attenuated vaccine. *Vaccine*, **30**, 4480–4489.
 63. Bolte, H., Rosu, M.E., Hagelauer, E., Garcia-Sastre, A. and Schwemmler, M. (2019) Packaging of the influenza virus genome is governed by a plastic network of RNA- and Nucleoprotein-Mediated interactions. *J. Virol.*, **93**, e01861-18.
 64. Seshimo, E., Momose, F. and Morikawa, Y. (2021) Identification of the 5'-Terminal packaging signal of the H1N1 influenza a virus neuraminidase segment at single-nucleotide resolution. *Front Microbiol.*, **12**, 709010.
 65. Miyamoto, S., Muramoto, Y., Shindo, K., Fujita-Fujiharu, Y., Morikawa, T., Tamura, R., Gilmore, J.L., Nakano, M. and Noda, T. (2022) Contribution of RNA-RNA interactions mediated by the genome packaging signals for the selective genome packaging of influenza a virus. *J. Virol.*, **96** e01641-21 .
 66. White, M.C., Steel, J. and Lowen, A.C. (2017) Heterologous packaging signals on segment 4, but not segment 6 or segment 8, limit influenza a virus reassortment. *J. Virol.*, **91**, e00195-17.
 67. White, M.C., Tao, H., Steel, J. and Lowen, A.C. (2019) H5N8 and H7N9 packaging signals constrain HA reassortment with a seasonal H3N2 influenza a virus. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 4611–4618.
 68. Paillart, J.C., Skripkin, E., Ehresmann, B., Ehresmann, C. and Marquet, R. (1996) A loop-loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 5572–5577.
 69. Sit, T.L., Vaewhongs, A.A. and Lommel, S.A. (1998) RNA-mediated trans-activation of transcription from a viral RNA. *Science*, **281**, 829–832.
 70. Lindenbach, B.D., Sgro, J.Y. and Ahlquist, P. (2002) Long-distance base pairing in flock house virus RNA1 regulates subgenomic RNA3 synthesis and RNA2 replication. *J. Virol.*, **76**, 3905–3919.
 71. Friebe, P., Boudet, J., Simorre, J.P. and Bartenschlager, R. (2005) Kissing-loop interaction in the 3' end of the hepatitis c virus genome essential for RNA replication. *J. Virol.*, **79**, 380–392.

72. Tssetsarkin, K.A., Liu, G., Shen, K. and Pletnev, A.G. (2016) Kissing-loop interaction between 5' and 3' ends of tick-borne langkat virus genome 'bridges the gap' between mosquito- and tick-borne flaviviruses in mechanisms of viral RNA cyclization: applications for virus attenuation and vaccine development. *Nucleic Acids Res.*, **44**, 3330–3350.
73. Imperatore, J.A., Cunningham, C.L., Pellegrine, K.A., Brinson, R.G., Marino, J.P., Evansek, J.D. and Mihailescu, M.R. (2022) Highly conserved s2m element of SARS-CoV-2 dimerizes via a kissing complex and interacts with host miRNA-1307-3p. *Nucleic Acids Res.*, **50**, 1017–1032.
74. Gultyaev, A.P., Tsyganov-Bodounov, A., Spronken, M.I., van der Kooij, S., Fouchier, R.A. and Olsthoorn, R.C. (2014) RNA structural constraints in the evolution of the influenza a virus genome NP segment. *RNA Biol.*, **11**, 942–952.
75. Gultyaev, A.P., Spronken, M.I., Richard, M., Schrauwen, E.J., Olsthoorn, R.C. and Fouchier, R.A. (2016) Subtype-specific structural constraints in the evolution of influenza a virus hemagglutinin genes. *Sci. Rep.*, **6**, 38892.
76. Michalak, P., Soszynska-Jozwiak, M., Biala, E., Moss, W.N., Keszy, J., Szutkowska, B., Lenartowicz, E., Kierzek, R. and Kierzek, E. (2019) Secondary structure of the segment 5 genomic RNA of influenza a virus and its application for designing antisense oligonucleotides. *Sci. Rep.*, **9**, 3801.
77. Lenartowicz, E., Keszy, J., Ruskowska, A., Soszynska-Jozwiak, M., Michalak, P., Moss, W.N., Turner, D.H., Kierzek, R. and Kierzek, E. (2016) Self-folding of naked segment 8 genomic RNA of influenza a virus. *PLoS One*, **11**, e0148281.
78. Ruskowska, A., Lenartowicz, E., Moss, W.N., Kierzek, R. and Kierzek, E. (2016) Secondary structure model of the naked segment 7 influenza a virus genomic RNA. *Biochem. J.*, **473**, 4327–4348.
79. Soszynska-Jozwiak, M., Pszczola, M., Piasecka, J., Peterson, J.M., Moss, W.N., Taras-Goslinska, K., Kierzek, R. and Kierzek, E. (2021) Universal and strain specific structure features of segment 8 genomic RNA of influenza a virus-application of 4-thiouridine photocrosslinking. *J. Biol. Chem.*, **297**, 101245.
80. Hagey, R.J., Elazar, M., Tian, S., Pham, E.A., Kladwang, W., Ben-Avi, L., Nguyen, K., Xiong, A., Rabinovich, M., Schaffert, S. et al. (2021) Identification and targeting of a pan-genotypic influenza a virus RNA structure that mediates packaging and disease. *Biorxiv*, 2021.08.21.457170, 21 August 2021 [cited 12 August 2022], preprint: not peer reviewed, available from: <https://doi.org/10.1101/2021.08.21.457170>.
81. Dadonaite, B., Gilbertson, B., Knight, M.L., Trifkovic, S., Rockman, S., Laederach, A., Brown, L.E., Fodor, E. and Bauer, D.L.V. (2019) The structure of the influenza a virus genome. *Nat Microbiol.*, **4**, 1781–1789.
82. Fournier, E., Moules, V., Essere, B., Paillart, J.C., Sirbat, J.D., Cavalier, A., Rolland, J.P., Thomas, D., Lina, B., Isel, C. et al. (2012) Interaction network linking the human H3N2 influenza a virus genomic RNA segments. *Vaccine*, **30**, 7359–7367.
83. Essere, B., Yver, M., Gavazzi, C., Terrier, O., Isel, C., Fournier, E., Giroux, F., Textoris, J., Julien, T., Socratous, C. et al. (2013) Critical role of segment-specific packaging signals in genetic reassortment of influenza a viruses. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E3840–E3848.
84. de Wit, E., Spronken, M.I., Rimmelzwaan, G.F., Osterhaus, A.D. and Fouchier, R.A. (2006) Evidence for specific packaging of the influenza a virus genome from conditionally defective virus particles lacking a polymerase gene. *Vaccine*, **24**, 6647–6650.
85. Yamagata, Y., Muramoto, Y., Miyamoto, S., Shindo, K., Nakano, M. and Noda, T. (2019) Generation of a purely clonal defective interfering influenza virus. *Microbiol. Immunol.*, **63**, 164–171.
86. Bdeir, N., Arora, P., Gartner, S., Hoffmann, M., Reichl, U., Pohlmann, S. and Winkler, M. (2019) A system for production of defective interfering particles in the absence of infectious influenza a virus. *PLoS One*, **14**, e0212757.
87. Gavazzi, C., Isel, C., Fournier, E., Moules, V., Cavalier, A., Thomas, D., Lina, B. and Marquet, R. (2013) 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.
88. Gavazzi, C., Yver, M., Isel, C., Smyth, R.P., Rosa-Calatrava, M., Lina, B., Moules, V. and Marquet, R. (2013) A functional sequence-specific interaction between influenza a virus genomic RNA segments. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 16604–16609.
89. Cobbin, J.C., Ong, C., Verity, E., Gilbertson, B.P., Rockman, S.P. and Brown, L.E. (2014) Influenza virus PB1 and neuraminidase gene segments can cosegregate during vaccine reassortment driven by interactions in the PB1 coding region. *J. Virol.*, **88**, 8971–8980.
90. Gilbertson, B., Zheng, T., Gerber, M., Printz-Schweigert, A., Ong, C., Marquet, R., Isel, C., Rockman, S. and Brown, L. (2016) 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.
91. Le Sage, V., Kanarek, J.P., Snyder, D.J., Cooper, V.S., Lakdawala, S.S. and Lee, N. (2020) Mapping of influenza virus RNA–RNA interactions reveals a flexible network. *Cell Rep.*, **31**, 107823.
92. Desiro, D., Holzer, M., Ibrahim, B. and Marz, M. (2019) SilentMutations (SIM): a tool for analyzing long-range RNA–RNA interactions in viral genomes and structured RNAs. *Virus Res.*, **260**, 135–141.
93. Baudin, F., Bach, C., Cusack, S. and Ruigrok, R.W. (1994) 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.
94. Cimino, G.D., Gamper, H.B., Isaacs, S.T. and Hearst, J.E. (1985) Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. *Annu. Rev. Biochem.*, **54**, 1151–1193.
95. Zhang, M., Li, K., Bai, J., Velema, W.A., Yu, C., van Damme, R., Lee, W.H., Corpuz, M.L., Chen, J.F. and Lu, Z. (2021) Optimized photochemistry enables efficient analysis of dynamic RNA structures and interactomes in genetic and infectious diseases. *Nat. Commun.*, **12**, 2344.
96. Zhang, M., Hwang, I.T., Li, K., Bai, J., Chen, J.F., Weissman, T., Zou, J.Y. and Lu, Z. (2022) Classification and clustering of RNA crosslink-ligation data reveal complex structures and homodimers. *Genome Res.*, **32**, 968–985.
97. Cao, C., Cai, Z., Xiao, X., Rao, J., Chen, J., Hu, N., Yang, M., Xing, X., Wang, Y., Li, M. et al. (2021) The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat. Commun.*, **12**, 3917.
98. Sugita, Y., Noda, T., Sagara, H. and Kawaoka, Y. (2011) Ultracentrifugation deforms unfixed influenza a virions. *J. Gen. Virol.*, **92**, 2485–2493.
99. Harris, A., Cardone, G., Winkler, D.C., Heymann, J.B., Brecher, M., White, J.M. and Steven, A.C. (2006) Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 19123–19127.
100. Slater, A., Nair, N., Suétt, R., Mac Donnchadha, R., Bamford, C., Jasim, S., Livingstone, D. and Hutchinson, E. (2022) Visualising viruses. *J. Gen. Virol.*, **103**, 001730.
101. Ye, Q., Krug, R.M. and Tao, Y.J. (2006) The mechanism by which influenza a virus nucleoprotein forms oligomers and binds RNA. *Nature*, **444**, 1078–1082.
102. Klumpp, K., Ruigrok, R.W. and Baudin, F. (1997) Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. *EMBO J.*, **16**, 1248–1257.
103. Tang, Y.S., Xu, S., Chen, Y.W., Wang, J.H. and Shaw, P.C. (2021) Crystal structures of influenza nucleoprotein complexed with nucleic acid provide insights into the mechanism of RNA interaction. *Nucleic Acids Res.*, **49**, 4144–4154.
104. Moreira, E.A., Weber, A., Bolte, H., Kolesnikova, L., Giese, S., Lakdawala, S., Beer, M., Zimmer, G., Garcia-Sastre, A., Schwemmler, M. et al. (2016) A conserved influenza a virus nucleoprotein code controls specific viral genome packaging. *Nat. Commun.*, **7**, 12861.
105. Kobayashi, M., Toyoda, T., Adyshev, D.M., Azuma, Y. and Ishihama, A. (1994) Molecular dissection of influenza virus nucleoprotein: deletion mapping of the RNA binding domain. *J. Virol.*, **68**, 8433–8436.
106. Albo, C., Valencia, A. and Portela, A. (1995) Identification of an RNA binding region within the N-terminal third of the influenza a virus nucleoprotein. *J. Virol.*, **69**, 3799–3806.
107. Elton, D., Medcalf, L., Bishop, K., Harrison, D. and Digard, P. (1999) Identification of amino acid residues of influenza virus nucleoprotein essential for RNA binding. *J. Virol.*, **73**, 7357–7367.

108. Ng,A.K., Zhang,H., Tan,K., Li,Z., Liu,J.H., Chan,P.K., Li,S.M., Chan,W.Y., Au,S.W., Joachimiak,A. *et al.* (2008) Structure of the influenza virus a H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J.*, **22**, 3638–3647.
109. Tarus,B., Chevalier,C., Richard,C.A., Delmas,B., Di Primo,C. and Slama-Schwok,A. (2012) Molecular dynamics studies of the nucleoprotein of influenza a virus: role of the protein flexibility in RNA binding. *PLoS One*, **7**, e30038.
110. Li,Z., Watanabe,T., Hatta,M., Watanabe,S., Nanbo,A., Ozawa,M., Kakugawa,S., Shimojima,M., Yamada,S., Neumann,G. *et al.* (2009) Mutational analysis of conserved amino acids in the influenza a virus nucleoprotein. *J. Virol.*, **83**, 4153–4162.
111. Isel,C., Munier,S. and Naffakh,N. (2016) Experimental approaches to study genome packaging of influenza a viruses. *Viruses*, **8**, 218.

Knowledge gaps

The IAV genome segments are selectively packaged into virions as a 7+1 supramolecular complex containing one copy of each segment. This selective assembly is mediated by a concerted effort of the terminal and internal segment-specific packaging signals, as evidenced by existing literature. These packaging signals form an extensive yet flexible network of intersegmental RNA-RNA interactions. Therefore, it has been hypothesised that the selective and coordinated genome packaging in IAV is governed by the packaging signals that maintain the 7+1 supramolecular complex through intersegmental RNA-RNA interactions. A major challenge in the field is the lack of functionally validated intersegmental RNA-RNA interactions responsible for genome packaging. Since packaging signals are also strain-specific, the lack of biologically relevant intersegmental RNA-RNA interactions impedes understanding of genetic reassortment in IAV.

Post-genome replication and nuclear export, it may be speculated that vRNPs undergo local structural remodelling to facilitate intersegmental RNA-RNA interactions during progressive genome assembly into the 7+1 supramolecular complex. The NP protein also plays a key role in genome packaging. While evidence on the role of NP in IAV genome packaging is still in its infancy, its impact on vRNP structural flexibility is established. The flexible nature of vRNP allows local structural modelling of the vRNP to expose short single-stranded loops that can engage in intersegmental RNA-RNA interactions during progressive genome assembly.

Therefore, it is imperative to study the impact of RNA secondary structures to understand how they contribute to the nature and flexibility of vRNP. This knowledge can prove crucial to understanding the restraints and mediators of selective genome packaging in IAV. Moreover, studying this process could be the first step in understanding the genomic constraints of reassortment between genetically diverse strains of IAV.

RNA structural biology and chemical probing

RNA structure and its implications on virus regulation

Since the late 1970s, when the crystal structure of tRNA was solved, RNA has gained massive popularity¹⁵². Its dynamic and complex 3D nature aligns with the molecule's functional complexity. Advances in the studies of RNA structure revealed answers to ancient questions and had compelling implications in gene regulation, microbiology, and even the Origins of life¹⁵³. Thus, RNA was no longer an underrated molecule and has been studied extensively since.

Viral RNA is the prime example of the best way to maximise the efficiency of a compact yet dynamic RNA molecule. In many eukaryotic RNA viruses, the replication fitness and propagation efficiency of the virus are encoded in the small yet structurally diverse genomes¹⁵⁴. Viral RNA secondary structures mediate the storage and regulation of genetic information¹⁵⁵. In IAV, a pseudoknot formed at the 3' splice in the M segment regulates splicing¹⁵⁶. Additionally, the panhandle structure formed by the partially complementary segment termini in IAVs binds with the RdRp in a sequence-specific manner to initiate transcription¹⁵⁷.

The genetic information of RNA is organised in two tiers: first in its sequence and, second in the structure. RNA structures can be described at five levels: primary, secondary, tertiary, quarternary and quinary¹⁵⁸.

The primary structure is the RNA sequence. The linear primary structure of an RNA lies in the sequence of its nucleotides and forms the basis for the folding of higher-order structures¹⁵⁹. Each base has three edges which can hydrogen bond to other bases: the Watson–Crick–Franklin edge, the Hoogsteen edge, and the sugar edge¹⁶⁰. Besides the classical canonical base pairing, RNA has other well-categorised non-canonical interactions that allow for structural diversity¹⁶¹. One example of alternative interactions is the G-U wobble base, which has similar bonding energy to an A-U pair and can fit within a Watson–Crick helix¹⁶².

RNA molecules fold back on themselves by forming intramolecular base pairs to form secondary structures such as stems, loops, bulges, junctions and pseudoknots (**Fig 15**). The resulting structures comprise two fundamental building blocks: paired regions (mostly A-form helices), and unpaired regions such as apical loops of hairpins, internal loops and single-stranded junctions. Secondary and tertiary structures are a combination of conserved motifs. Whether motif evolution is convergent or divergent is unclear, but such motifs exist in many organisms. These RNA structural motifs can be divided into two categories. Firstly, tertiary interactions localised about a central region, such as RNA junctions; and secondly, those that involve long-range interactions (i.e., kissing loops, pseudoknot-like interactions, and loop receptors. Since RNA junctions are localised around a specific set of nearby or neighbouring

nucleotides, they direct the local intramolecular folding of RNA. Junctions also promote coaxial stacking or bending of helices. In contrast, motifs that coordinate long-range contacts, such as kissing loops and loop receptor interactions, are often attributed to intermolecular interactions with partner polymers¹⁶³. The IAV vRNA segment termini undergoes base pairing to form a dsRNA due to the partially complementary UTRs resulting in a panhandle structure at the 5' and 3' segment termini with a hairpin structure at the opposite end. With such a plethora of possible interactions and the ability to transition between different states, higher-order RNA structures are incredibly diverse and flexible. As a result, many RNAs exist as ensembles of various conformations, usually with a dominant ground state and occasionally with rare excited state structures. The composition of the surrounding environment in which the RNAs exist usually modulates the ensemble composition¹⁶⁴.

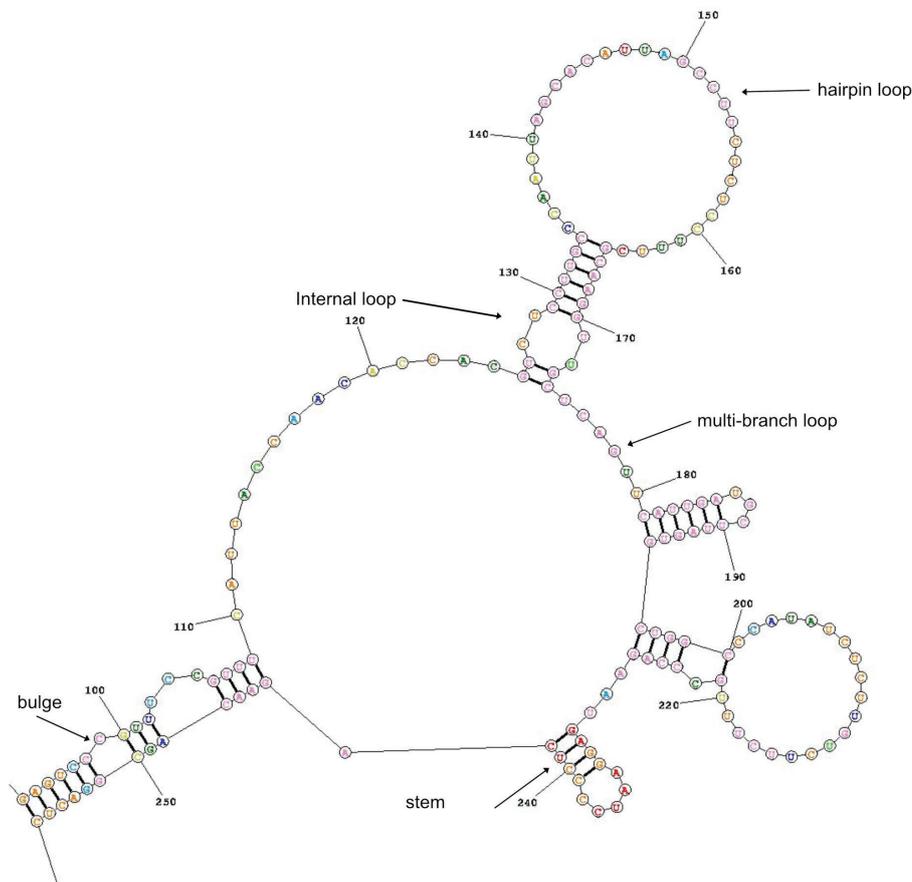


Fig 15 | RNA secondary structure of PR8_PB2 segment created by RNAstructure

Default parameters of the RNAstructure web server was used to predict a secondary structure of nucleotides 1- 250 from the 5' end of the vRNA sequence. The various RNA secondary structural elements created by the program are labelled in the figure.

Tertiary interactions compact the RNA and are typically achieved by long-range Watson–Crick and non-Watson–Crick interactions of elements within the pre-formed secondary structures. These interactions give rise to tertiary structural elements, including pseudoknots, which lock together two stem-loops by base pairing and sugar-phosphate interactions, often in a so-called kissing interaction^{165,166,158}. An interaction involving the PB1 and NS segment of an

H5N2 strain of IAV was previously suggested to involve kissing-loop interaction¹⁶⁷. Disruption of this intersegmental interaction impaired the co-packaging of other partner segments and increased the number of empty virus particles. Therefore, higher-order RNA structures involving flexible regions of IAV vRNPs are crucial to maintaining intersegmental interactions that determine genome packaging¹⁶⁷. As a result, it is important to understand the structure and flexibility of IAV vRNPs while studying segment-specific packaging signals.

RNA interaction partners such as proteins, or other RNAs recognise specific structural motifs to form quarternary structures with RNAs and can trigger refolding, cleavage, and chemical modifications upon binding¹⁶⁸. Often, the target regions must be unstructured, i.e. in loop regions, or engaged in weak structures in order to allow for interactions. The IAV NP protein oligomerises to form a scaffold for IAV vRNA resulting in right-handed helical vRNP with flexible regions. This facilitates structural heterogeneity in vRNP filaments through bending, allowing it to adopt various functionally relevant conformations throughout the different stages of the virus life cycle¹⁶⁹. Additionally, the heterotrimeric RdRp remains bound to the vRNP and acquires multiple conformations that facilitate viral RNA replication and initiation of transcription¹⁷⁰. Structured and unstructured regions, therefore, are interdependent while modulating their effector functions¹⁶⁴.

RNA is similar to DNA since they have primary structures comprising a long sequence of nucleotides that can base pair¹⁷¹. On the contrary, RNA is also analogous to proteins because they form sophisticated secondary and tertiary structures¹⁷¹. Furthermore, for both RNAs and proteins, secondary structure formation occurs mainly through hydrogen bonding (as well as base stacking for RNA). Additionally, tertiary structures are formed through interactions between secondary structure elements¹⁷¹. Given these similarities, methods to study RNA structures involve both the sequencing methods used for DNA and the biophysical and computational approaches used for characterising proteins¹⁷¹. The flexibility of an RNA moiety that allows it to adopt many conformations has made RNA structure probing technically challenging. Therefore, understanding the mechanisms of RNA-based virus regulation warrants a good knowledge of the corresponding RNA folding patterns and structures involved¹⁵⁵.

RNA structural probing strategies

Initially, RNA secondary structure probing was performed with ribonucleases. RNase T1 and RNase T2 isolated from *Aspergillus oryzae* were used to study the structure of yeast tRNAs¹⁷². RNase recognises ss/ds RNA regions and cleaves them at their recognition sites¹⁶⁸ (**Fig 16a**). The main limitations of RNases as structural probes are their size, making them sensitive to steric hindrance and preventing their use in cells or in viruses¹⁷³.

On the other hand, chemical reagents that modify specific bases according to their chemical reactivity and/or the structure in which they are found in the polyribonucleotide chain have

On the other hand, probes that are nonspecific to bases react with the RNA backbone to determine RNA structure and can, therefore, be used to study every nucleotide simultaneously irrespective of the nature of the base. For example, hydroxyl radicals use backbone cleavage to probe solvent accessibility¹⁷¹. Hydroxyl radical cleaves RNA by abstracting a hydrogen atom from the ribose moieties along the RNA backbone. As hydroxyl radical is exceedingly short-lived and reactive and attacks sites on the surface of the RNA molecule, there is almost no sequence dependence or base dependence in the cleavage reaction. Every position along the backbone is cleaved nearly equally¹⁷⁴. This chemistry probes the 3D structure of RNA as it discriminates the surface from the interior of large RNA molecules¹⁷⁵.

Following chemical modification, the RNA fragments are separated by polyacrylamide gel electrophoresis and by comparison with a sequencing ladder; thus, the positions of the cleavages in the RNA molecule were identified. An alternative method was to map enzymatic cleavage sites or chemical modification at Watson–Crick positions (kethoxal, DMS, and CMCT) with RT stops. This technique involves the extension of a radiolabeled primer with a reverse transcriptase. As the RT dissociates from the template at the modification sites, cDNAs of varying lengths are produced in different proportions. The resulting cDNA population is detected through denaturing PAGE followed either by autoradiography or phosphorimaging¹⁷³.

Novel genome-wide sequencing technologies and the availability of whole transcriptome data for various organisms demanded robust large-scale RNA structure prediction methods. The first *in vitro* approach for high-throughput transcriptome-wide RNA probing happened when NGS was used instead of gel or capillary electrophoresis for readout. Parallel analysis of RNA structure (PARS)¹⁷⁶, parallel analysis of RNA structures with temperature elevation (PARTE)¹⁷⁷, fragmentation sequencing (Frag-seq)¹⁷⁸ and ss/dsRNA-seq are experimental approaches that combine RNase treatment with NGS. The chemical inference of RNA followed by massive sequencing (CIRS-seq)¹⁷⁹, multiplexed accessibility probing-sequencing (MAP-seq)¹⁸⁰ and chemical modification-sequencing (ChemMod-seq)¹⁸¹ methods use CMCT and DMS for probing RNA structures. In contrast, hydroxyl radicals are used within the hydroxyl radical footprinting-sequencing (HRF-seq)¹⁸² method in the context of RNA tertiary structure analysis.

Selective 2'-hydroxyl acylation analysed by primer extension - mutational profiling (SHAPE-MaP)

RNA is a dynamic molecule, folding and changing its structure on several time-scales¹⁸³. However, a general chemical property of RNA is that it has a 2'-hydroxyl group in the ribose ring, and the reactivity of the 2'-hydroxyl in RNA is gated by local nucleotide flexibility¹⁸⁴. In other words, the 2'-hydroxyl is reactive at single-stranded and conformationally flexible positions but is unreactive at nucleotides constrained by base pairing. Thus, reagents that

modify the RNA 2'-hydroxyl (2'-OH) groups can read out RNA structure. The generic nature of the 2'-OH makes it an ideal candidate for RNA structure probing. SHAPE chemistry probes local nucleotide flexibility, while hydroxyl radical footprinting assesses solvent accessibility and, thus, the global RNA fold. It is also advantageous that both techniques are adaptable to various reaction conditions, including the presence of proteins. Moreover, since both techniques are insensitive to nucleotide identity, they can provide structural information at a per-nucleotide resolution¹⁸³.

In solution, 2'-OH functional groups of RNA have pKa values that range from 12–14. Nevertheless, RNA functional groups can alter their pKa values to approach biological conditions, and these changes are dependent on RNA structure. In the case of SHAPE reactivity and RNA structure, the 2'-OH pKa and reactivity is modulated by the propensity for a flexible 2'-OH to rarely sample a conformation that renders it active for reaction with a solution electrophile. The ability to score reactivities, and convert them to nucleotide flexibility is the biggest advantage of SHAPE when compared to other reagents used to probe RNA structure¹⁸³.

SHAPE is a chemical probing technique that uses electrophilic agents capable of binding to free or single-stranded 2'-OH regions in the sugar-phosphate backbone of RNA. **Figure 17** shows a schematic of the SHAPE-MaP strategy used in this study. A schematic of the SHAPE-MaP

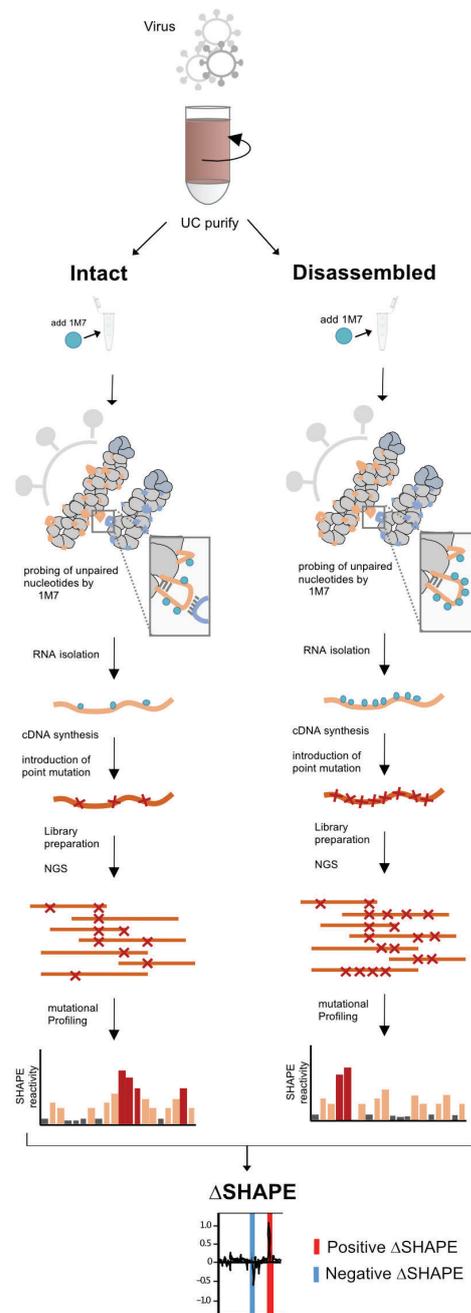


Fig 17 | Overview of the SHAPE-MaP assay

SHAPE-MaP was performed on purified viruses split into two aliquots - intact and disassembled. 1M7 was used as the SHAPE reagent. 1M7 interacts with single-stranded or unpaired nucleotides of the vRNP segments. Virus disassembly causes severance of vRNPs from its partner segments, increasing the expected number of single-stranded or unpaired nucleotides when compared to intact particles. Nucleotides potentially bound by NP or involved in intersegmental interactions are shielded from modification by 1M7 and therefore appear low reactive. Comparison of the SHAPE reactivity profiles of each segment in its intact and disassembled conditions with Δ SHAPE can potentially identify regions that are significantly different in nucleotide reactivities.

strategy adopted in this study. RNA has various nucleophilic functional groups that serve as reactive partners in acylating reactions, including the exocyclic amines on cytosine, adenine and guanine nucleobases, the 5'-OH and 3'-OH groups at the ends of the RNA strand and the 2'-OH groups¹⁷¹. Acylation reactions of acyl imidazole with single-stranded DNA and single-stranded RNA showed that acylation is highly selective towards RNA, indicating that the acylation reaction mainly occurs at the 2'-OH position of RNA¹⁷¹.

It should be noted that biologically derived RNAs are often phosphorylated at the terminal 5'-OH and 3'-OH positions, blocking these potential acylation sites. In contrast, synthetic RNAs are mostly designed without terminal phosphate groups¹⁷¹. Interestingly, the observation that DNA is poorly acylated relative to RNA suggests that the reactivity of the terminal 5'-OH and 3'-OH groups is lower than that of the 2'-OH groups¹⁸⁵.

SHAPE reagents probe flexible nucleotides, single-stranded, or unconstrained by tertiary interactions at a per nucleotide resolution. Consequently, the ability of the SHAPE reagent to bind to a nucleotide is quantitatively described by its reactivity and is, therefore, inversely proportional to the ability of a nucleotide to be involved in a secondary/tertiary interaction¹⁸⁶.

SHAPE probes self-quench by reacting with water and are characterised by hydrolysis half-life which determines the timescale of RNA dynamics probed, the type of environment they can be used in (*in vitro* vs. *in vivo*), and the meaning of the reactivity measured from these probes¹⁷¹. For RNA, the acylating reagent should ideally have high reactivity for the 2'-OH group and low reactivity towards water, which is the solvent in most RNA-acylation reactions. Highly reactive reagents react quickly with water, resulting in a rapid decline in concentrations prior to the reaction with the low-concentration RNA¹⁷¹. The reactivity of an acylating reagent can be assessed by determining its half-life in water¹⁷¹. A short half-life results in rapid hydrolysis of the reagent which can limit the desired concentrations of the probe, leading to reduced RNA acylation and eventually decreased nucleotide reactivities¹⁷¹. On the contrary, a long half-life in water usually means that the acylating reagent has low reactivity, resulting in low levels of RNA acylation, eventually making it difficult to calculate reactivity rates at positions of modified nucleotides¹⁸⁵.

The first SHAPE probe was NMIA (N-methyl-nitroisatoic anhydride). It reacts with hydroxyl groups to release CO₂ to form a 2-methylaminobenzoic acid ester. A unique property of NMIA that makes it an ideal RNA structure probe is that the product of the acylation reaction creates a handle that is inert to the reversible deacylation mechanism. The stability of the acylation adduct makes it ideal for manipulating the RNA after modification. The initial strength of SHAPE was demonstrated by characterising flexible regions in tRNA¹⁸⁷. Hydroxyl acylation by NMIA was shown to differentiate between flexible and constrained regions in RNA robustly¹⁸⁸. The first fast-reacting SHAPE reagent was 1M7 (1-Methyl-7-nitroisatoic anhydride), a para-nitro derivative of NMIA. 1M7 is significantly more labile to hydrolysis than NMIA. The half-

life of 1M7 is 14 seconds, compared to 20 minutes for NMIA. This drastic change in reactivity allows one to complete a SHAPE experiment in a few minutes¹⁸³.

SHAPE technology dramatically improved RNA secondary structure probing for the following two reasons: firstly, SHAPE reagents can acylate nucleotides independent of base identity providing a measure of the single or double-strandedness at per nucleotide resolution. Secondly, the methods used to analyse modified cDNAs have improved. The combination of SHAPE with NGS (SHAPE-seq) provides highly reproducible reactivity data over a wide range of RNA structural contexts without biases¹⁶⁸. Large-scale *de novo* identification of RNA functional motifs has improved with the SHAPE-MaP approach since chemically modified sites can be quantified in a single direct step by modifying the RNA backbone¹⁶⁸.

The Mutational Profiling (MaP) strategy uses specialised conditions that allow the RT to read through chemically modified positions. The enzyme incorporates a noncomplementary nucleotide or induces a deletion or other sequence change at the site of a chemical adduct. The locations of the SHAPE adducts are thus recorded in the resulting cDNA as mutational adducts relative to the parent RNA sequence. Although SHAPE data can also be read out by reverse transcription arrest (RT-stop) and sequencing library-ligation strategies, the MaP strategy (SHAPE-MaP) is simpler to implement, especially on long RNAs, and allows rare RNAs to be examined¹⁸⁹. In these MaP methods, the encoding of modifications as mutations rather than truncations during RT is promoted by using Mn^{2+} as the divalent cation in place of Mg^{2+} ¹⁷¹. Since misincorporation under these conditions does not halt RT, multiple RNA modifications can be detected per molecule¹⁷¹. This allows for more advanced data analysis approaches that separate sequencing reads into different groups according to mutation pattern before downstream analysis to uncover the signatures of tertiary interactions, multiple subpopulations of RNA structures in the probed ensemble¹⁹⁰, or even to detect paired bases¹⁹¹. RT-mutate methods have also been used to detect structural changes owing to single-nucleotide differences by splitting reads according to these differences before structural analysis¹⁷¹. However, an important caveat is that while using RT-mutate methods, it is important to use DNA sequencing read lengths that cover the entire RNA region of interest to obtain complete information on its structure¹⁷¹.

A key aspect of experimental design for both RT-stop and RT-mutate approaches is the choice of RT priming strategy^{171,192}. For RT-mutate methods, detection of longer regions is possible and can be expanded by current sequencing long-read length platforms. Multiple defined priming sites can be used for RNAs that exceed these length limitations, and reactivity data can be paired together within the different windows¹⁷¹. Multiple priming can also be performed using random RT primers that can bind at many positions throughout a mixed population of RNAs. This is particularly useful for transcriptome-wide studies or studying long RNAs¹⁷¹.

Finally, a no-reagent control sample is another key aspect of experimental design for both RT-stop and RT-mutate approaches. In this study, we have used DMSO as the no-reagent control samples. The RNA sample is folded in the same conditions but treated with a control solvent prior to RT in the exact same steps as the modified RNA sample. Including this sample allows data analysis to remove spurious false positive reactivities that could be due to natural RT fall-off or mutations in the specific sequence context of the RNAs being studied.

Using structural data from SHAPE-MaP to inform functional RNA elements

A common first step in analysing genome-wide SHAPE data sets involves the identification of structured regions. Highly structured regions can be good starting points for identifying functional viral RNA structural motifs. However, considering the number of highly structured elements alone tends to overpredict the number of functional elements in an RNA molecule. A major challenge, therefore, is to identify the specific RNA structures with effector functions and validate their functional relevance in the context of viral fitness. Structured regions can be identified by plotting median SHAPE reactivities over 30–75 nucleotide sliding windows and identifying regions with low median reactivities relative to the global median¹⁷¹. The median SHAPE reactivity identifying a highly structured region can be compared to the medians of well-characterised structured RNA elements. *In vitro* transcribed, full-length PB2 vRNA from strain PR8 folded in solution was probed using SHAPE reagents. SHAPE-guided modelling suggested several areas in this terminal region that contain a stable RNA secondary structure named PSL2¹⁷¹. Through mutational analysis, nucleotides in the PSL2 were previously identified as key players in packaging the PB2 segment¹⁹³. Also, PSL2 stem-loop structure was recovered in SHAPE-guided modelling of full-length PB2 RNA across diverse species and subtypes, including the highly pathogenic avian H5N1 and pandemic 1918 H1N1 strains¹⁹³.

Viral sequences can also be used to study mutation trends. If a mutation occurs on a functionally relevant paired base, there will usually be an evolutionary pressure to incur additional mutations that restore the base-pairing over time. This phenomenon is called covariation. Multiple sequence alignments are used to compute sequence covariation at a single-nucleotide resolution to identify these covaried and presumably functionally important structures¹⁹⁴. Comparative analyses of pairwise covariations in RNA sequence alignments have a successful history in consensus RNA secondary and 3D structure predictions¹⁹⁵. The lack of a significant covariation does not necessarily mean there is no conserved RNA structure¹⁹⁶. RNA structural constraints lead to the suppression of variation in the third (wobble) position of amino acid codons. Suppression of synonymous codon usage (SSCU) has been used to identify structured RNA elements in viral genomic RNAs^{197,198}. Algorithms such as R-scape support the presence of conserved RNA secondary structure elements identified through multiple sequence alignments⁹. For HIV, SIV, and poliovirus, RNA secondary structure conservation was analysed by evaluating the rate and pattern of nucleotide changes in viral

species-specific sequence alignments of protein-coding regions to yield a predicted evolutionary base-pairing probability for each nucleotide¹⁸⁹. This approach has been used extensively and proven useful in most SHAPE-based analyses of viral genomes. Another important method for identifying functional elements was developed with HCV¹⁸⁹. A preliminary covariation model from a SHAPE-directed HCV secondary structure model of a single strain coupled with a sequence-based alignment of multiple HCV genotype 2 viruses was built. Additional divergent HCV sequences were incorporated into this preliminary covariation model using algorithms developed to identify highly structured RNA motifs in shorter RNAs such as riboswitches and ribozymes. Seventeen stem loops containing multiple consecutive covarying nucleotides were identified. Four of the five novel elements studied were found important for viral fitness¹⁸⁹. In DENV and ZIKV, Huber *et al.* used information from more than 4000 and 500 sequences, respectively, to build covariation profiles to analyse regions with low synonymous mutation rates and low SHAPE reactivity i.e., highly structured regions and characterise structural similarity between the two viruses^{199,200}.

Another important development in the discovery of functional RNA motifs combines the identification of regions with low SHAPE reactivity with the identification of regions with low Shannon entropy²⁰¹. SHAPE reactivities can be used to constrain the genome-wide prediction of thermodynamically stable secondary structures based on nearest-neighbour rules. This will result in multiple possible conformations for long viral RNAs (especially >700 nucleotides). The probability of forming each base pair is then calculated across all possible structures in the ensemble structure generated²⁰². These base-pairing probabilities are used to calculate a Shannon entropy value for each nucleotide. Consequently, regions with low Shannon entropy are either likely to form a single stable structure or unlikely to base pair. Identification of regions that are both highly structured (low SHAPE reactivity) and have a well-determined structure (low Shannon entropy) in viral genomes has proved critical to the discovery of novel well-defined RNA structures that are critical for viral fitness in HIV, HCV, SINV, and DENV. SHAPE-directed structure models were combined with covariation analyses to study three lentivirus species to create structure-dependent genome-wide sequence alignment. This alignment was used to discover regions with statistically correlated SHAPE reactivity profiles and to construct consensus secondary structure models¹⁸⁹.

Another important approach for discovering functional RNA motifs in viral genomes involves comparing RNA structural elements in different biological conditions of a virus. Comparison of the DENV2 genome in the ex-virion and *in viro* states led to the discovery that flavivirus genomes are in their circularised form when packaged in virions and are in their linear form in the absence of viral proteins²⁰³. Comparing the CMV genome segment 3 structure in virion and in infected cell lysates revealed an RNA motif bound by the viral replicase in infected cells²⁰⁴. Dadonaite *et al.*, used SHAPE-MaP on ex virio, *in viro* and naked viral RNA experiments to show that different IAV segments acquire distinct RNA conformations and form intra- and intersegment RNA interactions within influenza virions¹⁷¹. The eight segments probed *in viro*

contained many highly structured regions; these regions were less structured than *in vitro* refolded RNAs, suggesting that binding of the NP partially remodels RNA structures²⁰⁵. Since such differences were observed in IAV vRNPs across different states, in this study, we decided to compare the nature of vRNPs in two alternative states, i.e., in intact virus particles (*in viro*) and after disassembly of virus particles, the details of which are described later.

Computational predictions of RNA secondary structures

Secondary structure prediction tools

Since experimental determination of RNA structures was challenging, computational predictions using sequence information were widely adopted. Computational pipelines for RNA structure prediction have improved drastically, primarily due to fundamental improvements in its underlying algorithms. Simultaneously, advances in structure probing technologies allows for both *in vivo* and *in vitro* high-throughput screening of the RNA 'structure-ome'. RNA structure probing technologies and structure predictions tools have been combined to increase the accuracy of 2D and 3D structure predictions¹⁶⁸.

RNA secondary structure can be predicted by thermodynamics-, base pairing simulations-, and AI-based approaches²⁰⁶. The most frequently used computational approaches for predicting RNA secondary structure are thermodynamics-based folding algorithms, including RNAstructure²⁰⁷ and ViennaRNA package²⁰⁸. A predicted secondary structure can guide further experiments or comparative sequence analysis and aid in designing antisense RNA molecules that could be tested as antivirals²⁰⁹.

The main advantage of these tools is that they can predict RNA secondary structures using only the sequence and does not require any experimental data. These algorithms sample every structure that can be obtained from the RNA sequence by following a set of folding rules (i.e., nucleotides allowed to pair) and search for the most probable native structure; i.e., the conformation with the minimum free energy (MFE)²¹⁰. MFE approaches give the most thermodynamically favoured conformation, which is theoretically the most common structure adopted by an RNA molecule when in equilibrium among an ensemble of potential structures¹⁷¹. To compute the free energy of an RNA secondary structure, thermodynamics-based algorithms use a set of parameters first determined by optical melting experiments. However, these methods are significantly limited by the length of an RNA sequence. They are accurate for shorter sequences, but accuracy drops drastically for sequences longer than 700 nucleotides. This limits their utility for long and complex RNAs, such as single-stranded RNA (ssRNA) viral genomes, especially for a virus like IAV, whose shortest segment (NS) is 890 nucleotides in length^{211,200}. Nevertheless, these algorithms continue to be used on viral genomes because of a lack of alternatives, however, with a locality constraint.

Algorithms based on multiple sequence analyses study viral RNA genomes for which multiple strains and phylogenomic data can be used to support structural evidence. When an RNA sequence contains many known homologs or shows strong structural conservation, information from homologous sequences can be used to build algorithms with high predictive accuracy. Some of these tools, such as RNAalifold²¹² and TurboFold²¹³, extract information from multiple sequence alignments to predict the secondary structure. The major difference

between these two approaches is that while RNAalifold uses the obtained consensus sequence to predict the structure, TurboFold individually computes all structures within the alignment. Other algorithms, such as Dynalign, Multalign and FoldalignM, generate the alignment and predict the structure simultaneously, making them more broadly applicable and computationally expensive^{214,215,216}.

Nevertheless, despite all the different methods mentioned above, the accuracy of each algorithm varies based on the state of RNA. Consequently, to date, no algorithm can use experimental data to predict the structure of every RNA with absolute accuracy¹⁷¹. This is because the computational models have inherent limitations that limit accuracy¹⁷¹. These include optimising free energy parameters to predict structures under specific *in vitro* folding conditions (temperature, salt concentration, pH) that may represent experimental probing conditions¹⁷¹. Additionally, non-canonical base-pairing interactions are not included in most models. Besides, experimental data quality influences predictive accuracy, and the relationship between data quality and prediction accuracy has been mostly assessed using simulated data¹⁷¹. Finally, accuracy assessments have been benchmarked only on a limited set of RNAs that are not representative of all biologically relevant RNAs^{171,217}. Improving these data-informed RNA structural modelling will likely evolve with the improvements in experimental techniques¹⁷¹.

However, integrating SHAPE reactivities as experimental constraints to guide RNA structure prediction has significantly improved modelling accuracy²¹⁸. Consequently, we decided to combine SHAPE reactivities with pairwise linkages of IAV vRNA long-range interaction predictions produced by Daniel Desirò, Roberto Koch Institute, Berlin, with vRNAsite package²¹⁹ to narrow down putative regions with functional implications in packaging.

RNA Structuromics: Optimising RNA secondary structure prediction with chemical probing data

Chemical probing, such as SHAPE or DMS, usually yields per-nucleotide reactivities that partially reflect a nucleotide's structural restraint. These reactivities are then used to either directly guide *in silico* RNA structure prediction methods or determine which structural conformation best fit the nature of experiments¹⁶⁸.

Preliminary attempts to guide RNA structure prediction with experimental probing data or covariation within homologous sequences were based on hard constraints. These constraints restrict the folding space of a nucleotide, e.g. through the exclusion or enforcement of specific base pairs²²⁰. However, such binary restraints are too stringent in structure predictions since experimental data may not always be error-free. For hard constraints, even small errors in the input can eventually result in wrong predictions. To overcome issues with ambiguous data that could result in erroneous RNA structure predictions, soft constraints that target the pseudo-free energy of loop motifs were used to replace hard constraints²⁰⁹. Specifically,

pseudo-free energy terms that use reactivities (r) when calculating free energies of nucleotides involved in base stacking have typically taken the form:

$$\Delta G_{SHAPE}(i) = m \times \ln [r(i) + 1] + b$$

where $\Delta G_{SHAPE}(i)$ is the pseudo-free energy term at nucleotide i , m and b are constant parameters, and $r_{(i)}$ is the reactivity at nucleotide (i) ¹⁷¹. With m positive and b negative, this term effectively penalises nucleotides within structures with high reactivity and thus favours nucleotides with low reactivities to be in structured positions²²¹. The use of chemical footprinting data as soft constraints for secondary structure prediction has been mainly driven by advances in SHAPE experiments¹⁶⁸. While SHAPE reactivities cannot be directly converted into pseudo-free energies, one can use SHAPE reactivities as likelihoods to enforce a nucleotide to be paired or unpaired. For instance, a nucleotide with a reactivity > 0.8 (highly reactive) can be forced to remain unpaired. In contrast, a nucleotide with a reactivity of < 0.4 (low reactive) would be forced to base-pair during structure prediction inputs. Subsequently, the corresponding pseudo energies are computed from these probabilities. However, converting probing data into probabilities to be paired or unpaired can be challenging sometimes since SHAPE reactivities, do not unambiguously distinguish paired from unpaired positions. For instance, when probing IAV vRNPs, since NP-bound nucleotides cannot react with 1M7, all nucleotides with reactivities < 0.4 may not imply single-strandedness. In fact, those nucleotides could either be unpaired or are simply shielded by NP from interacting with 1M7. Additionally, in many SHAPE datasets, there are often regions in the RNA of interest where nucleotides show intermediate reactivities. This also presents a challenge while using SHAPE data as a constraint to draw RNA structures¹⁶⁸.

