Analysis of Higher Order RNA Structures in the Influenza A Virus Genome

Bernadeta Dadonaite



Thesis submitted for the degree of Doctor of Philosophy

Sir William Dunn School of Pathology

Magdalen College

University of Oxford

Trinity term 2018

Table of Contents

TABLE OF CONTENTS **1**

TABLE OF FIGURES AND TABLES **5**

ABSTRACT **8**

DECLARATION **9**

ACKNOWLEDGMENTS **10**

ABBREVIATIONS **11**

1. CHAPTER 1: INTRODUCTION **14**

1.1 INFLUENZA VIRUS **14**

1.1.1 INFLUENZA VIRUS - A HISTORICAL PERSPECTIVE 14

1.1.2 INFLUENZA VIRUS CLASSIFICATION 15

1.1.3 INFLUENZA VIRUS STRUCTURE 16

1.1.4 INFLUENZA VIRUS LIFE CYCLE 18

1.2 RIBONUCLEOPROTEIN STRUCTURE **22**

1.2.1 RNA-DEPENDENT RNA POLYMERASE STRUCTURE 22

1.2.2 NUCLEOPROTEIN STRUCTURE 24

1.2.3 VRNA STRUCTURE 25

1.2.4 RNP ASSEMBLY 26

1.3 INFLUENZA VIRUS GENOME PACKAGING **28**

1.3.1 SELECTIVE GENOME PACKAGING 28

1.3.2 PACKAGING SIGNALS IN THE VRNA 28

1.3.3 INTER-SEGMENT RNA-RNA INTERACTIONS 30

1.3.4 VRNP INCORPORATION DURING VIRUS BUDDING 30

1.4 INFLUENZA VIRUS REASSORTMENT **31**

1.4.1 INFLUENZA VIRUS PANDEMICS OF THE PAST CENTURY 33

1.4.2 GENETIC FACTORS CONTRIBUTING TO VIRUS REASSORTMENT 37

1.5 SECONDARY RNA STRUCTURE PREDICTION AND MODELLING **39**

1.5.1 COMPUTATIONAL METHODS USED TO STUDY SECONDARY RNA STRUCTURES 39

1.5.2 EXPERIMENTAL APPROACHES TO SECONDARY RNA STRUCTURE PREDICTIONS 40

1.5.3 SECONDARY RNA STRUCTURES IN VIRUSES 43
1.6 THESIS AIMS 45

2. CHAPTER 2: MATERIALS & METHODS 46

2.1 PLASMIDS AND CLONING 46
2.1.1 PLASMIDS 46
2.1.2 PLASMID MAINTENANCE 46
2.1.3 PCR MUTAGENESIS 47
2.1.4 GIBSON ASSEMBLY 47
2.1.5 PRIMERS USED FOR MUTAGENESIS 49
2.2 PROTEIN ANALYSIS 50
2.2.1 SDS POLYACRYLAMIDE GEL ELECTROPHORESIS FOR VIRAL PROTEINS 50
2.3 RNA ANALYSIS 51
2.3.1 OLIGONUCLEOTIDE RADIOLABELLING 51
2.3.2 RNA POLYACRYLAMIDE GEL ELECTROPHORESIS 51
2.3.3 RNP RECONSTITUTIONS 51
2.3.4 PRIMER EXTENSION ANALYSIS 52
2.3.5 RT-QPCR FOR VRNA SEGMENTS 52
2.4 EUKARYOTIC CELL AND VIRUS CULTURE 55
2.4.1 EUKARYOTIC CELL CULTURE 55
2.4.2 VIRUSES 55
2.4.3 VIRUS AMPLIFICATION AND CELL INFECTION 56
2.4.4 VIRUS GROWTH CURVE ANALYSIS 56
2.4.5 PLAQUE ASSAYS 56
2.4.6 HEAMAGGLUTINATION ASSAY 57
2.4.7 VIRUS REVERSE GENETICS 57
2.4.8 VIRUS PURIFICATION BY ULTRACENTRIFUGATION 57
2.4.9 VIRUS GROWTH ANALYSIS IN THE PRESENCE OF ANTISENSE OLIGONUCLEOTIDES 58
2.5 SELECTIVE 2' -HYDROXYL ACYLATION ANALYZED BY PRIMER EXTENSION AND MUTATIONAL PROFILING (SHAPE-MAP) 58
2.5.1 VRNA SAMPLE PREPARATION 58
2.5.2 1M7 VRNA MODIFICATION *EX VIRIO* 60
2.5.3 1M7 AND NMIA VRNA MODIFICATION *IN VIRIO* 60
2.5.4 SHAPE-MAP LIBRARY PREPARATION 61
2.5.5 SHAPE-MAP SEQUENCING READ ANALYSIS 62
2.5.6 SECONDARY RNA STRUCTURE PREDICTION 62
2.6 SEQUENCING OF PSORALEN CROSSLINKED, LIGATED, AND SELECTED HYBRIDS (SPLASH) 63

2.6.1 Cross-linking 63
2.6.2 Size-selection 64
2.6.3 Cross-linked RNA enrichment 64
2.6.4 On-bead proximity ligation 64
2.6.5 AMT-B reverse crosslinking 65
2.6.6 SPLASH library and sequencing 66
2.6.7 SPLASH sequencing read analysis 67

3. CHAPTER 3: ESTABLISHMENT OF SHAPE-MAP FOR INFLUENZA VIRUS 68

3.1 Overview 68
3.2 Background 68
3.3 Aims 70
3.4 Results 71
3.4.1 *In virio* SHAPE modification 71
3.4.2 SHAPE-MaP reverse transcription 74
3.4.3 *Ex virio* SHAPE modification 76
3.4.4 SHAPE-MaP library preparation, sequencing and analysis 77
3.4.5 Discussion 81
3.4.6 Summary 82

4. CHAPTER 4: ANALYSIS OF THE INTRAMOLECULAR SECONDARY RNA STRUCTURES PRESENT IN THE INFLUENZA VIRUS GENOME 83

4.1 Overview 83
4.2 Background 83
4.3 Aims 86
4.4 Results 86
4.4.1 Different WSN genome segments have unique RNA conformations 86
4.4.2 Comparison between *in virio* and *ex virio* vRNA SHAPE reactivities 90
4.4.3 NP periodicity analysis 92
4.4.4 Secondary RNA structure modelling for the WSN genome 93
4.4.5 Functional roles of the secondary RNA structures in the WSN virus genome 99
4.4.6 SHAPE-MaP analysis of PR8, Udorn and FLB virus genomes 106
4.5 Discussion 119
4.6 Summary 125

5. CHAPTER 5: INTER-SEGMENT INTERACTION ANALYSIS OF THE INFLUENZA VIRUS RNA GENOME 126

5.1 OVERVIEW 126
5.2 BACKGROUND 126
5.3 AIMS 128
5.4 RESULTS 129
5.4.1 *IN VIRIO* VRNA CROSS-LINKING 129
5.4.2 *IN VIRIO* SPLASH LIBRARY AND SEQUENCING 132
5.4.3 SPLASH SEQUENCING AND BIOINFORMATICS 135
5.4.4 WSN INTER-SEGMENT RNA INTERACTION MAP 139
5.4.5 MODELLING THE STRUCTURES OF THE INTER-SEGMENT RNA INTERACTIONS 142
5.4.6 INTER-SEGMENT RNA INTERACTION ANALYSIS FOR PR8, UDORN AND FLB VIRUSES 144
5.5 DISCUSSION 149
5.6 SUMMARY 155

6. CHAPTER 6: ANALYSIS OF THE IMPORTANCE OF THE INTER-SEGMENT RNA INTERACTIONS IN THE INFLUENZA WSN VIRUS GENOME 156

6.1 OVERVIEW 156
6.2 BACKGROUND 156
6.3 AIMS 158
6.4 RESULTS 158
6.4.1 DESIGN OF INTER-SEGMENT INTERACTION MUTANTS 158
6.4.2 EFFECTS OF THE DISRUPTION OF THE INTER-SEGMENT RNA INTERACTIONS ON THE GROWTH OF THE WSN VIRUS 160
6.5 DISCUSSION 166
6.6 SUMMARY 169

7. CHAPTER 7: CONCLUSIONS AND FUTURE DIRECTIONS 170

REFERENCES 179

APPENDIX A: WSN SHAPE REACTIVITIES AND SUMMARY STATISTICS 201

APPENDIX B: SHAPE-MAP ANALYSIS SCRIPTS 243

APPENDIX C: SPLASH READ PROCESSING SCRIPTS 245

Table of figures and tables

Figure 1.1 Structure of influenza virus 17
Table 1.1 Proteins encoded by the influenza A virus genome 18
Figure 1.2 Influenza virus life cycle 22
Figure 1.3 Structure of influenza virus RNPs 27
Figure 1.4 IAV pandemics of the last century 36
Table 1.2 Examples of genome-wide methods used for higher-order RNA structure analysis 41
Table 2.1 PCR mutagenesis reaction 47
Table 2.2 Phusion HF reaction for plasmid backbone amplification 48
Table 2.3 Primers used for producing viruses with mutated secondary RNA structures in SHAPE-MaP experiments 49
Table 2.4 Primers used for producing viruses with mutated inter-segment interactions in SPLASH experiments 50
Table 2.5 Reverse transcription primers 53
Table 2.6 Reverse transcription reaction components 53
Table 2.7 qPCR master-mix 54
Table 2.8 qPCR primers and probes 54
Table 2.9 Media used for cell culture 55
Table 2.10 Primers used to produce linear templates of WSN genome segments and to add T7-Pol promoters 59
Table 2.11 HiScribe T7 reaction 60
Table 2.12 SPLASH buffers 63
Table 2.13 T4 PNK cyclic phosphate removal reaction 65
Table 2.14 T4 PNK 5′-phosphorylation addition reaction 65
Table 2.15 Rnl1 proximity ligation reaction 65
Figure 3.1 WSN virus purification 71
Figure 3.2 In virio SHAPE modification 73
Figure 3.3 Reverse transcription in the presence of Mn^{2+} 75
Figure 3.4 ivtRNA samples made using T7 RNA polymerase 76
Figure 3.5 SHAPE-MaP sample summary 77
Figure 3.6 SHAPE reactivity analysis 78
Figure 3.7 Probability density distribution of the mutation rates in 1M7 and DMSO-treated samples 79
Figure 3.8 Reproducibility of the SHAPE-MaP experiments 80
Figure 4.1 Median SHAPE reactivity profiles for all WSN genome segments 87

Figure 4.2 Variation in SHAPE reactivities within segments 88
Figure 4.3 SHAPE reactivity distribution in the 5′ end of the vRNA 89
Figure 4.4 Sliding window correlation between in virio and ex virio SHAPE reactivities of vRNA 92
Figure 4.5 NP Periodicity analysis 92
Figure 4.6 In virio secondary RNA structures in the WSN genome 94
Figure 4.7 ivtRNA secondary RNA structures 95
Figure 4.8 nkvRNA secondary RNA structures 96
Figure 4.9 Comparison between in virio and ex virio vRNA structures 97
Figure 4.10 Secondary RNA structures found in the NS segment in virio 98
Figure 4.11 SHAPE profiling recapitulates previously described secondary RNA structures 99
Figure 4.12 Growth analysis of the WSN viruses with disrupted secondary RNA structures in HA and PB2 segments 101
Figure 4.13 Growth analysis of the WSN viruses with disrupted secondary RNA structures in PB1 segment 102
Figure 4.14 Growth analysis of the WSN viruses with disrupted hairpin in NS segment 103
Figure 4.15 Analysis of vRNA packaging defects 104
Figure 4.16 Modelled secondary RNA structures in NP and M segments 104
Figure 4.17 ASO targeting of secondary RNA structures in the WSN genome 105
Figure 4.18 SHAPE reactivity normalisation 106
Table 4.1 SHAPE reactivity correlation and sequence similarity between H1N1 and H3N2 strains 107
Figure 4.19 SHAPE-MaP analysis for the PB2 segment of different IAV strains 108
Figure 4.20 SHAPE-MaP analysis for the PB1 segment of different IAV strains 109
Figure 4.21 SHAPE-MaP analysis for the PA segment of different IAV strains 110
Figure 4.22 SHAPE-MaP analysis for the HA segment of different IAV strains 111
Figure 4.23 SHAPE-MaP analysis for the NP segment of different IAV strains 112
Figure 4.24 SHAPE-MaP analysis for the NA segment of different IAV strains 113
Figure 4.25 SHAPE-MaP analysis for the M segment of different IAV strains 114
Figure 4.26 SHAPE-MaP analysis for the NS segment of different IAV strains 115
Figure 4.27 SHAPE-MaP analysis for FLB genome 117
Figure 4.28 Comparison between the SHAPE reactivities of the nucleotides located in the 5′ versus the 3′ region of the vRNA 118
Figure 4.29 Comparison between SHAPE-MaP and NP footprinting analysis 121
Figure 5.2 RNA mobility shift after AMT cross-linking 129
Figure 5.3 Cross-linking vRNA with AMT-B 131
Figure 5.4 AMT-B crosslink reversal 132

Figure 5.5 SPLASH library RNA cross-linking and ligation steps 133
Figure 5.6 SPLASH library PCR amplification and size selection products 134
Figure 5.7 cDNA preparation for SPLASH sequencing libraries 135
Figure 5.8 SPLASH sequencing and bioinformatic analysis 136
Figure 5.9 Summary of SPLASH data from three experimental replicates 137
Figure 5.10 Sequencing read analysis and the modelling of RNA interactions 139
Table 5.1 Inter-segment RNA-RNA interactions in the WSN virus genome 140
Figure 5.11 WSN inter-segment interaction map 141
Figure 5.12 Position of the inter-segment RNA interactions along the length of the vRNA 142
Table 5.2 Inter-segment interaction coordinates after IntaRNA modelling and the associated ΔG energies 143
Figure 5.13 Modelled inter-segment RNA interactions 144
Table 5.3 Inter-segment RNA-RNA interactions in the PR8 virus genome 145
Table 5.4 Inter-segment RNA-RNA interactions in the Udorn virus genome 145
Table 5.5 Inter-segment RNA-RNA interactions in the FLB virus genome 146
Figure 5.14 Inter-segment RNA interactions in the PR8 virus genome 147
Figure 5.15 Inter-segment RNA interactions in the Udorn virus genome 148
Figure 5.16 Inter-segment RNA interactions in the FLB virus genome 149
Figure 5.17 Different shapes of interaction loci 151
Figure 5.18 Common inter-segment RNA interactions and interaction loci in different IAVs 154
Figure 6.1 Summary of the viruses made for the mutagenesis of the inter-segment interactions 159
Figure 6.2 Individual-segment mutagenesis 161
Figure 6.3 Individual-segment mutagenesis for NS-M interaction 162
Figure 6.4 Defective particles produced by viruses with mutated inter-segment interactions 163
Figure 6.5 Extensive loci and individual segment mutagenesis 163
Figure 6.6 Restorative mutagenesis design 165
Figure 7.1 vRNP assembly model 172
Figure 7.2 SHAPE reactivities in inter-segment interaction sites 174
Figure 7.3 Long distance intramolecular RNA-RNA interactions 176

Abstract

Influenza viruses are respiratory pathogens, which constitute a major threat to public health. Up to 650,000 deaths and 5 million severe cases caused by seasonal influenza strains are recorded each year and potential pandemic influenza outbreaks pose a constant risk to the human population worldwide. Influenza viruses belong to the family *Orthomyxoviridae* and contain single stranded negative sense segmented RNA genomes. Each viral genome segment is maintained in a viral ribonucleoprotein complex (vRNP), which consists of the viral polymerase, oligomeric nucleoprotein and the viral RNA (vRNA). Little is known about the conformation of the vRNA in the context of the vRNP, although it has been proposed that secondary RNA structures in the vRNPs are essential for successful packaging of the influenza virus genome. The aim of the work presented in this thesis was to analyse the global secondary RNA structure of the influenza virus genome. To this end selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) method was used to perform secondary RNA structure analysis on naked and vRNP-associated vRNA. The analysis showed that multiple secondary RNA structures are present in the genomes of influenza viruses, each vRNA segment has a unique secondary RNA structure fold, and no clear NP binding footprint was observed in the SHAPE reactivities. The presence of the structured vRNA sites suggested that some secondary RNA structures may be accessible to form interactions between vRNPs, therefore, further analysis using sequencing of psoralen crosslinked, ligated, and selected hybrids (SPLASH) was performed to detect inter-segment RNA-RNA interactions. A redundant network of RNA-RNA interactions was detected to be present inside influenza virions and mutagenesis work was performed to show that disruption of inter-segment RNA interactions can lead to defects in the virus genome packaging and attenuated virus growth. The relevance of this work to the reassortment between different influenza strains and the emergence of new pandemic strains is discussed.

Declaration

I declare that the work presented in this thesis was performed by the author. Any contributors to the specific parts of this work are acknowledged in the text.

Bernadeta Dadonaite

August, 2018

Acknowledgments

I would firstly like to thank my DPhil supervisor, Professor Ervin Fodor, for not only welcoming me into his lab but for continuously taking leaps of faith and trusting me to follow my ideas. I have learned a lot from Ervin, both in science and personal relationships, and I know I will carry those lessons with me in the future to come.

Secondly, my DPhil would not have been what it was if not for Dr. David L.V. Bauer. David's continued optimism, help and attitude towards work and life overall, have helped me to move through the many ups and downs of the DPhil and showed me what a perfect teamwork can be. David has become my mentor and friend and a person I can only look up to.

The members of the Fodor lab have provided an invaluable support throughout my DPhil not only through scientific discussions but also by creating a great atmosphere in the lab, which was and is greatly appreciated. I would also like to thank Professor Alain Laederach for scientific support throughout my DPhil work and the many discussions, which have been a valuable learning tool for stepping into an unfamiliar scientific field.

I would not be the person I am if not for my family and their support, they have been the shoulder I cried and laughed on, have never doubted me and always supported my endeavours in life.

Finally, Mariya Lobanovska for being a good person and a better friend.

Abbreviations

1M6	1-methyl-6-nitroisatoic anhydride
1M7	1-methyl-7-nitroisatoic anhydride
3P	Viral polymerase
AMT	4′-Aminomethyltrioxsalen hydrochloride
APS	Ammonium persulphate
ASO	Antisense oligonucleotide
ATP	Adenosine triphosphate
CMCT	N-cyclohexyl-N′-(2-morpholinoethyl) carbodiimide metho-p-toluenesulfonate
CRE	cAMP response element
cRNP	Complementary ribonucleoprotein
CTD	C-terminal domain
DI	Defective interfering
DMEM	Dulbecco's Modified Eagle Medium
DMS	Dimethyl sulfate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron Microscopy
ER	Endoplasmic reticulum
FCS	Foetal Calf Serum
FISH	Fluorescence in situ hybridisation
FLB	B/Florida/4/06
GFP	Green fluorescent protein
HA	Haemagglutinin
HB	Hybridisation buffer
HF	High fidelity
HIV-1	Human immunodeficiency virus 1
HPAIV	High pathogenicity avian influenza virus
IAV	Influenza A virus
IBV	Influenza B virus
ICV	Influenza C virus
IDV	Influenza D virus
IRES	Internal ribosome entry site
LB	Lysis buffer
LIGR	Ligation of interacting RNA
LPAIV	Low pathogenicity avian influenza virus
MARIO	Mapping RNA interactome in vivo
MDBK	Madin-Darby Bovine Kidney
MDCK	Madin-Darby Canine Kidney

MEA	Minimum expected accuracy
MEM	Minimum essential medium
MFE	Minimum folding energy
MHV	Mouse hepatitis virus
MMLV	Murine leukemia virus
MOI	Multiplicity of infection
NA	Neuraminidase
NEP	Nuclear export protein
NES	Nuclear export signal
NMIA	N-methylisatoic anhydride
NP	Nucleoprotein
ORF	Open reading frame
PA	Polymerase acidic protein
PAGE	Polyacrylamide Gel Electrophoresis
PARIS	Psoralen analysis of RNA interactions and structures
PB1	Polymerase basic protein 1
PB2	Polymerase basic protein 2
PKB	Protein kinase buffer
PNK	T4 Polynucleotide Kinase
PR8	A/Puerto Rico/8/1934(H1N1)
qPCR	Quantitative OCR
RIG-I	Retinoic acid-inducible gene I
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
rpm	Revolutions per minute
RPM	Reads per million
RQ	Relative quantity
RT	Reverse transcription
SA	Sialic acid
SDS	Sodium dodecyl sulphate
SHAPE	Selective 2′-hydroxyl acylation analysed by primer extension and mutational profiling
SOC	Super optimal broth with catabolite repression
SPLASH	Sequencing of psoralen crosslinked, ligated, and selected hybrids
STMV	Satellite tobacco mosaic virus
TBE	Tris, Boric acid, EDTA
TEMED	Tetramethylethylenediamine
Udorn	A/Udorn/307/72(H3N2)
UTR	Untranslated region
UV	Ultraviolet
vmRNA	Viral messenger RNA
vRNP	Viral ribonucleoprotein

WB	Wash buffer
WSN	A/WSN/1933(H1N1)
WT	Wild type

1. Chapter 1: Introduction

1.1 Influenza virus

1.1.1 Influenza virus - a historical perspective

Influenza viruses are respiratory pathogens that cause epidemic outbreaks in most countries each year. The name *Influenza* comes from an Italian word meaning *to influence* and was first used during an unknown illness outbreak in the 14th century in the context of "*influenza di freddo*" - the influence of the cold (OED, 2017). Historical descriptions of influenza-like illnesses date back as far as the ancient Greek period, with a description of coughing outbreak found in the *Sixth Book of the Epidemics* written by the Hippocrates in ca. 400 BCE (Lina, 2008). Over the last two millennia multiple written records exists which are suggestive of pandemic influenza outbreaks, though perhaps the first clear description of an influenza pandemic comes from 1510, when an African-origin virus caused a wide-spread disease outbreak in Europe (Taubenberger & Morens, 2009). The influenza illness was initially thought to be caused by a bacterium *Bacillus influenzae* (now referred to as *Haemophilus influenzae*) when Richard Pfeiffer isolated a bacillus from patients with influenza-like illness (Pfeiffer, 1892). For several decades afterwards, a bacterium was thought to be the causative agent of influenza and support for this observation also came during the 1918 Spanish flu outbreak, as samples from many Spanish flu patients were found to contain *B. influenzae* in them. However, in the 1930s, Richard Shope showed that a filterable agent could cause an influenza-like illness in pigs (Lewis & Shope, 1931), which was exacerbated by the presence of bacteria (Shope, 1931), and so the first indication of virus-caused disease emerged. Subsequent to Shope's experiments, in 1933 Smith et *al.* filtered throat-washings from influenza patients and showed that they can be successfully passaged and cause respiratory disease in ferrets (Smith, Andrewes & Laidlaw, 1933), thus satisfying Koch's postulates and firmly proving the viral origin of the disease. Subsequent experiments also showed that the sera from patients of the 1918 influenza outbreak could neutralize the same virus that caused the influenza-like illness in pigs in Shope's experiments, suggesting the that two illnesses were likely caused by the same pathogen and already indicating the wider host range of the virus. Over the years, influenza virus has become one of the most widely studied viral pathogens, and its research has had a wide impact on the understanding of multiple other aspects of molecular biology including the

discovery of interferons (Isaacs & Lindenmann, 1957) and cell-mediated immune defenses (Effros, 1977; Zweerink, Courtneidge, Skehel, *et al.*, 1977).

1.1.2 Influenza virus classification

Influenza viruses are members of the *Orthomyxoviridae* family, which includes *Influenzavirus A, Influenzavirus B, Influenzavirus C, Influenzavirus D, Isavirus*, *Thogotovirus*, and *Quaranjavirus* genera. All orthomyxoviruses have single stranded negative sense segmented RNA genomes, which are replicated inside the nucleus of a cell. The genomes of influenza A (IAV) and B (IBV) viruses contain 8 segments, while the genomes of influenza C (ICV) and D (IDV) viruses are divided into 7 segments. Of the four influenza virus genera IAVs have the widest host range, which includes humans, chickens, pigs, horses, cats, dogs, whales, seals, bats and waterfowl, the latter being the major reservoir (Taubenberger & Kash, 2010). IBVs and ICVs are predominantly found in humans, though occasional isolates from other hosts such as whales and pigs have been reported (Guo, Jin, Wang, *et al.*, 1983; Hinshaw, Bean, Geraci, *et al.*, 1986; Bodewes, Morick, de Mutsert, *et al.*, 2013). The most recently discovered IDVs are primarily found in cattle and pigs (Chiapponi, Faccini, De Mattia, *et al.*, 2016; Foni, Chiapponi, Baioni, *et al.*, 2017). Influenza virus infections in non-human hosts are often asymptomatic, whereas in humans they usually cause respiratory tract illness of varied severity - ICVs cause only mild respiratory infections, whereas IAVs and IBVs viruses are responsible for seasonal epidemics that can lead to as many as 3-5 million severe cases of illness and between 290,000 and 650,000 deaths annually worldwide (WHO, 2017). IAVs are further subdivided based on their surface glycoproteins - haemagglutinin (HA) and neuraminidase (NA). There are 18 HA and 11 NA types (Tong, Zhu, Li, *et al.*, 2013; Mehle, 2014) discovered thus far and, theoretically, viruses with any combination of these two antigens can exist. Thus far though, only H1N1, H2N2, and H3N2 subtypes are known to circulate in the human population, with only occasional zoonotic infections and limited transmission of other subtypes recorded in humans (Taubenberger & Kash, 2010). IBVs are divided into two lineages, Victoria-like and Yamagata-like, which diverged some 30-40 years ago and have been cocirculating since (Biere, Bauer & Schweiger, 2010). Influenza viruses are named by their genus, host (if non-human), site of isolation, isolate number, the year of isolation and (if IAV) the HA and NA subtype, e.g. A/PR/8/34(H1N1) stands for isolate number 8 of H1N1 influenza A virus, isolated in Puerto Rico in 1934.

1.1.3 Influenza virus structure

In cell culture the virions of IAV and IBV are typically spherical particles, approximately 120 nm in length (Noda, 2012) (Figure 1.1). Primary isolates, however, are often found as filamentous structures, which can be as large as 1 μm in length, though it is not clear what advantage the filamentous morphology provides *in vivo* (Dadonaite, Vijayakrishnan, Fodor, *et al.*, 2016). Influenza viruses have host-cell derived envelopes in which three viral proteins reside - HA, NA and M2. HA and NA are the two major viral antigens, with HA required for cell receptor binding and NA for the release of budding viruses from the cell surface (Shaw & Palese, 2013). M2 is an ion channel required for the acidification of the viral particle upon entry and is also important for the generation of negative membrane curvature required for virus budding and scission (Pinto, Holsinger & Lamb, 1992; Schmidt, Mishra, Wang, *et al.*, 2013). Underneath the envelope a layer of the most abundant viral protein, M1, surrounds the viral genome. The genomes of IAVs and IBVs are divided into 8 viral RNA (vRNA) segments, each of which is contained in a ribonucleoprotein (RNP) complex (Figure 1.1 C). The RNPs consists of three components – a heterotrimeric viral polymerase (3P), an oligomer of the viral nucleoprotein (NP), and vRNA (Arranz, Coloma, Chichón, *et al.*, 2012; Coloma, Valpuesta, Arranz, *et al.*, 2009; Moeller, Kirchdoerfer, Potter, *et al.*, 2012). The structure of RNPs is discussed in a later section. Each of the 8 segments of IAV genome code for one or more different viral proteins. There are 10 major proteins encoded in the virus genome: 3 polymerase proteins, PB2, PB1 and PA, 2 glycoproteins, HA and NA, matrix protein M1, M2 ion channel, nucleoprotein (NP), non-structural protein 1 (NS1), and nuclear export protein (NEP) (Dawson, Lazniewski & Plewczynski, 2017) (Table 1.1). A number of additional proteins (Vasin, Temkina, Egorov, *et al.*, 2014), such as PA-X (Jagger, Wise, Kash, *et al.*, 2012), PB1-F2 (Chen, Calvo, Malide, *et al.*, 2001), M42 (Wise, Hutchinson, Jagger, *et al.*, 2012) and others, are also encoded by the viral genome, some of which are only present in a subset of influenza strains.

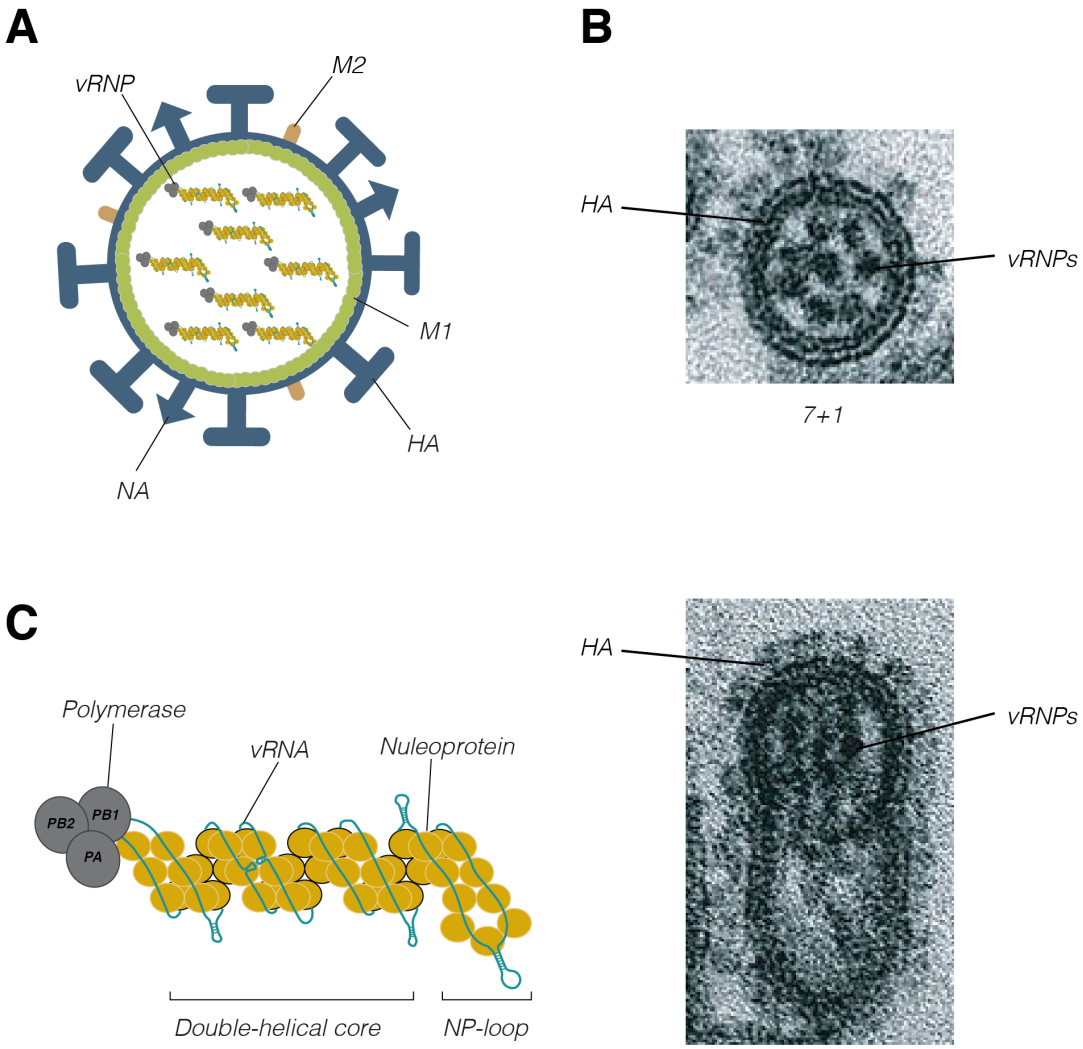


Figure 1.1 Structure of influenza virus

*(**A**) Structure of influenza virion. (**B**) Electron microscopy sections of budding IAV virions; adapted from Noda et al. (2006). 7+1 arrangement of vRNPs inside viral particles can be seen in the top cross-section. (**C**) Structure of an RNP.*

Table 1.1 Proteins encoded by the influenza A virus genome

Segment	*Protein*	*Primary function*
1	PB2	PB2 is a subunit of the viral polymerase, which binds the 5′-caps of host mRNAs
2	PB1	PB1 is the major catalytic subunit of the viral polymerase, which synthesises RNA
	PB1-N40	PB1-N40 regulates balanced expression between PB1 and PB1-F2
	PB1-F2	PB1-F2 is a pro-apoptotic factor
3	PA	PA is a subunit of the viral polymerase, which cleaves the 5′-caps from host mRNAs
	PA-X	PA-X is an endonuclease, which modulates the host response
4	HA	HA binds sialic acid receptors on target cells
5	NP	NP coats viral RNA and is a part of an RNP complex
6	NA	NA cleaves sialic acid from cell surface
7	M1	M1 is the most abundant protein in the virus, it lies underneath the viral envelope and influences virion morphology
	M2	M2 is an ion channel, required for virus release from endosomes and membrane scission
	M42	M42 is an ion channel, which can replace M2
8	NS1	NS1 is the major antagonist of the host innate immune responses
	NS2 (NEP)	NEP is responsible for vRNP export from the nucleus

1.1.4 Influenza virus life cycle

The life cycle of influenza virus is depicted in Figure 1.2. The infection begins with the virus HA glycoprotein binding to a receptor on a cell surface. Influenza viruses use sialic acids (SA), which are found on cell surface proteins, for cell entry - viruses infecting a human host preferentially bind to α2,6-linked SA, whereas avian-adapted viruses bind to α2,3-linked SA (Shelton, Ayora-Talavera, Ren, *et al.*, 2011). The binding to the cellular receptor triggers endocytosis, which brings the virus into the cell. Clathrin-mediated endocytosis is the major pathway described for influenza virus entry (Rust, Lakadamyali, Zhang, *et al.*, 2004), although, entry via caveolin-dependent endocytosis and micropinocytosis has also been observed (de Vries, Tscherne, Wienholts, *et al.*, 2011; de Vries, de Vries, Wienholts, *et al.*, 2012). Post cell entry, influenza viruses use the endosomal trafficking pathway to release the vRNPs into the cytoplasm. The low pH of the late endosomes triggers conformational changes in the HA, which causes the viral envelope to fuse with the endosomal membrane (Maeda & Ohnishi, 1980). In addition, the increase in protons inside the viral particle mediated by the M2 ion channel, further acidifies the endosome (Wang, Lamb & Pinto, 1995), which in turn causes the M1 protein to disassociate from the vRNPs (Ruigrok, Barge, Durrer, *et al.*, 2000), leading to the uncoating of the viral particle and

the release of the vRNPs into the cytoplasm. The vRNPs are too large for passive diffusion into the nucleus, therefore, karyopherin-dependent transport is used (Wang, Palese & O'Neill, 1997). Several nuclear localisation signals (NLSs) have been described in the NP (Bullido, Gómez-Puertas, Albo, *et al.*, 2000) as well as in all subunits of the 3P, however the NP-NLSs seem to be primarily used for the nuclear import of the vRNPs (O'Neill, Jaskunas, Blobel, *et al.*, 1995; Wu, Sun & Panté, 2007). The NLSs are recognised by the α and β karyopherins which, together with the Ran and P_{10} complex, dock to the nuclear pore complex and transfer the vRNPs into the nucleus in an ATP-dependent manner (O'Neill, Jaskunas, Blobel, *et al.*, 1995; Neumann, Castrucci & Kawaoka, 1997).

Once in the nucleus, transcription of the vRNA begins. The vRNPs interact with the C terminal domain (CTD) of the cellular DNA dependent RNA polymerase II (Pol II) via the 3P (Lukarska, Fournier, Pflug, *et al.*, 2017; Martínez-Alonso, Hengrung & Fodor, 2016). vRNP-Pol II association provides a source of 5'-caps for the viral messenger RNAs (vmRNAs). The 3P does not have the ability to synthesise the 5' cap structures, required for mRNA translation, therefore, through a process called cap-snatching, the 3P acquires the caps of cellular mRNAs (Herz, Stavnezer, Krug, *et al.*, 1981; Plotch, Bouloy, Ulmanen, *et al.*, 1981). The nascent mRNAs, produced by the Pol II, are captured by the cap-binding PB2 domain of the 3P (Guilligay, Tarendeau, Resa-Infante, *et al.*, 2008). The PA domain then cuts the mRNA some 8-14 nucleotides downstream of the cap and PB1 uses the cleaved caps as primers for producing vmRNAs (Dias, Bouvier, Crépin, *et al.*, 2009; Yuan, Bartlam, Lou, *et al.*, 2009). The produced vmRNAs are exported from the nucleus (Amorim, Read, Dalton, *et al.*, 2007; Read & Digard, 2010; Pereira, Read, Wise, *et al.*, 2017) and translated primarily in the same way as the cellular mRNAs are, although some studies have suggested that viral proteins, such as NS1 and 3P, may facilitate preferential translation of the mvRNAs over the mRNAs (Burgui, Yángüez, Sonenberg, *et al.*, 2007; la Luna, Fortes, Beloso, *et al.*, 1995; Park & Katze, 1995; Hooker, Sully, Handa, *et al.*, 2003). Translation of the vmRNAs leads to accumulation of viral proteins, which is thought to tilt the balance from transcription to replication. Unlike transcription, replication is a primer independent process. *De novo* initiation starts at the 3' terminus of the viral promoter; it is not clear yet what happens to the newly synthesised 5' terminus of the RNA as it exits the replicating polymerase but, potentially, it is captured by a new polymerase and is co-replicative coated with NP. In the 1st step of replication a cRNP is made, which acts as a replicative intermediate for the production of vRNPs. The production of vRNPs starts with internal initiation at the positions 4 and 5 of the cRNA promoter,

which produces a pppAG dinucleotide that then translocates to the very terminus of the cRNA and acts as a primer for the vRNA synthesis initiation (Velthuis & Fodor, 2016). The production of vRNA from cRNA requires the presence of a second 3P, although it is not clear if the second 3P acts as a *trans-activating* (York, Hengrung, Vreede, *et al.*, 2013) or a *trans-acting* (Jorba, Coloma & Ortin, 2009) polymerase. In the *trans-activating* polymerase model, the 2nd 3P is required for initiating conformational changes in the resident polymerase, which would favour the internal initiation from the 3' end of cRNA and the replication is performed by the resident polymerase itself. In the *trans-acting* polymerase model, the 2nd polymerase is responsible for the replication of the cRNA, by gaining access to the 3' end of the cRNA and starting internal initiation. The latter model also requires a 3rd polymerase, which would take the emerging 5'end of the vRNA from the *trans-acting* polymerase and form the vRNP.

After replication, the newly synthesised vRNPs are exported into the cytoplasm. The export of the vRNPs is thought to occur through a daisy-chain-like process, where the vRNP-resident NPs interact with the viral M1 protein that in turn interacts with the viral nuclear export protein (NEP), which contains nuclear export signals (NES) that are recognised by the Crm1 exportin (Paterson & Fodor, 2012). The latter step in the vRNP export pathway is still under debate because, while NEP-NES can cause nuclear export of non influenza virus-derived peptides (O'Neill, Talon & Palese, 1998), blocking the Crm1 pathway does not increase NEP accumulation in the nucleus, but does prevent vRNP export. Consequently, it has been suggested that other proteins, such as the NP, may be involved in the vRNP nuclear export (Elton, Simpson-Holley, Archer, *et al.*, 2001). In addition, interaction between NEP and the PB2 domain of the 3P has also been suggested to be important for the vRNP export (Brunotte, Flies, Bolte, *et al.*, 2014). Once in the cytoplasm, the vRNPs need to be transported to the plasma membrane, but the exact path of vRNP trafficking is not clear yet. After exiting the nucleus, the vRNPs have been observed to accumulate near the microtubule organising centre (MTOC) (Kawaguchi, Matsumoto & Nagata, 2012), though inhibition of microtubule polymerisation has little effect on the virus growth (Momose, Kikuchi, Komase, *et al.*, 2007; Amorim, Bruce, Read, *et al.*, 2011; Avilov, Moisy, Munier, *et al.*, 2011). At the MTOC, the vRNPs are thought to interact with Rab11 positive recycling endosomes (REs) (Momose, Sekimoto, Ohkura, *et al.*, 2011; Eisfeld, Kawakami, Watanabe, *et al.*, 2011). Rab11 pull-downs have shown that the association between vRNPs and Rab11 likely occurs via the PB2 (Amorim, Bruce, Read, *et al.*, 2011) and knocking down Rab11 in cells was shown to lead to

significant decrease in the virus titres (Eisfeld, Kawakami, Watanabe, *et al.*, 2011; Bruce, Digard & Stuart, 2010). As the Rab11 positive REs are involved in the trafficking of various cellular cargoes to the plasma membrane, it was thought that the vRNPs co-opt this pathway to be transported to the sites of virion assembly. However, a recent study suggested that perhaps the role of the endoplasmic reticulum (ER) in the vRNP trafficking process has been overlooked (Martini, Fournier, Sachse, *et al.*, 2017). Transmission electron microscopy and immunofluorescence experiments have shown that upon infection a dramatic rearrangement of ER around the MTOC takes place. Furthermore, Rab11-dependent budding of structures, called irregularly coated vesicles (ICVs), from the rearranged ER was observed. The ICVs were found to be coated with the vRNPs and moved towards the plasma membrane, where, through an unknown mechanism, the vRNPs were deposited on the plasma membrane.

Budding of the viral particles is mediated by the viral proteins that accumulate at the plasma membrane during the course of infection (Schmitt & Lamb, 2005). Overexpression of HA, M1, M2 and NA proteins can all lead to virus-like particle formation, suggesting their involvement in the budding process (Chen, Leser, Morita, *et al.*, 2007; Kerviel, Dash, Moncorgé, *et al.*, 2016; Wu, Yeh, Yang, *et al.*, 2010). Lipid rafts play an important role in virus budding by concentrating the HA and NA proteins in the budding sites and the M2 accumulates in their periphery (Simpson-Holley, Ellis, Fisher, *et al.*, 2002; Ohkura, Momose, Ichikawa, *et al.*, 2014). The M1 interacts with the cytoplasmic tails of the HA, NA and M2 proteins and polymerises at the plasma membrane (Chen, Leser, Jackson, *et al.*, 2008; McCown & Pekosz, 2006; Mitnaul, Castrucci, Murti, *et al.*, 1996). The polymerisation of M1 causes the elongation of the budding particle from the plasma membrane which, when the virus genome is incorporated, gets released by M2-mediated membrane scission (Martyna, Bahsoun, Badham, *et al.*, 2017). How the complex of 8 vRNPs gets incorporated into the budding virion is not clear, although vRNP association with M1 is likely involved (Gómez-Puertas, Albo, Pérez-Pastrana, *et al.*, 2000; Schmitt & Lamb, 2005). Electron tomography studies have shown that during budding the 8 vRNPs align at the distal tip of the virion forming a 7+1 complex, with one central RNP surrounded by seven others (Figure 1.1 **B**). It is not clear which vRNP takes the central position in the 7+1 complex, although, based on the vRNP size measurements, it seems that the central position may not be limited to an RNP of a single specific segment (Fournier, Moules, Essere, *et al.*, 2012b; Noda, Murakami, Nakatsu, *et al.*, 2018; Noda, Sagara, Yen, *et al.*, 2006; Noda, Sugita, Aoyama, *et al.*, 2012).

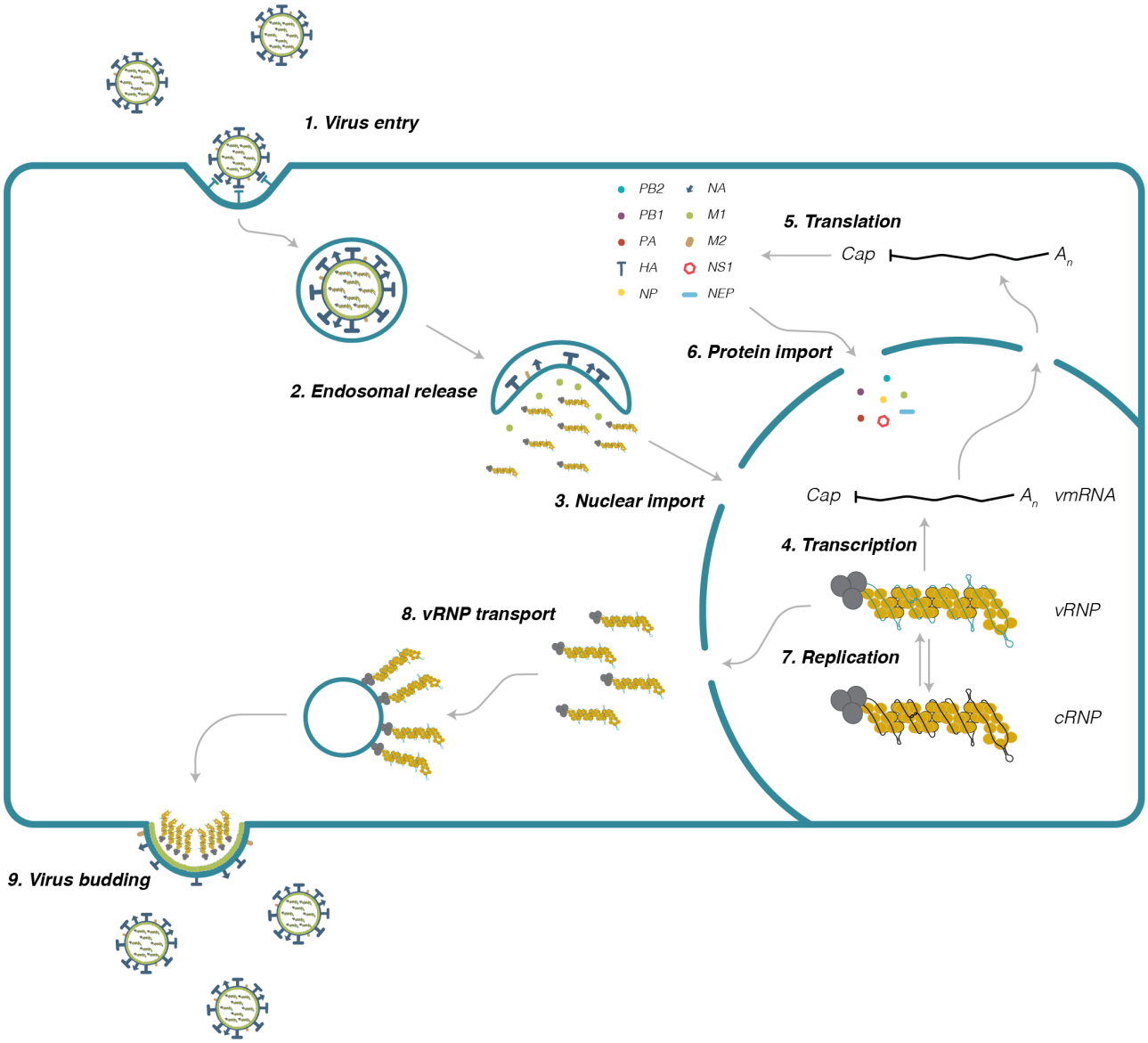


Figure 1.2 Influenza virus life cycle

*(**1**) The virus enters a cell via binding to sialic acid-containing receptors. (**2**) The vRNPs are released from the endosome after endosomal acidification. (**3**) vRNPs are imported into the nucleus where they are (**4**) transcribed to produce viral mRNAs (vmRNAs), which are exported to the cytoplasm and (**5**) translated. Some of the newly made viral proteins are (**6**) imported into the nucleus and a switch from transcription to (**7**) replication happens. cRNPs act as replicative intermediates during the replication step. Newly made vRNPs get exported from the nucleus and are (**8**) trafficked through the cytoplasm via a pathway involving intracellular vesicles. (**9**) Accumulation of viral proteins and vRNPs leads to budding of newly produced viruses from the infected cell.*

1.2 Ribonucleoprotein structure

1.2.1 RNA-dependent RNA polymerase structure

The polymerase of influenza virus consists of 3 subunits: polymerase basic protein 2 (PB2), polymerase basic protein 1 (PB1), and polymerase acidic protein (PA) (or P3 for influenza C viruses). Recently, high-resolution structures of ICV, IBV and IAV 3Ps have become available (Pflug, Guilligay, Reich, *et al.*, 2014; Chang, Sun, Liang, *et al.*, 2015; Hengrung, Omari, Martin, *et al.*, 2015). The PB1 subunit of the 3P forms the core, which contains the polymerase active site and has a structure similar to that of other RNA-dependent RNA polymerases, with the characteristic finger, palm and thumb subdomains. PB1 interacts with both the 3′ and 5′ promoter termini of the RNA and contains the priming loop in its thumb domain, that is important for the terminal initiation on the 3′ terminus of the promoter, during the synthesis of cRNA.