SHAPE data incorporation into thermodynamic RNA folding models have been done through addition of defined pseudo-free energy terms, or iteratively perturbing the energy model to generate structures that better match the SHAPE data¹⁷¹. Once reactivity information is incorporated into the energy model, various computational methods can be used to predict the experimentally restrained structural state²²². Maximum expected accuracy (MEA) methods use a probabilistic model to select the most probable structure across the ensemble. Multiple structures consistent with the experimental conditions can be predicted using MFE or MEA structure/s^{168,171}.

Despite the challenges mentioned above, many methods incorporating experimental probing data have been shown to improve predictive accuracy^{171,215}. The accuracy is usually quantified through sensitivity and positive predictive value measures with the incorporation of experimental data improving accuracy from 60 % to above 85 % when experimental data is included^{171,223}. These accuracy assessments can be compared with the base pairing patterns of structural predictions to solved RNA crystal structures since the latter reveal both secondary and tertiary structure base pairing partners^{171,218}. When such crystal structures are

unavailable, experimentally restrained structural predictions can be compared with comparative sequence analysis on highly conserved RNAs to improve accuracy¹⁷¹.

Despite substantial advances in studying the *in vivo* structure of RNA molecules, various challenges impede understanding the RNA structure²²⁴. Firstly, there are inherent limitations in RNA structure modelling algorithms since they exclude complex elements such as pseudoknots because of the computational cost¹⁷¹. Therefore, combining the data they generate into a biologically relevant structure is challenging²²⁴. Existing computational methods use sliding window-based solutions for predicting RNA containing multiple pseudoknots²²⁵. Although techniques for direct RNA–RNA interaction mapping theoretically have the potential to capture pseudoknots in RNAs, they do not preserve information about the relationship between individual RNA duplexes, which makes it impossible to determine whether two non-nested duplexes coexist as part of a pseudoknot, or whether they belong to two mutually exclusive alternative conformations¹⁷¹. In this regard, combining computational approaches for RNA structure ensemble deconvolution from chemical probing experiments with RNA–RNA interaction capture data might provide the means for identifying pseudoknots at scale^{171,224}.

A thorough characterisation of RNA structure ensembles and their dynamics in living cells mandates overcoming technical limitations²²⁴. This would imply better chemical probes (and RTs) are required to achieve a higher signal-to-noise ratio in MaP-based RNA chemical probing experiments. This would facilitate ensemble deconvolution by direct read clustering, lowering the sequencing depth required to detect less abundant RNA conformations²²⁴. On the contrary, the reaction time needed for the chemical probe to permeate the cell and modify the RNA efficiently also determines the quality of a SHAPE experiment. In general, *in situ* (*in vivo*/*in viro*) probing of RNA structure ensembles remains a substantial challenge¹⁷¹. RNA structural heterogeneity might arise due to numerous cellular determinants, affecting only a small fraction of the RNA population. Currently, the resolution of methods based on direct read clustering is limited to the reconstruction of conformations with sufficiently high stoichiometries (typically 10 % or higher), and such reconstructions likely represent an aggregate of highly similar yet structurally distinct conformations, hence providing only a global overview of all possible RNA ensembles^{171,226,227}. Combining these methods with computational approaches that depend on thermodynamics might help address this limitation by enabling the further deconvolution and refinement of these sub-ensembles²²⁴.

Another important caveat is that RNA structure mapping experiments are read out on the Illumina platform, which has a maximum achievable read length of 600 bp. Understandably, this is a major limitation for analysing viral RNAs since they are longer than 600 bps²²⁴. Using long-read platforms such as Oxford Nanopore and PacBio provides an opportunity to tackle this problem^{224,228,224}. One such upgrade is the Nano-DMS-MaP, which takes advantage of improvements in DNA basecaller accuracy, leading to a higher signal-to-noise ratio without

changes in the experimental or analytical pipeline. The technique, therefore, allows more accuracy in RNA structural determination by detecting long-range interactions by correlated chemical probing and computational deconvolution of structural ensembles^{171,228}.

The ability to deconvolve RNA structure ensembles is especially relevant when considering RNA structural motifs as a target for small-molecule drugs. Alternative structures might mediate different biological functions, so it is essential to identify the right conformation responsible for a specific pathological phenotype for precise target identification¹⁷¹. Consequently, characterising precise RNA structure ensembles within living cells is a key step towards mapping the target RNA molecule. Little is known about how small molecules can establish productive interactions with RNA or the features that define a candidate drug pocket within an RNA structure element²²⁴. The binding of small molecules to RNA has been shown to shield nucleotides from chemical probing^{229,230}, alter their reactivity²³¹, allowing the precise detection of RNA–small molecule interaction sites²²⁴. Such small molecules can be used to study and validate biological functions of RNA mediated by its structure²³². Successful small molecule candidates (e.g. antisense oligonucleotides) can eventually become therapeutics^{233,234}.

Hypothesis & Objectives

Genome packaging in IAV is rigorously selective, restricting random genetic reassortment among closely related strains under coinfection conditions. Compelling evidence suggests this process is mediated by discrete segment and strain-specific packaging signals at the segment termini and internal regions. Segment-specific packaging signals interact to form and maintain a 7+1 supramolecular complex comprising one copy of each IAV vRNPs during genome assembly and packaging. Mutations or alterations in these packaging signals have shown to cause an impediment in packaging of one or more segments resulting in virus attenuation or formation of empty virus particles. Therefore, we can speculate that this 7+1 complex is maintained primarily by a network of intersegmental RNA-RNA interactions while being supported by an auxiliary set of RNA-NP and protein-protein interactions (polymerase complex, cellular proteins etc). To tackle this, we propose a multifaceted approach combining RNA secondary structure analysis with genome packaging studies.

The structure of vRNPs is known to play a role in genome packaging, and Dadonaite *et al.*, used SHAPE-MaP to show that different IAV segments adopt various conformations to form inter- and intrasegmental interactions within the 7+1 complex of a virion. They provided a map of genome segment constellation for the H1N1/WSN strain, providing evidence for segment co-segregation through intersegmental RNA-RNA interactions.

In this study, we use reverse genetics to produce two parental (H1N1/PR8 and H3N2/MO) and eight single-segment reassortants from the parental strains of IAV. Consequently, we use SHAPE-MaP on the two parental (PR8 & MO) and two candidate single-segment reassortants chosen based on their replication fitness. We aim to study the distinct conformations of the 7+1 vRNP complex in each virus under intact and disassembled conditions. Comparison of distinct vRNP conformations between intact and disassembled conditions will inform us of the potential roles of RNA structures in genome assembly and packaging. In contrast, comparing parental strains and candidate single-segment reassortants will shed light on the significance of RNA structural elements in genetic reassortment.

Therefore, the aims of this project are:

1. Produce viral reassortants from PR8 (H1N1) and MO (H3N2) strains through 7:1 reverse genetics.
2. Estimate packaging efficiencies through competition coinfection assays.
3. Identify changes in vRNA structures on intact and disassembled virus particles with SHAPE-MaP.
4. Use pairwise computational predictions of long-range interactions between potential partner segments to inform functionally relevant packaging regions.

Materials and methods

Cells, plasmids, and viruses

Cell culture

MDCK.2 cells (ATCC CRL-2936) and HEK 293T cells (ATCC CRL-3216) were propagated and maintained in DMEM (Gibco 10564011) supplemented with 10 % (v/v) FBS (Biosera FB-1090/500) and 50 u.mL⁻¹ of Penicillin and Streptomycin (Gibco 15140122) at 37°C and 5 % CO₂.

MDCK.2 infected with IAV was maintained in DMEM (Gibco 10564011) supplemented with 0.2 % (v/v) BSA (Merck A9418) and 50 U/mL of Penicillin and Streptomycin (Gibco 15140122) with 1 µg.mL⁻¹ of TPCK-trypsin (ThermoFisher 20233). This solution will be called 'infection medium' in the following sections for ease.

Plasmid amplification

In this project, we chose A/Puerto Rico/8/1934/H1N1 (PR8) and A/Moscow/10/99/H3N2 (MO) strains for practical reasons since the plasmids were already available and the viruses could be handled in a BSL-2. Moreover, unpublished preliminary results suggested the existence of incompatibility at the RNA level between some segments of these two viruses. The pHW2000 plasmids corresponding to these strains were obtained from Prof. Bruno Lina (University of Lyon, France). The sequence of the PR8 strain from Lyon used in this project is different from the NCBI reference sequences ([NC_002016](#) – [NC_002023](#)) and is provided in <https://github.com/rithupaul/vRNAsite.git>. The sequence of H3N2/MO is available here ([CY121373](#) – [CY121380](#)).

Stocks of single unit pHW2000 plasmids (**supplementary fig 1: plasmid map**) with cDNA inserts of PR8 and MO segments were transformed into Top 10 bacteria (ThermoFisher - Invitrogen C404010) & DH5α and amplified by Maxiprep (PureLink™ K210007) according to manufacturer's instructions. The amplified plasmids were confirmed by Sanger sequencing (TubeSeq NightXpress, Eurofins Genomics Europe Shared Services GmbH) with the following forward and reverse primers: 5' AGTACTGGTCGACCTCCGAAG 3' and 5' CTGATCAGCGAGCTCTAGCATTTAG 3' respectively. Stock (1 µg.µL⁻¹) and working (100 ng.µL⁻¹) concentrations of the plasmids were measured with NanoDrop™ (ND-2000) and Qubit™ (ThermoFisher-Invitrogen™ Q32854) respectively. All plasmids were stored at -20°C until further use.

In vitro transcription to produce naked viral RNA

Twenty micrograms of each single unit pUC57 plasmid with cDNA inserts of PR8 and MO segments (**supplementary Fig 1: plasmid map**) was linearised with 4 µL of restriction enzyme overnight at 37°C. This was followed by phenol-chloroform extraction and ethanol

precipitation according to standard protocols. Linearised plasmids were confirmed by a 1 % native agarose gel electrophoresis in 1X TBE buffer. In-house T7 RNA polymerase was used for *in vitro* transcription according to published protocols¹³⁹. The IVTs were purified by size exclusion chromatography using an AKTA station on a TSK G4000SW column. Briefly, the following were added to 25 µg of linearised plasmids: 4 mM of each NTPs, 1 mM of spermidine, 5 mM of DTT, 1 % triton X-100 (v/v), 160 U of RNasin (Promega, Charbonnière-les-Bains, France), 50 µg.mL⁻¹ final BSA, 0.5 µL pyrophosphatase (Roche, Mannheim Germany) in 1X T7 buffer (0.4 M Tris-HCl pH 8, 0.15 M MgCl₂, 0.5 M NaCl) to make a final volume of 300 µL. This mixture was incubated for 3 H in a water bath at 37°C. Thirty-five µL of 10X DNase I buffer (Roche, Mannheim Germany), 3.5 µL of DNase I (Roche) and 11.5 µL of Milli-Q water were added before incubation for 1H in a water bath at 37°C. The reaction was then stopped by the addition of 150 µL of EDTA 250 mM, and a phenol/chloroform extraction was performed, followed by the addition of 3 vol. ethanol and 0.1 vol. sodium acetate 3 M at pH 5.0 and precipitation overnight at -20°C. After centrifugation at 21,000 g for 30 min at 4°C, the pellets were washed twice with cold (-20°C) 70 % ethanol and dried. The RNA pellet was dissolved in 250 µL of Milli-Q water and purified on a TSK Gel G4000SW (Tosoh Bioscience) column at a flow rate of 1 mL/min in a buffer containing 0.2 M sodium acetate and 1 % ethanol. Fractions containing the RNA of interest were pooled, ethanol-precipitated, and redissolved in 100 µL Milli-Q water. The different fractions were quantified, then the RNA integrity was analysed by 8 % denaturing polyacrylamide gel electrophoresis (PAGE)²³⁵. Concentrations were estimated with Nanodrop™ (ND-2000). IVTs were stored at -20°C and used as positive controls for PCR.

Virus production through 7:1 reverse genetics

Subconfluent HEK 293T cells seeded in 6-well plates were transfected with eight pHW2000 plasmids (7 PR8 + 1 MO) at concentrations of 150 ng/plasmid using Lipofectamine™ 2000 (ThermoFisher-Invitrogen 11668019) as the transfecting reagent and Opti-MEM™ (Gibco 31985062) according to manufacturer's instructions. VLPs released into the supernatants were harvested at 24, 48 & 72 hpt and centrifuged at 3200 × g for 15 min at 4°C before storage at -80°C. Plaque assays were performed on the supernatants for VLP titre calculation. 48 hpt VLP supernatants were harvested and passaged twice in MDCK.2 cells supplemented with 1 µg.mL⁻¹ of TPCK-trypsin (ThermoFisher 20233). Supernatants from the second virus passage were titrated with plaque assays. Individual plaques were isolated in 0.2 % BSA-PBS solutions and used as inoculum to produce DI-free clonal virus stocks. For simplicity, all the rescued reassortant viruses will be referred to as R1 through R8 based on the canonical nomenclature of IAV vRNAs. For instance, R1 corresponds to the reassortant virus (PR8_PB2_{MO}) with segment 1 (PB2) derived from MO and the remaining seven segments (2 through 8) from PR8.

Virus stock production

Subconfluent MDCK.2 cells supplemented with $1 \mu\text{g.mL}^{-1}$ of TPCK-trypsin (ThermoFisher 20233) were infected with purified plaques at an MOI of 0.001 of the PR8, MO, R1 and R6. Supernatants were harvested at 48 hpi and centrifuged at $3200 \times g$ for 15 min at 4°C to remove cell debris. Indirect RNA-Seq was performed by Dr. Jonas Fuchs (University of Freiburg, Germany) according to published protocols that confirmed virus stock sequences²³⁶. Single-use aliquots of virus stocks were quantified and stored at -80°C for competition assays and SHAPE-MaP.

Growth kinetics

Subconfluent MDCK.2 cells in 12-well plates were infected with PR8, MO, R1 through R8 and PR8_HA_{tag} (described below) at a MOI of 0.001. Briefly, media was aspirated from MDCK.2 cells, washed once with pre-warmed PBS to which 500 μL of viruses at an MOI of 0.001 were added and incubated at 37°C and 5 % CO_2 . Viruses were removed, cells were washed once with pre-warmed PBS and 1 mL of media with $1 \mu\text{g.mL}^{-1}$ of TPCK-trypsin (ThermoFisher 20233) was added to each well. Supernatants were harvested at 12, 24, 36, 48, 60 and 74 hpi and quantified with plaque assays. Growth kinetics were plotted with GraphPad Prism v10.

Plaque assays for virus titre determination

Serial 10-fold dilutions of viruses were prepared in DMEM (Gibco 10564011) supplemented with 0.2 % (v/v) BSA (Merck A9418), 50 U.mL^{-1} of Penicillin and Streptomycin (Gibco 15140122) and $1 \mu\text{g.mL}^{-1}$ of TPCK-trypsin (ThermoFisher 20233). Virus dilutions were incubated with MDCK.2 cells at 37°C and 5 % CO_2 for 1 hour of virus adsorption. Virus dilutions were removed, and cells were washed with D-PBS (Gibco 14190144). An overlay medium of DMEM containing 2 % Oxoid agar (ThermoFisher-Oxoid LP0028B), 5 % NaHCO_3 (Merck 172577), 10 % BSA (Merck A9418), 1 % DEAE-Dextran (ThermoFisher 15455949), 1M HEPES (Carl Roth® 7365-45-9), and $1 \mu\text{g.mL}^{-1}$ of TPCK-trypsin (ThermoFisher 20233) was used for plaque formation. Plaques were counted three days later, and virus titres were calculated. Virus titres were plotted with GraphPad Prism v10.

Immunofluorescence assay

Subconfluent MDCK.2 cells were cultured in a 24-well plate in DMEM (Gibco 10564011) supplemented with 10 % (v/v) FBS (Biosera FB-1090/500) and 50 u.mL^{-1} of Penicillin and Streptomycin (Gibco 15140122) at 37°C and 5 % CO_2 . The cells were infected with all ten viruses (PR8, MO, R1 - R8) for 24 H. The medium was removed, and cells were washed with $\sim 300 \mu\text{L}$ of PBS. Cells were fixed with 300 μL of 4 % paraformaldehyde (PFA) for 10 min at RT. The PFA solution was removed from the cells and washed thrice with 300 μL of PBS. The cells were then permeabilised with 300 μL of PBS containing 0.5 % Triton-X for 5 min at RT. The

cells were washed thrice again with 300 μ L of PBS. 250 μ L of blocking buffer containing 5 % BSA diluted in PBS was added to cells for 5 min. Cells were incubated with 250 μ L of primary antibody (goat anti-rabbit Cy2 #68) diluted 1:1000 in blocking buffer for 60 min at RT. The primary antibody was removed, and cells were washed thrice with PBS. The secondary antibody (rabbit anti-NP #268) was added at a dilution of 1:200 (250 μ L) and incubated for 30 min at RT. The cells were washed once again with 300 μ L of PBS followed by one wash with 500 μ L of PBS containing 0.01 % of DAPI (1:10,000). The DAPI was removed with one PBS wash of the cells before viewing under the fluorescence microscope.

Competition experiments

Mutagenesis of HA_{PR8} and production of PR8_HA_{tag} virus

Site-directed mutagenesis was performed on the PR8_HA segment in the pHW2000 plasmid with the Phusion DNA polymerase (ThermoFisher Scientific F530S) using the overlap extension PCR protocol²³⁷. The forward and reverse primers used for mutagenesis were: 5' GTCATCACTCGAAAGATTTGAAATATTTCCCAAAG 3' and 5' CTTTGGGAAATATTTCAAATCTTTGAGTGATGAC 3' respectively. To 50 ng of pHW2000_PR8_HA plasmid, 1 μ L of each dNTP, 10 μ L of 5X Phusion buffer, 0.4 μ L of Phusion Pol and 35.6 μ L of milliQ water was added. The plasmid was amplified at the following conditions: 98°C for 60 sec, 98°C for 30 sec, 53°C for 50 sec, 72°C for 2 min, 72°C for 5 min followed by 4°C until use. The total number of cycles was 32. The amplicon was digested with 1 μ L of Dpn I at 37°C for 1 H. The digested amplicon (30 μ L) was transformed into 50 μ L of XL-10 competent cells and incubated on LB agar plates overnight at 37°C. The transformed colonies were picked and confirmed by Sanger sequencing (TubeSeq NightXpress, Eurofins Genomics Europe Shared Services GmbH). A synonymous mutation C425T was introduced in the pHW2000_PR8_HA plasmid destroying the BstBI restriction site to produce the pHW2000_PR8_HA_{tag} segment. Using 7:1 reverse genetics with this mutated plasmid, a modified PR8 virus (hereafter referred to as PR8_HA_{tag}) was rescued. Virus stocks of PR8_HA_{tag} were produced according to the protocol described above and plaque assays determined virus titre.

Single cycle replication kinetics

MDCK.2 cells were seeded in a 12-well plate. Virus inoculums at MOIs 1, 5 and 15 were prepared in infection medium with 1 μ g.mL⁻¹ of TPCK-trypsin (ThermoFisher 20233). Media from subconfluent MDCK.2 cells was replaced with 1mL of ice-cold PBS in each well while the plate was placed on ice for 3 min. The ice-cold PBS was replaced by 350 μ L of prepared virus inoculum and incubated 45 min on ice followed by 45 min at 37°C and 5 % CO₂. Following incubation, the virus suspensions were removed and infected cells were washed five times with 1 mL of pre-warmed D-PBS. 1mL of fresh infection medium without TPCK-trypsin was

added to each well for incubation at 37°C and 5 % CO₂. Supernatants were collected at 0, 4, 8, 12 & 24 hpi and plaque assays were performed for virus titration.

Competition co-infections

The three coinfection conditions were:

Control coinfection - PR8_HA_{tag} with PR8,

Condition A - PR8_HA_{tag} with R1 and

Condition B - PR8_HA_{tag} with R6

MDCK.2 cells were seeded in a 12-well plate at the recommended seeding volume. Subconfluent cells were coinfecting with viruses, each at an MOI of 5 according to the previously described single-cycle infection protocol. Coinfection supernatant was harvested at 12 hpi and viral titres were calculated with plaque assays.

Individual plaques from each condition were isolated and maintained in 1 mL D-PBS (Gibco 14190144) with 0.5 % BSA (Merck A9418). Subsequently, 500 µL of each isolated plaque was reinfected into subconfluent MDCK.2 cells in a 24-well plate for amplification. Following virus adsorption for 1H at 37°C and 5 % CO₂, cells were washed, and virus inoculum was replaced with infection medium containing 1 µg.mL⁻¹ of TPCK-trypsin (ThermoFisher 20233). Supernatants were harvested at 48 hpi and centrifuged at 3200 × g for 10 min before storing them at -80°C.

RNA extraction from these plaques was performed with the Direct-zol RNA MiniPrep kit (Ozyme ZR1055) according to the manufacturer's instructions, and concentrations were measured with the Qubit RNA HS kit (ThermoFisher – Invitrogen Q32852). Segment-specific PCR amplification (**Supplementary Fig 1**) was performed on 10 µL of extracted RNA with the Qiagen One-Step RT-PCR kit (Qiagen 210212) according to the manufacturer's instructions followed by 2 % native agarose gel electrophoresis. Sanger sequencing by Eurofins GATC (TubeSeq NightXpress, Eurofins Genomics Europe Shared Services GmbH) confirmed the plaque phenotype.

Chemical probing and computational biology

SHAPE-MaP

Virus production and purification

Two hundred millilitres of subconfluent MDCK.2 cells seeded in eight T175 cell culture flasks were infected with PR8, MO, R1 and R6 viruses, each at an MOI of 0.001. Supernatants were harvested at 48 hpi and centrifuged twice at 3200 × g for 10 min at 4°C to remove cell debris. 200 mL of each virus supernatant was distributed into ultracentrifugation tubes (Beckmann Coulter C13926) each containing 5 mL of 30 % sucrose (BioRad 161-0720) cushions in SHAPE buffer (0.1M KCl, 0.05M HEPES-KOH pH 8 @ RT, 5 mM MgCl₂). The viruses were purified and concentrated by ultracentrifugation (Beckmann Coulter Optima™ XE-90 Ultracentrifuge) in SW 32Ti rotors (Beckmann Coulter 369694) at 106559 × g for 120 min. Five hundred

microlitres of SHAPE buffer were added to 250 μ L of the virus pellet in each ultracentrifugation tube and incubated for 30 min at 4°C. The virus pellets containing SHAPE buffer were pooled and split equally into two aliquots before RNA modification. The viruses were kept intact in one of the aliquots, whereas in the other, they were disassembled before RNA modification.

Chemical modification of intact viruses

To the aliquot for intact samples, 250 μ L of SHAPE buffer was added and incubated for 15 min at room temperature followed by 15 min at 37°C. The intact sample was split into equal volumes for no-reagent control (DMSO) and test (1M7) modifications. Three millilitres of anhydrous DMSO (ThermoFisher-Invitrogen D12345) were used to prepare 100 mM of 1M7 (Sigma-Merck 908401). Test samples were modified with 1M7 at a final concentration of 10 mM for 75 s at 37°C. No-reagent control samples were treated with anhydrous DMSO for 75 s at 37°C.

Chemical modification of disassembled particles

To disassemble virus particles, 250 μ L of 10X disassembly buffer (*1X SHAPE buffer, 10 % (w/v) Triton X-100, 50 % (w/v) glycerol, 10 mg.mL⁻¹ lysolecithin*) was added to the samples (second aliquot) and incubated for 15 min at room temperature followed by 15 min at 37°C. Following virus disassembly, the samples were split into two equal volumes for control (DMSO) and test (1M7) modifications. Test samples were modified with 1M7 at a final concentration of 10 mM for 75 s at 37°C. No-reagent control samples were treated with the same volumes of anhydrous DMSO for 75 s at 37°C as test samples.

Mutational Profiling

One mg/mL of Proteinase K (ThermoFisher EO0491) was added to each sample and mixed thoroughly followed by addition of 3 volumes of RNA/DNA Shield (Ozyme ZR1100-250). Samples were thoroughly mixed and RNA was extracted with the Zymo RCC-5 (Zymo R1019) according to manufacturer's instructions. Reverse transcription was performed in specific buffer conditions with SuperScript II. We used a mix of 99X of random nonamer primers (S1254S by NEB) and 1X of a specific nonamer primer (from IDT) complementary to the 3' extremity of IAV RNAs. RNA-cDNA hybrids were purified with Micro BioSpin P-6 Gel Columns (7326222 by BioRad). Second DNA strand synthesis was performed with the NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module (E611L by NEB). Purification of dsDNA was done with NucleoSpin Gel and PCR Clean-Up kit (740609.250 by Macherey-Nagel) followed by fragmentation of dsDNA with a sonicator (Covaris E220 – Focused Ultrasonicator). cDNA libraries were prepared and dual indexed with the Microplex kit C05010007 by Diagenode. Double-size selection of the libraries was with SPRISelect beads (Beckman Coulter) followed by high-throughput sequencing on a HiSeq 4000 (2x100 bp). The samples were sequenced by the GenomEast platform, a member of the 'France Génomique' consortium (ANR-10-INBS-0009) at the IGBMC, Strasbourg, France.

Shapemapper2

ShapeMapper automates the calculation of RNA chemical probing reactivities from SHAPE-MaP experiments. The software only runs on 64-bit Linux systems and is available at <https://github.com/Weeks-UNC/shapemapper2>.

Reads are scanned from left to right, and if a window of `--window-to-trim` nucleotides is found with less than `--min-qual-to-trim` average phred score, that window and any nucleotides to the right are removed from the read. It is removed if the remaining read is shorter than `--min-length-to-trim`. Paired reads were provided as input (with `--folder` or with `--R1 <r1.fastq> --R2 <r2.fastq>`). Read pairs are merged using Bbmerge v37.78 with the `vstrict=t` option. Unmerged reads are not discarded. Merged reads were aligned to reference sequences with Bowtie2 v2.3.4.3. Since we used random primers, the software also performed random primer trimming using `--random-primer-len`, which was 9 due to the nonamer primers mentioned before. Mutations overlapping $(--random-primer-len + 1)$ nucleotides on the 3' end of reads were discarded, and $(--random-primer-len + 1)$ nucleotides were also excluded from contributing to effective read depth. `--min-seq-depth` Minimum read depth required to include a mutation at a given position was set at 2000. At each position in a given RNA, mutation rate for each provided sample is calculated as the mutation count divided by the effective read depth at that position. Lowercase sequence is excluded from reactivity profiles. Positions with effective read depth in any sample below `--min-depth 2000` were excluded. If any untreated control sample with untreated mutation rate above maximum background `--max-bg` (default=0.05) was also excluded.

Mutation rate standard errors are calculated as $\frac{\sqrt{\text{mutation rate}}}{\sqrt{\text{effective depth}}}$. Since, we used modified and untreated DMSO samples, mutation rates were calculated as $rate_{\text{modified}} - rate_{\text{untreated}}$. Reactivity standard errors were calculated as $\sqrt{(\text{stderr}_{\text{modified}})^2 + (\text{stderr}_{\text{untreated}})^2}$. By default, ShapeMapper normalises all profiles together by using the combined set of reactivities to compute a single normalisation factor that is applied to all profiles. This was disabled by passing the `--indiv-norm` option.

Output plots named `*_profiles.pdf` include panels for read depths and mutation rates. Reactivity standard errors are shown as error bars in `*_profile.pdf` plots. The mean \pm SD reactivity rates of replicates with the highest correlation coefficients were used to create average reactivity profiles of each segment with Shapemapper2. Figures were designed with RNAVigate²³⁸ by Ms. Béatrice Chane Woong-Min.

Median SHAPE reactivities of a given segment were calculated over 50-nucleotide windows and plotted relative to the global median with MS Office Excel 2016 from average reactivity profiles. These graphs were aesthetically enhanced with Affinity Designer (©Pantone 2019).

Δ SHAPE

Δ SHAPE presents a statistically rigorous approach to analyse changes in nucleotide reactivities of samples between two conditions. It is available at <https://github.com/Weeks-UNC/deltaSHAPE>. We used Δ SHAPE to compare nucleotide reactivities between the intact and disassembled conditions. It compares reactivity differences relative to the associated errors, and also compares the magnitude of each difference relative to every other change in reactivity. Nucleotides that exhibit strong, meaningful changes in reactivity are thus identified. Δ SHAPE is run on Python 2.7 with two additional modules matplotlib and NumPy.

The default parameters for Δ SHAPE analysis are set such that SHAPE reactivities must differ by at least 1.96 SD and Δ SHAPE values must be at least 1 SD away from the mean Δ SHAPE value; at least 3 nucleotides meeting these criteria are required to occur in a 5-nucleotide window in order to be highlighted by the program²³⁹. To run deltaSHAPE.py with these default parameters, one must enter the following command:

```
python deltaSHAPE.py file1.map file2.map
```

where 'file1.map' and 'file2.map' are the names of .map files corresponding to each experiment being compared. deltaSHAPE.py automatically generates a file named 'differences.txt', which is a tab-delimited file providing the position, sequence, and statistical data for nucleotides identified as showing significant changes between experiments²³⁹.

vRNAsite

The workflow of vRNAsite can be separated into six main steps, indexing, folding, averaging, extracting, trimming and plotting. The code, supplement and documentation of vRNAsite are open-source and can be accessed at <https://github.com/desiro/vRNAsite>²¹⁹. The required input is a multiple FASTA file representing the different segments of IAV, preceded by the --fasta command parameter. The parameters --reverse and --complement build the reverse complement of the input sequence. The parameter --reversePositions reverses the output from the IRI found on the negative-sense strand to provide the positive-sense strand nt positions.

To extract IRIs from the contact maps, vRNAsite utilises the start and end nucleotide positions of the potential IRI from both input RNA sequences in the following, referred to as *outer boundaries* of a contact zone. Contact zones are the areas which contribute to a peak in average free energy on the IRI contact map. Two or more contact zones can overlap if multiple possible IRIs share the same nucleotide positions on the two interacting RNA segments. vRNAsite uses a watershed segmentation algorithm to detect and separate contact zones

from each other²¹⁹. The interacting sequence parts were extracted and folded with the RNACofold algorithm of the ViennaRNA Package²⁰⁸.

Results & Discussion

In vitro 7:1 reverse genetics

Production of PR8, MO and eight single-segment reassortants (R1 – R8)

In this study, we used the gold-standard 7:1 reverse genetics system to rescue IAV VLPs²⁴⁰. **Fig 18** shows the titres of viruses produced from 24H, 48H, and 72H rescue supernatants in HEK 293T cells.

Although MO, R2 (PR8_PB1_{MO}) and R5 (PR8_NP_{MO}) did not form plaques on MDCK.2 cells, IFA with anti-NP staining confirmed cellular infection (**Fig 19**).

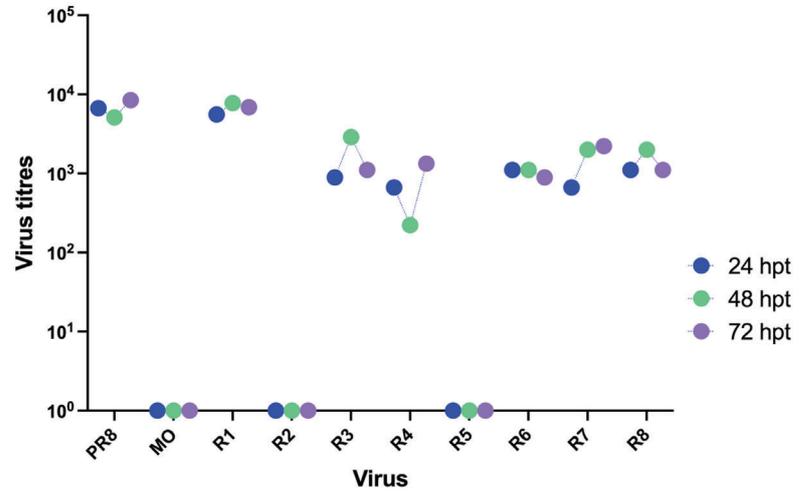


Fig 18 | Titres of virus supernatants following form 7:1 reverse genetics in HEK 293T cells

Titres of VLPs from transfection supernatants at 24, 48 and 74 hours are shown. MO, R2 and R5 did not form plaques. IFA was performed to check for cellular infection.

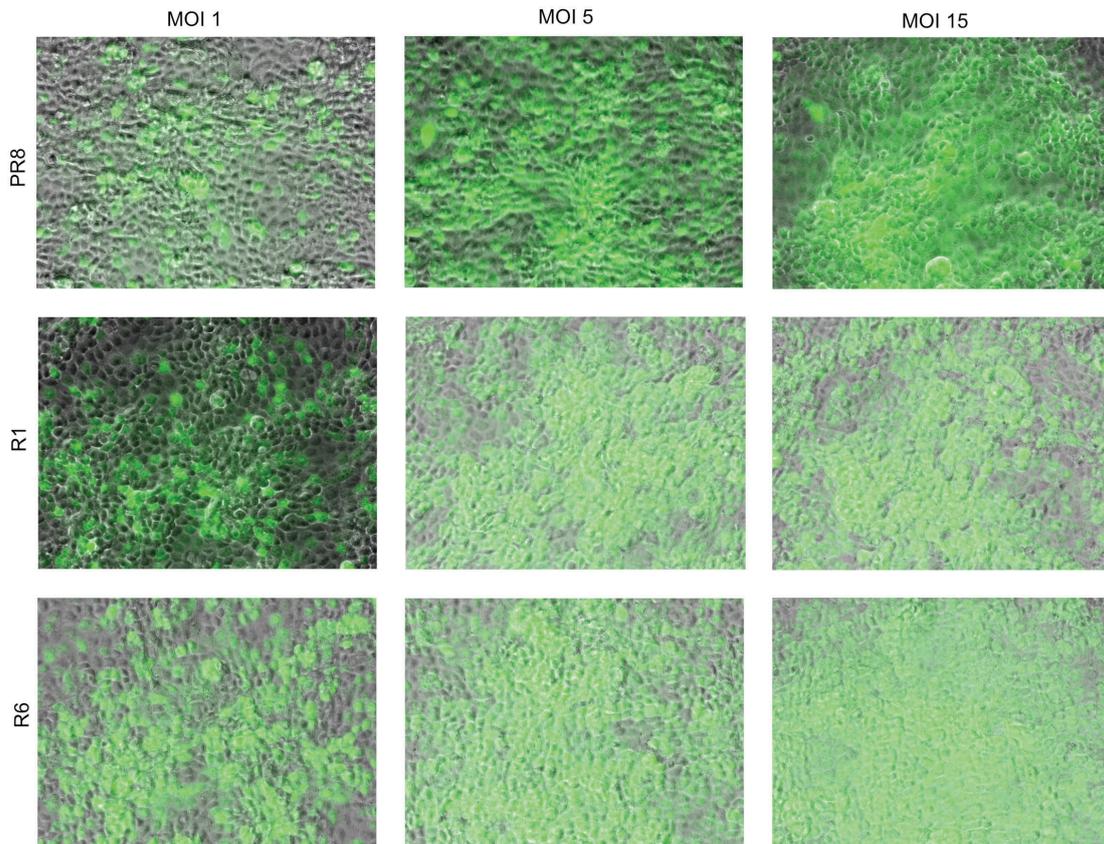


Fig 19 | Immunofluorescence staining of MDCK.2 infected with PR8, R1 and R6

IFA with anti-NP staining of MDCK.2 infected with PR8, R1 and R6 at MOIs 1, 5 and 15 are shown respectively. MOI 5 was chosen for competition coinfection assays since > 80 % of all cells were infected with each virus.

Hence, it was speculated that the rescue supernatants from HEK 293T were too low in titres and required passage/s in MDCK.2 cells for virus amplification before titration. As a result, all ten rescue supernatants were passaged twice in MDCK.2. MO, R2 (PR8_PB1_{MO}) and R5 (PR8_NP_{MO}) formed plaques of varying sizes following the two virus passages and their respective titres are shown in **Fig 20**. Although HEK 293T cells exhibit high transfection efficiencies and produce VLPs since many cells are readily transfected, MDCK cells demonstrate high *in vitro* replication efficiency for influenza viruses. As a result, for downstream experiments that require virus titres greater than 10^3 , it is better to passage the viruses in MDCK^{241,242}. An alternative approach could be using HEK293T and MDCK.2 co-cultures for simultaneous transfection and amplification of VLPs, thereby reducing the chances of cell culture-induced compensatory mutations that promote virus adaptation²⁴³.

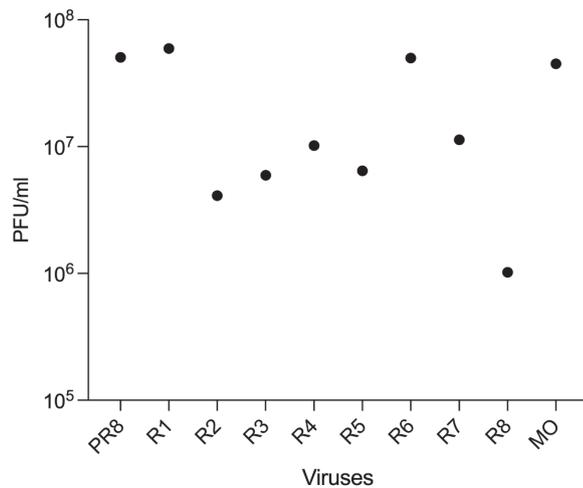


Fig 20 | Virus titres of PR8, MO and the eight reassortants

Titres of PR8, MO and R1 - R8 following two virus passages in MDCK.2 are shown. Plaque purified viruses were chosen for subsequent growth kinetics.

The successful replication of IAV depends on the expression of four different and crucial proteins - PB2, PB1, PA, and NP. We successfully rescued all eight single-segment reassortants in the first attempt. Unlike Hatta *et al.*, wherein despite being able to rescue parental Mem/88 (H3N2) and Mal/NY (H2N2) viruses from plasmids, they were only able to rescue four single-segment reassortant viruses possessing the PB2, NP, M, or NS gene of Mem/88 virus in a Mal/NY background²⁴⁴. While there were no segment mismatches in this study, as evidenced by the rescue of all single-segment reassortants, we wanted to check for any protein mismatches that suboptimal physical or functional interactions between cognate and non-cognate proteins may have introduced. Such protein mismatches become evident only when the progeny virions spread to infect and replicate within new cells. Therefore, after the rescue, we studied the growth patterns in MDCK.2 and compared the fitness of the progeny reassortants with the parental strains - PR8 and MO. The following section describes and discusses the fitness of each chimeric virus in greater depth.

Two candidate single-segment reassortants R1 (PR8_PB2_{MO}) and R6 (PR8_NA_{MO}) were chosen for SHAPE-MaP

The single-segment reassortants were cultured in MDCK.2 along with PR8 and MO, and their growth kinetics are shown in **figs 21a & 21b**. As expected, the parental viruses PR8 and MO showed the highest titres. The single-segment reassortants exhibited differential growth

capacities, with some replicating more efficiently (R1 & R6) than others (R3 & R5). This could be due to varying degrees of physical interactions between the cognate and non-cognate proteins that consecutively impair their function and attenuate growth. The two most common protein mismatches are HA-NA imbalance and incompatibilities among the polymerase subunits that prevent the formation and reduce the activity of the heterotrimeric complex.

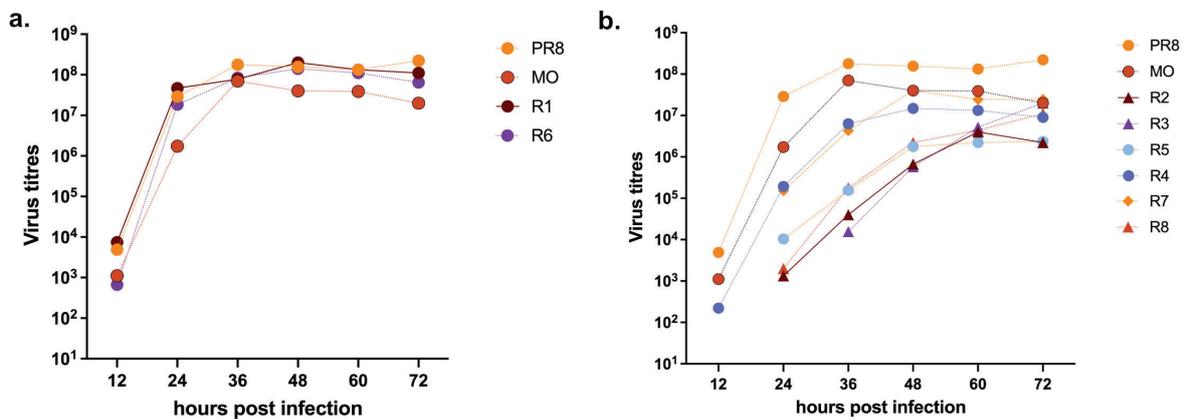


Fig 21 | Growth kinetics of PR8, MO and the eight reassortants (R1 - R8)

a. Growth curves of candidate reassortants R1 and R6 are shown with respect to parental viruses PR8 and MO. R1 and R6 were used for subsequent coinfection and SHAPE-MaP due to their similarity in growth pattern compared to PR8 and MO. **b.** The remaining six reassortants R2-R5, R7 and R8 were attenuated compared to parental PR8 and MO viruses. These reassortants showed differential growth patterns due to suboptimal compatibilities between the cognate and non-cognate proteins derived from the MO/H3N2 strain.

In this study, the parental viruses were closely followed by two chimeras R1 (PR8_PB2_{MO}) and R6 (PR8_NA_{MO}). Rescue and growth of R1 and R6 with high titres implied that incorporating PB2_{MO} and NA_{MO} segments in the PR8 background could successfully produce viruses with no vRNP or protein incompatibilities. Concurrent findings were observed by Lowen *et al.*, where PB2 and NA segments from another H3N2 strain (A/Panama/2007/99) were preferentially selected in an (A/Netherlands/602/2009) H1N1 background⁵. However, it would be interesting to study if the reverse, i.e., an H1N1 PB2 & NA segment, when introduced in an H3N2 background, would i) be preferentially selected and ii) produce such fit viruses. If these viruses could be rescued successfully, it would imply that the compatibility of packaging signals between PB2 and NA segments of PR8 and MO was reciprocated, increasing the reassortment potential of H1N1 and H3N2 strains involving these segments. However, if they could not be rescued, it would suggest that the reassortment of PB2 and NA segments of PR8 (H1N1) is restricted in a MO (H3N2) background due to incompatibilities in packaging signals⁵. On the contrary, if they could be rescued yet only produced attenuated progeny following reinfection, it would suggest incompatibilities of PR8 PB2 and NA in a milieu of MO proteins⁵. Either way, we know that PB2 and NA segments of the MO (H3N2) are not restricted to reassort in a PR8 (H1N1) background and it remains to be investigated if the reverse holds true.

Contrary to other segments, HA and NA are more prone to reassortment. The ‘Asian flu’ of 1957 was caused by a novel reassortant strain H2N2 with avian-like HA, NA and PB1 genes and the remaining segments from a human IAV²⁴⁵. The NA of H2N2 reassorted with the HA and PB1 of an avian strain to generate the H3N2 of Hong Kong Flu. The acquisition of novel surface antigens is beneficial for virus evolution and hence positively selected when natural coinfections with compatible vRNPs occur in mixing vessels such as swine. Therefore, it is unsurprising that reassortants with non-cognate HA/NA segments may be typically more fit than reassortants with other segments. However, there is evidence of differential HA and NA reassortment rates based on the degree of divergence in their packaging signals. White *et al.* reported that the reassortment of NA segments with heterologous packaging signals was more flexible and favoured, unlike the reassortment of HA segments with heterologous packaging signals²⁴⁶.

Since R1 and R6 exhibited optimal segment and protein compatibilities, they were chosen as candidate reassortants, in addition to PR8 and MO, for all subsequent chemical probing and genome packaging studies.

Single cycle replication kinetics showed R1 had slightly better *in vitro* viral fitness

Before coinfection assays, a single-cycle replication assay at multiple MOIs was performed to ensure sufficiently high infection levels per cell. Since at a higher MOI, the number of viruses per cell is high, the replication efficiency can be high. MOI 5 was chosen for coinfection assays, since the viral titres were not significantly different between the MOI 5 and 15. This would allow better control over the accumulation of DIPs, which is often a risk with *in vitro* IAV cultures. The intrinsic ability of DIPs to interfere and outcompete the number of standard infectious IAV particles during high MOI infections could skew reassortment frequencies and bias the outcome²⁴⁷. **Figures 22 a - c** show single-cycle growth kinetics of PR8, R1, and R6 at MOIs 1, 5 and 15, respectively. All coinfections described in the next section were performed with two viruses infecting MDCK.2 each at an MOI of 5.

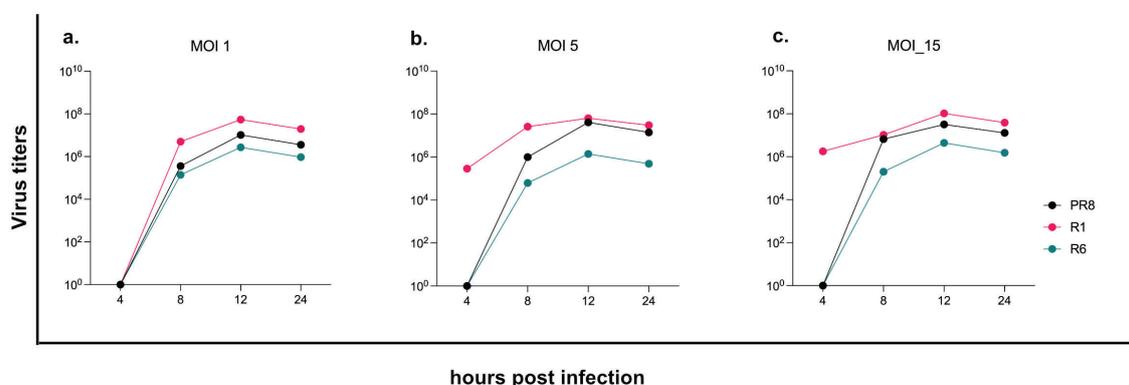


Fig 22 | Single-cycle infection of PR8, R1 and R6.

Titers from virus supernatants harvested at 12 hours post infections in infection media without TPCK-trypsin are shown. The MOI 5 was chosen for all subsequent competition coinfection conditions.

R1 showed better replication efficiency than the parental PR8 strain by having 10^5 PFU.mL⁻¹ at 4 hpi (**Figs 22 b and c**). This could be attributed to an enhanced polymerase activity conferred by the distinct non-cognate polymerase complex comprising MO_PB2, PR8_PB1 and PR8_PA subunits. An increased polymerase efficiency enables faster replication and transcription rates of vRNPs, eventually resulting in higher production of infectious virions²⁴⁸. To confirm this, the efficiency of the polymerase complexes can be compared with the mini replicon assay²⁴⁹. Phipps and team observed similar findings in reassortants that contained a PB2 subunit from A/Panama/2007/99 (H3N2) and the PB1 and PA proteins of a variant of A/Netherlands/602/2009 (H1N1). They showed through a minireplicon assay that the PB2 protein from the H3N2 virus had higher luciferase activity when combined with other subunits derived from the H1N1 virus⁵. The activity of an H1N1 PB2 has been documented to be restricted by glutamic acid at position 627, which, when replaced with lysine, dramatically enhances enhanced polymerase activity²⁴⁸. This difference in amino acids between the H1N1 and H3N2 PB2 subunits was attributed to the enhanced polymerase activity. Consequently, the H3N2 PB2 subunit was preferentially incorporated in an H1N1 background. On the contrary, a 7:1 reassortant with PB2 from A/WSN/33 inserted into the PR8 background was attenuated despite 97 % homology between the PR8 and WSN PB2 proteins²⁵⁰. Therefore, to ensure efficient multimerisation of the heterotrimeric polymerase complex, it was concluded that the individual subunits underwent co-evolution.