The PB2 is a flexible subunit, the arrangement of which differs substantially between the available IAV and IBV versus the ICV 3P structures (Hengrung, Omari, Martin, *et al.*, 2015; Pflug, Guilligay, Reich, *et al.*, 2014). PB2 binds the cap structures of cellular mRNAs, a step required for the cap-snatching process (Reich, Guilligay, Pflug, *et al.*, 2014). The structure of the ICV 3P suggests that, in the absence of the vRNA promoter, the PB2 cap-binding domain faces the PA endonuclease domain in a conformation that does not allow cap-snatching to take place, however, upon the promoter binding, as suggested by the influenza A and B polymerase structures, PB2 undergoes conformational rearrangement which brings the PB2 cap-binding domain closer to the PA-endonuclease making the polymerase transcriptionally active. The PB2-627 domain is named after its 627 residue, which has been shown to be an important host-range determinant. The polymerases of influenza viruses adapted to grow in avian hosts usually have a glutamate residue in the 627 position and viruses with 627E grow poorly in mammalian hosts. However, when the 627E residue is replaced with a lysine, the 627K containing polymerases can replicate well in mammalian-hosts. It is not clear what function these changes in the 627 residue perform, in fact, *in vitro* a 3P lacking the 627-domain can successfully transcribe and replicate. To explain the molecular basis of PB2 residue 627 mediated host restriction, factors such as host-temperature adaptation (Massin, Van der Werf & Naffakh, 2001; Aggarwal, Dewhurst, Takimoto, *et al.*, 2011) and host-factors have been suggested to be involved (Long, Giotis, Moncorgé, *et al.*, 2016; Weber, Sediri, Felgenhauer, *et al.*, 2015; Cauldwell, Long, Moncorge, *et al.*, 2014).

PA has 2 subdomains: an N-terminal endonuclease domain and a C-terminal domain, which are connected by a flexible linker. The endonuclease domain plays a key role in the cap-

snatching of cellular mRNAs by performing the mRNA cleavage step of the process (Dias, Bouvier, Crépin, *et al.*, 2009; Yuan, Bartlam, Lou, *et al.*, 2009). The C-terminus of PA interacts extensively with the viral promoter and is also important for the association between the viral polymerase and the CTD of cellular Pol II during transcription (Lukarska, Fournier, Pflug, *et al.*, 2017; Serna Martin, Hengrung, Renner, *et al.*, 2018).

1.2.2 Nucleoprotein structure

The genomes of all the segmented negative strand RNA viruses (sNSVs) are coated with viral nucleoproteins (NPs) (Reguera, Cusack & Kolakofsky, 2014). Crystal structures are now available for many sNSVs, including IAV (Ye, Krug & Tao, 2006; Ng, Zhang, Tan, *et al.*, 2008) and IBV (Ng, Lam, Zhang, *et al.*, 2012), which all have a characteristic globular core domain containing a positively charged surface that is involved in the binding of RNA. The structure of the influenza NP can by further subdivided into the head and body domains. The head domain contains a flexible tail-loop structure that is important for NP oligomerisation during RNP assembly (Arranz, Coloma, Chichón, *et al.*, 2012). During the oligomerisation a tail-loop of one NP molecule inserts itself into an insertion groove of another NP molecule, which is positioned in the body domain. The flexibility of the tail-loop is thought to explain the structural flexibility observed in the RNPs, which in themselves have both an NP-loop and double-helical central region and are also often seen variably bent and twisted in negative stain electron microscopy (Reguera, Cusack & Kolakofsky, 2014). Furthermore, the NP molecules of the non-segmented negative strand RNA viruses lack the flexible tail-loops, and consequently, form RNPs with quite rigid helical rod conformations (Ruigrok, Crépin & Kolakofsky, 2011). Purified IAV or IBV NP molecules form trimers or tetramers, respectively, though only the monomeric form of NP is thought to be loaded onto the vRNA during RNP assembly. The switch between monomeric and oligomeric NP in cells is thought to be controlled by factors such as NP phosphorylation (Chenavas, Estrozi, Slama-Schwok, *et al.*, 2013) and association with cellular proteins (Hu, Gor, Morikawa, *et al.*, 2017; Momose, Naito, Yano, *et al.*, 2002; Kawaguchi, Momose & Nagata, 2011).

1.2.3 vRNA structure

The structure of the vRNA is not visible in the central helical or the terminal NP-loop regions of the available RNP structures (Moeller, Kirchdoerfer, Potter, *et al.*, 2012; Coloma, Valpuesta, Arranz, *et al.*, 2009; Arranz, Coloma, Chichón, *et al.*, 2012) and the only protein-associated vRNA structure that is available to this date is that of the promoter region, which is visible in the polymerase crystal structures (Pflug, Guilligay, Reich, *et al.*, 2014). Over the years multiple studies have proposed several conformations for the promoter RNA including panhandle, fork, corkscrew and hook-like structures (Fodor, Pritlove & Brownlee, 1995; Flick, Neumann, Hoffmann, *et al.*, 1996; Fodor, Pritlove & Brownlee, 1994). The promoter visible in the polymerase structure forms a hook structure with the 1st 10 nucleotides of the 5′ vRNA followed by a duplex region formed by complementary base-pairing between nucleotides 11-15 and 10-15 of the 5′ and 3′ vRNA termini, respectively.

It is unclear what is the conformation in the vRNA in the context of RNPs. Previous studies have suggested that the NP binds the phosphate backbone of the vRNA with ~24 nucleotide periodicity (Ortega, Martín-Benito, Zürcher, *et al.*, 2000). The binding of NP does not show any sequence specificity, but it shows preferential binding to single stranded RNA (Yamanaka, Ishihama & Nagata, 1990). More recent studies have suggested that the binding of NP to vRNA in the context of RNPs is not regular, with some regions being more likely to have NP present than others and the regions with less NP binding perhaps being able to form secondary RNA structures (Williams, Townsend, Wylie, *et al.*, 2018; Lee, Le Sage, Nanni, *et al.*, 2017). The binding of NP does not protect the vRNA from nuclease digestion and the RNPs are sensitive to RNase V1 treatment (Yamanaka, Ishihama & Nagata, 1990), which suggests that both single and double stranded regions of vRNA exist. While *in vitro* transcribed vRNA, lacking any added proteins, forms multiple secondary RNA structures (Lenartowicz, Kesy, Ruszkowska, *et al.*, 2016) it is though that most of these secondary structures are absent in the presence of NP. Clear evidence exists showing that at least some secondary RNA structures can be formed in vRNPs. Mutagenesis studies have suggested that structures as complex as pseudoknots can be formed in RNPs (Gultyaev, Tsyganov-Bodounov, Spronken, *et al.*, 2014) in addition to a number of stem-loop structures (Spronken, van de Sandt, de Jongh, *et al.*, 2017; Kobayashi, Dadonaite, van Doremalen, *et al.*, 2016). Furthermore, non-native secondary RNA structures can be introduced into RNPs, where they are maintained,

exposed and can be recognised by RNA structure-specific proteins (York, Hengrung, Vreede, *et al.*, 2013).

1.2.4 RNP assembly

The structures of both full-length and a mini-RNP are available (Moeller, Kirchdoerfer, Potter, *et al.*, 2012; Coloma, Valpuesta, Arranz, *et al.*, 2009; Arranz, Coloma, Chichón, *et al.*, 2012) (Figure 1.3). The mini-RNP, which only has a 254 nucleotide vRNA in it, forms a nine NP unit oligomer, which is circularised by the polymerase that interacts with two NP monomers (Coloma, Valpuesta, Arranz, *et al.*, 2009). In contrast to the circular mini-RNP, the full length RNPs have polymerase at one end, a double-helix forming NP core and an NP-loop at the opposite end (Arranz, Coloma, Chichón, *et al.*, 2012; Moeller, Kirchdoerfer, Potter, *et al.*, 2012) (Figure 1.3 B). The length of an RNP seems to depend on the length of the RNA it is associated with, but even RNPs containing the same type of vRNA can have an NP-loop region formed by a variable number of NP units (York, Hengrung, Vreede, *et al.*, 2013). The contacts between the 3P and NP in the RNP structures are not fully resolved, but mutational studies have shown all 3P subunits (Biswas, Boutz & Nayak, 1998; Poole, Elton, Medcalf, *et al.*, 2004; Vidic, Noiray, Bagchi, *et al.*, 2016), could potentially interact with NP, with structural modelling favouring the PA-NP interaction (Reich, Guilligay, Pflug, *et al.*, 2014). Mixing of purified NP, vRNA and 3P *in vitro* does not seem to be sufficient to produce vRNPs (Chenavas, Estrozi, Slama-Schwok, *et al.*, 2013; Tarus, Bakowiez, Chenavas, *et al.*, 2012). By contrast, co-expression of these components in cells leads to production of transcription and replication-competent vRNPs. The latter observation suggests that a co-replicational loading of NP may take place. When the newly synthesised 5′ end of the viral RNA is made it is bound by a new 3P, which likely recruits the 1st NP monomer to the replication site. How the 1st NP is recruited is not clear, but it likely occurs through weak 3P-NP surface interactions. As the synthesised vRNA grows, the subsequent NPs oligomerise via the NP tail-loop and insertion groove interactions (Turrell, Lyall, Tiley, *et al.*, 2013). It has been observed that the NP oligomerisation speed increases with and increasing length of the associated RNA (Tarus, Bakowiez, Chenavas, *et al.*, 2012). It is likely the that the fixing of the 5′ and 3′ ends of the vRNA by the 3Ps, while the NP-loaded vRNA continues to grow, causes the supercoiling of RNP and formation of the double-helical RNP core. The binding of the 3P to the vRNA ends maintains the coiled RNP structure, and removal of the 3P from an RNP leads to its unwinding (Klumpp, Ruigrok

& Baudin, 1997). Interestingly, NP is only required to replicate vRNA, which is longer than 101 nucleotides (Turrell, Lyall, Tiley, *et al.*, 2013), suggesting that it acts as an essential elongation factor during the replication of RNPs.

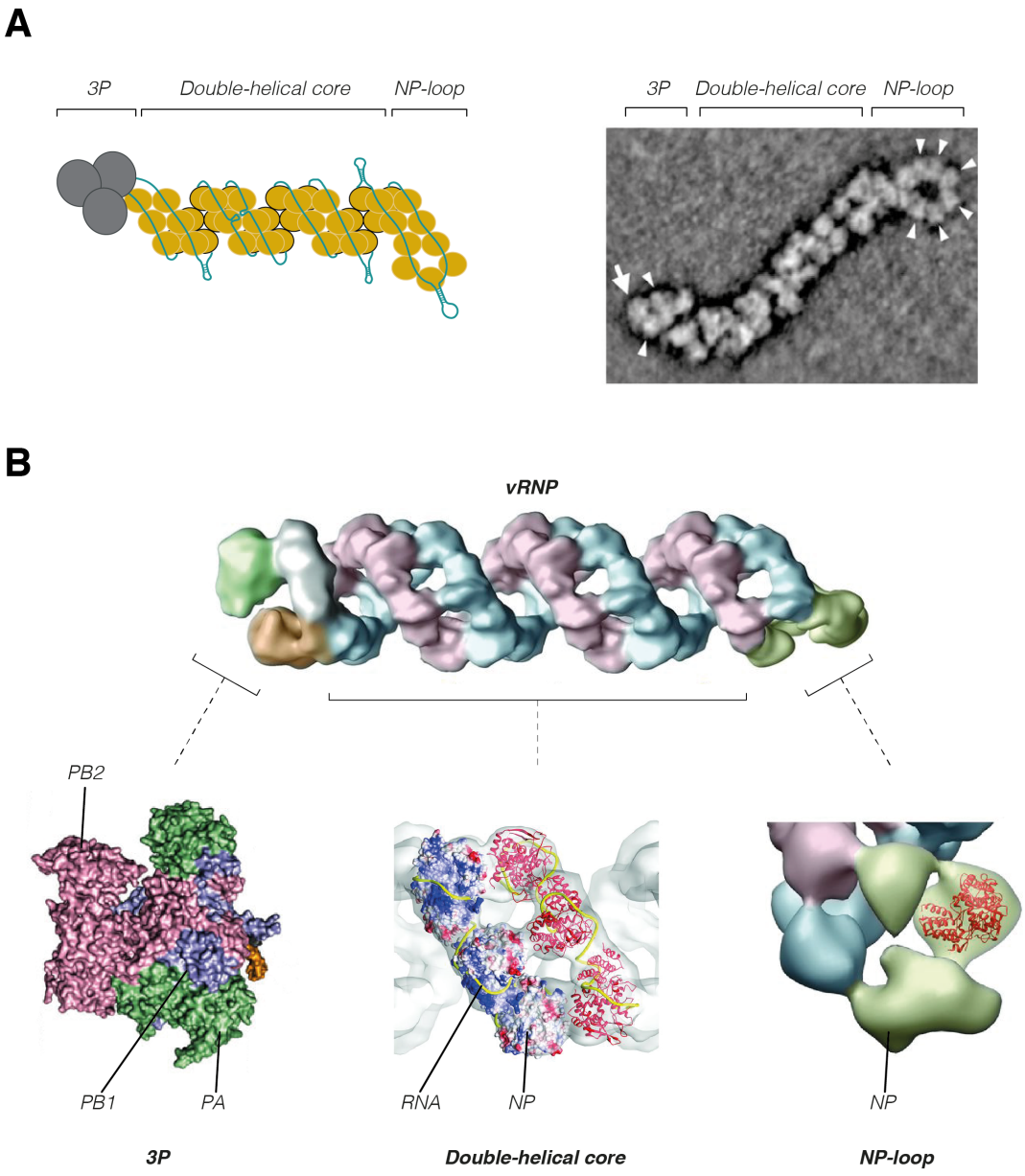


Figure 1.3 Structure of influenza virus RNPs

*(**A**) Cartoon illustration of an RNP and the structure of an RNP as seen using negative staining microscopy; adapted from Sugita et al. (2013). (**B**) Cryo-EM structure of an RNP with zoomed-in 3P, helical-core (the potential path of RNA is modelled in yellow) and NP-loop sections; adapted from Arranz et al. (2012) and Velthuis & Fodor (2016).*

1.3 Influenza virus genome packaging

Each segment of the influenza virus genome codes for one or more viral proteins that are required for successful virus replication. While genome segmentation provides a way for increasing the virus genetic diversity, through the process of genetic reassortment, it also poses a challenge in the need to package all 8 segments in each virion. The mechanism of the packaging of the vRNPs into the budding viruses has not yet been fully understood but the importance of packaging signals in the viral RNA has been widely accepted (Giese, Bolte & Schwemmle, 2016; Gerber, Isel, Moules, *et al.*, 2014).

1.3.1 Selective genome packaging

There are two possibilities of packaging a segmented genome - doing so randomly or selectively. Random packaging would assemble any combination of the viral segments into the viral progeny, whereas selective packaging would ensure that only a single unit of each vRNA segment is selected. Examples of both kinds of genome packaging modes exist among segmented RNA viruses (McDonald & Patton, 2011; Luque, Rivas, Alfonso, *et al.*, 2009), but for influenza viruses a mechanism of selective genome packaging has gained the most support (McGeoch, Fellner & Newton, 1976; Chou, Vafabakhsh, Doğanay, *et al.*, 2012). A random genome packaging would predict an approximate 1 in 400 infectious versus non-infectious viral particles, but the observed 1 in 10 to 1 in 100 ratio during the influenzas infections contradicts this prediction (Donald & Isaacs, 1954). Experiments using specific vRNA FISH probes also have suggested that individual virions tend to contain only unique vRNAs (Chou, Vafabakhsh, Doğanay, *et al.*, 2012) and it has been noticed that defective interfering RNAs (DIs), which are truncated versions of different vRNA segments, can only outcompete the packaging of the vRNA segments from which they are derived (Duhaut & MCCAULEY, 1996; Odagiri & Tashiro, 1997).

1.3.2 Packaging signals in the vRNA

As the only known difference between the different vRNPs is the vRNA they are associated with, the major driver of the selective packaging has been proposed to be the packing signals

present in the vRNA. Analysis of DIs has been a valuable tool in understanding the involvement of the vRNA in the packaging of the influenza virus genome. DIs retain the terminal regions (consisting of the conserved promoter sequence and segment-unique untranslated regions (UTRs)) of the vRNA but have variable-size truncations in the internal parts of a segment (Dimmock & Easton, 2014). Early studies using viruses which had their NA segment replaced with NA-derived DIs have shown that when such virus constructs are used to infect NA-expressing cells, the progeny viruses maintain the packaging of the 7 viral segments plus the NA-DI, suggesting that selection for packaging the terminal regions of vRNA of each segment exists (Liu & Air, 1993; Yang, Bansal, Liu, *et al.*, 1997). Similar experiments have been conducted for other viral segments and led to the same results, although it has been noted that in addition to the promoter and the UTR sequences the efficiency of segment packaging increases substantially when the DIs contain parts of protein-coding regions as well, the length of which is segment-specific (Fujii, Goto, Watanabe, *et al.*, 2003; Santos Afonso, Escriou, Leclercq, *et al.*, 2005; Fujii, Fujii, Noda, *et al.*, 2005; Liang, Hong & Parslow, 2005; Marsh, Rabadán, Levine, *et al.*, 2008; Muramoto, Takada, Fujii, *et al.*, 2006; Watanabe, Watanabe, Noda, *et al.*, 2003). In the literature, the packaging signals present in the non-coding regions of the vRNA are often referred to as the vRNA selection signals and the regions important for packaging present in the protein-coding parts of the segments are referred to as the bundling signals (Goto, Muramoto, Noda, *et al.*, 2013).

The packaging signals for the different segments have been widely-studied through the use of vRNA reporter constructs, which contain reporter-coding sequences flanked by variable length native vRNA termini. Through the use of such reporters up to 300 nucleotides from the ends of vRNAs have been shown to be important for the packaging of different segments, although, depending on the segment, virus strain and the specific reporter used in a study, these packaging signals could differ substantially between studies (Fujii, Goto, Watanabe, *et al.*, 2003; Duhaut & Dimmock, 1998; Nayak, Hui & Barman, 2004; Eisfeld, Neumann & Kawaoka, 2014). The importance of the terminal regions in the segment packaging has also been highlighted by a sequence conservation analysis, which showed that the terminal regions accumulate fewer than expected synonymous mutations in them (Gog, Afonso, Dalton, *et al.*, 2007). Further mutagenesis analysis has shown that mutations in some, though not all, sites where the synonymous mutations seem to be supressed can lead to observable packaging defects (Hutchinson, Wise, Kudryavtseva, *et al.*, 2009; Hutchinson, Curran, Read, *et al.*, 2008; Fujii, Fujii, Noda, *et al.*, 2005; Ozawa, Maeda,

Iwatsuki-Horimoto, *et al.*, 2009). Multiple mutagenesis studies have now shown that the nucleotide sequence is an important determinant of the vRNA packaging; just a single point mutation has been observed to affect the packaging of vRNA. However, no clear pattern in position or sequence of packaging defects-inducing mutations has emerged so far, leading to suggestion that the packaging signals may not be continuous and may comprise multiple sites within a given segment.

1.3.3 Inter-segment RNA-RNA interactions

While no study has yet definitely shown a physical interaction between the vRNAs in the context of RNPs, the specific RNA-RNA interactions formed by the vRNA packaging signals in different segments have been proposed to drive the selection of the different vRNPs during genome packaging. An elegant study by Gavazzi et *al.* (2013) has shown that mutagenesis of regions in segments PB1 and NS, which can form complementary base-pairs, can lead to packaging defects. Importantly, such defects were restored upon the introduction of base-pairing restoring mutations. Individually, the PB1 and NS regions involved in the interactions could be predicted to form small hairpin structures, suggesting that perhaps secondary RNA structures extending from the RNPs could mediate their bundling during genome assembly. Interestingly, undefined string-like structures (Noda, Sugita, Aoyama, *et al.*, 2012) and high electron density regions (Fournier, Moules, Essere, *et al.*, 2012a) formed between RNPs in virion electron tomography studies have also been suggested to be the sites of inter-segment RNA-RNA interaction formation.

1.3.4 vRNP incorporation during virus budding

The actual incorporation of the vRNA versus cellular RNA into the virus particles requires the recognition of both the packaging signals in the vRNA and the associated viral proteins. Without a promoter-bound viral polymerase the vRNA will not be selectively incorporated into the virions, regardless of the presence of UTRs and/or the bundling signals in it (Luytjes, Krystal, Enami, *et al.*, 1989; Neumann & Hobom, 1995). In addition, a recent study has suggested that conserved residues in the NP are also important for the selection of the 8 different RNPs into a budding virion (Moreira, Weber, Bolte, *et al.*, 2016). Electron tomography images of incorporated vRNP inside budding virions have shown that the RNPs are arranged in a distinct 7+1 arrangement

(Figure 1.1 B), with a single central RNP surrounded by seven others (Noda, Sagara, Yen, *et al.*, 2006; Noda, Sugita, Aoyama, *et al.*, 2012). The process behind the formation of such arrangement is not clear but two possible paths of assembling RNPs together have been suggested - the 'daisy chain' and the 'master segment' assembly. In the daisy chain process, the 1st RNP interacts with a 2nd which in turn interacts with a 3rd and so on, whereas in the master segment process, a single, central (the +1), RNP contacts all other RNPs and mediates the selection of a full set of 8 RNPs. The question still remains as to where, during the process of the vRNP trafficking in a cell, the interactions between the RNPs take place. Two studies have observed that vRNPs already exit the nucleus in bundles of several vRNPs (Lakdawala, Wu, Wawrzusin, *et al.*, 2014; Martini, Fournier, Sachse, *et al.*, 2017), although they may separate and interact with alternative RNPs while in the cytoplasm. A recent observation of high densities of vRNPs being transported to ICVs (Martini, Fournier, Sachse, *et al.*, 2017), suggests that perhaps the formation of the 7+1 complex happens on the surface of the ICVs, although RNP bundling on the plasma membrane just before the initiation of budding, cannot be excluded.

1.4 Influenza virus reassortment

The changes in influenza virus genome occur through two major processes – genetic drift and shift. Genetic drift refers to the changes in the virus genome, which occur due to the low fidelity of the viral polymerase, which introduces around 2 to 3 point mutations with every round of genome replication (Pauly, Procario & Lauring, 2017). This rapid accumulation of mutations, while potentially damaging, is also a key to escaping the host's immune responses. Accumulation of mutations in seasonal influenza viruses allows them to produce new versions of HA and NA antigens (known as antigenic drift), which are not recognized by the host's immune system, allowing the infection to proceed. The constant antigenic drift in influenza viruses is the reason for the need for annual vaccination against seasonal influenza viruses and is also the reason why, in some years, vaccines are poorly matched to the circulating influenza strains (Paules, Sullivan, Subbarao, *et al.*, 2018).

While genetic drift accounts for gradual changes in the virus genome, genetic shift leads an introduction of completely new versions of one or more vRNA segments and the emergence

of new influenza strains. Genetic shift happens through a process called reassortment, which is a consequence of the segmented virus genome. Reassortment happens when viruses of different strains infect the same cell, replicate and produce progeny viruses which package a mixed set of vRNPs, originating from both parental strains. So far, reassortment is known to only happen between the strains within a genus and has been described for IAV (Taubenberger & Kash, 2010; Lowen, 2017), IBV (Dudas, Bedford, Lycett, *et al.*, 2015) and ICV (Matsuzaki, Sugawara, Furuse, *et al.*, 2016). No reassortment between the genera has been observed yet. Understanding of the processes governing the reassortment between different strains is of particular importance for IAVs, as reassortment between the viruses circulating in the human and non-human hosts underlies the emergence of new pandemic strains.

The major host of IAVs are avian species, in which H1-H16 and N1-N9 antigens have been detected. At least 105 different species of birds have been found to carry IAVs, although the different HA/NA subtypes are not evenly distributed across the different species (Munster & Fouchier, 2009). In birds, IAVs seem to predominantly infect the intestinal epithelial cells and the fecal-oral route is the major path for the virus transmission. Most infections in avian species are asymptomatic, caused by low pathogenic avian influenza viruses (LPAIV). However, occasionally, the LPAI strains of H5 and H7 viruses acquire a polybasic cleavage site, which allows the HA to be cleaved by abundant furin proteases (Gröber, Vey, Angliker, *et al.*, 1992) and the virus to spread systemically in many organs (Schrauwen, Herfst, Leijten, *et al.*, 2012). The polybasic site-containing viruses are called highly pathogenic avian influenza viruses (HPAIVs), and can have up to 100% mortality rates in poultry (Alexander, 2000). HPAIVs are known to also infect humans, though no sustained human to human transmission of HPAIVs has yet been observed (Belser, Bridges, Katz, *et al.*, 2009). Reassortment among avian IAVs seems to be a common and abundant occurrence with seemingly little preference for specific segment combinations to be maintained, suggesting that most of the segments are functionally equivalent and there is little pressure for the segments to co-segregate (Wang, Soll, Dugan, *et al.*, 2008; Dugan, Chen, Spiro, *et al.*, 2008). Interestingly, when avian species jump to mammalian hosts they often acquire mutations in their genomes, which subsequently restrict them from replicating in the avian hosts again (Taubenberger, Reid, Lourens, *et al.*, 2005; Subbarao, London& Murphy, 1993; Sorrell & Perez, 2007). It is thought that all strains of IAV found in mammals have been derived from an avian parental strain at some point (Shaw & Palese, 2013) and therefore, an understanding of when, where and how avian IAV strains

jump into new hosts is important in order to be able to predict and monitor the emergence of novel influenza strains. Often described as 'mixing vessels', pigs are a known site where reassortment can readily occur (Ma, Kahn & Richt, 2008) - the 2009 Swine flu pandemic being the prime example (Neumann, Noda & Kawaoka, 2009) (see Figure 1.4). Initially, it was thought that the presence of both avian virus-preferred α2,3-galactose-linked receptors as well as human-adapted virus-preferred α2,6-galactose-linked receptors in the pigs may explain the common mixing of viruses in them (Ito, Couceiro, Kelm, *et al.*, 1998); however, later studies have shown that the pig respiratory tract has human-like receptor distribution (Nelli, Kuchipudi, White, *et al.*, 2010; Van Poucke, Nicholls, Nauwynck, *et al.*, 2010), leaving the question open. Transmission of viruses from wild birds to other domesticated animals such as turkeys, chickens (Swayne, Pantin-Jackwood, Kapczynski, *et al.*, 2009), horses (Morens & Taubenberger, 2010) and dogs (Maas, Tacken, Ruuls, *et al.*, 2007) has been reported, however, so far, only zoonotic transmission from domestic poultry and swine sources to humans has been well characterised.

1.4.1 Influenza virus pandemics of the past century

While influenza-like illnesses have been observed already in the ancient Greece, the influenza pandemic in Europe in 1510 marks the first clearly recognisable record of an IAV outbreak (Taubenberger & Morens, 2009). The first pandemic from which patient samples were available originated in Russia in 1889 and was likely caused by a virus with either H3 or H2 antigens (MASUREL & MARINE, 1973). The subsequent century has seen 4 major pandemics, (Figure 1.4) and while the exact origin and timing of all the pandemics is not certain, developments in sequencing and molecular dating methods have shed some light on the past events. The first, the largest and the most lethal pandemic recorded in the past century was the 1918 *Spanish flu* pandemic. Killing an estimated 50 million people worldwide, the *Spanish flu* is still often cited as a reminder of the potential devastation an influenza pandemic could cause (Oxford & Gill, 2018). As no pre-1918 sequences of influenza viruses are available the exact origin of the 1918 virus is still unclear. However, in 1997 Taubenberger et *al.* (Taubenberger, Reid, Krafft, *et al.*, 1997) succeeded in cloning some of the 1918 virus genes from paraffin-imbedded lung samples of the 1918 victims and showed that it was an H1N1 virus, which contained an avian-like HA sequence suggesting its avian origin. Subsequent analysis of the 1918 virus polymerase genes also showed a high sequence similarity to avian strains, indicating that the 1918 virus may not have been a

reassortant but had entered the human population directly from an avian host (Taubenberger, Reid, Lourens, *et al.*, 2005). However, due to the lack of viral sequence availability, the latter view is still under debate and the reassortment of avian-strains circulating in mammalian population pre-1918 still cannot be excluded (Smith, Bahl, Vijaykrishna, *et al.*, 2009).

After the 1918 pandemic the H1N1 virus circulated in the human population for the next 39 years. During this time the virus genome slowly drifted and the virus showed similar epidemiology as seasonal influenza strains. In 1957, the circulating H1N1 strains had been completely replaced by a new H2N2 pandemic virus. First reports of the new pandemic came from China in the early spring in 1957 and by the summer the virus had spread to 20 other countries leading to what has been called the *Asian flu* pandemic (Dunn, 1958). The pandemic H2N2 virus had emerged from a reassortment event between the seasonal 1918-like virus and an avian H2N2 strain, which produced a virus containing avian-origin PB1, HA and NA genes (Kawaoka, Krauss & Webster, 1989; Schäfer, Kawaoka, Bean, *et al.*, 1993). Pigs have been implicated as the potential site of reassortment between the H1N1 and H2N2 strains, as the H2 viruses are known to circulate in the swine population (Ma, Kahn & Richt, 2008).

Just 11 years after the Asian flu outbreak another virus, an H3N2 strain, caused the 3rd pandemic of the 20th century. The H2N2 strain reassorted with an avian virus and acquired its PB1 and HA genes. Given that the NA antigen remained the same and provided some cross-protection against H3N2 virus, this pandemic was not as severe as some of the previous ones (Morens, Taubenberger & Fauci, 2009). H3N2 viruses are still present today and circulate as seasonal influenza strains (Russell, Jones, Barr, *et al.*, 2008).

After the H3N2 outbreak no major pandemic was seen until 2009, although small H1N1 virus-caused outbreaks, were recorded in 1976 and 1977. In 1976 230 people died from an H1N1 virus infection in a Fort Dix military base in the United States and in 1977 an H1N1 *Russian flu* outbreak was detected in Siberia, China and Hong Kong. While the H1N1 viruses had disappeared from circulation in the human population after the Asian flu pandemic, the re-emergence of H1N1 viruses suggested that they were likely circulating in a non-human host, such as pigs (Zimmer & Burke, 2009). Interestingly, the Russian flu stain had a striking resemblance to the H1N1 strains

isolated just before the 1957 pandemic, leading to suggestions that the Russian flu could have been a consequence of a virus that escaped from a laboratory source (Scholtissek, Hoyningen & Rott, 1978).

The last recorded major pandemic happened in 2009 and was caused by an H1N1 virus, which originated from multiple reassortment events in a swine host, consequently called - the *Swine flu* pandemic. The 2009 outbreak started in Mexico (Fraser, Donnelly, Cauchemez, *et al.*, 2009) when a Eurasian avian-like swine H1N1 virus reassorted with a *triple reassortant* swine strain. The triple reassortant swine IAV strain has been first reported in pigs in 1998 (Brown, Harris, MCCAULEY, *et al.*, 1998), and is a virus which contains HA, NP, M and NS segments from the classical swine H1N1 virus (which itself is a descendant of the 1918 pandemic virus), PB2 and PA segments from a North American avian H1N1 strain, and PB1 and NA segments from the seasonal human H3N2 strain. The 2009 pandemic H1N1 strain (pH1N1) emerged when a Eurasian avian-like virus, which has not been previously seen in North American pigs, has donated its NA and M segments to the triple reassortant background. It is unclear when the reassortment took place or for how long has the pH1N1 virus circulated in pigs before the jump to the human population occurred (Smith, Vijaykrishna, Bahl, *et al.*, 2009).

It is difficult to predict which IAV strain will lead to the next pandemic. Since the first zoonotic outbreak of H5N1 virus in Hong Kong in 1997 (Subbarao, Klimov, Katz, *et al.*, 1998), the virus has spread from Asia to Europe and is now strictly monitored for its potential pandemic spread. The H5N1 viruses have particularly high mortality rates not only in humans but in domestic poultry and waterfowl as well, however no sustained human-to-human transmission has been detected thus far. As of July 20th, 2018, 860 human H5N1 cases have been reported of which 454 were lethal (Goubau, Schlee, Deddouche, *et al.*, 2014b). Since 2013, another avian virus, H7N9 has also been heavily monitored for the its pandemic potential - out of 1,625 confirmed H7N9 cases 623 people have died and the virus has been detected in China, Malaysia and Canada (FAO, 2018).

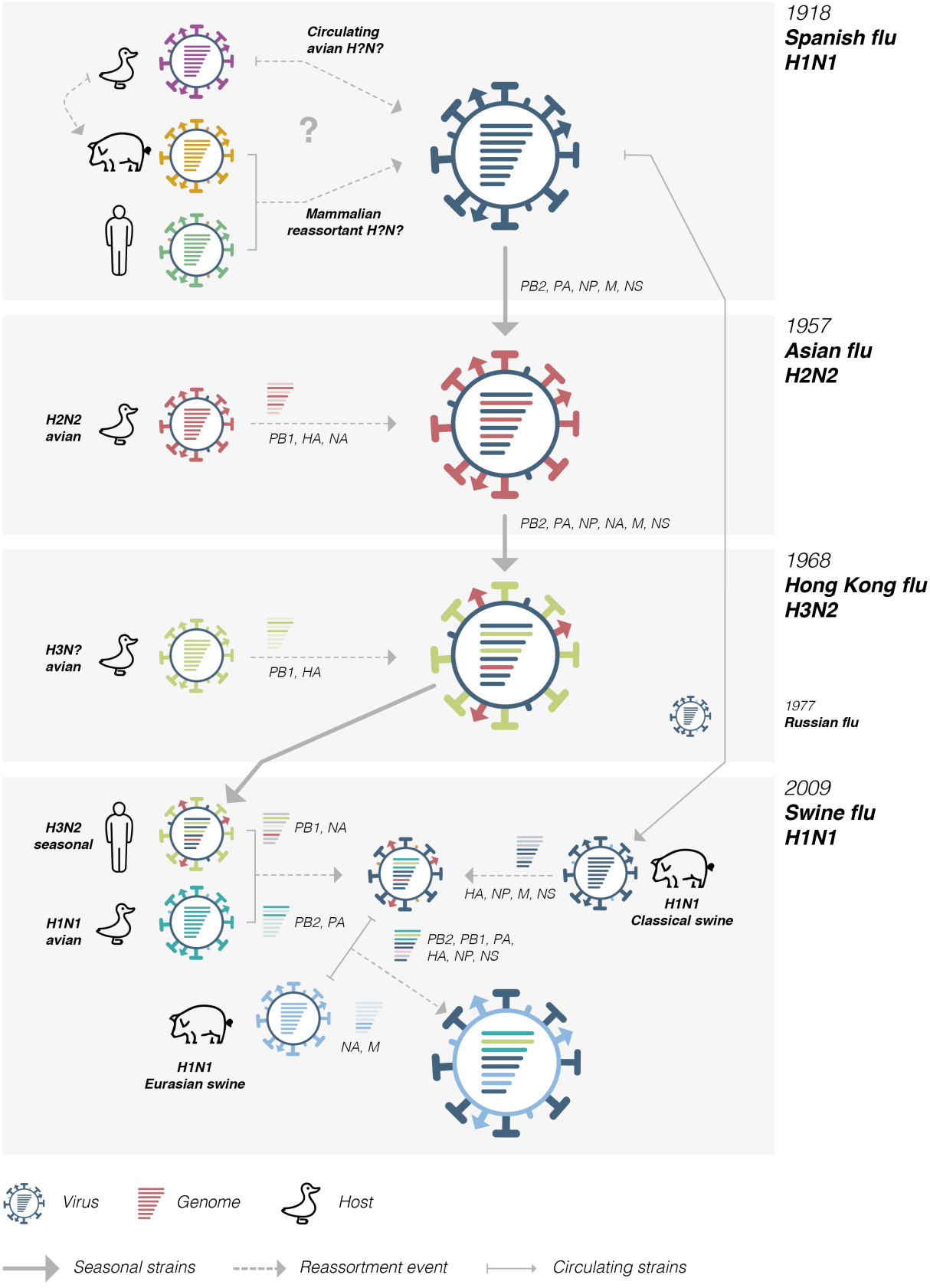


Figure 1.4 IAV pandemics of the last century

The "Spanish flu" pandemic occurred in 1918 and it killed approximately 50 million people worldwide. The 1918 virus originated from either a direct infection of humans by an avian virus (Taubenberger, Reid, Lourens, et al.*, 2005) or an avian virus, which reassorted with mammalian IAVs in an unknown host (Smith, Bahl, Vijaykrishna,* et al.*, 2009). The 1918 virus was an H1N1 strain and its derivatives*

circulated in the human population for the next 39 years. The next pandemic occurred in 1957 when, by then seasonal, H1N1 virus from humans reassorted with an avian H2N2 virus and produced an H2N2 virus with avian PB1, HA and NA segments. 11 years later, during the "Hong Kong flu" pandemic in 1968, the circulating H2N2 strain again exchanged its PB1 and HA segments with an avian strain. Finally, the most recent "Swine flu" pandemic occurred in 2009. Unlike in the previous two major pandemics, the H1N1 swine flu virus was a product of reassortment between viruses that have been circulating in pigs. A "triple reassortant" H1N2, which emerged in pigs in the late 90s (Ito, Couceiro, Kelm, et al.*, 1998), was a strain with the human H3N2 PB1 and NA segments, avian-origin PB2 and PA segments and the rest of the segments originating from the "classical swine" virus that had its origin in the 1918 outbreak. The triple reassortant underwent another reassortment event with a Eurasian IAV strain that circulated in pigs at that time and provided NA and M segments to ultimately produce the pandemic H1N1 (pH1N1) Swine flu (Smith, Vijaykrishna, Bahl,* et al.*, 2009). Currently, both H3N2 and pH1N1-derived IAV strains circulate in human population and causing seasonal epidemics each year. While, probably inevitable, it is unclear when, where and which strains will reassort and produce the next pandemic IAV virus.*

1.4.2 Genetic factors contributing to virus reassortment

While theoretically all IAV strains have the potential to reassort, not all possible combinations of viruses have been observed in nature. Moreover, experimental evidence suggests that the mixing between the segments from different strains is not a completely random event, with some segments showing a preferred co-segregation over others (Essere, Yver, Gavazzi, *et al.*, 2013; White, Steel & Lowen, 2017; Cobbin, Verity, Gilbertson, *et al.*, 2013). One of the difficulties in predicting the potential for a reassortment between the different IAV strains is the lack of comprehensive understanding of the factors that may contribute to segment incompatibility. From the molecular point of view, such incompatibilities can come either at RNA or protein level.

Given that the packaging signals mediate the selection of vRNAs into the viral progeny, incompatibility between the packaging signals among different strains would restrict the ability of viruses to reassort. Indeed, several studies have shown the requirement for compatible packaging signals to drive reassortment. In one study it was shown that during the reassortment between avian A/Finch/England/2051/91 (H5N2) and human A/Moscow/10/99 (H3N2) viruses the avian HA segment required the presence of avian PA and M segments in order to be incorporated into the human virus background (Essere, Yver, Gavazzi, *et al.*, 2013). Replacing the terminal regions of the avian HA with a human HA sequence allowed to efficiently incorporate otherwise avian HA into the human virus. Furthermore, mutations in a human M segment could be shown to eliminate the requirement for human HA packaging signals to be present for avian HA incorporation. A different study has observed that the A/Netherlands/602/2009 (H1N1) packaging signals limit incorporation of its HA segment into the A/Panama/2007/99 (H3N2) background; however, the Netherlands virus packaging signals did not restrict the incorporation of its NS or NA

segments into the Panama strain (White, Steel & Lowen, 2017). Packaging signals have also been shown to be partially responsible for the lack of reassortment between IAV and IBV HA segments (Baker, Nogales, Finch, *et al.*, 2014). Finally, a preferred co-segregation between the PB1 and NA segments of H3N2 strains has been observed during the production of seed strains for seasonal vaccine manufacturing (Cobbin, Verity, Gilbertson, *et al.*, 2013). Additional, experiments have shown that this segment co-segregation is driven by the NA vRNA sequence (Gilbertson, Zheng, Gerber, *et al.*, 2016; Cobbin, Ong, Verity, *et al.*, 2014).

Even if the packaging signals of the different viruses were compatible, mismatches between the encoded proteins could also restrict the ability of viruses to reassort. Such protein-mismatch has been especially well described between the subunits of 3P (Li, Hatta, Watanabe, *et al.*, 2008; Hatta, Halfmann, Wells, *et al.*, 2002; Song, Pascua, Lee, *et al.*, 2011; Octaviani, Goto & Kawaoka, 2011; Hara, Nakazono, Kashiwagi, *et al.*, 2013) as well as between the HA and NA proteins (Wagner, Matrosovich & Klenk, 2002; Rudneva, Kovaleva, Varich, *et al.*, 1993; Rudneva, Sklyanskaya, Barulina, *et al.*, 1996; Yen, Liang, Wu, *et al.*, 2011; Xu, Zhu, McBride, *et al.*, 2012). It has been shown, for example, that replacement of pH1N1 PA segment with a seasonal H1N1 protein significantly decreases polymerase activity, as measured by a luciferase minigenome reporter assay (Song, Pascua, Lee, *et al.*, 2011) and similar detrimental effects have been observed when the pandemic PB2 subunit was replaced (Octaviani, Goto & Kawaoka, 2011). The balance between the HA and NA segments is essential, due to the opposing functions of the two proteins in the virus life cycle. HA is required for attachment and entry into a cell, whereas NA mediates the exit and release. The rate between the sialic-acid binding and release has to be balanced to prevent virus from aggregating on the infected cell surfaces. A study using A/Memphis/8/88 (H3N2) and A/mallard/New York/6750/78 (H2N2) viruses was not able to generate viruses with either HA or NA genes from the Memphis virus in the New York background, but viruses with both glycoproteins from the same strain could be made (Hatta, Halfmann, Wells, *et al.*, 2002). Other studies have observed that replication of viruses with non-functional NA segments can be restored by mutations in the HA segment, which decrease its receptor binding affinity (Mitnaul, Matrosovich, Castrucci, *et al.*, 2000; Kaverin, Gambaryan, Bovin, *et al.*, 1998). Interestingly, NP has also been implicated in the control of HA/NA balance through the modulation of NA gene expression (Brooke, Ince, Wei, *et al.*, 2014), highlighting the complexity of gene-protein interactions required to fine-tune balance between multiple viral components during infection.

1.5 Secondary RNA structure prediction and modelling

1.5.1 Computational methods used to study secondary RNA structures

The simplest computational method used to predict secondary RNA structures is a free energy minimisation approach, which predicts the secondary RNA structure fold which is associated with the lowest minimum free energy (MFE) value and, theoretically, is the most likely structure to form. MFE predictions assume that the folding of an RNA is local and influenced solely by the identity of a base and its neighbouring nucleotides. To calculate the MFE structure a difference in the Gibbs free energies between the folded versus unfolded RNA is calculated (ΔG) and the structure associated with the lowest ΔG is considered to be the most prevalent structure present in solution. MFE approach is quite accurate for predicting secondary RNA structures of small naked RNAs and is particularly useful for RNAs for which no related RNA sequences are available and thus comparative approaches are not possible (Mathews, 2004). The major limitation of using MFE is that the predictions of protein-associated and longer RNA sequences are not very accurate, e.g. the accuracy of MFE predicted structure of an *E. coli* small rRNA is only 47.1% (Bellaousov & Mathews, 2010).

While most free energy minimisation algorithms will present a single MFE structure, in reality multiple secondary RNA structures with very similar low ΔG energies can always be predicted, furthermore, the assumption of a single secondary RNA structure conformation being present may also not be true for some RNAs (Garmann, Gopal, Athavale, *et al.*, 2015). A more versatile approach to structure prediction uses a partition function, which, rather than predicting a single RNA structure, calculates the probabilities for base-pairing between each nucleotide in a sequence (McCaskill, 1990). The calculated partition function can be used to visualise RNA structures in terms of the probability of their formation and can be suggestive of multiple RNA conformations. The nucleotides with above 99% base-pairing probabilities calculated using partition function have been found to be correctly paired 91% of the time (Mathews, 2004). Partition function calculation can be further used to predict the maximum expected accuracy (MEA) structure, which shows the structure that is based on the most probable base-pair combinations (Mathews, Moss & Turner, 2010).

When multiple related RNA sequences are available, a comparative sequence analysis approach for secondary RNA structure prediction has served as a useful and perhaps the most accurate tool. The basic assumption when comparing related sequences is that RNA structures of functional importance should be conserved despite the changes in their sequences. Consequently, if a change in nucleotide sequence occurs in one position of a functionally important RNA structure, a restorative change should occur in an interacting nucleotide to preserve the base-pairing between the two (referred to as covariance). The best example of the power of covariance analysis is the early predictions of transfer RNA structure (Madison, Everett & Kung, 1966; Levitt, 1969), which were determined to be highly accurate when the first crystal structures became available (Robertus, Ladner, Finch, *et al.*, 1974). However, the covariance analysis is still a difficult task as it requires the availability of multiple RNA sequences and their alignments that have to be based on structure rather than sequence conservation in order to accurately detect the compensating base-pair changes.

1.5.2 Experimental approaches to secondary RNA structure predictions

The major disadvantage of computational methods used to predict secondary RNA structures is the lack of context in which the RNAs exist. Most, if not all, RNA molecules involved in biological processes associate with proteins at some point in their existence (Flores & Ataide, 2018; Belfetmi, Zargarian, Tisné, *et al.*, 2016; Noller & Woese, 1981) and some RNA molecules have different conformations in the presence and absence of ligands (Garst, Edwards & Batey, 2011), while yet others may change their structure according to the changes in temperature (Kortmann & Narberhaus, 2012). The need to understand the secondary RNA structures formed by RNAs in different contexts has led to the development of a number of experimental approaches that allow to get experiment-based information about RNA conformation. Overall, the methods used for secondary RNA structure probing can be broadly divided into those that use enzymes to cleave RNA and those which use chemical reagents to modify RNA. Table 1.2 summarises some of the currently available experimental methods for secondary RNA structure analysis.

Table 1.2 Examples of genome-wide methods used for higher-order RNA structure analysis

		Method	Reagent	Conditions	Examples	Reference
Intramolecular RNA structure analysis	***Enzymatic probing***	FragSeq	RNase P1	*in vitro*	Non-coding RNA probing from *M. musculus* neuronal and embryonic stem cells	Underwood et *al*. (2010)
		PARS	RNase V1 and S1	*in vitro*	*S.cerevisiae* mRNA probing	Kertesz et *al.* (2010)
		PARTE	RNase V1 and temperature change	*in vitro*	*S.cerevisiae* total RNA probing	Wan et *al.* (2012)
		PIP-seq	and protein cross-linking	*in vitro*	HeLa cell transcriptome	Silverman et *al.* (2014)
	Chemical probing	SHAPE-MaP	1M6, 1M7, NMIA	*in vitro* & *in vivo*	Xist RNA *in vivo* probing	Smola et *al.* (2015)
		SHAPE-Seq	1M7	*in vitro* & *in vivo*	5S rRNA, RNase P and the btuB riboswitch probing in *E. coli*	Watters et *al.* (2016)
		CIRS-seq	DMS, CMCT	*in vitro*	*M. musculus* embryonic stem cell transcriptome probing	Incarnato et *al.* (2014)
		Structure-seq	DMS	*In vivo*	*O. sativa* mRNA probing	Ritchey et *al.* (2017)
		Mod-seq	DMS	*In vivo*	*S. cerevisiae* rRNA probing	Talkish et *al.* (2014)
		icSHAPE	NAI-N3	*in vitro* & *in vivo*	*M. musculus* embryonic stem cell transcriptome probing	Spitale et *al.* (2015)
Intermolecular and long-range RNA-RNA interactions	***Small molecule probing***	LIGR-Seq	AMT and circRNA ligase	*In vivo*	Transcriptome-wide RNA probing in *H. sapiens* cells	Sharma et *al.* (2016)
		SPLASH	biotinylated AMT	*In vivo*	Transcriptome-wide RNA probing in H. sapiens, *S. cerevisiae* and *E. coli* cells	Aw et *al.* (2016)
		RNA proximity ligation	T4 RNA ligase 1	*In vivo*	Abundant RNA structure probing in *S. cerevisiae* and *H. sapiens* cells	Ramani et *al.* (2016)
		PARIS	AMT and RNA ligase I	*In vivo*	Transcriptome-wide RNA probing in *H. sapiens* and *M. musculus* cells	Lu et *al.* (2016)
		MARIO	crosslinking and biotinylation	*In vivo*	*M. musculus* brain and embryonic stem cells transcriptome RNA probing	Nguyen et *al.* (2016)

The enzyme-based structure probing methods use RNases, which have a specific preference for cutting RNA depending on its structural features, e.g. RNase A cleaves at single stranded cytidine and uridine nucleotides, RNase V1 cleaves only base-paired nucleotides and RNase T1 cleaves only at single stranded guanines. The RNA digestion products can be analysed by denaturing gel electrophoresis and the digestion pattern can be used to predict the secondary RNA structure of a digested RNA. As with most RNA structure probing methods, enzyme digestion-based structure predictions have been adapted to be used together with various high-throughput sequencing methods, which allows parallel analysis of multiple RNAs (Kertesz, Wan, Mazor, *et al.*, 2010). The disadvantage of using enzymatic digestion approach is that it is limited to *in vitro* analysis due to the need for specific RNases and therefore, chemical modification-based methods have become more widely used.

The chemical probing-based methods use reagents that can selectively modify either the RNA base or its backbone. Dimethyl sulfate (DMS), for example, selectively modifies single stranded adenosine and cytidine nucleotides, whereas guanosine and uridine can be modified by CMCT (1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-*p*-toluenesulfonate). By contrast, a group of reagents, collectively called SHAPE reagents (from the method called selective 2'-OH acylation analyzed by primer extension), including 1M7 (1-methyl-7-nitroisatoic anhydride), 1M6

(1-methyl-6-nitroisatoic anhydride), NMIA (N-methylisatoic anhydride) and others, selectively modify the ribose 2′ -OH group of single stranded nucleotides (Bevilacqua, Ritchey, Su, *et al.*, 2016). When these structure-specific reagents react with an RNA molecule they leave a 'chemical map', which can be read by a simple primer extension. The reaction between the probing reagent and the nucleobase forms an adduct that acts as a barrier for the reverse transcriptase, causing it to prematurely terminate. Just like for the enzymatic probing, the truncated primer extension products can be analysed on denaturing gels together with sequencing ladders and interpreted to predict secondary RNA structures. Multiple high-throughput derivatives of chemical probing methods have now been developed, which proved to be especially useful for the probing of multiple RNAs form various biological sources such as in animal and plant cells (Spitale, Flynn, Zhang, *et al.*, 2015; Ding, Tang, Kwok, *et al.*, 2014; Kwok, Ding, Tang, *et al.*, 2013; Talkish, May, Lin, *et al.*, 2014) as well as viruses (Watts, Dang, Gorelick, *et al.*, 2009; Archer, Simpson, Watts, *et al.*, 2013; Pollom, Dang, Potter, *et al.*, 2013; Burrill, Westesson, Schulte, *et al.*, 2013).

Recent years have also seen the development of methods used to study not only secondary intramolecular RNA structures but the intermolecular ones as well (Table 1.2). In cells, RNA-RNA interactions are an important part of many biological processes. For example, microRNAs (miRNA), small 21–25-nucleotide long RNAs, base-pair with mRNAs and can mark an mRNA for degradation or repress its translation. miRNA-mRNA interactions have been shown to regulate multiple processes ranging from *C. elegans* neuronal development to flowering in *A. thaliana* plants (He & Hannon, 2004). In addition, interactions between small noncoding RNAs and pre-mRNAs are essential for correct splicing and maturation of mRNAs (Engreitz, Sirokman, McDonel, *et al.*, 2014). To understand the mechanistic basis of RNA-RNA interactions taking place in cells, various in cell small molecule probing methods have been developed (Sharma, Sterne-Weiler, O'Hanlon, *et al.*, 2016; Lu, Zhang, Lee, *et al.*, 2016; Aw, Shen, Wilm, *et al.*, 2016). Most of the methods for studying RNA-RNA interaction use an intercalating reagent, called psoralen, or its derivatives, which can cross-link base-paired RNAs and therefore capture RNA-RNA interaction taking place at any given time. On its own, psoralen-based crosslinking is somewhat limited as to how much resolution it can provide about the structural basis of an RNA-RNA interaction, however, it does reveal which RNAs interact with each other and, combined with chemical probing methods, can even be used for potentially modelling tertiary RNA structures (Sharma, Sterne-Weiler, O'Hanlon, *et al.*, 2016).

1.5.3 Secondary RNA structures in viruses

The importance of secondary RNA structures in viral genomes has by now been widely recognised. Overall, positive sense RNA viruses are considered to have more structured genomes compared to negative-sense RNA viruses, and the structuredness of the viral genome may be linked to the viral persistence (Simmonds, Tuplin & Evans, 2004; Davis, Sagan, Pezacki, *et al.*, 2008). Given the relatively small genome sizes of RNA viruses and therefore, a somewhat restricted protein-coding capacity, the presence of secondary RNA structures allows the RNA viruses to expand their genome functionality. Perhaps the 'classical' example of secondary RNA structure present in a virus genome is the internal ribosome entry site (IRES), which was first discovered in picornaviruses (Pelletier & Sonenberg, 1988) and later shown to be present in many other viral (Chiba, Jamal & Suzuki, 2018; Friis, Rasmussen & Belsham, 2012; Redondo, Angel Sanz, Steinberger, *et al.*, 2012) and cellular RNAs (Stoneley & Willis, 2004). IRES is a highly structured region of RNA, usually present upstream of the protein coding sequence in the 5' untranslated region of the viral RNA. The presence of IRES is used to recruit the ribosomal translation machinery to the viral RNAs, which otherwise lack the 5'-cap structures found in cellular RNAs and would therefore, be unrecognized by a ribosome.

Multiple studies have shown that secondary RNA structures play important roles in many stages of a viral genome replication- CRE elements in picornaviruses serve in primer formation (Steil & Barton, 2009), long-distance RNA-RNA interactions in the red clover necrotic mosaic and the barley yellow dwarf viruses are required for ribosomal-frameshifting and regulation of negative strand RNA synthesis (Barry & Miller, 2002; Tajima, Iwakawa, Kaido, *et al.*, 2011), replication elements in the flock house virus are important for premature termination and sub genomic RNA production (Lindenbach, Sgro & Ahlquist, 2002). In addition, multiple other elements, in a range of RNA viruses, have been shown to play important roles during the viral life cycle, though the precise step modulated by a secondary RNA structure has not always been well defined (Friebe, Boudet, Simorre, *et al.*, 2005; Sztuba-Solinska, Diaz, Kumar, *et al.*, 2016; Nicholson & White, 2015). In addition, secondary RNA structures have been shown to be important for virus assembly (Beren, Dreesens, Liu, *et al.*, 2017; Larson & McPherson, 2001) and genome packaging (Stewart, Bingham, White, *et al.*, 2016; Clever, Miranda & Parslow, 2002; Kim, Firth, Atasheva, *et al.*, 2011; Sung & Roy, 2014) in a number of different RNA viruses. While secondary RNA structures do

allow to expand the coding capacity of small viral genomes they can also be a double-edged sword as cells contain pattern-recognition receptor (PRRs). PRRs specifically recognize pathogen-derived secondary RNA elements and trigger signaling pathways, which activate host defense mechanisms (Jensen & Thomsen, 2012). For example, MDA-5 recognizes long double-stranded RNAs (Pichlmair, Schulz, Tan, *et al.*, 2009), RIG-I recognizes 5′-triphosphates and 5′-diphosphates of double or single stranded RNAs (Hornung, Ellegast, Kim, *et al.*, 2006; Wang, Ludwig, Schuberth, *et al.*, 2010; Goubau, Schlee, Deddouche, *et al.*, 2014a), TLR3 and TLR7 recognize double and single-stranded RNAs, respectively (Iwasaki & Medzhitov, 2004), and dsRNA-dependent protein kinase R recognizes greater than 30 nucleotide-long dsRNAs (Lemaire, Anderson, Lary, *et al.*, 2008).

Advances in chemical probing methods used to analyse secondary RNA structures have also enabled better understanding of the secondary RNA structures present inside the viral genomes. Through the use of selective 2-hydroxyl acylation analyzed by primer extension (SHAPE) method structural models of human immunodeficiency virus-1 (HIV-1) (Watts, Dang, Gorelick, *et al.*, 2009; Siegfried, Busan, Rice, *et al.*, 2014), simian immunodeficiency virus (Pollom, Dang, Potter, *et al.*, 2013), hepatitis C virus (Mauger, Golden, Yamane, *et al.*, 2015), poliovirus (Burrill, Westesson, Schulte, *et al.*, 2013), tomato bushy stunt virus (Wu, Grigull, Ore, *et al.*, 2013) and satellite tobacco mosaic virus (Larman, Dethoff & Weeks, 2017) RNA genomes are now available. The genome-wide analysis of the secondary RNA structures present in these viruses has not only confirmed multiple previously described secondary RNA structures but also identified new ones and has discovered some unexpected new functions for the structural RNA organization. For example, genome-wide analysis of the polio RNA structure has identified previously undescribed conserved secondary RNA structures in the virus open reading frame, which were shown to be important for virus replication and infectivity (Burrill, Westesson, Schulte, *et al.*, 2013). Perhaps the best studied secondary RNA structure amongst all the RNA viruses is that of HIV-1, for which experimental data for the vRNA inside the viral particles (Wilkinson, Gorelick, Vasa, *et al.*, 2008), deproteinated vRNA (Watts, Dang, Gorelick, *et al.*, 2009), dimerized vRNA (Kenyon, Prestwood, Le Grice, *et al.*, 2013) and vRNA in complex with Gag and Tat proteins are available (Kenyon, Prestwood & Lever, 2015). One of the most intriguing aspects of the vRNA structural organization, which was discovered through the SHAPE analysis of the HIV-1 genome, was the fact that the structure of RNA can influence the structure of the proteins it encodes. It was shown that the structured regions of HIV-1 vRNA fall into unstructured regions of proteins, such

as flexible linkers between structured protein domains. Further analysis of ribosomal toeprinting showed that the ribosomes slow down when reading through the structured regions of vRNA, which researchers have suggested may provide time for the structured protein domains to fold, thus implying a direct relationship between the RNA and protein structure (Watts, Dang, Gorelick, *et al.*, 2009). Overall, genome-wide studies of secondary RNA structures in viruses can clearly provide new insights into various aspects of viral biology and with the ever developing high-throughput structural analysis methods, which could be combined with classical structural biology approaches, an unprecedented ability to acquire high-resolution RNA structures becomes available.

1.6 Thesis aims

Given the lack of knowledge about the conformation of the vRNA in the context of RNPs, this thesis project aimed to do a global analysis of the secondary RNA structures present in the influenza virus genome. Multiple high-throughput methods were used to analyse of both intra- and inter-segment secondary RNA structures present inside the viral particles and the importance of the discovered secondary RNA structures was assessed using virus mutagenesis experiments.

2. Chapter 2: Materials & Methods

2.1 Plasmids and cloning

2.1.1 Plasmids

Bidirectional pHW2000 plasmids used for A/WSN/33 (WSN) protein and vRNA expression have been described before in Hoffmann et *al*. (2000). pcDNA plasmids expressing WSN PA, PB1, PB2 and NP proteins have been described before in Fodor et *al.* (2002) (Fodor, Crow, Mingay, *et al.*, 2002). pPOLI plasmids encoding vRNAs for different WSN segments were described previously in Fodor et *al.* (1999).

2.1.2 Plasmid maintenance

All plasmids were amplified and purified from *E. coli* DH5α cells. 1 ng of plasmid was added to 50 μL of chemically competent DH5α cells. The cells were incubated on ice for 20 min and heat-shocked for 30 s at 42°C. Cells were placed back on ice for 2 min and 200 μL of SOC Medium (Sigma-Aldrich) were added. Cells were incubated for 45 min at 37°C in a shaking incubator. 50 μL of transformed cells were plated on agar plates containing 100 μL/mL of ampicillin and placed at 37°C overnight. Single colonies were selected from the agar plates they were used to inoculate 5 mL of lysogeny broth (10 g bacto-tryptone, 5 g yeast extract,10 g NaCl per litre; LyB) containing 100 μL/mL of ampicillin. The inoculated LyB cultures were incubated in a shaking incubator at 37°C overnight. To purify the plasmids from the overnight cultures QIAprep Spin Miniprep Kit (QIAGEN) were used according to the manufacturer's instructions. The plasmids were eluted in 30-50 μL of water and the purified plasmid concentration was measured using NanoDrop 2000 spectrophotometer. For larger cultures Genopure Plasmid Maxi Kit (Roche) was used according to the manufacturer's instructions. All plasmids were stored at -20°C.

2.1.3 PCR mutagenesis

PCR mutagenesis was used to introduce mutations into pPOLI and pHW2000 plasmids when fewer than 8 mutations were required. The PCR reaction was prepared according to Table 2.1. 18 cycles of PCR amplification were performed by setting the temperature cycling to 95°C for 30 s, 50°C for 30 s and 72°C for 16 min.

Table 2.1 PCR mutagenesis reaction

Reagent	Amount (final concentration)
Plasmid	50 ng
dNTPs	1µL (10 µM)
Fw primer	1µL (10 µM)
Rv primer	1µL (10 µM)
10x Pfu buffer	5 µL (1x)
100% DMSO	1.5 µL (3%)
Pfu turbo	1 µL
H_20	38.5 µL

Following the PCR, the reactions were digested using DpnI and purified using QIAquick PCR Purification Kit, according to manufacturer's instructions. 2 µL of the purified reaction were used to transform competent DH5α cells. Mutated plasmids were amplified as described in section 2.1.2. All mutated plasmids were checked for the presence of the desired mutations and any additional changes by sequencing the plasmids using the Source BioScience services.

2.1.4 Gibson assembly

Gibson assembly was used for introducing more than 8 mutations into pHW2000 plasmids. Overlapping backbone primers and insert oligonucleotides for Gibson assembly were designed with overlap annealing temperatures between 50-53°C. For the assembly, plasmid backbones were amplified using Phusion HF polymerase with the reactions prepared according to Table 2.2. The PCR reaction was set for 30 cycles to 98°C for 30 s, 61°C for 30 s and 72°C for 5 min and a 2 min final extension at 72°C. The amplified plasmid backbones were gel-purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel), according to manufacturer's instructions. The plasmid

backbones and insert oligonucleotides were used to assemble the mutated plasmids using Gibson Assembly Master Mix (NEB), according to manufacturer's instructions. The assembly was performed for 1 h at 50°C. The assembly reactions were used to transform competent DH5α cells and the final plasmids were amplified as described in section 2.1.2. All mutated plasmids were checked for the presence of the desired mutations and any additional changes by sequencing the plasmids using the Source BioScience services.

Table 2.2 Phusion HF reaction for plasmid backbone amplification

Reagent	**Amount**
5x Phusion HF buffer	10 µL
10mM dNTPs	1 µL
10uM Fw primer	2.5 µL
10uM Rv primer	2.5 µL
100ng pHW2000	1 µL
DMSO	1.5 µL
Phusion HF	0.5 µL
H_2O	top to 50 µL

2.1.5 Primers used for mutagenesis

The primers used to produce viruses with mutations for SHAPE-MaP and SPLASH experiments are given in Table 2.3 and Table 2.4 , respectively.

Table 2.3 Primers used for producing viruses with mutated secondary RNA structures in SHAPE-MaP experiments

Stars next to the mutant names indicate the constructs that were produced using Gibson assembly protocol, other constructs were made using PCR mutagenesis.

Primer name	*Sequence*	*Mutant name*
NS_366_Rv_Flank	GATGGTGCCTTCCTCGGTGAAGGCCCTTAGTAATATTAG	NS 364-411 (17mut)*
NS_366_411_insert	GCAGTGATGGTAATGGACTGATCTCTCCCACGATGGTGCCTTCCTC	
NS_366_Fw_Flank	CCATTACCATCACTGCCAGGACATACTGATGAGGATG	
NS_364_411_new_Rv	CAGTATGTCCTGGAAGAGATGGTAGAGGAGAAATCCGCCAACAATTGTCC	NS 364-411 (5mut)
NS_364_411_new_Fw	GGACAATTGTTGGCGGATTTCTCCTCTACCATCTCTTCCAGGACATACTG	
PB1_1309_1391_Rv	GGCCAAAAACATGCGAGGATTTTGATTTTCATTCCATTTGGTG	PB1 1309-1391 (4mut)
PB1_1309_1391_Fw	CACCAAATGGAATGAAAATCAAAATCCTCGCATGTTTTTGGCC	
PB1 118 161 Rv	GATGCACGAATTGACTTTGAATCGGGAAGGATAAAGAAAGAGG	PB1 118-161 (3mut)
PB1 118 161 Fw	CCTCTTTCTTTATCCTTCCCGATTCAAAGTCAATTCGTGCATC	
NP 1475-1525_Rv	CATGGCGACCAAAGGAACTAAACGTTCTTACGAACAGATGG	NP 1475-1525 (3mut)
NP 1475-1525_Fw	CCATCTGTTCGTAAGAACGTTTAGTTCCTTTGGTCGCCATG	
NP 1221-1275_Rv	GTGCGGGGAACGACCCTAAGAAAACTGGAGGACCTATATACAGGAGAGTCGATGG	NP 1221-1275 (3mut)
NP 1221-1275_Fw	CCATCGACTCTCCTGTATATAGGTCCTCCAGTTTTCTTAGGGTCGTTCCCCGCAC	
M 110-207 RV	CGCTTTAAATACGGATTAAAGAGGGGGCCTTCTACGGAAGG	M 110-207 (4mut)
M 110-207 Fw	CCTTCCGTAGAAGGCCCCCTCTTTAATCCGTATTTAAAGCG	
PB2 31 79_Rw	GGAACTCTAGCATACTTACTGATAGTCAAACAGCGACCAAAAG	PB2 31-79 (3mut)
PB2 31 79_Fw	CTTTTGGTCGCTGTTTGACTATCAGTAAGTATGCTAGAGTTCC	
PB2 1301-1385 1mut rv	GCAAGCCGTGGATATTTGTAAGGCTGCAATGG	PB2 1301-1385 (6mut)
PB2 1301-1385 1mut Fw	CCATTGCAGCCTTACAAATATCCACGGCTTGC	
PB2 1301-1385 3mut Rv	CTGAGAATTAGTTCATCGTTTAGTTTTGGTGGATTCACATTTAAG	
PB2 1301-1385 3mut fw	CTTAAATGTGAATCCACCAAAACTAAACGATGAACTAATTCTCAG	
PB2 1301-1385 2mut Rv	GAACAAGCGGATCATCTGTGAAGAGAGAGGAAGAGG	
PB2 1301-1385 2mut Fw	CCTCTTCCTCTCTCTTCACAGATGATCCGCTTGTTC	
HA 996-1050 Rv	GCTGCAAGGCCCAAGGTTAAGGACCAACATGGTAGGATGAACTATTACTGG	HA 996-1050 (5mut)
HA 996-1050 Fw	CCAGTAATAGTTCATCCTACCATGTTGGTCCTTAACCTTGGGCCTTGCAGC	
HA 133-180 Rv	GGCGATCTACTCAACTGTTGCAAGTTCACTGGTGCTTTTGG	HA 133-180 (2mut)
HA 133-180 Fw	CCAAAAGCACCAGTGAACTTGCAACAGTTGAGTAGATCGCC	

Table 2.4 Primers used for producing viruses with mutated inter-segment interactions in SPLASH experiments

Stars next to the mutant names indicate the constructs that were produced using Gibson assembly protocol, other constructs were made using PCR mutagenesis.

Primer name	Sequence	Mutant name
SP1_PA_half_resto re_rv	CCTGATTAATGACCCTTGGGTTTTGC	PA 84-138 (2mut)
SP1_PA_half_resto re_fw	GCAAAACCCAAGGGTCATTAATCAGG	
SP10-PB2-Rv	ATAAAACTGCTTCCATTCGCAGCCGCACCCCCCAAGCAAAGTGG	PB2 437-480 (5mut)
SP10-PB2-fw	CCACTTTGCTTGGGGGGTGCGGCTGCGAATGGAAGCAGTTTTAT	
SP6-PB2-Rv	GGAGACGTGGTTTTAGTAATGAAACGGAAACGGAACTCGAGTATACTTACTG	PB2 116-133 (2mut)
SP6-PB2-Fw	CAGTAAGTATACTCGAGTTCCGTTTCCGTTTCATTACTAAAACCACGTCTCC	
SP2_NS	CTCAAAATGACCATGGCTTCAGTACCTGCATCGC	NS 593-637 (2mut)
SP2_NS	GCGATGCAGGTACTGAAGCCATGGTCATTTTGAG	
SP2-M-Rv	GCTGGATCGAGTGAGCAAGCAGCTGAAGCGATGGATATTGCTAG	M 382-426 (3mut)
SP2-M-Fw	CTAGCAATATCCATCGCTTCAGCTGCTTGCTCACTCGATCCAGC	
NA_1151_1225_flank_fw	CAGGATAGCACAAGCGTGATATTAACCGGCAATT	NA 1173-1206 (8mut)*
NA_1151_1225_flank_rv	GCTTGTGCTATCCTGCCCAGCAACAACTTTATAG	
NS_rv_M_NS_ID7	CAGTTAGGTACCGCGATGCTGGTGCTGAAGCGATGGTCATTTTGAG	NS 593-637 (5mut), restorative
NS_fw_M_NS_ID7	CTCAAAATGACCATCGCTTCAGCACCAGCATCGCGGTACCTAACTG	
M_fw_NS_M_ID7	GGCTGGATCCAGTGAGCAAGCGGCTGAAGCGATGGATATTGCTAGTC	M 382-426 (6mut), restorative
M_rv_NS_M_ID7	GACTAGCAATATCCATCGCTTCAGCCGCTTGCTCACTGGATCCAGCC	
NP_rv_NP_HA_ID94	CTCCATCTGTTCGTAAGAACGTTTAGTTCCTTTTGTCGCCATGATTTCG	NP 1479-1515 (3mut), restorative
NP_fw_NP-HA_ID94	CGAAATCATGGCGACAAAAGGAACTAAACGTTCTTACGAACAGATGGAG	
HA_rv_NP_HA_ID94	CCCCAGGGAGACAAAAAGAACTAGTGTACTGGCGACAGTTGAG	HA 112-145 (4mut), restorative
HA_fw_NP_HA_ID94	CTCAACTGTCGCCAGTACACTAGTTCTTTTTGTCTCCCTGGGG	
PA_1375_1447_flank_Rev	CTTCAGGAATGGCTCTATCCTGGCATTTACTTCTTTGGACATTTGAG	PA 1375-1447 (27mut)*
PA_1375_1447_ins1_HiFi	GTCAGGGAGGCGCAACGGTCTCGGCGTGGACTTCAGGAATGGCTCTATCC	
PA_1375_1447_flank_Fw	GCGCCTCCCTGACGGACCACCGTGCAGTCAACGGTCCAAATTCCTGC	
NS_597_706_flank_rv	CGATCTGTTTCCCGGCCCTTGTAGCGGTCTCAATGTCCAGACCGAGAGTG	NS 593-706 (39mut)*
NS_597_706_ins1_hifi	AGCTTCGTCGCTCTCCTCTTTGAGGATCCTTTCGACGATCTGTTTCCCGG	
NS_597_706_flank_fw	GGCTAGCGTGCCCGCTAGCCGCTACCTAACTGACATGACT	
HA_794_885_ins1	ATGTTTTGAAATGGAAGGTTACTATTAATTGACCCTTGTGGGGTCTGGCA	HA 794-885 (28mut)*
HA_794_885_flank_fw	CCTTCCATTTCAAAACATTCATCCGGTGACCATTGGAGAGTGCCCAAAATATG	
HA_794_885_flank_Rv	TGGGGTCTGGCATTTTGTATTGCATTCATGCATGCTCGCGTTTGAGGTGA	

2.2 Protein analysis

2.2.1 SDS polyacrylamide gel electrophoresis for viral proteins

To analyse purified virion protein content the virus samples were resuspended in Laemmli buffer (200 mM Tris- HCl (pH 6.8), 40 % glycerol, 2 % SDS, 4 % β-mercaptoethanol, 50 mM EDTA, bromophenol blue) and heated to 95°C for 10 min. Mini-protein gel apparatus (BioRad) was used to run SDS-PAGE gels. 4% stacking gel (125 mM Tris-HCl pH 6.8, 0.1 % SDS, 0.1 % APS, 0.01 % TEMED, 4% 29 acrylamide:1 bis-acrylamide) and 12% resolving gel (375 mM Tris-HCl pH 8.8, 0.1 % SDS, 0.1 % APS, 0.01 % TEMED, 12 % 29 acrylamide:1 bis-acrylamide) were prepared. 10 μL of virus sample and 5 μL of protein ladder (Precision Plus Protein All Blue Standards, Bio-Rad) were used. Gels were run for 1 h 20 min at 130 V. Gels were stained using SilverXpress silver staining kit (Invitrogen) according to the manufacturer's instructions.

2.3 RNA analysis

2.3.1 Oligonucleotide radiolabelling

6 μL of water, 1 μL of 10x PNK buffer A (Thermo Fisher), 1 μL of PNK (Thermo Fisher), 1 μL of primer (10 pmol/μL) and 1 μL of [γ-^{32}P]ATP were mixed together. The reaction was incubated for 1 h at 37°C and the radiolabelled oligonucleotide was purified using QiaQuick Nucleotide Removal Kit (Qiagen), according to manufacturer's instructions. Labelled primers were dissolved in 30 μL of nuclease-free water; the 5S rRNA primer (5'-TCCCAGGCGGTCTCCCATCC-3') was dissolved in 50 μL of nuclease-free water.

2.3.2 RNA polyacrylamide gel electrophoresis

For primer extension analysis (described in section 2.3.4) 8 % polyacrylamide gels were prepared. For analysis of vRNA extracted from virions 3-3.5 % polyacrylamide gels were prepared. Gel mix contained acrylamide:bisacrylamide (19:1), 7 M urea and 1x TBE buffer (90 mM Tris base, 90 mM boric acid, 2.5 mM EDTA) and was polymerised with 0.01% TEMED and 0.1% ammonium persulphate. For the purified vRNA, 150-200 ng of vRNA were used. The vRNAs and the RiboRuler High Range RNA Ladder (Thermo Scientific) were mixed with 2x RNA loading dye (95% formamide, 0.025% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol FF, 0.5 mM EDTA). The gel was run for 3 h at 150V and subsequently stained with SilverXpress silver staining kit (Invitrogen), according to the manufacturer's instructions. For the primer extension samples 8 μL of the reverse transcription reaction were used. The gels were run for 1 h 10 min at 1300V and phosphor screen imaging plates (Fujifilm) were used to detect the radiolabelled cDNA. The phosphor screens were scanned with a Fujifilm FLA-5000 scanner and quantitated with Aida Image Analyser software (Raytest). Where needed autoradiography was done on Biomax MS film (Carestream), using Kodak Biomax intensifying screens.

2.3.3 RNP reconstitutions

To analyse the vRNA expression from pHW2000 plasmids RNP reconstitutions were performed. To this end, HEK-293T cells (described in section 2.4.1) were transfected with pHW2000 plasmids coding for each of the polymerase segments, NP segment and a plasmid coding for the mutated segment. Transfections were performed in 24-well plates and 250 ng of each plasmid were used. A 1:1 ratio of Lipofectamine-2000 reagent to µg of DNA was used and transfection was performed according to manufacturer's instructions. 24 hours post transfection the supernatant was removed from the cells, 250 µL of TRIzol reagent (Thermo Fisher) were added to each well and the plates were incubated with agitation at room temperature for 5 min. The RNA was subsequently extracted according to the manufacturer's instructions and dissolved in 10 µL of nuclease-free water. The extracted RNA was frozen at -80°C until further use.

2.3.4 Primer extension analysis

For primer extension ~100 ng of purified RNA were used. The RNA was mixed with 0.25 µL of labelled primer (as described in section 2.3.1) and the sample was adjusted to a final volume of 5 µL. The sample was heated to 95°C for 5 min and cooled on ice. Reverse transcription master mix (20 mM DTT, 1x First Strand buffer (Invitrogen), 10 U/µL SuperScript III (Invitrogen), 1 mM dNTPs) was pre-warmed at 50°C for 2 min and 5 µL of the master-mix were added to the RNA sample. The reverse transcription was performed for 1 h at 50°C and SuperScript was inactivated at 70°C for 10 min. cDNA samples were resuspended in formamide loading dye (9 mL formamide, 200 µL 0.5 M EDTA (pH 8), bromophenol blue, xylene blue), heated to 95°C and frozen at -20°C until RNA polyacrylamide gel electrophoresis was performed.

2.3.5 RT-qPCR for vRNA segments

The vRNA was extracted from infected cell supernatants using Direct-zol™ RNA MiniPrep columns (Zymo Research), according to the manufacturer's instructions. ~ 40 ng of extracted vRNA was used for reverse transcription (RT) reaction. 0.25 µL of each RT primer, annealing to the vRNA (Table 2.5), were added to the extracted RNA and the sample volume was adjusted to 5 µL with water. The sample was heated to 95°C for 2 min and cooled on ice. RT master-mix was prepared according to Table 2.6 and incubated for 2 min at 37°C. 5 µL of RT master-mix were

added to the sample and the sample was incubated at 37°C for 1 h followed by 10 min at 75°C. The final sample was diluted to 80 µL with nuclease free water.

1 µL of diluted RT reaction product was used for qPCR reaction which was set up according to the Table 2.7 using Brilliant III Ultra-Fast Probe High ROX QPCR Master Mix (Agilent). 3 technical replicates were performed for each segment per sample. The TaqMan probe and primer sequences are provided in Table 2.8. The cycling conditions were: 95°C for 3 min followed by 40 cycles of 95°C for 5 s and 60°C for 10 s. qPCR was performed on ABI StepOnePlus machine. Relative segment quantities (RQ) were calculated as:

$$2^{Ct_{reference\ segment} - Ct_{segment\ of\ interest}}$$

The RQ values were subsequently used to assess the changes in the packaging of vRNA in the mutant versus the wild type viruses. The changes were measured by calculating the ratio between the segment RQ value of the mutant versus the wild type virus. If the content of the vRNA packaged into the virion was not affected by the introduced mutations, the RQ values for the different segments would be similar in the wild-type and mutat viruses giving a ratio close to 1.

Table 2.5 Reverse transcription primers

Primers target vRNA; + sign before a nucleotide indicates locked nucleic acid bases.

Primer	Sequence
vRNA_3_ga	GTTCAGACGTGTGCTCTTCCGATCTAGCG+AAAGCAGG
vRNA_3_aa	GTTCAGACGTGTGCTCTTCCGATCTAGC+A+AAAGCAGG

Table 2.6 Reverse transcription reaction components

Reagent	Amount
SSRTIII 10x buffer	2 µL
DTT (100mM)	1 µL
dNTP (10 mM)	0.5 µL
Water	1 µL
Rnasin (10 U/µL)	0.25 µL
SuperScript III	0.25 µL

Table 2.7 qPCR master-mix

Reagent	Amount
TaqMan probe (10uM)	0.25 µL
Primer fw (10uM)	0.4 µL
Primer rv (10uM)	0.4 µL
cDNA	1 µL
qPCR master mix	5 µL
H20	2.95 µL

Table 2.8 qPCR primers and probes

The primers named with '_Taq' are TaqMan probes labelled with fluorophores (JOE, 6FAM, TAM) and quenchers (BHQ1, BHQ2).

Primer	Sequence
PB2_Fw	CGACTGTTCGTCTCTCCCAC
PB2_Rv	AGAGGTGCTTACGGGCAATC
PB2_Taq	[JOE]ACAATGGTTGGGAGAAGAGCAACAGCT[BHQ1]
PB1_Fw	CACTGAACCCATTTGTCAACC
PB1_Rv	TGGCTTGTATTCAAGATGGATCG
PB1_Taq	[TAM]GCAGTGATAATGCCAGCACATGGTCC[BHQ2]
PA_Fw	TTTGGACCGCTGAGAACAGG
PA_Rv	GGATTCGAACCGAACGGCTAC
PA_Taq	[6FAM]CAACACCACGACCACTTAGACTTCCGG[BHQ1]
HA_Fw	GTTTGGTGTTTCTACAATGTAGGACC
HA_Rv	CGAAGACAGACACAACGGGA
HA_Taq	[TAM]CTGGAAGCAGTGAGTCGCA[BHQ2]
NP_Fw	ACCCCTCCAGAAGTTCCGAT
NP_Rv	GCTCACTGATGCAGGGTTCA
NP_Taq	[6FAM]GGCCGCAGGTGCTGCAGTCA[BHQ1]
NA_Fw	CTGTGTATAGCCCACCCACG
NA_Rv	GCAACCAAGGCAGCATTACC
NA_Taq	[JOE]TGCTGGGCAGGACTCAACTTCAGT[BHQ1]
M_Fw	GCTCACTCGATCCAGCCATT
M_Rv	AAGTGGCATTTGGCCTGGTA
M_Taq	[6FAM]TGCTGACTCCCAGCATCGGTCTCA[BHQ1]
NS_Fw	TCGATGTCCAGACCGAGAGT
NS_Rv	GAACTAGGTGATGCCCCATTCC
NS_Taq	[JOE]CGGCTTCGCCGAGATCAGAAGTCC[BHQ1]
PA_fw2	CGAAATTGGAGTAACAAGGAGAGAAG
PA_rv2	CTTATGGTGAATAGCCTGGTTTTGATCC
NA_fw2	ATCGAGAAGGGGAAGGTTACTAAATCAATAG
NA_rv2	CTAGGTTTTGGTCGAAGGACACC
NA_Taq2	[JOE]GATACCGGCAAAGTGATGTGTGTGTGCAG[BHQ1]
PA_Taq2	[6FAM]GCCACAAAGGCCGACTACACTCTCGATG[BHQ1]

2.4 Eukaryotic cell and virus culture

The different media used for cell culture experiments are summarised in Table 2.9.

Table 2.9 Media used for cell culture

Medium	Composition
Minimal Essential Medium (MEM)	NaHCO3, Earle's salts, phenol red (from Sigma)
Dulbecco's Modified Eagle Medium (DMEM)	Sodium pyruvate, glucose, phenol red, NaHCO3, L-glutamine (from Sigma)
Cell culture MEM	MEM, 2 mM L-glutamine, 10% foetal calf serum (FCS)
Cell culture DMEM	DMEM, 10% FCS
MEM infection medium	For WSN: MEM, 2 mM L-glutamine, 0.5% Foetal Calf Serum (FCS) For PR8, Udorn, FLB: MEM, 2 mM L-glutamine, 0.8 µg/mL TPCK-trypsin
Agarose overlay	1:1 ratio of 2% agarose dissolved in PBS and MEM infection medium

2.4.1 Eukaryotic cell culture

Human embryonic kidney 293T (HEK-293T) cells were maintained in cell culture DMEM medium. Madin-Darby Bovine Kidney (MDBK) and Madin-Darby Canine Kidney (MDCK) epithelial cells were grown in cell culture MEM medium. All cells were regularly passaged by detaching them from a cell culture flask surface using TrypLE Express trypsin and diluting 1 in 10 in fresh cell culture medium.

2.4.2 Viruses

The following viruses were used in experiments: A/Udorn/307/72 (Udorn), A/Puerto Rico/8/1934 (PR8), A/WSN/33 (WSN) (H1N1), and B/Florida/4/06 (FLB).

2.4.3 Virus amplification and cell infection

For the virus stock amplification MDBK (for WSN virus) or MDCK (for Udorn, PR8, and FLB viruses) cells were used. For all infections, cells were split a day before, so that on the day of infection the cell confluency would be ~70%. On the day of infections supernatants were removed from the cells and the cells were washed once with PBS. MEM infection medium, described in Table 2.9, were used for infections. Viruses were diluted in a small amount of the infection medium (enough volume to cover the cell monolayer) to a multiplicity of infection (MOI) of 0.01. The infection was performed by applying the virus-containing medium to the cells and incubating the cells for 1 h at room temperature. Subsequently, the infection inoculum was removed, cells were washed once with PBS and fresh infection medium was added. Infected cells were incubated in humidified incubators at 37°C in 5% CO_2 until clear cytopathic effects were observed and most of the cell monolayer has detached from the flask surface (48-72 hours post infection). The virus-containing cell supernatants were collected, centrifuged at 4000 rpm for 10min, aliquoted and stored at -80°C. The viral titres were determined using plaque assays as described in section 2.4.5.

2.4.4 Virus growth curve analysis

The growth kinetics of different viruses were analysed by infecting cells as described in section 2.4.3 using an MOI of 0.001. Infections were performed in 6-well plates. Cell supernatants were collected at desired time points and the viral titres were determined by plaque assays as described in section 2.4.5.

2.4.5 Plaque assays

For every virus sample the plaque assays were performed in triplicates in 12-well dishes. Cell infection was performed as described in section 2.4.3 using virus stocks that were serially diluted in MEM infection medium. After the infection step, cells were washed twice with PBS and 500 μL of 1% agarose overlay was applied to each well. The overlay was allowed to set at room temperature and the plates were placed in humidified incubators at 37°C in 5% CO_2 for 72h. The

overlay was removed and the plaques were stained using a coomassie blue solution (0.02% w/v Coomassie brilliant blue R, 7.5% acetic acid, 50% ethanol, in water).

2.4.6 Heamagglutination assay

Supernatants from infected cells were serially diluted (1 in 2 dilution series) in PBS in 96 well-plates. An equal volume of 0.5% chicken red blood cells was added to each well. The plates were placed at 4°C and haemagglutination units were counted after 45 min.

2.4.7 Virus reverse genetics

12-plasmid (Fodor, Devenish, Engelhardt, *et al.*, 1999) or 8-plasmid (Hoffmann, Neumann, Kawaoka, *et al.*, 2000) systems were used to rescue wildtype or mutated viruses. For 12-plasmids system pPOLI-PB2, pPOLI-PB1, pPOLI-PA, pPOLI-HA, pPOLI-NP, pPOLI-NA, pPOLI-M, pPOLI-NS, pcDNA-PB2, pcDNA-PB1, pcDNA-PA and pcDNA-NP were used. For 8-plasmids system pHW2000-PB2, pHW2000-PB1, pHW2000-PA, pHW2000-HA, pHW2000-NP, pHW2000-NA, pHW2000-M, pHW2000-NS were used. HEK-293T cells were transfected with 0.5 μg of each plasmid using Lipofectamine-2000 reagent, according to manufacturer's instructions, and plated in 6-well dishes. For negative control samples, one of the plasmids encoding a polymerase subunit was omitted. After transfection cells were placed in humidified incubators at 37°C in 5% CO2. After 48 h the supernatant from transfected cells was collected and 10 μL of it were used to infect MDCK cells as described in section 2.4.3. 48 hours post infection the supernatants from MDCKs were collected and stored at -80°C. Viral titres were measured using plaque assays as described in section 2.4.5.

2.4.8 Virus purification by ultracentrifugation

For virus purification for use in SHAPE-MaP and SPLASH experiments, for each virus 12 T175 flasks of MDCKs were infected as described in section 2.4.3 using MOI of 0.01. A final volume of 19 mL of MEM infection medium in each flask was used. 48 hours post infection, cell supernatants were transferred to 50 mL falcon tubes and clarified by spinning at 4,000 rpm for 10

min at 4°C. Clarified supernatants were transferred into thick-wall ultracentrifuge tubes and placed in a SW32 rotor (Beckman Coulter). The tubes were spun at 10,000 rpm for 15 min at 4°C. After the initial purification the supernatants were layered on 30% sucrose cushions (prepared in sterilised NTC buffer (100 mM NaCl, 20 mM Tris-HCl (pH 7.4), 5 mM $CaCl_2$). Samples were spun at 25,000 rpm for 1.5 h at 4°C. The supernatants and the sucrose cushion were aspirated and the virus pellet was resuspended in cold resuspension buffer (0.01 M Tris-HCl (pH 7.4), 0.1 M NaCl, 0.0001 M EDTA) overnight at 4°C.

2.4.9 Virus growth analysis in the presence of antisense oligonucleotides

MDBK cells were plated in 12-well dishes at 60-70% confluency. Cell supernatants were removed and the cells were washed once with PBS. Cell monolayers were transfected with 1μM of antisense oligonucleotide (ASO) using Lipofectamine-3000 reagent, according to manufacturer's instructions. 4 hours post transfection, the cells were washed twice with PBS and infected with WSN virus at an MOI of 0.01 as described in 2.4.3. 24 hours post infection cell supernatants were collected and the virus titres were determined using plaque assay as described in section 2.4.5. For each ASO three transfection replicates were performed.

2.5 Selective 2' -hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP)

2.5.1 vRNA sample preparation

In virio vRNA samples were prepared by purifying viruses as described in section 2.4.8 and SHAPE modification was performed directly on the purified virions as described in section 2.5.3.

Naked vRNA samples (*nkvRNA*) were prepared by purifying viruses as described in section 2.4.8. The purified viruses were treated with proteinase K (0.250 μg/mL proteinase K, 10 mM Tris-HCl (pH 7.0), 100 mM NaCl, 1 mM EDTA) for 40 min at 37°C and the vRNA was extracted using Direct-zol™ RNA MiniPrep columns (Zymo Research), according to the manufacturer's instructions. The samples were frozen at -80°C until further use.

In vitro transcribed vRNA (*ivtRNA*) samples were prepared by firstly making linear DNA templates of each WSN genome segments that contained T7 DNA dependent RNA polymerase (T7-Pol) promoters. To this end, pPOlI plasmids, coding for the different WSN genome segments, were used. The T7-Pol templates were produced using primers that annealed to the viral segment termini and had a T7-Pol promoter sequence, which was introduced upstream of the 5′ vRNA (Table 2.10). The PCR amplification was performed using Phusion HF polymerase as described in Table 2.2 and a reaction clean-up was performed using QIAquick PCR Purification Kit (Qiagen), according to the manufacturer's instructions.

HiScribe T7 polymerase (NEB) was used to produce vRNAs according to conditions in Table 2.11; the reaction was performed for 2 h at 37 °C. Produced vRNAs were treated with TURBO DNase (Thermo Fisher), purified using RNeasy Mini Kit (QIAGEN) and the correct transcripts were confirmed by running the samples on a 3.5% PAGE-Urea gel as described in 2.3.2. The vRNA transcripts were frozen at -80°C until further use.

Table 2.10 Primers used to produce linear templates of WSN genome segments and to add T7-Pol promoters

Segment	*Primer name*	*Primer*
PB2	T7-PB2 Fw	TAATACGACTCACTATAGGGAGTAGAAACAAGGTCG
	PB2 Rv	AGCGAAAGCAGGTCAATTATATTC
PB1	T7-PB1 Fw	TAATACGACTCACTATAGGGAGTAGAAACAAGGCA
	PB2 Rv	AGCGAAAGCAGGCAAACCA
PA	T7-PA Fw	TAATACGACTCACTATAGGGAGTAGAAACAAGGTAC
	PA Rv	AGCGAAAGCAGGTACTG
HA	T7-HA Fw	TAATACGACTCACTATAGGGAGTAGAAACAAGGGTG
	HA Rv	AGCAAAAGCAGGGGAAAATAAAAAC
NP	T7-NP Fw	TAATACGACTCACTATAGGGAGTAGAAACAAGGGTA
	NP Rv	AGCAAAAGCAGGGTAGATAATC
NA	T7-NA Fw	TAATACGACTCACTATAGGGAGTAGAAACAAGGAG
	NA Rv	AGCGAAAGCAGGAGTTTAAATG
M	T7-M Fw	TAATACGACTCACTATAGGGAGTAGAAACAAGGTAG
	M Rv	AGCAAAAGCAGGTAGATATTG
NS	T7-NS Fw	TAATACGACTCACTATAGGGAGTAGAAACAAGGGTG
	NS Rv	AGCAAAAGCAGGGTGACAAAG

Table 2.11 HiScribe T7 reaction

Reagent	Amount
Template RNA	8 µL (300 ng)
10x T7 buffer	2 µL
NTPs	2 µL of each NTP from a 100 mM stock
HiScribe T7	1.5 µL
RNasin	0.5 µL

2.5.2 1M7 vRNA modification *ex virio*

1 µg of purified *nkvRNA* or 1 µg *ivtRNA* (for each vRNA transcript) was used for 1M7 and DMSO reactions. For both 1M7 and DMSO reactions, 3 µL of water were added to 9 µL of RNA. The samples were heated to 95°C for 2 min and placed on ice for 2 min. 6 µL of folding buffer (333 mM HEPES (pH=8), 333 mM NaCl, 33 mM $MgCl_2$) were added to each sample. The vRNA was folded by incubating it at 37 °C for 20 min. 1.676 µL of 100 mM 1M7 or 1.67ul of DMSO were added to a 0.65 mL PCR tubes.15 µL of RNA from the RNA folding reactions were added the 1M7 or DMSO containing PCR tubes, mixed and immediately transferred to 37°C heat block for 75 s. After RNA modification the sample volume was adjusted to 50 µL with H_2O and RNA Clean & Concentrator kit (Zymo Research) was used for the reaction clean-up. For the *ivtRNA* reactions the different vRNA segment samples were pooled together after the modification step in equimolar rations.

2.5.3 1M7 and NMIA vRNA modification *in virio*

Each purified virus sample from section 2.5.1 was divided into two equal-volume aliquots in 0.65 mL thin-wall PCR tubes. Into each aliquot 1/10th of its volume of 100 mM 1M7 or DMSO was added. The virus samples were incubated at 37 °C for 75 s for the SHAPE modification to take place. The samples were then treated with proteinase K and the modified vRNA was extracted as described in section 2.5.1 for the *nkvRNA*.

Modification with N-Methylisatoic Anhydride (NMIA, Thermo Fisher) was done exactly as the 1M7 modification, except the incubation time during the modification step was increased to 45 min.

The successful modification was checked by performing ^{32}P-labelled primer extensions as described in section 2.3.4, using NA segment-targeting primer (5'-AATTGGTTCCAAAGGAGACG-3').

2.5.4 SHAPE-MaP library preparation

100 ng of purified and modified vRNA was used to prepare SHAPE-MaP sequencing libraries. RNA samples were diluted to 9.65 μL volume in 0.65mL PCR tubes. 1.35 μL of random nonamer primer (50 μM stock, Sigma) was added to each sample. Samples were heated to 65°C for 5min and cooled on ice for 2 min. SHAPE-MaP reaction buffer was prepared by mixing 1:1 ratio of SHAPE-MaP pre-buffer (250 mM Tris (pH=8), 375 mM KCl, 50 mM DTT, 2.5 mM of each dNTP) and 30 mM $MnCl_2$. 8 μL of SHAPE-MaP reaction buffer were added to each sample. The samples were incubated at 25°C for 2 min and 1 μL of SuperScript II reverse transcriptase (200 U/μL, Thermo Fisher) was added to each sample. Samples were incubated at 25°C for 10 min, 42°C for 3 h, 70°C for 15 min and held at 4°C overnight. The RT reactions were purified over G-25 micro spin columns (GE Healthcare) and adjusted to the final 68 μL volume with water. Second stand DNA synthesis was performed using NEBNext mRNA Second Strand Synthesis Module (NEB), according to manufacturer's instructions. PureLink PCR micro elution columns (Invitrogen) were used to purify dsDNA and the samples were diluted to 0.2 ng/μL in water. Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1024) was used to produce indexed reads, according to manufacturer's instructions; 13 PCR cycles were used for the final amplification step.

The excess primers and primer dimers were removed from the libraries using Agencourt AMPure XP beads (Beckman Coulter). 45 μL of the beads were added to the final PCR reactions and the samples were incubated for 5 min at room temperature. The samples were placed on magnetic stands and the supernatant was discarded. While still on the magnetic stand, the samples were washed with 80% ethanol 3 times. After the final wash, the DNA-bound beads were

left to dry for 15 min at room temperature. The samples were resuspended in 17 μL of water and the DNA was allowed to dissociate from the beads for 2 min. 15 μL of the supernatant was transferred into a new microcentrifuge tube and the sample concentration and fragment sizes were measured using a Qubit fluorometer and Agilent High Sensitivity DNA Kit, respectively. All samples were pooled into a single 2 nM library and sequenced using 1x150 bp reads with NextSeq 500/550 High Output v2 kit on a NextSeq 500 (Illumina) platform (final library diluted to 1.8 pM was loaded into the sequencer). The NMIA-modified libraries were sequenced using 2x150 bp reads on a HiSeq4000 instrument (Illumina) by the Wellcome Trust Centre for Human Genetics.

2.5.5 SHAPE-MaP sequencing read analysis

The sequencing reads were trimmed using Trim Galore tool for Cutadapt (Martin, 2011) and genome alignment and mutation counting was performed using Shape Mapper 2 software (Busan & Weeks, 2018). Commands for the trimming and alignment are provided in Appendix B.

2.5.6 Secondary RNA structure prediction

SHAPE reactivities were used to model secondary RNA structures using RNAStructure (Bellaousov, Reuter, Seetin, *et al.*, 2013) package. *Fold* function was used to predict minimum folding energy containing secondary RNA structures. *Partition* function was used to calculate partition functions for structures. *ProbabilityPlot* function was used to generate the pairing probabilities between the nucleotides from the partition function. *MaxExpect* function was used to generate maximum expected accuracy structures from the partition functions. Scripts provided by Prof. Alain Laederach were used to draw secondary RNA structures based on the connectivity table and probability plot files. The structure esthetics were adjusted using Adobe Illustrator. Commands for the structure modeling are provided in Appendix B.

2.6 Sequencing of psoralen crosslinked, ligated, and selected hybrids (SPLASH)

MDBK or MDCK cells were infected at MOI of 0.01 as described in section 2.4.3; 12 T175 flasks were used for each virus analysed. 48 hours post infection the cell supernatants were collected and the viruses were purified and concentrated as described in section 2.4.8. The components of the buffers used in the SPLASH protocol are provided in Table 2.12.