Protein incompatibilities may have caused attenuation of the remaining six reassortants

The remaining progeny single-segment reassortants R2 – R5, R7 and R8, were attenuated to varying degrees (**Fig 21 b**). This indicates negative epistasis from introducing a non-cognate MO segment in the PR8 background, resulting in suboptimal protein compatibility²⁷. In parallel, the introduction of the MO segment also destroys some positive epistasis between the remaining PR8 segments. To check if viral fitness is influenced by genetic diversity of non-cognate segment in the reassortant, a pairwise sequence alignment was performed on segments of PR8 and MO. The attenuated progeny reassortants have been discussed below in their descending order of initial virus titres.

R4 (PR8_HA_{MO}) - The HA protein is responsible for virus attachment and entry. The HA segment is the most common segment to reassort. Essere et.al., showed that they could rescue a reassortant with the HA segment of A/Finch/England/2051/91 (HA_{EN}) in a MO (H3N2) background with the 7:1 reverse genetics systems. However, under competitive settings where HA_{EN} and HA_{MO} plasmids were co-transfected in 293T cells, they found that HA_{MO} segment with the homologous packaging signals was preferentially incorporated over HA_{EN} segment with heterologous packaging signals. Conversely, introducing HA_{MO} packaging signals in an otherwise HA_{EN} segment increased the incorporation of the resulting HA_{EN/MO} chimeric segment, signifying that heterologous packaging signals limit reassortment²⁷.

R7 (PR8_M_{MO}) - It was hypothesised that M1 and/or M2 proteins can alter HA/NA balance by affecting one or both glycoproteins. The M segment was also found highly important to the transmissibility of the A/NL/602/09 (NL602) pandemic isolate. The matrix proteins of PR8 and NL602 have 13 amino acid differences (95 % identity), while the M2 proteins of the strains differ by 14 amino acids (86 % identity). The M-NA functional interaction appears to have a prominent role in transmitting this strain. It may be a highly important means through which influenza viruses restore HA/NA balance following reassortment or transfer to new host environments²⁵¹. It was confirmed that the M and NA gene segments of the NL602 virus contribute to transmissibility, and a cognate HA segment was found to be required for optimal transmissibility²⁵¹.

R5 (PR8_NP_{MO}) – Reduced levels of NP have shown to drastically reduce genome replication in IAV while also increasing defective interfering particle formation with segments 1 – 3⁶⁹. This suggests that unusual NP levels within viruses impair genome replication and induces preferential production of DIPs. Since we did not perform a total particle analysis with HA titres on any of these reassortants to measure the total virus particles/infectious virion ratios²³⁷, we are unable to establish if this occurred. However, it would be interesting to look at this phenomenon.

R3 (PR8_PA_{MO}) - R3 was the least efficient reassortant with poor titres and a slow growth rate. Potential protein incompatibility in the reassorted polymerase with the MO_{PA} segment hinders the formation of an efficient polymerase complex that decreases virus replication. We hypothesised that the PB2, PB1 segments of PR8 and PA of MO interacted to form a dysfunctional polymerase complex with impaired transcriptional and translational abilities. A similar phenomenon was observed in reassortants when the PA segment of H3N2 strain A/NT/60/1968 was introduced in an H1N1 and H5N1 background²⁴. The presence of 184N/383N sequences unique to human H3N2 strains was attributed to progeny virus attenuation. Analyse MO PA for these residues. Previous studies have attributed impaired formation of a functional vRNP complex as a major constraint to successful reassortment. This is often caused by the incompatibility of the interacting polymerase subunits. Restoration of the PB2-PA interaction with compensatory mutations in either segments have led to renewed virus titers²⁵. Attenuated R3 is the evidence of segment mismatch that imposes a constraint on genetic reassortment thus restricting the virus heterogeneity. PA is the primary genetic determinant of the high reassortment exhibited by GFHK99 virus in MDCK cells and defines a need for cooperation between coinfecting viruses. To verify the contribution of the PA segment by a second approach, the authors evaluated the impact of VAR virus coinfection on WT vRNA synthesis for the MaMN99:GFHK99 PA strain. The results confirm that the GFHK99 PA segment increases reliance on multiple infection, specifically in mammalian cells²⁵².

Summary and limitations of the 7:1 reverse genetics system

The advantage of the 7:1 reverse genetics system is, however, an artificial coercion of preferred genotype combinations. These are usually not confounded by factors such as host cellular restriction factors, natural infection conditions, etc. Studying the molecular determinants of reassortment restriction using reverse genetics does not represent restrictions during co-infection because such studies do not consider the important role of competition among homologous segments³¹.

Mismatches between specific packaging signals can also reduce virus viability. Given the relatively large size and likely discontinuity of the packaging signals, such incompatibility is unlikely to exclude reassortment totally. However, even a modest reduction in the fitness of a virus that is viable in the laboratory is likely to reduce successful transmission in the wild and thus potentially make the viral reproductive ratio less than one ($R_0 < 1$), preventing large outbreaks¹⁹⁷.

Activity tests of the RNP complexes composed of different proteins from the two viruses (Yokohama/H3N2 and Prague/H7N7) in a minigenome assay showed that all RNP combinations found in the recovered reassortants had higher activity than that of the parental Prague RNP, while RNP combinations not found in the reassortants had 40 % of the activity of parental Prague RNP²⁵. These results suggested that the formation of a functional RNP complex is a prerequisite for reassortment between the two influenza A viruses tested in this study under the conditions used²⁵. An alternative approach to producing reassortants may be checking for protein compatibility between cognate and non-cognate proteins PA, PB2, PB1 and NP proteins. Protein combinations that exhibit optimal activity, like parental strains, could then be used to produce reassortant viruses by reverse genetics of segments and checked for fitness in cells. For instance, the co-incorporation of PB2 and PA from the same virus appears important for efficient virus replication, especially in the reassortment between human H3N2 strains and other influenza strains²⁴. Interestingly, the 1968 H3N2 (Hong Kong flu) strain simultaneously incorporated PB2 and PA genes from the 1957 H2N2 (Asian flu) strain, which was originally derived from the 1918 H1N1 strain (Spanish flu). Similarly, natural-reassortant H3N2 viruses isolated from swine possess PB2 and PA genes from the same virus. Simultaneous incorporation of PB2 and PA was also found in the 2009 H1N1 strain. These observations imply the importance of functional cooperation between PB2 and PA in generating viable reassortant viruses²⁴.

Notably, the four RNP combinations containing Prague PB2 and Yokohama PA were either nonfunctional or significantly impaired in their activity. Thus, the authors have identified an interdependence between PB2 and PA in which Prague PB2 cannot cooperate well with Yokohama PA²⁵. Fodor *et al.* proposed that PB1 and PA are transported into the nucleus as a dimer and that PB2 enters the nucleus as a monomer, where it assembles with the PB1-PA dimer. Considering that there are no direct physical interactions between PB2 and PA, the

data suggested that the dimer formation between Yokohama PA and Prague or Yokohama PB1 triggers a conformational change unsuitable to accommodate Prague PB2²⁵.

As evidenced from our data and the discussion above, it is obvious that genetic exchange between representative strains of H1N1 and seasonal H3N2 strains yields a diverse range of reassortant but attenuated genotypes relative to the parental strains. Lowen *et al.* propose that the observed fitness defects are likely to be highly significant on an epidemiological scale but also could be overcome through reduced competition from parental strains and/or mutation of mismatched gene segments⁵. This occurs since natural infection would impose selection pressures which would cause weak reassortants to be negatively selected and eventually eliminated from circulation. We observe this common phenomenon with IAV strains that once circulated prominently in the human population but were eventually eliminated. In fact, the H3N2/MO is one such good example that once circulated but no longer exists in the human population.

Coinfection with two competing viruses

To evaluate the extent to which IAV relies on multiple infection for productive infection, we initially used reassortment and coinfection as readouts. Coinfection is a prerequisite to reassortment, and higher frequencies of reassortment are guaranteed when coinfections occur at high MOI³⁰. However, under routine experimental conditions, a small proportion of cells remain infected only by either one of the viruses. Since singly-infected cells yield only parental progeny, as a result, the progeny virions produced by single-infected cells would also contribute in numbers to the reassortant virion population and become indistinguishable from true reassortants arising from coinfecting cells (**Supplementary fig 2a**).

Introduction of a mutational tag in the HA segment of PR8

To circumvent this problem, we adopted the strategy used by Lowen *et al.* and introduced a 'tag' (synonymous mutation C425T in the 5' to 3' cDNA HA_{PR8} insert of pHW2000 plasmid)³⁰. Since PR8_HA_{tag} has only a synonymous point mutation in its HA segment, it is expected to be genotypically distinct yet phenotypically identical to PR8. However, when PR8_HA_{tag} replaces PR8 in coinfecting cells, two additional virus genotypes, which can only be produced by reassortment in coinfecting cells are produced (**Supplementary fig 2b**). Segment-specific PCR in RNA extracted from individual plaques was used to identify these reassortant genotypes. This, in turn, would allow for distinguishing virus genotypes that originate from coinfecting cells versus single-infected cells and thus help identify true reassortment rates and packaging efficiencies of the competing genome segments.

The mutated plasmid was amplified, and the presence of the tag was confirmed by Sanger sequencing. The tagged plasmid was used instead of wild-type (HA_{PR8}) in a 7:1 reverse genetics system to rescue a new virus hereafter referred to as PR8_HA_{tag} with titres of 6.8E+07 PFU/mL. Next generation sequencing confirmed the PR8_HA_{tag} virus and showed no unintentional substitutions.

Primer optimisation for segment-specific amplification

Segment-specific primers were optimised on *in vitro* transcripts (**Supplementary Fig 1**).

To ascertain that this approach was not impeded by an intrinsic packaging advantage/disadvantage of the modified HA_{tag} segment, a control coinfection condition where cells were coinfecting with PR8 and PR8_HA_{tag} was performed. The test coinfections named B and C involved PR8_HA_{tag} with R1 and R6 respectively. 127 plaques were isolated, amplified and tested with segment-specific PCRs (**Supplementary Fig 1**). Results obtained from each coinfection are discussed below.

Table 1: Plaques obtained from competition coinfection assays

Name	Condition	# of plaques	Median RNA concentrations (ng.μL ⁻¹)
A	PR8_HA _{tag} and PR8	44	9.1
B	PR8_HA _{tag} and R1	34	10.9
C	PR8_HA _{tag} and R6	35	10.7

Control coinfection A. PR8_HA_{tag} outperformed PR8

Potential differences in replicative ability between the two test viruses could result in unequal numbers of gene segments of the two viruses, which would obscure our results. We thus performed coinfections under single-cycle conditions with homologous viruses that differ only by a silent mutation in the HA segment.

Since the RNA interactome of PR8_HA_{tag} was expected to be identical to that of PR8, an equal distribution of plaques containing PR8 and PR8_HA_{tag} was expected. Contrary to this, 22 out of 44 plaques (50 %) had the HA_{tag} segment and only 4 plaques (9 %) had the HA_{wt} segment (**Fig 23**).

Additionally, 36 % of all plaques (N = 16) had both HA_{wt} segment and HA_{tag} segment (**Fig 23**). RNA from these 16 plaques was re-extracted and re-sequenced confirming both HA_{wt} and HA_{tag} within the same plaque. It was intriguing that HA_{tag} segment either outcompeted HA_{wt} and or co-packaged itself with its wild-type counterpart in 38 % of the plaques.

Although it was unexpected to find any changes in the packaging efficiency of the HA_{tag} segment, we cannot rule out any potential effects of the mutation on genome packaging. Alternatively, the HA_{tag} segment may provide an intrinsic replication advantage due to RNA structural differences or may have been added in excess over the HA_{wt} segment during virus inoculation leading to an increased incorporation efficiency. A growth curve of PR8_HA_{tag} to assess and compare its fitness is underway.

All plaque phenotypes from this condition ranged in diameters of 2 – 3 mm and no differences were observed between plaque sizes of PR8 and PR8_HA_{tag}. Since, differing plaque phenotypes are indicative of diverse virus genotypes²⁵³, one plausible explanation for the presence of both HA_{wt} and HA_{tag} within the same plaque could be that some of the isolated plaques originated from fusion of two or more microplaques formed separately by both viruses. Another scenario could be aggregates of PR8 and PR8_HA_{tag} resulting in the formation of a single plaque which was subsequently isolated and amplified giving rise to clonal isolates of both virus genotypes. A potential solution could be serial dilution of one amplified plaque to reduce its virus titres down to one plaque/well followed by segment-specific PCR.

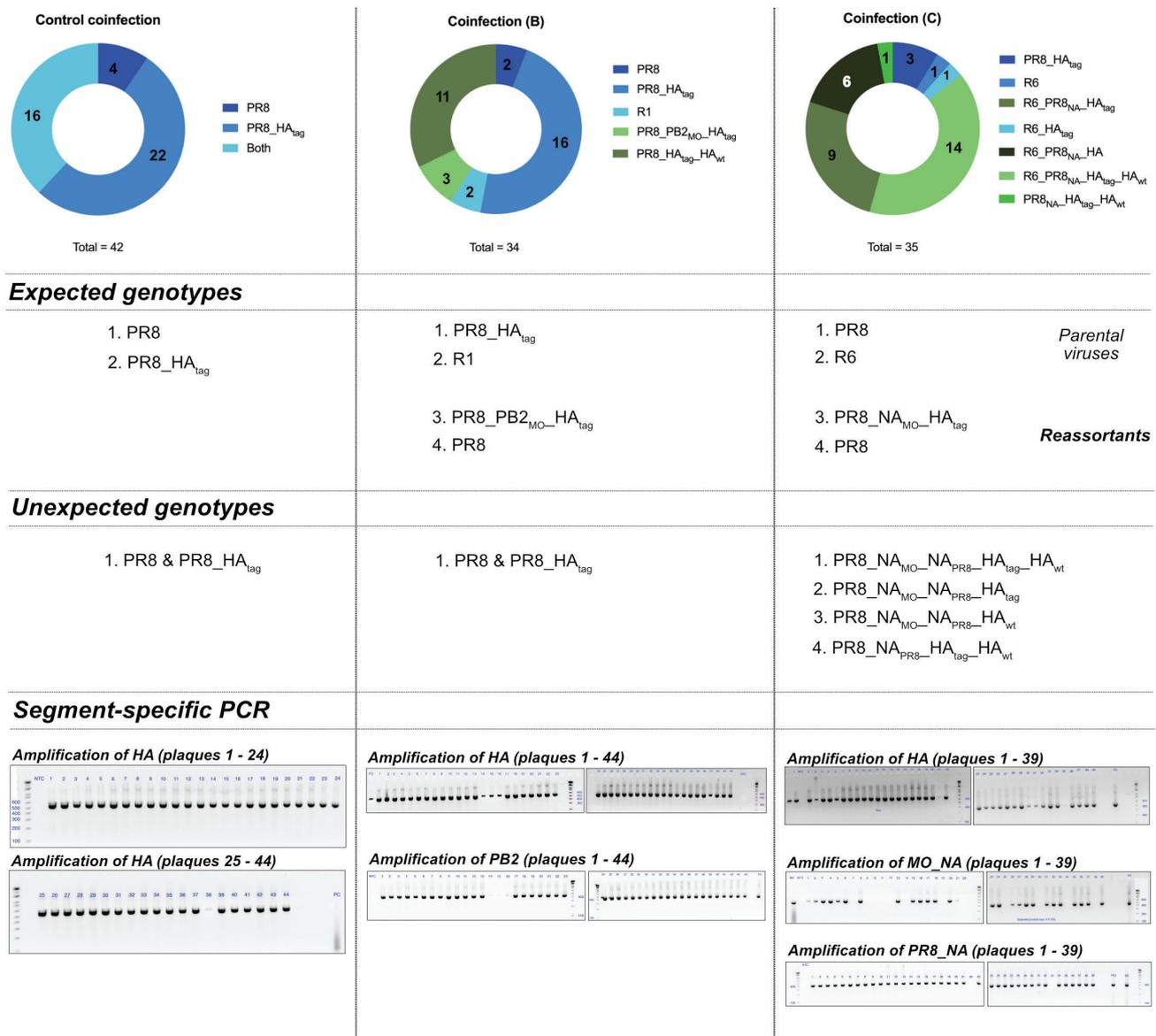


Figure 23 | Competitive coinfection assays

Coinfection conditions. PR8, PR8_HA_{tag}, R1 and R6 were coinfecting each at an MOI of 5 in MDCK.2 cell lines.

Control coinfection was homologous coinfection involving PR8 and PR8_HA_{tag}. 44 plaques were isolated. The number of plaques with the genotype PR8_HA_{tag} were higher than PR8. 11 plaques had both PR8 and PR8_HA_{tag}.

Coinfection B involved PR8_HA_{tag} and R1. Like the control coinfection, PR8_HA_{tag} outperformed PR8 in this condition. However, there was a packaging preference for the PB2 segment of PR8 than MO.

Coinfection C involved PR8_HA_{tag} and R6. Surprisingly, this coinfection setting generated the most number of reassortants with some plaques having NA_{PR8}, NA_{MO}, HA_{tag} and HA_{wt}. Intriguingly, the parental strain PR8 was not recovered from any plaques. Isolating a higher number of plaques and their subsequent genotyping could fix this issue. With the high reassortment frequency in the quadruple reassortant PR8_NA_{MO}_NA_{PR8}_HA_{tag}_HA_{wt}, it seems like there could be more than 8 segments in this reassortant. This may have been caused by an HA-NA imbalance owing to the introduction of the coinfection of R6 with PR8_HA_{tag} impairing the virion release following budding resulting in multiple particle aggregation within the same plaque.

However, if this trend recurs in conditions B and C, it would be compelling to consider the possibility of a novel and unexpected reassortant virion with more than eight vRNPs; specifically, a reassortant with two HA segments. While this phenomenon has been previously reported in viruses produced by reverse genetics²⁵⁴, the implications of the nine-segmented flu virus and its impact on HA-NA receptor balance are avenues that require consideration³⁴.

Specific residues of the influenza A virus hemagglutinin viral RNA are important for efficient packaging into budding virions. To further define nucleotides of the HA coding sequence important for vRNA packaging, synonymous mutations were introduced into the full-length HA cDNA of influenza A/WSN/33 and A/Puerto Rico/8/34 viruses, and mutant viruses were rescued. Covariation studies suggested the presence of local structured domains along the entirety of the HA segment, which contribute to the virus fitness but are not essential for the virus individually²⁵⁵. Similar findings were observed by Gog *et al.*, confirming the ability to identify functionally important RNA conservation and highlighting the surprising finding that single nucleotide changes can dramatically affect segment packaging¹⁹⁷.

Coinfection B. Heterologous coinfection of MDCK.2 with PR8_HA_{tag} and R1

The expected reassortants from this coinfection were PR8_PB2_{MO}_HA_{tag} (N = 3) and PR8 (N = 2). Noticeably, the numbers of progeny reassortant genotypes are low compared to the parental viruses namely PR8_HA_{tag} (N = 16) and R1 (N = 2) (**Fig 23**).

Additionally, like condition A, 32 % of the plaques (N = 11) had both PR8 and PR8_HA_{tag}. This recurring pattern affirmed an unexpected yet improved packaging efficiency of the HA_{tag} segment (**Fig 23**).

Only 5 out of 34 plaques had a PB2_{MO} segment and the remaining plaques had PB2_{PR8}. It is likely that PR8_{PB2} is preferred over PB2_{MO} in an H1N1 background and is positively selected during packaging (**Fig 23**). Lowen *et al.* also reported this finding⁵, where although PB2_{MO} can be efficiently packaged in a 7:1 system when its H1N1 counterpart is unavailable, it produces an attenuated yet efficient virus. However, the H1N1 PB2 is preferably packaged over H3N2 PB2 when available⁵.

An alternative scenario is that the infection input levels of the R1 virus were rather low compared to the PR8_HA_{tag} virus. Under these conditions, only few cells might be co-infected with both viruses while many cells would be infected only with the PR8_HA_{tag} virus, leading to an overproduction of the latter genotype from single-infected cells. However, in coinfecting cells where reassortment happens, the PB2_{MO} segment would be as efficiently packaged as the PB2_{PR8} segment, which is reflected by a balanced distribution of the reassortant genotypes PR8 and PR8_PB2_{MO}_HA_{tag}. Thus, to quantify the intrinsic packaging efficiencies of the PB2 segments, co-transfection assays with PR8 and MO pHW2000 plasmids under competitive

settings (7:2), which ensure similar input levels of the competing PB2 segments, would be highly valuable and are underway.

In this experimental setup, it has to be noted that, we have only counted virus plaques and not measured the ratios of the vRNAs competing for packaging. We could speculate that if there could be multi-particle aggregation of the PR8_HA_{tag} because this virus was more sticky, the distribution of infectious virions would not be same in the two coinfection conditions. This would imply that at the presumed MOI of 5 for PR8 and PR8_HA_{tag}, there could have been an unintentional introduction of more viruses of PR8_HA_{tag}, increasing its odds for cellular infection compared to PR8.

Another alternate scenario which may have led to these ambiguous findings could be a fusion of microplaques containing both PR8 and PR8_HA_{tag}. Since these plaque phenotypes looked identical at 60 H after plaque assay when they were picked, it is possible that each plaque contained aggregates of microplaques of PR8 and PR8_HA_{tag}. Consequently, the PCR amplification and sequencing strategy we adopted may have to be complemented with other techniques, such as quantifying vRNA levels in each plaque through PCR²⁵⁶.

Coinfection C. Heterologous coinfection of MDCK.2 with PR8_HA_{tag} and R6

Of the 35 plaques, the expected reassortants were PR8_NA_{MO}_HA_{tag} (N = 1) and PR8 (N = 0). Unlike the previous coinfection conditions where PR8_HA_{tag} was significantly higher than PR8, the ratio of PR8_HA_{tag} : PR8 was lower in this condition (3:1).

However, there were a higher number of additional virus genotypes in this coinfection than in the previous ones. There were 9 plaques with PR8_NA_{MO}_NA_{PR8}_HA_{tag} and 6 plaques with PR8_NA_{MO}_NA_{PR8}_HA_{wt} indicating the presence of NA segments derived from both parental viruses within the same plaque. Alternatively, a neuraminidase inhibition assay would also establish if there were two NA segments from PR8 and MO within the same plaque. It would however still be a challenge to estimate if this observation was due to the presence of viral aggregates²⁵⁷ of one or more genotypes within the same plaque. The potential viral aggregates would have to be disrupted with bacterial NA and confirmed for individual genotypes before analysing the segment composition from each genotype.

We also observed plaques with what seemed like more than eight segmented viruses (**Fig 23**). There are two possibilities for this phenomenon. The second most common type of protein mismatch frequently detected following IAV reassortment is an imbalance between HA avidity and NA activity. Since HA mediates attachment to host cells via interactions with sialic acid, and NA cleaves sialic acid to facilitate release of virions at the end of the viral life cycle, the functions of these two proteins are interdependent²⁵⁸. As a result, reassortant genotypes that combine noncognate HA and NA genes often yield sticky viruses that do not bind efficiently to target cells or do not spread efficiently due to aggregation at the cell surface²⁶.

It is likely that the introduction of the NA_MO protein in a virus with PR8_HA protein may have tampered with the HA-NA balance in the reassortants produced. This may have tipped the reassortants with the NA_MO protein to impair the release of viruses. As a result, the progeny virions produced ended up stuck to each other. Eventually, this was reflected when each isolated plaque contained more than one virus stuck together (multiple-particle aggregation). Since we used only NA-segment-specific primer without a quantitative PCR, we would never know the ratio of segments present in a plaque owing to one or more virus. A preprint published recently highlights this issue where they profiled reassortants between naturally occurring LPAs only to find that two-to-three particle aggregation contributed to genome mixing in 75-99 % of the reassortants profiled²⁵⁷.

SHAPE-MaP

Chemical probing on intact and disassembled particles

In viro SHAPE-MaP with 1M7 was performed on purified intact and disassembled samples of PR8, MO, R1 and R6 infected in MDCK.2 cells. Four biological replicates were performed, and viral titres of samples were measured before and after ultracentrifugation steps to ensure there was no significant loss of virus titres (**Fig 24**).

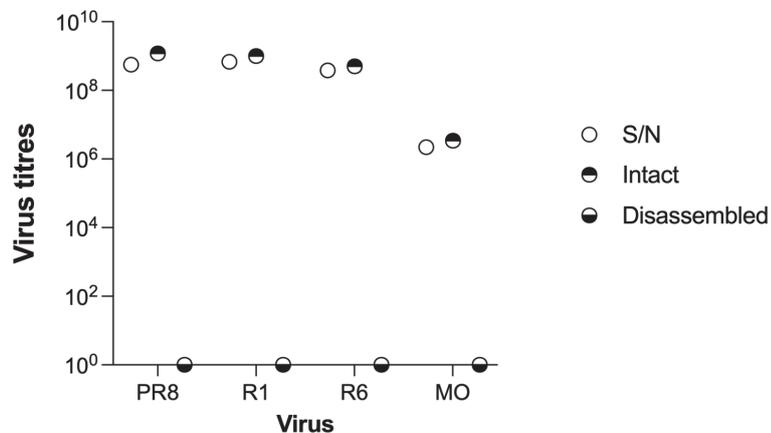


Fig 24 | Virus titres during SHAPE-MaP protocol.

S/N indicates titres obtained after harvesting supernatant. The remaining virus titres shown were measured after ultracentrifugation and treatment of samples with disassembly buffer. Since, there were no live viruses in the disassembled sample, total disassembly was confirmed.

This technique on intact virus particles probes RNA secondary structures of the 7+1 complex in its native state following virion budding and release. On the contrary, virion disassembly causes severance of vRNPs from its partner segment/s in the 7+1 complex. Hypothetically, chemical probing on disassembled particles followed by the comparison of the SHAPE reactivity profile of each disassembled vRNP with its corresponding intact reactivity profiles can be used to distinguish putative regions of intersegmental RNA-RNA interactions from intrasegmental interactions. Such regions were expected to exhibit a lower SHAPE reactivity in the intact particles when compared to their disassembled counterparts. This strategy of probing packaged vRNPs reduces the potential ambiguity of transient RNA interactions that occur in cells during vRNP assembly with cellular traffickers and or other viral proteins that may disappear post-virion budding and release. Besides, since there is very little knowledge on the nature of the hierarchical or progressive assembly of vRNPs within cells, SHAPE reactivities of vRNPs within cells would be more complicated to interpret and may bias the assumed RNA secondary structure.

The disassembled samples showed no plaques ensuring functional disassembly of the virion following treatment with the disassembly buffer²⁵⁹. Due to the low virus yields of MO (<10⁹ PFU.mL⁻¹), RNA and cDNA library concentrations were also consistently low. Higher titres of MO virus could possibly be achieved by increasing the TPCK-trypsin concentrations or using a higher MOI for infection (while avoiding DI/semi-infectious particle formation) may guarantee sufficient RNA yields for good quality SHAPE data. However, under our conditions, the data quality of the MO virus was poor and failed to produce good reactivity profiles. Subsequent

analyses were performed only on PR8, R1 and R6 SHAPE-MaP data. The following section summarises the findings from SHAPE-MaP on intact and disassembled particles of PR8, R1 and R6.

Shapemapper2 produced reactivity profiles for each segment

SHAPE-MaP gives the quantitative measurement of RNA structural information at a per nucleotide resolution for each segment of the sampled viruses. This measurement is reported as reactivity that describes how reactive each RNA nucleotide is to a given probe. Since probes tend to react with unpaired positions, high reactivities generally correlate with unstructured regions of the RNA, while low reactivities correlate with structured regions. The following quality-control checks are automatically implemented in ShapeMapper2²⁶⁰: **(1)** read-depth check, at least 80 % of nucleotides meet a minimum sequencing depth of 2000 in all samples; **(2)** positive mutation rates above background check, at least 50 % of good-depth nucleotides have a higher mutation rate in the SHAPE-modified sample than in the untreated sample; **(3)** high background mutation rates check, no more than 5 % of good-depth nucleotides have an untreated mutation rate above 0.05 (an unusually high number of high-background nucleotides can indicate the presence of native modifications, sequence variants, or instrument run failure); and **(4)** number of highly reactive nucleotides check, at least 8 % of good-depth nucleotides have a modified mutation rate above 0.006 after background subtraction. Failure to pass these checks warrants close user scrutiny²⁶⁰.

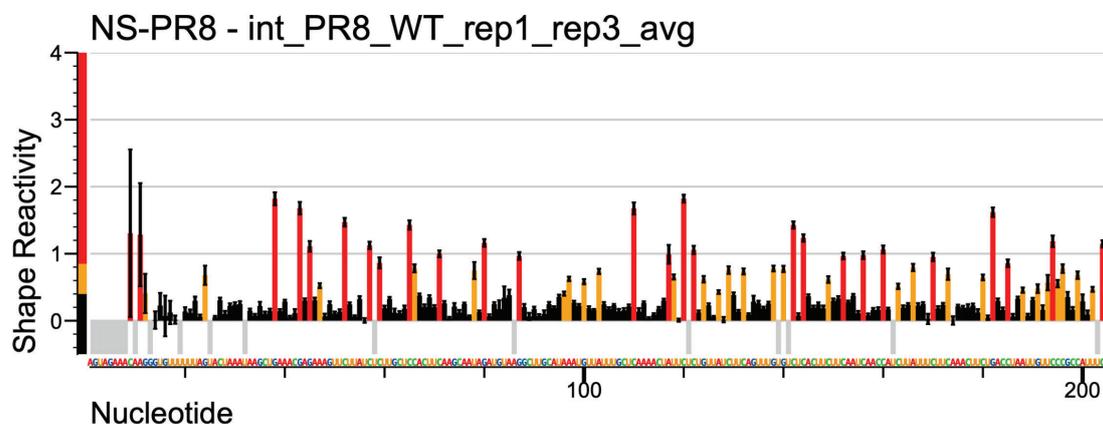


Fig 25 | Reactivity profile produced by Shapemapper2.

Average reactivity profile (replicates 1 & 3) of the NS segment of intact PR8 virus are shown here. The red bars indicate regions of high reactivity, orange - intermediate reactivities and black - regions with low reactivity to 1M7. Grey bars indicates nucleotides with no reactivity data.

Shapemapper2 calculated reactivity rates at a per nucleotide resolution for each sample (**Fig 25**). Correlation coefficients of all replicates were calculated for each sample condition. Replicates with the highest correlation coefficients were used to plot average reactivity profiles and were graphically represented as circos plots (**Figs 26 – 28**).

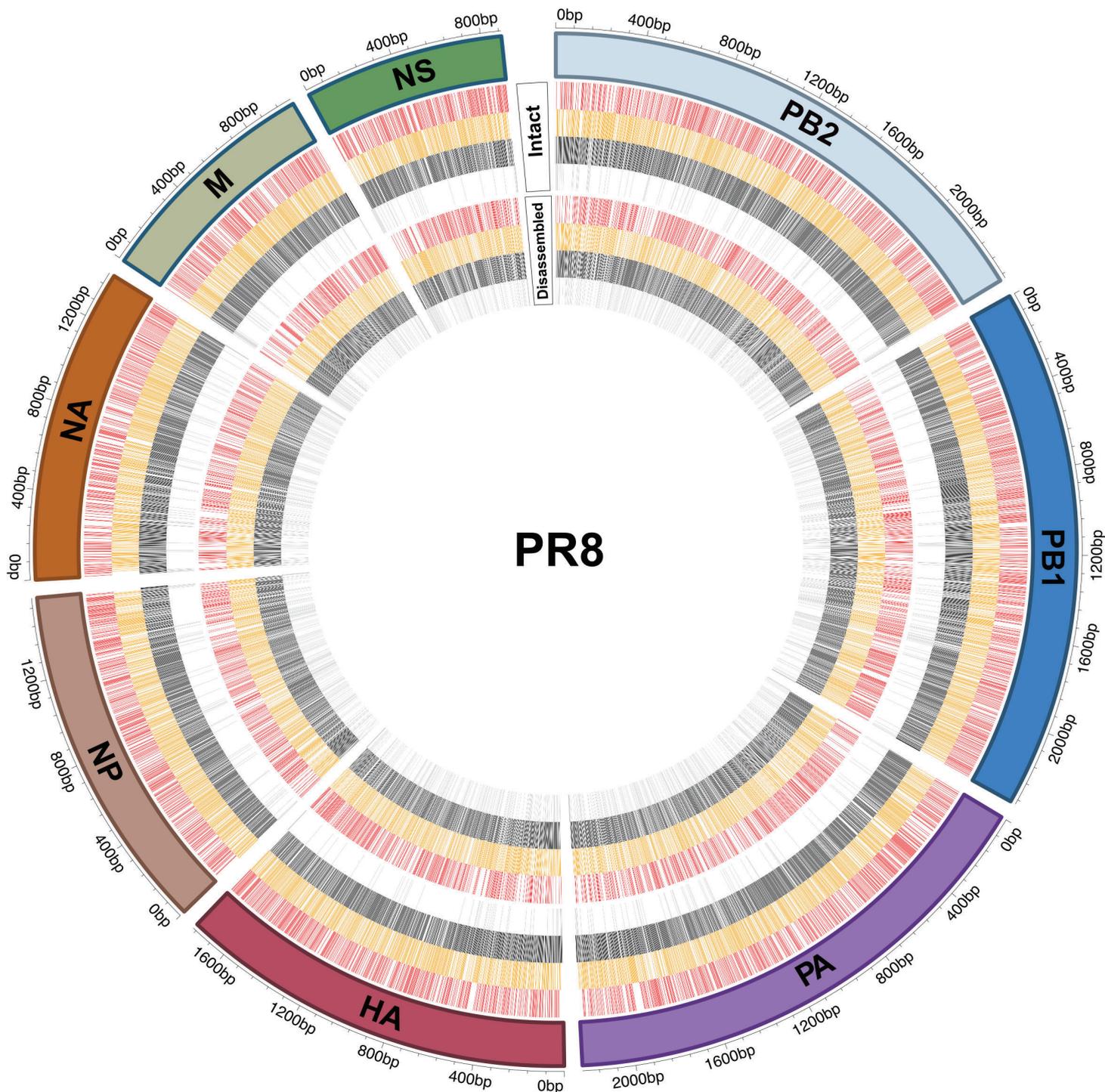


Fig 26 | Circos plot showing SHAPe-MaP reactivities of all segments of PR8/H1N1 virus

The circos plot represents average SHAPe reactivities of every nucleotide of each segment of PR8/H1N1 virus. The red bars represent highly reactive nucleotides, yellow represent nucleotides with intermediate reactivities, black represent non-reactive nucleotides. The first panel (closest to the segment) represent SHAPe reactivities of intact virus while the second panel represent SHAPe reactivities of disassembled particles. The grey bars indicate nucleotide positions with no data.

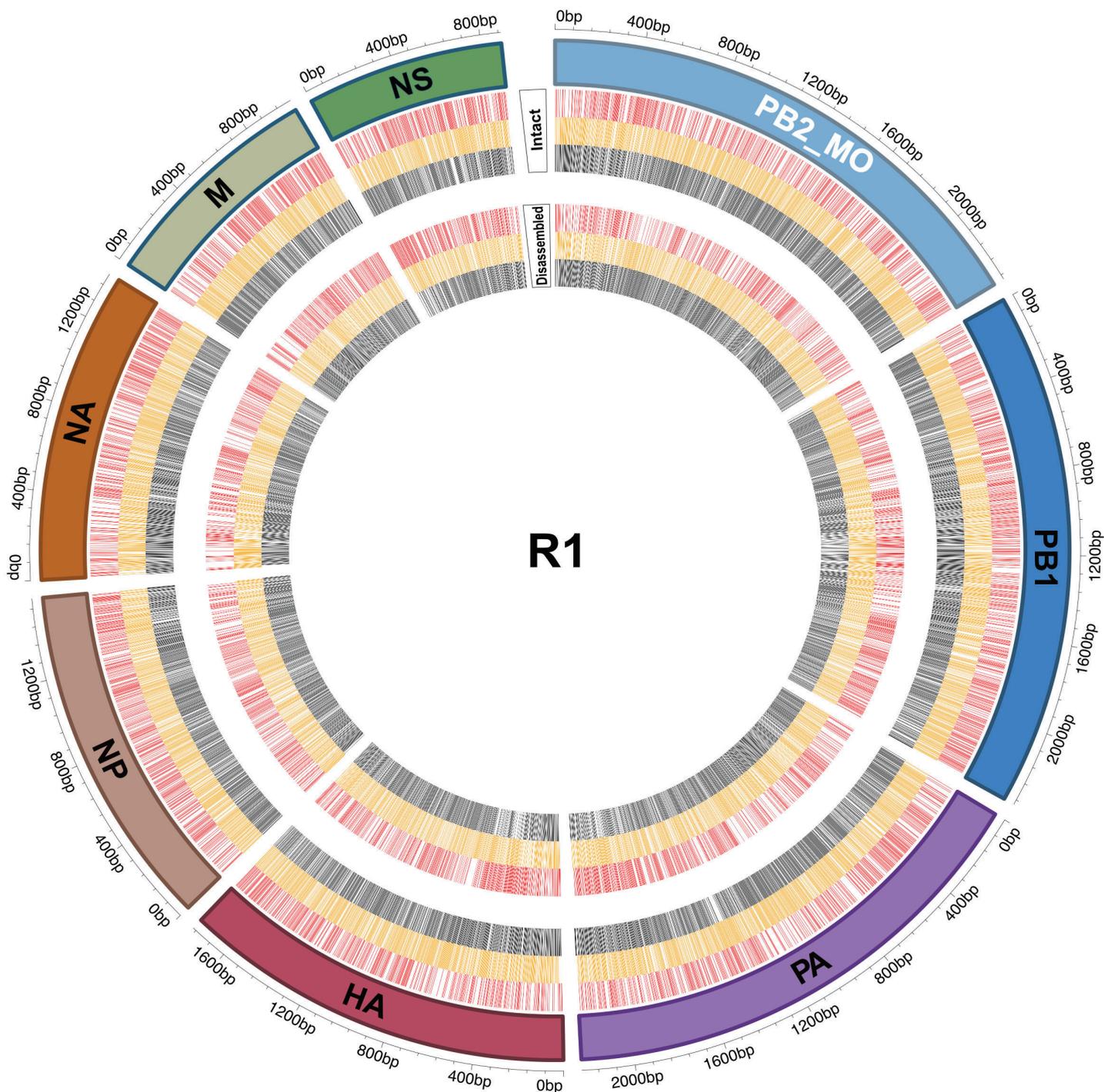


Fig 27 | Circos plot showing SHAPe-MaP reactivities of all segments of R1 virus

The circos plot represents average SHAPE reactivities of every nucleotide of each segment of R1 (PR8_PB2_{MO}) virus. The red bars represent highly reactive nucleotides, yellow represent nucleotides with intermediate reactivities, black represent non-reactive nucleotides. The first panel (closest to the segment) represent SHAPE reactivities of intact virus while the second panel represent SHAPE reactivities of disassembled particles. The grey bars indicate nucleotide positions with no data.

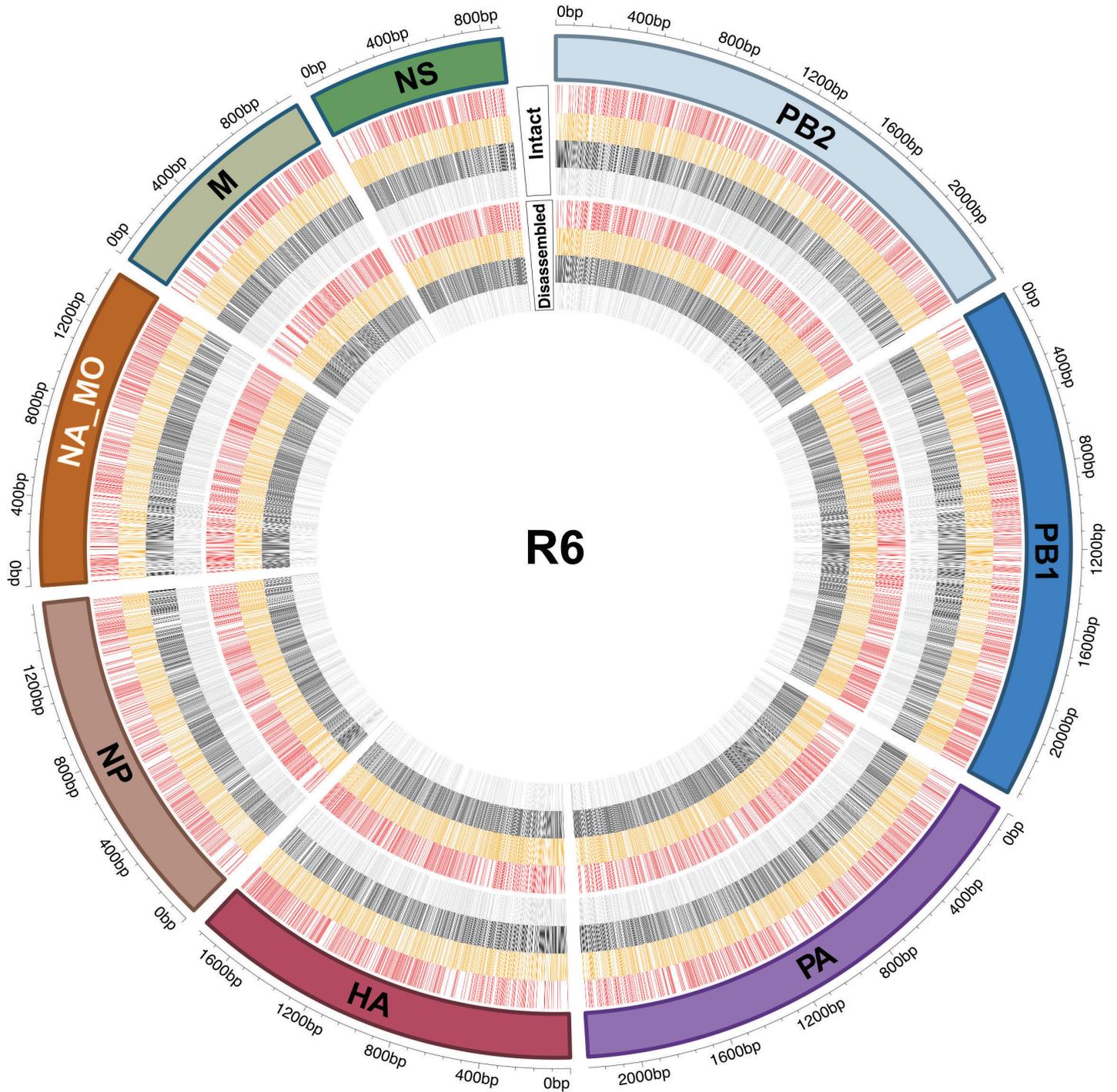


Fig 28 | Circos plot showing SHAPE-MaP reactivities of all segments of R6 virus

The circos plot represents average SHAPE reactivities of every nucleotide of each segment of R1 (PR8_{NA_{MO}}) virus. The red bars represent highly reactive nucleotides, yellow represent nucleotides with intermediate reactivities, black represent non-reactive nucleotides. The first panel (closest to the segment) represent SHAPE reactivities of intact virus while the second panel represent SHAPE reactivities of disassembled particles. The grey bars indicate nucleotide positions with no data.

We also generated median SHAPE graphs for each segment of PR8, R1 and R6 (**Figs 29 – 31 respectively**). Median graphs showed global rearrangement of RNA structural elements of the reassortants R1 and R6 when compared to PR8. This rearrangement was not only restricted to the non-cognate segment introduced in each virus, but it also affected the remaining genome. This is reminiscent of the plasticity of the RNA interaction network of IAV. A closer look at the segments of each virus also revealed while some regions underwent structural rearrangement between intact and disassembled conditions (indicated by boxes in the graphs), structural arrangements of other regions were retained. This implies that some regions are preferentially rearranged in the event of an introduction of non-cognate segment may be due to new partner segment interactions that are created in reassortants. However, SHAPE-MaP cannot give us partner details and therefore, we decided to complement this with computational predictions by vRNA site.

Although R1 and R6 differed from PR8 by only one segment, the SHAPE reactivity plots of the reassortants varied from PR8 and showed regional differences across all eight segments. This suggested a global rearrangement of the vRNP structures to accommodate MO segment 1 or 6 in the PR8 background to ensure successful packaging of the eight segments. However, a detailed and direct pairwise comparison of intact and disassembled vRNPs was performed with Δ SHAPE for each virus.

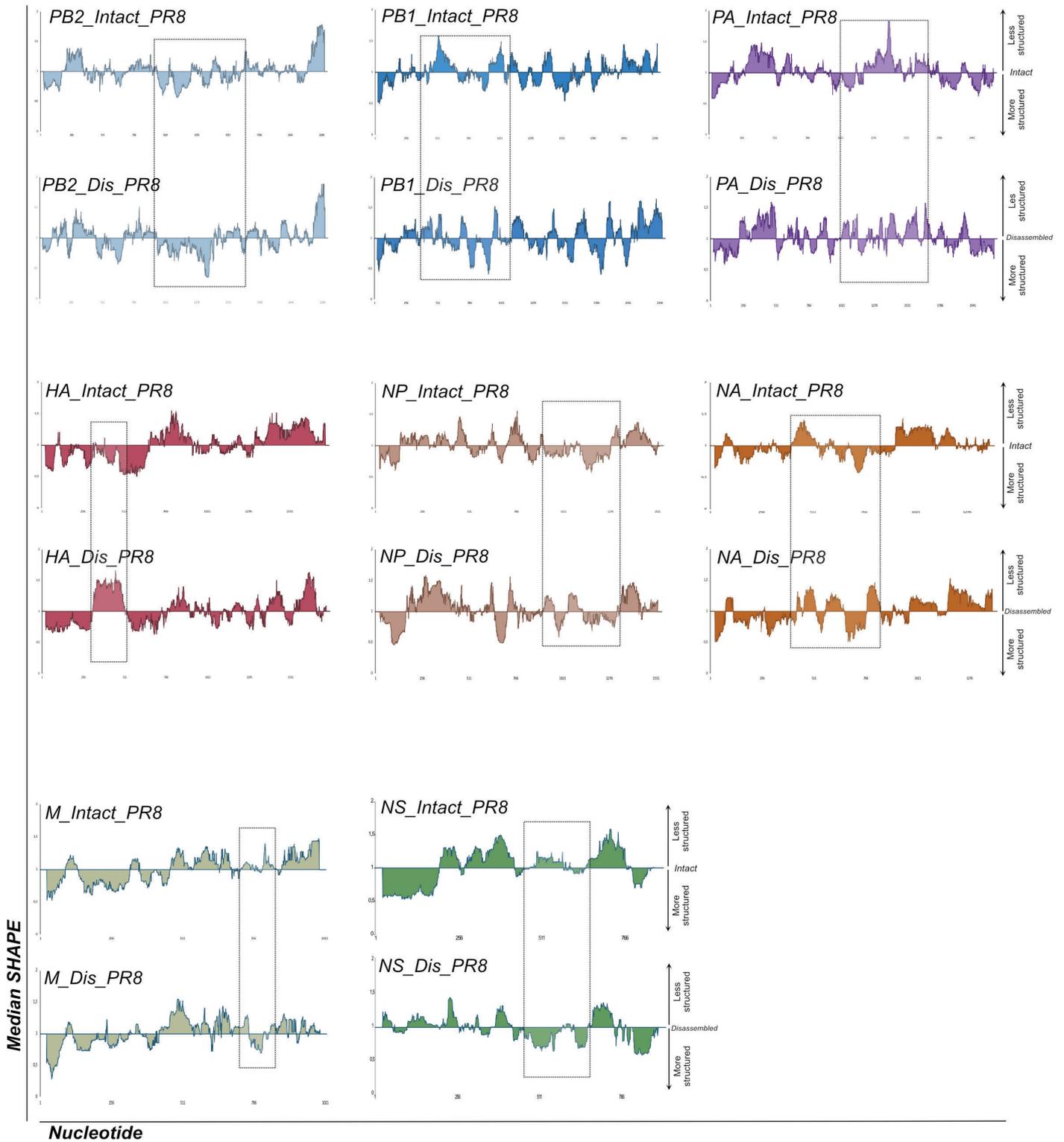


Figure 29 | Median SHAPE - MaP reactivities of vRNP segments one through eight of intact and disassembled PR8

Average per nucleotide reactivities were calculated from biological replicates 1 and 3 for each segment. Medians were calculated over 50 nucleotide windows and plotted relative to the global median of each segment. Median graphs of each segment is represented by a different colour. For each segment, the top and bottom profiles represent intact and disassembled conditions respectively. Nucleotide regions of low SHAPE-MaP reactivities fall below the global median. Therefore, regions below 1 are more structured than the remaining segment. Disassembly of particles causes severance of vRNPs from the 7+1 complex and consequently causes local structural rearrangement within each vRNP. These differences in structured regions for each segment between intact and disassembled regions are depicted by the dotted boxes. HA segment shows regions (256 - 511) with low SHAPE-MaP reactivities in the disassembled state which disappeared in its intact state. This suggests that the region of 256-511 of the HA segment underwent structural rearrangement from a flexible state with high SHAPE reactivity to a more restricted state with low SHAPE reactivity when the virus particle was disassembled.