Table 2.12 SPLASH buffers

	Reagent	Concentration
Lysis buffer (LB)	Tris-Cl pH=7	50 mM
	EDTA	10 mM
	SDS	1%
	LB is supplemented with RNasin on the day of use	

	Reagent	Concentration
Proteinase K buffer (PKB)	NaCl	100 mM
	Tris-Cl pH=7	10 mM
	SDS	0.50%
	EDTA	1 mM

	Reagent	Concentration
Hybridisation Buffer (HB)	NaCl	750 mM
	SDS	1%
	Tris-Cl pH=7	50 mM
	EDTA	1 mM
	Formamide	15%
	HB is supplemented with RNasin on the day of use	

	Reagent	Concentration
2x SSC Wash Buffer (WB)	20x saline-sodium citrate (SSC) &NaCl (Invitrogen)	dilute to 2x
	SDS	0.50%

2.6.1 Cross-linking

Purified virus sample was transferred to a well in a 24-well dish. 0.01% of digitonin (dissolved in water) and 50 μg/mL of AMT-B (dissolved in DMSO) were added to the virus sample. The plate was incubated at 37°C for 5min, with an occasional agitation. The plate was transferred onto ice, its lid was removed and replaced with a 4mm glass cover slide. The plate was positioned ~10 cm away from a UV lamp (UVP Ultra Violet Product Handheld UV 292 Lamp, Fisher). The sample was irradiated for 45 min with a 365 nm UV wavelength. Cross-linked samples were digested with proteinase K (0.250 μg/mL proteinase K, 10 mM Tris-HCl (pH 7.0), 100 mM NaCl, 1 mM EDTA) for 40 min at 37°C and the RNA was extracted using TRIzol LS Reagent (Thermo Fisher), according to the manufacturer's instructions.

2.6.2 Size-selection

The purified RNA from section 2.6.1 was dissolved in 18 μL of water and NEBNext Magnesium RNA Fragmentation Module (NEB) was used to fragment the RNA according to the manufacturer's instructions (4 min fragmentation was used). The fragmented RNA was selected for less than 200 nucleotide length fragments using RNA Clean & Concentrator™-5 (Zymo Research), according to the manufacturer's instructions.

2.6.3 Cross-linked RNA enrichment

100 μL of Hydrophilic Streptavidin Magnetic Beads (NEB) were washed 3 times with 1 mL of LB. Beads were resuspended in 100 μL of LB supplemented with 1:200 of RNasin (Promega). The beads were transferred into 15 mL falcons and 2 mL of HB supplemented with 1:200 of RNasin were added. ~2 μg of fragmented RNA from section 2.6.2 were added to the beads. To hybridise the RNA to the beads, the beads were incubated for 30 min at 37°C with an end-to-end rotation. After hybridisation the beads were placed on a magnetic stand for 2 min and the supernatant was removed. 1 mL of warm WB was added to the beads and the samples were transferred to 1.5 mL micro-centrifuge tubes. The beads were washed 5 times with WB by incubating them for 5 min in a shaking thermal block and removing the supernatants each time using a magnetic stand.

2.6.4 On-bead proximity ligation

The beads from section 2.6.3 were washed twice with ice-cold T4 PNK buffer (NEB) by incubating the beads at 4°C with an end-to-end rotation for 5 min, spinning the beads down and removing the supernatant. To remove the 3' cyclic phosphate groups the beads were incubated for 4 h at 37°C with T4 PNK (NEB), according to the Table 2.13.

Table 2.13 T4 PNK cyclic phosphate removal reaction

Reagent	Volume
H2O	66 µL
T4 PNK buffer (10x)	8 µL
RNasin	2 µL
T4 PNK (10U/ µL)	4 µL

To add 5′-monophosphate to the RNA, the above reaction was subsequently mixed according to Table 2.14 and incubated for 1 h at 37°C.

Table 2.14 T4 PNK 5'-phosphorylation addition reaction

Reagent	Volume
T4 PNK reaction	80 µL
T4 PNK buffer (10x)	2 µL
T4 PNK (10U/ µL)	5 µL
ATP (10 mM)	10 µL
H2O	3 µL

The beads were washed twice with cold T4 PNK buffer and resuspended in 110.66 µL of water. The sample was divided into two equal volumes. Proximity ligation was performed using RNA ligase 1 (Rnl1, NEB) according to Table 2.15 and the reactions were incubated for 16 h at 16°C.

Table 2.15 Rnl1 proximity ligation reaction

Reagent	Volume
T4 PNK reaction	55.33 µL
Rnl1 buffer (10x)	8 µL
Rnl1 (30U/ µL)	6.67 µL
ATP (10 mM)	8 µL
RNasin	2 µL
H2O	add 6.67 µL to reaction without Rnl1

2.6.5 AMT-B reverse crosslinking

Beads from section 2.6.4 were washed twice with 1 mL of WB and resuspended in 100 μL of PKB. The reactions were heated for 10 min at 95°C, spun down and chilled on ice. 100 μL of TRIzol reagent (Thermo Fisher) were added to each sample and the samples were vortexed for 3 min. Samples were incubated for 10 min at room temperature, 100 μL of chloroform were added and the samples were vortexed for 10 s. Samples were centrifuged for 15 min at 20,000 rpm at 4°C. Aqueous phase of the sample was used to extract RNA with RNA Clean & Concentrator™-5 (Zymo Research), according to manufacturer's instructions. The samples were eluted in 100 μL of water and placed on ice. While on ice, the samples were placed into UVP CL-1000L UV Crosslinker and irradiated with 254 nm wavelength for 5 min.

After the irradiation the RNA was precipitated with ethanol by adding 10 μL of 3 M sodium acetate, 1 μL of GlycoBlue and 300 μL of ice cold 100% ethanol and incubating the RNA for 1 h at -80°C. The samples were subsequently centrifuged for 30 min at 20,000 rpm at 4°C and washed once with 70% ethanol. RNA pellets were dried at room temperature and dissolved in 8ul of water.

2.6.6 SPLASH library and sequencing

For sequencing, the RNA libraries were prepared using a SMARTer smRNA-Seq kit (Takara bio), according to the manufacturer's instructions, using 12 PCR cycles for the final amplification step. After the final PCR amplification the reactions were precipitated with 100% ethanol as described in section 2.6.5 and dissolved in 12 μL of water. The samples were size-selected by running them on a Novex 6% TBE Gel at 150V until the bromophenol blue dye reached the bottom of the gel. The gel was stained using 1x SYBR Gold Nucleic Acid Gel Stain (Thermo Fisher) for 10min in TBE and visualised with transillumination. The DNA bands between 250-400 nucleotides were cut out from the stained gel and the selected gel was crushed by spinning it through a thin-walled PCR tube, which was pierced with a 25G needle at the bottom. The DNA was eluted from the gel by adding 450 μL of elution buffer (1M sodium acetate, 1mM EDTA), incubating samples for 2 h on a shaking block at 37 °C and spinning the gel and the eluate through a Corning Costar Spin-X column (Sigma). 1 mL of ice-cold 100% ethanol and 0.5 μL of GlycoBlue were added to the sample and the DNA was precipitated as described in section 2.6.5. The DNA pellets were air-dried, resuspended in 8 μL of water and the sample concentration was measured using Qubit

Fluorometer. The sample fragment size distribution was measured using Agilent High Sensitivity DNA Kit. The DNA libraries were pooled to a single 2nM sample for multiplexing and sequenced using NextSeq 500/550 High Output v2 kit (150 cycles) on a NextSeq 500 (Illumina) platform using final dilutions of 1.8 pM or 2 pM.

2.6.7 SPLASH sequencing read analysis

All scripts used for SPLASH sequencing analysis are provided in Appendix C and the steps involved in the sequencing data processing are explained in Chapter 5. In brief, the reads were trimmed using Trim Galore tool for Cutadapt (Martin, 2011) and aligned using STAR (Dobin, Davis, Schlesinger, *et al.*, 2013). Identified chimeric reads in the sequencing data were deduplicated based on their CIGAR strings, which were also used to extract their alignment coordinates. The alignment coordinates for each read were used to draw a read matrix from which the most common inter-segment interactions were selected. Sequence and SHAPE reactivities for the most common interactions were extracted based on their coordinates and these were used to model the interactions using IntaRNA(v2.2) (Mann, Wright & Backofen, 2017). IntaRNA structures were drawn using VARNA (Darty, Denise & Ponty, 2009) and the esthetics were added using scripts provided by Prof. Alain Laederach and Adobe Illustrator. The interaction map was drawn using Circos software (Krzywinski, Schein, Birol, *et al.*, 2009) and additional graphics were added using Adobe Illustrator. The probability density distribution plots were drawn using ggplot2 library in RStudio and the statistical differences between the distributions were calculated using Wilcoxon signed-rank test.

3. Chapter 3: Establishment of SHAPE-MaP for influenza virus

3.1 Overview

Secondary RNA structures play important roles in multiple biological processes (Nicholson & White, 2015; Mandal & Breaker, 2004; Mortimer, Kidwell & Doudna, 2014). Consequently, multiple methods have been developed to enable us to better understand the structure of RNA in different contexts (Bevilacqua, Ritchey, Su, *et al.*, 2016). More recently, methods have been developed which allow to experimentally probe secondary RNA structures present inside virions (Wilkinson, Gorelick, Vasa, *et al.*, 2008). Furthermore, some secondary RNA structure probing methods can now be coupled with high-throughput sequencing, allowing to analyse multiple secondary RNA structures present in different RNA molecules at the same time. This chapter describes the establishment of SHAPE-MaP (selective 2'-hydroxyl acylation analysed by primer extension and mutational profiling) method to probe naked and RNP-associated vRNAs.

3.2 Background

Multiple approaches are available for secondary RNA structure analysis and they can be broadly divided into *in silico* and RNA probing-based methods. Generally, *in silico* approaches use free energy minimization-based models to predict the folding of an RNA. The basic principle of RNA folding assumes the canonical base-pairing between G:C and A:T bases, sometimes allowing for noncanonical interactions, such as G:U, to occur as well. *In silico* predictions can be further coupled with additional parameters, including folding temperature, ion concentration, sequence covariance, etc. (Lorenz, Wolfinger, Tanzer, *et al.*, 2016). However, often the information about the additional parameters influencing the RNA folding is not available, and therefore, *in silico* predictions are frequently found to be inaccurate (Doshi, Cannone, Cobaugh, *et al.*, 2004). These inaccuracies are especially prominent when it comes to predicting RNA structures for larger molecules that are present in complex systems, such as cellular environments. Generally, cellular RNAs are associated with proteins, which dramatically influence their folding in ways that are difficult to predict just based on the RNA sequence itself. Consequently, methods which can be used to probe RNAs in their native state have been developed.

RNA probing methods fall into two categories - enzymatic and chemical probing. Enzymatic probing works by analysing the fragments of RNA, which have been produced by RNA digestion with a range of RNases. The RNases used in the enzymatic probing have different preferences for cutting an RNA that depend on its conformation (e.g. whether the RNA is single or double stranded) and so the RNA digestion patterns can be interpreted to predict the RNA structure. By contrast, chemical probing uses chemical reagents that selectively modify individual RNA bases depending on their conformation. The reagents used for chemical probing can either modify the RNA base itself or the sugar-backbone of an RNA molecule. For example, the use of DMS (dimethyl sulphate) to probe RNA will methylate the single stranded adenosines and cytidines on the Watson-Crick side of the base, whereas the use of SHAPE reagents will modify the 2′-OH groups on the ribose backbone, independently of the type of base present (Bevilacqua, Ritchey, Su, *et al.*, 2016).

SHAPE-MaP is an RNA probing method, which allows to selectively modify only flexible, most likely single-stranded, regions of RNA (Siegfried, Busan, Rice, *et al.*, 2014). The SHAPE reagents used in the SHAPE-MaP procedure react with the 2′-OH groups in the RNA backbone (Merino, Wilkinson, Coughlan, *et al.*, 2005). The base-paired RNA nucleotides are constrained in the ability of the 2′-OH to sample different chemical conformations. As only certain 2′-OH conformations favour the reaction with the SHAPE reagents, only the single stranded flexible nucleotides will be able to adopt these conformations and react with the SHAPE reagents. Several different reagents are available for SHAPE probing. All SHAPE reagents have similar anhydride structures but differ in their functional groups, which determines their half-life; probing with different SHAPE reagents can be used to get information about the structural dynamics of an RNA molecule (Wilkinson, Vasa, Deigan, *et al.*, 2009). 1M7 (1-methyl-7-nitroisatoic anhydride) is the most widely used SHAPE reagent with a half-life of 17 s, other commonly used reagents are 1-methyl-6-nitroisatoic anhydride (1M6) and N-methylisatoic anhydride (NMIA), with 31 and 260 s half-times, respectively (Smola, Rice, Busan, *et al.*, 2015). The reaction between the RNA base and the SHAPE reagent forms a chemical adduct, which can be detected using a reverse transcription (RT) reaction. In the SHAPE-MaP procedure RT is performed in the presence of Mn^{2+} ions, which make the reverse transcriptase more error prone. Under such conditions the reverse transcriptase will introduce a point mutation or cause deletions or insertions at the site of SHAPE modification to occur. When SHAPE modification and RT steps are performed in parallel to vehicle control (RNA

treated with DMSO), mutations introduced by SHAPE reagent treatment (as opposed to background RT mistakes) can be quantified. The mutation rates for every nucleotide of an RNA are then converted into SHAPE reactivity values, which are now well-established pseudoenergy constains for many RNA folding algorithms (Busan & Weeks, 2018; Smola, Rice, Busan, *et al.*, 2015).

While a number of previous studies have looked at the secondary RNA structures present in the influenza virus genome, most of these studies have used either *in silico*-based predictions or probing of naked vRNAs (Moss, Priore & Turner, 2011; Moss, Dela-Moss, Kierzek, *et al.*, 2012; Kobayashi, Dadonaite, van Doremalen, *et al.*, 2016; Ferhadian, Contrant, Printz-Schweigert, *et al.*, 2018; Gavazzi, Yver, Isel, *et al.*, 2013). As the vRNA is always coated with the viral NP and polymerase proteins, such approaches are clearly limited in their ability to understand the native vRNA structures. It has generally been considered that the binding of NP during vRNA replication melts any secondary RNA structures in the vRNPs (Baudin, Bach, Cusack, *et al.*, 1994). However, the fact that vRNPs purified from IAV virions are sensitive to the RNase V1 (Yamanaka, Ishihama & Nagata, 1990), which cuts only dsRNA, suggests that dsRNA is present in vRNPs. Furthermore, recent studies looking at the binding of the NP in the vRNPs, show that regions of low-NP binding in the vRNPs exist and that the vRNA in the NP-depleted sites can form secondary RNA structure (Lee, Le Sage, Nanni, *et al.*, 2017; Williams, Townsend, Wylie, *et al.*, 2018). The SHAPE reagents have been previously used to probe naked vRNAs and cRNAs (Ruszkowska, Lenartowicz, Moss, *et al.*, 2016; Lenartowicz, Kesy, Ruszkowska, *et al.*, 2016) but no RNP-associated vRNA analysis has been performed. Studies on other enveloped (HIV-1, KSHV and MHV) and nonenveloped (STMV) viruses (Wilkinson, Gorelick, Vasa, *et al.*, 2008; Sztuba-Solinska, Rausch, Smith, *et al.*, 2017; Larman, Dethoff & Weeks, 2017; Yang, Liu, Wudeck, *et al.*, 2015) have used SHAPE experiments to analyse the *in virio* structures of the viral genomes. This chapter describes the establishment of the use of SHAPE-MaP to study the structure of the influenza virus genome both *in virio* (inside the virus particles) and *ex virio* (using vRNAs lacking any associated proteins).

3.3 Aims

- Determine whether SHAPE reagents are suitable for the study of the *in virio* structure of the influenza virus genome.

- Establish SHAPE-MaP protocol for analysis of *in virio* and *ex virio* secondary RNA structures in the influenza viruses.

3.4 Results

3.4.1 *In virio* SHAPE modification

The ability of SHAPE reagents to modify vRNA inside the viral particles was assessed first. To this end, WSN virus was grown in MDBK cells and purified over a 30% sucrose cushion (Figure 3.1 A).

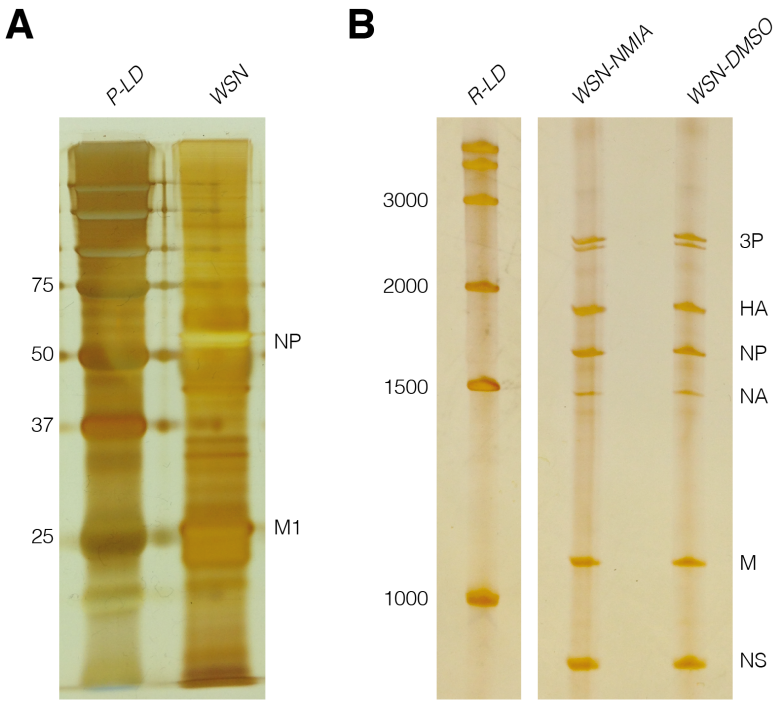


Figure 3.1 WSN virus purification

*WSN virus was purified over 30% sucrose cushion and protein (**A**) and RNA (**B**) contents were analysed on SDS-PAGE or polyacrylamide gels, respectively. RNA samples were extracted from purified viruses after treatment with NMIA or DMSO. RNA ladder and RNA samples were run on the same gel. P-LD, protein ladder, size shown in kilodaltons; R-LD RNA ladder, size shown in nucleotides.*

Purified virions were resuspended in a small volume of resuspension buffer and divided into two equal aliquots; one aliquot was used for SHAPE reagent treatment and another one for a control, DMSO-treated, reaction. Initial experiments were preformed using commercially available SHAPE reagent, NMIA, later experiments were done using custom synthesised 1M7 reagent. It was

important to perform SHAPE modification reactions in the smallest volumes possible, as the reagents hydrolyse quickly in water. After the *in virio* modification the virions were treated with proteinase K and the vRNA was extracted (Figure 3.1 B).

Purified vRNA from SHAPE and control-treated samples does not show any differences in the migration pattern on a polyacrylamide gel (see Figure 3.1 B), therefore, primer extension experiments were performed. Primers aligning to the NA segment were radiolabelled with ^{32}P and RT reaction on modified and control treated vRNA was performed. The reaction between the SHAPE reagent and the ribose 2'-OH group forms an adduct, which acts as a barrier for the reverse transcriptase, causing a premature termination to occur. In accordance, when the primer extension products were separated on a polyacrylamide gel, samples treated with the SHAPE reagents had increased number of premature termination products and lacked the full-length extension product (Figure 3.2). It is worth noting, that the lack of full-length extension product in the SHAPE reagent-treated samples also indicates that the SHAPE reagents can penetrate virions.

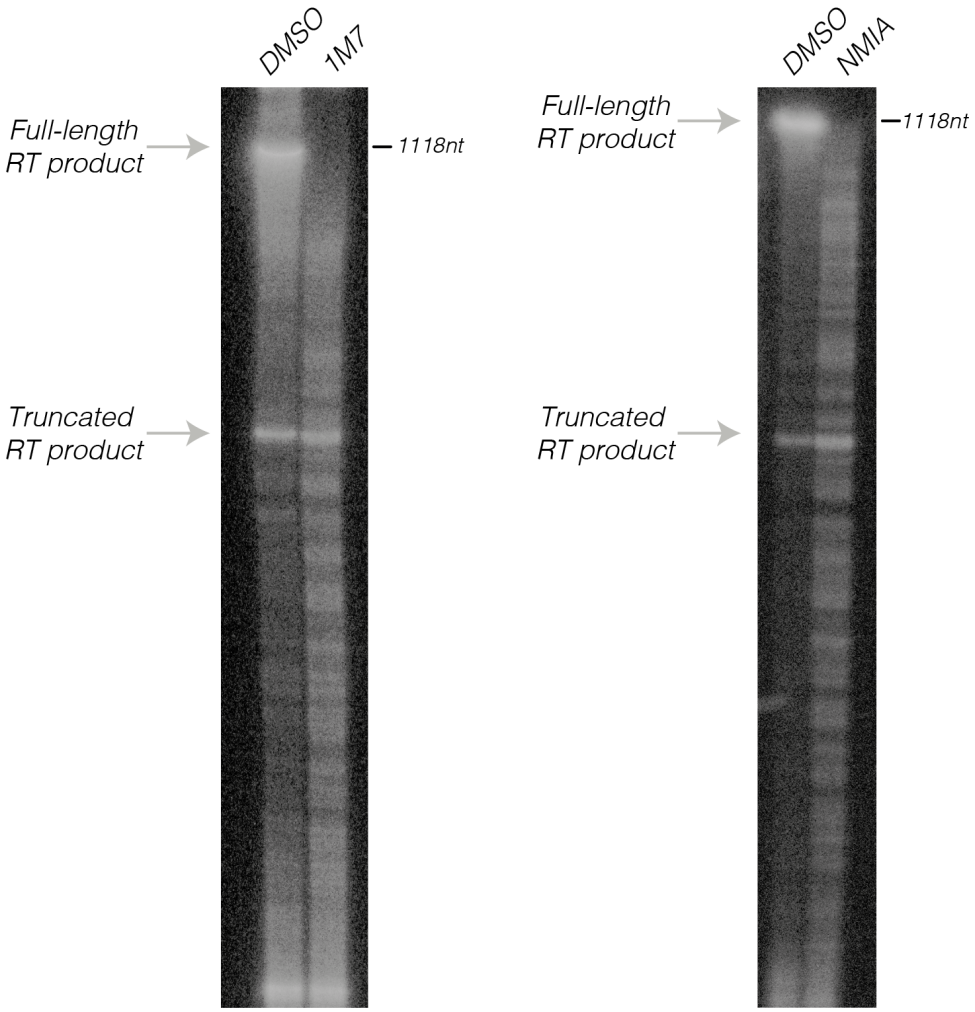


Figure 3.2 In virio SHAPE modification

RT reaction products from the vRNA samples treated with different SHAPE reagents (1M7 and NMIA) or DMSO. The lack of full-length RT product and multiple premature termination bands in the SHAPE reagent-treated samples indicate that vRNA can be successfully modified in virio. RT was performed using P^{32} labeled primer (5'-AATTGGTTCCAAAGGAGACG-3') targeting NA segment vRNA.

3.4.2 SHAPE-MaP reverse transcription

The SHAPE-MaP protocol is a derived-version of the originally published SHAPE analysis (Merino, Wilkinson, Coughlan, *et al.*, 2005), which relies on a mutagenic behaviour of the reverse transcriptase in the presence of Mn^{2+} ions. The Mn^{2+} ions cause the reverse transcriptase to introduce mismatch, deletion, insertion or cause a premature termination at the site of 2'-O-adduct, left by the SHAPE reagent modification Therefore, next the conditions of the RT reaction in the presence of Mn^{2+} ions were tested. *In virio* modification was performed as described in the previous section and the RT reaction step was performed in the presence of 12 mM $MnCl_2$. Both DMSO or SHAPE reagent-treated reactions showed an increased rate of premature termination in the presence of Mn^{2+} ions compared to reactions in the absence of Mn^{2+}, though the effect was greatly amplified in the SHAPE reagent-treated sample (Figure 3.3). It should be noted that the exact type of RT mutagenesis cannot be distinguished on a radiolabelled gel.

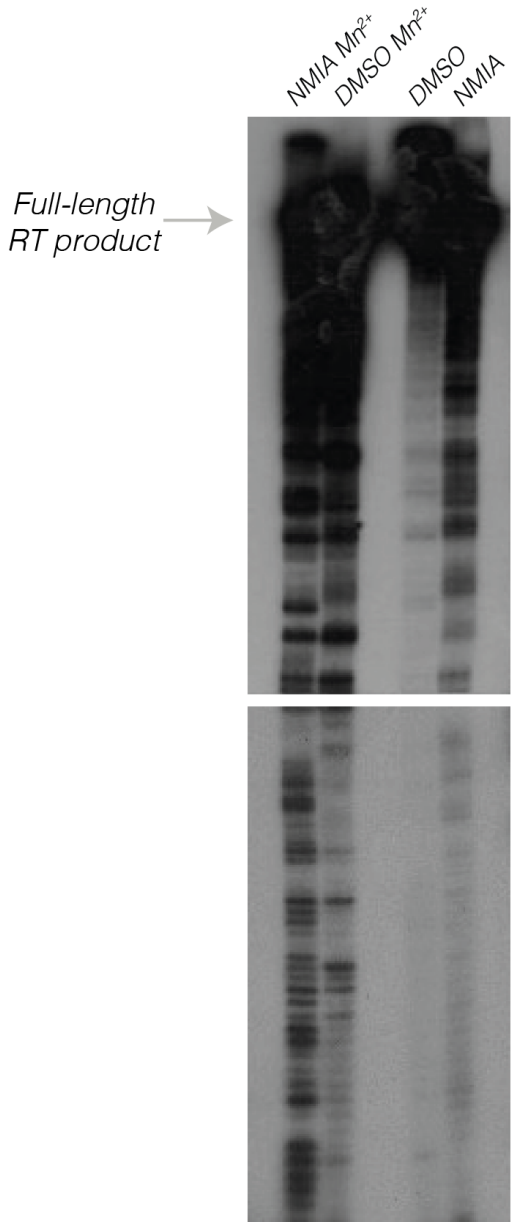


Figure 3.3 Reverse transcription in the presence of Mn^{2+}

Presence of Mn^{2+} causes increased premature termination and this effect is amplified in the SHAPE reagent-treated samples. The top and bottom of the gel were exposed to separate X-ray films.

3.4.3 *Ex virio* SHAPE modification

To be able to assess the effect of NP binding on the structure of the vRNA, two different *ex virio* samples were prepared. Naked viral RNA (*nkvRNA*) was prepared by treating the purified WSN virions with proteinase K, extracting the vRNA, folding it at 37°C, and finally modifying it with 1M7 reagent. *In vitro* transcribed vRNA (*ivtRNA*) was prepared by transcribing each vRNA segment from a DNA template using a T7 RNA polymerase (Figure 3.4). Individual vRNA segments were folded at 37°C and modified with 1M7. After modification all segments where pooled together for SHAPE-MaP sequencing library preparation. A summary of the different samples used for the SHAPE-MaP analysis is provided in Figure 3.5.

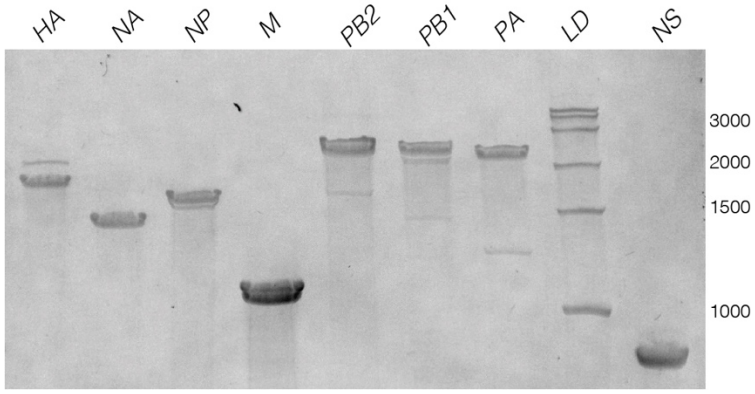


Figure 3.4 ivtRNA samples made using T7 RNA polymerase

Different T7 transcribed vRNA segments were separated on a 3.5% polyacrylamide gel and stained with SYBR Gold stain. LD, ladder, size shown in nucleotides.

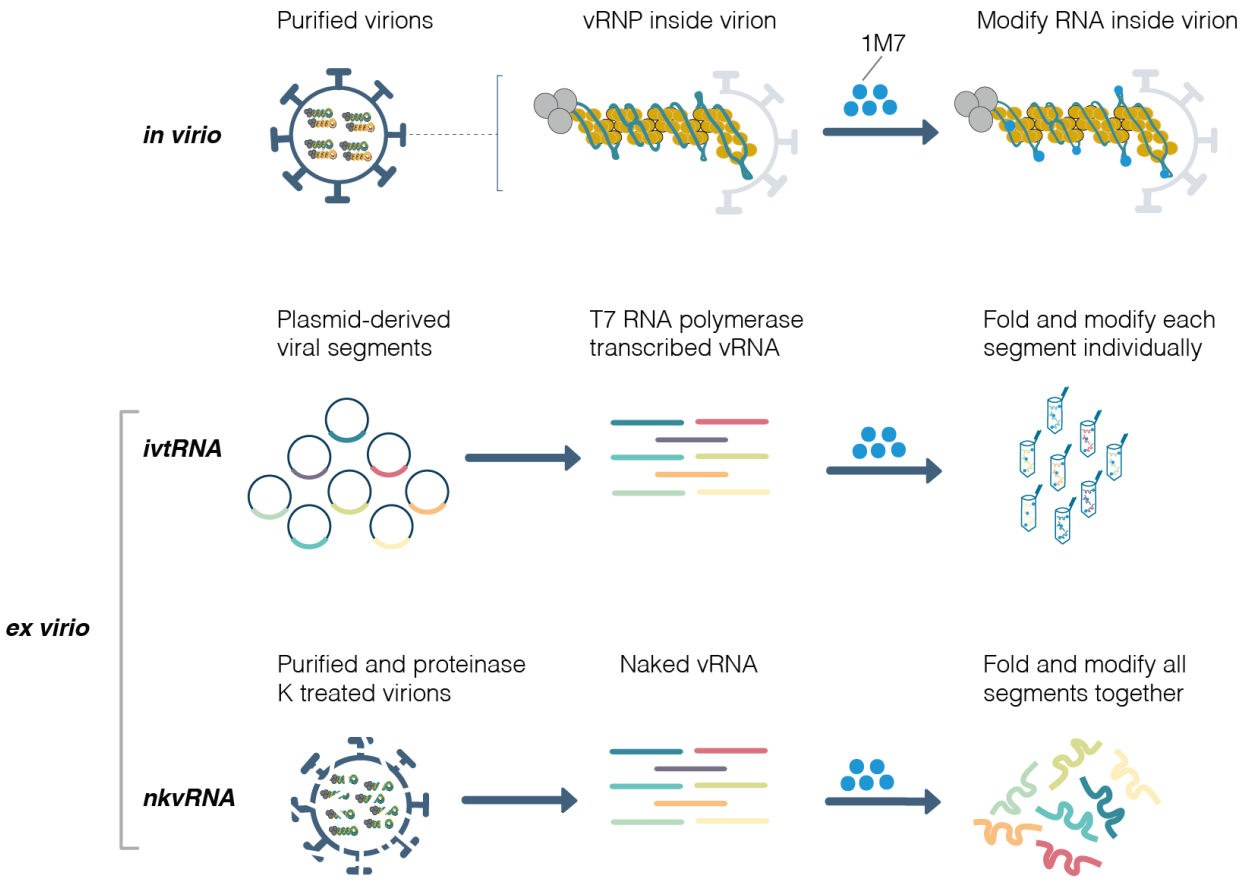


Figure 3.5 SHAPE-MaP sample summary

In virio samples are prepared by directly treating purified virions with the SHAPE reagents. ivtRNA samples are prepared by transcribing individual vRNA segments from DNA plasmids, folding the transcripts and modifying them individually. nkvRNA samples are from proteinase K treated and purified vRNA extracted from purified virions.

3.4.4 SHAPE-MaP library preparation, sequencing and analysis

SHAPE-MaP libraries for all vRNA samples were prepared as described previously (Smola, Rice, Busan, *et al.*, 2015); the steps involved in the SHAPE reactivity analysis are summarized in Figure 3.6. In brief, the SHAPE-modified RNA was used for RT in the presence of Mn^{2+} ions. The DNA-RNA hybrid, which is created during the RT reaction, was converted to a dsDNA using NEBNext Second Strand Synthesis module (NEB). The module uses RNAse H to fragment the RNA strand in the DNA-RNA hybrid, DNA polymerase I to produce the second DNA strand using the RNA fragments and DNA ligase to ligate the gaps between the newly transcribed DNA. The dsDNA was used for sequencing library preparation with Nextera kit (Illumina). The Nextera kit uses transposomes to preform a tagmentation reaction, during which the transposomes cut the dsDNA and add a known DNA sequence to the dsDNA termini. Tagmentation effectively both

fragments the DNA and adds a DNA sequence, which is complementary to the illumina primers used in the PCR amplification. During a limited-cycle PCR amplification multiplexing indices and sequencing primer sequences were added. Sequencing for the NMIA reagent-treated samples was performed using 2x150 bp reads on MySeq system, however, the terminal ends and A/T rich regions within some viral segments did not provide enough sequencing depth to reliably calculate SHAPE reactivities (see HA segment in Appendix A. 4). Therefore, later experiments with the 1M7 reagent were performed using either 2x150 bp sequencing on a HiSeq4000 instrument or 1x150 bp on a NextSeq500 instrument. All 1M7 experiments provided sufficient depth to achieve reliable sequencing depths for most nucleotides, however, between 4 and 26 nucleotides at the termini of each segment were still underrepresented in the sequencing data, most likely due to ineffective terminal priming during RT and bias in the transposome tagmentation sites.

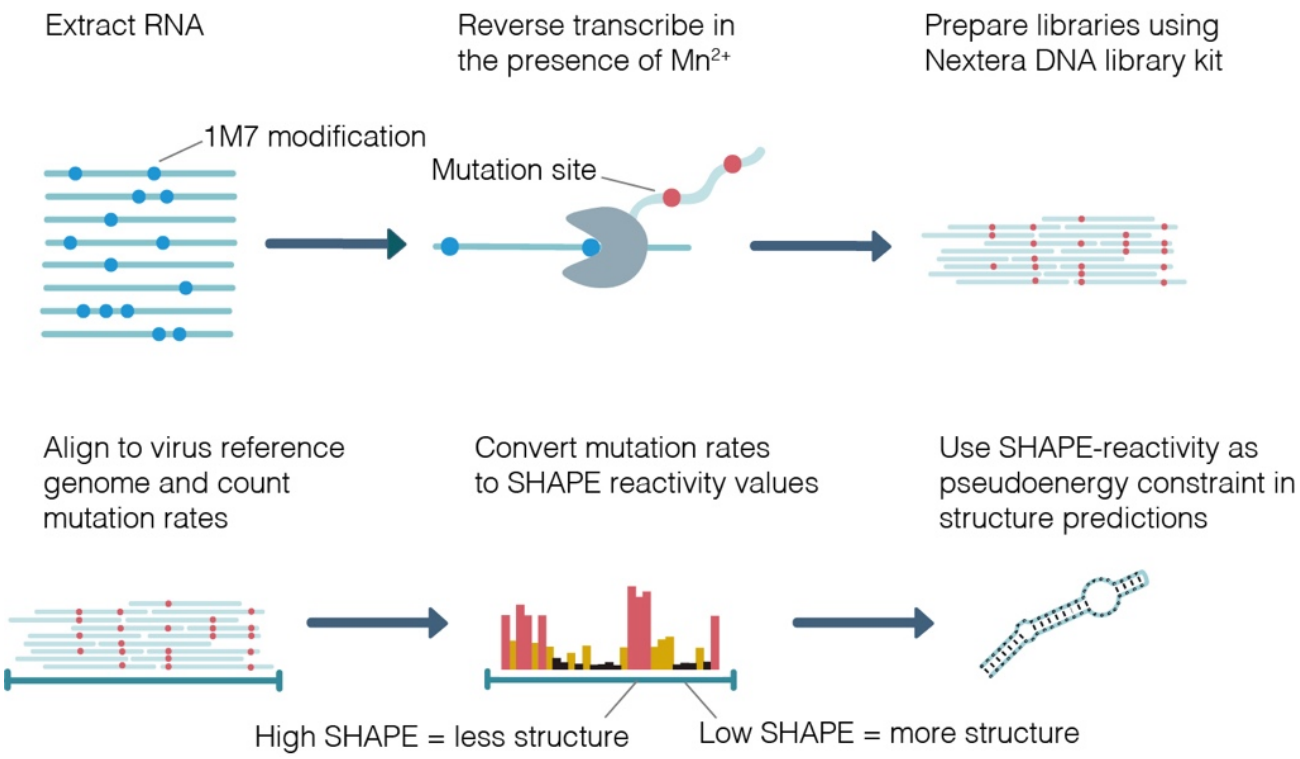


Figure 3.6 SHAPE reactivity analysis

Modified RNA is transcribed to cDNA in the presence of Mn^{2+}, which causes reverse transcriptase to introduce mutations at the sites of SHAPE modification. The cDNA is used to prepare a DNA library for sequencing. Sequencing reads are aligned to a reference genome and mutation rates for each nucleotide are calculated. Mutation rates are converted into the SHAPE reactivity values. SHAPE reactivities are used for SHAPE-informed RNA structure folding.

Overall, three replicates of NMIA-treated *in virio,* three replicates of 1M7-treated *in virio*, two replicates of 1M7-treated *ivtRNA*, and a single 1M7-treated *nkvRNA* experiments were

performed. The summary statistics and the raw SHAPE reactivities for the individual segments are provided in the Appendix A. Overall, the mutation rates in the SHAPE modified samples were significantly higher (p-value < 2.2^{-16}) as compared to the DMSO treated samples (Figure 3.7). The mutation rates were used to calculated SHAPE reactivities for every nucleotide using the following formula:

$$SHAPE\ reactivity = mutr_{SHAPE} - mutr_{DMSO}$$

, where $mutr_{SHAPE}$ is mutation rate in the SHAPE reagent-treated sample and $mutr_{DMSO}$ is mutation rate in the DMSO-treated control. The reactivities were normalised to an approximate 0-2 scale by excluding the outlier values (defined as nucleotides with reactivities greater than 1.5x of the interquartile range) and dividing by the average reactivity of the 10% most highly reactive nucleotides. Only the nucleotides with sequencing depths greater than 5000 were included in SHAPE analysis. Any nucleotides that had background mutation rates greater than 0.05 were excluded from the analysis.

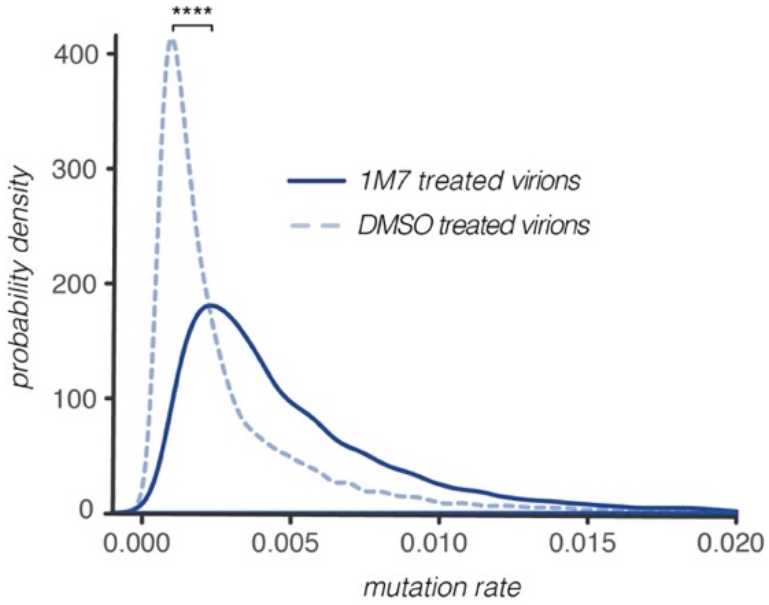


Figure 3.7 Probability density distribution of the mutation rates in 1M7 and DMSO-treated samples

*Distribution shows fewer mutated nucleotides in the DMSO-treated samples and more nucleotides associated with higher mutation rates in the 1M7 treated samples. **** P < 0.0001 Wilcoxon rank sum test.*

The obtained data using different reagents were highly reproducible and showed high correlation within *in virio* and *ex virio* groups (Figure 3.8) (the biological significance of the latter observation will be discussed in the next chapter).

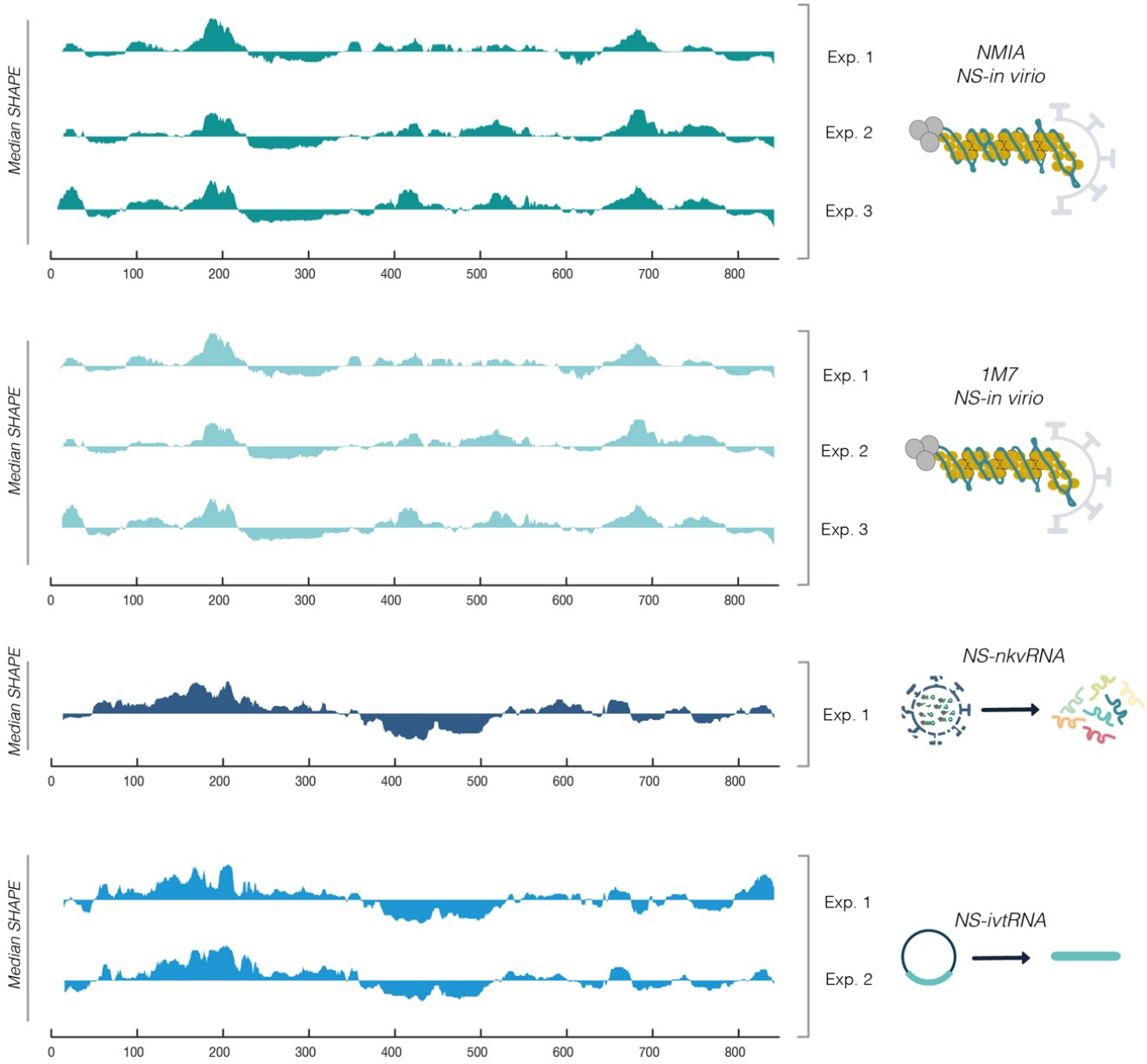


Figure 3.8 Reproducibility of the SHAPE-MaP experiments

Median SHAPE reactivities were calculated over sliding 50nt windows and plotted relative to the global median SHAPE reactivity of the segment. Representative median plots for the NS segment are shown. The data is plotted in a 5'-3' direction for the vRNA.

3.4.5 Discussion

The aim of this chapter was to establish the use of SHAPE-MaP method for the analysis of secondary RNA structure in the influenza virus genome. While previous studies have performed *in virio* SHAPE analysis for several different viruses (Wilkinson, Gorelick, Vasa, *et al.*, 2008; Sztuba-Solinska, Rausch, Smith, *et al.*, 2017; Larman, Dethoff & Weeks, 2017; Yang, Liu, Wudeck, *et al.*, 2015), no such analysis for influenza virus has been done so far. Therefore, the ability of the SHAPE reagents to penetrate the viral envelope and the matrix layer was established first. It was shown that the SHAPE reagents can successfully modify the vRNA directly inside the virions, in the context of the vRNP complex. A number of independent replicates of the SHAPE-MaP analysis showed highly reproducible results, implying the feasibility of the method as well as sufficient underlying stability in the vRNA structure for further analysis.

It is important to note that the exact effect of nucleoprotein (NP) binding on the SHAPE reactivities is unknown. Previous studies have shown that protein binding can affect SHAPE reactivities. Direct protein binding to single-stranded RNA can result in low SHAPE reactivities as a result of the 2'-OH protection by the bound protein, however, weaker protein binding may not have the same effect (Smola, Calabrese & Weeks, 2015); little is known about the strength of NP binding to the vRNA across the vRNPs. To avoid the issue of NP having an effect on SHAPE reactivities, experiments using *nkvRNA* and *ivtRNA* were performed. While little is known about how NP is loaded onto the vRNA during genome replication, it is reasonable to assume that this process is co-replicative - as the newly made vRNA emerges from the polymerase it wraps around the NP. The initial folding of an RNA molecule is primarily a local process (Fallmann, Will, Engelhardt, *et al.*, 2017) as the interactions between nucleotides in close proximity are more energetically favourable. Consequently, it is likely that as the vRNA emerges from the polymerase it is capable of forming secondary RNA structures if no NP monomers are immediately available. The comparison between *in virio* and *ex virio* SHAPE conditions is therefore useful in highlighting the local secondary RNA structures, which should be preserved regardless of the presence or absence of NP.

3.4.6 Summary

- SHAPE reagents can be successfully used to modify vRNA inside viral particles.
- SHAPE-MaP method was used to acquire SHAPE reactivity profiles for all genome segments *in virio* and *ex virio*.

4. Chapter 4: Analysis of the intramolecular secondary RNA structures present in the influenza virus genome

4.1 Overview

Influenza viruses have single stranded negative sense RNA genomes consisting of eight vRNA segments, each of which codes for one or more viral proteins. Every individual vRNA is maintained in an RNP complex, however, little is known about the secondary RNA structures present in the context of vRNPs. This chapter investigates the secondary RNA structures present in the genomes of WSN (A/WSN/33), PR8 (A/PR/8/34(H1N1)), Udorn (A/Udorn/307/72) and FLB (B/Florida/4/06) viruses using the SHAPE-MaP approach. The data presented here show that influenza viruses contain multiple secondary RNA structures in their genomes, some of which play important roles in the virus life cycle.

4.2 Background

A vRNP complex in influenza viruses consists of three elements - the heterotrimeric viral polymerase (3P), the viral nucleoprotein (NP), and the vRNA. The 3P binds the 5′ and 3′ termini of the vRNA, and the rest of vRNA is wrapped around a double helix-forming oligomer of NP (Eisfeld, Neumann & Kawaoka, 2014). vRNA is synthesised in the nucleus of a cell from a cRNA template, which acts as replicative intermediate. It is thought that the NP loading onto the vRNA occurs as soon as the newly made 5′ end of the vRNA gets bound by the *trans*-activating 3P, which also recruits the first NP molecule. The formation of the rest of the vRNP requires oligomerisation of NP, which occurs when the tail-loop of an incoming NP monomer is inserted into the insertion-groove of the resident NP (Velthuis & Fodor, 2016). The polymerisation of NP ultimately forms a double-helix structure, with the 3P positioned at one end of it and an NP loop structure formed at the opposite end. While structural studies have given us an overall understanding of the protein arrangements in the RNPs, the highest available 20-40 Å resolution structures do not provide enough detail to see the vRNA (Arranz, Coloma, Chichón, *et al.*, 2012; Moeller, Kirchdoerfer, Potter, *et al.*, 2012).

Perhaps the best understood structure in the vRNA is that of its promoter sequence, in which the first 10 nucleotides of the 5′ termini form a hook-like structure, which is followed by double stranded region formed by the base-pairing between positions 11-16 and 10-15 of the 5′ and 3′ termini, respectively. Nucleotides 1-9 of the 3′ vRNA end appear to be single stranded (Pflug, Lukarska, Resa-Infante, *et al.*, 2017). While the binding of NP to the vRNA likely influences the folding of the vRNA, conserved secondary RNA structures present in M and NP segments were predicted *in silico*. Further analysis using synonymous mutagenesis to disrupt the predicted structures has shown that they are important for vRNA packaging (Gultyaev, Tsyganov-Bodounov, Spronken, *et al.*, 2014), replication and splicing (Spronken, van de Sandt, de Jongh, *et al.*, 2017; Kobayashi, Dadonaite, van Doremalen, *et al.*, 2016). Experiments, in which different vRNA segments were co-incubated and allowed to interact *ex virio*, have shown that multiple RNA-RNA interactions can occur between the genomic segments (Gavazzi, Isel, Fournier, *et al.*, 2013; Fournier, Moules, Essere, *et al.*, 2012a; 2012b). One such interaction was later shown to be formed through a kissing-loop-like contact between hairpins extending from the PB1 and NS segments. Mutations introduced to disrupt the potential base-pairing region between the two hairpins led to attenuated virus growth and defects in genome packaging. Significantly, when both segments were mutated in a way that allowed to restore the kissing-loop structure both virus growth and genome packaging were recovered (Gavazzi, Yver, Isel, *et al.*, 2013). Interestingly, non-native structural motifs can be successfully inserted into the viral genome as well. A bacteriophage PP7 hairpin has been inserted into vRNPs and cRNPs (York, Hengrung, Vreede, *et al.*, 2013). PP7-containing RNPs can be successfully recovered from infected cells using a PP7-binding PP7CP protein, indicating not only that the hairpin can form and is preserved during the RNP replication but also that the PP7 structure protrudes from the RNPs and is accessible to protein binding.

While, compared to many other viruses, still relatively little is known about the RNA structure of the influenza virus genome, multiple potential functions for secondary RNA structures in influenza virus have been proposed. Secondary RNA structures have been shown to form important motifs that regulate the splicing of M1 and NS1 mRNAs (Moss, Dela-Moss, Kierzek, *et al.*, 2012; Lenartowicz, Kesy, Ruszkowska, *et al.*, 2016; Moss, Priore & Turner, 2011; Dubois, Terrier & Rosa-Calatrava, 2014), and potentially translation initiation at alternative open reading frames in the PB1 mRNA (Priore, Kauffmann, Baman, *et al.*, 2015). In addition, a conserved double stranded motif in the 5′ region of positive-sense viral RNA was shown to be selectively recognised by the

viral NS1 protein (Marc, Barbachou & Soubieux, 2013), the exact consequence of which is not yet clear. The secondary RNA structures present in the negative sense vRNA have been of particular interest due to their potential involvement in the selective packaging of the viral genome (Gerber, Isel, Moules, *et al.*, 2014). Influenza viruses need to package all eight different vRNPs into the budding virions to be able to initiate a successful infection. The packaging of the different vRNPs has been overwhelmingly shown to be a selective process, i.e. only one of each of the eight segments get packaged into a virion (Giese, Bolte & Schwemmle, 2016; Chou, Vafabakhsh, Doğanay, *et al.*, 2012; Hutchinson, Kirchbach, Gog, *et al.*, 2010). As the only known difference between the eight different vRNPs is the sequence of vRNA, it has been proposed that selective interactions between the vRNAs determine the selectivity of the packaging process. Indeed, multiple studies have shown that synonymous mutations in the virus genome can affect the packaging of different segments (Hutchinson, Wise, Kudryavtseva, *et al.*, 2009; Hutchinson, Curran, Read, *et al.*, 2008; Marsh, Rabadán, Levine, *et al.*, 2008). Most of these mutations have been mapped to the terminal regions of each segment (ranging from 12 to 222 terminal nucleotides for different segments (Eisfeld, Neumann & Kawaoka, 2014)). Furthermore, these terminal packaging signals are sufficient to promote packaging of non-viral RNAs and defective viral genomes into the budding virions (Hutchinson, Kirchbach, Gog, *et al.*, 2010; Odagiri & Tashiro, 1997). On the other hand, an increasing number of studies have now shown that central regions of vRNA segments can affect packaging as well, and these seem to be especially important for the co-segregation of segments from different influenza strains (Cobbin, Ong, Verity, *et al.*, 2014; Cobbin, Verity, Gilbertson, *et al.*, 2013; Fujii, Fujii, Noda, *et al.*, 2005; Gultyaev, Tsyganov-Bodounov, Spronken, *et al.*, 2014), highlighting the fact that packaging may also play an important role in the reassortment between different influenza strains.

A better structural insight into the influenza virus RNA conformation could serve as an important step in advancing our understanding of multiple stages in the influenza virus life cycle - from replication to packaging and reassortment. This chapter describes the use of the SHAPE-MaP method to analyse the global RNA structures of the genomes of several different influenza strains. The analysis shows that while the binding of NP decreases the formation of secondary RNA structures in the vRNA, still multiple secondary RNA structures in the virus genome are maintained. The SHAPE-MaP analysis recapitulates several previously described secondary RNA

structures, as well as uncovers a number of new structures, some of which are shown to be important for the virus life cycle and are also found in different virus strains.

4.3 Aims

- Determine the conformation of vRNA in the context of RNPs
- Determine the effect of NP binding on the structure of the vRNA
- Identify secondary RNA structures in the viral genomes based on SHAPE reactivities
- Determine if secondary RNA structures are important for the viral life cycle
- Compare vRNA structures between different influenza virus strains

4.4 Results

4.4.1 Different WSN genome segments have unique RNA conformations

The *in virio* SHAPE-MaP reactivity profiles were acquired as described in chapter 3. Independent experimental replicates of *in virio* SHAPE probing produced highly similar SHAPE reactivity profiles, indicating that the average genome RNA structure tends to be preserved. Different genome segments show very different SHAPE reactivity profiles, suggesting that each segment has a unique RNA fold (Figure 4.1). In addition, within all segments, regions of extensive low SHAPE reactivity can be found, implying that highly structured RNA sites within the WSN genome exist (Figure 4.2). Interestingly, except for the NS segment, the first 200 nucleotides of the 5′ end in all segments have more nucleotides associated with low SHAPE reactivities, as compared to the next 200 nucleotides or the last 200 nucleotides at the 3′ end, indicating that the segment 5′ ends are more structurally constrained than the rest of the segment (Figure 4.3).



Figure 4.1 Median SHAPE reactivity profiles for all WSN genome segments

Median SHAPE reactivities were calculated over sliding 20 nucleotide windows and plotted in relation to the global median for a given genome segment. Reactivity profiles for the three replicates of in virio experiments using 1M7 reagent are shown.

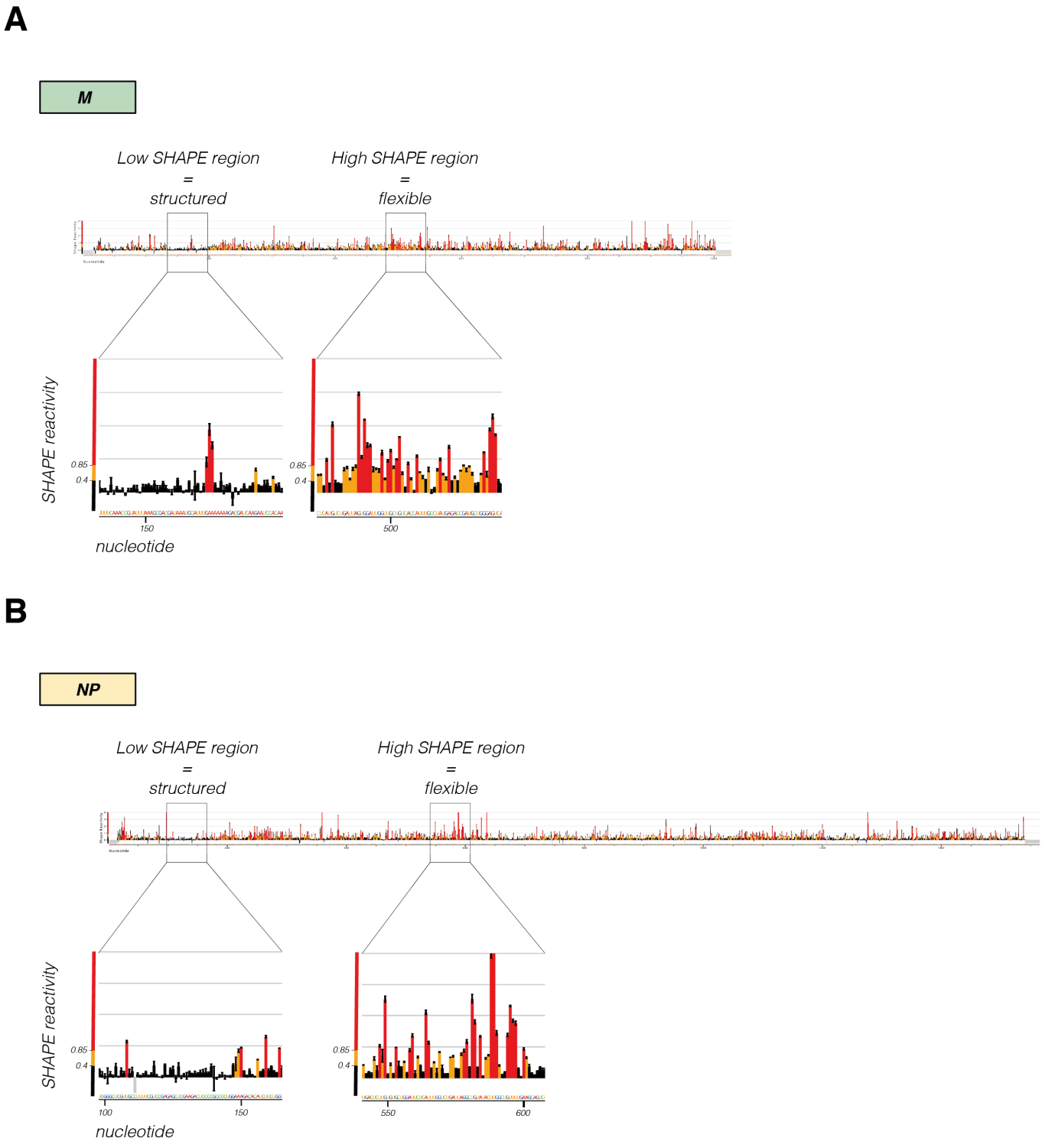


Figure 4.2 Variation in SHAPE reactivities within segments

*Examples of low and high SHAPE reactivity regions in M (**A**) and NP (**B**) segments are shown.*

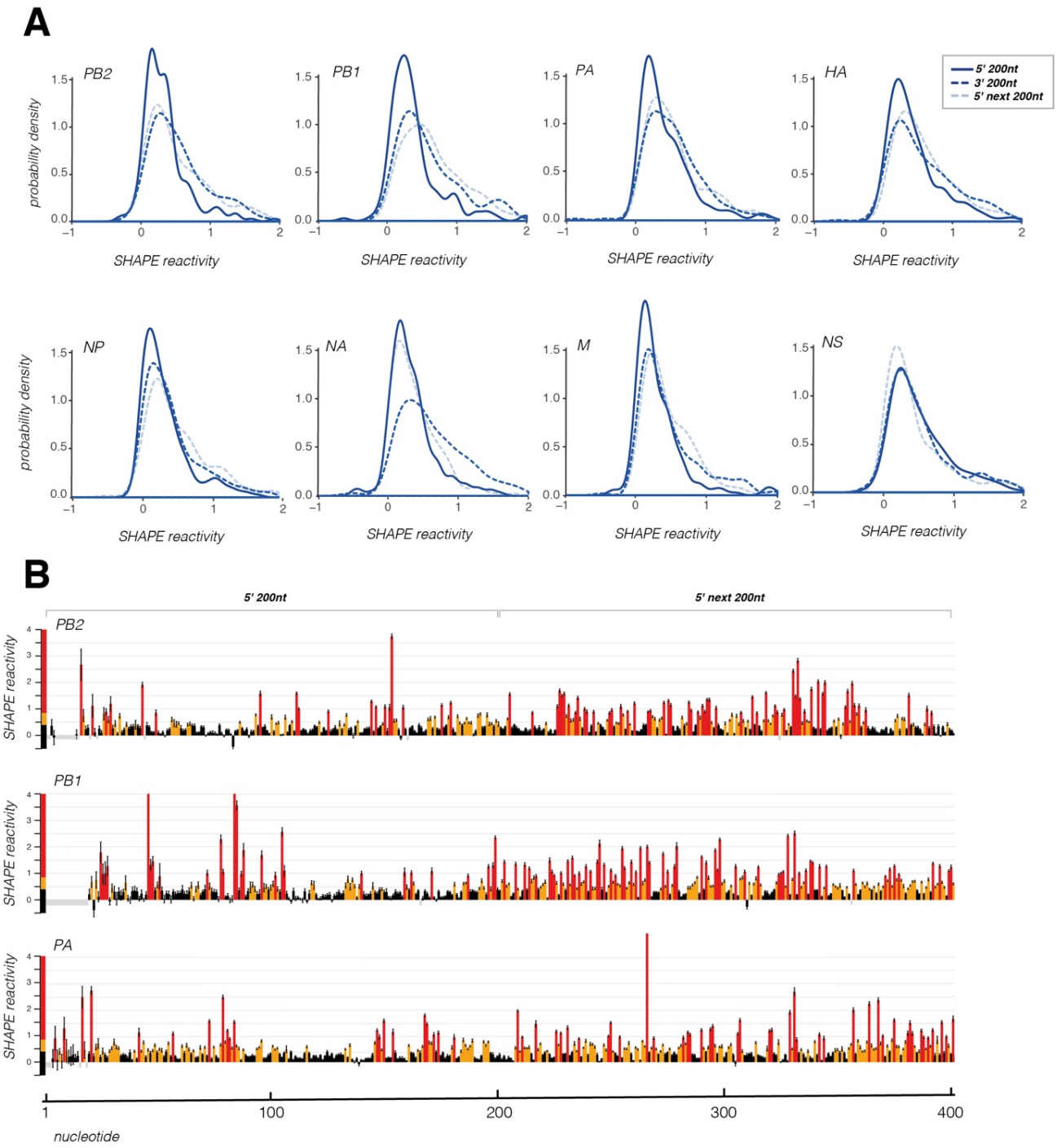


Figure 4.3 SHAPE reactivity distribution in the 5′ end of the vRNA

*(**A**) Probability density distributions of the SHAPE reactivities in the 200 nucleotides of the 5′ end, the next 200 nucleotides and the last 200 nucleotides in the 3′ end of all viral genome segments. (**B**) Examples of raw SHAPE reactivity profiles of the WSN polymerase segments, illustrating the distributions of the first 200 nucleotides and the next 200 nucleotides in the 5′ end.*

4.4.2 Comparison between *in virio* and *ex virio* vRNA SHAPE reactivities

The *ex virio* SHAPE profiles (for *nkvRNA* and *ivtRNA*) were acquired as described in chapter 3. A high correlation of SHAPE reactivities was found between the two different *ex virio* conditions; the average correlation coefficient between *ivtRNA* and *nkvRNA* SHAPE reactivities was between 0.91 and 0.95. By contrast, the correlation between *in virio* and *ex virio* SHAPE reactivities was in the range of 0.52-0.62 (Figure 4.4). While overall the correlation between the RNP-associated and *ex virio* RNA SHAPE reactivities was lower, there were nevertheless regions spanning generally less than 150 bases, which had high (>0.7) correlation coefficients, indicating that structural commonalities between naked and NP-bound vRNAs exist.

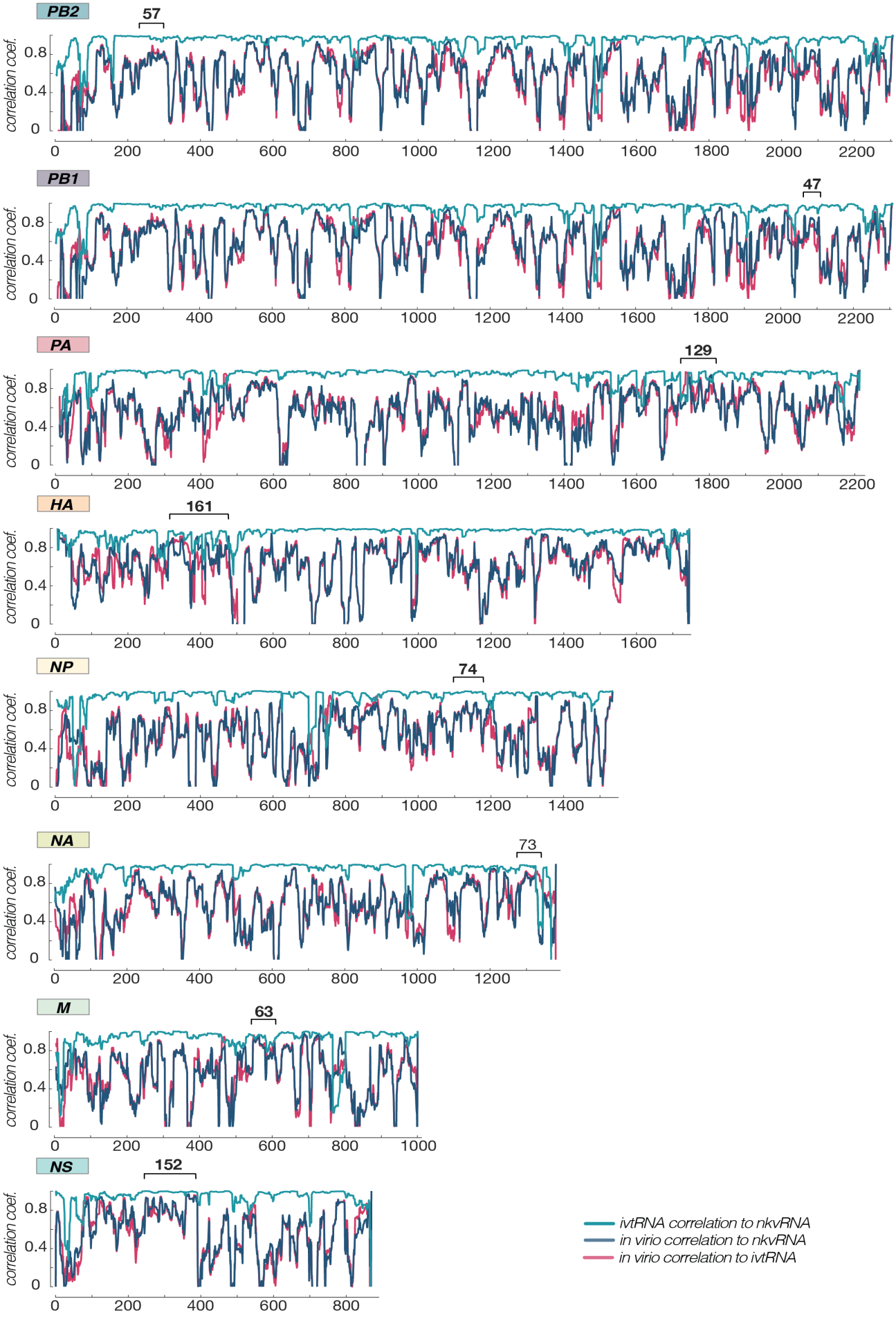

PB2
57
PB1
47
PA
129
HA
161
NP
74
NA
73
M
63
NS
152
correlation coef.
0
0.4
0.8
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
ivtRNA correlation to nkvRNA
in virio correlation to nkvRNA
in virio correlation to ivtRNA

Figure 4.4 Sliding window correlation between in virio and ex virio SHAPE reactivities of vRNA

The Pearson correlation coefficient was calculated over 20 nucleotide windows. Several regions of >0.7 correlation between in virio and ex virio samples are marked with numbers above the correlation graphs, indicating the length of the regions with high correlation.

4.4.3 NP periodicity analysis

The binding of NP has an unknown effect on SHAPE reactivities, however, it is possible that its binding shields the 2'-OH from the SHAPE reagent, leading to low SHAPE reactivities in NP-bound regions. If the NP is binding to the vRNA at regular intervals these NP-caused low SHAPE reactivities could be detected as a pattern in the SHAPE signal. Therefore, a Fourier transform analysis was used to identify periodicities in the SHAPE reactivity signal. While a periodicity of around 18-20 nucleotides could be identified in the *in virio* SHAPE profiles, a similar frequency was also found in periodicity analysis for *ivtRNA* data, suggesting that this pattern is not caused by NP binding.

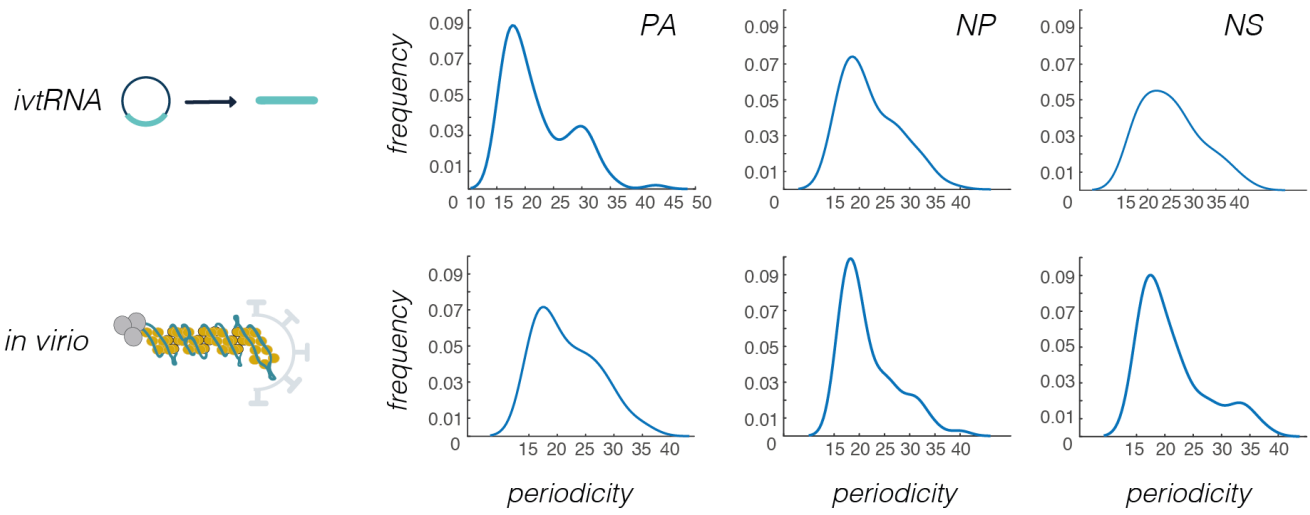


Figure 4.5 NP Periodicity analysis

ivtRNA and in virio samples were used to analyse periodicity signals present in the SHAPE reactivity data. Representative frequency plots for three genome segments are shown.

4.4.4 Secondary RNA structure modelling for the WSN genome

SHAPE reactivities are commonly used as pseudoenergy constrains in secondary RNA structure modelling algorithms (Spasic, Assmann, Bevilacqua, *et al.*, 2018; Deigan, Li, Mathews, *et al.*, 2009). Therefore, SHAPE-informed *ex virio* and *in virio* secondary RNA structures for all WSN genome segments were modelled next. As the comparison between *ex virio* and *in virio* SHAPE reactivities suggested that high correlation can be found within <150 nucleotide regions (see Figure 4.4), to find local secondary RNA structures, a maximum base-pairing window during the structure modelling was set to 150 bases. While a small maximum base-pairing window biases the predictions to only short-distance RNA interactions, this bias is consistent with the co-replicative loading of NP model, which would lead to predominantly local secondary structure formation. For the modelling purposes, the promoter regions were set to be single stranded. During the modelling, the SHAPE reactivities were used in computing the partition function for all structures. Rather than providing a single secondary structure fold that is based on the minimum folding energies (MFE), the partition function allows to calculate the probabilities of base-pairing for every nucleotide in the sequence based on the pool of multiple suboptimal MFE structures. The partition function calculation was converted into a probability plot, which was used to predict the maximum expected accuracy structures (MEA) for each segment. The MEA structures and the associated probability plots were visualised in arcplots, in which each arc indicates the pairs of interacting nucleotides and colours are used to indicate associated base-pairing probabilities (Figure 4.6, Figure 4.7, Figure 4.8).

As expected the SHAPE-informed structure folding showed that *nkvRNA* and *ivtRNA* segments adopt highly similar RNA structures. A number of common structures can also be identified between the *in virio* and *ex virio* conditions (Figure 4.9 A), though overall, the probability of base-pairing between nucleotides decreases significantly *in virio* (Figure 4.9 B). The decreased base-pairing probabilities *in virio* are likely an indication of a destabilising effect imposed on the vRNA structure by the presence NP. However, while clearly fewer structures are capable of forming *in virio*, local secondary RNA structures can nevertheless be clearly detected in the vRNPs (Figure 4.10). Furthermore, in addition to predicting many previously uncharacterised secondary RNA structures, SHAPE profiling can also recapitulate RNA structures that have already been described and shown to have functional effects on virus replication (Figure 4.11). These include a hairpin in

the 5′ end of the M segment, which was shown to be important for vRNA replication and M2 mRNA splicing, and a pseudoknot structure in the NP segment.

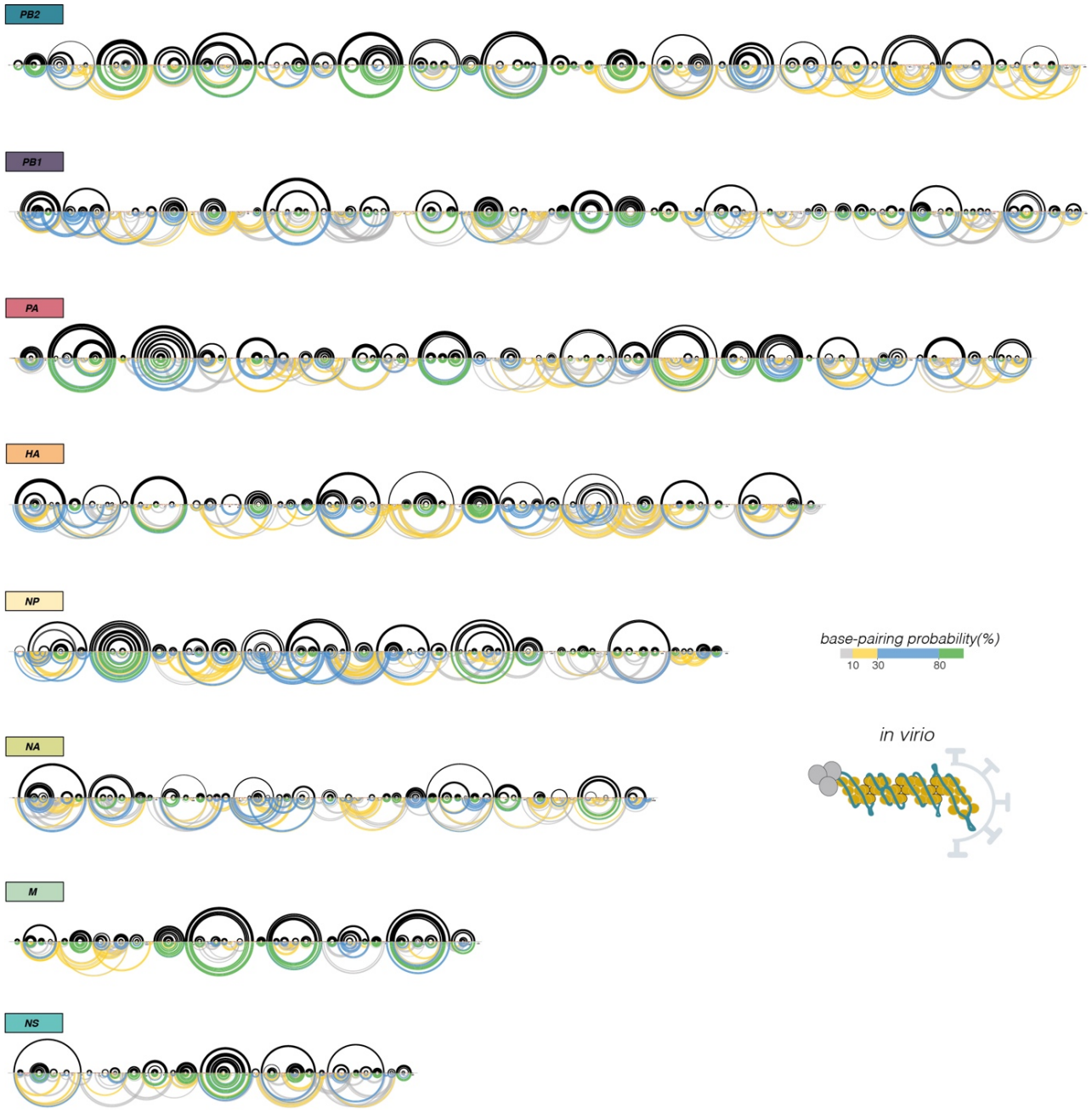


Figure 4.6 In virio secondary RNA structures in the WSN genome

The horizontal baseline depicts the sequence of each vRNA, the arcs connect the base-paired nucleotides. The black arcs above the horizontal line depict the minimum predicted accuracy structure for a segment and the coloured arcs below the horizontal line indicate the base-pairing probabilities associated with multiple suboptimal minimum free energy structures.

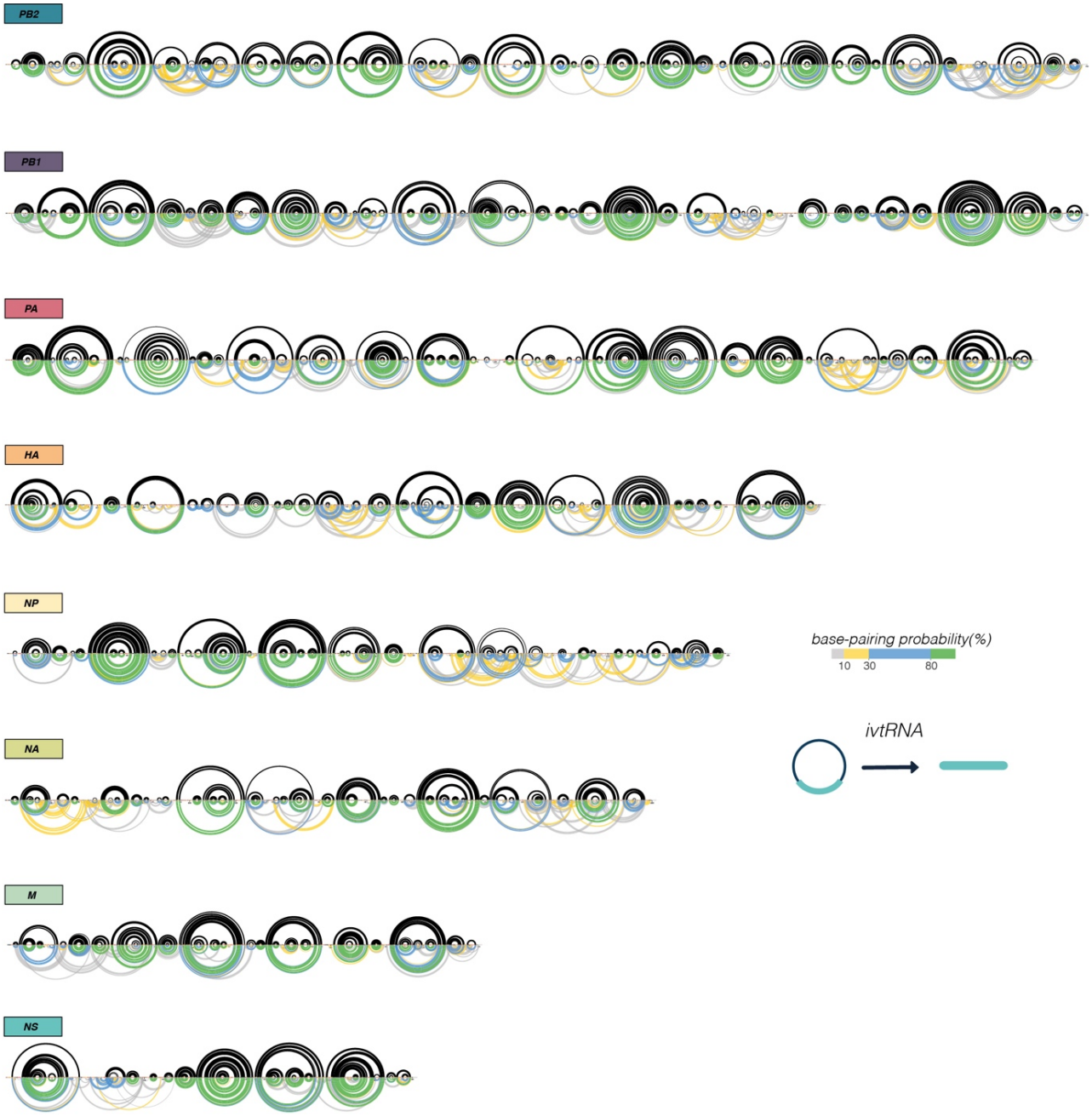


Figure 4.7 ivtRNA secondary RNA structures

The horizontal baseline depicts the sequence of each vRNA, the arcs connect the base-paired nucleotides. The black arcs above the horizontal line depict the minimum predicted accuracy structure for a segment and the coloured arcs below the horizontal line indicate the base-pairing probabilities associated with multiple suboptimal minimum free energy structures.

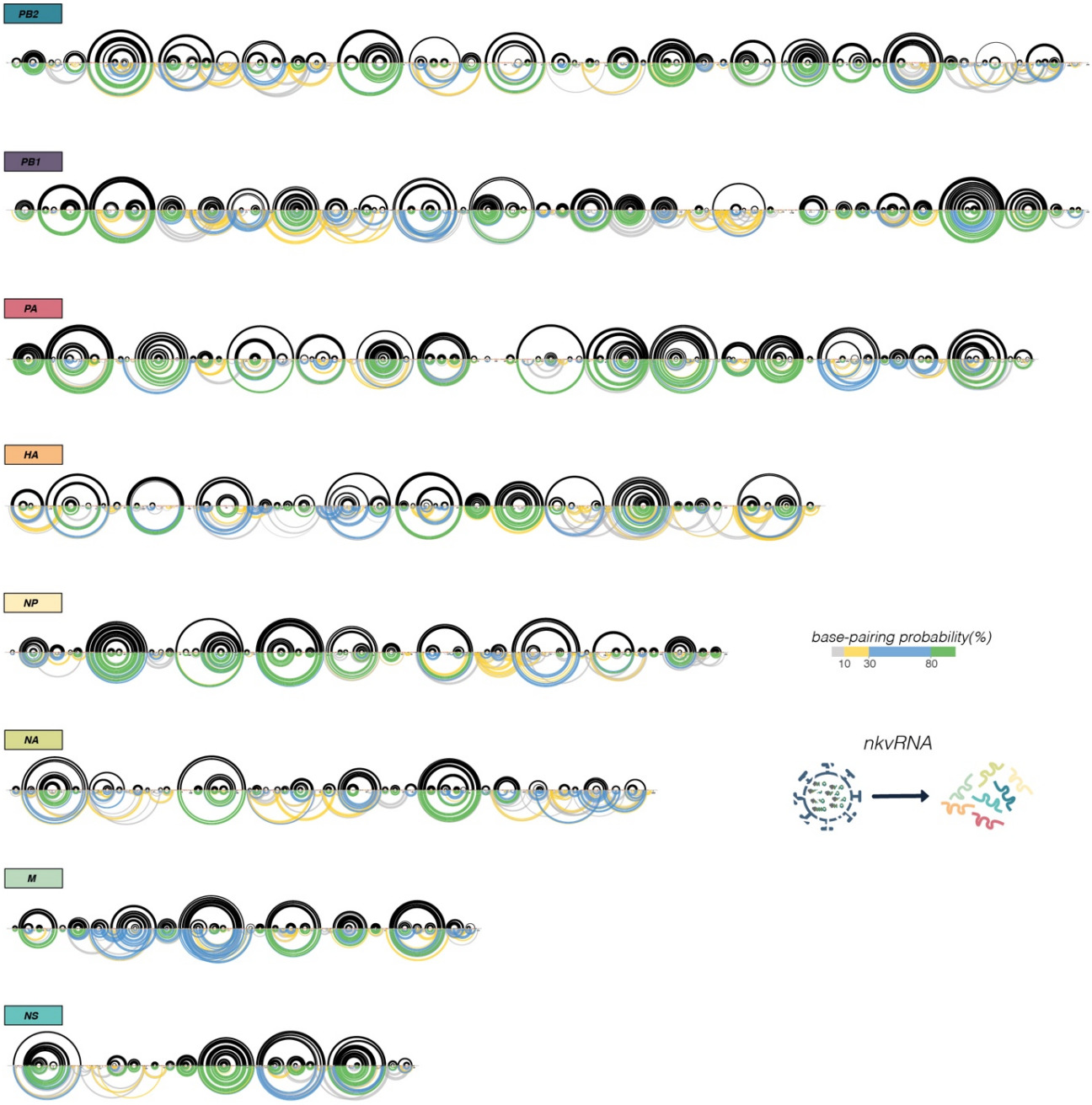


Figure 4.8 nkvRNA secondary RNA structures

The horizontal baseline depicts the sequence of each vRNA, the arcs connect the base-paired nucleotides. The black arcs above the horizontal line depict the minimum predicted accuracy structure for a segment and the coloured arcs below the horizontal line indicate the base-pairing probabilities associated with multiple suboptimal minimum free energy structures.

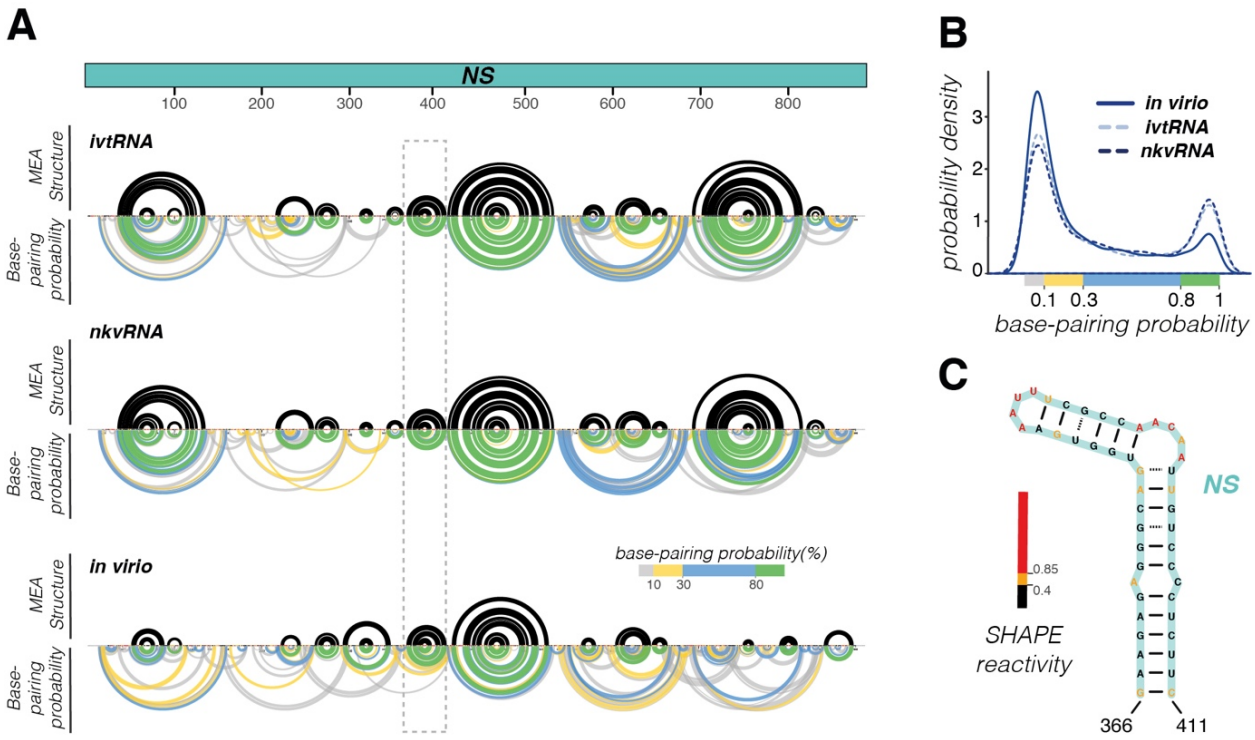


Figure 4.9 Comparison between in virio and ex virio vRNA structures

*(**A**) Arcplots depicting the RNA structure of NS segment under different experimental conditions. Only the arcs associated with high probability minimum accuracy structures are shown. A common high-probability hairpin structure found under all conditions is highlighted with a dashed rectangle and drawn in (**C**). (**B**) A summary probability density distribution of nucleotide base-pairing probabilities in different conditions shows more nucleotides associated with high structure formation probabilities ex virio and more nucleotides associates with low structure formation probabilities in virio.*

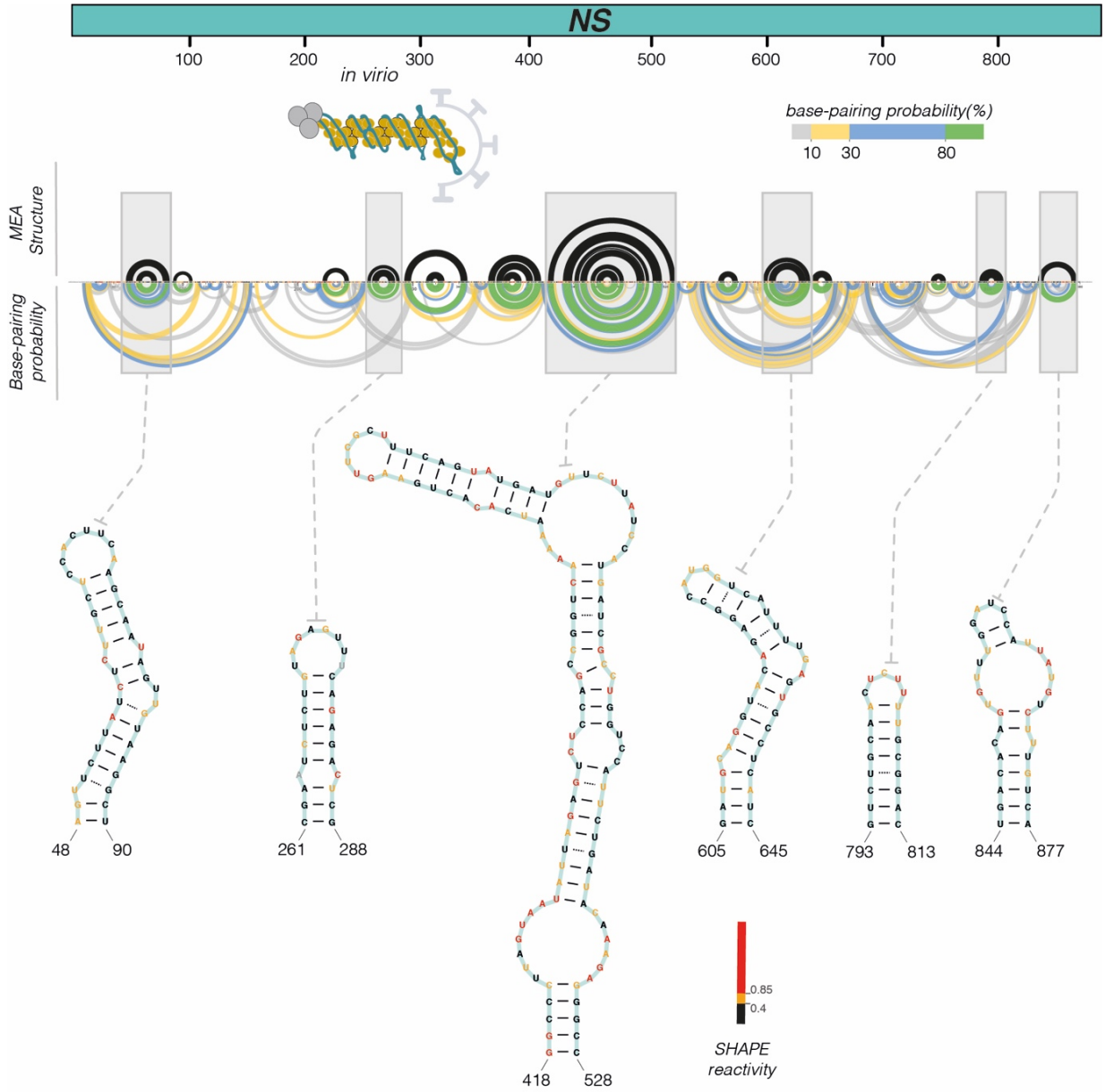


Figure 4.10 Secondary RNA structures found in the NS segment in virio

High base-pairing probability RNA structures are highlighted on the arcplot and the associated SHAPE-informed structure predictions are shown. The nucleotides in the modelled RNA structures are coloured by their SHAPE reactivities.

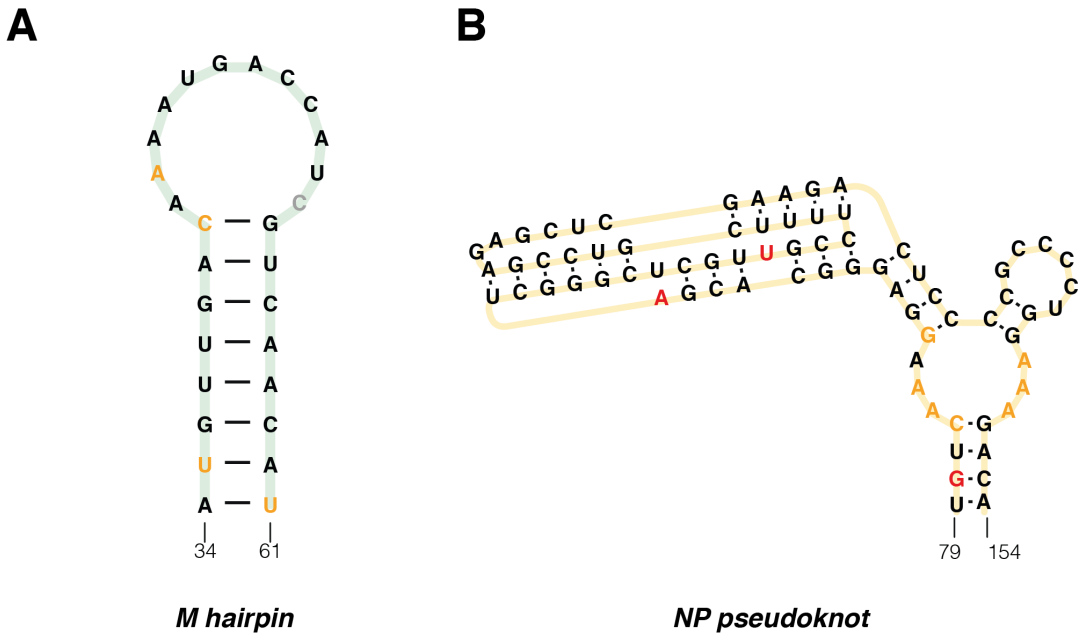


Figure 4.11 SHAPE profiling recapitulates previously described secondary RNA structures

*(**A**) A hairpin in the M segment previously shown to be important for genome packaging in vRNA and splicing in mRNA sense by Kobayashi et al (2016) and Spronken et al. (2018). (**B**) A pseudoknot in the NP segment previously shown to be important for vRNA packaging by Gultyaev et al. (2014) and Williams et al. (2017).*

4.4.5 Functional roles of the secondary RNA structures in the WSN virus genome

SHAPE-MaP analysis has identified multiple secondary RNA structures in the WSN virus genome, therefore next, their role in the viral life cycle was tested. To this end, WSN viruses with secondary RNA structure-disrupting mutations were made, and their growth was assessed. Because many of the identified secondary RNA structures were within the protein coding regions, in order to exclude any effects of mutations due to the changes in the protein coding capacity, all nucleotide changes were designed to be synonymous. In addition, the changes in codons were designed in such way that the new codon was either more commonly used in mammalian cells or, if the codon in question was already optimal, the next most codon was chosen. The changes in GC content of the sequence were not considered.