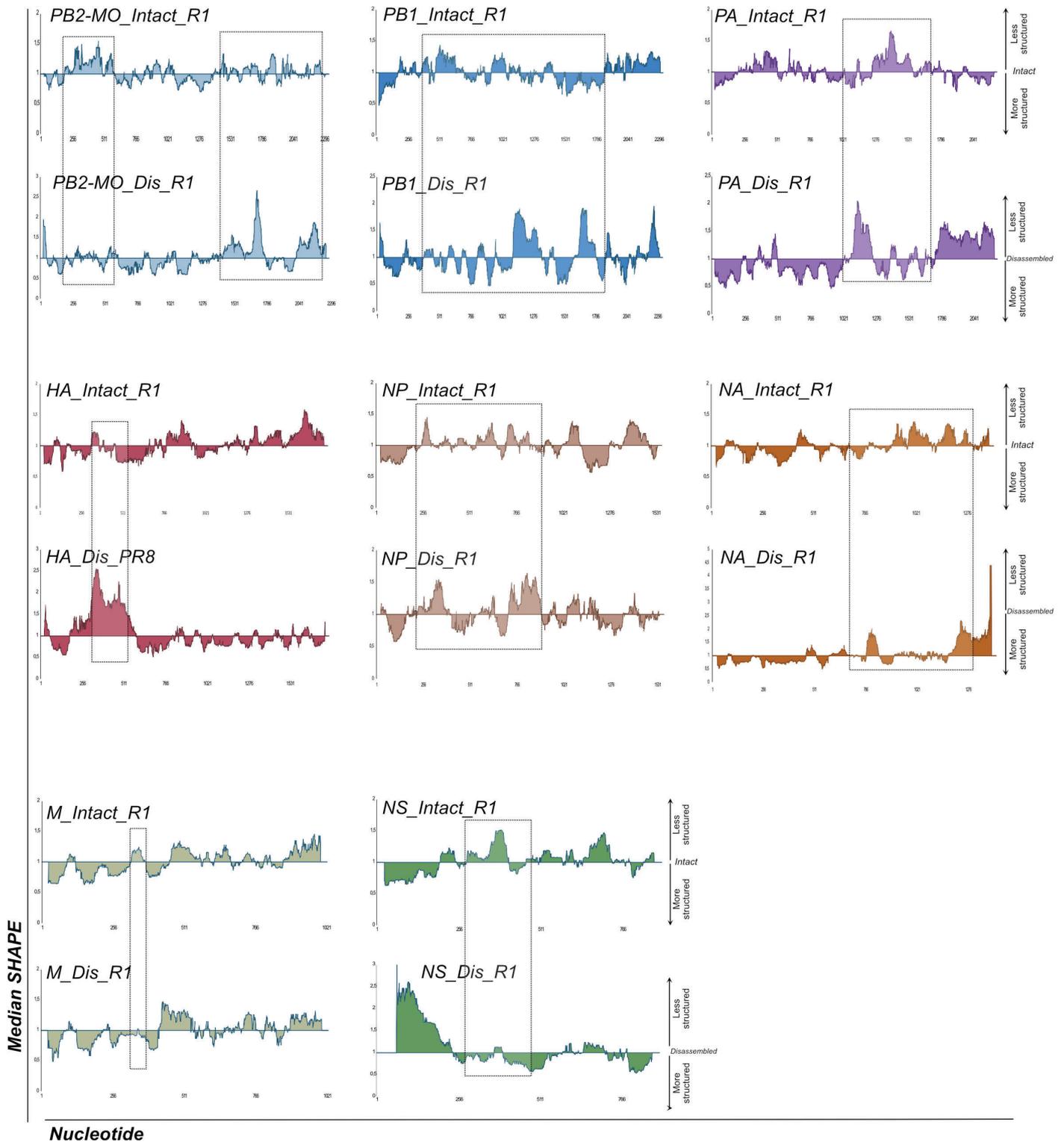


Figure 30 | Median SHAPE - MaP reactivities of vRNP segments one through eight of intact and disassembled R1

Average per nucleotide reactivities were calculated from biological replicates 1 and 3 for each segment. Medians were calculated over 50 nucleotide windows and plotted relative to the global median of each segment. Median graphs of each segment is represented by a different colour. Note that the PB2 segment in this virus is of MO origin (PB2_MO) while others are from PR8. For each segment, the top and bottom profiles represent intact and disassembled conditions respectively. Nucleotide regions of low SHAPE-MaP reactivities fall below the global median. Therefore, regions below 1 are more structured than the remaining segment. Disassembly of particles causes severance of vRNPs from the 7+1 complex and consequently causes local structural rearrangement within each vRNP. These differences in structured regions for each segment between intact and disassembled regions are depicted by the dotted boxes. M segment shows a small region between 256 - 511 with low SHAPE-MaP reactivities in the disassembled state which disappeared in its intact state. This suggests that this region of the M segment underwent structural rearrangement from a flexible state with high SHAPE reactivity in the intact particle to a more restricted state with low SHAPE reactivity after disassembly.

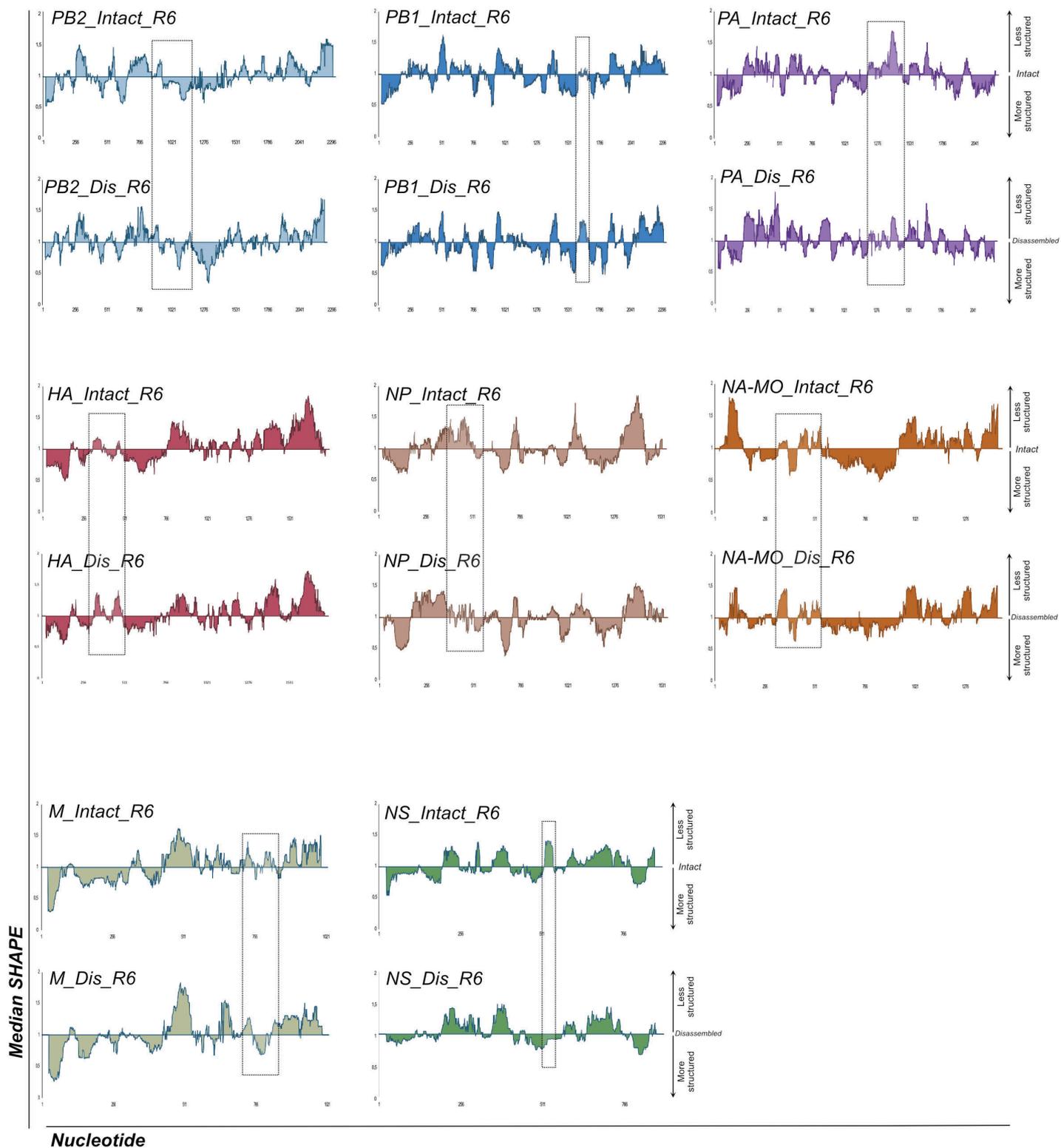


Figure 31 | Median SHAPE - MaP reactivities of vRNP segments one through eight of intact and disassembled R6

Average per nucleotide reactivities were calculated from biological replicates 1 and 3 for each segment. Medians were calculated over 50 nucleotide windows and plotted relative to the global median of each segment. Median graphs of each segment is represented by a different colour. For each segment, the top and bottom profiles represent intact and disassembled conditions respectively. Note that the NA segment in this virus is of MO origin (NA_MO) while others are from PR8. Nucleotide regions of low SHAPE-MaP reactivities fall below the global median. Therefore, regions below 1 are more structured than the remaining segment. Disassembly of particles causes severance of vRNPs from the 7+1 complex and consequently causes local structural rearrangement within each vRNP. These differences in structured regions for each segment between intact and disassembled regions are depicted by the dotted boxes. Regions around 511 of the NS have low SHAPE-MaP reactivities in the disassembled state which become highly reactive in the intact particles. This suggests structural rearrangement of the NS vRNP within the virus particle during incorporation into the 7+1 complex.

Custom Δ SHAPE parameters were used to compare intact and disassembled particles

Δ SHAPE is a statistical tool developed by Weeks *et al.*,²⁶¹ to contrast SHAPE reactivities of an RNA molecule under two different conditions. It compares the raw SHAPE reactivity of a nucleotide (*i*) under two conditions and its corresponding magnitudes of error to identify putative regions of interest with statistical significance. In this case, we used Δ SHAPE to compare differences in SHAPE reactivities of nucleotides between intact and disassembled samples.

$$\Delta SHAPE_i = SHAPE_{i (disassembled)} - SHAPE_{i (intact)}$$

In the original publications aimed at identifying protein binding sites, statistical thresholds of 'default Δ SHAPE' were derived from a well-studied group of RNP complexes^{261,239}. The SHAPE reactivities were first averaged on three nucleotides (nucleotides *i-1*, *i* and *i+1*) before computing the differences and statistical significance was assessed by computing a standard score ($|S|$) and a Z factor. Putative binding sites were identified as regions within five-nucleotide sliding windows for which at least three nucleotides had and $|S| \geq 1$. As we aimed to compare RNA structures, keeping a per nucleotide resolution was essential, and we did not average the SHAPE reactivities. We computed the Z score on non-averaged SHAPE reactivities, and rather than using a standard score; we included structurally relevant Δ SHAPE thresholds: namely, we imposed thresholds on the absolute and relative reactivity differences²⁶²:

$$|\Delta SHAPE| \geq 0.2$$

And

$$\frac{|\Delta SHAPE_i|}{1/2(SHAPE_{i (disassembled)} + SHAPE_{i (intact)})} \geq 0.4$$

Finally, we imposed that at least three nucleotides in a sliding window of nine nucleotides have a Z score > 0 and fulfil the abovementioned criteria (**Supplementary figs 6 - 11**).

Consequently, all the findings mentioned in the next section are identified through 'custom Δ SHAPE'.

vRNPs of intact particles showed higher SHAPE reactivities than their disassembled counterparts

The reactivity of 1M7 is influenced by local nucleotide flexibility. Hence, 1M7-mediated SHAPE reactivity is independent of nucleotide identity and inversely proportional to the unpaired nature of a nucleotide. The number of base-paired nucleotides is higher in intact particles since there are inter and intrasegmental interactions than in disassembled particles with only intrasegmental interactions. Consequently, these base-paired nucleotides would be shielded from reacting with 1M7. Therefore, we expected a higher number of regions with 1M7 reactive nucleotides in disassembled vRNPs distributed globally across each segment as

opposed to its intact counterparts. As a result, compared with intact viruses, we expected an overall increase in the number of positive Δ SHAPE regions in the genome of disassembled particles.

Positive Δ SHAPE \rightarrow increased reactivity in disassembled particles

Negative Δ SHAPE \rightarrow increased reactivity in intact particles

However, Δ SHAPE patterns on all viruses showed intriguing results. Contrary to our expectations, in the PR8 virus, while $\sim 8\%$ of the total genome showed regions with negative Δ SHAPE, only 4% showed positive Δ SHAPE (Fig 32).

This implies that intact PR8 viruses have more open single-stranded regions that are reactive to 1M7 than disassembled particles. Conversely, upon disassembly of PR8, some of these regions in the genome become non-reactive to 1M7. This would imply local structural remodelling of the vRNPs upon disassembly into a more compact form, thus restricting the number of unpaired nucleotides available for reacting with

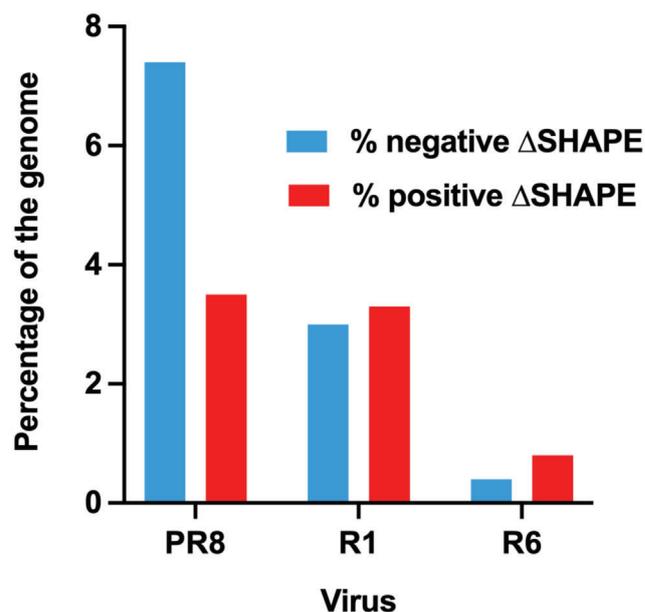


Fig 32 | Comparison of negative and positive Δ SHAPE regions.

8% of the PR8 genome was more reactive in intact particles than in disassembled where only 4% of the genome was reactive. While, in both reassortants R1 and R6, we noticed a drop in the regions with negative Δ SHAPE. However, all viruses showed a general trend of having more reactive regions in intact particles than their disassembled counterparts.

1M7. Similar findings were observed by Dadonaite *et al.*, where they observed that loci identified as potential RNA-RNA interaction sites by SPLASH contained more RNA structural elements than the remaining genome²⁰⁵. vRNA from Flaviviruses and satellite tobacco mosaic virus (STMV) also had higher SHAPE reactivities (less RNA structural elements) within virions than *in vitro* or *ex virio* conditions^{263,199}. These results are consistent with the idea that viral RNA in cells or in virions may be bound by proteins and hence more compact, protecting them from reacting with 1M7¹⁸⁹. Simulation studies conducted by Perlmutter *et al.*, also observed that compact RNA structures or more base-paired RNA led to optimal genome packaging in viruses¹³¹. Studies show that the amount of RNA bases that can be packaged depends on how efficiently RNA secondary structures are formed. It has been demonstrated that nonviral RNA is usually less compact than the viral RNA genomes when comparing RNA molecules of the same length and similar amount of base-pairing¹³⁶.

During disassembly, severing vRNPs from the supramolecular 7+1 complex may drive it to return to an alternate conformation where some local structural remodelling occurs, some of the RNA structural elements are preserved while others giving rise to some new elements. This alternate conformation may ensure that the vRNP retains its structural integrity as an independent entity in the biochemical environment. It could also be speculated that upon disassembly, the vRNPs undergo local refolding to conceal potential RNA secondary structures that may have mediated intersegmental RNA-RNA interactions through kissing loops in the 7+1 supramolecular complex. Eventually, disassembly and results in vRNP compaction mediated either through NP remodelling along the length of vRNP. Atomic force microscopy on IAV vRNPs have shown they are highly flexible, a trait conferred by inherent dynamic nature of the viral NP protein²⁶⁴. There is also evidence of tolerance of RNA disorder to a certain degree within the NP protein of IAV²⁶⁵. NP protein induces local melting of RNA secondary structures to cause changes in the vRNP conformation²⁶⁶. This could facilitate the movement of short stretches of nucleotides within the RNA-NP binding regions to conceal or expose regions of interactions while maintaining the helical rings. However, the impact of such vRNP remodelling on its length, stability and other tertiary interactions remains elusive.

The plasticity of vRNP structure is critical for RNA synthesis, nuclear import/export, and virus assembly¹⁶⁹. Mutations in NS1 and PA that inhibit particle formation without apparently affecting vRNA synthesis or trafficking have been identified confirming this apparent plasticity¹⁹⁷. Since we probed intact virus particles following budding and release, we can assume that the intrinsic conformation of the 7+1 supramolecular complex during genome assembly was preserved within these particles. Therefore, the RNA structural elements observed in intact particles represent the 7+1 complex, albeit other confounding factors such as vRNP distortion following budding to accommodate into a spherical virus²⁶⁷. Therefore, we would assume that packaging signals dispersed along the length of the genome when coupled with an ordered RNP structure would result in detectable periodicity in the spacing of conserved residues¹⁹⁷. However, this may not be true because packaging signals can be masked by flexibility in the RNP structure and/or its tertiary interactions with other ligands. This provides further insights into our understanding of genome packaging by considering intramolecular charge repulsion compensation between NP and vRNA by accounting for the compact conformations of viral genomes. Therefore, we may speculate that while sequence-specific packaging signals determine selective packaging, the flexible nature of vRNP can accommodate a repertoire of RNA structural elements formed by long-range intramolecular interactions that account for the plasticity of the genome packaging network.

Δ SHAPE patterns of R1 and R6

Meanwhile, contrary to Δ SHAPE patterns of PR8, there was an increase in the number of positive Δ SHAPE regions of R1 and R6 when compared to the negative Δ SHAPE regions. However, this difference was too small to be significant. Nevertheless, it was perplexing to

observe similar Δ SHAPE patterns between R1 and R6 which contrasted with the patterns of PR8. Also, there was a decrease in the overall number of negative Δ SHAPE regions in the genome of R1 and R6 when compared to PR8 (**Fig 32**).

To check if the vRNP structural rearrangements were localised only to the non-cognate H3N2 segment of R1 and R6, a segment-wise breakdown of all significant Δ SHAPE regions was performed (data not shown). As evidenced from the median graphs, the distribution of significant Δ SHAPE regions in a non-segment-specific manner confirms the global remodelling of RNA structures across all eight segments to maintain the 7+1 complex. This evokes the redundancy theory of the IAV genome packaging model which explains that selective and coordinated packaging is mediated by a plethora of tiered RNA-RNA interactions chosen in order of maximum efficiency for virus fitness and adaptation²³⁷.

Further examination of the Δ SHAPE regions showed an interesting nucleotide distribution pattern. There was a strong bias towards GC nucleotides in all the negative Δ SHAPE regions, whereas, in all the positive Δ SHAPE regions, there was a strong predominance of AU regions (**Figs 33**). GC interactions, although short, are known to be stronger and, therefore, confer more stability and are preferred in kissing loop interactions²⁶⁸. Arguably, it could be postulated that the higher number of such GC regions in intact particles plays a crucial role in forming stem loops that mediate short yet strong intersegmental RNA-RNA interactions. Such GC-rich stem loops may be refolded into the vRNP structure upon virus disassembly and masked by the NP protein to prevent damage. During this process, the vRNP maintains flexibility by local NP remodelling to expose AU/GU regions.

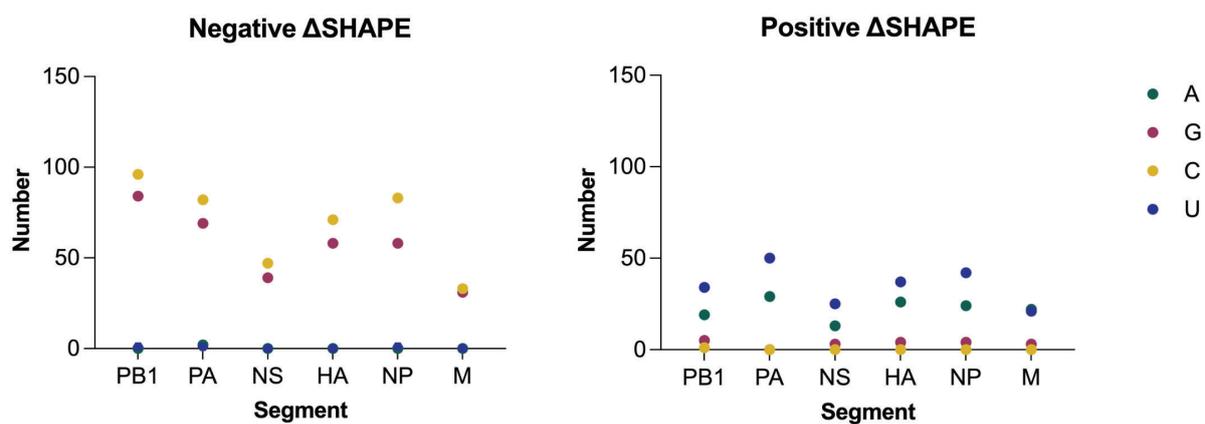


Fig 33 | Nucleotide distribution among negative and positive ASHAPE regions.

There was a bias towards GC nucleotides in all the negative ASHAPE regions whereas, in all the positive ASHAPE regions, there was a strong predominance of AU regions.

Caveats of SHAPE-MaP setup

Although high-throughput RNA probing approaches have advanced our understanding of RNA structure, the experimental setup has some caveats. RNA is extracted from its cellular

environment for *in vitro* probing assays, sometimes combined with de- and re-naturing steps. Therefore, it lacks RNA-binding proteins or other factors that influence structure formation. Consequently, the probed structure may not always be identical to the native structure¹⁶⁸. However, in our experiments the RNA structure was directly probed into native, intact viral particles. For disassembling viral particles, we used a well-established protocol²⁵⁹ that was also used as the first step of vRNP purification in cryo-EM studies of vRNPs¹⁶⁹ and contained no protein removal or denaturation/renaturation steps.

The mechanism of SHAPE probing can make reactivity interpretation challenging. For example, unpaired bases in an RNA hairpin loop can still be constrained by base stacking interactions that make them weakly reactive to SHAPE probes²⁶⁹. However, if interpreted appropriately, this can reveal higher-order structures and tertiary contacts, manifesting as regions of intermediate to weak reactivities¹⁷¹. However, any conformational change involving an RNA-protein or intramolecular interaction also reflect as low SHAPE reactivities. Therefore, SHAPE reactivities are intrinsically biased by tertiary interactions of an RNA molecule and should consequently be supported by detailed knowledge of the structure of probed RNA.

Additionally, chemical probing datasets are inherently noisy, and reproducibility remains an issue. How to deal with cases where the RNA forms an ensemble of diverse structures needs to be determined. Probing data may be less useful in such cases. Even an equilibrium of just two structures could result in the pairing probabilities of exactly 50 % for every nucleotide, thus yielding a completely uninformative probing signal. On the upside, current methods do not yet make the best possible use of probing data since they assume a binary distinction between paired and unpaired positions. Probing reactivity will depend on more structural details and should, therefore, give information on more classes of structural context.

Although various SHAPE reagents have been used to probe different RNA secondary structures, in this study, we used 1M7. The short reaction time (75s) of 1M7 and no additional quenching steps make it very easy to handle under experimental settings²⁷⁰. Additionally, comparative studies on 1M7 with other SHAPE reagents showed that it selectively probes unpaired nucleotides without any nucleotide bias and shows flexibility on all four ribonucleotides with similar reactivity values²⁷¹. Busan *et al.*, compared SHAPE reagents such as 1M7, NAI, and 5NIA to check for nucleotide biases. They showed that while all three reagents could discriminate unpaired from paired nucleotides successfully, NAI and 5NIA exhibited profound biases to nucleotide identity. For instance, NAI could effectively distinguish unpaired and paired adenosine residues but less effectively from unpaired and paired guanosine and cytosine. Likewise, 5NIA showed excellent discrimination for all four ribonucleotides but reacted with adenosine at higher relative rates than the other three nucleotides²⁷¹. Hence, 1M7 is a suitable probing reagent for local nucleotide flexibility of RNA secondary structures independent of nucleotide identity. However, ribonucleotide flexibility

depends on base pairing and any interactions that restrict nucleotide dynamics such as tertiary interactions with other RNA/proteins. Hence, reactivities produced by 1M7 could potentially indicate ribonucleotide constraints in the structure independent of base pairing. Consequently, it can only be used as a proxy to study secondary structures when coupled with best practices such as good read depth, reproducible experimental conditions and stringent probing times and temperatures¹⁸⁶.

Despite using the best practices, in this study, the mutation rates of samples were often only around 1 % above background, which eventually led to the failure of the final quality control metric of Shapemapper2. On reaching out to P. Irving [University of North Carolina at Chapel Hill] about this discrepancy, he mentioned that ideally, mutation rates with 1M7 should be around 3 % above background. A recurring problem with 1M7 is its intrinsic low reactivity when compared with other reagents such as 5NIA (3 - 5 % nucleotide reactivity above background) or 2A3 (8 - 12 % nucleotide reactivity above background), and its ability to become extremely unstable with exposure to moisture. Since, in this dataset, the overall reactivity rates were low, we suspect a problem with the stability of the reagent and or its inability to probe the highly structured vRNPs.

To circumvent this problem, one could use an additional reagent with different probing chemistry (eg. DMS) along with 1M7 to study the structure of vRNPs. DMS methylates N1 of adenine and N3 of cytosine on the Watson–Crick base pairing face of non-structured RNA²⁷². Although DMS exhibits nucleotide specificities to base-paired adenosine and cytosine, it can be used to study vRNP-NP interactions or interactions with other proteins and potential structural changes of the vRNP. Coupling DMS methylation and SHAPE acylation data from the same samples to guide RNA secondary structure modelling may provide a better outcome.

Additionally, the current SHAPE-MaP data on disassembled particles indicate potential structural vRNP remodelling outside the supramolecular complex, it would be interesting to study the nature of individual vRNPs with an additional structural/biophysical technique such as single-particle high-speed atomic force microscopy. Furthermore, virus disassembly using previously published protocols showed residual M1 protein associated with vRNP cores which may also influence the structural conformation undetected by SHAPE-MaP. Hence, the non-reactive regions observed in the reactivity profiles may reflect spurious RNA-RNA or RNA-protein interactions. Such tertiary interactions that have no scope of being identified or distinguished by SHAPE-MaP warrant further studies with other complementary techniques before narrowing down RNA-RNA interactions obtained only from SHAPE-MaP.

vRNAsite

vRNAsite is a novel tool that efficiently predicts intermolecular RNA-RNA interaction (IRI) sites between two or more RNA sequences²¹⁹. It creates a minimum free energy (MFE) landscape

of IRIs with the guidance of a sliding window between any combination of two RNA sequences. With the default window size of 20, the algorithm returns strong, short, consecutive interactions, like the IRI presented by Gavazzi *et al.*²⁷³ while considering the thermodynamic influence of surrounding nucleotides.

To predict IRIs, vRNAsite first creates IRI contact maps, which are MFE matrices representing the average of the overlapping window predictions. These average free energy matrices represent the possible IRIs between two given input sequences. From an experimentally verified IRI the authors could infer, that an IRI has to be at most below a certain energy threshold. A significant increase of possibly insignificant IRIs at a peak average MFE of around -17.0 kcal/mol and higher was observed. Therefore, these were used to estimate a good threshold for the peak of average energies, and only 4 % of the predictions were considered insignificant at that threshold.

Collaborating with us, Daniel Desiro [Robert Koch Institute, Berlin] first generated sets of Raw vRNA site tables. MFE is the minimum free energy of the "RNA" in kcal/mol that was predicted with the viennaRNA package²⁰⁸ and the corresponding dot-bracketed folding structures. "a_structure" and "b_structure" shows if the two interacting regions can form an intramolecular structure (e.g. hairpin). The "peak" column is the minimum of the average MFE in the interacting region. Next, he generated vRNAsite plots, interactive "bokeh" plots, that can be opened in a web browser. An example dataset of the vRNAsite tables and bokeh plots created by Daniel for the HA segment of MO can be viewed here: <https://github.com/rithupaul/vRNAsite.git>. They show the interaction landscape as "peak" values. Finally, he generated Circos plots of predicted intersegmental RNA-RNA interactions using two different "peak" thresholds, -10.0 and -17.0 kcal/mol (**Figs 34 a-d**).

We then used the predicted intersegmental RNA-RNA interactions at a threshold of -17.0 kcal/mol to see if they fell in any regions identified as positive Δ SHAPE. We first tried to find any positive Δ SHAPE regions that could be shared between the cognate segments across the three viruses PR8, R1, and R6. We repeated this same process for the negative Δ SHAPE regions. The positive and negative Δ SHAPE regions shared among viruses PR8, R1 and R6 are shown in the Venn diagrams (**Figs 35 - 36**).

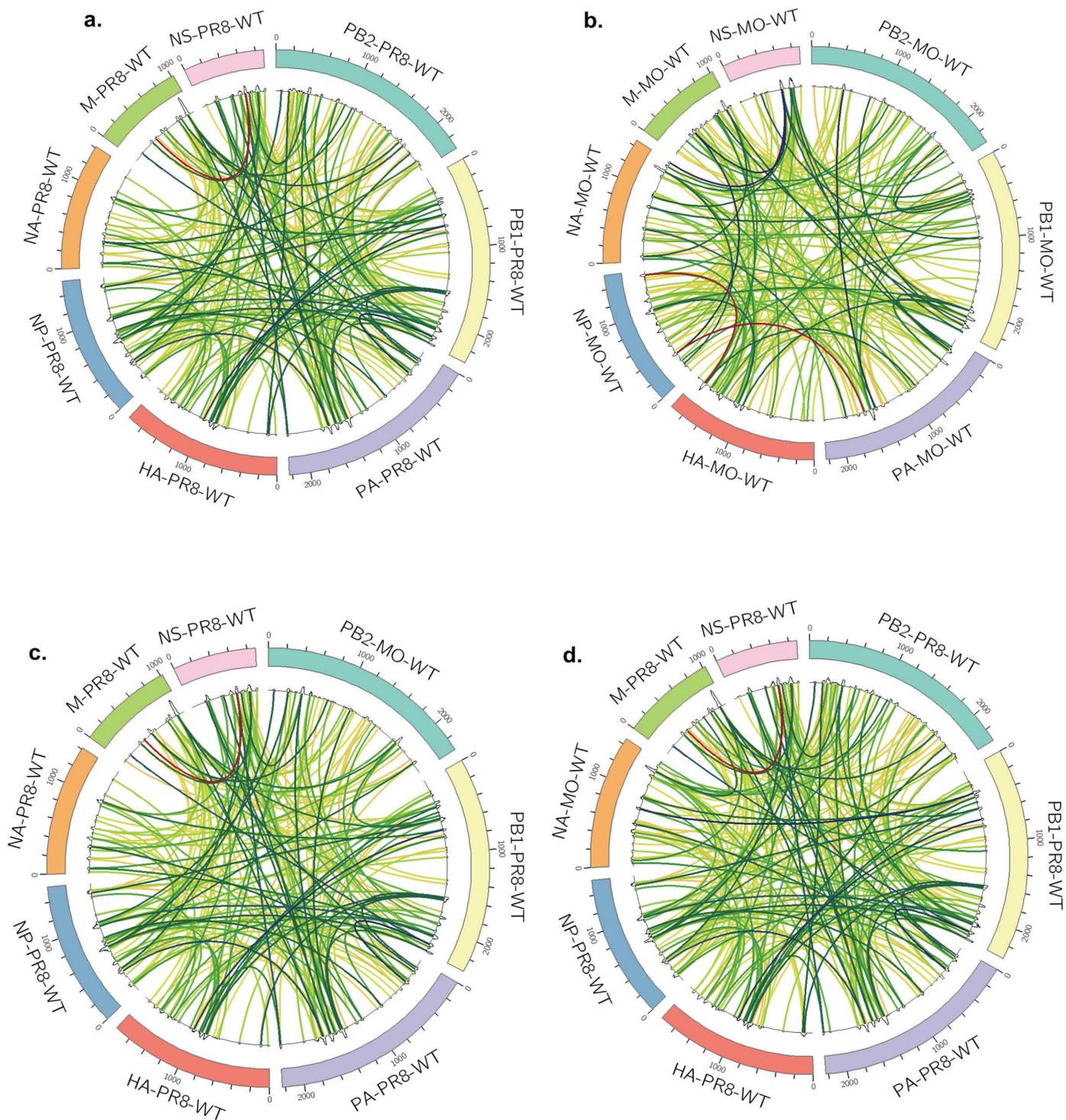


Figure 34 | Circos plots generated by vRNAsite for PR8 (a), MO (b), R1 (c) and R6 (d)

Circos plots were produced with vRNAsite to predict intersegmental RNA-RNA interactions using two different "peak" thresholds, -10.0 and -17.0 kcal/mol. The following circos plots show all possible intersegmental RNA-RNA interactions that occur at an MFE of -17.0 kcal/mol. The colour bars are the same as the ones in the bokeh plots.

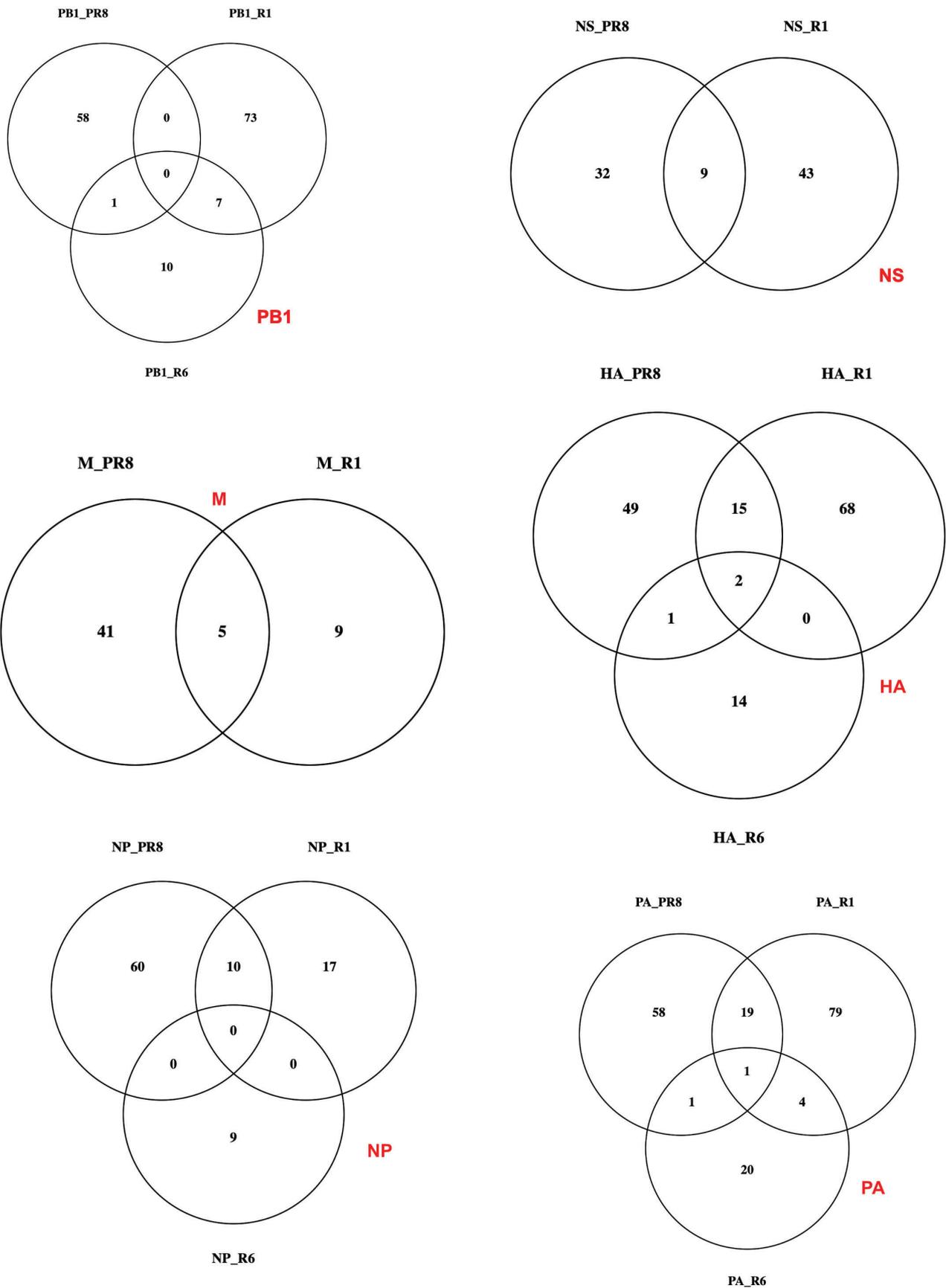


Fig 35 | Venn diagram showing all the positive Δ SHAPE regions predicted by vRNAsite that are shared by viruses

All the positive Δ SHAPE regions identified were searched in the database of putative interactions predicted with vRNAsite. Each Venn diagram represents a segment and the number of interactions that are shared among PR8, R1 and R6. These shared regions could potentially indicate putative RNA-RNA intersegmental interactions that would require further investigation to study their role in genome packaging.

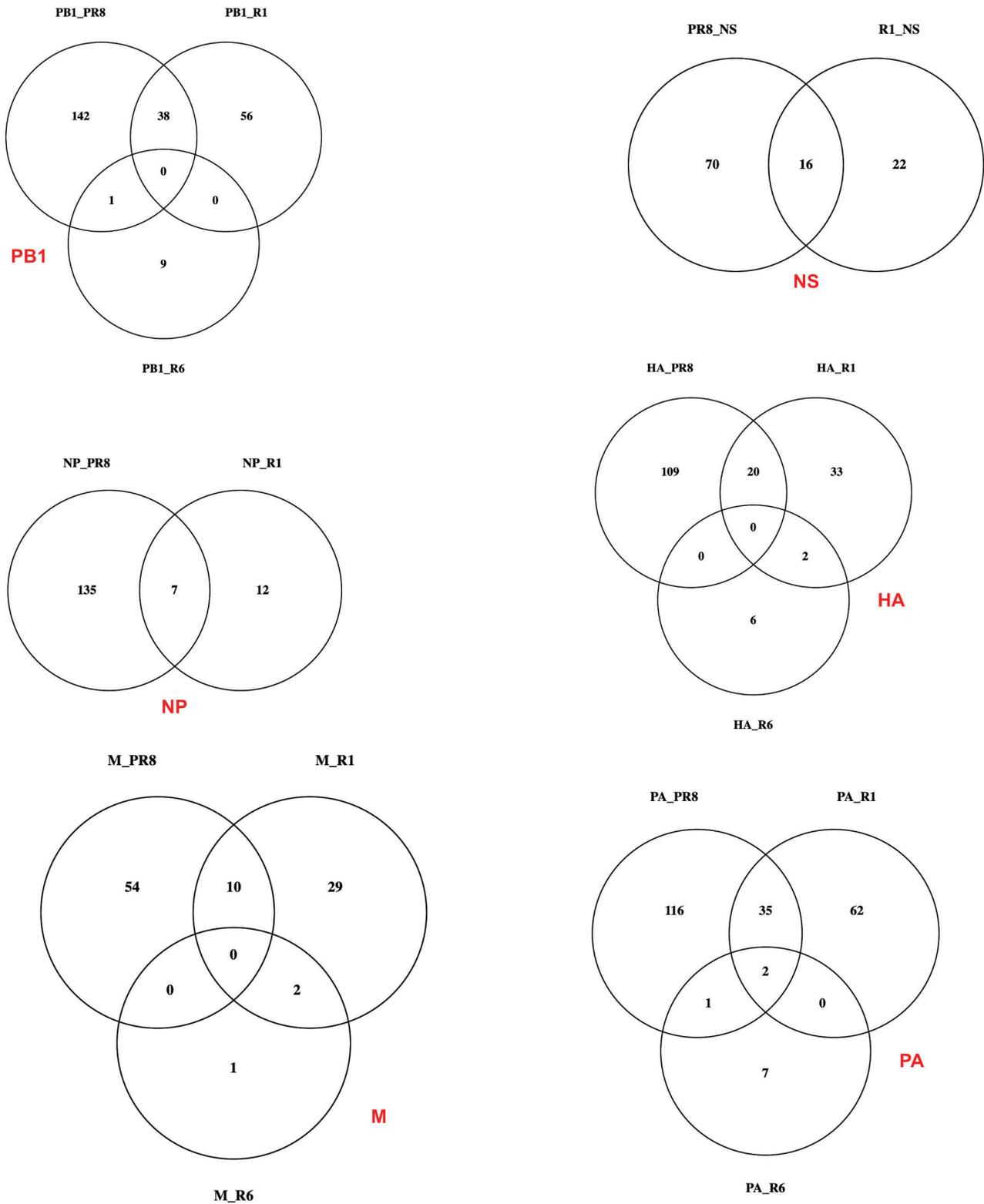


Fig 36 | Venn diagram showing all the negative Δ SHAPE regions predicted by vRNA site that are shared by viruses

All the negative Δ SHAPE regions identified were searched in the database of putative interactions predicted with vRNA site. Each Venn diagram represents a segment and the number of interactions that are shared among PR8, R1 and R6. These shared regions could potentially indicate putative RNA-RNA intersegmental interactions that would require further investigation.

Conclusions & Prospects

This study aimed to study the growth patterns and RNA secondary structures of H1N1/PR8, H3N2/MO, and two single-segment reassortants R1 (PR8_PB2_{MO}) and R6 (PR8_NA_{MO}) produced by 7:1 reverse genetics from PR8 and MO. We used SHAPE-MaP to study the RNA secondary structures in intact and disassembled viral particles of PR8, R1, and R6.

R1 and R6 reassortants were chosen with the parental strains PR8 and MO for further studies based on their growth properties. The growth curves of R1 and R6 were similar to the parental strains PR8 and MO, indicating virus fitness. The remaining six reassortants were attenuated. This finding suggested that the H3N2/MO PB2 and NA segments and their corresponding proteins are compatible with H1N1/PR8. However, we need further experiments to determine if the opposite (i.e. if individual H3N2/MO proteins are compatible in an H1N1/PR8 background) will hold true. For instance, we could perform reverse genetics experiments with the H3N2/MO as the backbone and check growth patterns of the reassortants produced. This would be a prospective direction to study potential genetic constraints that restrict reassortment between H1N1 and H3N2 strains of IAV.

Moreover, we have only performed reverse genetics with 7:1 segment combinations in this study. Since there is ample evidence of segment co-segregation in IAV, we could perform alternative 5:3 (non-cognate PB2, PB1, PA), 6:2 (non-cognate HA, NA or NP, M etc) or 4:4 segment combinations to study reassortment at differing capacities. Some reassortants produced due to these combinations likely outperform the growth kinetics of R1 and R6 obtained in this study. Therefore, we could improve our understanding of potential segment-specific driving factors for reassortment between H1N1 and H3N2 strains.

We used the single-segment reassortants R1 and R6 in competitive coinfection settings to look for suboptimal compatibility between the non-cognate and cognate segments in the reassortants. We found very intriguing results from this set of experiments. Firstly, although PR8_HA_{tag} differed from the parental PR8 by only a synonymous mutation, the HA_{tag} segment seemed to have better packaging efficiency during all three conditions of coinfection experiments. Further studies on the replicative ability of this virus when compared with PR8 may give reason to explain this intriguing phenomenon. It would be interesting to check if indeed the introduction of a synonymous mutation in the PR8_HA segment improved the packaging efficiency of the segment. This could have further potential implications on the supposed nature of the covariation in RNA segments and its impact on genome packaging of IAV. Secondly, we observed a few reassortant virus genotypes with more than eight segments typical for IAVs. While this phenomenon is not uncommon and has been reported previously, in this study, we have not been able to distinguish this phenomenon from the presence of multi-particle aggregates within a plaque or the incorporation of more than eight segments within a virion. As a result, ruling out the possibility of multi-particle aggregates and checking for the presence of more than eight segments within a virus would be an exciting avenue for further studies.

Following coinfection assays, we used purified virus supernatants to modify with the SHAPE reagent 1M7 for chemical probing followed by mutational profiling. This is a unique SHAPE-MaP study on IAV reassortants besides the collaborative FluCode project between our laboratory and the one of Prof. Martin Schwemmle in the University of Freiburg. In this study, we compared RNA secondary structures of intact and disassembled IAV virus particles with Δ SHAPE. Contrary to our hypothesis, we observed a higher number of negative Δ SHAPE regions across all segments, which is indicative of more relaxed vRNP structures in intact particles. We also observed extensive local and global structural remodelling of vRNPs upon disassembly of virus particles which leads us to speculate we need better tools to study genetic reassortment and genome packaging extensively. Reproducibility of chemical probing data is quite tricky, and the fact that we used intact virus particles to probe the flexible and dynamic IAV vRNP made direct inference of nucleotide reactivities and their implications on the paired/single-stranded nature of vRNA a challenge. Perhaps this could be mitigated by complementing this dataset with DMS-MaP to provide a better understanding of potential RNA-NP interactions and distinguish them from RNA-RNA interactions identified by SHAPE-MaP.

Our SHAPE-MaP results suggest that there could be local structural remodelling of vRNP between intact and disassembled states. The flexible nature of vRNP facilitates key processes of IAV genome replication. Additionally, vRNA packaged into virus particles are known to be more compact than other RNA molecules. Arguably, there may be reason to speculate that during the genome packaging of IAV vRNPs, they undergo local structural remodelling to expose packaging signals that facilitate segment bundling and progressive assembly. However, there may be better solutions than using SHAPE-MaP alone to study this and it would require high-resolution biophysical techniques, preferably in a single virus particle format.

Perhaps it is less of a challenge to first study genome packaging in individual strains such as H1N1/PR8, H3N2/Udorn, etc, and then use that repertoire of knowledge to inform the rules of packaging during reassortment. Since IAV genome packaging is unfailingly rigorous and selective, having a fundamental understanding of the rules of genome packaging before reassortment studies may prove helpful in bridging some of the existing knowledge gaps. Admittedly, the redundancy exhibited by the IAV genome packaging network makes this more challenging than it seems. Nevertheless, if we identify putative RNA structures and validate their biological relevance in genome packaging while also showing the lack of these interactions can cause packaging defects in viruses, we are off to a promising start.

Different laboratories have developed a diverse suite of bioinformatics and functional analysis tools, and collectively, these efforts support the validation of numerous functional RNA structural elements across viral genomes¹⁸⁹. Nonetheless, it remains challenging to identify

the subset of functionally important motifs in the context of extensive secondary structure folding. Implementing virus functional assays is time-consuming and often inefficient, which has limited the number of elements that have been functionally investigated¹⁸⁹. New higher-throughput strategies for viral functional analysis are critically needed¹⁸⁹. We also need novel computational strategies using previously identified functional packaging signals as learning models to study patterns in putative RNA motifs with implications in packaging. Understanding the molecular mechanism of genome packaging is crucial as it provides promising avenues for antiviral research.