Viruses with disrupted secondary RNA structures exhibited a variety of growth defects. Viruses with mutations in the HA segment showed no significant changes in their growth compared to the wild-type (Figure 4.12 A). The *PB2 31-78(5mut)* virus was attenuated at the early stages of replication, showing delayed growth kinetics; by contrast, the *PB2 1301-1385(6mut)*

mutant produced significantly more plaques at the 12 and 24-hour time points compared to the wild-type (Figure 4.12 B). Disruption of the *PB1 1309-1391(4mut)* stem-loop led to around 10,000-fold reduction in plaque forming units at 48 hours post infection, while introduction of just 3 mutations into the PB1 118-161 hairpin prevented any recovery of the virus (Figure 4.13). Similarly, no virus could be rescued when 17 mutations were introduced into the NS 364-411 hairpin, though introduction of just 5 mutations to the same hairpin did not lead to any significant defects in virus growth (Figure 4.14). Further investigation into the replication of the *NS 364-411(17mut)* vRNA using an RNP reconstitution assay, showed that the introduction of the 17 mutations into the segment led to a more than two-fold reduction in the ability of the viral polymerase to replicate the *NS 364-411(17mut)* segment (Figure 4.14 **C**). In addition, when the vRNA was extracted from the *NS 364-411(5mut)* and *PB1 1309-1391(4mut)* virus stocks and the relative quantities of each segment were measured in relation to the HA segment RNA, significant changes in the abundance of several vRNA segments were detected for *PB1 1309-1391(4mut)* but not *NS 364-411(5mut)* virus (Figure 4.15). In addition to the *NS 364-411(17mut)* and PB1 *118-161(3mut)* viruses, also no virus was recovered for *NP 1475-1525(3mut), NP 1221-1275(3mut)* and *M 110-217(4mut)* constructs (Figure 4.16).

An alternative approach to investigate the importance of secondary structures in vRNA was the use of anti-sense oligonucleotides (ASOs), targeting those secondary RNA structures that were found to be critical for the recovery of infectious virus. The ASOs were designed to include locked nucleic acid nucleotides, which should increase their affinity to the complementary sequence and allow to disrupt the native RNA structure. Preliminary experiments were then conducted in which MDBK cells were transfected with ASOs, 4 hours later the cells were infected with the wild-type WSN, and 24 hours post infection cell supernatants were collected and plaque assays were performed. The presence of secondary RNA structure-targeting ASOs in the cell culture was found to lead to around 6 to 4-fold reduction in the virus titres, compared to a scrambled ASO control (Figure 4.17).

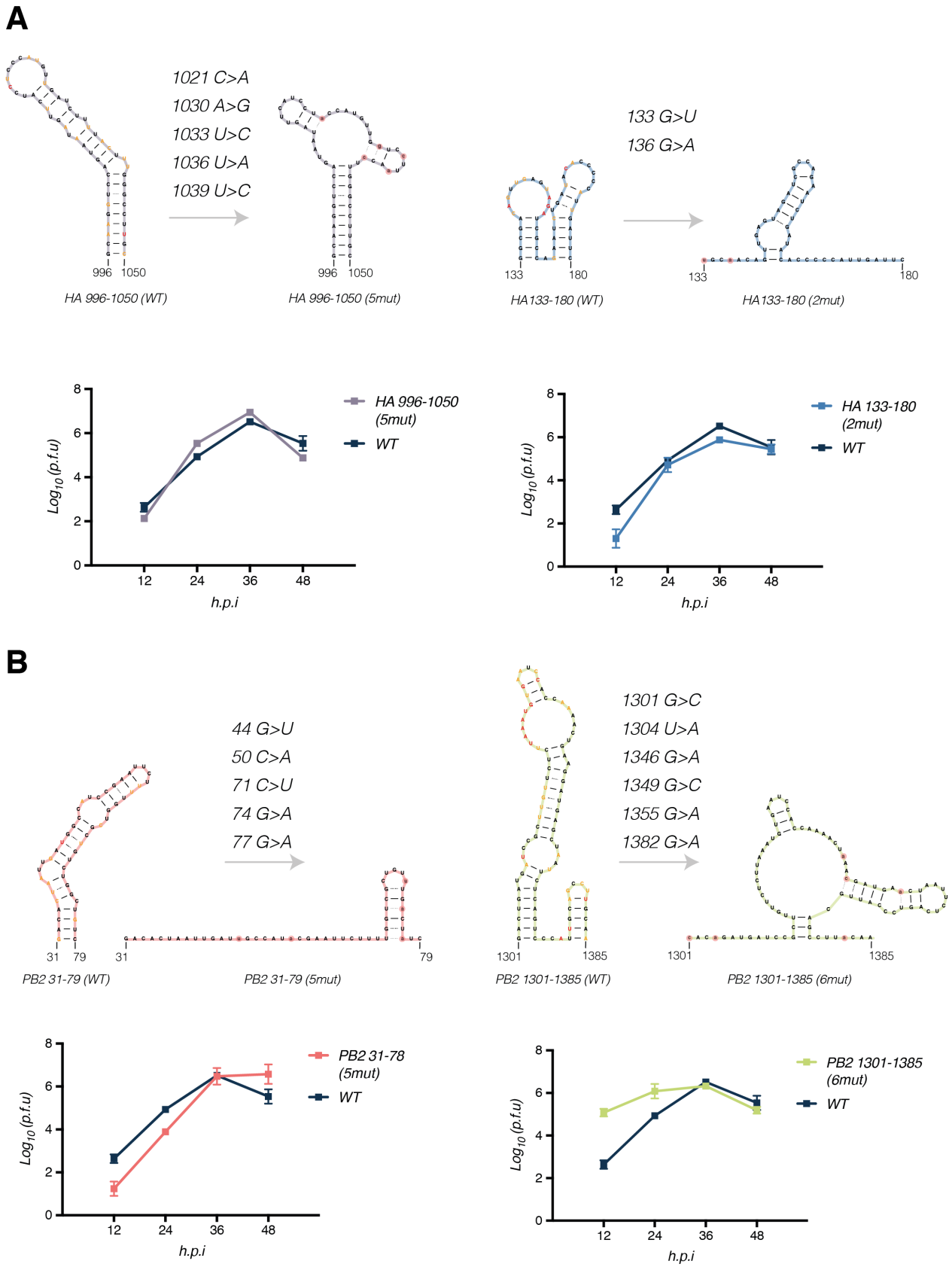


Figure 4.12 Growth analysis of the WSN viruses with disrupted secondary RNA structures in HA and PB2 segments

*Synonymous mutations were introduced into the secondary RNA structures present in HA (**A**) and PB2 (**B**) segments. Structures above the growth curves show the predicted effect of mutations on secondary RNA structure formation; nucleotides highlighted with red circles in the mutated structures indicate sites where the mutations were introduced. For each growth curve two independent cell infection replicates were performed. Infections were performed with MOI= 0.001.*

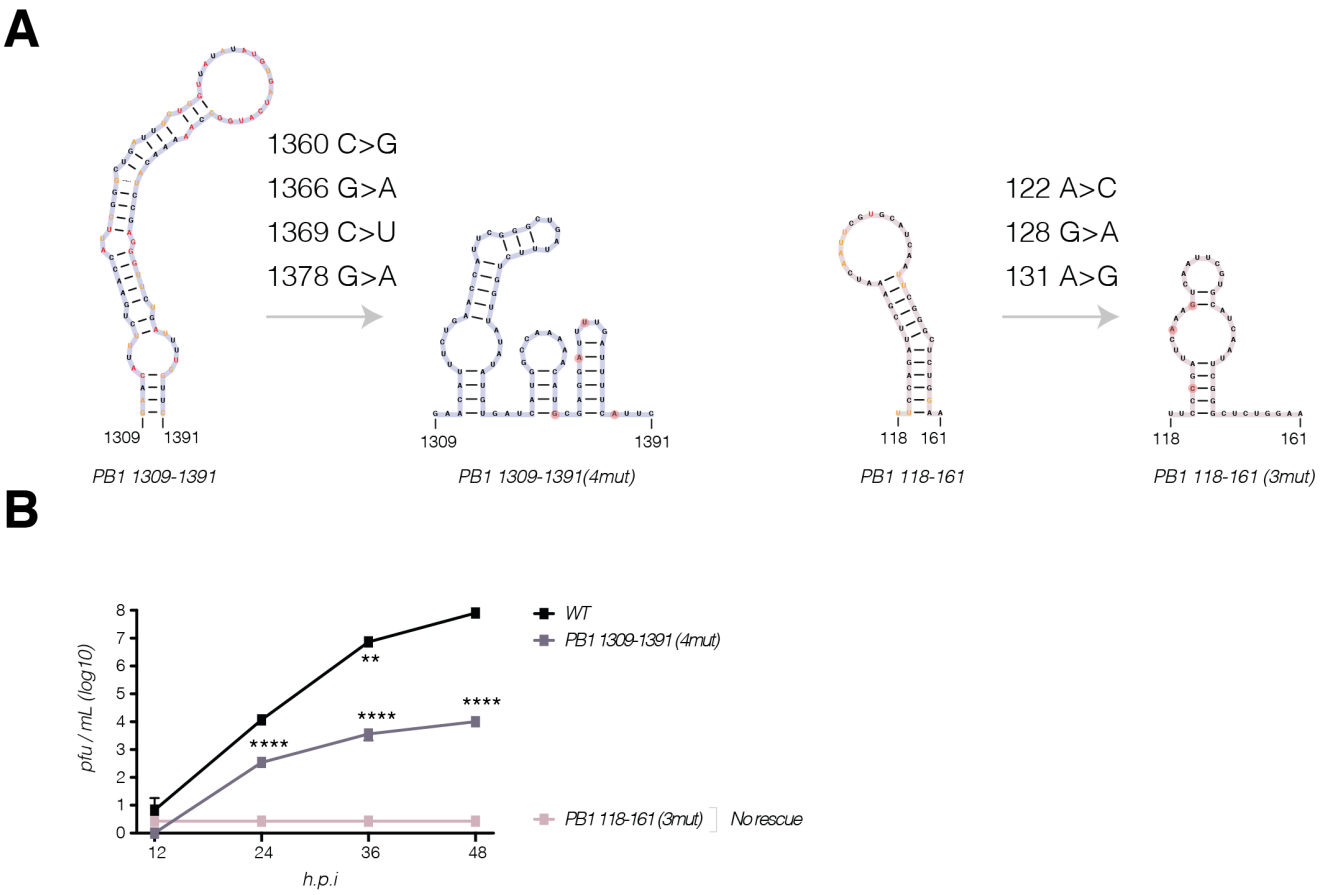


Figure 4.13 Growth analysis of the WSN viruses with disrupted secondary RNA structures in PB1 segment

*(**A**) SHAPE-directed models of secondary structures present in PB1 vRNA. Nucleotides are coloured by their SHAPE reactivities. Arrows point to the predicted structures formed upon introduction of mutations, which are highlighted in red circles. (**B**) Growth curves for viruses with synonymous mutations in secondary RNA structures modelled in **A.** Infections were performed with MOI= 0.001.*

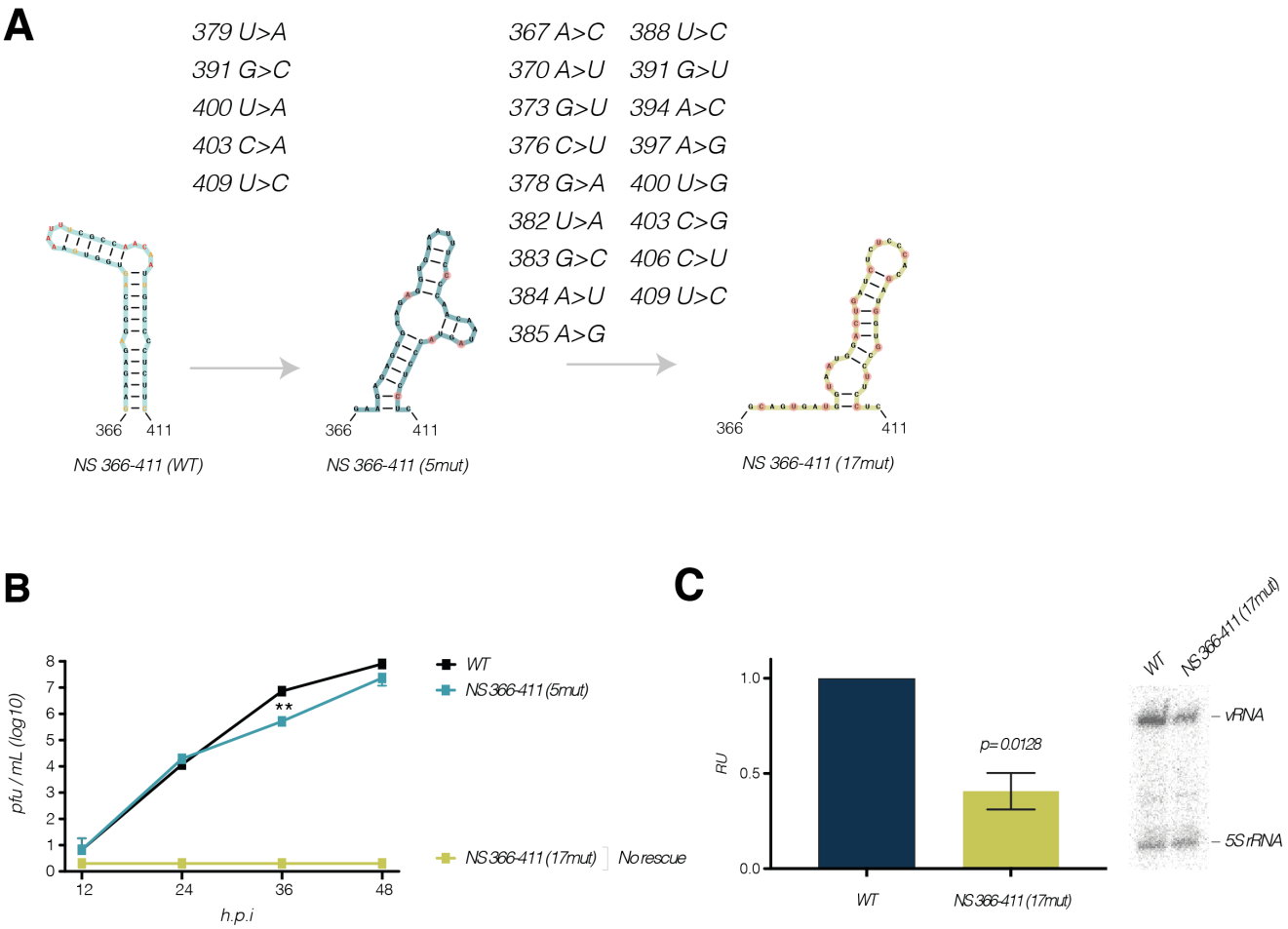


Figure 4.14 Growth analysis of the WSN viruses with disrupted hairpin in NS segment

*(**A**) Models of NS 366-411 hairpin and the structural changes introduced by its mutagenesis. Nucleotides in the wild-type hairpin are coloured by their SHAPE reactivities. Arrows point to the predicted structures formed upon introduction of mutations, which are highlighted in red circles. (**B**) Growth curves for viruses with synonymous mutations in NS 366-411 hairpin. Infections were performed with MOI= 0.001. (**C**) RNP reconstitution for NS 364-411(17mut) construct. Total RNA was extracted from cells transfected with the viral polymerase, NP and NS 364-411(17mut) or WT NS segment-encoding plasmids and primer extension was preformed using NS segment vRNA targeting primer. Quantitation for two replicates is shown on the left, a representative primer extension is shown on the right. Band intensities were normalised to the 5S rRNA signal.*

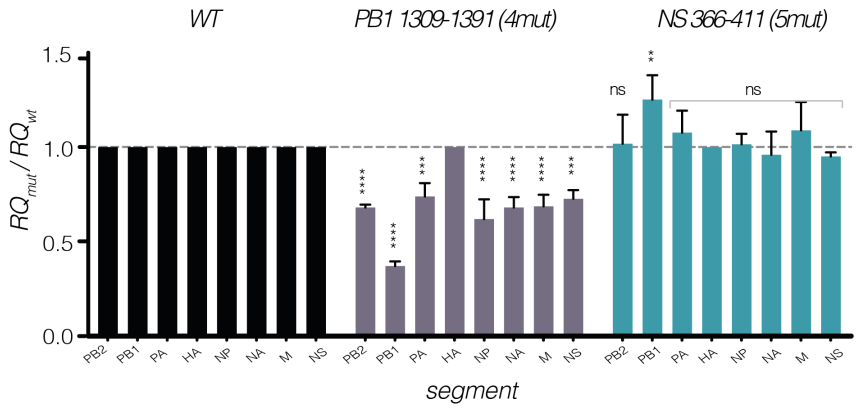


Figure 4.15 Analysis of vRNA packaging defects

RT-qPCR analysis of the vRNA content for the PB1 1309-1391(4mut) and NS 364-411(5mut) viruses. For each virus RQ values were calculated relative to the HA segment and the changes compared to the wild-type are shown. Error bars indicate standard deviation; data for three technical replicates is shown.

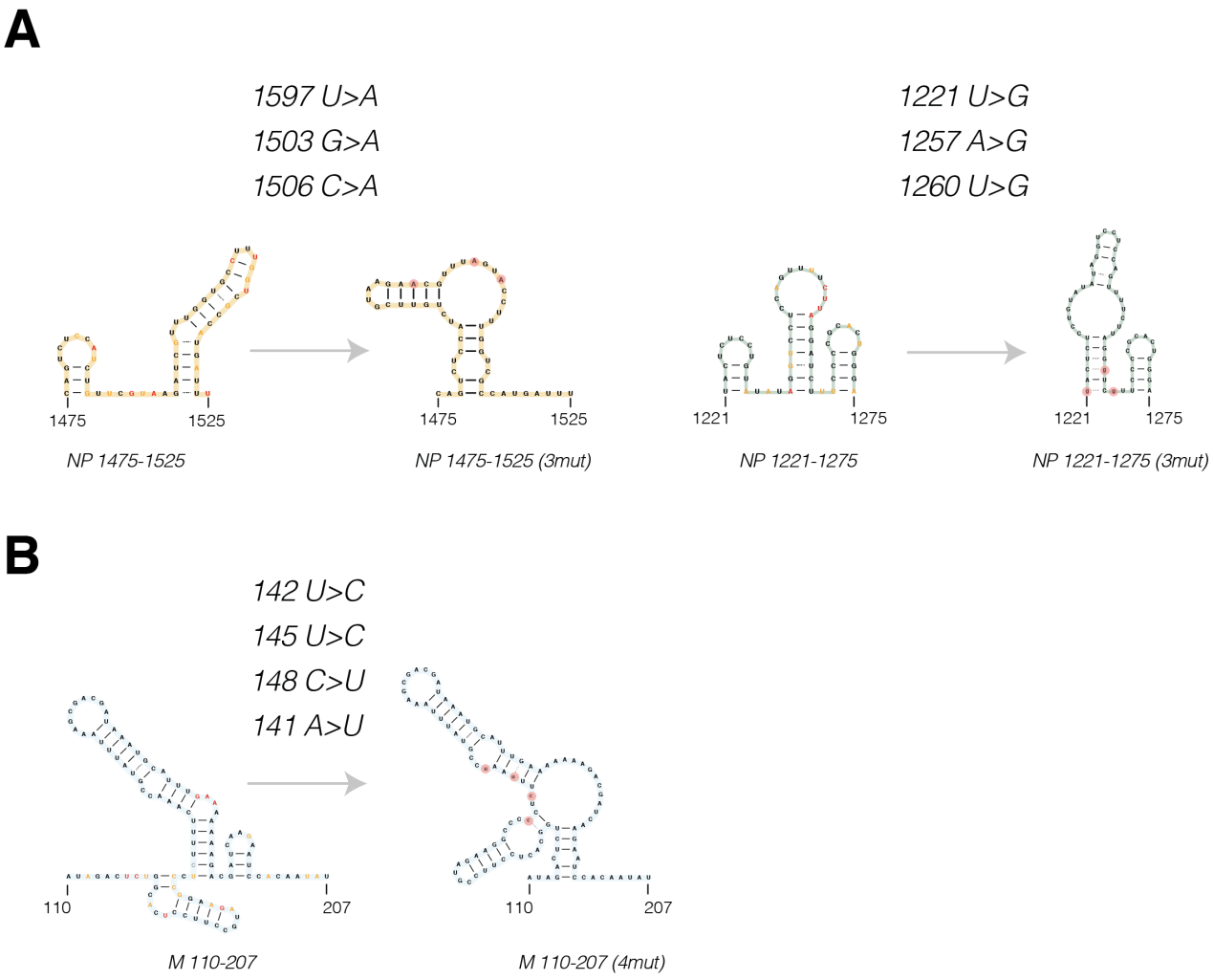


Figure 4.16 Modelled secondary RNA structures in NP and M segments

*SHAPE-directed models of secondary structures in NP (**A**), and M (**B**) segments which were found to be critical for infectious virus recovery. Nucleotides are coloured by their SHAPE reactivities. Arrows point to the predicted structures formed upon the introduction of mutations, which are highlighted in red circles.*

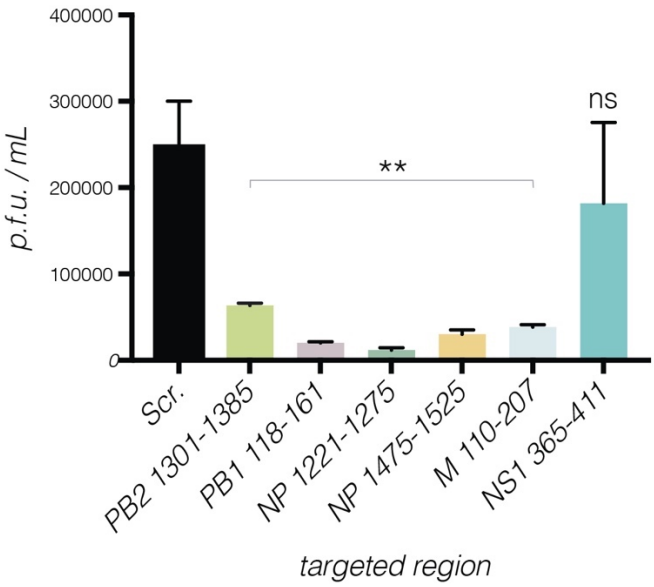


Figure 4.17 ASO targeting of secondary RNA structures in the WSN genome

*Cells were transfected with ASOs and infected with WSN virus (MOI=0.01). 24 hours post infection cell supernatants were harvested and the virus titres were measured. Scr., scrambled ASO designed by shuffling NP 1221-1275 ASO sequence. Error bars show standard deviation from three independent replicates. ** P< 0.01, Student's t-test.*

4.4.6 SHAPE-MaP analysis of PR8, Udorn and FLB virus genomes

Having analyzed the structure of the WSN genome using SHAPE, next the same analysis was performed for PR8, Udorn and FLB viruses. For each virus a single *in virio* SHAPE-MaP experiment was performed. As the background mutation rates (caused by the mutagenic RT conditions) for the different virus experiments were not equal, the final SHAPE reactivities were normalized to the distributions containing the same medians (Figure 4.18). In particular, for the PR8 virus fewer than expected high SHAPE reactivity nucleotides were detected during the SHAPE analysis. Generally, it is required for at least 8% of the modified nucleotides to have mutation rates above 0.006, however, for the PR8 sample, between 3.6% and 5.8% of nucleotides in the different segments satisfied this criterion. After the SHAPE reactivity calculation, the lower than expected modification rates in the PR8 sample led to a number of nucleotides having negative SHAPE reactivities. The normalization of the SHAPE reactivities across the different samples to the same median removed most of the negative SHAPE values from the PR8 data, though the PR8 reactivity distribution remained skewed. Multiple factors could explain the lower SHAPE modification rates in the PR8 data, including poor mixing of the reagent with the sample or partially hydrolyzed reagent. As the Udorn, FLB and PR8 experiments were performed a few days apart, it is unlikely that the condition of the SHAPE reagent itself was compromised. A repeat of the SHAPE modification step for the PR8 virus, however, did not improve the overall modification rates, suggesting that perhaps the PR8 particles themselves are less permissive to the SHAPE reagent penetration, leading to reduced 1M7-induced mutation rates in the virus.

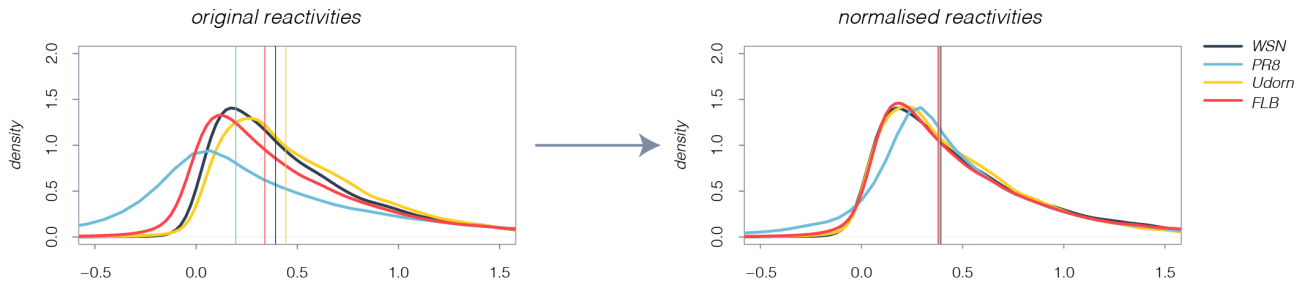


Figure 4.18 SHAPE reactivity normalisation

Reactivities across the 1M7 modified samples were normalised to the same median (indicated by the vertical lines in the plots) by adding and dividing each dataset by a constant value.

Comparison between the SHAPE reactivities for the WSN and PR8 viruses indicated that the median SHAPE reactivity correlation for all segments was generally greater than 0.5, indicating similar SHAPE profiles (Table 4.1). Using the SHAPE reactivities, multiple common secondary RNA structures were modelled for the two viruses (Figure 4.19 - Figure 4.26). A comparison between the H3N2 (Udorn) and H1N1 (WSN and PR8) strains showed that the most evolutionarily divergent segments between the H3N2 and H1N1 strains had the lowest SHAPE reactivity correlations and few common secondary RNA structures (Table 4.1 and Figure 4.19-Figure 4.26). Namely, the SHAPE reactivity correlations between H1N1 and H3N2 NA or HA segments were negative and for the PB1 segment the correlation fell to 0.36, as compared to the rest of the segments, which, like for the WSN and PR8 comparison, had median correlation above 0.5 (Table 4.1).

Table 4.1 SHAPE reactivity correlation and sequence similarity between H1N1 and H3N2 strains

*(**A**) Median SHAPE reactivity correlation coefficients for different viral genome segments. (**B**) Nucleotide sequence similarities between H1N1 and H3N2 strains. The segments showing the greatest differences are indicated in bold.*

A

Median Pearson correlation		
Segment	***WSN-PR8***	***WSN-Udorn***
PB2	0.66	0.53
PB1	***0.64***	***0.36***
PA	0.59	0.55
HA	***0.58***	***-0.02***
NP	0.56	0.51
NA	***0.47***	***-0.08***
M	0.68	0.69
NS	0.54	0.53

B

Sequence similarity		
Segment	***WSN-PR8***	***WSN-Udorn***
PB2	96.20	90.65
PB1	***97.31***	***82.83***
PA	96.87	92.52
HA	***93.35***	***40.76***
NP	95.72	90.93
NA	***92.97***	***32.58***
M	96.98	92.52
NS	96.18	91.91

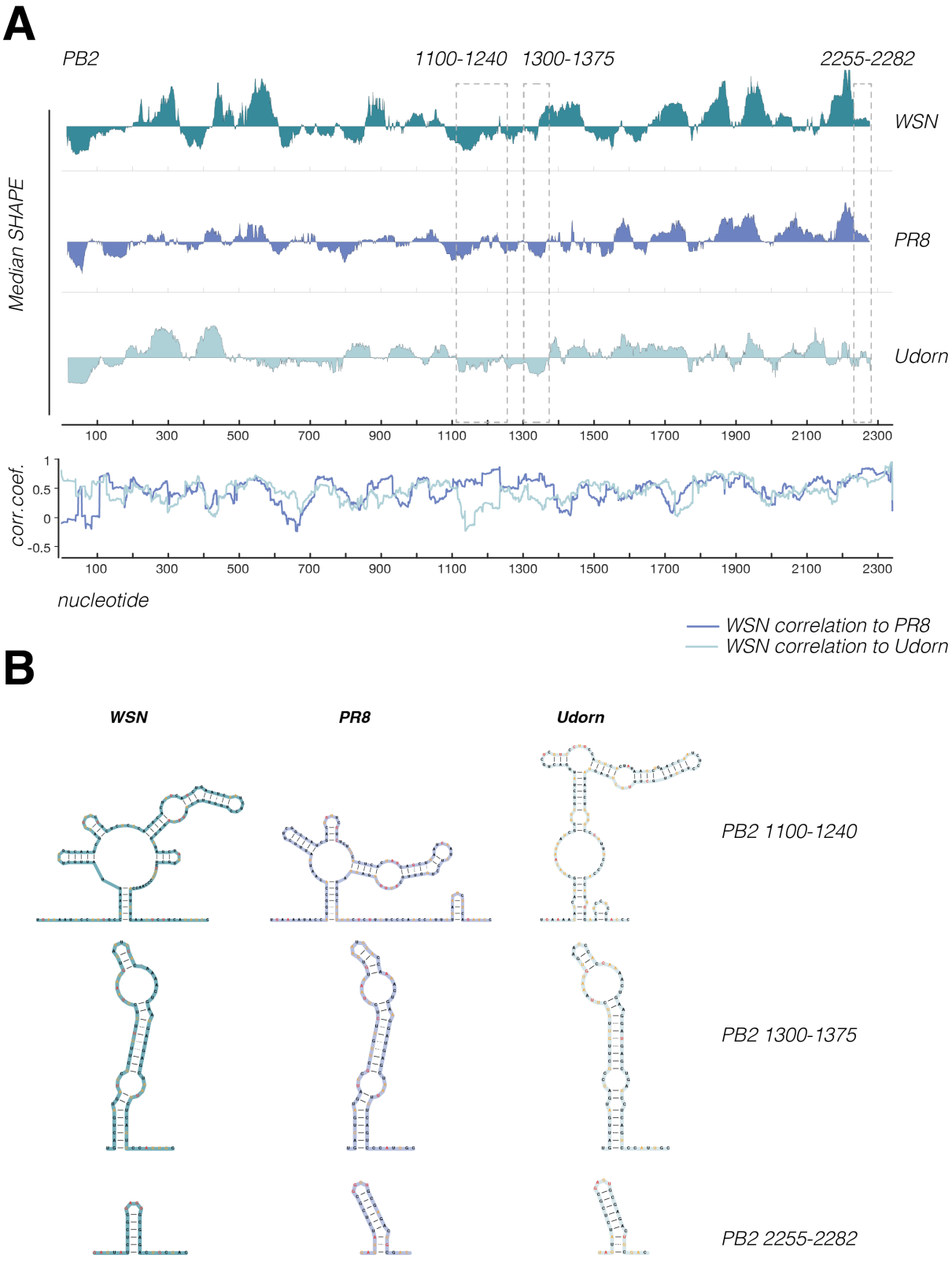
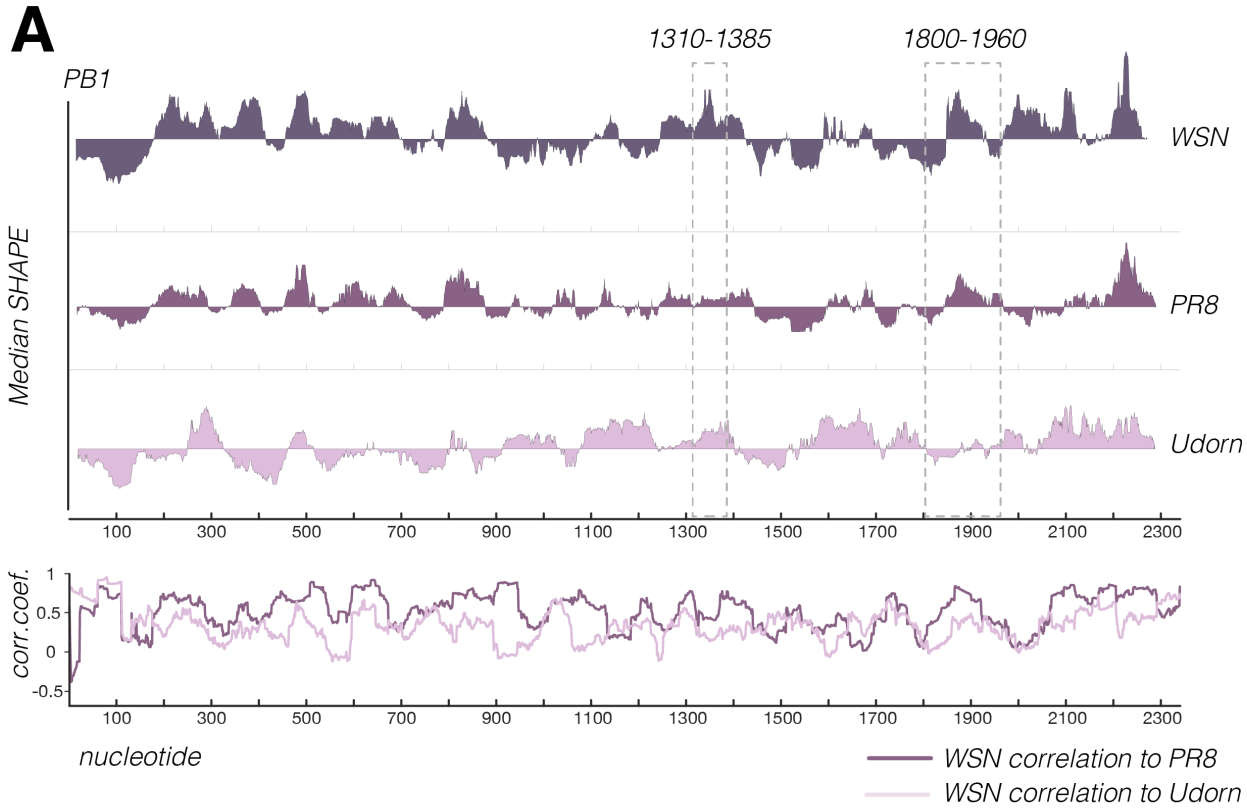


Figure 4.19 SHAPE-MaP analysis for the PB2 segment of different IAV strains

*(**A**) Median SHAPE reactivities for WSN, PR8 and Udorn strains. Median SHAPE reactivities were calculated over sliding 50 nucleotide windows and plotted relative to the global SHAPE median for the segment. 50 nucleotide sliding-window Pearson correlation analysis for the SHAPE reactivities between the different viruses is shown below the median SHAPE reactivity plots. (**B**) Modelled secondary RNA structures formed in the regions highlighted with the dashed rectangles in **A.** Nucleotides in the secondary RNA structures are coloured by their SHAPE reactivities, where high SHAPE reactivities are indicated in red, low reactivities in black and average reactivities in orange colour.*



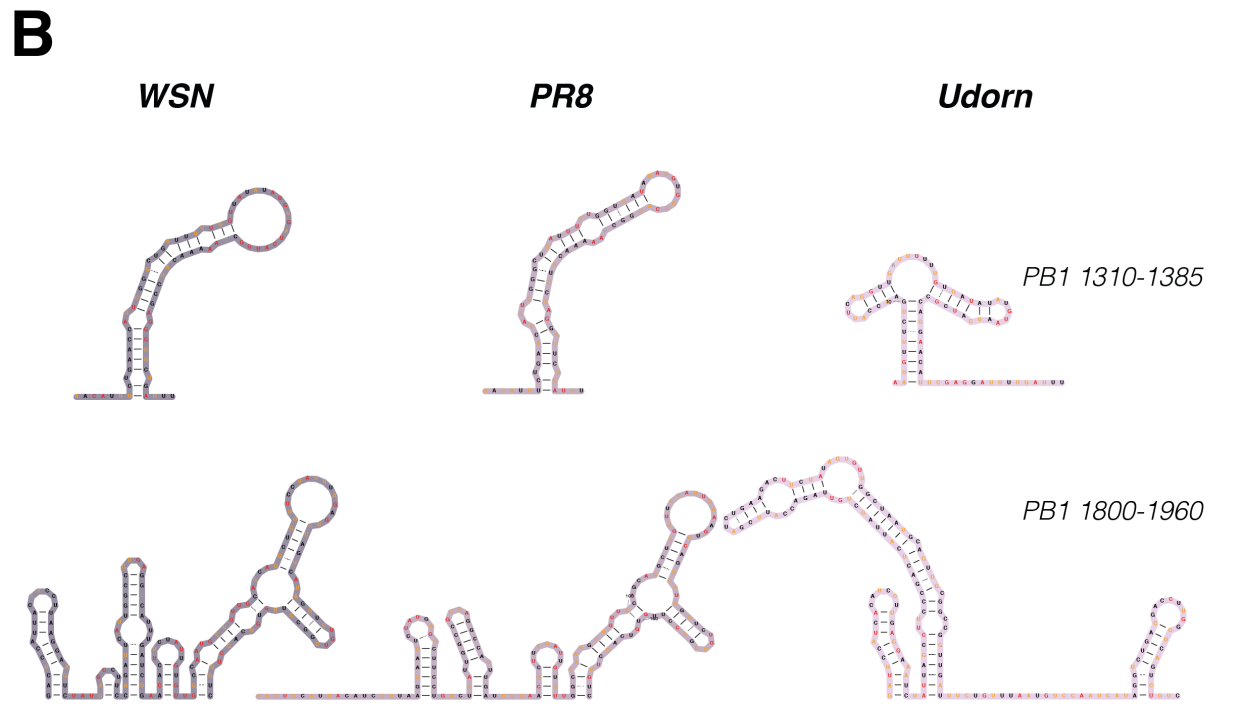


Figure 4.20 SHAPE-MaP analysis for the PB1 segment of different IAV strains

*(**A**) Median SHAPE reactivities for WSN, PR8 and Udorn strains. Median SHAPE reactivities were calculated over sliding 50 nucleotide windows and plotted relative to the global SHAPE median for the segment. 50 nucleotide sliding-window Pearson correlation analysis for the SHAPE reactivities between the different viruses is shown below the median SHAPE reactivity plots. (**B**) Modelled secondary RNA structures formed in the regions highlighted with the dashed rectangles in **A.** Nucleotides in the secondary RNA structures are coloured by their SHAPE reactivities, where high SHAPE reactivities are indicated in red, low reactivities in black and average reactivities in orange colour.*

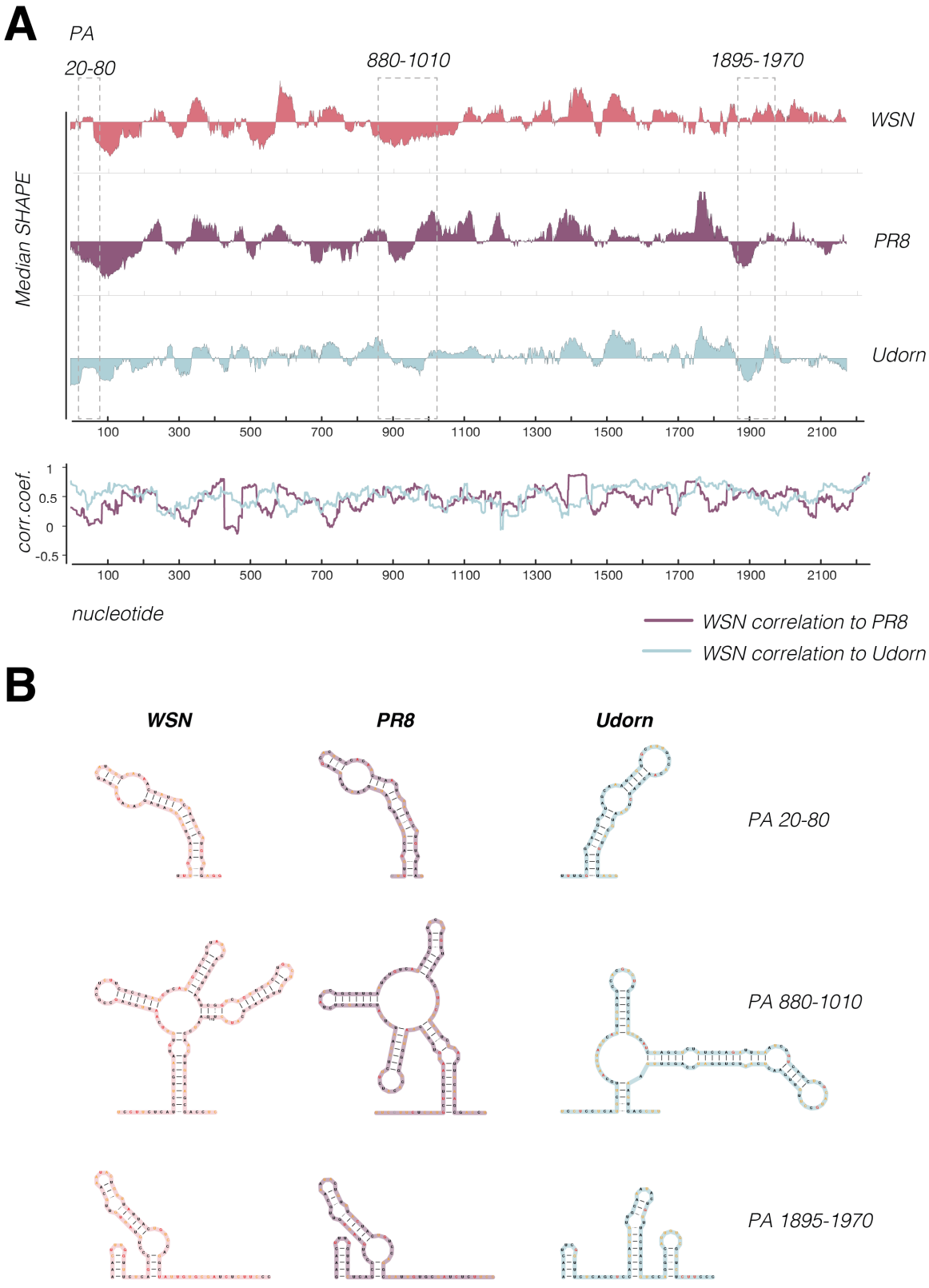
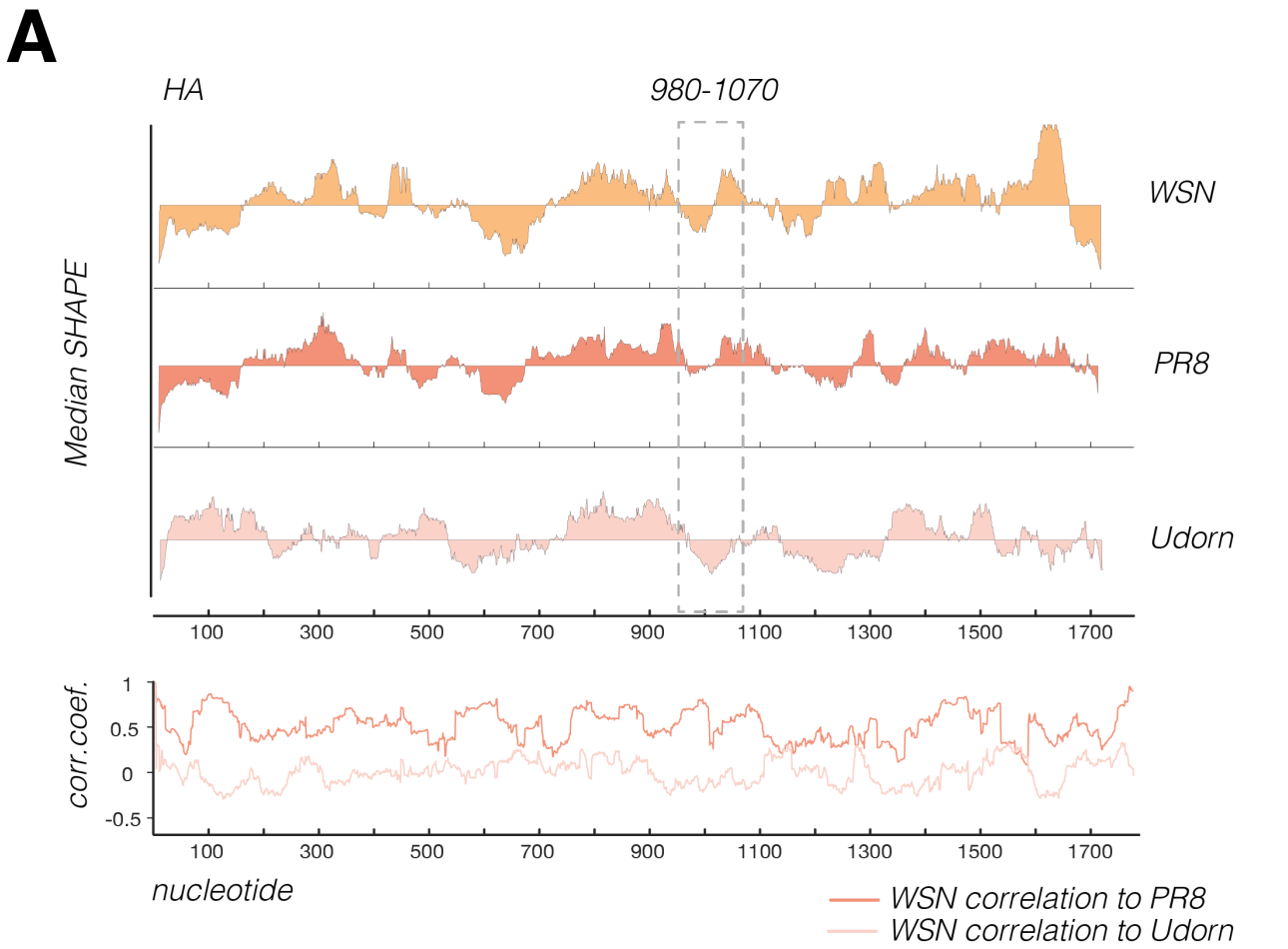


Figure 4.21 SHAPE-MaP analysis for the PA segment of different IAV strains

*(**A**) Median SHAPE reactivities for WSN, PR8 and Udorn strains. Median SHAPE reactivities were calculated over sliding 50 nucleotide windows and plotted relative to the global SHAPE median for the segment. 50 nucleotide sliding-window Pearson correlation analysis for the SHAPE reactivities between the different viruses is shown below the median SHAPE reactivity plots. (**B**) Modelled secondary RNA structures formed in the regions highlighted with the dashed rectangles in **A.** Nucleotides in the secondary RNA structures are coloured by their SHAPE reactivities, where high SHAPE reactivities are indicated in red, low reactivities in black and average reactivities in orange colour.*



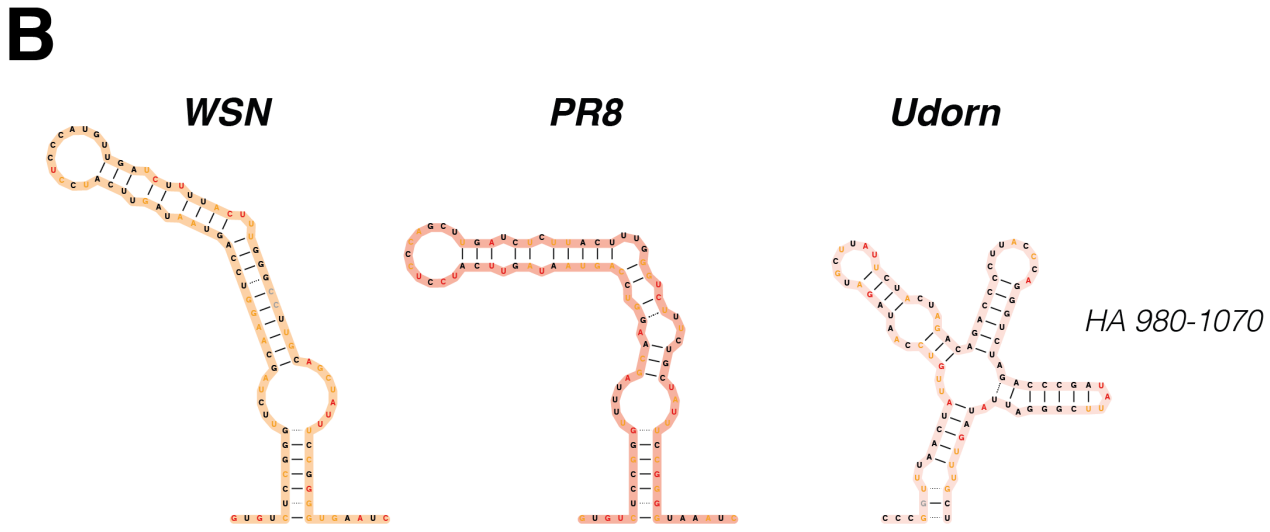
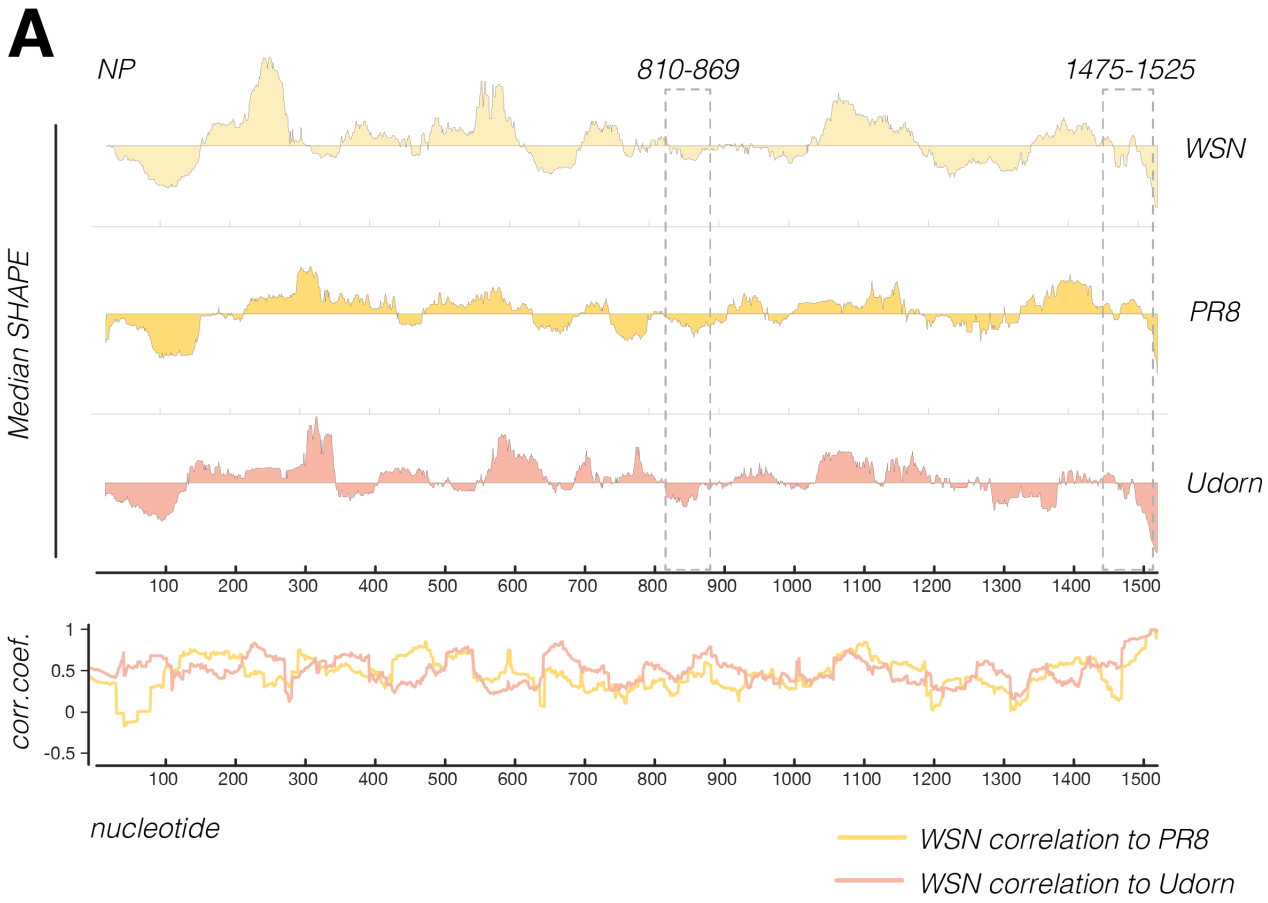