To conclude, we have not delineated intersegmental RNA-RNA interactions that mediate genome packaging in this study. While a certain set of interactions are indispensable to genome packaging, and their absence can lead to absolute loss of segments, we may have to suspect that the proximity of genetically similar strains in a coinfection setting can expand the repertoire of functional RNA structural motifs, increasing the reassortment potential while simultaneously diversifying the genotypes of progeny reassortants. It may be speculated that, instead of a finite discrete set of intersegmental RNA-RNA interactions, we may be looking at a dynamic discontinuous redundant set of regions that exhibit a hierarchy in packaging. A set of specific high-affinity packaging signals coupled with low-affinity transient interactions mediated by the vRNA and NP and vRNA acting separately or synergistically mediate genome packaging. There could be a dynamic set of high-affinity and low-affinity packaging signals specific to each strain. This set of interactions can ensure selective packaging of each vRNP occurs while accommodating the flexibility to incorporate non-cognate segments in the event of coinfection. This ensures reassortment potential under favourable conditions when genetically homologous strains are infected in close proximity. Therefore, it is safe to assume that while studying IAV genetic reassortment is more complicated than anticipated, every new piece of information adds to the existing database of information while creating potential novel research avenues.

References

1. Monaghan, K. J. SARS: DOWN BUT STILL A THREAT. in *Learning from SARS: Preparing for the Next Disease Outbreak: Workshop Summary* (National Academies Press (US), 2004).
2. Trifonov, V., Khiabani, H. & Rabadan, R. Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus. *N Engl J Med* **361**, 115–119 (2009).
3. Rha, B. *et al.* Update on the Epidemiology of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Infection, and Guidance for the Public, Clinicians, and Public Health Authorities — January 2015. *MMWR Morb Mortal Wkly Rep* **64**, 61–62 (2015).
4. Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. *N Engl J Med* **371**, 1481–1495 (2014).
5. Phipps, K. L. *et al.* Seasonal H3N2 and 2009 Pandemic H1N1 Influenza A Viruses Reassort Efficiently but Produce Attenuated Progeny. *J Virol* **91**, e00830-17 (2017).
6. Jakob, C., Paul-Stanslaus, R., Schwemmle, M., Marquet, R. & Bolte, H. The influenza A virus genome packaging network - complex, flexible and yet unsolved. *Nucleic Acids Res* **50**, 9023–9038 (2022).
7. Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc* **10**, 1643–1669 (2015).
8. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral Mutation Rates. *Journal of Virology* **84**, 9733–9748 (2010).
9. Steinhauer, D. A., Domingo, E. & Holland, J. J. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* **122**, 281–288 (1992).
10. Paillart, J.-C. *et al.* First Snapshots of the HIV-1 RNA Structure in Infected Cells and in Virions *. *Journal of Biological Chemistry* **279**, 48397–48403 (2004).
11. Chan, J. M., Carlsson, G. & Rabadan, R. Topology of viral evolution. *Proceedings of the National Academy of Sciences* **110**, 18566–18571 (2013).
12. Vijaykrishna, D., Mukerji, R. & Smith, G. J. D. RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion. *PLoS Pathog* **11**, e1004902 (2015).
13. Lowen, A. C. It's in the mix: Reassortment of segmented viral genomes. *PLOS Pathogens* **14**, e1007200 (2018).

14. Lowen, A. C. Constraints, Drivers, and Implications of Influenza A Virus Reassortment. *Annu Rev Virol* **4**, 105–121 (2017).
15. Renaud, C., Kuypers, J. & Englund, J. A. Emerging oseltamivir resistance in seasonal and pandemic influenza A/H1N1. *J Clin Virol* **52**, 70–78 (2011).
16. Deyde, V. M. *et al.* Surveillance of Resistance to Adamantanes among Influenza A(H3N2) and A(H1N1) Viruses Isolated Worldwide. *The Journal of Infectious Diseases* **196**, 249–257 (2007).
17. Schäffr, J. R. *et al.* Origin of the Pandemic 1957 H2 Influenza A Virus and the Persistence of Its Possible Progenitors in the Avian Reservoir. *Virology* **194**, 781–788 (1993).
18. Jester, B. J., Uyeki, T. M. & Jernigan, D. B. Fifty Years of Influenza A(H3N2) Following the Pandemic of 1968. *Am J Public Health* **110**, 669–676 (2020).
19. Van Poucke, S. *et al.* Role of Substitutions in the Hemagglutinin in the Emergence of the 1968 Pandemic Influenza Virus. *J Virol* **89**, 12211–12216 (2015).
20. null null. Clinical Aspects of Pandemic 2009 Influenza A (H1N1) Virus Infection. *New England Journal of Medicine* **362**, 1708–1719 (2010).
21. null, null. Emergence of a Novel Swine-Origin Influenza A (H1N1) Virus in Humans. *New England Journal of Medicine* **360**, 2605–2615 (2009).
22. Van Schalkwyk, A., Coetzee, P., Ebersohn, K., Von Teichman, B. & Venter, E. Widespread Reassortment Contributes to Antigenic Shift in Bluetongue Viruses from South Africa. *Viruses* **15**, 1611 (2023).
23. Kopanke, J., Lee, J., Stenglein, M. & Mayo, C. In Vitro Reassortment between Endemic Bluetongue Viruses Features Global Shifts in Segment Frequencies and Preferred Segment Combinations. *Microorganisms* **9**, 405 (2021).
24. Hara, K., Nakazono, Y., Kashiwagi, T., Hamada, N. & Watanabe, H. Co-incorporation of the PB2 and PA polymerase subunits from human H3N2 influenza virus is a critical determinant of the replication of reassortant ribonucleoprotein complexes. *J Gen Virol* **94**, 2406–2416 (2013).
25. Li, C., Hatta, M., Watanabe, S., Neumann, G. & Kawaoka, Y. Compatibility among polymerase subunit proteins is a restricting factor in reassortment between equine H7N7 and human H3N2 influenza viruses. *J Virol* **82**, 11880–11888 (2008).

26. Handel, A., Akin, V., Pilyugin, S. S., Zarnitsyna, V. & Antia, R. How sticky should a virus be? The impact of virus binding and release on transmission fitness using influenza as an example. *J R Soc Interface* **11**, 20131083 (2014).
27. Essere, B. *et al.* Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proc. Natl. Acad. Sci. U.S.A.* **110**, (2013).
28. Taylor, K. Y. *et al.* Influenza A virus reassortment is strain dependent. *PLoS Pathog* **19**, e1011155 (2023).
29. Folimonova, S. Y. Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein. *Journal of Virology* **86**, 5554–5561 (2012).
30. Marshall, N., Priyamvada, L., Ende, Z., Steel, J. & Lowen, A. C. Influenza Virus Reassortment Occurs with High Frequency in the Absence of Segment Mismatch. *PLoS Pathog* **9**, e1003421 (2013).
31. McDonald, S. M., Nelson, M. I., Turner, P. E. & Patton, J. T. Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nat Rev Microbiol* **14**, 448–460 (2016).
32. Walker, P. J. *et al.* Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). *Arch Virol* **166**, 2633–2648 (2021).
33. Bouvier, N. M. & Palese, P. THE BIOLOGY OF INFLUENZA VIRUSES. *Vaccine* **26**, D49–D53 (2008).
34. de Vries, E., Du, W., Guo, H. & de Haan, C. A. M. Influenza A Virus Hemagglutinin–Neuraminidase–Receptor Balance: Preserving Virus Motility. *Trends Microbiol* **28**, 57–67 (2020).
35. Kosik, I. & Yewdell, J. W. Influenza Hemagglutinin and Neuraminidase: Yin–Yang Proteins Coevolving to Thwart Immunity. *Viruses* **11**, 346 (2019).
36. A revision of the system of nomenclature for influenza viruses: a WHO Memorandum. *Bull World Health Organ* **58**, 585–591 (1980).
37. Melidou, A. *et al.* Influenza returns with a season dominated by clade 3C.2a1b.2a.2 A(H3N2) viruses, WHO European Region, 2021/22. *Euro Surveill* **27**, 2200255 (2022).
38. Reina, J. [The Victoria and Yamagata Lineages of Influenza B Viruses, unknown and undervalued]. *Rev Esp Quimioter* **35**, 231–235 (2022).
39. Tong, S. *et al.* A distinct lineage of influenza A virus from bats. *Proceedings of the National Academy of Sciences* **109**, 4269–4274 (2012).

40. Sederdahl, B. K. & Williams, J. V. Epidemiology and Clinical Characteristics of Influenza C Virus. *Viruses* **12**, 89 (2020).
41. Liu, R., Sheng, Z., Huang, C., Wang, D. & Li, F. Influenza D Virus. *Curr Opin Virol* **44**, 154–161 (2020).
42. Dedola, S. *et al.* Sialic acids in infection and their potential use in detection and protection against pathogens. *RSC Chemical Biology* **5**, 167–188 (2024).
43. Suzuki, Y. *et al.* Sialic Acid Species as a Determinant of the Host Range of Influenza A Viruses. *J Virol* **74**, 11825–11831 (2000).
44. Stencel-Baerenwald, J. E., Reiss, K., Reiter, D. M., Stehle, T. & Dermody, T. S. The sweet spot: defining virus–sialic acid interactions. *Nat Rev Microbiol* **12**, 739–749 (2014).
45. Jilani, T. N., Jamil, R. T., Nguyen, A. D. & Siddiqui, A. H. H1N1 Influenza. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2024).
46. Sreenivasan, C. C., Sheng, Z., Wang, D. & Li, F. Host Range, Biology, and Species Specificity of Seven-Segmented Influenza Viruses—A Comparative Review on Influenza C and D. *Pathogens* **10**, 1583 (2021).
47. Lee, D.-H., Criado, M. F. & Swayne, D. E. Pathobiological Origins and Evolutionary History of Highly Pathogenic Avian Influenza Viruses. *Cold Spring Harb Perspect Med* **11**, a038679 (2021).
48. Kanaujia, R. *et al.* Avian influenza revisited: concerns and constraints. *VirusDis.* **33**, 456–465 (2022).
49. Pacific, W. H. O. R. O. for the W. Avian Influenza Weekly Update 2024. (2024).
50. Yang, W., Schountz, T. & Ma, W. Bat Influenza Viruses: Current Status and Perspective. *Viruses* **13**, 547 (2021).
51. Voorhees, I. E. H. *et al.* Spread of Canine Influenza A(H3N2) Virus, United States. *Emerg Infect Dis* **23**, 1950–1957 (2017).
52. Sack, A. *et al.* Equine Influenza Virus—A Neglected, Reemergent Disease Threat. *Emerg Infect Dis* **25**, 1185–1191 (2019).
53. Tenforde, M. W. *et al.* Effect of Antigenic Drift on Influenza Vaccine Effectiveness in the United States—2019–2020. *Clinical Infectious Diseases* **73**, e4244–e4250 (2021).
54. Boni, M. F. Vaccination and antigenic drift in influenza. *Vaccine* **26**, C8–C14 (2008).
55. Nolan, T. *et al.* Efficacy of a Cell-Culture–Derived Quadrivalent Influenza Vaccine in Children. *New England Journal of Medicine* **385**, 1485–1495 (2021).

56. Horimoto, T. & Kawaoka, Y. Designing vaccines for pandemic influenza. *Curr Top Microbiol Immunol* **333**, 165–176 (2009).
57. Seladi-Schulman, J., Steel, J. & Lowen, A. C. Spherical Influenza Viruses Have a Fitness Advantage in Embryonated Eggs, while Filament-Producing Strains Are Selected In Vivo. *Journal of Virology* **87**, 13343–13353 (2013).
58. Chu, C. M., Dawson, I. M. & Elford, W. J. FILAMENTOUS FORMS ASSOCIATED WITH NEWLY ISOLATED INFLUENZA VIRUS. *The Lancet* **253**, e2 (1949).
59. Carter, T. & Iqbal, M. The Influenza A Virus Replication Cycle: A Comprehensive Review. *Viruses* **16**, 316 (2024).
60. Dadonaite, B., Vijayakrishnan, S., Fodor, E., Bhella, D. & Hutchinson, E. C. Filamentous Influenza Viruses. *J Gen Virol* **97**, 1755–1764 (2016).
61. Fontana, J., Cardone, G., Heymann, J. B., Winkler, D. C. & Steven, A. C. Structural Changes in Influenza Virus at Low pH Characterized by Cryo-Electron Tomography. *Journal of Virology* **86**, 2919–2929 (2012).
62. Eisfeld, A. J., Neumann, G. & Kawaoka, Y. At the centre: influenza A virus ribonucleoproteins. *Nat Rev Microbiol* **13**, 28–41 (2015).
63. Ortega, J. *et al.* Ultrastructural and Functional Analyses of Recombinant Influenza Virus Ribonucleoproteins Suggest Dimerization of Nucleoprotein during Virus Amplification. *Journal of Virology* **74**, 156–163 (2000).
64. 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 U S A* **84**, 8140–8144 (1987).
65. 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).
66. Area, E. *et al.* 3D structure of the influenza virus polymerase complex: Localization of subunit domains. *Proceedings of the National Academy of Sciences* **101**, 308–313 (2004).
67. Murti, K. G., Webster, R. G. & Jones, I. M. Localization of RNA polymerases on influenza viral ribonucleoproteins by immunogold labeling. *Virology* **164**, 562–566 (1988).
68. Nakano, M. *et al.* Ultrastructure of influenza virus ribonucleoprotein complexes during viral RNA synthesis. *Commun Biol* **4**, 1–10 (2021).
69. Nilsson-Payant, B. E. *et al.* Reduced Nucleoprotein Availability Impairs Negative-Sense RNA Virus Replication and Promotes Host Recognition. *J Virol* **95**, e02274-20 (2021).

70. Wilson, I. A., Skehel, J. J. & Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373 (1981).
71. McCauley, J. *et al.* Influenza virus haemagglutinin signal sequence. *FEBS Letters* **108**, 422–426 (1979).
72. Boulay, F., Doms, R. W., Webster, R. G. & Helenius, A. Posttranslational oligomerization and cooperative acid activation of mixed influenza hemagglutinin trimers. *Journal of Cell Biology* **106**, 629–639 (1988).
73. Chandrasekaran, A. *et al.* Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat Biotechnol* **26**, 107–113 (2008).
74. Rust, M. J., Lakadamyali, M., Zhang, F. & Zhuang, X. Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nat Struct Mol Biol* **11**, 567–573 (2004).
75. Rossman, J. S., Leser, G. P. & Lamb, R. A. Filamentous influenza virus enters cells via macropinocytosis. *J Virol* **86**, 10950–10960 (2012).
76. Pinto, L. H., Holsinger, L. J. & Lamb, R. A. Influenza virus M2 protein has ion channel activity. *Cell* **69**, 517–528 (1992).
77. Lakdawala, S. S. *et al.* Influenza A Virus Assembly Intermediates Fuse in the Cytoplasm. *PLoS Pathogens* **10**, e1003971 (2014).
78. Bui, M., Whittaker, G. & Helenius, A. Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. *J Virol* **70**, 8391–8401 (1996).
79. Chou, Y. *et al.* Colocalization of different influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS Pathog* **9**, e1003358 (2013).
80. Oka, M. & Yoneda, Y. Importin α : functions as a nuclear transport factor and beyond. *Proc Jpn Acad Ser B Phys Biol Sci* **94**, 259–274 (2018).
81. Pflug, A. *et al.* Capped RNA primer binding to influenza polymerase and implications for the mechanism of cap-binding inhibitors. *Nucleic Acids Res* **46**, 956–971 (2018).
82. Carter, T. & Iqbal, M. The Influenza A Virus Replication Cycle: A Comprehensive Review. *Viruses* **16**, 316 (2024).
83. Krischuns, T. *et al.* The host RNA polymerase II C-terminal domain is the anchor for replication of the influenza virus genome. *Nat Commun* **15**, 1064 (2024).

84. De Vlught, C., Sikora, D. & Pelchat, M. Insight into Influenza: A Virus Cap-Snatching. *Viruses* **10**, 641 (2018).
85. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**, 847–858 (1981).
86. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* **15**, 500–506 (2008).
87. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**, 355–360 (2014).
88. Thierry, E. *et al.* Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains. *Mol Cell* **61**, 125–137 (2016).
89. Reich, S. *et al.* Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**, 361–366 (2014).
90. Fodor, E. *et al.* A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J Virol* **76**, 8989–9001 (2002).
91. Pritlove, D. C., Poon, L. L. M., Devenish, L. J., Leahy, M. B. & Brownlee, G. G. A Hairpin Loop at the 5' End of Influenza A Virus Virion RNA Is Required for Synthesis of Poly(A)⁺ mRNA In Vitro. *J Virol* **73**, 2109–2114 (1999).
92. Kouba, T., Drncová, P. & Cusack, S. Structural snapshots of actively transcribing influenza polymerase. *Nat Struct Mol Biol* **26**, 460–470 (2019).
93. Wandzik, J. M., Kouba, T. & Cusack, S. Structure and Function of Influenza Polymerase. *Cold Spring Harb Perspect Med* **11**, a038372 (2021).
94. Wandzik, J. M. *et al.* A Structure-Based Model for the Complete Transcription Cycle of Influenza Polymerase. *Cell* **181**, 877–893.e21 (2020).
95. Poon, L. L. M., Pritlove, D. C., Fodor, E. & Brownlee, G. G. Direct Evidence that the Poly(A) Tail of Influenza A Virus mRNA Is Synthesized by Reiterative Copying of a U Track in the Virion RNA Template. *J Virol* **73**, 3473–3476 (1999).
96. Pritlove, D. C., Poon, L. L. M., Fodor, E., Sharps, J. & Brownlee, G. G. Polyadenylation of Influenza Virus mRNA Transcribed In Vitro from Model Virion RNA Templates: Requirement for 5' Conserved Sequences. *J Virol* **72**, 1280–1286 (1998).

97. Huang, X. *et al.* An NS-segment exonic splicing enhancer regulates influenza A virus replication in mammalian cells. *Nat Commun* **8**, 14751 (2017).
98. Lamb, R. A. & Choppin, P. W. Identification of a second protein (M2) encoded by RNA segment 7 of influenza virus. *Virology* **112**, 729–737 (1981).
99. Lamb, R. A. & Lai, C.-J. Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. *Cell* **21**, 475–485 (1980).
100. De Magistris, P. The Great Escape: mRNA Export through the Nuclear Pore Complex. *International Journal of Molecular Sciences* **22**, 11767 (2021).
101. Dou, D., Revol, R., Östbye, H., Wang, H. & Daniels, R. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Front. Immunol.* **9**, (2018).
102. Mikulášová, A., Varecková, E. & Fodor, E. Transcription and replication of the influenza a virus genome. *Acta Virol* **44**, 273–282 (2000).
103. Hay, A. J., Lomniczi, B., Bellamy, A. R. & Skehel, J. J. Transcription of the influenza virus genome. *Virology* **83**, 337–355 (1977).
104. Vreede, F. T., Jung, T. E. & Brownlee, G. G. Model Suggesting that Replication of Influenza Virus Is Regulated by Stabilization of Replicative Intermediates. *J Virol* **78**, 9568–9572 (2004).
105. Hay, A. J., Skehel, J. J. & McCauley, J. Characterization of influenza virus RNA complete transcripts. *Virology* **116**, 517–522 (1982).
106. Alenquer, M. *et al.* Influenza A virus ribonucleoproteins form liquid organelles at endoplasmic reticulum exit sites. *Nat Commun* **10**, 1629 (2019).
107. Amorim, M. J. *et al.* A Rab11- and microtubule-dependent mechanism for cytoplasmic transport of influenza A virus viral RNA. *J Virol* **85**, 4143–4156 (2011).
108. Huang, S. *et al.* A second CRM1-dependent nuclear export signal in the influenza A virus NS2 protein contributes to the nuclear export of viral ribonucleoproteins. *J Virol* **87**, 767–778 (2013).
109. Vale-Costa, S. *et al.* ATG9A regulates the dissociation of recycling endosomes from microtubules to form liquid influenza A virus inclusions. *PLoS Biol* **21**, e3002290 (2023).
110. Carrique, L. *et al.* Host ANP32A mediates the assembly of the influenza virus replicase. *Nature* **587**, 638–643 (2020).

111. Zhu, Z., Fodor, E. & Keown, J. R. A structural understanding of influenza virus genome replication. *Trends in Microbiology* **31**, 308–319 (2023).
112. Vreede, F. T., Gifford, H. & Brownlee, G. G. Role of Initiating Nucleoside Triphosphate Concentrations in the Regulation of Influenza Virus Replication and Transcription. *J Virol* **82**, 6902–6910 (2008).
113. Honda, A., Mizumoto, K. & Ishihama, A. Identification of the 5' terminal structure of influenza virus genome RNA by a newly developed enzymatic method. *Virus Res* **55**, 199–206 (1998).
114. te Velthuis, A. J. W., Robb, N. C., Kapanidis, A. N. & Fodor, E. The role of the priming loop in influenza A virus RNA synthesis. *Nat Microbiol* **1**, 1–7 (2016).
115. Prokudina-Kantorovich, E. N. & Semenova, N. P. Intracellular Oligomerization of Influenza Virus Nucleoprotein. *Virology* **223**, 51–56 (1996).
116. York, A., Hengrung, N., Vreede, F. T., Huiskonen, J. T. & Fodor, E. Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proceedings of the National Academy of Sciences* **110**, E4238–E4245 (2013).
117. Deng, T., Vreede, F. T. & Brownlee, G. G. Different de novo initiation strategies are used by influenza virus RNA polymerase on its cRNA and viral RNA promoters during viral RNA replication. *J Virol* **80**, 2337–2348 (2006).
118. Resa-Infante, P., Jorba, N., Coloma, R. & Ortin, J. The influenza virus RNA synthesis machine. *RNA Biology* **8**, 207–215 (2011).
119. Elton, D. *et al.* Interaction of the Influenza Virus Nucleoprotein with the Cellular CRM1-Mediated Nuclear Export Pathway. *J Virol* **75**, 408–419 (2001).
120. Shimizu, T., Takizawa, N., Watanabe, K., Nagata, K. & Kobayashi, N. Crucial role of the influenza virus NS2 (NEP) C-terminal domain in M1 binding and nuclear export of vRNP. *FEBS Lett* **585**, 41–46 (2011).
121. Brunotte, L. *et al.* The nuclear export protein of H5N1 influenza A viruses recruits Matrix 1 (M1) protein to the viral ribonucleoprotein to mediate nuclear export. *J Biol Chem* **289**, 20067–20077 (2014).
122. Ullrich, O., Reinsch, S., Urbé, S., Zerial, M. & Parton, R. G. Rab11 regulates recycling through the pericentriolar recycling endosome. *J Cell Biol* **135**, 913–924 (1996).
123. Guichard, A., Nizet, V. & Bier, E. Rab11-mediated trafficking in host-pathogen interactions. *Nat Rev Microbiol* **12**, 624–634 (2014).

124. Veler, H. *et al.* The C-Terminal Domains of the PB2 Subunit of the Influenza A Virus RNA Polymerase Directly Interact with Cellular GTPase Rab11a. *Journal of Virology* **96**, e01979-21 (2022).
125. de Castro Martin, I. F. *et al.* Influenza virus genome reaches the plasma membrane via a modified endoplasmic reticulum and Rab11-dependent vesicles. *Nat Commun* **8**, 1396 (2017).
126. Dolnik, O., Gerresheim, G. K. & Biedenkopf, N. New Perspectives on the Biogenesis of Viral Inclusion Bodies in Negative-Sense RNA Virus Infections. *Cells* **10**, 1460 (2021).
127. Schmidt, N. W., Mishra, A., Wang, J., DeGrado, W. F. & Wong, G. C. L. Influenza Virus A M2 Protein Generates Negative Gaussian Membrane Curvature Necessary for Budding and Scission. *J. Am. Chem. Soc.* **135**, 13710–13719 (2013).
128. Nayak, D., Shivakoti, S., Balogun, R. A., Lee, G. & Zhou, Z. H. Structure, disassembly, assembly, and budding of influenza viruses. in *Textbook of Influenza* 35–56 (John Wiley & Sons, Ltd, 2013). doi:10.1002/9781118636817.ch3.
129. Di Mattia, J. *et al.* Nonconcomitant host-to-host transmission of multipartite virus genome segments may lead to complete genome reconstitution. *Proceedings of the National Academy of Sciences* **119**, e2201453119 (2022).
130. Bonnamy, M., Brousse, A., Pirolles, E., Michalakakis, Y. & Blanc, S. The genome formula of a multipartite virus is regulated both at the individual segment and the segment group levels. *PLOS Pathogens* **20**, e1011973 (2024).
131. Perlmutter, J. D., Qiao, C. & Hagan, M. F. Viral genome structures are optimal for capsid assembly. *eLife* **2**, e00632 (2013).
132. Lin, H.-K., Schoot, P. van der & Zandi, R. Impact of charge variation on the encapsulation of nanoparticles by virus coat proteins. *Phys. Biol.* **9**, 066004 (2012).
133. Belyi, V. A. & Muthukumar, M. Electrostatic origin of the genome packing in viruses. *Proceedings of the National Academy of Sciences* **103**, 17174–17178 (2006).
134. Buzón, P., Maity, S. & Roos, W. H. Physical virology: From virus self-assembly to particle mechanics. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **12**, e1613 (2020).
135. Bruinsma, R. F., Wuite, G. J. L. & Roos, W. H. Physics of viral dynamics. *Nat Rev Phys* **3**, 76–91 (2021).
136. Roos, W. H., Bruinsma, R. & Wuite, G. J. L. Physical virology. *Nature Phys* **6**, 733–743 (2010).

137. Ruigrok, R. W. H., Crépin, T. & Kolakofsky, D. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol* **14**, 504–510 (2011).
138. Luo, M., Terrell, J. R. & Mcmanus, S. A. Nucleocapsid Structure of Negative Strand RNA Virus. *Viruses* **12**, 835 (2020).
139. Fournier, E. *et al.* A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res* **40**, 2197–2209 (2012).
140. Perlmutter, J. D. & Hagan, M. F. The Role of Packaging Sites in Efficient and Specific Virus Assembly. *Journal of Molecular Biology* **427**, 2451–2467 (2015).
141. Zlotnick, A. Are weak protein-protein interactions the general rule in capsid assembly? *Virology* **315**, 269–274 (2003).
142. Garmann, R. F., Goldfain, A. M. & Manoharan, V. N. Measurements of the self-assembly kinetics of individual viral capsids around their RNA genome. *Proc Natl Acad Sci U S A* **116**, 22485–22490 (2019).
143. Borodavka, A., Dykeman, E. C., Schrimpf, W. & Lamb, D. C. Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. *eLife* **6**, e27453 (2017).
144. Borodavka, A., Dykeman, E. C., Schrimpf, W. & Lamb, D. C. Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. *Elife* **6**, e27453 (2017).
145. Sung, P. *et al.* Recruitment of multi-segment genomic RNAs by Bluetongue virus requires a preformed RNA network. *Nucleic Acids Research* gkae404 (2024) doi:10.1093/nar/gkae404.
146. Fajardo, T., Sung, P.-Y. & Roy, P. Disruption of Specific RNA-RNA Interactions in a Double-Stranded RNA Virus Inhibits Genome Packaging and Virus Infectivity. *PLoS Pathog* **11**, e1005321 (2015).
147. Xia, X. *et al.* RNA genome packaging and capsid assembly of bluetongue virus visualized in host cells. *Cell* **187**, 2236-2249.e17 (2024).
148. Sung, P.-Y. *et al.* A multidisciplinary approach to the identification of the protein-RNA connectome in double-stranded RNA virus capsids. *Nucleic Acids Res* **51**, 5210–5227 (2023).
149. Dall’Ara, M., Ratti, C., Bouzoubaa, S. E. & Gilmer, D. Ins and Outs of Multipartite Positive-Strand RNA Plant Viruses: Packaging versus Systemic Spread. *Viruses* **8**, 228 (2016).

150. Choi, Y. G. & Rao, A. L. N. Packaging of Brome Mosaic Virus RNA3 Is Mediated through a Bipartite Signal. *J Virol* **77**, 9750–9757 (2003).
151. Gottlieb, P., Qiao, X., Strassman, J., Frilander, M. & Mindich, L. Identification of the Packaging Regions within the Genomic RNA Segments of Bacteriophage Φ 6. *Virology* **200**, 42–47 (1994).
152. Barciszewska, M. Z., Perrigue, P. M. & Barciszewski, J. tRNA – the golden standard in molecular biology. *Mol. BioSyst.* **12**, 12–17 (2015).
153. Alberts, B. *et al.* The RNA World and the Origins of Life. in *Molecular Biology of the Cell. 4th edition* (Garland Science, 2002).
154. Allan, M. F., Brivanlou, A. & Rouskin, S. RNA levers and switches controlling viral gene expression. *Trends in Biochemical Sciences* **48**, 391–406 (2023).
155. Fairman, C. W., Lever, A. M. L. & Kenyon, J. C. Evaluating RNA Structural Flexibility: Viruses Lead the Way. *Viruses* **13**, 2130 (2021).
156. Moss, W. N. *et al.* The 3' Splice Site of Influenza A Segment 7 mRNA Can Exist in Two Conformations: A Pseudoknot and a Hairpin. *PLOS ONE* **7**, e38323 (2012).
157. Cheong, H.-K., Cheong, C. & Choi, B.-S. Secondary Structure of the Panhandle RNA of Influenza Virus A Studied by NMR Spectroscopy. *Nucleic Acids Research* **24**, 4197–4201 (1996).
158. Assmann, S. M., Chou, H.-L. & Bevilacqua, P. C. Rock, scissors, paper: How RNA structure informs function. *Plant Cell* **35**, 1671–1707 (2023).
159. Stombaugh, J., Zirbel, C. L., Westhof, E. & Leontis, N. B. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res* **37**, 2294–2312 (2009).
160. Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512 (2001).
161. Leontis, N. B., Stombaugh, J. & Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* **30**, 3497–3531 (2002).
162. Vicens, Q. & Kieft, J. S. Thoughts on how to think (and talk) about RNA structure. *Proceedings of the National Academy of Sciences* **119**, e2112677119 (2022).
163. Grabow, W. W. & Jaeger, L. RNA Self-Assembly and RNA Nanotechnology. *Acc. Chem. Res.* **47**, 1871–1880 (2014).
164. Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nat Rev Genet* **15**, 469–479 (2014).

165. Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B. & Ehresmann, C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc Natl Acad Sci U S A* **91**, 4945–4949 (1994).
166. Paillart, J. C., Skripkin, E., Ehresmann, B., Ehresmann, C. & Marquet, R. A loop-loop ‘kissing’ complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A* **93**, 5572–5577 (1996).
167. 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).
168. Lorenz, R., Wolfinger, M. T., Tanzer, A. & Hofacker, I. L. Predicting RNA secondary structures from sequence and probing data. *Methods* **103**, 86–98 (2016).
169. Gallagher, J. R., Torian, U., McCraw, D. M. & Harris, A. K. Structural studies of influenza virus RNPs by electron microscopy indicate molecular contortions within NP supra-structures. *J Struct Biol* **197**, 294–307 (2017).
170. Fan, H. *et al.* Structures of influenza A virus RNA polymerase offer insight into viral genome replication. *Nature* **573**, 287–290 (2019).
171. Strobel, E. J., Yu, A. M. & Lucks, J. B. High-throughput determination of RNA structures. *Nat Rev Genet* **19**, 615–634 (2018).
172. Sato-Asano, K. & Egami, F. Ribonucleases in Takadiastase. *Nature* **185**, 462–463 (1960).
173. Gilmer, O. *et al.* Chemical and Enzymatic Probing of Viral RNAs: From Infancy to Maturity and Beyond. *Viruses* **13**, 1894 (2021).
174. Ehresmann, C. *et al.* Probing the structure of RNAs in solution. *Nucleic Acids Res* **15**, 9109–9128 (1987).
175. Isel, C. *et al.* Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J* **18**, 1038–1048 (1999).
176. Wan, Y., Qu, K., Ouyang, Z. & Chang, H. Y. Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat Protoc* **8**, 849–869 (2013).
177. Wan, Y. *et al.* Genome-wide Measurement of RNA Folding Energies. *Molecular Cell* **48**, 169–181 (2012).
178. Underwood, J. G. *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat Methods* **7**, 995–1001 (2010).

179. Incarnato, D., Neri, F., Anselmi, F. & Oliviero, S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol* **15**, 491 (2014).
180. Kobschull, J. M. *et al.* High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron* **91**, 975–987 (2016).
181. Hector, R. D. *et al.* Snapshots of pre-rRNA structural flexibility reveal eukaryotic 40S assembly dynamics at nucleotide resolution. *Nucleic Acids Res* **42**, 12138–12154 (2014).
182. Kielbinski, L. J. & Vinther, J. Massive parallel-sequencing-based hydroxyl radical probing of RNA accessibility. *Nucleic Acids Res* **42**, e70 (2014).
183. Spitale, R. C., Flynn, R. A., Torre, E. A., Kool, E. T. & Chang, H. Y. RNA structural analysis by evolving SHAPE chemistry. *WIREs RNA* **5**, 867–881 (2014).
184. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).
185. Velema, W. A. & Kool, E. T. The chemistry and applications of RNA 2'-OH acylation. *Nat Rev Chem* **4**, 22–37 (2020).
186. Leonard, C. W. *et al.* Principles for Understanding the Accuracy of SHAPE-Directed RNA Structure Modeling. *Biochemistry* **52**, 588–595 (2013).
187. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127**, 4223–4231 (2005).
188. Rice, G. M., Leonard, C. W. & Weeks, K. M. RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA* **20**, 846–854 (2014).
189. Boerneke, M. A., Ehrhardt, J. E. & Weeks, K. M. Physical and Functional Analysis of Viral RNA Genomes by SHAPE. *Annual Review of Virology* **6**, 93–117 (2019).
190. Homan, P. J. *et al.* Single-molecule correlated chemical probing of RNA. *Proc Natl Acad Sci U S A* **111**, 13858–13863 (2014).
191. Krokhotin, A., Mustoe, A. M., Weeks, K. M. & Dokholyan, N. V. Direct identification of base-paired RNA nucleotides by correlated chemical probing. *RNA* **23**, 6–13 (2017).
192. Aviran, S. & Pachter, L. Rational experiment design for sequencing-based RNA structure mapping. *RNA* **20**, 1864–1877 (2014).

193. Hagey, R. J. *et al.* Programmable antivirals targeting critical conserved viral RNA secondary structures from influenza A virus and SARS-CoV-2. *Nat Med* **28**, 1944–1955 (2022).
194. Lindgreen, S., Gardner, P. P. & Krogh, A. Measuring covariation in RNA alignments: physical realism improves information measures. *Bioinformatics* **22**, 2988–2995 (2006).
195. Michel, F. & Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* **216**, 585–610 (1990).
196. Rivas, E., Clements, J. & Eddy, S. R. Estimating the power of sequence covariation for detecting conserved RNA structure. *Bioinformatics* **36**, 3072–3076 (2020).
197. Gog, J. R. *et al.* Codon conservation in the influenza A virus genome defines RNA packaging signals. *Nucleic Acids Research* **35**, 1897–1907 (2007).
198. Moss, W. N., Priore, S. F. & Turner, D. H. Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA* **17**, 991–1011 (2011).
199. Huber, R. G. *et al.* Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat Commun* **10**, 1408 (2019).
200. Yang, S. L., Ponti, R. D., Wan, Y. & Huber, R. G. Computational and Experimental Approaches to Study the RNA Secondary Structures of RNA Viruses. *Viruses* **14**, 1795 (2022).
201. Wang, Y. *et al.* Stable stem enabled Shannon entropies distinguish non-coding RNAs from random backgrounds. *BMC Bioinformatics* **13**, S1 (2012).
202. Martin, J. S. Describing the Structural Diversity within an RNA's Ensemble. *Entropy* **16**, 1331–1348 (2014).
203. Dethoff, E. A. *et al.* Pervasive tertiary structure in the dengue virus RNA genome. *Proceedings of the National Academy of Sciences* **115**, 11513–11518 (2018).
204. Watters, K. E. *et al.* Probing of RNA structures in a positive sense RNA virus reveals selection pressures for structural elements. *Nucleic Acids Res* **46**, 2573–2584 (2018).
205. Dadonaite, B. *et al.* The structure of the influenza A virus genome. *Nat Microbiol* **4**, 1781–1789 (2019).
206. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

207. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
208. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Research* **31**, 3429–3431 (2003).
209. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences* **101**, 7287–7292 (2004).
210. Zuker, M. & Stiegler, P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* **9**, 133–148 (1981).
211. Mathews, D. H., Moss, W. N. & Turner, D. H. Folding and Finding RNA Secondary Structure. *Cold Spring Harb Perspect Biol* **2**, a003665 (2010).
212. Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R. & Stadler, P. F. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9**, 474 (2008).
213. Harmanci, A. O., Sharma, G. & Mathews, D. H. TurboFold: iterative probabilistic estimation of secondary structures for multiple RNA sequences. *BMC Bioinformatics* **12**, 108 (2011).
214. Fu, Y., Sharma, G. & Mathews, D. H. Dynalign II: common secondary structure prediction for RNA homologs with domain insertions. *Nucleic Acids Res* **42**, 13939–13948 (2014).
215. Xu, Z. & Mathews, D. H. Multalign: an algorithm to predict secondary structures conserved in multiple RNA sequences. *Bioinformatics* **27**, 626–632 (2011).
216. Torarinsson, E., Havgaard, J. H. & Gorodkin, J. Multiple structural alignment and clustering of RNA sequences. *Bioinformatics* **23**, 926–932 (2007).
217. Yesselman, J. D. *et al.* Updates to the RNA mapping database (RMDB), version 2. *Nucleic Acids Res* **46**, D375–D379 (2018).
218. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* **11**, 959–965 (2014).
219. Desirò, D. M. M. The Complexity of Packaging Mechanisms in Segmented RNA Viruses.
220. Lorenz, R., Hofacker, I. L. & Stadler, P. F. RNA folding with hard and soft constraints. *Algorithms for Molecular Biology* **11**, 8 (2016).
221. Hajdin, C. E. *et al.* Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc Natl Acad Sci U S A* **110**, 5498–5503 (2013).

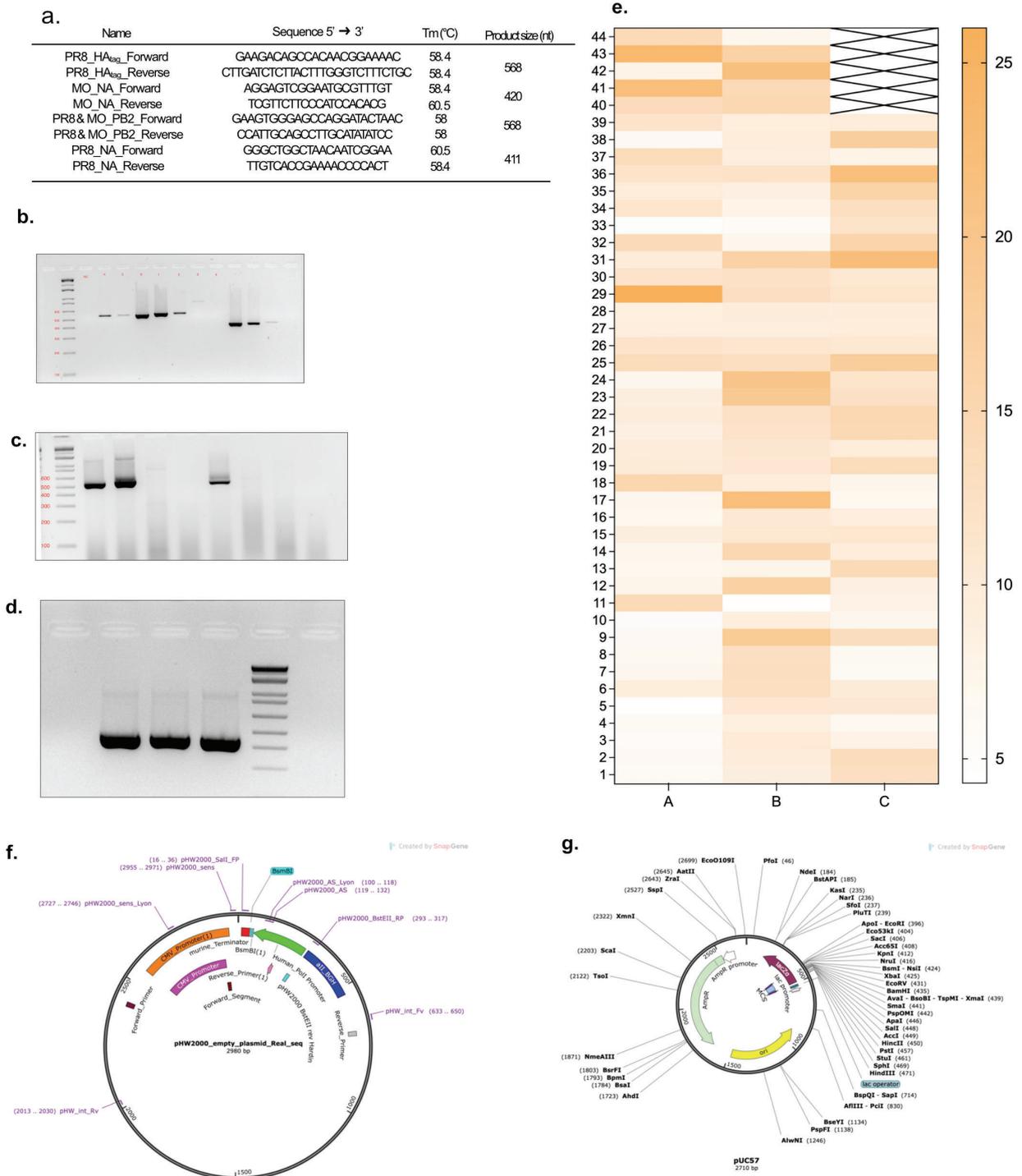
222. Seetin, M. G. & Mathews, D. H. RNA structure prediction: an overview of methods. *Methods Mol Biol* **905**, 99–122 (2012).
223. Wu, Y. *et al.* Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic Acids Res* **43**, 7247–7259 (2015).
224. Spitale, R. C. & Incarnato, D. Probing the dynamic RNA structurome and its functions. *Nat Rev Genet* **24**, 178–196 (2023).
225. Smith, M. A., Gesell, T., Stadler, P. F. & Mattick, J. S. Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Research* **41**, 8220–8236 (2013).
226. Wu, M. T.-P. & D’Souza, V. Alternate RNA Structures. *Cold Spring Harb Perspect Biol* **12**, a032425 (2020).
227. Morandi, E. *et al.* Genome-scale deconvolution of RNA structure ensembles. (2021) doi:10.1038/s41592-021-01075-w.
228. Bohn, P., Gribling-Burrer, A.-S., Ambi, U. B. & Smyth, R. P. Nano-DMS-MaP allows isoform-specific RNA structure determination. *Nat Methods* **20**, 849–859 (2023).
229. Ennifar, E. *et al.* Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell. *Nucleic Acids Res* **34**, 2328–2339 (2006).
230. Sengupta, A., Rice, G. M. & Weeks, K. M. Single-molecule correlated chemical probing reveals large-scale structural communication in the ribosome and the mechanism of the antibiotic spectinomycin in living cells. *PLoS Biol* **17**, e3000393 (2019).
231. Zeller, M. J. *et al.* SHAPE-enabled fragment-based ligand discovery for RNA. *Proc Natl Acad Sci U S A* **119**, e2122660119 (2022).
232. Martin, S., Blankenship, C., Rausch, J. W. & Sztuba-Solinska, J. Using SHAPE-MaP to probe small molecule-RNA interactions. *Methods* **167**, 105–116 (2019).
233. Zhu, Y., Zhu, L., Wang, X. & Jin, H. RNA-based therapeutics: an overview and prospectus. *Cell Death Dis* **13**, 1–15 (2022).
234. Sparmann, A. & Vogel, J. RNA-based medicine: from molecular mechanisms to therapy. *The EMBO Journal* **42**, e114760 (2023).
235. Quignon, E. *et al.* Structural Impact of the Interaction of the Influenza A Virus Nucleoprotein with Genomic RNA Segments. *Viruses* **16**, 421 (2024).
236. Girard, J. *et al.* Disruption of influenza virus packaging signals results in various misassembled genome complexes. *J Virol* **97**, e01076-23 (2023).

237. Bolte, H., Rosu, M. E., Hagelauer, E., García-Sastre, A. & Schwemmler, M. Packaging of the Influenza Virus Genome Is Governed by a Plastic Network of RNA- and Nucleoprotein-Mediated Interactions. *J Virol* **93**, e01861-18 (2019).
238. Irving, P. S. & Weeks, K. M. RNAVigate: efficient exploration of RNA chemical probing datasets. *Nucleic Acids Research* **52**, 2231–2241 (2024).
239. Smola, M. J. & Weeks, K. M. In-cell RNA structure probing with SHAPE-MaP. *Nat Protoc* **13**, 1181–1195 (2018).
240. Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G. & Webster, R. G. A DNA transfection system for generation of influenza A virus from eight plasmids. *Proceedings of the National Academy of Sciences* **97**, 6108–6113 (2000).
241. Neumann, G. *et al.* Generation of influenza A viruses entirely from cloned cDNAs. *Proc Natl Acad Sci U S A* **96**, 9345–9350 (1999).
242. Chua, S. C. J. H., Tan, H. Q., Engelberg, D. & Lim, L. H. K. Alternative Experimental Models for Studying Influenza Proteins, Host–Virus Interactions and Anti-Influenza Drugs. *Pharmaceuticals* **12**, 147 (2019).
243. 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).
244. Hatta, M., Halfmann, P., Wells, K. & Kawaoka, Y. Human Influenza A Viral Genes Responsible for the Restriction of Its Replication in Duck Intestine. *Virology* **295**, 250–255 (2002).
245. Honigsbaum, M. Revisiting the 1957 and 1968 influenza pandemics. *The Lancet* **395**, 1824–1826 (2020).
246. White, M. C., Steel, J. & Lowen, A. C. Heterologous Packaging Signals on Segment 4, but Not Segment 6 or Segment 8, Limit Influenza A Virus Reassortment. *J Virol* **91**, e00195-17 (2017).
247. Wu, M. *et al.* Defective Interfering Particles of Influenza Virus and Their Characteristics, Impacts, and Use in Vaccines and Antiviral Strategies: A Systematic Review. *Viruses* **14**, 2773 (2022).
248. Mehle, A. & Doudna, J. A. Adaptive strategies of the influenza virus polymerase for replication in humans. *Proc Natl Acad Sci U S A* **106**, 21312–21316 (2009).

249. te Velthuis, A. J. W., Long, J. S. & Barclay, W. S. Assays to Measure the Activity of Influenza Virus Polymerase. in *Influenza Virus: Methods and Protocols* (ed. Yamauchi, Y.) 343–374 (Springer, New York, NY, 2018). doi:10.1007/978-1-4939-8678-1_17.
250. Chen, K.-Y., Santos Afonso, E. D., Enouf, V., Isel, C. & Naffakh, N. Influenza virus polymerase subunits co-evolve to ensure proper levels of dimerization of the heterotrimer. *PLoS Pathog* **15**, e1008034 (2019).
251. Campbell, P. J. *et al.* The M segment of the 2009 pandemic influenza virus confers increased neuraminidase activity, filamentous morphology, and efficient contact transmissibility to A/Puerto Rico/8/1934-based reassortant viruses. *J Virol* **88**, 3802–3814 (2014).
252. Phipps, K. L. *et al.* Collective interactions augment influenza A virus replication in a host-dependent manner. *Nat Microbiol* **5**, 1158–1169 (2020).
253. Goh, K. C. M. *et al.* Molecular determinants of plaque size as an indicator of dengue virus attenuation. *Sci Rep* **6**, 26100 (2016).
254. Gao, Q., Lowen, A. C., Wang, T. T. & Palese, P. A Nine-Segment Influenza A Virus Carrying Subtype H1 and H3 Hemagglutinins. *J Virol* **84**, 8062–8071 (2010).
255. Gulyaev, A. P. *et al.* Subtype-specific structural constraints in the evolution of influenza A virus hemagglutinin genes. *Sci Rep* **6**, 38892 (2016).
256. Vester, D. *et al.* Real-time RT-qPCR assay for the analysis of human influenza A virus transcription and replication dynamics. *Journal of Virological Methods* **168**, 63–71 (2010).
257. Cowell, T. W. *et al.* Unbiased, Cell-free Profiling of Single Influenza Genomes at High-throughput. Preprint at <https://doi.org/10.1101/2024.02.03.578479> (2024).
258. Wagner, R., Matrosovich, M. & Klenk, H.-D. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev Med Virol* **12**, 159–166 (2002).
259. Yamanaka, K., Ishihama, A. & Nagata, K. Reconstitution of influenza virus RNA-nucleoprotein complexes structurally resembling native viral ribonucleoprotein cores. *Journal of Biological Chemistry* **265**, 11151–11155 (1990).
260. Busan, S. & Weeks, K. M. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA* **24**, 143–148 (2018).
261. Smola, M. J., Calabrese, J. M. & Weeks, K. M. Detection of RNA–Protein Interactions in Living Cells with SHAPE. *Biochemistry* **54**, 6867–6875 (2015).

262. Gilmer, O. *et al.* Structural maturation of the HIV-1 RNA 5' untranslated region by Pr55Gag and its maturation products. *RNA Biology* **19**, 191–205 (2022).
263. Archer, E. J. *et al.* Long-range architecture in a viral RNA genome. *Biochemistry* **52**, 3182–3190 (2013).
264. Chenavier, F. *et al.* Cryo-EM structure of influenza helical nucleocapsid reveals NP-NP and NP-RNA interactions as a model for the genome encapsidation. *Sci. Adv.* **9**, eadj9974 (2023).
265. Tang, Y.-S., Xu, S., Chen, Y.-W., Wang, J.-H. & Shaw, P.-C. Crystal structures of influenza nucleoprotein complexed with nucleic acid provide insights into the mechanism of RNA interaction. *Nucleic Acids Res* **49**, 4144–4154 (2021).
266. Baudin, F., Bach, C., Cusack, S. & Ruigrok, R. W. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *The EMBO Journal* **13**, 3158–3165 (1994).
267. Slater, A. *et al.* Visualising Viruses. *J Gen Virol* **103**, 001730 (2022).
268. Lodmell, J. S., Ehresmann, C., Ehresmann, B. & Marquet, R. Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site. *RNA* **6**, 1267–1276 (2000).
269. Bindewald, E. *et al.* Correlating SHAPE signatures with three-dimensional RNA structures. *RNA* **17**, 1688–1696 (2011).
270. Mortimer, S. A. & Weeks, K. M. A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. *J. Am. Chem. Soc.* **129**, 4144–4145 (2007).
271. Busan, S., Weidmann, C. A., Sengupta, A. & Weeks, K. M. Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry* **58**, 2655–2664 (2019).
272. Wells, S. E., Hughes, J. M. X., Haller Igel, A. & Ares, M. [32] Use of dimethyl sulfate to probe RNA structure *in vivo*. in *Methods in Enzymology* vol. 318 479–493 (Academic Press, 2000).
273. 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 Research* **41**, 1241–1254 (2013).

Appendices



Supplementary Fig 1 | 7:2 reverse genetics and coinfection assays

a. Primers used for segment-specific amplification.

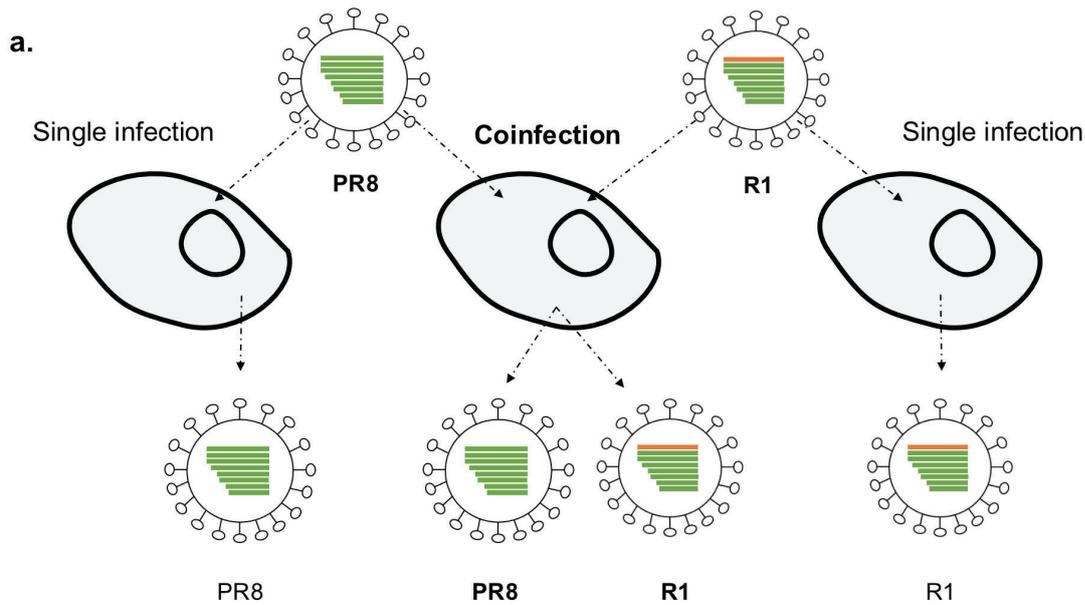
b. Primer optimisation for coinfection condition A of PR8 and PR8_HA_{tag}.

c. Primer optimisation of PR8 and MO_NA primers for coinfection condition B. The PR8_NA primer pair showed non-specific amplification of MO_NA (lane 5).

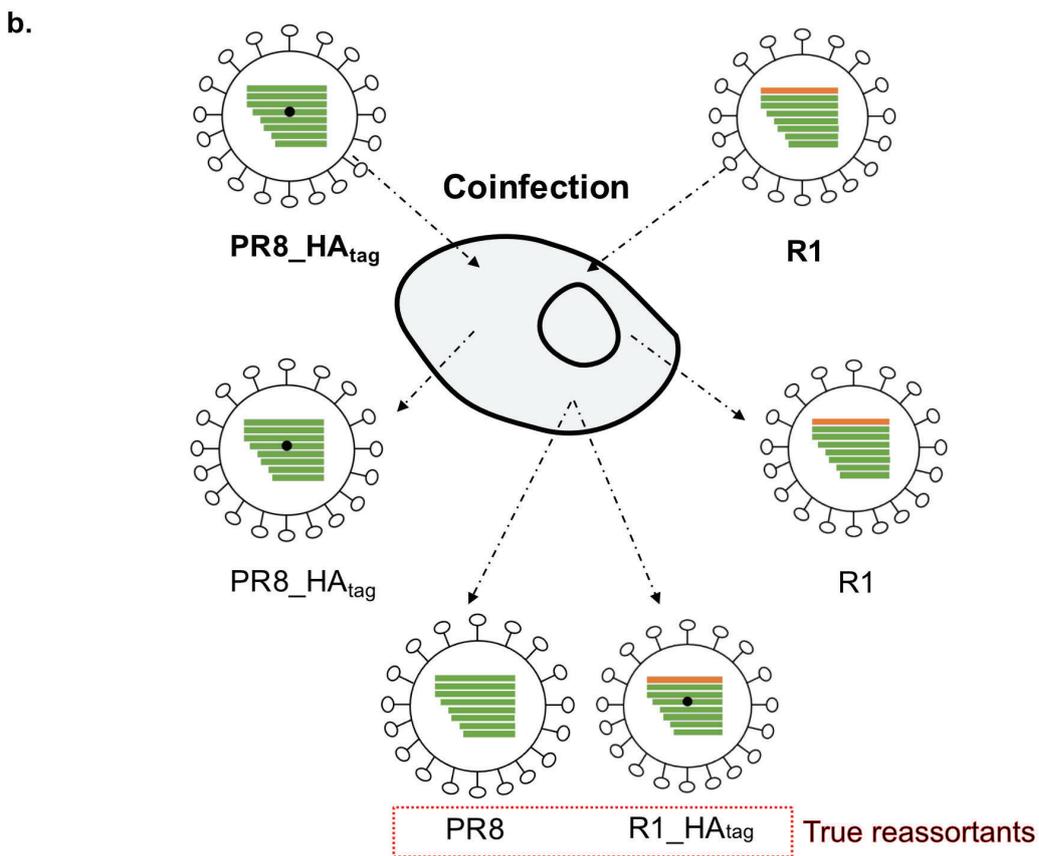
d. Primer optimisation for PR8 and PB2_MO. IVTs were used as test samples for all primer optimisations.

e. RNA concentrations of plaques from each coinfection.

f & g. pHW2000 and pUC plasmids used for 7:1 reverse genetics and in vitro transcription (respectively) of PR8 and MO



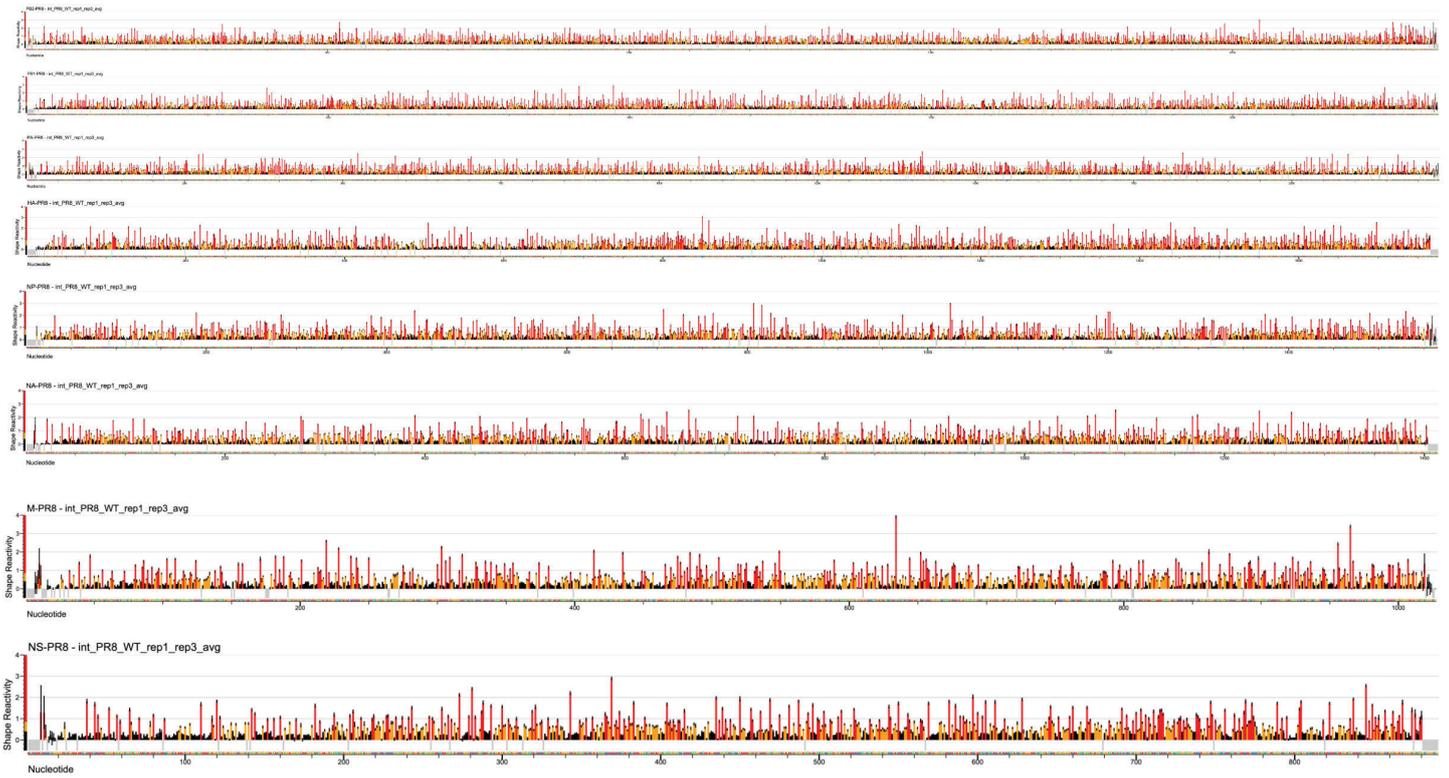
18



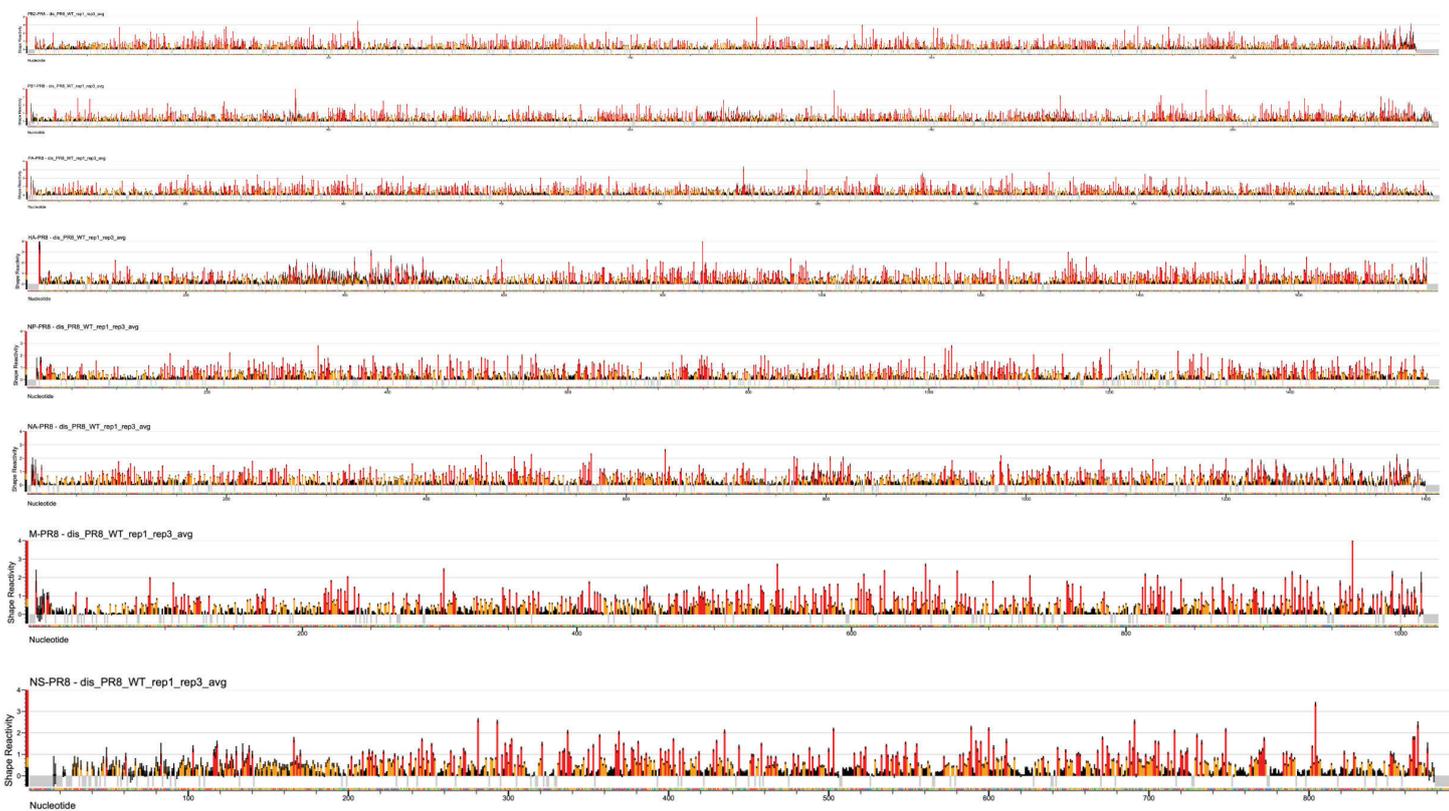
Supplementary figure 2 | Competition coinfection strategy

a. Normal coinfection conditions. Harvested supernatant will have virus genotypes produced by both coinfecting and single-infected cells which could inadvertently increase the number of observed virus genotypes and be misconstrued for reassortants produced within coinfecting cells. **b.** In order to circumvent this potential challenge, we substituted the parental PR8 with PR8_HA_{tag}. When coinfecting a cell, the segments reassort to produce the parental viruses like scenario (a), in addition to producing two new reassortant genotypes which are indicative to true reassortment within coinfecting cells. The presence of these true reassortants helps overcome the bias produced in scenario (a).

a. Segments 1 through 8 of intact PR8



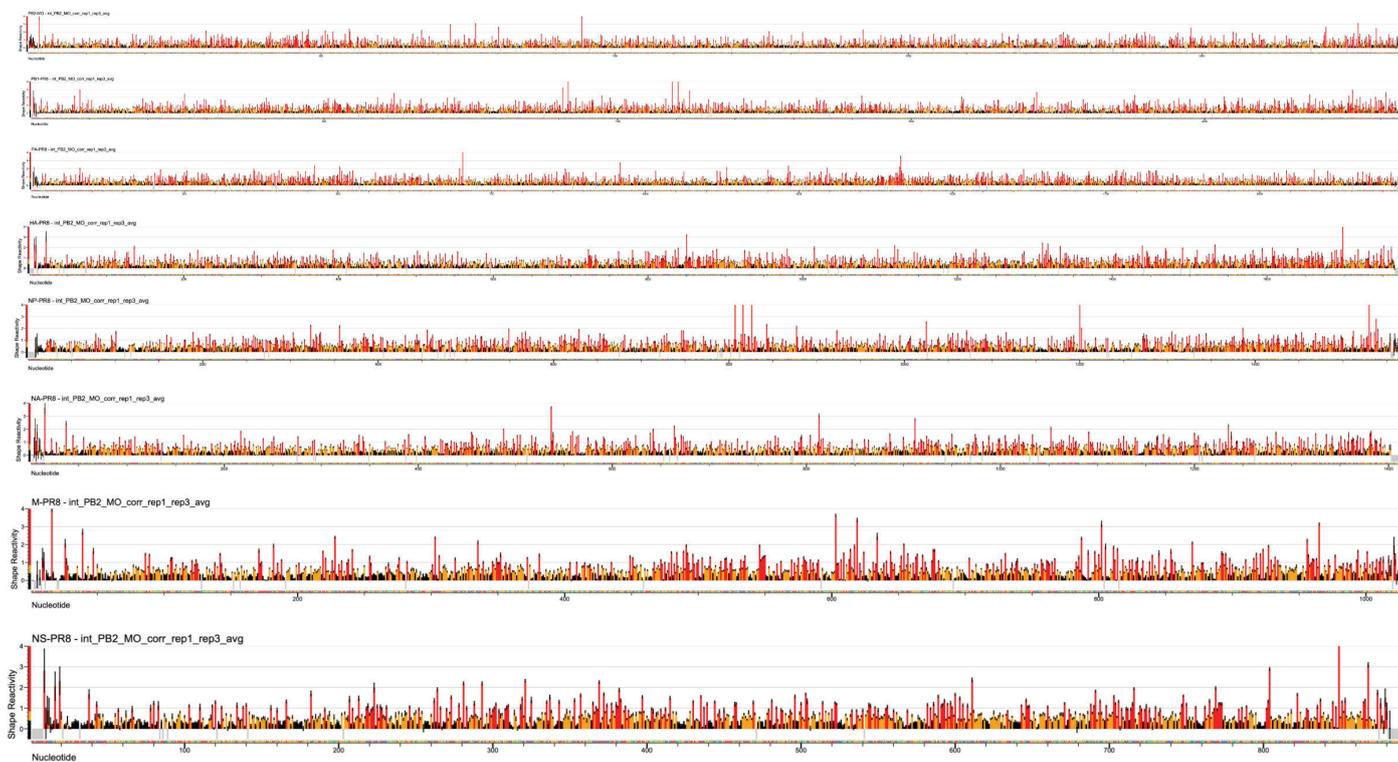
b. Segments 1 through 8 of disassembled PR8



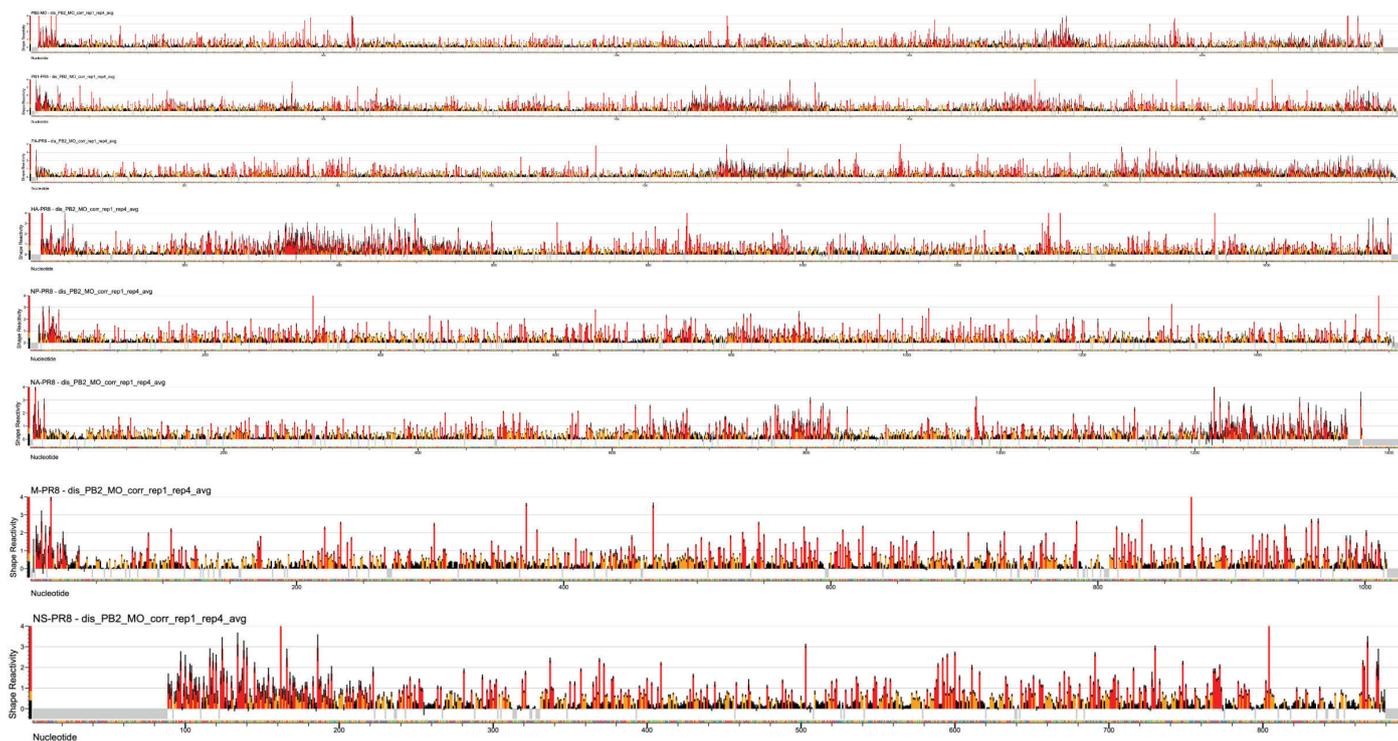
Supplementary fig 3 | SHAPE reactivity profiles of intact (a) and disassembled (b) PR8 virus.

Normalised SHAPE-MaP reactivities of each nucleotide with estimated standard errors produced by Shapemapper2 are shown here. The grey bars indicate positions with no data. The black bars indicate nucleotides with low SHAPE-MaP reactivities ranging between 0 - 0.4. These regions are involved in base-pairing or are shielded from 1M7 modification by tertiary interactions. The orange bars indicate regions that exhibit intermediate reactivities ranging between 0.4 - 0.8. The red bars indicate highly reactive nucleotides with reactivity rates greater than 0.8. These regions are flexible and unpaired or single-stranded, hence are modified readily by 1M7.

a. Segments 1 through 8 of intact R1



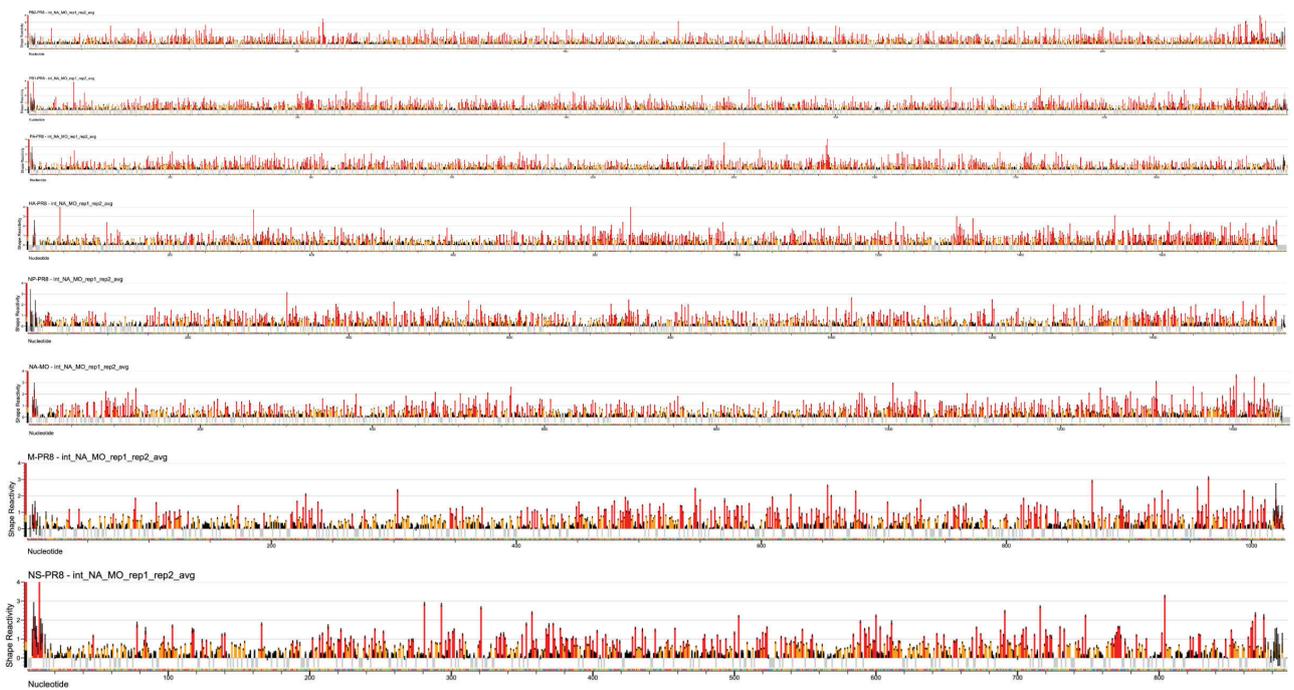
b. Segments 1 through 8 of disassembled R1



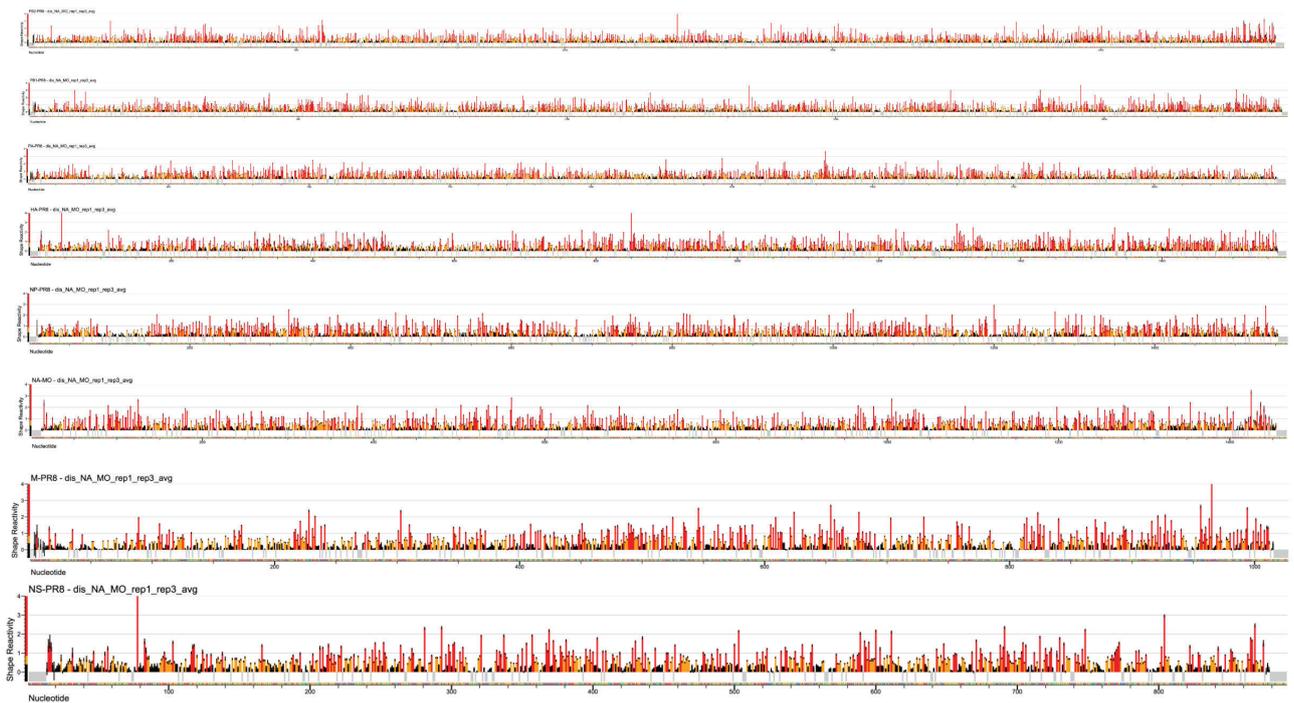
Supplementary fig 4 | SHAPE reactivity profiles of intact (a) and disassembled (b) R1 virus.

Normalised SHAPE-MaP reactivities of each nucleotide are shown with estimated standard errors produced by Shapemapper2 are shown here. The grey bars indicate positions with no data. The black bars indicate nucleotides with low SHAPE-MaP reactivities ranging between 0 - 0.4. These regions are involved in base-pairing or are shielded from 1M7 modification by tertiary interactions. The orange bars indicate regions that exhibit intermediate reactivities ranging between 0.4 - 0.8. The red bars indicate highly reactive nucleotides with reactivity rates greater than 0.8. These regions are flexible and unpaired or single-stranded, hence are modified readily by 1M7.

a. Segments one through eight of intact R6

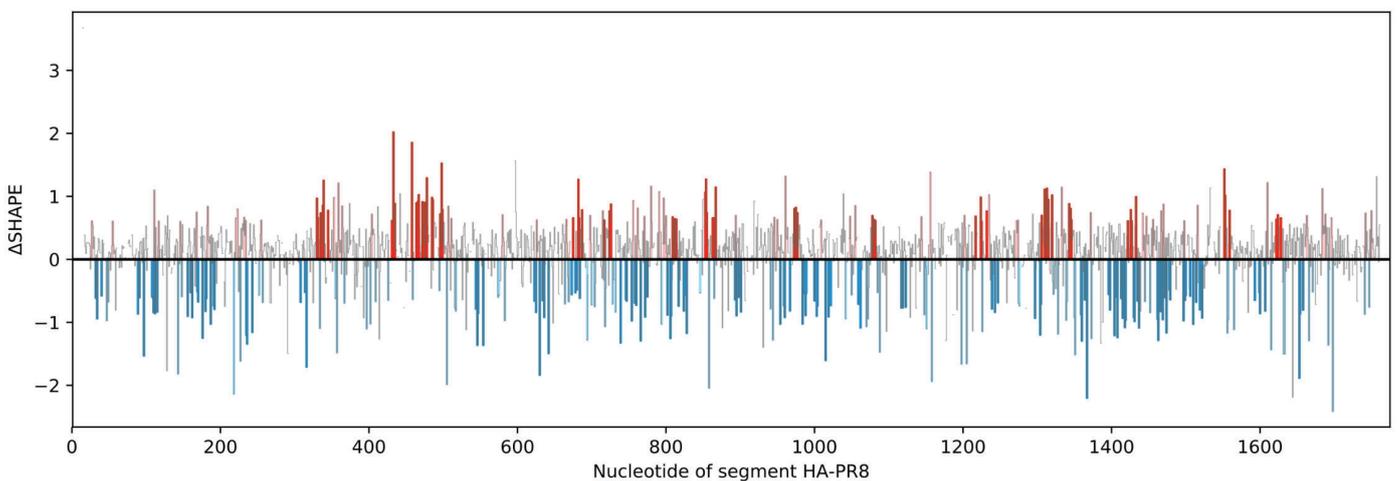
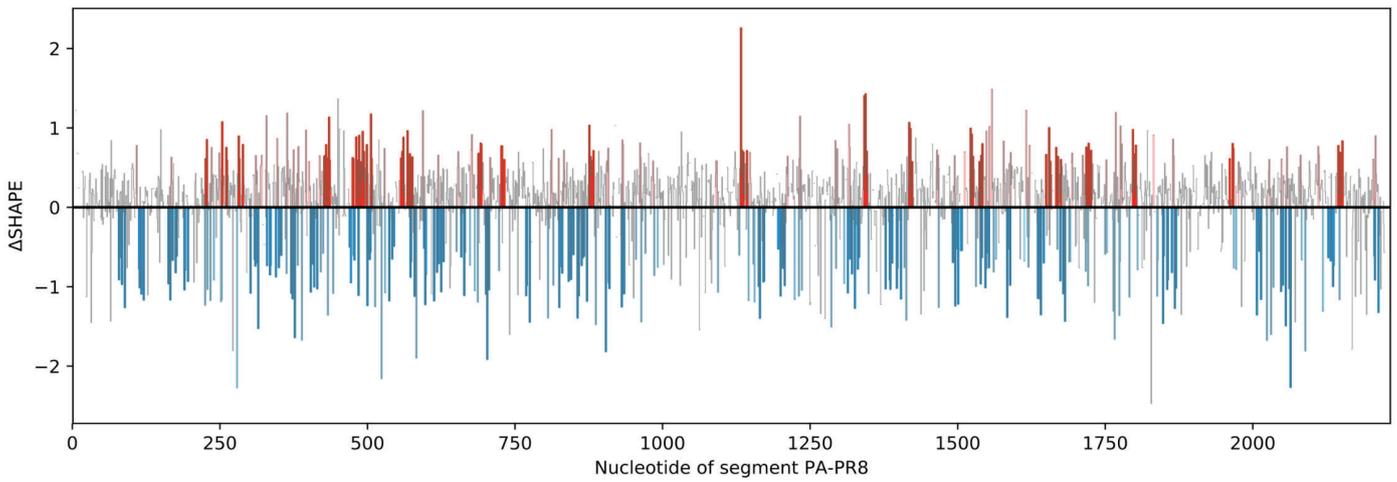
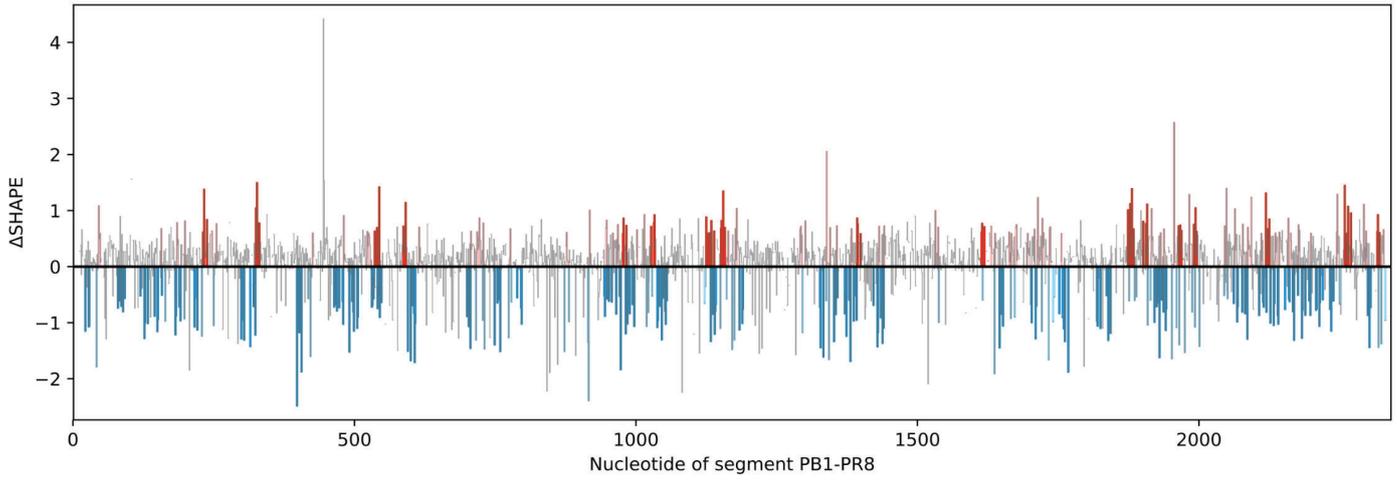
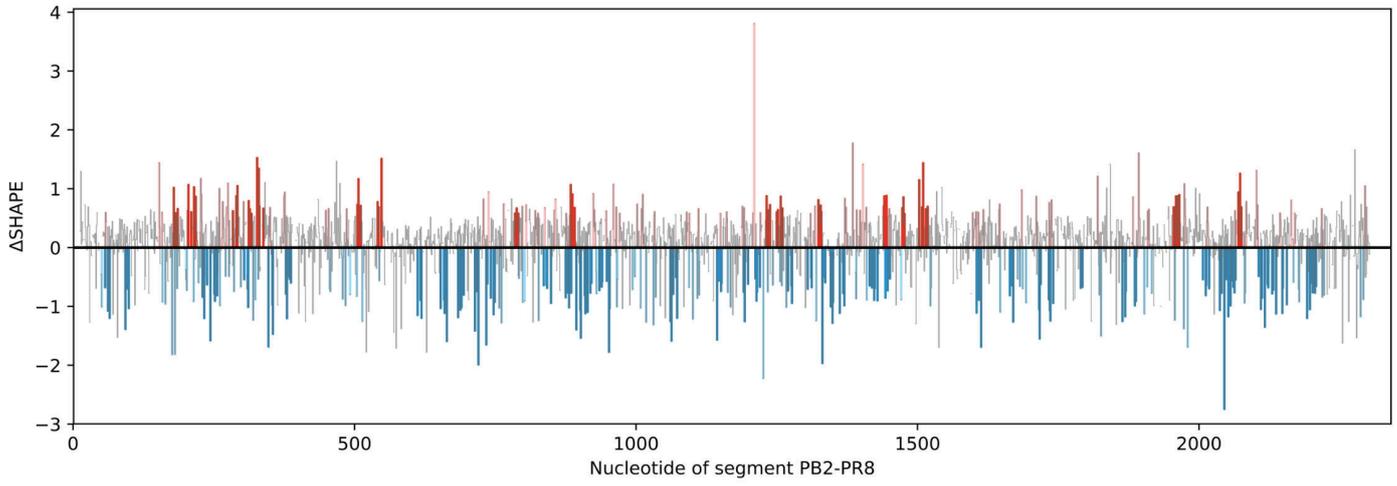


b. Segments one through eight of disassembled R6

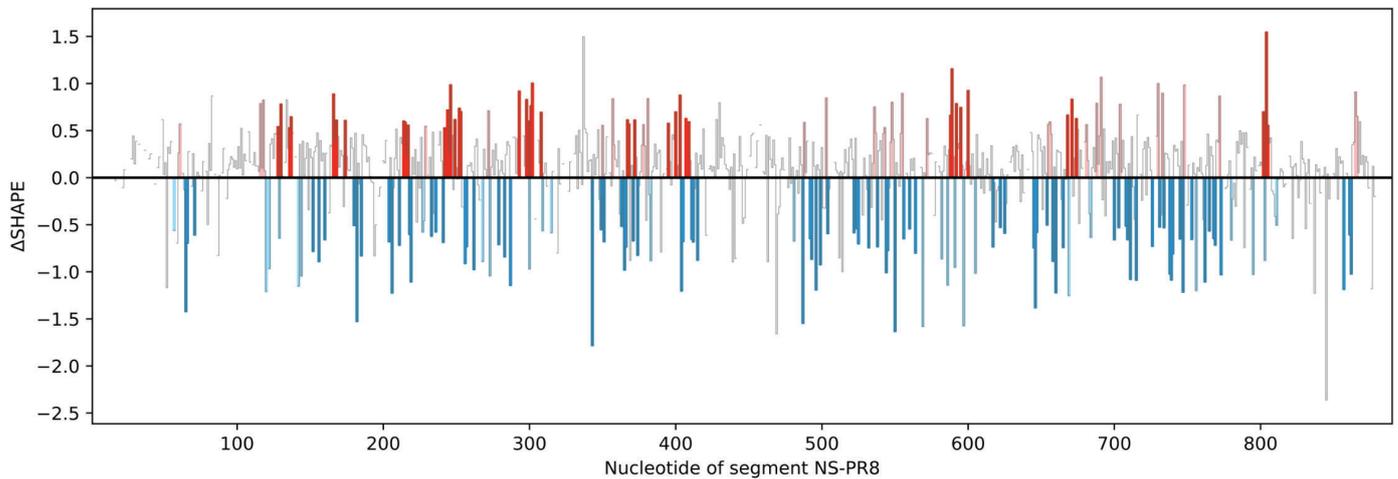
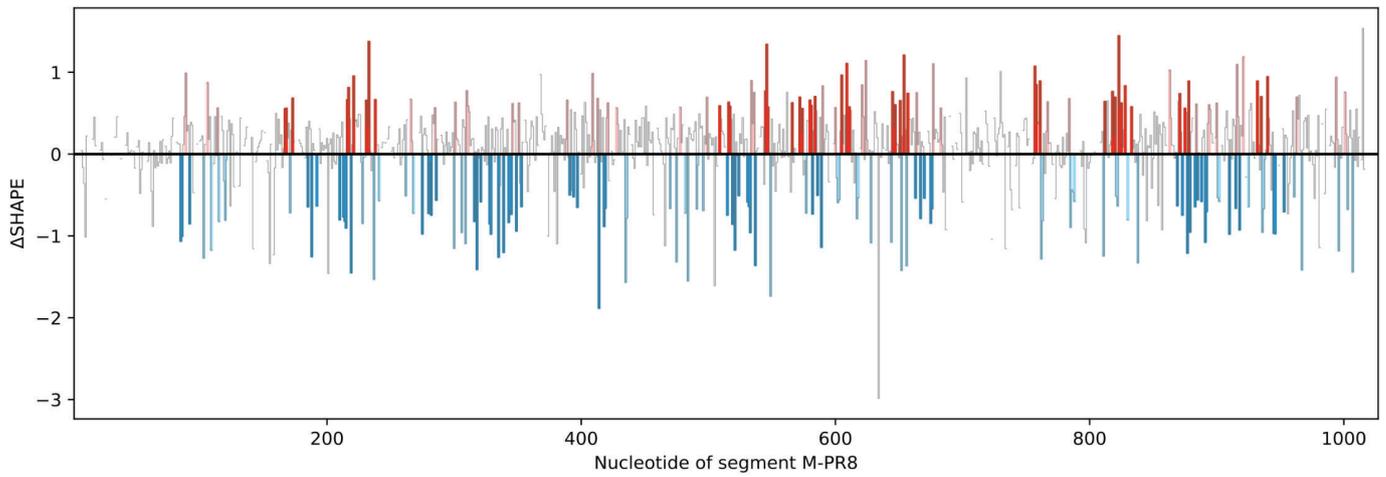
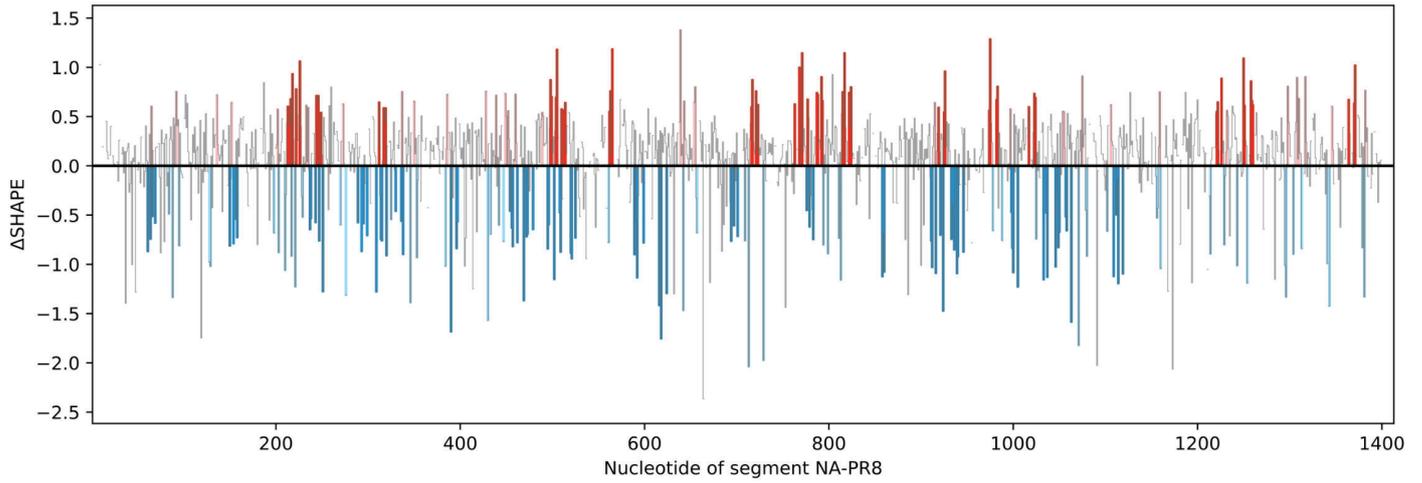
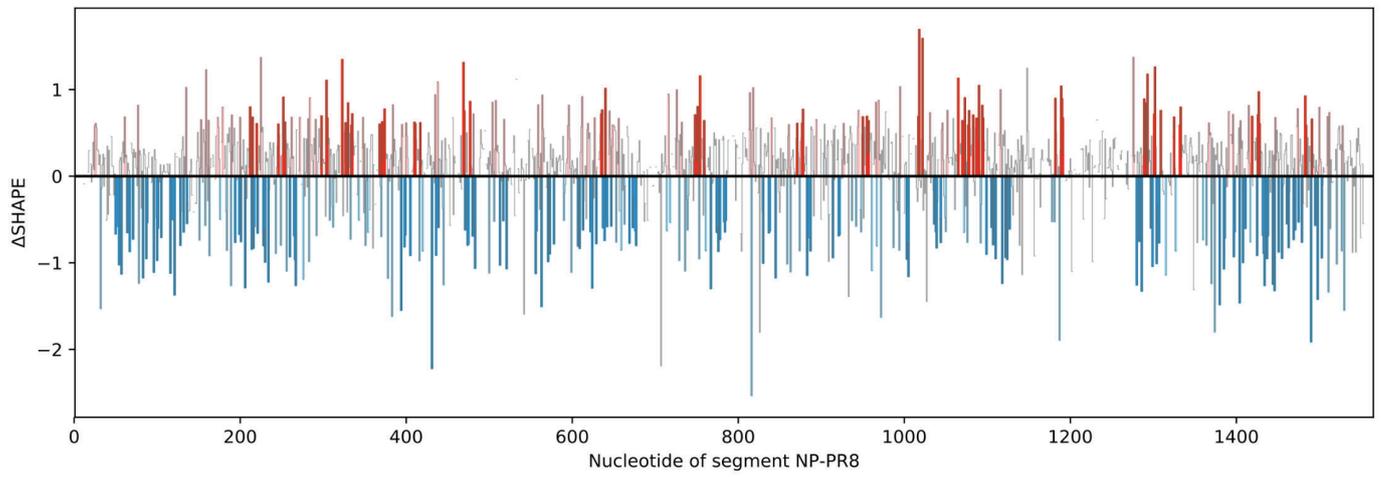


Supplementary fig 5 | SHAPE reactivity profiles of intact (a) and disassembled (b) R6 virus.