Figure 4.22 SHAPE-MaP analysis for the HA segment of different IAV strains

*(**A**) Median SHAPE reactivities for WSN, PR8 and Udorn strains. Median SHAPE reactivities were calculated over sliding 50 nucleotide windows and plotted relative to the global SHAPE median for the segment. 50 nucleotide sliding-window Pearson correlation analysis for the SHAPE reactivities between the different viruses is shown below the median SHAPE reactivity plots. (**B**) Modelled secondary RNA structures formed in the regions highlighted with the dashed rectangles in **A.** Nucleotides in the secondary RNA structures are coloured by their SHAPE reactivities, where high SHAPE reactivities are indicated in red, low reactivities in black and average reactivities in orange colour.*



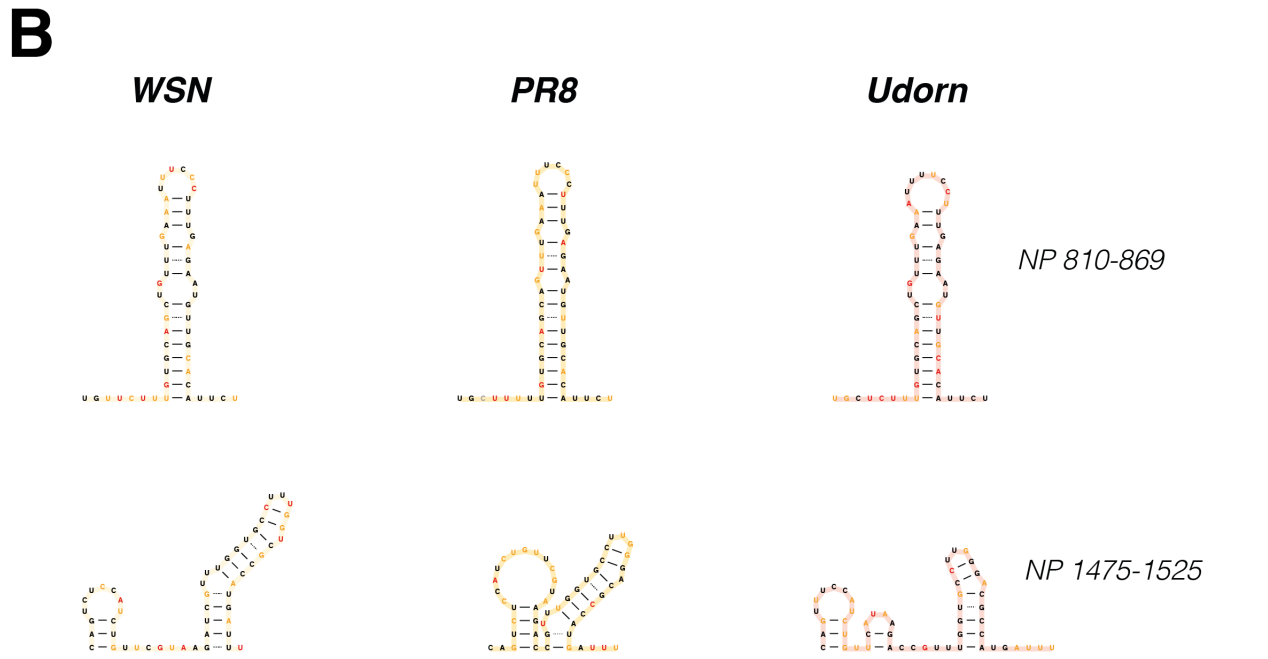


Figure 4.23 SHAPE-MaP analysis for the NP segment of different IAV strains

*(**A**) Median SHAPE reactivities for WSN, PR8 and Udorn strains. Median SHAPE reactivities were calculated over sliding 50 nucleotide windows and plotted relative to the global SHAPE median for the segment. 50 nucleotide sliding-window Pearson correlation analysis for the SHAPE reactivities between the different viruses is shown below the median SHAPE reactivity plots. (**B**) Modelled secondary RNA structures formed in the regions highlighted with the dashed rectangles in **A.** Nucleotides in the secondary RNA structures are coloured by their SHAPE reactivities, where high SHAPE reactivities are indicated in red, low reactivities in black and average reactivities in orange colour.*

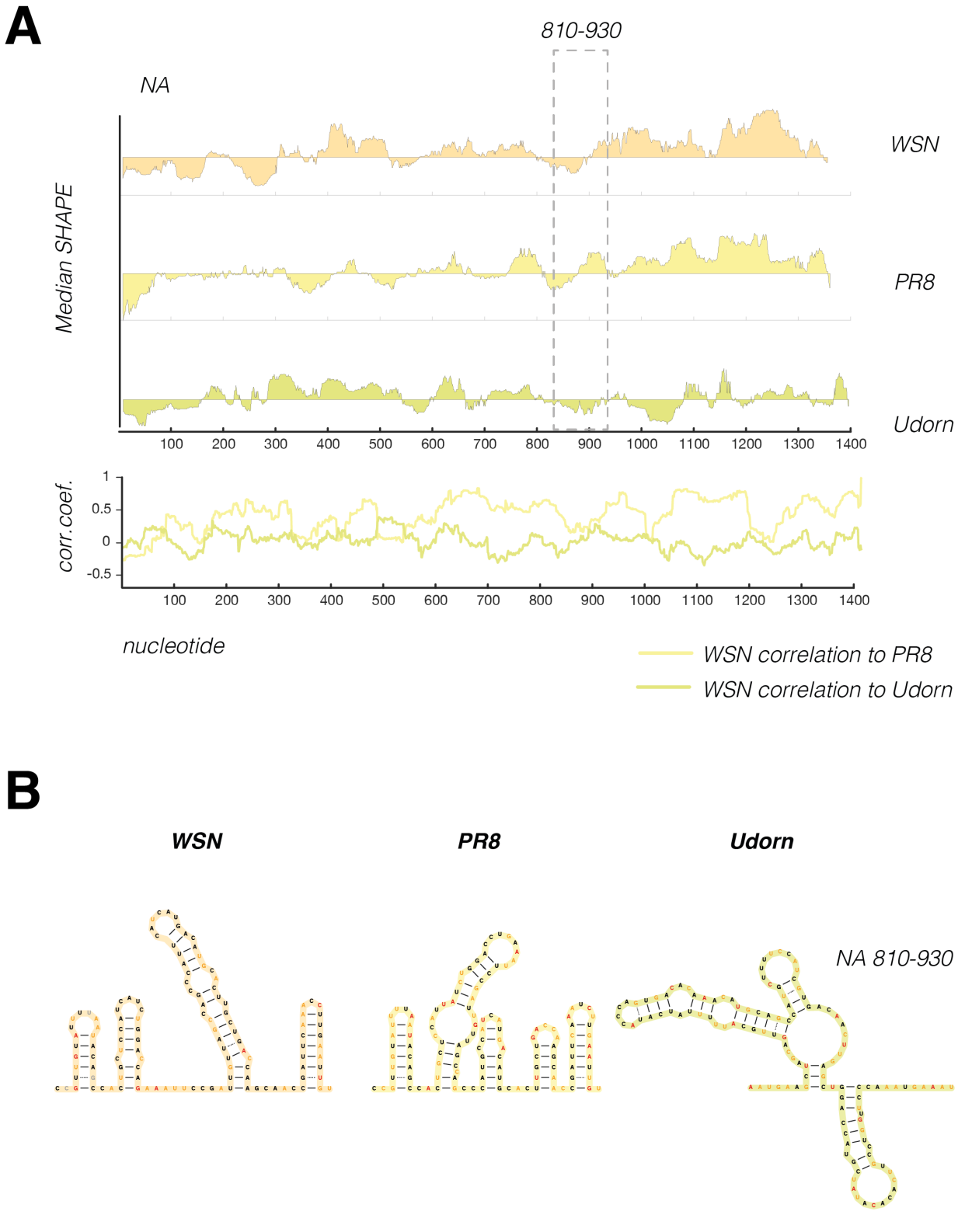


Figure 4.24 SHAPE-MaP analysis for the NA segment of different IAV strains

*(**A**) Median SHAPE reactivities for WSN, PR8 and Udorn strains. Median SHAPE reactivities were calculated over sliding 50 nucleotide windows and plotted relative to the global SHAPE median for the segment. 50 nucleotide sliding-window Pearson correlation analysis for the SHAPE reactivities between the different viruses is shown below the median SHAPE reactivity plots. (**B**) Modelled secondary RNA structures formed in the regions highlighted with the dashed rectangles in **A.** Nucleotides in the secondary RNA structures are coloured by their SHAPE reactivities, where high SHAPE reactivities are indicated in red, low reactivities in black and average reactivities in orange colour.*

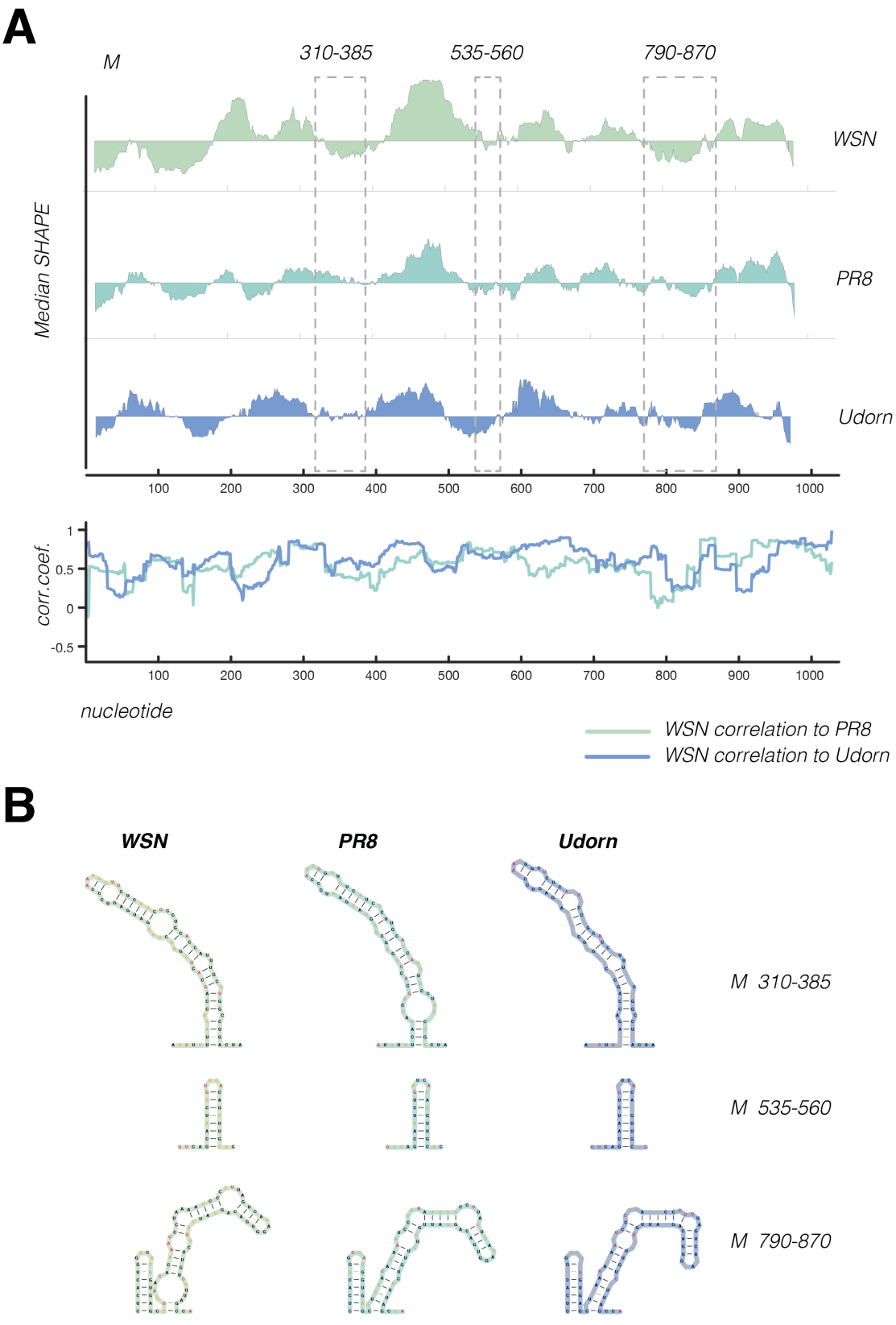


Figure 4.25 SHAPE-MaP analysis for the M segment of different IAV strains

*(**A**) Median SHAPE reactivities for WSN, PR8 and Udorn strains. Median SHAPE reactivities were calculated over sliding 50 nucleotide windows and plotted relative to the global SHAPE median for the segment. 50 nucleotide sliding-window Pearson correlation analysis for the SHAPE reactivities between the different viruses is shown below the median SHAPE reactivity plots. (**B**) Modelled secondary RNA structures formed in the regions highlighted with the dashed rectangles in **A.** Nucleotides in the secondary RNA structures are coloured by their SHAPE reactivities, where high SHAPE reactivities are indicated in red, low reactivities in black and average reactivities in orange.*

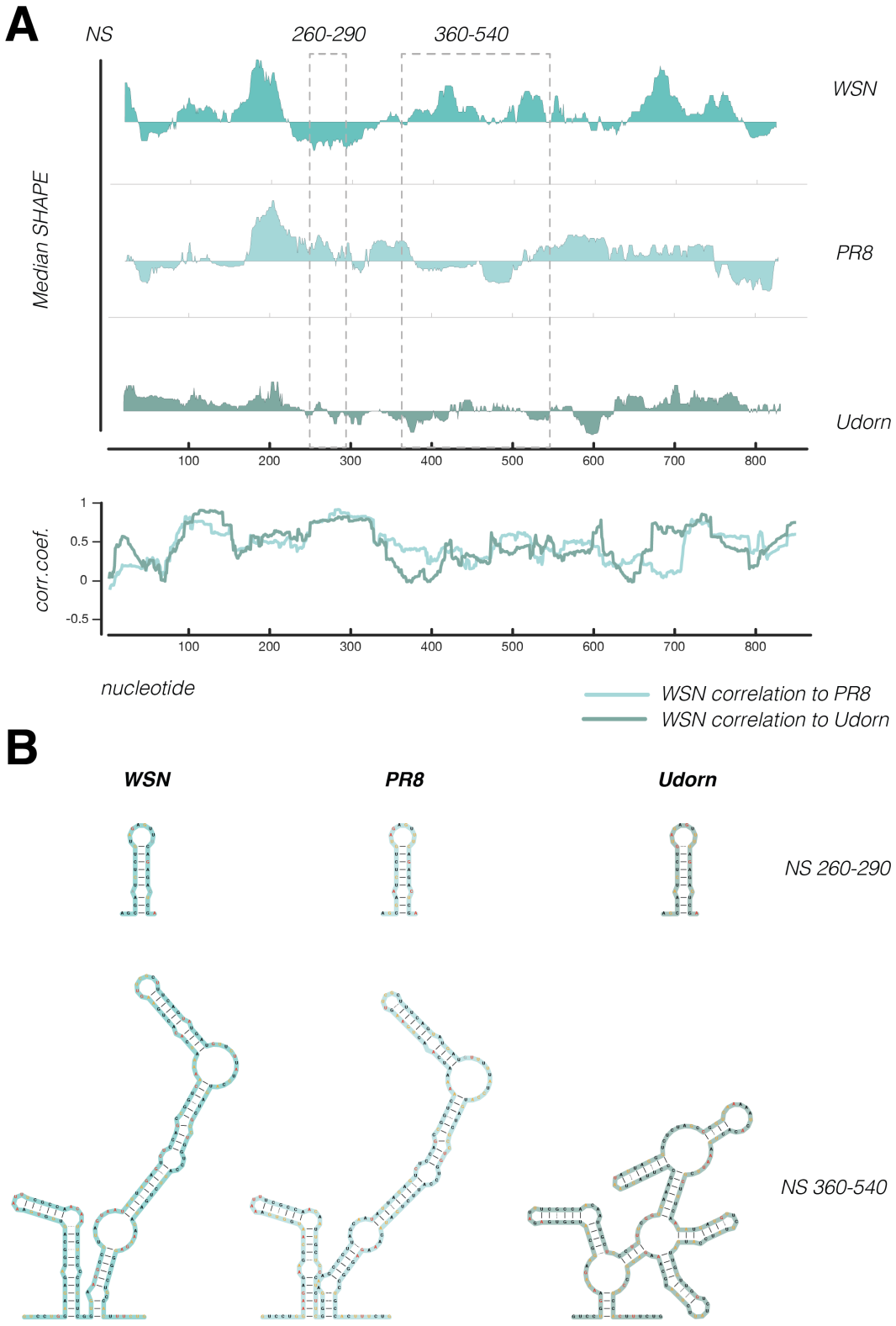


Figure 4.26 SHAPE-MaP analysis for the NS segment of different IAV strains

*(**A**) Median SHAPE reactivities for WSN, PR8 and Udorn strains. Median SHAPE reactivities were calculated over sliding 50 nucleotide windows and plotted relative to the global SHAPE median for the segment. 50 nucleotide sliding-window Pearson correlation analysis for the SHAPE reactivities between the different viruses is shown below the median SHAPE reactivity plots. (**B**) Modelled secondary RNA structures formed in the regions highlighted with the dashed rectangles in **A.** Nucleotides in the secondary RNA structures are coloured by their SHAPE reactivities, where high SHAPE reactivities are indicated in red, low reactivities in black and average reactivities in orange colour.*

The divergence between the genes of the influenza A and B viruses has occurred at least 4000 years ago (Suzuki & Nei, 2002). As there is little nucleotide sequence conservation between IAV and IBV no SHAPE reactivity correlation analysis was performed between the two subtypes. SHAPE-MaP analysis for the FLB genome showed that, as for IAVs, different genome segments of the FLB virus had distinct SHAPE reactivity profiles and contained regions of low SHAPE reactivities, which could be modelled to predict the presence of secondary RNA structures in them, as is illustrated in Figure 4.27.

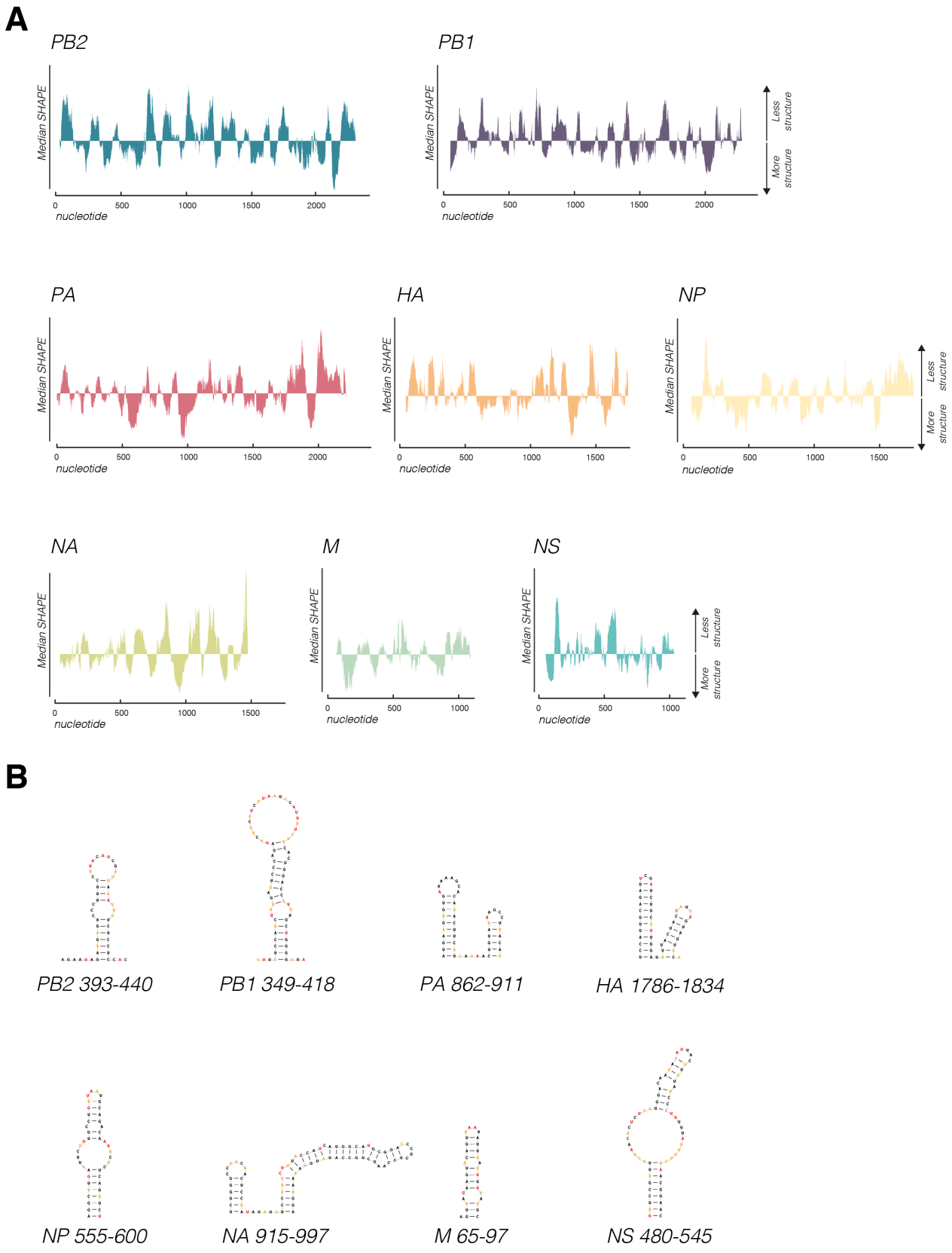


Figure 4.27 SHAPE-MaP analysis for FLB genome

*(**A**) Median SHAPE reactivities for the different FLB genome segments. Median SHAPE reactivities were calculated over sliding 50 nucleotide windows and plotted relative to the global SHAPE median for each segment. (**B**) Examples of secondary RNA structures present in the FLB genome, coloured by nucleotide SHAPE reactivities.*

Comparison between the SHAPE reactivities of the nucleotides located closer to the 5′ end versus those located closer to the 3′ end, indicated that, as for the WSN, there is a trend for the 5′ region of the different segments to be more structured than the 3′ end in other strains as well, and this difference is especially clear in the polymerase segments (Figure 4.28).

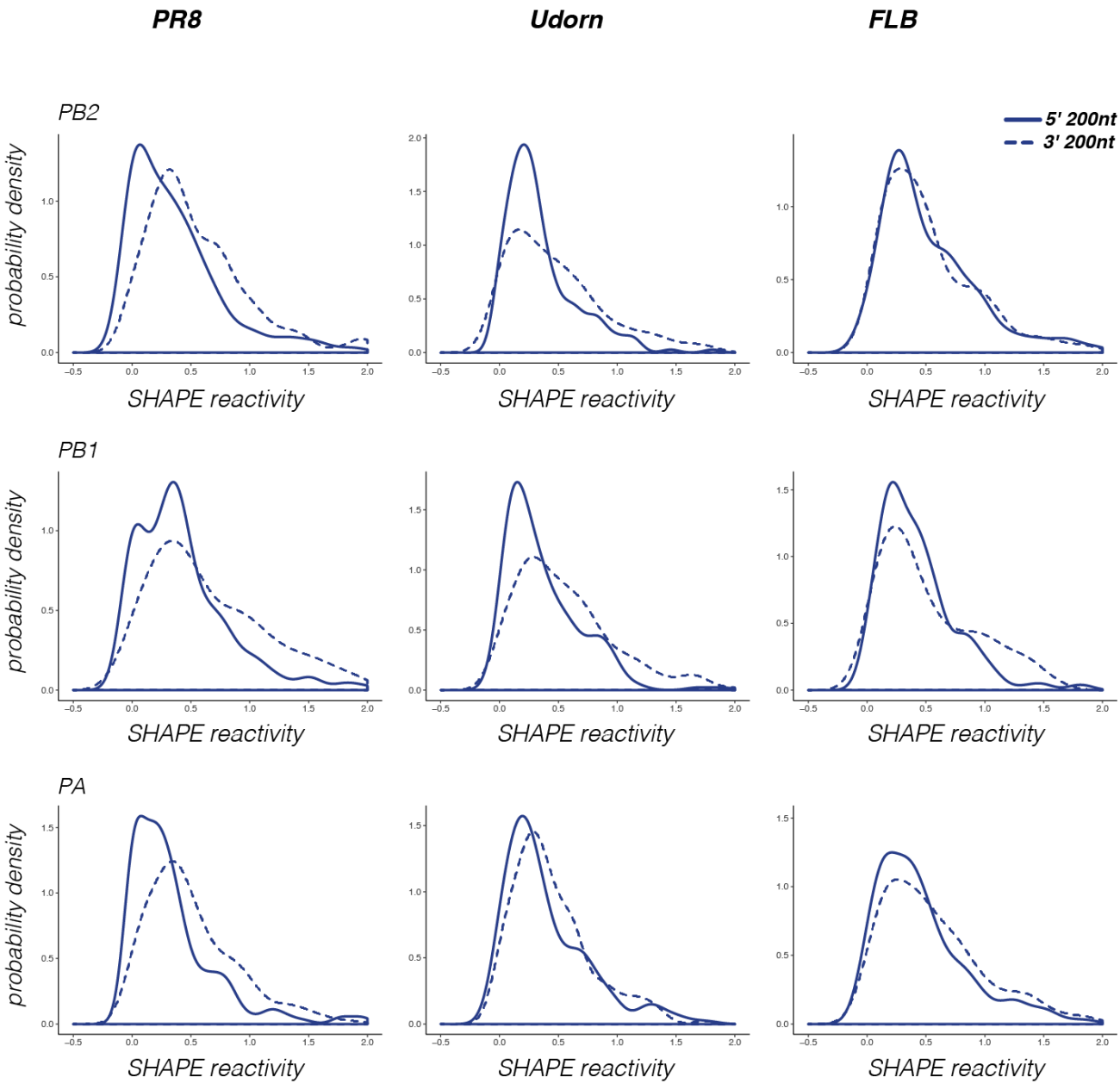


Figure 4.28 Comparison between the SHAPE reactivities of the nucleotides located in the 5′ versus the 3′ region of the vRNA

Probability distributions for the nucleotides of the polymerase segments are shown, the missing terminal nucleotides from the sequencing data were excluded from the analysis.

4.5 Discussion

This chapter presented the analysis of the secondary RNA structures present in influenza viruses using the SHAPE-MaP method. SHAPE reactivities were acquired for every vRNA segment of the different viruses, which were in turn used to model the secondary RNA structures. For the WSN virus both *ex virio* and *in virio* conditions were compared and the importance of the secondary RNA structures in the virus life cycle was assessed.

In virio SHAPE reactivity profiling for WSN virus showed high reproducibility across multiple independent replicates, which suggested that the vRNA structure in the vRNPs is constrained to a conformation, which is preserved over time. For all the viruses tested in this work, different vRNA segments were found to have significantly different SHAPE profiles, indicating that each segment has a unique RNA fold (Figure 4.19, Figure 4.27). While the latter observation is not surprising *ex virio,* because each segment has a different nucleotide sequence and would be expected to have a different structure, it is however an interesting observation *in virio*, where the binding of NP would perhaps be expected to melt the secondary RNA structures and lead to a predominantly single stranded RNA. RNAs that are all single stranded would be expected to have very similar SHAPE profiles (SHAPE reactivities are not base-type dependent), as would a single stranded vRNA with the NP protein bound at regular intervals.

The periodicity analysis for the vRNA SHAPE reactivities did not find any pattern that was unique to the *in virio* SHAPE data but absent *ex virio*. It was hypothesised that if NP is binding to the vRNA at regular intervals it should lead to a footprint in the SHAPE reactivities; the absence of a footprint unique to the *in virio* data therefore suggest that NP is unlikely to be bound to the vRNA at fixed intervals. In fact, this might imply that the sites of NP binding to the vRNA during the RNP replication are influenced by the presence/absence of secondary RNA structures at a given position. Alternatively, NP could have a nonuniform effect on the SHAPE reactivities across the vRNA. For example, if NP is binding vRNA more tightly at one site than another perhaps only the tight binding of NP will be detected as low-SHAPE region in which case periodicity would be difficult to detect. A nonuniform binding of NP would be supported by the recent NP-footprinting data in which the NP-footprint was highly uneven across the vRNA and NP binding was not just

present or absent (Williams, Townsend, Wylie, *et al.*, 2018). There is no evidence in the literature which shows that NP can "melt" double stranded RNA, but it is known that NP preferentially binds single stranded RNA (Yamanaka, Ishihama & Nagata, 1990). Consequently, a replication model can be envisaged in which, 1) the 5' end of the vRNA emerging from a replicating polymerase starts folding onto itself, 2) the promoter region of the emerging vRNA is captured by the trans-activating polymerase, 3) the vRNA continues to fold onto itself until the first NP molecule becomes available (probably recruited via weak polymerase-NP interactions) and 3) a subsequent oligomerisation of NP via its groove and tail-loop interactions occurs more rapidly, leading to a more sparsely distributed secondary structure formation towards the 3' end of the vRNA. Such model would explain the observation that the nucleotides towards the 5' region of the vRNA tend to be more structured compared to the nucleotides towards the 3' end (Figure 4.3).

A recent study investigating NP binding to vRNA in the RNPs, has found that regions of vRNA are present which are free of NP (Williams, Townsend, Wylie, *et al.*, 2018). The study concluded that the regions lacking NP are sites where secondary RNA structures in the influenza virus genome could potentially form. It is worth noting that the data from the NP foot-printing study match well with the SHAPE analysis presented in this chapter, especially for the regions with low NP, for which the SHAPE data clearly shows the presence of secondary RNA structures (Figure 4.29). The presence of local secondary RNA structures in the WSN genome is also supported by the comparison between *in virio* and *ex virio* SHAPE analysis data. *In virio* and *ex virio* vRNA SHAPE reactivities correlate well within <150 nucleotide windows, suggesting common structures are being formed.

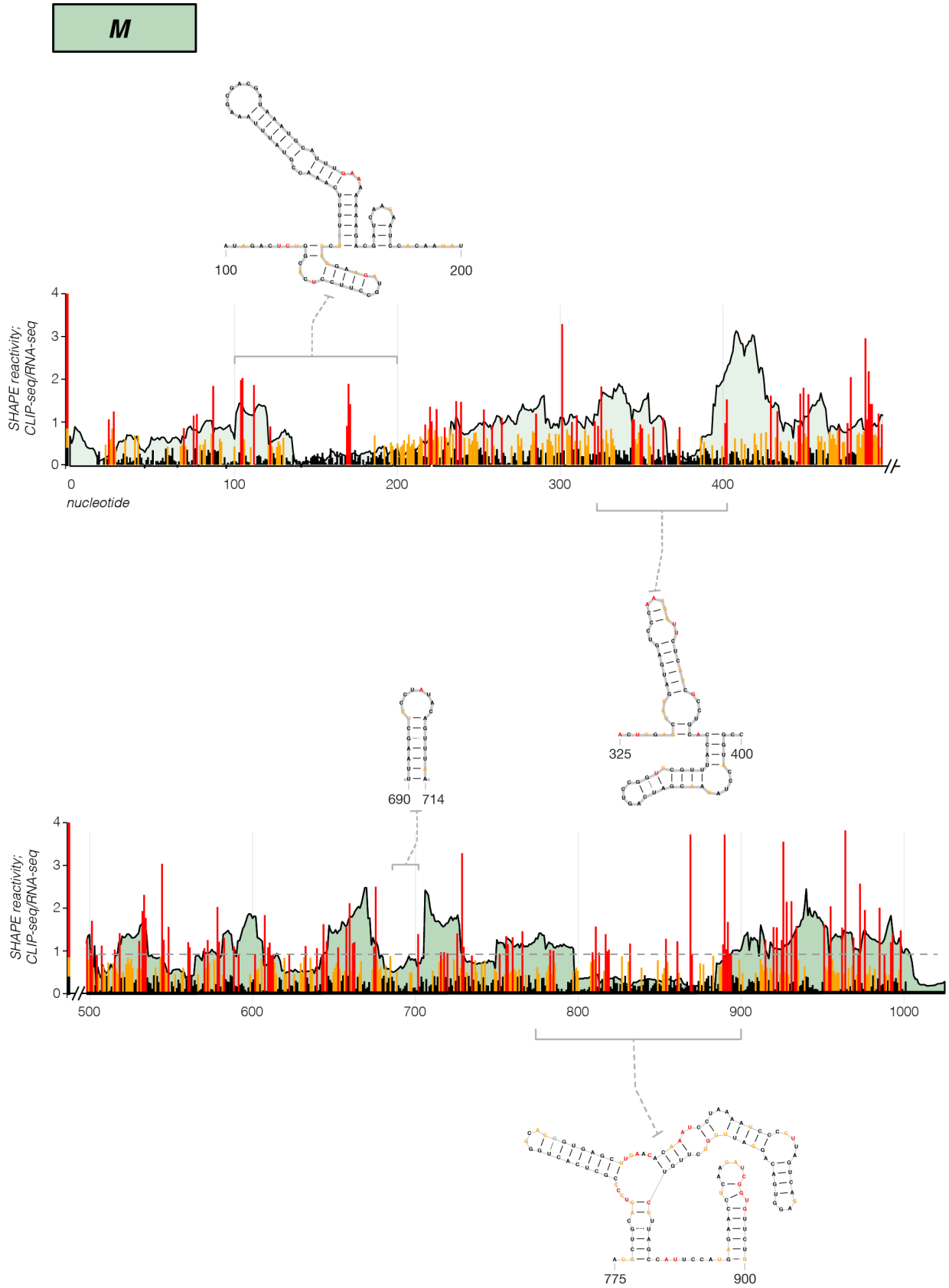


Figure 4.29 Comparison between SHAPE-MaP and NP footprinting analysis

NP footprinting data is shown in the green graph and is presented as a ratio between of NP-CLIP and total RNA-Seq reads, the dashed line indicates the average NP-CLIP/ RNA-Seq ratio; values which fall below the line are regions with lower than average NP binding. NP foot-printing data was adapted from Williams et al. (2017).

For the WSN virus two different *ex virio* conditions were used, *nkvRNA* and *ivtRNA*. The two conditions resulted in almost identical SHAPE reactivity profiles though there was substantial difference in the experimental procedures: for the *nkvRNA* all eight vRNA segments were folded and SHAPE modified together, whereas for the *ivtRNA* all segments were modified independently of each other. It is possible that the eight vRNAs in *nkvRNA* condition could interact with each other, in which case the regions in which the RNA-RNA interactions occur should have lower SHAPE reactivities compared to the *ivtRNA* condition. Furthermore, if the same RNA-RNA interactions are preserved *in virio*, the SHAPE reactivity profiles in the interacting regions would be expected to be more similar between *in virio* and *nkvRNA* compared to *in virio* and *ivtRNA* profiles. In fact, a similar approach has been successfully employed previously, where *in vitro* interactions between naked vRNAs were used to identify interactions occurring between vRNPs during a viral infection (Gavazzi, Yver, Isel, *et al.*, 2013; Gavazzi, Isel, Fournier, *et al.*, 2013). The correlation analysis, however, did not identify any such regions, either indicating that the vRNAs do not interact or, more likely, due to experimental conditions, which were not optimised for RNA-RNA interactions between *nkvRNA* segments to occur. In particular, the molar concentration of *nkvRNA* should be titrated to the point where intersegment RNA-RNA interactions are favoured over self-folding. In addition, in order to use the SHAPE method to identify intersegment interaction sites, a more informative approach would be to look for differences in the SHAPE profiles between the vRNA modified *in virio* and the vRNA modified using individually purified vRNPs (either form cells or virions). Such differences would indicate true changes in RNA-RNA interactions rather than possible differences in the intramolecular RNA folding.

When SHAPE-informed RNA structure modelling was used to predict secondary RNA structures present inside the WSN genome, it was shown that compared to the *ex virio* RNA structure, *in virio* fewer high-probability secondary RNA structures were found to form. This observation supports previous suggestions that NP helps to maintain the vRNA in a less structured conformation and acts as an elongation factor required for the viral polymerase during genome replication (Yamanaka, Ishihama & Nagata, 1990; Baudin, Bach, Cusack, *et al.*, 1994; Turrell, Lyall, Tiley, *et al.*, 2013). While SHAPE reactivities *in virio* were found to be consistent in different experiments, it is worth noting that the calculated SHAPE profiles are averages of all the potential vRNA conformations; it is difficult to know if there is a single vRNA conformation in the RNPs or if

there is a pool of multiple conformations, which are always present in the RNP pool and therefore, always average-out to give the same SHAPE reactivity profile. Consequently, the vRNA structure was modelled using a partition function, which, rather than providing one definitive RNA structure, assigns probabilities for base-pairing between nucleotides, given their sequence and SHAPE reactivities.

Several structures in the WSN genome, which were associated with high base-pairing probabilities, were further investigated using a synonymous mutagenesis approach. Disruptions of several structures in the genome had no or only a marginal effect on the virus growth, implying that not every secondary RNA structure necessarily has a biological function (Figure 4.12). Surprisingly, *the PB2 1301-1385(6mut)* virus grew better than the wild-type at early time points during infection. It would be interesting to test this mutant virus for the production of defective particles as the high titres early during infection might be due to complementation. Complementation can occur when a virus stock has multiple virions which lack the full set of genomic segments. When the virus titre is measured the defective virions will not produce plaques, consequently, in subsequent experiments, when equal amounts of infectious virus are counted for the infection, the actual number viral particles used could differ. Multiple defective viral particles could complement each other's replication through co-infection, leading to greater viral titres.

Multiple viruses with mutated secondary RNA structure could not be rescued, suggesting that those secondary RNA structures are important for virus replication (Figure 4.16 A). Further experiments need to be done to determine which steps of the viral life cycle these secondary RNA structures are involved in. Interestingly, introduction of 17 single nucleotide mutations into NS 366-411 hairpin led to significant reduction in vRNA production in the RNP reconstitution assay. This might explain why the *NS 364-411(17mut)* virus could not be rescued but it is not clear why destabilising the NS 364-411 hairpin causes such a reduction in replication. Disruption of the PB1 1309-1391 stem-loop led to changes in the vRNA content inside virions, suggesting that some secondary RNA structures are important for viral genome packaging.

In addition, ASOs were used to target virus during replication. The advantage of using ASOs to target secondary RNA structures is that the wild-type virus can be used for these assays, which avoids unknown and structure-unrelated effects of mutagenesis. Preliminary experiments suggested that the presence of structure-disrupting ASOs can significantly decrease virus titres. Further experiments should be conducted, with optimised transfection and infection conditions and a control ASO, which would target non-structured regions of vRNA. The latter control would be an important condition to include, as DNA-RNA hybrids in cells are known to be targets for cellular RNases (though the presence of LNAs should reduce this effect), so the presence of ASOs may simply cause vRNA degradation rather than allow to assess the effects of RNA structure perturbation on virus replication. It is worth noting that effective ASOs could potentially provide a good antiviral strategy for the treatment of influenza infections.

Several secondary RNA structures analysed in the WSN genome were also found in PR8 and Udorn viruses. These include HA 996-1050 and PB1 1309-1391 stem-loops, found in both WSN and PR8, and PB2 1301-1385 and NS 366-411, found in all three strains. Further analysis will need to be done to determine if these secondary RNA structures are important for replication of viruses other than WSN. To assess the importance of other secondary RNA structures identified in this work and to extend the analysis to other virus strains, sequence conservation and covariance analysis should also be performed. It should be noted though, that secondary RNA structures that do not show any conservation in related viruses, can nevertheless be functionally important. SHAPE reactivity and sequence conservation analyses for alphaviruses has shown that non-conserved structural elements in the alphaviruses genomes can have significant effects on virus growth (Kutchko, Madden, Morrison, *et al.*, 2018) and perhaps such structural differences in the RNA structure could explain the differences in host-range and cell tropism among these viruses. Interestingly, avian influenza A viruses are known to have more structured genomes than the strains infecting humans, as a result of a higher GC content in their genomes (Priore, Moss & Turner, 2013; Greenbaum, Levine, Bhanot, *et al.*, 2008). GC nucleotides can increase the thermodynamic stability of an RNA; this might be important for preserving secondary RNA structures and replication at higher temperatures in the avian hosts (Rabadán, Levine & Robins, 2006).

Comparison between the SHAPE profiles of different IAV strains showed that the most significant differences between the H1N1 and H3N2 strains are found in the HA, NA and PB1

segments, the most evolutionarily divergent segments between the H1N1 and H3N2 strains. WSN and PR8 viruses are related to the H1N1 viruses which have emerged after the 1918 pandemic, whereas Udorn virus belongs to the H3N2 group which has emerged two pandemics later. In 1957, a 1918-derived H1N1 strain has acquired HA, NA and PB1 segments from an avian source to produce an H2N2 virus, which, in 1968, acquired HA and PB1 segments from an avian H3 virus, creating the Udorn-like H3N2 group (Taubenberger & Kash, 2010).

4.6 Summary

- Each genome segment in influenza viruses has a unique RNA conformation
- No clear periodicity unique to the *in virio* SHAPE reactivities was detected, suggesting irregular NP binding
- Multiple secondary RNA structures are present in the genomes of the different viruses tested
- Secondary RNA structures play an important role in the life cycle of the WSN virus
- Some secondary RNA structures shown to attenuate the WSN replication are also found in other IAV strains
- HA, NA and PB1 segments show major structural differences between H1N1 and H3N2 strains

5. Chapter 5: Inter-segment interaction analysis of the influenza virus RNA genome

5.1 Overview

RNA-RNA interactions in the influenza virus genome have been proposed to be important for its genome packaging and the reassortment between different influenza strains. This chapter describes how a modified sequencing of psoralen crosslinked, ligated, and selected hybrids (SPLASH) method, previously used to study RNA-RNA interactions in cells, was used to detect inter-segment RNA interactions in the genomes of different influenza viruses. A redundant network of highly complementary RNA-RNA interactions was found to be present *in virio*, which was not fully conserved in the different virus strains.

5.2 Background

Interactions between different RNAs are known to play important roles in various biological processes, including cellular RNA splicing, translation, degradation, and localisation (Guil & Esteller, 2015). RNA-RNA interactions have been shown to be important for various viruses as well. Formation of RNA dimers via a kissing-loop interaction during HIV replication, is known to be important for HIV genome packaging (Chang & Tinoco, 1997). In rotaviruses and orbiviruses, RNA-RNA interactions mediate the packaging of the eleven segments into the viral capsids (Fajardo, Sung, Celma, *et al.*, 2017; Teodoro Fajardo, 2016). In plant viruses, specific RNA-RNA interactions can activate genome replication (Guenther, Sit, Gracz, *et al.*, 2004). Though no structural evidence for intermolecular RNA-RNA interactions in influenza viruses exists so far, a number of studies have suggested that RNA-RNA interactions may mediate the packaging of the influenza virus genome as well. A key study by Gavazzi et *al.* (2013) has shown that a kissing-loop-like interaction between the PB1 and NS segments could mediate virus genome packaging in certain strains. Other studies have also shown that mutations in one IAV genome segment can directly affect the packaging of a different segment (Gilbertson, Zheng, Gerber, *et al.*, 2016), suggesting that RNA-RNA interactions may be involved. While no structural evidence for intersegment RNA-RNA interactions mediating IAV genome packaging exists so far, scanning

transmission electron microscopy studies have observed filament-like structures linking the packaged vRNPs, which may be indicative of the existence of inter-segment RNA-RNA interactions *in virio* (Noda, Sugita, Aoyama, *et al.*, 2012).

The importance of the intermolecular RNA interactions in many biological processes has consequently stimulated the development of multiple methods for detecting RNA-RNA interactions *in vivo*. Furthermore, with the advancement of high-throughput sequencing, methods have been developed that enable identification of RNA-RNA interactions on a whole genome scale. SPLASH (Aw, Shen, Wilm, *et al.*, 2016), LIGR-seq (Sharma, Sterne-Weiler, O'Hanlon, *et al.*, 2016), PARIS (Lu, Zhang, Lee, *et al.*, 2016), and MARIO (Nguyen, Cao, Yu, *et al.*, 2016) are just some of the more recent methods developed to detect RNA-RNA interactions in cells. The former three methods all use a photoactivatable intercalating reagent, called psoralen, to crosslink any base-paired RNAs. Psoralen has been used to study RNA and DNA structures for many years (Thompson & Hearst, 1983; Skripkin, Isel, Marquet, *et al.*, 1996; Wassarman, 1993), as, upon UV irradiation, it crosslinks pyrimidines through a formation of covalent adducts. In RNA, primarily uridine-uridine bases are crosslinked, although cytidines can be crosslinked at low frequencies as well (Cimino, Gamper, Isaacs, *et al.*, 1985). The advantage of using psoralen is the reversibility of the crosslink by irradiation using short UV wavelengths (Figure 5.1). Furthermore, psoralen has been previously used to show the circular arrangement of the vRNA in the influenza virus RNPs, which occurs due to the base-pairing between the promoter regions of the 5′ and 3′ ends in the vRNA (Hsu, Parvin, Gupta *et al.*, 1987).

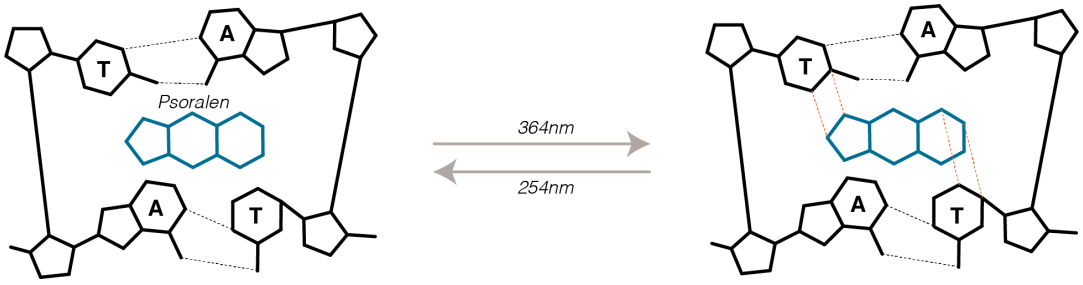


Figure 5.1 Psoralen crosslinking

Upon long-wavelength UV irradiation psoralen forms covalent adducts with pyrimidines. Short UV light can be used to reverse the cross-link.