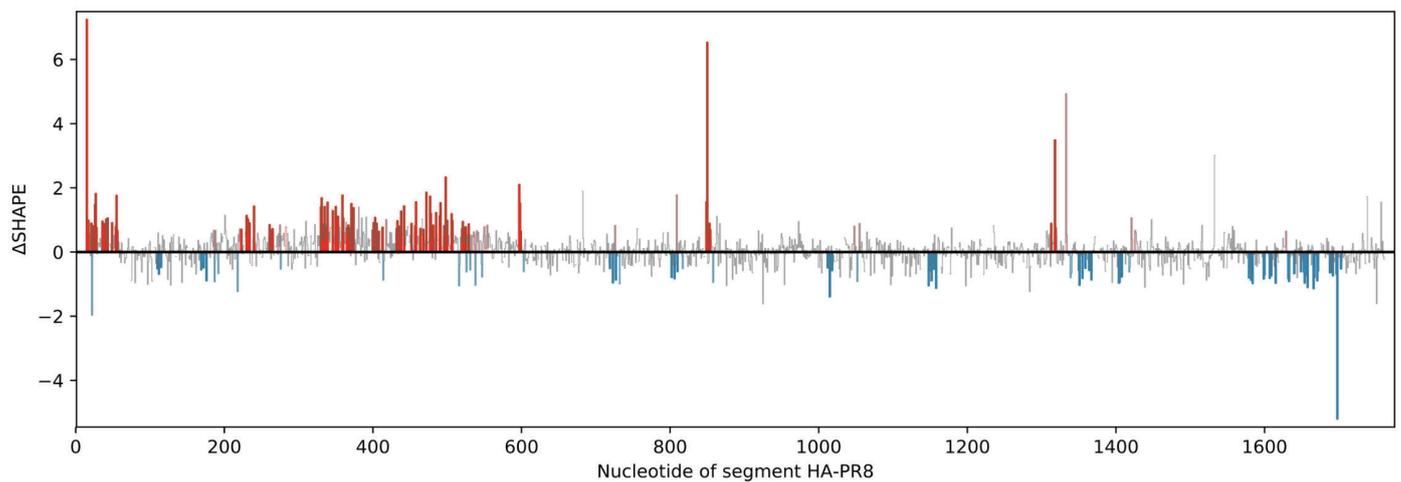
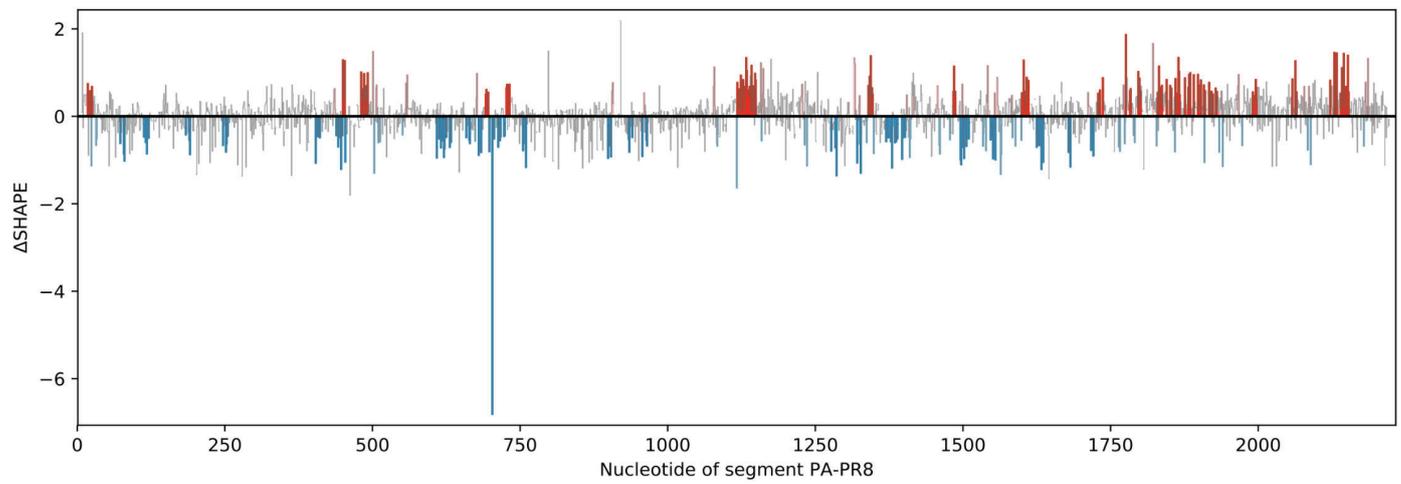
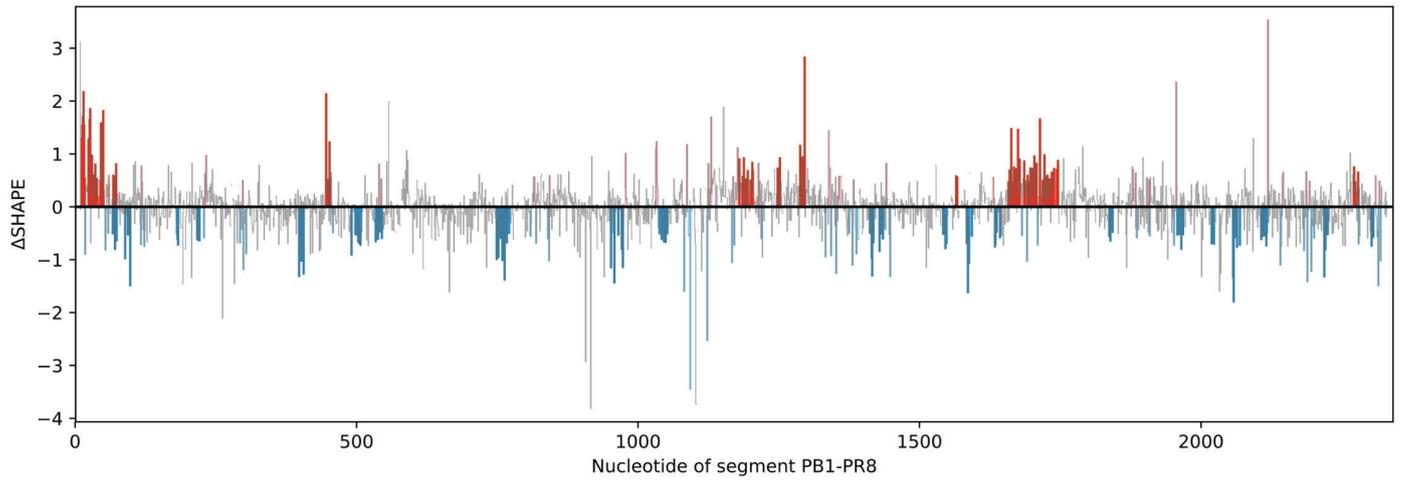
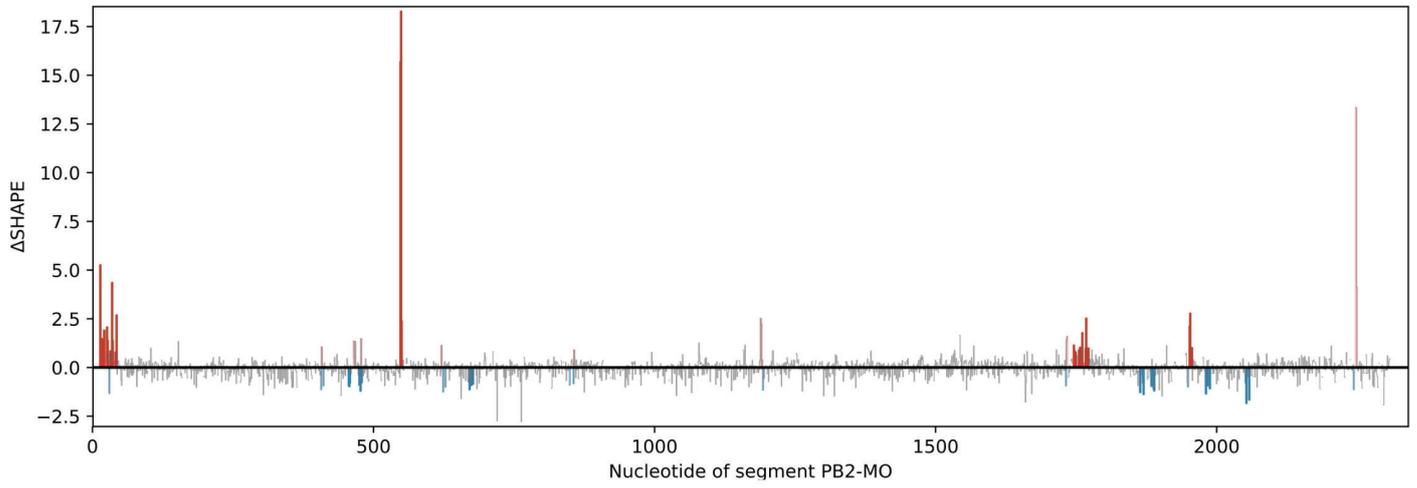
Normalised SHAPE-MaP reactivities of each nucleotide are shown with estimated standard errors produced by Shapemapper2 are shown here. The grey bars indicate positions with no data. The black bars indicate nucleotides with low SHAPE-MaP reactivities ranging between 0 - 0.4. These regions are involved in base-pairing or are shielded from 1M7 modification by tertiary interactions. The orange bars indicate regions that exhibit intermediate reactivities ranging between 0.4 - 0.8. The red bars indicate highly reactive nucleotides with reactivity rates greater than 0.8. These regions are flexible and unpaired or single-stranded, hence are modified readily by 1M7.



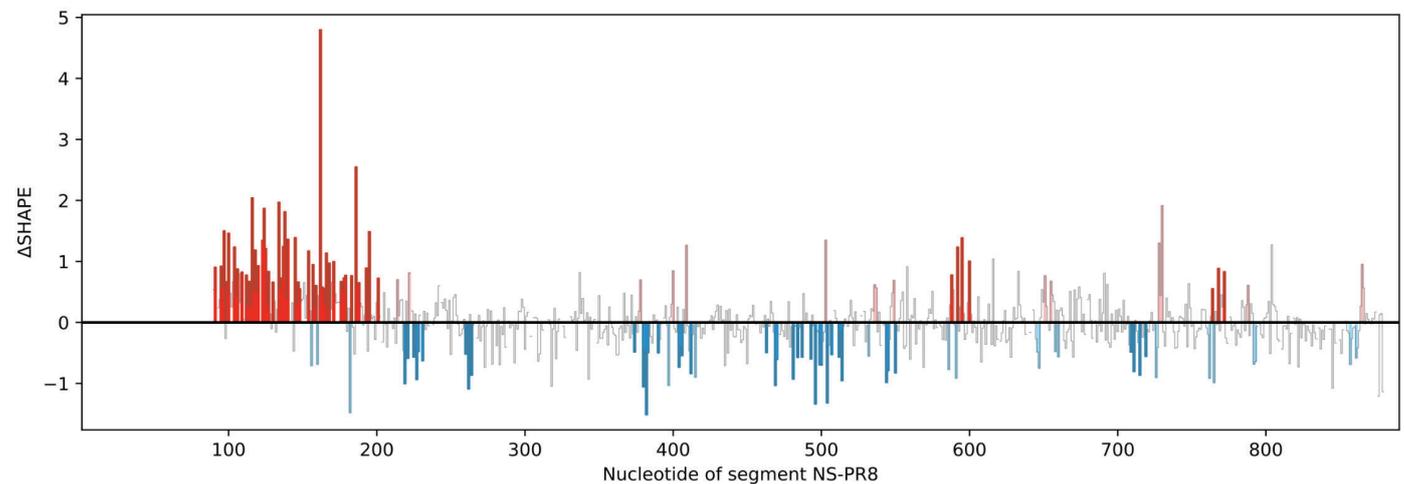
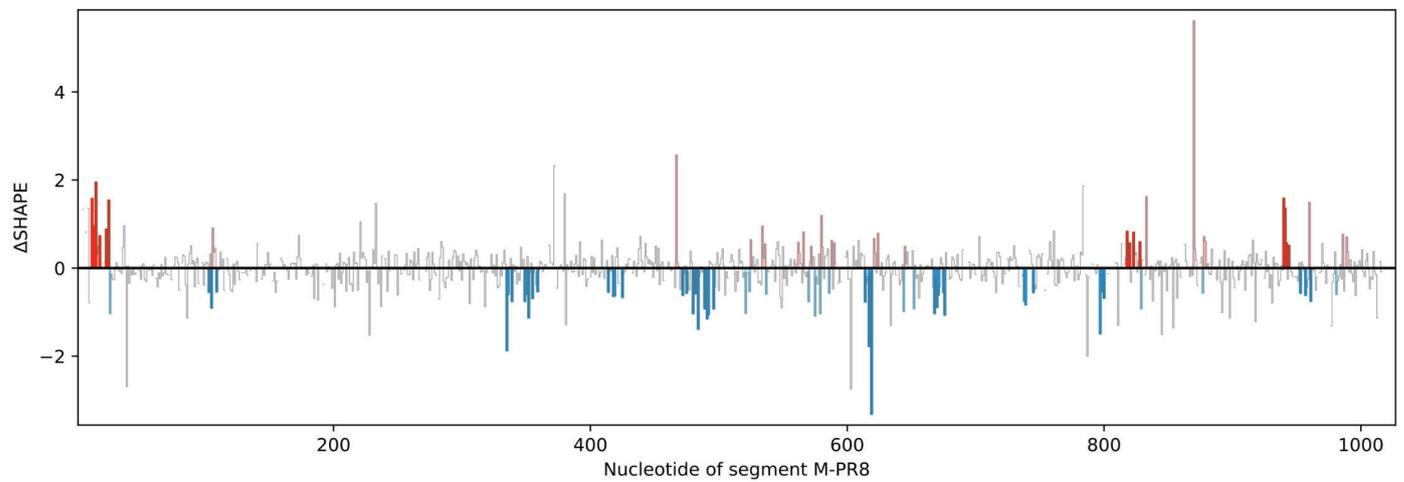
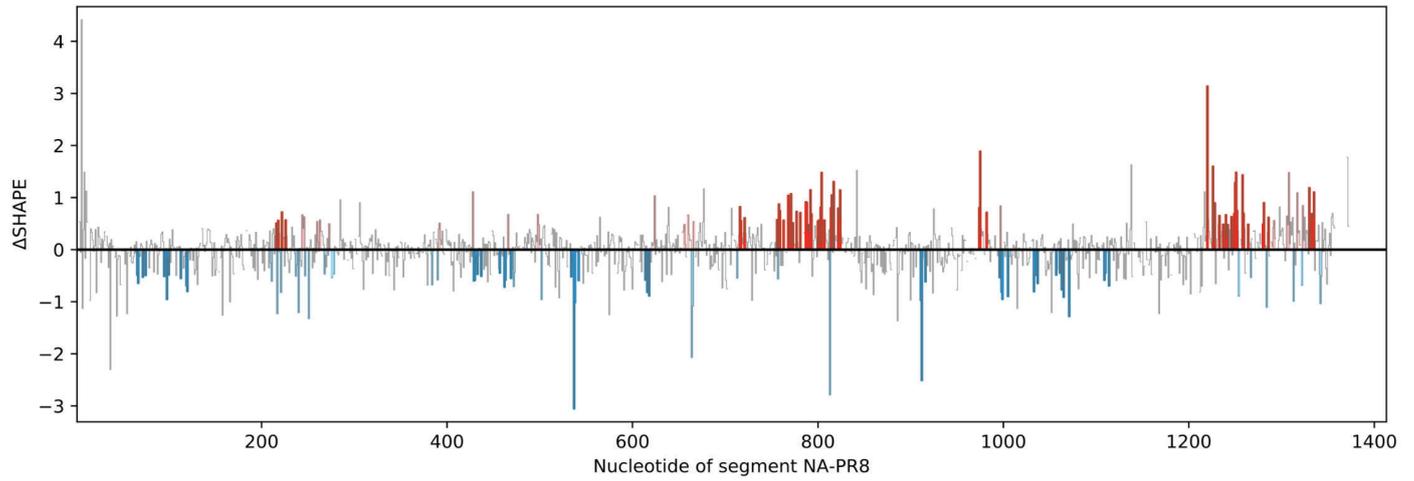
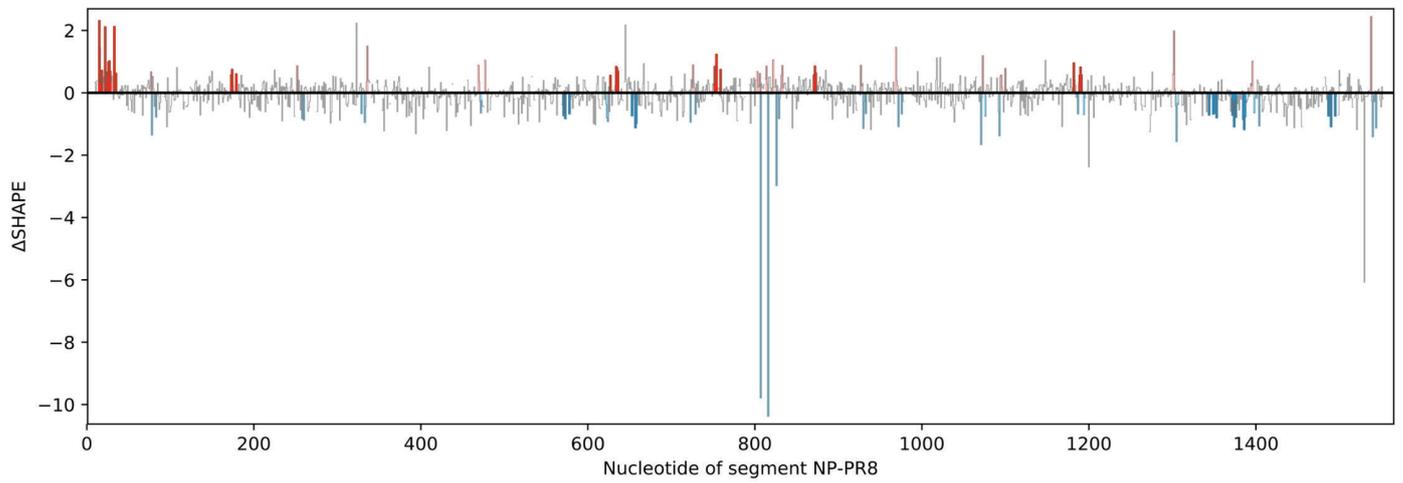
Supplementary fig 6 | Δ SHAPE results of segments one through four of PR8



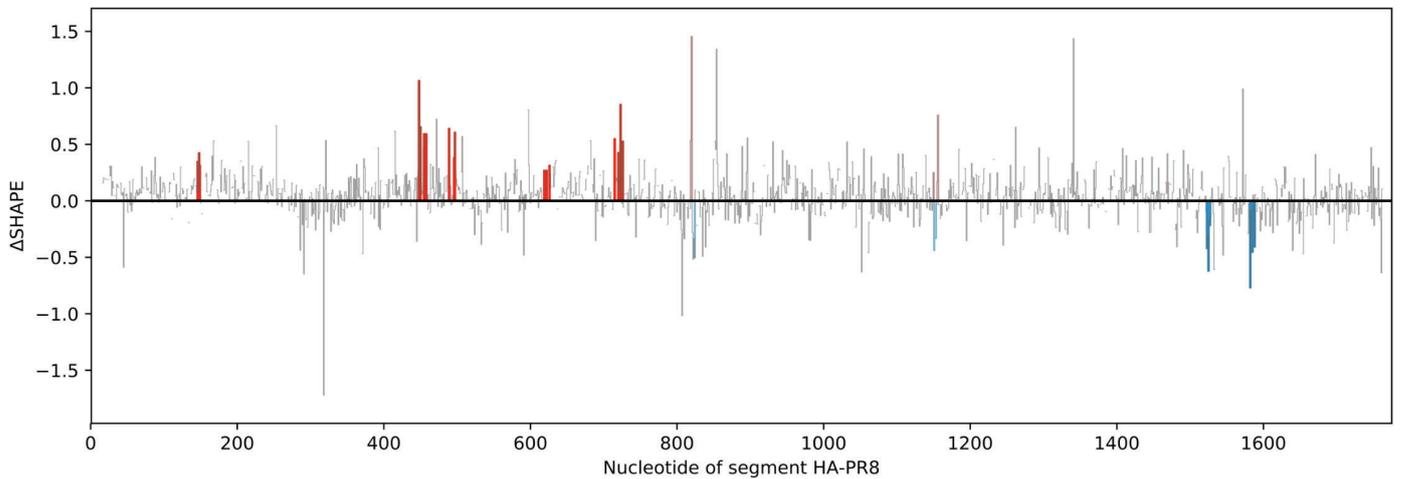
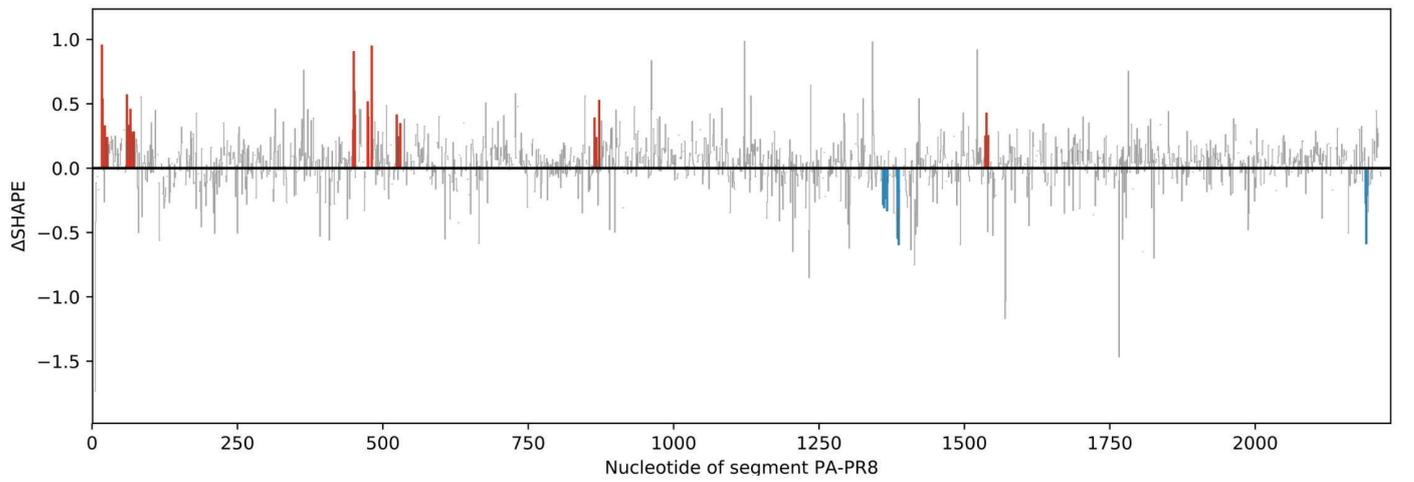
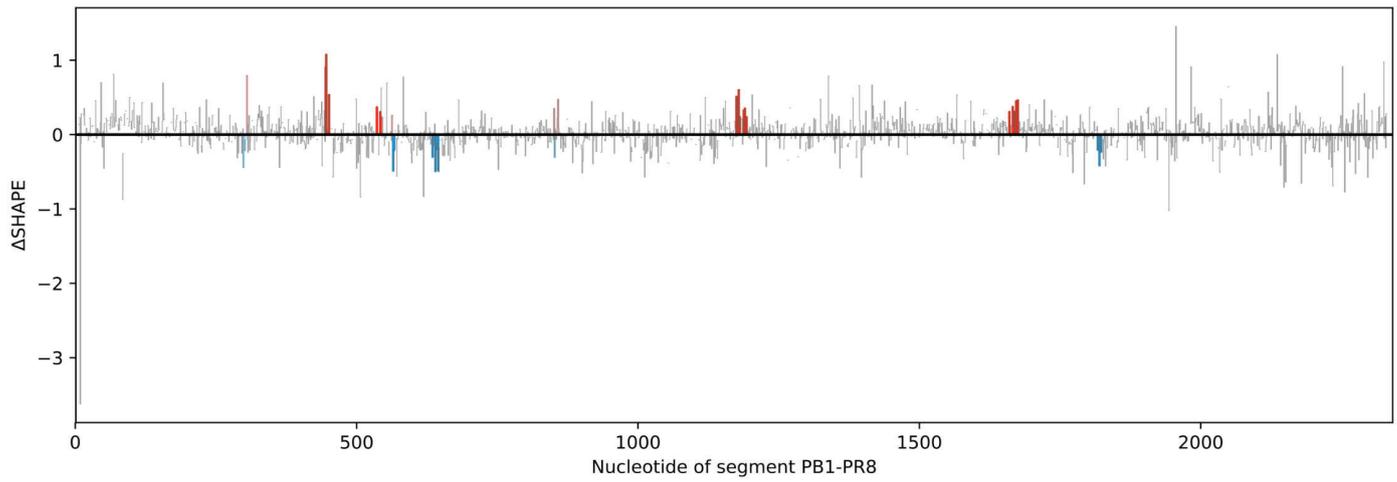
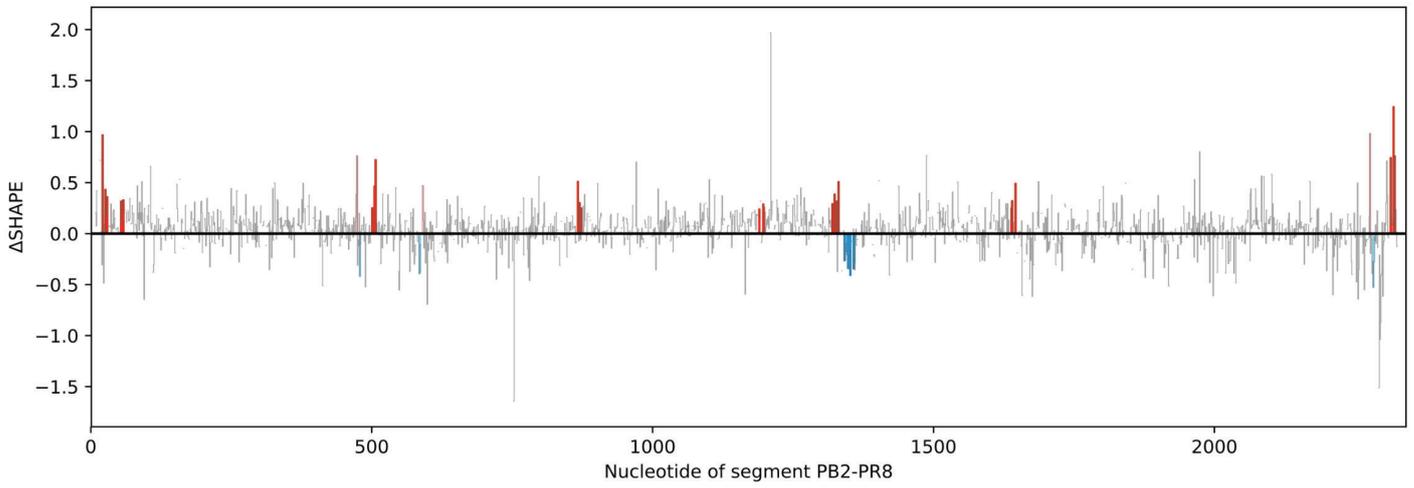
Supplementary fig 7 | Δ SHAPE results of segments five through eight of PR8



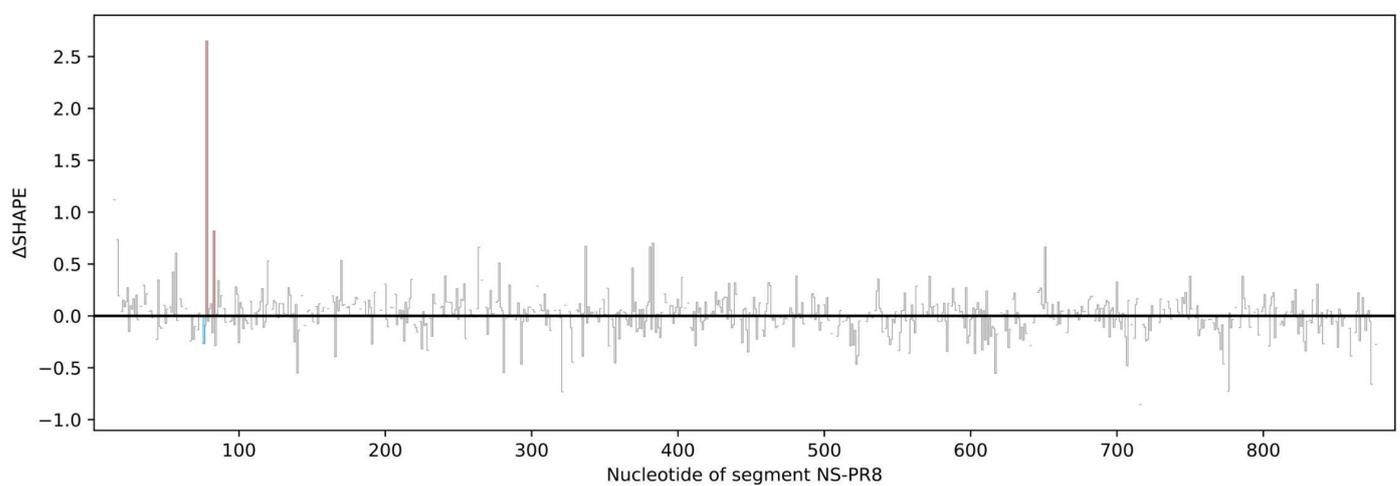
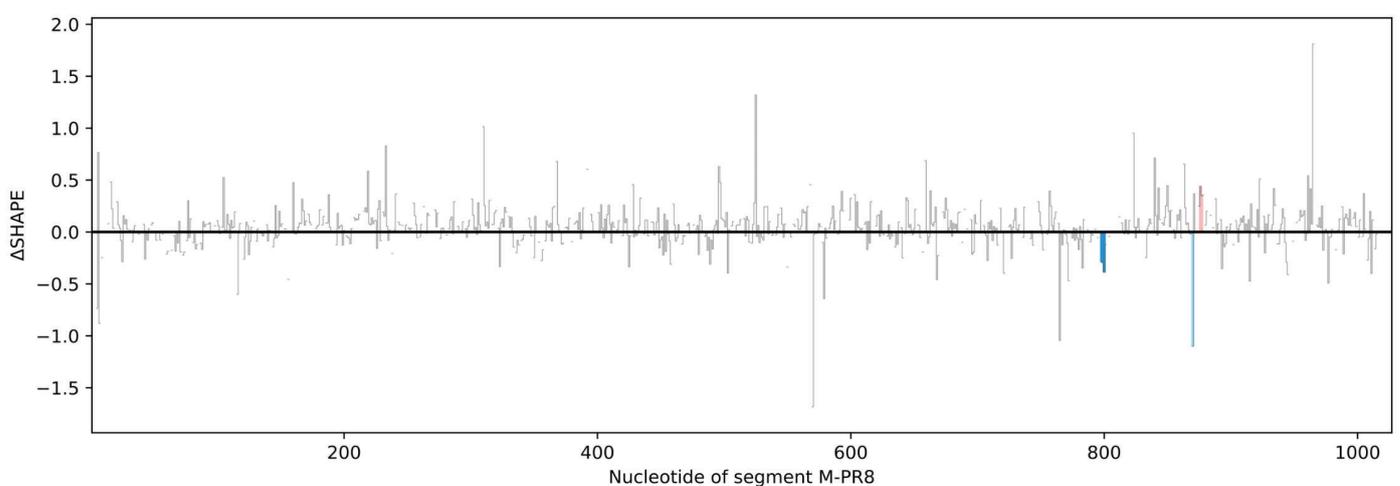
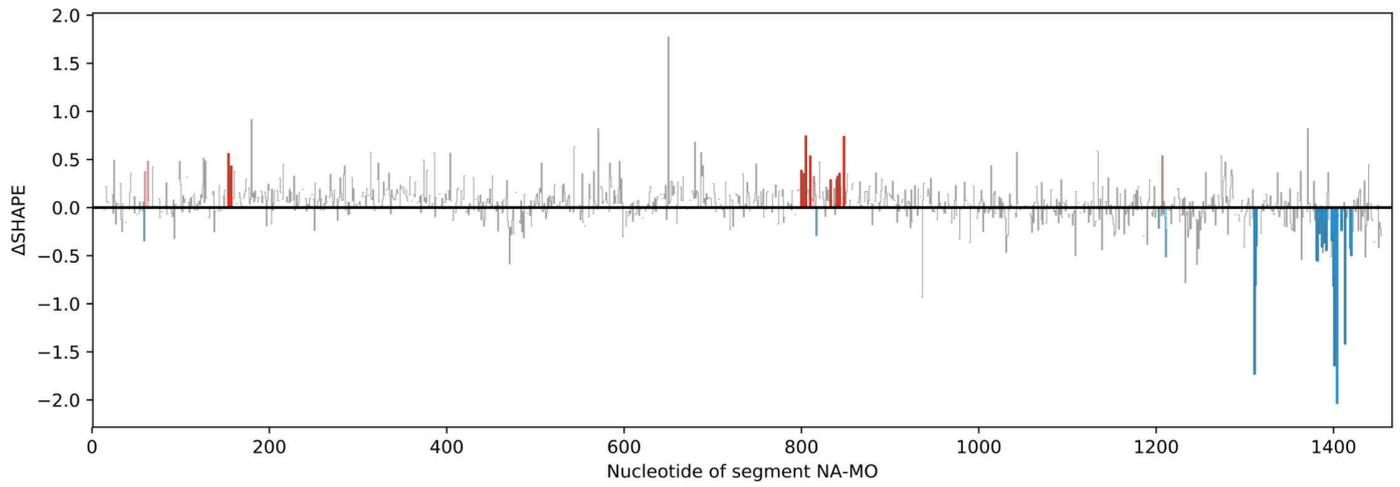
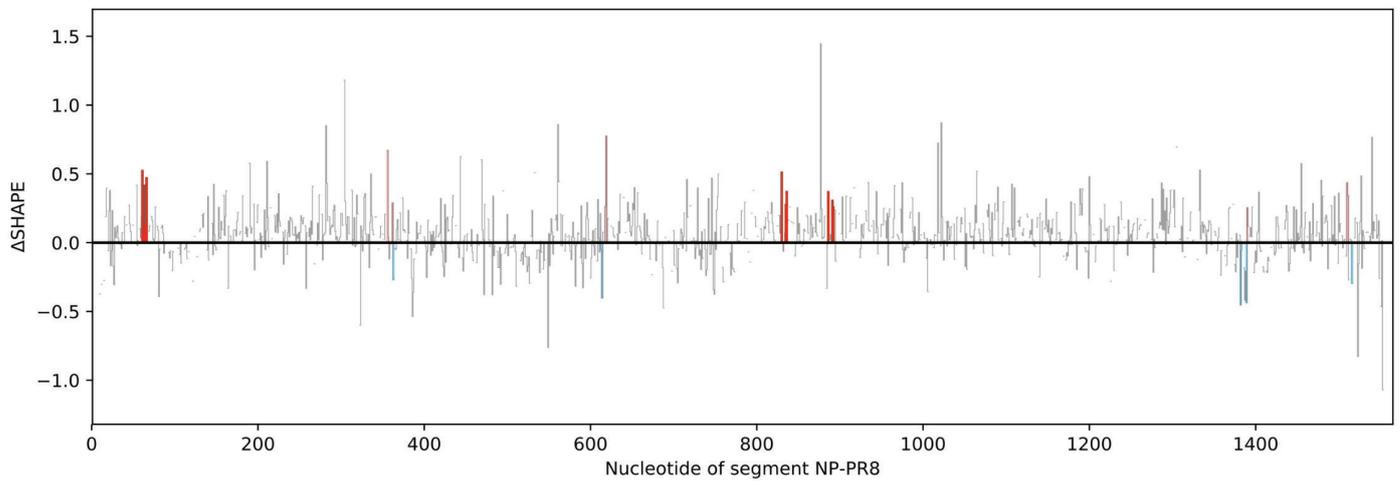
Supplementary fig 8 | Δ SHAPE results of segments one through four of R1



Supplementary fig 9 | Δ SHAPE results of segments five through eight of R1



Supplementary fig 10 | Δ SHAPE results of segments one through four of R6



Supplementary fig 11 | Δ SHAPE results of segments five through eight of R6

Résumé en français

Préface

Au cours des deux dernières décennies, les pandémies et les épidémies causées par le SRAS-CoV-2 (COVID-19), le VIH-1, le virus Zika, le virus de la grippe, le virus Monkeypox et le virus Ebola ont pesé sur l'infrastructure mondiale des soins de santé, l'économie, et l'environnement sociopolitique^{134,210,172,51}. L'incidence des agents pathogènes émergents et réémergents ainsi que le fardeau clinique qui en découle ont également augmenté, nécessitant de meilleurs programmes de préparation aux pandémies. Un aspect essentiel de ces programmes de lutte contre les futures pandémies est une compréhension approfondie des schémas d'évolution des virus. L'étude de la transmission inter-espèce et des facteurs d'adaptation de l'hôte qui permettent l'évolution du virus est essentielle pour atténuer les futures pandémies.

On sait que les virus à ARN évoluent plus rapidement que les virus à ADN. Parfois, cela peut conduire à l'émergence de nouvelles variantes phénotypiques capables d'échapper à l'immunité adaptative préexistante dans les populations humaines. Comme le montrent les exemples cités plus haut, cela peut entraîner des pandémies avec une morbidité et une mortalité importantes. Par conséquent, les mécanismes moléculaires d'acquisition de nouvelles mutations qui améliorent la virulence, et permettent la transmission entre espèces sont souvent le point de départ de la compréhension de la variabilité génétique et de l'évolution des virus.

Les virus à ARN dont le génome est segmenté utilisent le réassortiment génétique comme mécanisme évolutif auxiliaire. Plusieurs sous-types d'un virus peuvent infecter simultanément des hôtes naturels sensibles. De telles infections produisent de nouveaux virions descendants avec des segments de génome provenant des deux souches parentales. Cela accélère la variabilité génétique en facilitant l'acquisition de multiples mutations favorables qui favorisent la transmission inter-espèces, et l'émergence de nouvelles souches au cours d'un seul événement de réassortiment. Ces nouvelles souches réassorties sont antigéniquement distinctes de leurs souches parentales, augmentant donc la gravité de la maladie lorsqu'elles sont introduites dans une population immunologiquement naïve.

Bien que le réassortiment ait été démontré pour de nombreux virus segmentés tels que la fièvre catarrhale ovine, la grippe, les rotavirus et même certains virus multipartites, les mécanismes moléculaires sous-jacents restent largement inconnus. Parfois, les réassortiments produits peuvent être d'une virulence sous-optimale, voire non infectieux, en raison des discordances génétiques et antigéniques conférées par la nouvelle constellation génomique. Par conséquent, les événements de réassortiment qui produisent avec succès de nouvelles souches à potentiel pandémique sont rares. Cela signifie que le réassortiment n'est pas aléatoire, car la nature des segments génomiques en interaction le contraint. Par conséquent, certains de ces virus présentent un mécanisme sélectif d'empaquetage du

génomique. Ce mécanisme d'emballage incorpore sélectivement des segments de génome compatibles dans un virion bourgeonnant. Par conséquent, la compréhension des règles de l'encapsulation coordonnée du génome ouvre la voie au développement d'outils de prévision des pandémies, d'antiviraux ciblés et de stratégies vaccinales efficaces pour lutter contre l'évolution des virus.

Le virus de la grippe A (IAV) est un virus enveloppé à ARN de sens négatif. Son génome est divisé en huit segments, chacun codant pour au moins une protéine virale essentielle. Il est donc crucial d'emballer sélectivement une copie de chaque segment pour produire un virion infectieux au cours du cycle de vie du virus. Par conséquent, lors de l'emballage du génome dans le cytosol, les segments s'assemblent progressivement en un complexe octamérique composé de sept segments entourant un segment central. L'effet synergique des interactions ARN-ARN et ARN-protéines entre les segments maintient la conformation de ce complexe supramoléculaire octamérique. Des études approfondies ont permis d'identifier des régions qui pourraient être impliquées dans la formation et le maintien de ce complexe octamérique. De plus, d'autres études à haut débit ont également suggéré que l'IAV possède un mécanisme redondant d'emballage du génome avec un réseau flexible d'interactions ARN-ARN intersegmentaires. Si cela implique un mécanisme de sélection rigoureux et infaillible lors de l'emballage du génome, cela suggère également que le réseau a évolué pour résister à toute pression négative qui pourrait potentiellement nuire à la virulence du virus.

Dans cette étude, nous avons utilisé l'acylation 2'-hydroxyle sélective analysée par l'extension d'amorce et le profilage mutationnel (SHAPE-MaP) pour étudier l'étendue des structures secondaires de l'ARN sur la formation du complexe octamérique dans l'IAV. Nous avons utilisé deux états alternatifs du virus - intact et désassemblé - pour cartographier les régions potentielles qui pourraient être impliquées dans les interactions ARN-ARN intersegmentaires. Nous avons également utilisé des prédictions informatiques par paires pour étayer les données expérimentales SHAPE-MaP en cartographiant les régions potentielles cruciales pour l'emballage du génome.

Introduction

L'évolution des virus est facilitée par l'acquisition de nouvelles mutations bénéfiques qui permettent la transmission interspécifique et une virulence accrue. Alors que les taux de mutation varient selon les virus, les virus à ARN ont des taux de mutation plus élevés allant de 10^{-6} à 10^{-4} substitutions par site nucléotidique¹⁸⁹. Cela est dû à l'absence d'activité de relecture exonucléase du RdRp codé par le virus, entraînant parfois un gain de fonction²⁰⁸. Ces taux de mutation élevés sont favorisés car ils confèrent une plus grande capacité d'adaptation aux virus. L'ARN est également une biomolécule structurellement dynamique avec différentes fonctions effectrices régulées par la nature de ses conformations structurelles secondaires et tertiaires. Par exemple, les tiges-boucles hautement conservées dans l'UTR 5' du génome du VIH-1 régulent des étapes vitales telles que la transcription, la polyadénylation, l'épissage, la traduction, la transcription inverse et l'empaquetage de l'ARN¹⁶⁰. Par conséquent, les mutations qui altèrent ces structures auraient un impact sur les fonctions effectrices correspondantes, soit en exacerbant l'effet, soit en entravant la propagation du virus.

En général, les mutations bénéfiques aléatoires de l'ARN sont acquises progressivement sur plusieurs générations. Parfois, le transfert horizontal d'éléments du génome entre des virus de différentes lignées facilite également l'évolution³³. Ce phénomène est souvent observé dans les virus segmentés où les génomes d'ARN existent sous forme de plusieurs molécules distinctes. L'infection simultanée d'hôtes sensibles par deux sous-types génétiquement éloignés d'un virus segmenté facilite le réassortiment génétique. Les virions ainsi produits contiennent des segments de génome des deux souches parentales des virus infectants. Il s'agit d'un mécanisme évolutif unique des virus à ARN segmenté et il est essentiel à l'émergence de nouveaux virus et à la transmission interspécifique^{231,128,127}. Souvent, le réassortiment accélère l'évolution du virus par rapport aux événements de mutation incrémentielle. Cela est dû à l'acquisition de multiples mutations par le biais d'un seul événement de réassortiment où un ou plusieurs segments peuvent être échangés dans un virus progéniteur. En conséquence, les virus à ARN segmenté présentent un taux d'évolution plus élevé que les virus non segmentés. Par conséquent, l'étude du réassortiment est essentielle pour détecter de nouveaux réassortis ayant une puissance accrue de circulation dans une population immunologiquement naïve et pour étudier les réassortis qui pourraient échapper à la reconnaissance des anticorps, ou résistants aux antiviraux²³¹.

Une autre préoccupation concerne le développement d'une résistance aux antiviraux due à l'acquisition de gènes résistants par réassortiment génétique, comme décrit dans les IAV. Cela a été observé dans la résistance à l'adamantane obtenue par la présence d'un segment non apparenté mais compatible de la protéine M1. De même, de nouvelles mutations résistantes aux médicaments peuvent être acquises après réassortiment, comme le montre l'émergence de virus H3N2 résistants à l'amantadine et de souches résistantes à l'oseltamivir^{177,50}.

Le phénomène par lequel le réassortiment génère un nouveau variant viral modifiant l'épitope est appelé décalage antigénique. Cela permet aux nouveaux variants de contourner les anticorps neutralisants préexistants produits activement par infection naturelle des hôtes ou obtenus passivement par vaccination. La nouveauté génomique générée par le décalage antigénique invalide toutes les méthodes de contrôle des virus préexistantes. Par conséquent, la compréhension du décalage antigénique est un aspect important de l'atténuation de l'évolution des virus.

La « grippe asiatique » de 1957 a été causée par un virus H2N2 IAV généré par le réassortiment du virus H1N1 humain précédemment en circulation et un virus H2N2 aviaire qui a fourni les gènes PB1, HA et NA à la souche pandémique. L'analyse phylogénétique a identifié trois lignées distinctes de la protéine HA acquise chez les humains et chez deux espèces aviaires. La mortalité mondiale pendant la pandémie a été estimée à 1,1 million ; cependant, cette souche a disparu de la population humaine en 1968¹⁹¹. Elle a été suivie par la « grippe de Hong Kong » en 1968, causée par un virus H3N2 contenant une HA provenant d'un virus H1N1 aviaire et la neuraminidase N2 du virus H2N2 de 1957¹⁰³. Il s'agissait de la plus grande pandémie depuis la grippe asiatique. Le gène HA contenait deux mutations dans son site de liaison au récepteur provenant de virus aviaires, modifiant sa spécificité de liaison au récepteur d'une liaison préférentielle pour les acides sialiques liés en α -2,3, aux acides sialiques liés en α -2,6²²⁴ (**Fig 2**). Le virus pandémique H1N1 de 2009 (H1N1pdm) dérive six gènes de lignées virales porcines nord-américaines à triple réassortiment et deux gènes (codant les protéines NA et M) de la lignée virale porcine eurasiennne^{155,154} (**Fig 3**).

Il existe plusieurs facteurs déterminants pour la réussite du réassortiment génétique. Les goulots d'étranglement sélectifs et stochastiques peuvent avoir un impact négatif sur le réassortiment génétique et diminuer la diversité génétique des réassortiments. Lors de la co-infection d'une cellule hôte par deux ou plusieurs IAV, certains génotypes réassortis ne sont pas générés efficacement en raison d'incompatibilités entre les ARN hétérologues. Les interactions de faible affinité entre les ARN non apparentés (dérivés de différents virus parentaux) sont facilement supplantées par les interactions optimales de plus haute affinité entre les ARN apparentés (provenant du même virus parental)¹²⁷.

Selon la mise à jour ICTV 2021, les virus de la grippe appartiennent au règne des Orthornavirae et à la famille des *Orthomyxoviridae* (**Fig 4**). Sur la base des différences antigéniques des protéines NP et M1 des virus de la grippe, ils sont classés en quatre types : A, B, C et D²³⁷. La grippe est un agent pathogène humain courant qui provoque une maladie respiratoire contagieuse souvent appelée grippe. Elle se caractérise par des maladies légères à graves telles que la pneumonie et peut être associée à des taux de mortalité élevés.

Les virus de la grippe A (IAV) sont divisés en sous-types en fonction des deux protéines de surface HA et NA. HA se lie aux acides sialiques cellulaires pour l'entrée du virus, et NA clive

les acides sialiques après avoir bourgeonné de la cellule hôte pour faciliter la libération du virion²³⁵. Il existe respectivement 18 et 11 sous-types différents d'hémagglutinine et de neuraminidase. Ils sont nommés H1 à H18 et N1 à N11 et sont indiqués entre parenthèses pour tous les IAV¹⁰⁹. La nomenclature des souches de grippe A impose les informations suivantes en plus des sous-types HA et NA : le type antigénique (A, B, C ou D), l'hôte d'origine pour les virus non humains (porc, aviaire) ; l'origine géographique ou le lieu d'isolement, le numéro de la souche et l'année d'isolement de la souche, par exemple A/poulet/Novosibirsk/65/2005 (H5N1)¹. Les IAV peuvent être subdivisés en clades et sous-clades génétiques, autrement appelés respectivement groupes et sous-groupes. Les clades et sous-clades génétiquement différents peuvent ne pas être antigéniquement différents. Les sous-types actuels d'IAV qui circulent régulièrement dans la population humaine comprennent H1N1 et H3N2¹³⁷.

Les IAV sont pléiomorphes avec des formes sphériques dans les souches adaptées au laboratoire (~100 nm de diamètre) ou des formes filamenteuses dans les isolats cliniques (~100 nm de large et jusqu'à 20 µm de long)^{195,40}. L'enveloppe virale provient de la membrane plasmique de l'hôte et contient trois protéines transmembranaires : HA ; NA ; et la protéine matricielle 2 (M2) (**Fig 5**). Sous l'enveloppe virale se trouve la couche matricielle, une monocouche de M1 oligomérisé, qui fournit un support structurel³¹. Il a été proposé que la protéine M1 soit un déterminant dans la formation de virions filamenteux. Le réseau de revêtement membranaire de la protéine M1 impose vraisemblablement la forme de virions allongés ou filamenteux⁴⁴. D'autre part, l'état de faible énergie des vésicules membranaires correspond à une morphologie sphérique. Il semble probable que la disparition des virions allongés à faible pH corresponde à leur conversion en une forme sphérique. De même, la sphéricité des particules géantes produites par la fusion virion-virion résulte probablement d'un remodelage de la morphologie en sablier qui serait initialement réalisée. Dans les deux cas, la morphologie sphérique résulte probablement de la relaxation des interactions M1-enveloppe et des interactions M1-M1 en réponse à un faible pH⁶³.

Le génome viral de 13,5 kb est attaché à la couche matricielle et divisé en huit segments associés aux protéines NP pour former des ribonucléoprotéines virales (vRNP) (**Figure 6**)⁵⁵. Au total, huit vRNP codent pour 10 protéines majeures essentielles à la réplication virale : PB2 ; PB1 ; PA ; HA ; NP ; NA ; M1 ; M2 ; protéine non structurale 1 (NS1) ; et protéine d'exportation nucléaire (NEP/NS2)³¹. Les IAV possèdent trois protéines membranaires intégrales : une protéine canal ionique (M2) et deux glycoprotéines, l'hémagglutinine (HA), nécessaire à l'entrée dans les cellules hôtes, et la neuraminidase (NA), impliquée dans la libération des virions progéniteurs de la cellule hôte. Le génome code également jusqu'à 11 protéines accessoires non essentielles : PB2-S1 ; PB1-N40 ; PB1-F2 ; PA-X ; PA-N155 ; PA-N182 ; eNP ; NA43 ; M42 ; NS3 ; et tNS¹³¹ (**Fig 7**).

Le SHAPE est une technique de cartographie chimique qui utilise des agents électrophiles capables de se lier aux régions 2'-OH libres ou monocaténaire dans le squelette sucre-phosphate de l'ARN (**Figs 16 & 17**). L'ARN possède plusieurs groupes fonctionnels nucléophiles qui pourraient potentiellement servir de partenaires réactifs dans les réactions d'acylation, notamment les amines exocycliques sur les nucléobases cytosine, adénine et guanine, les groupes 5'-OH et 3'-OH aux extrémités du brin d'ARN et les groupes 2'-OH²¹¹. Les réactions d'acylation du réactif acylimidazole avec de l'ADN monocaténaire et de l'ARN monocaténaire d'une séquence analogue ont montré que l'acylation est hautement sélective envers l'ARN, indiquant que la réaction d'acylation se produit principalement à la position 2'-OH de l'ARN, plutôt qu'aux positions d'amine exocyclique ; des observations similaires ont été faites pour des nucléotides simples²¹¹. En plus des groupes 2'-OH, l'acylation peut également se produire au niveau des groupes terminaux 5'-OH et 3'-OH de l'ARN, ce qui doit être pris en compte lors de la conception d'une expérience d'acylation²¹¹. Il convient de noter que les ARN d'origine biologique sont souvent phosphorylés au niveau des positions terminales 5'-OH et 3'-OH, bloquant ces sites d'acylation potentiels, alors que les ARN synthétiques sont principalement conçus sans groupes phosphate terminaux²¹¹. Il est intéressant de noter que l'observation selon laquelle l'ADN est peu acylé par rapport à l'ARN suggère que la réactivité des groupes terminaux 5'-OH et 3'-OH est inférieure à celle des groupes 2'-OH²²⁶. Le SHAPE mesure donc la flexibilité locale des régions de l'ARN avec une résolution par nucléotide. Les réactifs SHAPE sondent les nucléotides flexibles, monocaténaire ou non contraints par des interactions tertiaires. Par conséquent, la capacité du réactif SHAPE à se lier à un nucléotide est décrite quantitativement par sa réactivité et est donc inversement proportionnelle à la capacité d'un nucléotide à être impliqué dans une interaction secondaire/tertiaire¹¹⁶.

La difficulté de déterminer expérimentalement les structures de l'ARN a conduit à l'utilisation généralisée de prédictions informatiques utilisant des informations de séquence. Les méthodes informatiques de prédiction de la structure de l'ARN ont évolué rapidement au cours des dernières décennies, principalement en raison d'améliorations fondamentales des algorithmes sous-jacents. Dans le même temps, les progrès des technologies de cartographie des structures ont permis un criblage à haut débit du « structure-ome » de l'ARN in vivo et in vitro. Ces deux approches ont été combinées ces dernières années pour augmenter la précision des prédictions de structure 2D et 3D¹²⁶.

La structure secondaire de l'ARN peut être prédite par des approches basées sur la thermodynamique, des simulations d'appariement de bases et l'IA2. Les approches informatiques les plus fréquemment utilisées pour prédire la structure secondaire de l'ARN sont les algorithmes de repliement basés sur la thermodynamique, notamment RNAstructure¹⁷⁹ et ViennaRNA package⁸⁹. Une structure secondaire prédite peut guider d'autres expériences ou des analyses de séquences comparatives et aider à la conception de molécules d'ARN antisens qui pourraient être testées comme antiviraux¹³².

Leur principal avantage est qu'ils peuvent prédire les structures secondaires de l'ARN en utilisant uniquement la séquence d'intérêt sans nécessiter de données expérimentales ou d'informations complémentaires. Ces algorithmes échantillonnent chaque structure qui peut être obtenue à partir de la séquence d'ARN en suivant un ensemble de règles de repliement (c'est-à-dire les nucléotides autorisés à s'apparier) et recherchent la structure native la plus probable, c'est-à-dire la conformation avec l'énergie libre minimale (MFE)²⁶². Les approches MFE renvoient la conformation la plus favorisée thermodynamiquement, qui est théoriquement la structure la plus occupée par un ARN en équilibre²¹¹. Pour calculer l'énergie libre d'une structure secondaire de l'ARN, les algorithmes basés sur la thermodynamique utilisent un ensemble de paramètres qui ont d'abord été déterminés par des expériences de fusion optique. Cependant, les méthodes de programmation dynamique basées sur la thermodynamique sont fortement limitées par la longueur de la séquence d'ARN. Elles sont précises pour les séquences courtes, mais les performances chutent considérablement pour celles de plus de 700 nucléotides. Cela limite leur utilité pour les ARN longs et complexes, tels que les génomes viraux à ARN simple brin (ssRNA)^{133, 253}. Néanmoins, ces algorithmes continuent d'être utilisés sur les génomes viraux, généralement avec une contrainte de localité, principalement en raison d'un manque d'alternatives.

Lacunes dans les connaissances

Les segments du génome de l'IAV sont sélectivement empaquetés dans les virions sous la forme d'un complexe supramoléculaire 7+1 contenant une copie de chaque segment. Cet assemblage sélectif est médié par un effort concerté des signaux d'empaquetage terminaux et internes spécifiques aux segments, comme le montre la littérature existante. Ces signaux d'encapsulation forment un réseau étendu mais flexible d'interactions ARN-ARN entre les segments. On a donc émis l'hypothèse que l'empaquetage sélectif et coordonné du génome chez l'IAV est régi par les signaux d'empaquetage qui maintiennent le complexe supramoléculaire 7+1 par le biais d'interactions ARN-ARN intersegmentaires. Un défi majeur dans ce domaine est le manque d'interactions ARN-ARN intersegmentaires fonctionnellement validées responsables de l'empaquetage du génome. Étant donné que les signaux d'empaquetage sont également spécifiques à la souche, l'absence d'interactions ARN-ARN intersegmentaires biologiquement pertinentes entrave la compréhension du réassortiment génétique chez l'IAV.

Après la réplication du génome et l'exportation nucléaire, on peut supposer que les vRNP subissent un remodelage structural local pour faciliter les interactions ARN-ARN intersegmentaires au cours de l'assemblage progressif du génome dans le complexe supramoléculaire 7+1. Il est donc essentiel d'étudier les structures des ARNv pour comprendre le mécanisme d'empaquetage du génome. La protéine NP joue également un rôle clé dans l'empaquetage du génome. Bien que les preuves du rôle de la NP dans l'empaquetage du génome de l'IAV n'en soient qu'à leurs débuts, son impact sur la flexibilité structurale de la vRNP est établi. La nature flexible de la vRNP permet un modelage structural local de la vRNP afin d'exposer de courtes boucles simple brin qui peuvent s'engager dans des interactions ARN-ARN intersegmentaires au cours de l'assemblage progressif du génome.

Il est donc impératif d'étudier l'impact des structures secondaires de l'ARN pour comprendre comment elles contribuent à la nature et à la flexibilité des vRNP. Ces connaissances peuvent s'avérer cruciales pour comprendre les contraintes et les médiateurs de l'empaquetage sélectif du génome chez l'IAV. En outre, l'étude de ce processus pourrait constituer la première étape dans la compréhension des contraintes génomiques du réassortiment entre des souches génétiquement différentes de l'IAV.

Hypothèse et objectifs

L'empaquetage du génome des IAV est rigoureusement sélectif, limitant le réassortiment génétique aléatoire entre des souches étroitement apparentées dans des conditions de coinfection. Des preuves irréfutables suggèrent que ce processus est médié par des signaux d'empaquetage discrets, spécifiques aux segments et aux souches, situés aux extrémités des segments et dans les régions internes. Les signaux d'encapsulation spécifiques aux segments interagissent les uns avec les autres pour former et maintenir un complexe supramoléculaire 7+1 comprenant une copie de chaque vRNP de l'IAV pendant l'assemblage et l'encapsulation du génome. Des mutations ou des altérations de ces signaux d'empaquetage ont montré qu'elles entravaient l'empaquetage d'un ou de plusieurs segments, ce qui entraînait une atténuation du virus ou la formation de particules virales vides. Par conséquent, nous pouvons supposer que ce complexe 7+1 est maintenu principalement par un réseau d'interactions ARN-ARN intersegmentaires, tout en étant soutenu par un ensemble auxiliaire d'interactions ARN-NP et protéine-protéine (complexe polymérase, protéines cellulaires, etc.). Nous proposons une approche multidimensionnelle combinant l'analyse de la structure secondaire de l'ARN et l'étude de l'empaquetage du génome pour aborder cette question.

On sait que la structure des vRNP joue un rôle dans l'empaquetage du génome, et Dadonaite *et al.* ont utilisé le SHAPE-MaP pour montrer que différents segments de l'IAV adoptent diverses conformations pour former des interactions inter- et intrasegmentaires au sein du complexe 7+1 d'un virion. Ils ont fourni une carte de la constellation des segments du génome pour la souche H1N1/WSN, apportant la preuve de la coségrégation des segments par le biais d'interactions ARN-ARN intersegmentaires.

Dans cette étude, nous utilisons le SHAPE-MaP sur deux virus parentaux (PR8 & MO) et deux réassortants produits par génétique inverse de PR8 et MO. Nous visons à étudier les conformations distinctes du complexe vRNP 7+1 dans chaque virus dans des conditions intactes et désassemblées. La comparaison des conformations distinctes des vRNP entre les conditions intactes et désassemblées nous informera sur les rôles potentiels des structures d'ARN dans l'assemblage du génome et l'empaquetage, tandis que la comparaison entre les souches parentales et les réassortiments mettra en lumière l'importance des éléments structuraux d'ARN dans le réassortiment génétique.

Les objectifs de ce projet sont donc les suivants

1. Produire des réassortants viraux à partir des souches PR8/Lyon (H1N1) et MO (H3N2) par la génétique inverse 7+1.
2. Identifier les changements dans les structures de l'ARNv sur les particules virales intactes et désassemblées à l'aide du SHAPE-MaP.

3. Utiliser les prédictions computationnelles par paire des interactions à longue portée entre les segments partenaires potentiels pour informer les régions d'encapsulation pertinentes sur le plan fonctionnel.

Matériels et méthodes

Cellules, plasmides et virus

Culture cellulaire

Les cellules MDCK.2 (ATCC CRL-2936) et les cellules HEK 293T (ATCC CRL-3216) ont été multipliées et maintenues dans du DMEM (Gibco 10564011) supplémenté avec 10 % (v/v) de FBS (Biosera FB-1090/500) et 50 U.mL⁻¹ de pénicilline et de streptomycine (Gibco 15140122) à 37°C et 5 % de CO₂.

Les MDCK.2 infectées par l'IAV ont été maintenues dans du DMEM (Gibco 10564011) complété par 0,2 % (v/v) de BSA (Merck A9418) et 50 U.mL⁻¹ de pénicilline et de streptomycine (Gibco 15140122) avec 1 µg.mL⁻¹ de TPCK-trypsine (ThermoFisher 20233). Cette solution sera appelée « milieu d'infection » dans les sections suivantes pour plus de facilité.

Amplification des plasmides

Dans ce projet, nous avons choisi les souches A/Puerto Rico/8/1934/H1N1 (PR8) et A/Moscou/10/99/H3N2 (MO) pour des raisons pratiques puisque les plasmides étaient déjà disponibles et que les virus pouvaient être manipulés dans un BSL-2. En outre, des résultats préliminaires non publiés suggéraient l'existence d'une incompatibilité au niveau de l'ARN entre certains segments de ces deux virus. Les plasmides pHW2000 correspondant à ces souches ont été obtenus auprès du professeur Bruno Lina (Université de Lyon, France). La séquence de la souche PR8 de Lyon utilisée dans ce projet est différente des séquences de référence NCBI (NC_002016 - NC_002023). La séquence de H3N2/MO est disponible ici (CY121373 - CY121380).

Des stocks de plasmides pHW2000 à unité unique (fig. supplémentaire : carte des plasmides) avec des inserts d'ADNc des segments PR8 et MO ont été transformés en bactéries Top 10 (ThermoFisher - Invitrogen C404010) & DH5α et amplifiés par Maxiprep (PureLink™ K210007) selon les instructions du fabricant. La séquence de chacun des plasmides amplifiés a été confirmée par séquençage Sanger (TubeSeq NightXpress, Eurofins Genomics Europe Shared Services GmbH) avec les amorces sens et antisens suivantes : 5' AGTACTGGTCGACCTCCGAAG 3' et 5' CTGATCAGCGAGCTAGCATTTAG 3' respectivement. Les concentrations de stock (1 µg.µL⁻¹) et de travail (100 ng.µL⁻¹) des plasmides ont été mesurées avec NanoDrop™ (ND-2000) et Qubit™ (ThermoFisher-Invitrogen™ Q32854) respectivement. Tous les plasmides ont été conservés à -20°C jusqu'à leur utilisation ultérieure.

Transcription in vitro pour produire de l'ARN viral nu

Vingt microgrammes de chaque unité unique de plasmide pUC57 avec des insertions d'ADNc des segments PR8 et MO (fig. supplémentaire : carte des plasmides) ont été linéarisés avec 4

μL d'enzyme de restriction pendant une nuit à 37°C . Cette opération a été suivie d'une extraction phénol-chloroforme et d'une précipitation à l'éthanol selon les protocoles standard. La linéarisation des plasmides a été confirmée par électrophorèse sur gel d'agarose natif à 1 % dans un tampon TBE 1X. L'ARN polymérase T7 interne a été utilisée pour la transcription in vitro conformément aux protocoles publiés⁶². Les IVT ont été purifiés par chromatographie d'exclusion de taille en utilisant une station AKTA sur une colonne TSK G4000SW. Brièvement, les éléments suivants ont été ajoutés à 25 μg de plasmides linéarisés : 4 mM de chaque NTPs, 1 mM de spermidine, 5 mM de DTT, 1 % de triton X-100 (v/v), 160 U de RNasin (Promega, Charbonnière-les-Bains, France), 50 $\mu\text{g}\cdot\text{mL}^{-1}$ de BSA finale, 0.5 μL de pyrophosphatase (Roche, Mannheim Allemagne) dans un tampon T7 1X (0,4 M Tris-HCl pH 8, 0,15 M MgCl_2 , 0,5 M NaCl) pour obtenir un volume final de 300 μL . Ce mélange a été incubé pendant 3 h dans un bain-marie à 37°C . Trente-cinq μL de tampon Dnase I 10X (Roche, Mannheim Allemagne), 3,5 μL de Dnase I (Roche) et 11,5 μL d'eau milliQ ont été ajoutés avant incubation pendant 1h dans un bain-marie à 37°C . La réaction a ensuite été stoppée par l'ajout de 150 μL d'EDTA 250 mM, et une extraction phénol/chloroforme a été réalisée, suivie de l'ajout de 3 vol d'éthanol et de 0,1 vol d'acétate de sodium 3 M à pH 5,0 et d'une précipitation pendant une nuit à -20°C . Après centrifugation à 21 000 g pendant 30 min à 4°C , les culots ont été lavés deux fois avec de l'éthanol froid (-20°C) à 70 % et séchés. Le culot d'ARN a été dissous dans 250 μL d'eau milliQ et purifié sur une colonne TSK Gel G4000SW (Tosoh Bioscience) à un débit de $1\text{ mL}\cdot\text{min}^{-1}$ dans un tampon contenant de l'acétate de sodium 0,2 M et 1 % d'éthanol. Les fractions contenant l'ARN d'intérêt ont été regroupées, précipitées à l'éthanol et redissoutes dans 100 μL d'eau milliQ. Les différentes fractions ont été quantifiées, puis l'intégrité de l'ARN a été analysée par électrophorèse sur gel de polyacrylamide dénaturant à 8 % (PAGE)¹⁶⁶. Les concentrations ont été estimées à l'aide d'un Nanodrop™ (ND-2000). Les IVT ont été conservées à -20°C et utilisées comme contrôles positifs pour la PCR.

Production de virus par génétique inverse 7+1

Des cellules HEK 293T subconfluentesensemencées dans des plaques à 6 puits ont été transfectées avec huit plasmides pHW2000 (7 PR8 + 1 MO) à des concentrations de 150 ng/plasmide en utilisant la Lipofectamine™ 2000 (ThermoFisher-Invitrogen 11668019) comme réactif de transfection et Opti-MEM™ (Gibco 31985062) selon les instructions du fabricant. Les VLP libérées dans les surnageants ont été récoltées à 24, 48 et 72 hpt et centrifugées à $3200 \times g$ pendant 15 minutes à 4°C avant d'être stockées à -80°C . Des tests de plaque ont été effectués sur les surnageants pour le calcul du titre de VLP. Les surnageants de VLP à 48 hpt ont été récoltés et passés deux fois dans des cellules MDCK.2 supplémentées avec $1\mu\text{g}\cdot\text{mL}^{-1}$ de TPCK-trypsine (ThermoFisher 20233). Les surnageants du deuxième passage du virus ont été titrés avec des tests de plaques. Les plaques individuelles ont été isolées dans des solutions BSA-PBS à 0,2 % et utilisées comme inoculum pour produire des stocks de virus clonaux exempts de DI. Par souci de simplicité, tous les virus réassortis sauvés seront désignés par les lettres R1 à R8 sur la base de la nomenclature canonique des ARNV de l'IAV. Par

exemple, R1 correspond au virus réassorti (PR8_PB2_{MO}) dont le segment 1 (PB2) est dérivé de MO et les sept segments restants (2 à 8) de PR8.

Production d'un stock de virus

Des cellules MDCK.2 subconfluentes supplémentées avec 1 $\mu\text{g.mL}^{-1}$ de TPCK-trypsine (ThermoFisher 20233) ont été infectées avec des plaques purifiées à une MOI de 0,001 des PR8, MO, R1 et R6. Les surnageants ont été récoltés à 48HPI et centrifugés à 3200 \times g pendant 15 minutes à 4°C pour éliminer les débris cellulaires. Jonas Fuchs (Université de Fribourg, Allemagne) a confirmé les séquences de chaque stock de virus par méthode indirecte à l'aide protocoles déjà publiés 70. Des aliquotes de stocks de virus à usage unique ont été quantifiées et conservées à -80°C pour les tests de compétition et SHAPE-MaP.

Cinétique de croissance

Des cellules MDCK.2 subconfluentes dans des plaques à 12 puits ont été infectées par PR8, MO, R1 à R8 et PR8_HA_{tag} (décrits ci-dessous) à une MOI de 0,001. Brièvement, le milieu a été aspiré des cellules MDCK.2, lavé une fois avec du PBS préchauffé auquel 500 μL de virus à une MOI de 0,001 ont été ajoutés et incubés à 37°C et 5 % de CO₂. Les virus ont été retirés, les cellules ont été lavées une fois avec du PBS préchauffé et 1 mL de milieu avec 1 $\mu\text{g.mL}^{-1}$ de TPCK-trypsine (ThermoFisher 20233) a été ajouté à chaque puits. Les surnageants ont été récoltés à 12, 24, 36, 48, 60 et 74 hpi et quantifiés à l'aide de tests de plaque. La cinétique de croissance a été tracée avec GraphPad Prism v10.

Essais en plaques pour la détermination du titre viral

Des dilutions en série de 10 fois des virus ont été préparées dans du DMEM (Gibco 10564011) complété par 0,2 % (v/v) de BSA (Merck A9418), 50U.mL⁻¹ de pénicilline et de streptomycine (Gibco 15140122) et 1 $\mu\text{g.mL}^{-1}$ de TPCK-trypsine (ThermoFisher 20233). Les dilutions de virus ont été incubées avec des cellules MDCK.2 à 37°C et 5 % de CO₂ pour une heure d'adsorption du virus. Les dilutions virales ont été éliminées et les cellules ont été lavées avec du D-PBS (Gibco 14190144). Un milieu de recouvrement de DMEM contenant 2 % d'agar Oxoid (ThermoFisher-Oxoid LP0028B), 5 % de NaHCO₃ (Merck 172577), 10 % de BSA (Merck A9418), 1 % de DEAE-Dextran (ThermoFisher 15455949), 1M HEPES (Carl Roth® 7365-45-9), et 1 $\mu\text{g.mL}^{-1}$ de TPCK-trypsine (ThermoFisher 20233) a été utilisé pour la formation de la plaque. Les plaques ont été comptées trois jours plus tard et les titres de virus ont été calculés. Les titres de virus ont été tracés avec GraphPad Prism v10.

Expériences de compétition

Mutagenèse de HA_PR8 et production du virus PR8_HA_{tag}

La mutagenèse dirigée sur le site a été réalisée sur le segment PR8_HA du plasmide pHW2000 avec l'ADN polymérase Phusion (ThermoFisher Scientific F530S) en utilisant le protocole PCR d'extension de chevauchement¹⁸. Les amorces sens et antisens utilisées pour la mutagenèse

étaient les suivantes : 5' GTCATCACTCGAAAGATTTGAAATTTCCCAAAG 3' et 5' CTTTGGGAAATTTCAAATCTTTGAGTGATGAC 3' respectivement. À 50 ng de plasmide pHW2000_PR8_HA, on a ajouté 1 µL de chaque dNTP, 10 µL de tampon de phusion 5X, 0,4 µL de Phusion Pol et 35,6 µL d'eau milliQ. Le plasmide a été amplifié dans les conditions suivantes : 98°C pendant 60 secondes, 98°C pendant 30 secondes, 53°C pendant 50 secondes, 72°C pendant 2 minutes, 72°C pendant 5 minutes, puis 4°C jusqu'à utilisation. Le nombre total de cycles était de 32. L'amplicon a été digéré avec 1 µL de Dpn I à 37°C pendant 1 heure. L'amplicon digéré (30 µL) a été transformé dans 50 µL de cellules compétentes XL-10 et incubé sur des plaques de gélose LB pendant une nuit à 37°C. Les colonies transformées ont été prélevées et confirmées par séquençage Sanger (TubeSeq NightXpress, Eurofins Genomics Europe Shared Services GmbH). Une mutation synonyme C425T a été introduite dans le plasmide pHW2000_PR8_HA, détruisant le site de restriction BstBI pour produire le segment pHW2000_PR8_HAtag. En utilisant la génétique inverse 7+1 avec ce plasmide muté, un virus PR8 modifié (ci-après dénommé PR8_HAtag) a été récupéré. Des stocks de virus PR8_HAtag ont été produits selon le protocole décrit ci-dessus et des tests de plaque ont permis de déterminer le titre viral.

Cinétique de réplication à cycle unique

Les cellules MDCK.2 ont étéensemencées dans une plaque à 12 puits. Les inoculums de virus aux MOI 1, 5 et 15 ont été préparés dans un milieu d'infection avec 1 µg.mL⁻¹ de TPCK-trypsine (ThermoFisher 20233). Le milieu de cellules MDCK.2 subconfluentes a été remplacé par 1 mL de PBS glacé dans chaque puits, la plaque étant placée sur de la glace pendant 3 minutes. Le PBS glacé a été remplacé par 350 µL d'inoculum de virus préparé et incubé 45 minutes sur la glace suivies de 45 minutes à 37°C et 5 % CO₂. Après incubation, les suspensions virales ont été retirées et les cellules infectées ont été lavées cinq fois avec 1 mL de D-PBS préchauffé. 1 mL de milieu d'infection frais sans TPCK-trypsine a été ajouté à chaque puits pour une incubation à 37°C et 5 % CO₂. Les surnageants ont été prélevés à 0, 4, 8, 12 et 24 hpi et des tests de plaque ont été effectués pour titrer le virus.

Co-infections de compétition

Les trois conditions de coinfection étaient les suivantes

Co-infection de contrôle - PR8_HAtag avec PR8,

Condition A - PR8_HAtag avec R1 et

Condition B - PR8_HAtag avec R6

Les cellules MDCK.2 ont étéensemencées dans une plaque à 12 puits au volume d'ensemencement recommandé. Les cellules subconfluentes ont été co-infectées avec des virus, chacun à une MOI de 5 selon le protocole d'infection à cycle unique décrit précédemment. Le surnageant de co-infection a été récolté à 12 hpi et les titres viraux ont été calculés à l'aide de tests de plaque.

Les plaques individuelles de chaque condition ont été isolées et maintenues dans 1 mL de D-PBS (Gibco 14190144) avec 0,5 % de BSA (Merck A9418). Ensuite, 500 µL de chaque plaque

isolée ont été réinfectés dans des cellules MDCK.2 subconfluentes dans une plaque à 24 puits pour l'amplification. Après adsorption du virus pendant 1H à 37°C et 5 % CO₂, les cellules ont été lavées et l'inoculum de virus a été remplacé par un milieu d'infection contenant 1 µg.mL⁻¹ de TPCK-trypsine (ThermoFisher 20233). Les surnageants ont été récoltés à 48HPI et centrifugés à 3200 × g pendant 10 minutes avant d'être stockés à -80°C.

L'extraction de l'ARN de ces plaques a été réalisée avec le kit Direct-zol RNA MiniPrep (Ozyme ZR1055) selon les instructions du fabricant, et les concentrations ont été mesurées avec le kit Qubit RNA HS (ThermoFisher - Invitrogen Q32852). L'amplification par PCR spécifique à un segment (tableau des amorces) a été réalisée sur 10 µL d'ARN extrait avec le kit Qiagen One-Step RT-PCR (Qiagen 210212) selon les instructions du fabricant, suivie d'une électrophorèse sur gel d'agarose natif à 2 %. Le séquençage Sanger par Eurofins GATC (TubeSeq NightXpress, Eurofins Genomics Europe Shared Services GmbH) a confirmé le phénotype de la plaque.

Sondage chimique et biologie computationnelle

SHAPE-MaP

Production et purification du virus

Deux cents millilitres de cellules MDCK.2 subconfluentes ensemencées dans huit flacons de culture cellulaire T175 ont été infectés par les virus PR8, MO, R1 et R6, chacun à une MOI de 0,001. Les surnageants ont été récoltés à 48 hpi et centrifugés deux fois à 3200 × g pendant 10 minutes à 4°C pour éliminer les débris cellulaires. 200 mL de chaque surnageant de virus ont été répartis dans des tubes d'ultracentrifugation (Beckmann Coulter C13926) contenant chacun 5 mL de coussin de sucrose à 30 % (BioRad 161-0720) dans un tampon SHAPE (0,1M KCl, 0,05M HEPES-KOH pH 8 @ RT, 5 mM MgCl₂). Les virus ont été purifiés et concentrés par ultracentrifugation (Beckmann Coulter Optima™ XE-90 Ultracentrifuge) dans des rotors SW 32Ti (Beckmann Coulter 369694) à 106559 × g pendant 120 minutes. Cinq cents microlitres de tampon SHAPE ont été ajoutés à 250 µL du culot viral dans chaque tube d'ultracentrifugation et incubés pendant 30 minutes à 4°C. Les culots de virus contenant le tampon SHAPE ont été regroupés et divisés en deux fractions aliquantes égales avant la modification de l'ARN. Les virus ont été conservés intacts dans l'une des fractions, tandis que dans l'autre, ils ont été désassemblés avant la modification de l'ARN.

Modification chimique de virus intacts

Pour la fraction des échantillons intacts, 250 µL de tampon SHAPE ont été ajoutés et incubés pendant 15 minutes à température ambiante suivi de 15 minutes à 37°C. L'échantillon intact a été divisé en volumes égaux pour le contrôle sans réactif (DMSO) et les modifications de test (1M7). Trois millilitres de DMSO anhydre (ThermoFisher-Invitrogen D12345) ont été utilisés pour préparer 100 mM de 1M7 (Sigma-Merck 908401). Les échantillons testés ont été modifiés avec du 1M7 à une concentration finale de 10 mM pendant 75 secondes à 37°C. Les

échantillons de contrôle sans réactif ont été traités avec du DMSO anhydre pendant 75 secondes à 37°C.

Modification chimique des particules désassemblées

Pour désassembler les particules virales, 250 µL de tampon de désassemblage 10X (tampon SHAPE 1X, 10 % (p/v) Triton X-100, 50 % (p/v) Glycérol, 10 mg.mL⁻¹ Lysolecithine) ont été ajoutés aux échantillons (seconde fraction) et incubés pendant 15 minutes à température ambiante suivies de 15 minutes à 37°C. Après le désassemblage du virus, les échantillons ont été divisés en deux volumes égaux pour les modifications du contrôle (DMSO) et du test (1M7). Trois millilitres de DMSO anhydre (ThermoFisher-Invitrogen D12345) ont été utilisés pour préparer 100 mM de 1M7 (Sigma-Merck 908401). Les échantillons testés ont été modifiés avec du 1M7 à une concentration finale de 10 mM pendant 75 secondes à 37°C. Les échantillons de contrôle sans réactif ont été traités avec les mêmes volumes de DMSO anhydre pendant 75 secondes à 37°C que les échantillons de test.

Profilage mutationnel

Un mg.mL⁻¹ de protéinase K (ThermoFisher E00491) a été ajouté à chaque échantillon et soigneusement mélangé, suivi de 3 volumes de RNA/DNA Shield (Ozyme ZR1100-250). Les échantillons ont été soigneusement mélangés et l'ARN a été extrait avec le Zymo RCC-5 (Zymo R1019) conformément aux instructions du fabricant. La transcription inverse a été réalisée dans des conditions de tampon spécifiques avec SuperScript II. Nous avons utilisé un mélange de 99X d'amorces nonamères aléatoires (S1254S de NEB) et 1X d'une amorce nonamère spécifique (d'IDT) complémentaire de l'extrémité 3' des ARN IAV. Les hybrides ARN-ADNc ont été purifiés avec des colonnes de gel Micro BioSpin P-6 (7326222 de BioRad). La synthèse du second brin d'ADN a été réalisée avec le module de synthèse du second brin d'ARN non directionnel NEBNext Ultra II (E611L de NEB). La purification de l'ADN double brin a été réalisée avec le kit NucleoSpin Gel and PCR Clean-Up (740609.250 de Macherey-Nagel) suivie d'une fragmentation de l'ADN double brin avec un sonicateur (Covaris E220 - Focused Ultrasonicator). Les bibliothèques d'ADNc ont été préparées et indexées en double avec le kit Microplex C05010007 de Diagenode. La sélection en double taille des bibliothèques a été effectuée avec des billes SPRISelect (Beckman Coulter) suivie d'un séquençage à haut débit sur un HiSeq 4000 (2×100 pb). Les échantillons ont été séquencés par la plateforme GenomEast, membre du consortium « France Génomique » (ANR-10-INBS-0009) à l'IGBMC, Strasbourg, France.

Shapemapper2

ShapeMapper automatise le calcul des réactivités de sondage chimique de l'ARN à partir d'expériences MaP. Le logiciel ne fonctionne que sur les systèmes Linux 64 bits et est disponible sur <https://github.com/Weeks-UNC/shapemapper2>.

Les lectures sont analysées de gauche à droite, et si une fenêtre de nucléotides --window-to-trim est trouvée avec un score phred moyen inférieur à --min-qual-to-trim, cette fenêtre est

tous les nucléotides à droite sont supprimés de la lecture. Elle est supprimée si la lecture restante est plus courte que `--min-length-to-trim`. Des lectures appariées ont été fournies en entrée (avec `--folder` ou avec `--R1 <r1.fastq> --R2 <r2.fastq>`). Les paires de lectures sont fusionnées à l'aide de Bbmerge v37.78 avec l'option `vstrict=t`. Les lectures non fusionnées ne sont pas supprimées. Les lectures fusionnées ont été alignées sur les séquences de référence avec Bowtie2 v2.3.4.3. Étant donné que nous avons utilisé des amorces aléatoires, le logiciel a également effectué une découpe aléatoire des amorces à l'aide de `--random-primer-len`, qui était de 9 en raison des amorces nonamères mentionnées précédemment. Les mutations chevauchant (`--random-primer-len + 1`) nucléotides à l'extrémité 3' des lectures ont été écartées, et (`--random-primer-len + 1`) nucléotides ont également été exclus de la contribution à la profondeur de lecture effective. `--min-seq-depth` La profondeur de lecture minimale requise pour inclure une mutation à une position donnée a été fixée à 2000. À chaque position dans un ARN donné, le taux de mutation pour chaque échantillon fourni est calculé comme le nombre de mutations divisé par la profondeur de lecture effective à cette position. La séquence en minuscules est exclue des profils de réactivité. Les positions avec une profondeur de lecture effective dans tout échantillon inférieur à `--min-depth 2000` ont été exclues. Si un échantillon de contrôle non traité avec un taux de mutation non traité supérieur au niveau de fond maximal `--max-bg` (par défaut = 0,05) était également exclu.

Les erreurs standard du taux de mutation sont calculées comme $(\sqrt{\text{taux de mutation}})/(\sqrt{\text{profondeur effective}})$. Étant donné que nous avons utilisé des échantillons de DMSO modifiés et non traités, les taux de mutation ont été calculés comme $[\text{taux}]_{\text{modifié}} - [\text{taux}]_{\text{non traité}}$. Les erreurs standard de réactivité ont été calculées comme $\sqrt{(\text{stderr} [\text{modifié}]^2 + \text{stderr} [\text{non traité}]^2)}$. Par défaut, ShapeMapper normalise tous les profils ensemble en utilisant l'ensemble combiné de réactivités pour calculer un facteur de normalisation unique qui est appliqué à tous les profils. Cela a été désactivé en passant l'option `--indiv-norm`.

Les tracés de sortie nommés `*_profiles.pdf` incluent des panneaux pour les profondeurs de lecture et les taux de mutation. Les erreurs standard de réactivité sont affichées sous forme de barres d'erreur dans les tracés `*_profile.pdf`. Les taux de réactivité moyens \pm écart type des réplicats avec les coefficients de corrélation les plus élevés (fig.) ont été utilisés pour créer des profils de réactivité de chaque segment d'échantillon avec Shapemapper2. Les figures ont été conçues avec RNAVigate94 par Mme Chane Woong-Min.

Δ SHAPE

Δ SHAPE présente une approche statistiquement rigoureuse pour analyser les changements de réactivité des nucléotides d'échantillons entre deux conditions. Il est disponible sur <https://github.com/Weeks-UNC/deltaSHAPE>. Nous avons utilisé Δ SHAPE pour comparer les réactivités des nucléotides entre les conditions intactes et désassemblées. Il compare les différences de réactivité par rapport aux erreurs associées, et compare également l'ampleur

de chaque différence par rapport à chaque autre changement de réactivité. Les nucléotides qui présentent des changements forts et significatifs de réactivité sont ainsi identifiés. Δ SHAPE est exécuté sur Python 2.7 avec deux modules supplémentaires matplotlib et NumPy.

vRNAsite

Le flux de travail de vRNAsite peut être séparé en six étapes principales : indexation, pliage, calcul de la moyenne, extraction, découpage et traçage. Le code, le supplément et la documentation de vRNAsite sont open source et sont accessibles à l'adresse <https://github.com/desiro/vRNAsite>. L'entrée requise est un fichier FASTA multiple représentant les différents segments de l'IAV, précédé du paramètre de commande --fasta. Les paramètres --reverse et --complement construisent le complément inverse de la séquence d'entrée. Le paramètre --reversePositions inverse la sortie de l'IRI trouvé sur le brin de sens négatif pour fournir les positions nt du brin de sens positif.

Pour extraire les IRI des cartes de contact, vRNAsite utilise les positions nucléotidiques de début et de fin de l'IRI potentiel des deux séquences d'ARN d'entrée dans ce qui suit, appelées limites extérieures d'une zone de contact. Les zones de contact sont les zones qui contribuent à un pic d'énergie libre moyenne sur la carte de contact IRI. Deux ou plusieurs zones de contact peuvent se chevaucher si plusieurs IRI possibles partagent les mêmes positions nucléotidiques sur les deux segments d'ARN en interaction. vRNAsite utilise un algorithme de segmentation par bassin versant pour détecter et séparer les zones de contact les unes des autres. Les parties de séquence en interaction ont été extraites et pliées avec l'algorithme RNACofold du ViennaRNA Package.

Conclusions et perspectives

Cette étude visait à étudier les modèles de croissance et les structures secondaires de l'ARN de H1N1/PR8, H3N2/MO et de deux réassortants à segment unique R1 (PR8_PB2MO) et R6 (PR8_NAMO) produits par la génétique inverse 7 : 1 à partir de PR8 et MO. Nous avons utilisé SHAPE-MaP pour étudier les structures secondaires de l'ARN dans les particules virales intactes et démontées de PR8, R1 et R6.

Les réassortants R1 et R6 ont été choisis avec les souches parentales PR8 et MO pour des études plus approfondies basées sur leurs propriétés de croissance. Les courbes de croissance de R1 et R6 étaient similaires à celles des souches parentales PR8 et MO, indiquant l'aptitude du virus. Les six réassortants restants étaient atténués. Cette découverte suggère que les segments H3N2/MO PB2 et NA et leurs protéines correspondantes sont compatibles avec H1N1/PR8. Cependant, nous avons besoin d'expériences supplémentaires pour déterminer si l'inverse (c'est-à-dire si les protéines individuelles H3N2/MO sont compatibles dans un contexte H1N1/PR8) sera vrai. Par exemple, nous pourrions réaliser des expériences de génétique inverse avec le H3N2/MO comme épine dorsale et vérifier les schémas de croissance des réassortants produits. Il s'agirait d'une orientation prospective pour étudier les contraintes génétiques potentielles qui limitent le réassortiment entre les souches H1N1 et H3N2 de l'IAV.

De plus, nous n'avons effectué de génétique inverse qu'avec des combinaisons de segments 7 : 1 dans cette étude. Puisqu'il existe de nombreuses preuves de co-ségrégation de segments dans IAV, nous pourrions effectuer une alternative 5:3 (PB2, PB1, PA non apparentés), 6:2 (HA, NA ou NP, M, etc. non apparentés) ou 4 : 4 combinaisons de segments pour étudier le réassortiment à différentes capacités. Certains réassortants produits grâce à ces combinaisons surpassent probablement les cinétiques de croissance de R1 et R6 obtenues dans cette étude. Par conséquent, nous pourrions améliorer notre compréhension des facteurs potentiels spécifiques à chaque segment pour le réassortiment entre les souches H1N1 et H3N2.

Nous avons utilisé les réassortants à segment unique R1 et R6 dans des contextes de co-infection compétitifs pour rechercher une compatibilité sous-optimale entre les segments non apparentés et apparentés dans les réassortants. Nous avons trouvé des résultats très intrigants à partir de cet ensemble d'expériences. Premièrement, bien que PR8_HAtag ne différât du PR8 parental que par une mutation synonyme, le segment HAtag semblait avoir une meilleure efficacité d'empaquetage dans les trois conditions des expériences de co-infection. D'autres études sur la capacité de répllication de ce virus par rapport au PR8 pourraient expliquer ce phénomène intrigant. Il serait intéressant de vérifier si effectivement l'introduction d'une mutation synonyme dans le segment PR8_HA améliore l'efficacité du packaging du segment. Cela pourrait avoir d'autres implications potentielles sur la nature

supposée de la covariation dans les segments d'ARN et son impact sur l'empaquetage du génome de l'IAV. Deuxièmement, nous avons observé quelques génotypes de virus réassortis comportant plus de huit segments typiques des IAV. Bien que ce phénomène ne soit pas rare et ait déjà été rapporté, dans cette étude, nous n'avons pas pu distinguer ce phénomène de la présence d'agrégats multiparticulaires au sein d'une plaque ou de l'incorporation de plus de huit segments au sein d'un virion. En conséquence, exclure la possibilité d'agrégats multiparticulaires et vérifier la présence de plus de huit segments au sein d'un virus serait une voie passionnante pour des études plus approfondies.

Après des tests de co-infection, nous avons utilisé des surnageants de virus purifiés pour les modifier avec le réactif SHAPE 1M7 pour un sondage chimique suivi d'un profilage mutationnel. Il s'agit d'une étude SHAPE-MaP unique sur les réassortants IAV en plus du projet collaboratif FluCode entre notre laboratoire et celui du professeur Martin Schwemmler de l'Université de Fribourg. Dans cette étude, nous avons comparé les structures secondaires d'ARN de particules virales IAV intactes et démontées avec Δ SHAPE. Contrairement à notre hypothèse, nous avons observé un nombre plus élevé de régions Δ SHAPE négatives dans tous les segments, ce qui indique des structures vRNP plus détendues dans les particules intactes. Nous avons également observé un remodelage structurel local et global important des vRNP lors du désassemblage des particules virales, ce qui nous amène à supposer que nous avons besoin de meilleurs outils pour étudier de manière approfondie le réassortiment génétique et l'empaquetage du génome. La reproductibilité des données de sondage chimique est assez délicate, et le fait que nous ayons utilisé des particules virales intactes pour sonder le vRNP flexible et dynamique de l'IAV a rendu l'inférence directe des réactivités nucléotidiques et de leurs implications sur la nature appariée/simple brin de l'ARNv un défi. Cela pourrait peut-être être atténué en complétant cet ensemble de données avec DMS-MaP afin de mieux comprendre les interactions potentielles ARN-NP et de les distinguer des interactions ARN-ARN identifiées par SHAPE-MaP.

Nos résultats SHAPE-MaP suggèrent qu'il pourrait y avoir un remodelage structurel local du vRNP entre les états intact et démonté. La nature flexible du vRNP facilite les processus clés de la réplication du génome IAV. De plus, on sait que les ARNv conditionnés dans des particules virales sont plus compacts que les autres molécules d'ARN. On peut sans doute avoir des raisons de supposer que lors de l'empaquetage du génome des vRNP IAV, celles-ci subissent un remodelage structurel local pour exposer des signaux d'empaquetage qui facilitent le regroupement de segments et l'assemblage progressif. Cependant, il peut y avoir de meilleures solutions que d'utiliser SHAPE-MaP seul pour étudier cela et cela nécessiterait des techniques biophysiques à haute résolution, de préférence dans un format de particule virale unique.

Il est peut-être moins difficile d'étudier d'abord l'empaquetage du génome dans des souches individuelles telles que H1N1/PR8, H3N2/Udorn, etc., puis d'utiliser ce répertoire de

connaissances pour éclairer les règles d'empaquetage lors du réassortiment. Étant donné que l'empaquetage du génome de l'IAV est toujours rigoureux et sélectif, avoir une compréhension fondamentale des règles d'empaquetage du génome avant les études de réassortiment peut s'avérer utile pour combler certaines des lacunes existantes dans les connaissances. Certes, la redondance présentée par le réseau d'empaquetage du génome IAV rend cela plus difficile qu'il n'y paraît. Néanmoins, si nous identifions des structures d'ARN putatives et validons leur pertinence biologique dans l'empaquetage du génome tout en montrant que l'absence de ces interactions peut provoquer des défauts d'empaquetage dans les virus, nous prenons un départ prometteur.

Différents laboratoires ont développé une suite diversifiée d'outils bioinformatiques et d'analyse fonctionnelle et, collectivement, ces efforts soutiennent la validation de nombreux éléments structurels fonctionnels de l'ARN dans les génomes viraux¹⁸⁸. Néanmoins, il reste difficile d'identifier le sous-ensemble de motifs fonctionnellement importants dans le contexte d'un repliement étendu de la structure secondaire. La mise en œuvre d'analyses fonctionnelles virales prend du temps et est souvent inefficace, ce qui a limité le nombre d'éléments ayant fait l'objet d'études fonctionnelles¹⁸⁸. De nouvelles stratégies à plus haut débit pour l'analyse fonctionnelle virale sont absolument nécessaires¹⁸⁸. Nous avons également besoin de nouvelles stratégies informatiques utilisant des signaux d'empaquetage fonctionnels précédemment identifiés comme modèles d'apprentissage pour étudier les modèles de motifs d'ARN putatifs ayant des implications en matière d'empaquetage. Comprendre le mécanisme moléculaire de l'empaquetage du génome est crucial car il ouvre des voies prometteuses pour la recherche antivirale.

Pour conclure, nous n'avons pas délimité les interactions intersegmentaires ARN-ARN qui interviennent dans l'empaquetage du génome dans cette étude. Bien qu'un certain ensemble d'interactions soient indispensables à l'empaquetage du génome et que leur absence puisse conduire à une perte absolue de segments, nous devons peut-être soupçonner que la proximité de souches génétiquement similaires dans un contexte de co-infection peut élargir le répertoire de motifs structurels fonctionnels d'ARN, augmentant ainsi le potentiel de réassortiment tout en diversifiant simultanément les génotypes des réassortiments de descendance. On peut supposer qu'au lieu d'un ensemble discret fini d'interactions intersegmentaires ARN-ARN, nous pourrions envisager un ensemble dynamique, discontinu et redondant de régions qui présentent une hiérarchie dans l'empaquetage. Un ensemble de signaux d'empaquetage spécifiques de haute affinité couplés à des interactions transitoires de faible affinité médiées par l'ARNv et le NP et l'ARNv agissant séparément ou en synergie médient l'empaquetage du génome. Il pourrait y avoir un ensemble dynamique de signaux d'empaquetage de haute et de faible affinité spécifiques à chaque souche. Cet ensemble d'interactions peut garantir l'empaquetage sélectif de chaque vRNP tout en offrant la flexibilité d'incorporer des segments non apparentés en cas de co-infection. Cela garantit un potentiel de réassortiment dans des conditions favorables lorsque des souches

génétiqnement homologues sont infectées à proximité. Par conséquent, il est prudent de supposer que même si l'étude du réassortiment génétique de l'IAV est plus compliquée que prévu, chaque nouvel élément d'information s'ajoute à la base de données d'informations existante tout en créant de nouvelles pistes de recherche potentielles.