Chapter 4 described how SHAPE analysis can be used to assess the structuredness of the RNP-associated vRNAs. While extensive regions of low SHAPE reactivities in vRNPs were identified not all such regions could form local intramolecular secondary RNA structures. In addition, the fact that secondary RNA structures can be formed in RNPs, suggests that some of these structures may be exposed to the cellular environment. The exposed vRNA could potentially interact with other RNAs, including other vRNAs, and form inter-segment RNA-RNA interactions. While the SHAPE-MaP method provides information about the structuredness of an RNA, it cannot distinguish between intra- versus intermolecular RNA interactions.

This chapter presents how the SPLASH method has been adapted and used to detect RNA-RNA interactions inside influenza virions. Firstly, SPLASH method was adapted for the *in virio* use; the ability of different psoralen compounds to get inside the viral particles and different sequencing library preparations were tested, as well as a custom sequencing read analysis pipeline was developed. Next, SPLASH method was used to analyse the inter-segment RNA interactions present in WSN, PR8, Udorn and FLB viruses.

5.3 Aims

- Test psoralen crosslinking *in virio*
- Modify SPLASH sequencing library preparation for efficient use with influenza viruses
- Develop sequencing data analysis pipeline for mapping inter-segment interactions
- Identify inter-segment RNA-RNA interactions present in the genomes of WSN, PR8, Udorn and FLB viruses
- Model the structures of RNA-RNA interactions in the different viruses

5.4 Results

5.4.1 *In virio* vRNA cross-linking

As SPLASH method uses a water-soluble psoralen derivative 4'-aminomethyl trioxsalen (AMT) for RNA-RNA crosslinking, first the ability to cross-link vRNAs in virio was tested. To this end, WSN virus was purified over a 30% sucrose cushion and resuspended in a small volume of resuspension buffer. The purified virus stock was divided into equal aliquots, the aliquots were transferred into a 24-well plate and 50 μg/mL of AMT per well was added. The samples were incubated at 37°C for 5 min and the sample plate was transferred onto ice. The lid of the 24-well plate was removed and replaced with a 0.4 mm glass plate. The glass plate was used to filter-out any shorter UV wavelengths, which may cross-link proteins to RNA or cause RNA damage. Still on ice, the samples were placed at approximately 10 cm distance from a hand-held UV lamp and irradiated for different time spans with 365 nm wavelength light. After irradiation, the samples were digested with proteinase K and the vRNA was extracted. Crosslinking was detected by running the purified vRNA on a 3% polyacrylamide gel and observing any shifts in RNA mobility. 45 min irradiation produced a clear shift in RNA mobility, with four particularly distinct slow-mobility bands appearing after the irradiation (Figure 5.2).

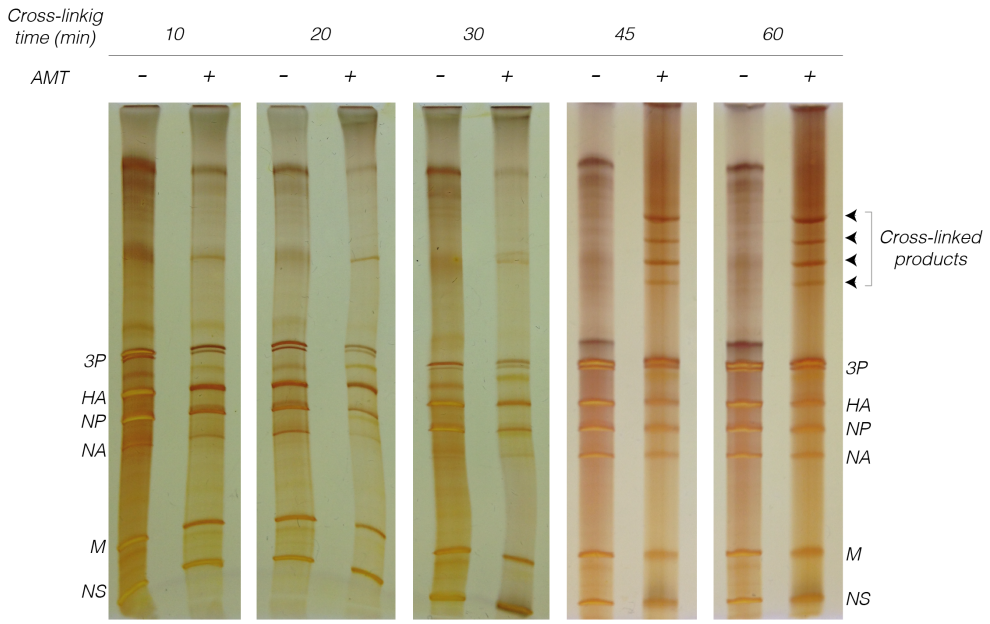


Figure 5.2 RNA mobility shift after AMT cross-linking

Purified WSN viruses were crosslinked with AMT for different times in the presence or absence of AMT. vRNA was extracted from the UV-irradiated samples and analysed on 3% polyacrylamide gels. Distinct cross-linking products are indicated with black arrowheads.

Next the ability of a biotinylated AMT (AMT-B) to crosslink the vRNA was tested. AMT-B has a polyethylene glycol (PEG_3) spacer, which separates the AMT from the biotin tag and makes the biotin tag more accessible after the crosslink. The presence of the PEG_3 and biotin, however, also makes the crosslinker a much larger molecular weight compound (688.79 g/mol vs 293.75 g/mol), which is less efficient at cross-linking, as well as more difficult to introduce into cells. The original SPLASH protocol used a light 0.01% digitonin treatment, to introduce AMT-B into cells, therefore, the same protocol was applied for the *in virio* experiments. When AMT-B-treated virus RNA was separated on a polyacrylamide gel, no distinct shift in the RNA mobility was observed (Figure 5.3 A). However, when the RNA was extracted from the AMT-B cross-linked viruses and tested for biotin incorporation, in the presence of AMT-B, a clear biotin signal was observed (Figure 5.3 B). The lack of any distinct changes in the vRNA mobility likely reflects the low efficiency of cross-linking when AMT-B is used. Notably, AMT-B does not precipitate with the vRNA during RNA extraction if no UV crosslinking is performed (Figure 5.3 **C**, last lane).

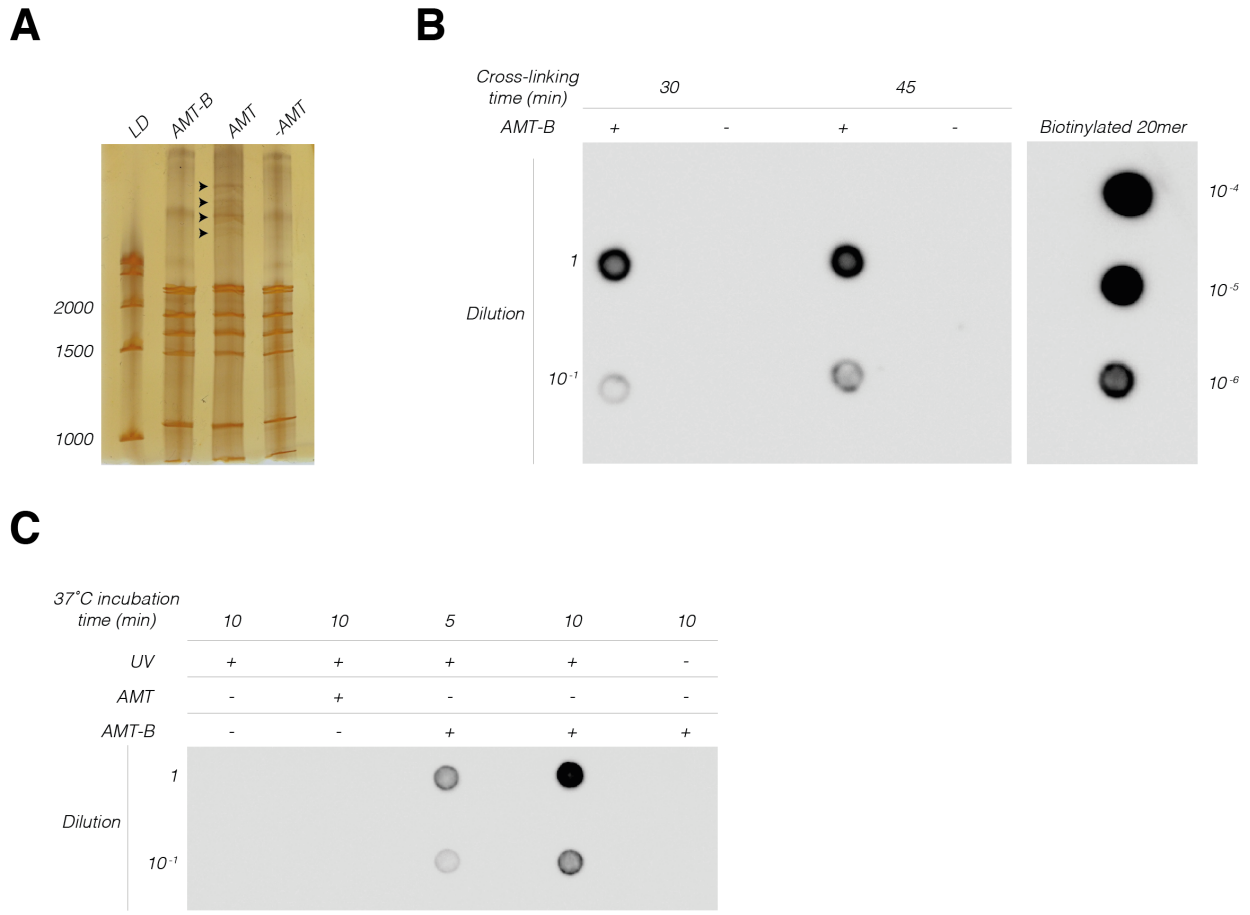


Figure 5.3 Cross-linking vRNA with AMT-B

*(**A**) vRNA was extracted from purified viruses that were irradiated with a 365nm UV light in the presence of AMT, AMT-B, or no crosslinker (-AMT). Changes in vRNA mobility were checked on a 3% polyacrylamide gel. Black arrowheads indicate shifted bands. (**B**) An anti-biotin dot-blot for vRNA crosslinked in the presence or absence of AMT-B for different time spans. A biotinylated 20mer oligonucleotide was used as a positive control. (**C**) Different virus pre-incubation times with AMT-B before the UV cross-linking were tested. Note the lack of biotin signal in the absence of UV irradiation. LD, ladder, size shown in nucleotides.*

An important advantage of using AMT, over other intercalating reagents, to cross-link RNA, is the ability to reverse the crosslink. The reversal of the vRNA crosslink was therefore tested. To this end the AMT-B cross-linked vRNA was irradiated at a 254 nm wavelength and the presence of biotin in the irradiated samples was tested using an anti-biotin dot-blot (Figure 5.4). Longer short UV wavelength irradiation times were found to increase the efficiency of the crosslink reversal. While the efficiency of the reversal increased over time, RNA irradiation with a 254 nm UV light is known to cause extensive damage to the RNA, even after very short exposure times (Kladwang, Hum & Das, 2012). Therefore, the earliest time point (5 min), for which reversal was observed, was used in the later protocol.

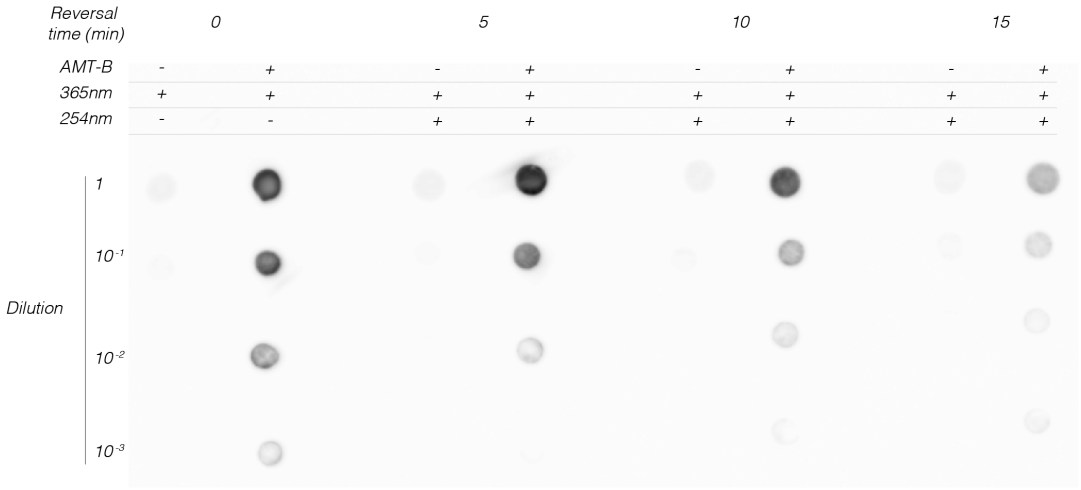


Figure 5.4 AMT-B crosslink reversal

AMT-B cross-linked vRNA was irradiated with a 254 nm wavelength and an anti-biotin dot-blot was used to test the efficiency of crosslink reversal.

5.4.2 *In virio* SPLASH library and sequencing

An overview of SPLASH library preparation steps is provided in Figure 5.5. Three replicates of SPLASH experiments using WSN virus and a single experiment for each of PR8, Udorn and FLB were performed. All SPLASH experiments followed the same RNA cross-linking, enrichment and ligation steps:

1. The virus was purified over a sucrose cushion
2. vRNA was cross-linked *in virio* using AMT-B
3. Crosslinked vRNA was extracted from the virions, fragmented with Mg^{2+} at 95°C, and <200 nucleotide fragments were selected
4. Crosslinked vRNA was enriched on magnetic streptavidin beads
5. vRNA was prepared for ligation by removing 3′-phosphoryl groups and phosphorylating the 5′-OH groups
6. Samples were divided into two equal aliquots and proximity ligation was performed in the presence or absence of ssRNA ligase
7. AMT-B crosslink was removed by irradiating the RNA at short UV wavelength

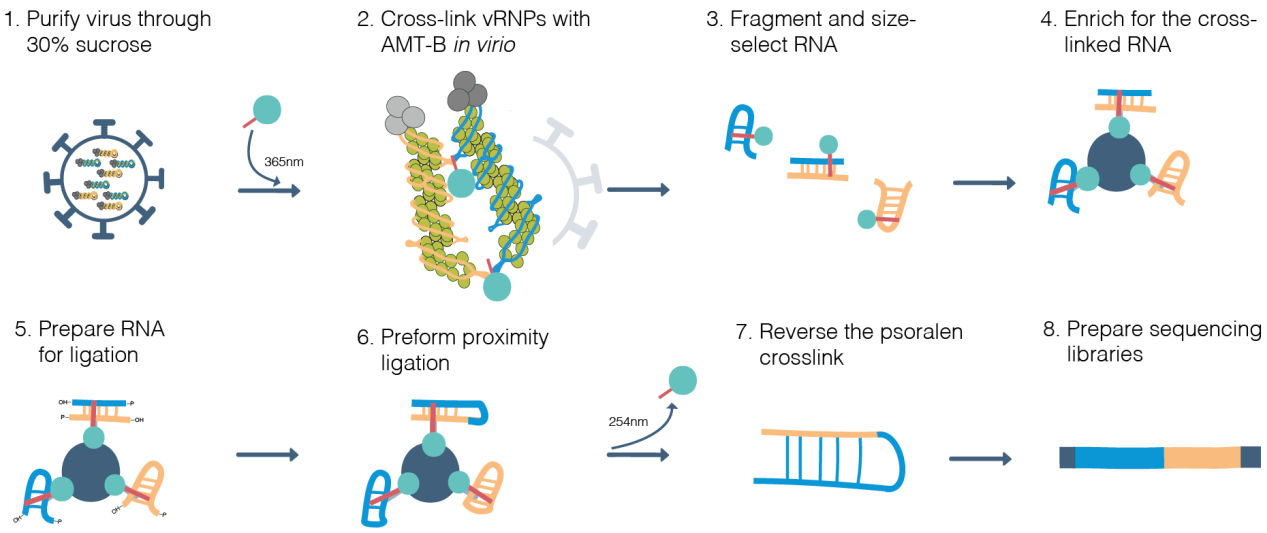


Figure 5.5 SPLASH library RNA cross-linking and ligation steps

The proximity-ligated RNA (referred to as 'chimeric RNA') was subsequently used in the sequencing library preparation. For the first experimental replicate using WSN virus, the library was prepared as described in the originally published SPLASH protocol. In brief, a universal adaptor was ligated onto the 3' end of the RNA and RT was performed using primers annealing to the universal adaptor. The cDNA products were circularised and PCR amplification was performed using primers with multiplexing indexes. PCR-amplified products were size selected for fragments ranging between 200 and 350 nucleotides, and sequenced (**Error! Reference source not found.**).

The issue with the original protocol was found to be that it produced sequencing libraries with poor read diversity (analysis of sequencing reads, discussed below, showed that most reads were PCR duplicates, and very few unique reads were present after the deduplication). The likely reason for the poor diversity was that after the adaptor ligation and reverse transcription, additional steps of size selection and gel purification were required, which led to a significant product loss for samples that already had a very small amount of input (after reverse cross-linking, the amount of chimeric RNA recovered was typically not detectable using a NanoDrop spectrophotometer). Consequently, the library preparation for all the subsequent experiments was changed. After reverse cross-linking the chimeric RNA fragments were used in the SMARTer smRNA-Seq kit protocol, which is optimised for small inputs and works well with small RNA fragments. The SMARTer protocol uses 3' polyadenylation, to create sites for a poly(dT) RT primer

annealing. The RT step is performed with modified MMLV reverse transcriptase (from the SMARTer smRNA-Seq kit), which adds non-templated nucleotides to the cDNA (usually several cytosine nucleotided) (Figure 5.7). The non-templated nucleotides act as annealing sites for a transcript-switching oligo, which is used by the MMLV reverse transcriptase to add a known sequence to the 5′ end of the cDNA; ultimately, cDNA with known sequences at both ends is made.

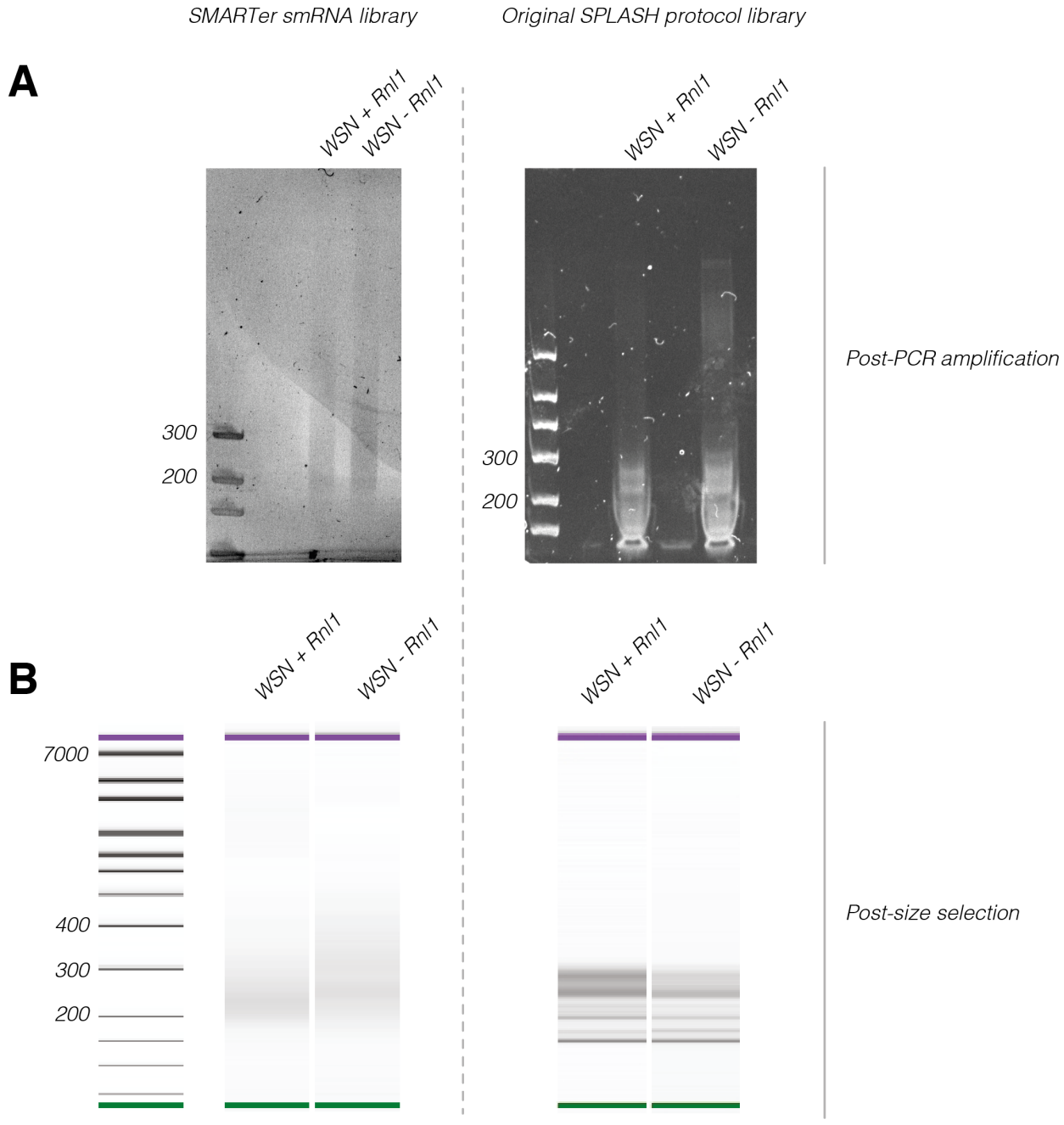


Figure 5.6 SPLASH library PCR amplification and size selection products

*(**A**) dsDNA products from the libraries produced using SMARTer smRNA kit (left) and the originally published SPLASH protocol (right). (**B**) Bioanalyzer 2100 high sensitivity DNA gel trace for samples post size-selection. Rnl1, RNA ligase 1.*

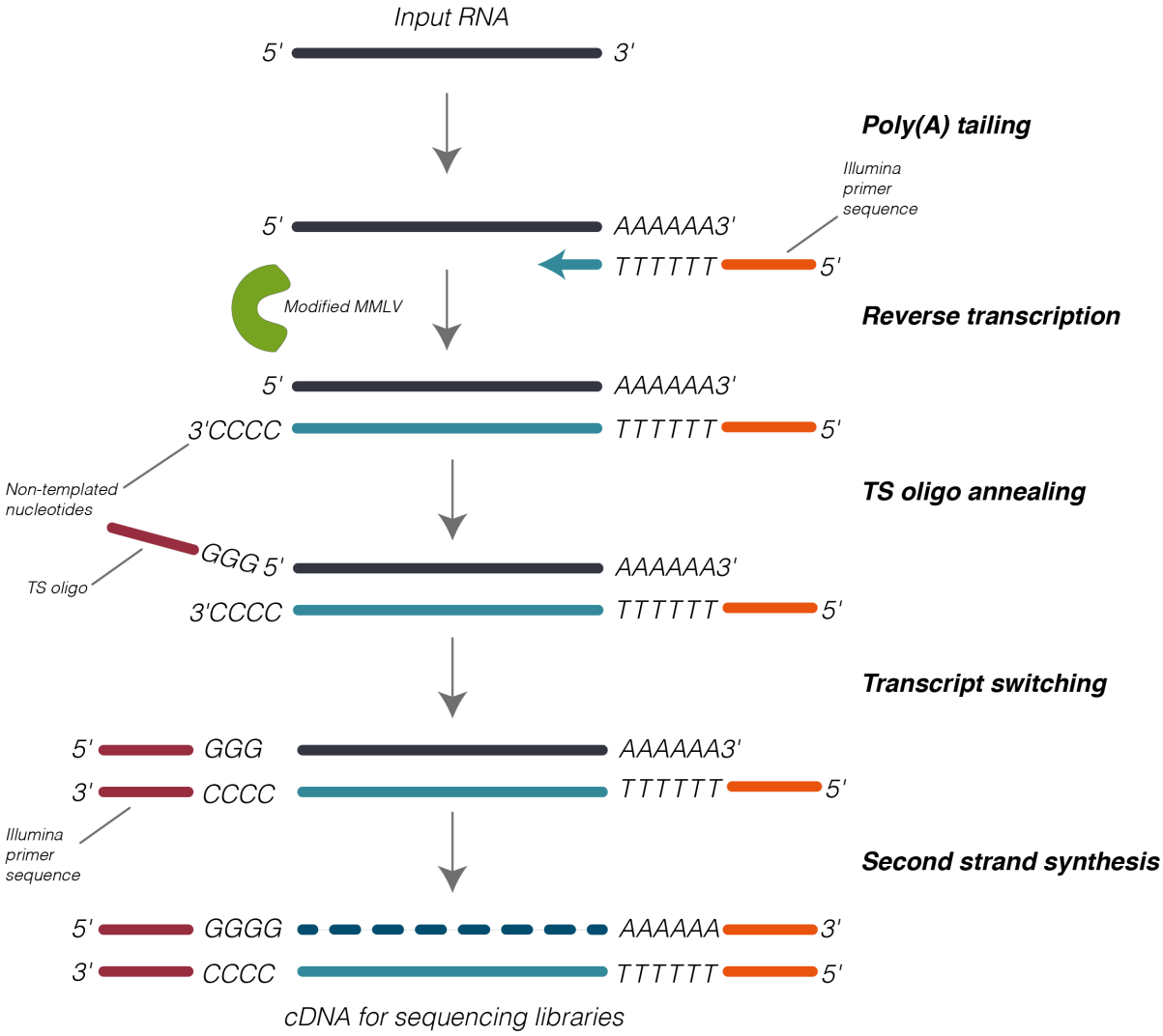


Figure 5.7 cDNA preparation for SPLASH sequencing libraries

The input RNA is poly-adenylated using a poly-A polymerase. A poly-T primer, which also contains Illumina sequencing primer sequence, is used for the first strand cDNA synthesis. The RT step is performed using a modified MMLV reverse transcriptase, which adds nintemplated nucleotides to the synthesisted cDNA. The non-templated nucleotides are used to anneal the transcript-switching (TS) iligo, which also contains Illumina sequencing primer sequence.

5.4.3 SPLASH sequencing and bioinformatics

An overview of the steps involved in SPLASH sequencing and bioinformatic analysis is provided in Figure 5.8. Sequencing for SPLASH experiments was performed using a NextSeq500 instrument with 1x150bp reads. The reads were trimmed using Skewer software (Jiang, Lei, Ding, *et al.*, 2014) and aligned to the reference genome using the STAR aligner (Dobin, Davis, Schlesinger, *et al.*, 2013). For the alignment, the --chimSegmentMin parameter, which turns on the chimeric read search, was set to 20, which defines a requirement for at least 20 nucleotides of each segment in a chimeric fragment to align to the reference genome (the probability of 20 identical

nucleotides occurring in a 13,583-nucleotide genome is 1.2×10^{-8}). The number of chimeric reads after trimming, alignment and deduplication is given in Figure 5.9 A. As mentioned previously, the first WSN experimental replicate had a low unique chimeric read count - while the total mapped number of reads was 201,926, only 2,894 of them were found to be unique after deduplication. Even though the individual WSN experiments were performed months apart, a high read overlap was found between them, indicating that the experiments were reproducible. As the first WSN replicate produced very few unique reads, further analysis for the WSN samples was performed using only the reads that were common between the 2nd and 3rd experimental replicates.

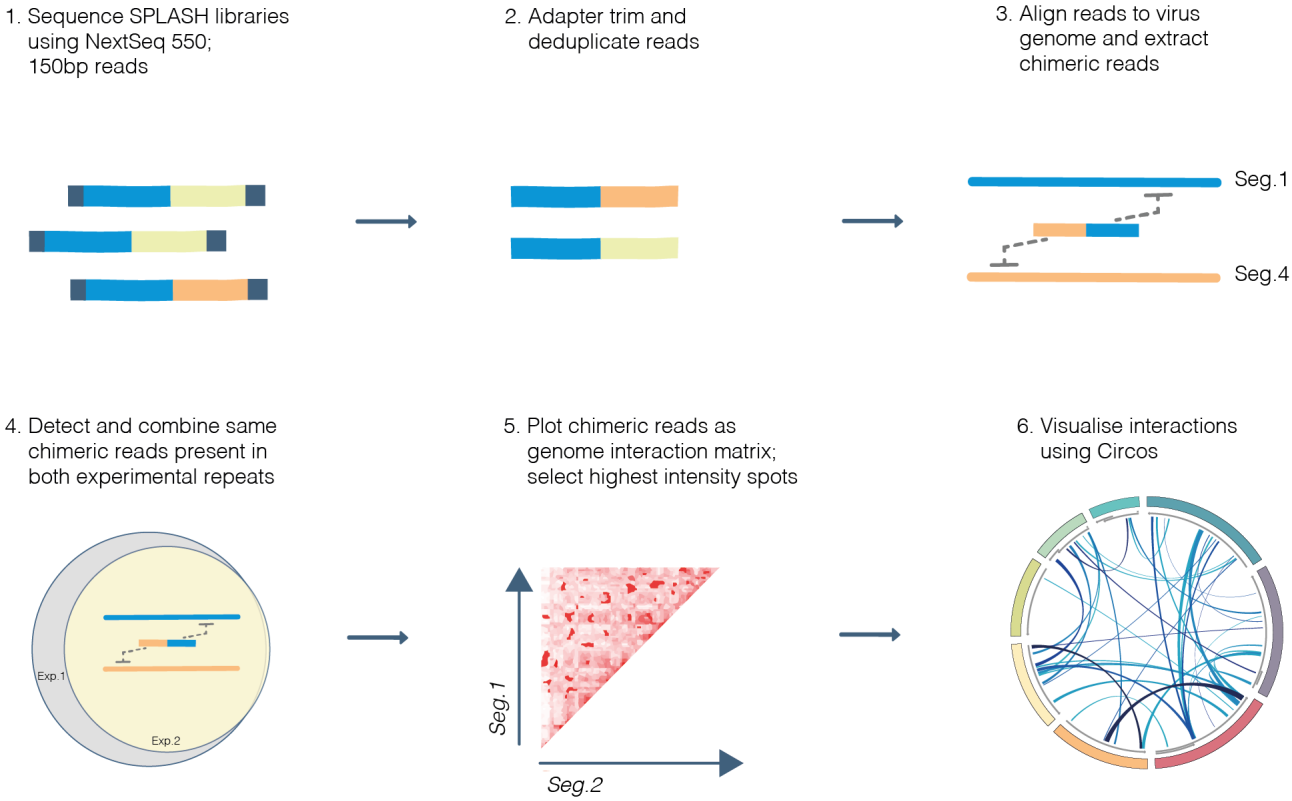


Figure 5.8 SPLASH sequencing and bioinformatic analysis

A

Virus	No. of chimeric reads with ligase	No. of chimeric reads without ligase	Enrichment over background	Overlap (%)
WSN 1	2,894 (2.44)	11 (0.05)	48.80	99.8 *overlap with exp.2
WSN 2	375,070 (2.42)	6,546 (0.07)	34.57	84.1 *overlap with exp.3
WSN 3	135,569 (1.44)	6,286 (0.04)	36.00	97.6 *overlap with exp.2
PR8	257,680 (1.35)	51 (0.02)	67.50	
Udorn	10,233 (0.81)	3,102 (0.1)	8.10	
FLB	1,190 (0.16)	1,191 (0.03)	5.33	

B

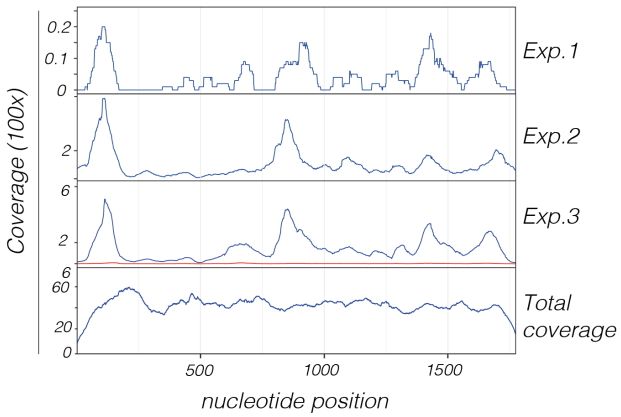


Figure 5.9 Summary of SPLASH data from three experimental replicates

*(**A**) Total number of unique chimeric reads aligning to the virus genome in the presence or absence of ligase during the proximity ligation step. For each sample, the number of chimeric reads as a percentage of the total reads aligning to a virus genome is given in the brackets. Enrichment over background was calculated as percentage of chimeric reads in the presence of ligase over the percentage of chimeric reads in the absence of ligase. Percentage overlap between the WSN sample replicates was calculated by counting reads aligning within 20 nucleotides of their start and end coordinates. (**B**) Comparison of the coverage of the chimeric reads aligning to the reference genome in the three experimental WSN replicates. Total coverage is shown for the sequencing done using the SMARTer smRNA protocol without enrichment for the cross-linked RNA. A representative coverage plot for HA segment is shown.*

Processing of the chimeric reads was done using custom R scripts. The read matrix and link colouring script was written by David VL Bauer, the IntaRNA input and output processing script was written by the author and David VL Bauer, the CIGAR parsing script was written by the author. The scripts used for the SPLASH analysis are provided in Appendix B.

The chimeric read coordinates were extracted by parsing the aligned read CIGAR strings. The read coordinates were then used to draw a read matrix, in which every nucleotide in a read was converted into a pixel (Figure 5.10). The overlapping reads in the matrix formed loci of different pixel intensities. A read threshold for picking less than 30 top interactions in the

sequencing data was set, which was >3500 reads per million (RPM) for WSN and FLB and >3000 RPM for PR8 and Udorn viruses. For all loci that passed the threshold, Gaussian curves through their x and y-axes were drawn and the average peak height of the two distributions was used to measure the intensity of each locus. Locus intensity was considered to be a good approximation of the frequency of the interaction *in virio*. A distance of 1.5x standard deviations of the Gaussian distribution was used to determine the interaction coordinates for each interacting vRNA segment. The coordinates were used to draw Circos plots (Krzywinski, Schein, Birol, *et al.*, 2009), which illustrate the interaction sites between different viral segments. The interaction arcs were coloured by the measured loci peak intensities.

While SPLASH analysis can identify a region involved in an RNA-RNA interaction, it does not model the interaction itself. Therefore, the interaction coordinates estimated from the chimeric read matrix were further processed with an IntaRNA algorithm, which uses sequence hybridisation energy, sequence accessibility and a predefined interaction seed length to model RNA-RNA interactions (Mann, Wright & Backofen, 2017). The interaction seed was set to 7 nucleotides and the previously acquired SHAPE reactivities were also included in modelling.

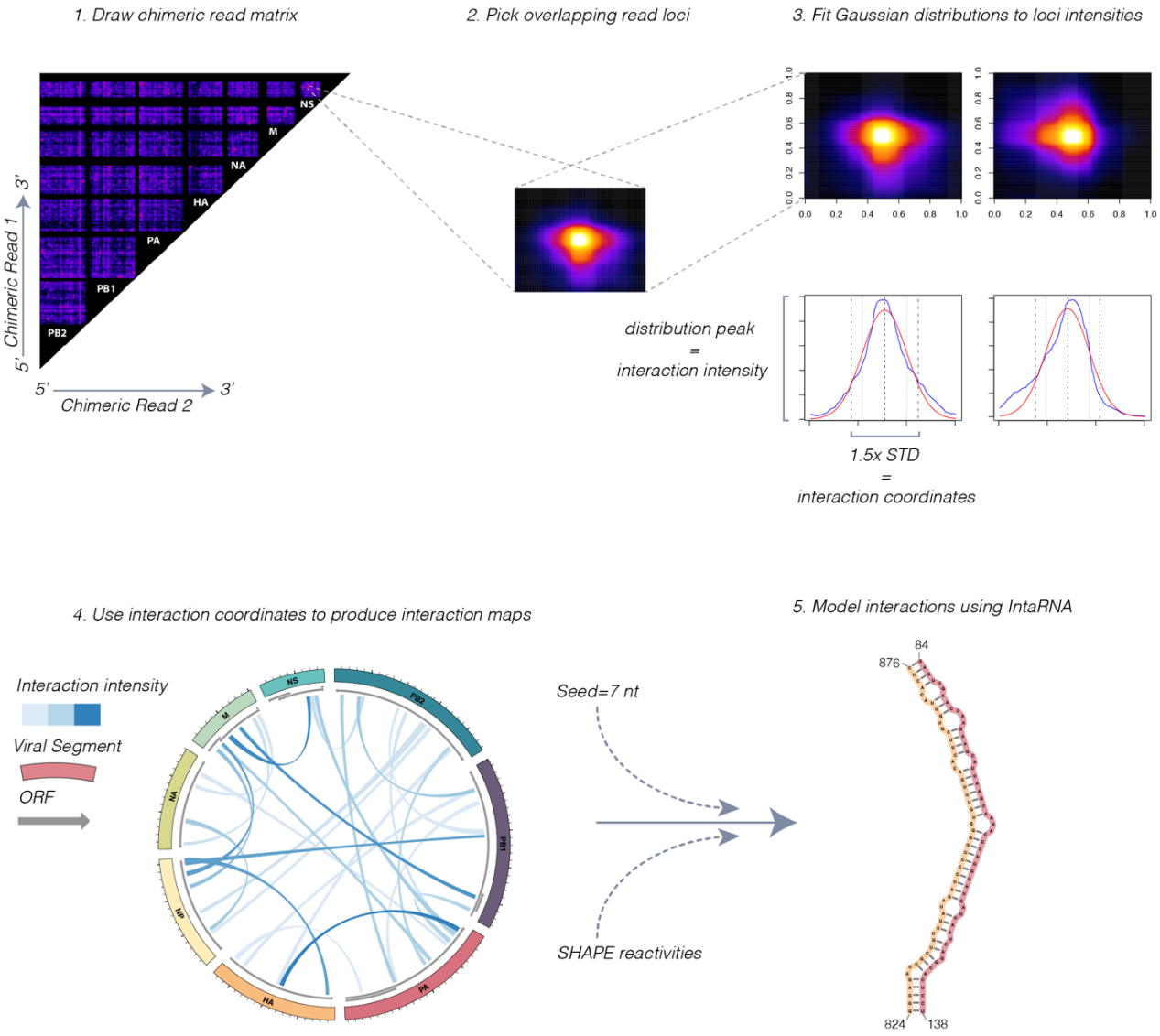


Figure 5.10 Sequencing read analysis and the modelling of RNA interactions

The chimeric reads are used to draw an interaction matrix, from which distinct loci are selected. For each locus a Gaussian curve is drawn to quantitate interaction intensities and coordinates. The coordinates are used to produces Circos maps and to model RNA-RNA interaction structures.

5.4.4 WSN inter-segment RNA interaction map

A total of 26 inter-segment interactions were identified for the WSN virus. The identified interactions are summarized in Table 5.1 and the raw interaction map, i.e. with interactions presented before IntaRNA processing, is given in Figure 5.11.

All genome segments were found to be involved in inter-segment RNA interactions. Most of the interactions were found to stem from the PA segment (10 interactions) followed by the PB2

(9), PB1, NP and M (7), HA (5), NA (4), and NS (3) segments. A number of interactions stemmed from the same loci within a segment (e.g. PB1 1038-1080 interaction with NP 1441-1515 overlapped with HA 96-146 interaction with NP 1474-1556). Interactions were found to be distributed throughout the segments' lengths, although the terminal regions exhibited more interactions than the central region (Figure 5.12).

Table 5.1 Inter-segment RNA-RNA interactions in the WSN virus genome

Interaction ID	Seg.1	Seg.1 start	Seg.1 end	Seg.2	Seg.2 start	Seg.2 end	Intensity
1	PA	1932	1982	HA	1632	1708	3509.88
2	PB1	499	573	PA	1356	1420	3623.68
3	PB2	2028	2104	HA	874	938	3697.32
4	PB1	290	358	HA	1471	1529	3826.73
5	PB2	1053	1005	PB1	948	1032	3855.74
6	PB2	1573	1641	NP	282	338	3862.43
7	PA	395	439	NA	1169	1235	3951.69
8	NA	966	1024	M	923	983	4190.44
9	PB2	1139	1197	PA	195	263	4293.08
10	PB2	1434	1542	NP	646	702	5002.64
11	NA	54	102	M	833	891	5250.32
12	PB2	418	480	NS	635	701	5489.07
13	PB2	2006	2062	PA	837	895	7211.66
14	PB1	2208	2268	PA	307	377	7477.19
15	PB2	95	151	PA	1388	1444	8519.22
16	PA	731	793	M	192	244	8557.15
17	PA	151	235	NS	713	773	8612.94
18	NP	343	395	NA	398	470	9376.05
19	PB2	362	414	PB1	292	350	9953.97
20	NP	1076	1134	M	79	141	10507.34
21	HA	96	146	NP	1474	1556	13512.94
22	PB1	1038	1080	NP	1441	1515	13524.10
23	NP	1274	1346	M	216	260	13731.61
24	PB1	1991	2053	M	579	633	14686.62
25	M	360	436	NS	588	640	15617.09
26	PA	82	152	HA	814	880	16415.90

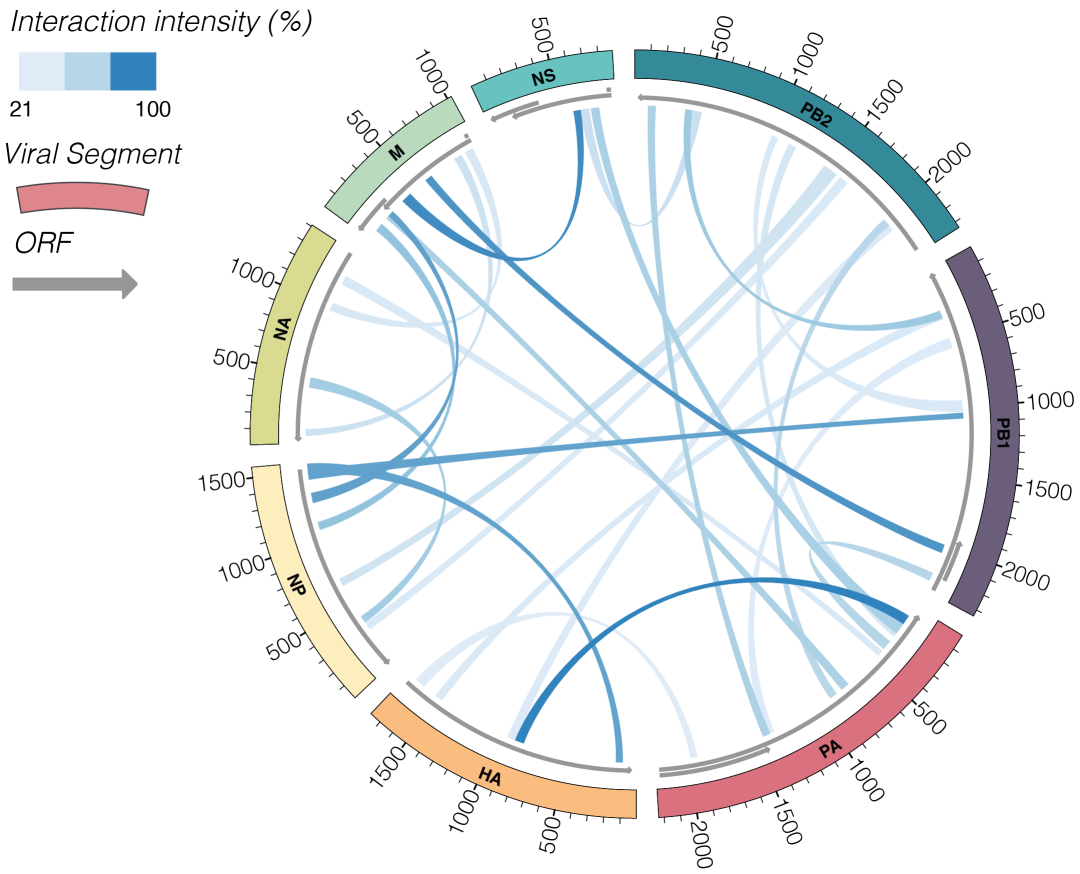


Figure 5.11 WSN inter-segment interaction map

Interactions from Table 6.1 were visualised using Circos plot. Each viral segment is shown in the perimeter of the circle, grey arrows indicate open reading frames (ORF). Links in the middle of the circle indicate SPLASH-identified interactions, which were coloured by interaction intensities as measured during the matrix interaction selection. Colour scale shows the difference in the intensities on a percentage scale. Each viral genome segment is plotted clockwise in a 5′ to 3′ orientation.

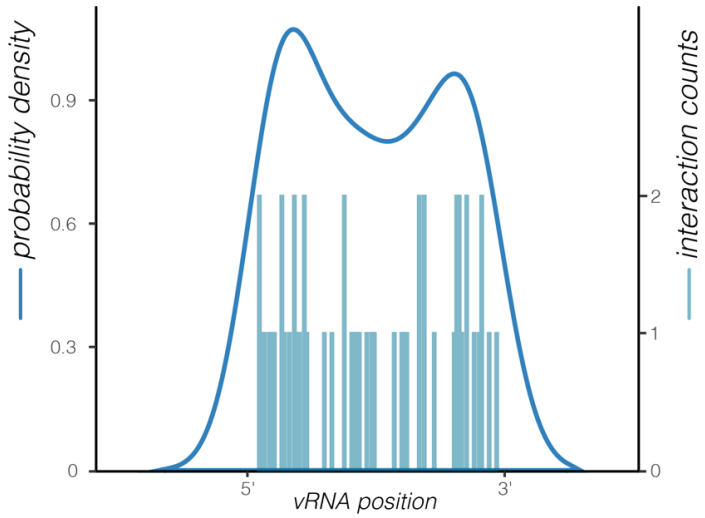


Figure 5.12 Position of the inter-segment RNA interactions along the length of the vRNA

The length of each genome segment was standardised to a length of 1 unit and the interaction coordinates were converted accordingly. The midpoint for each interaction is plotted.

5.4.5 Modelling the structures of the inter-segment RNA interactions

While the SPLASH analysis can identify the regions that are involved in the inter-segment RNA interactions, it does not provide any structural information about the conformation of the RNA structures involved in intermolecular base-pairing. Therefore, IntaRNA algorithm was used next, to assess if the identified regions can form interactions via complementary base-pairing and to more precisely model the interaction structures. Previously acquired SHAPE reactivities and SPLASH analysis-defined inter-segment pair sequences were used as the inputs into the IntaRNA predictions, and a requirement for at least 7 nucleotide complementarity to be present in the interacting region was set. For comparison, modelling using a permutated dataset was also performed, for which the SPLASH-identified interactions were shuffled randomly amongst themselves. A permutated dataset, therefore, preserved the overall sequence properties (e.g. GC nucleotide content) of the dataset and allowed to be assess the specificity of the interacting pairs. The IntaRNA analysis showed that the top SPLASH-identified interactions were all able to form complementary base-pairs, unlike the interactions in the permutated dataset. The modelled structures spanned between 14 and 71 nucleotides and, compared to the permutated dataset, the interactions were enriched for low ΔG energy pairings (Figure 5.13, Table 5.2).

Table 5.2 Inter-segment interaction coordinates after IntaRNA modelling and the associated ΔG energies

Interaction ID	*Seg.1*	*Seg.1 start*	*Seg.1 end*	*Seg.2*	*Seg.2 start*	*Seg.2 end*	*ΔG (kcal/mol)*
1	PA	1943	1975	HA	1651	1683	-27.98
2	PB1	499	569	PA	1357	1413	-22.76
3	PB2	2060	2099	HA	879	909	-22.68
4	PB1	304	343	HA	1475	1517	-26.91
5	PB2	1016	1041	PB1	983	1010	-16.91
6	PB2	1603	1634	NP	289	318	-22.80
7	PA	409	438	NA	1173	1206	-23.81
8	NA	985	1021	M	925	961	-24.41
9	PB2	1140	1191	PA	206	262	-6.94
10	PB2	1442	1491	NP	656	701	-19.99
11	NA	66	80	M	849	863	-8.01
12	PB2	437	480	NS	639	674	-21.58
13	PB2	2026	2040	PA	860	873	-18.31
14	PB1	2221	2253	PA	329	357	-22.82
15	PB2	116	133	PA	1405	1422	-20.18
16	PA	732	764	M	210	243	-23.05
17	PA	188	225	NS	714	742	-17.80
18	NP	359	390	NA	401	434	-17.78
19	PB2	380	393	PB1	314	328	-15.84
20	NP	1098	1125	M	91	116	-21.66
21	HA	112	145	NP	1479	1515	-13.85
22	PB1	1049	1072	NP	1452	1475	-12.41
23	NP	1298	1313	M	226	243	-18.30
24	PB1	2003	2046	M	582	618	-25.53
25	M	382	426	NS	593	637	-22.92
26	PA	84	138	HA	824	876	-36.23

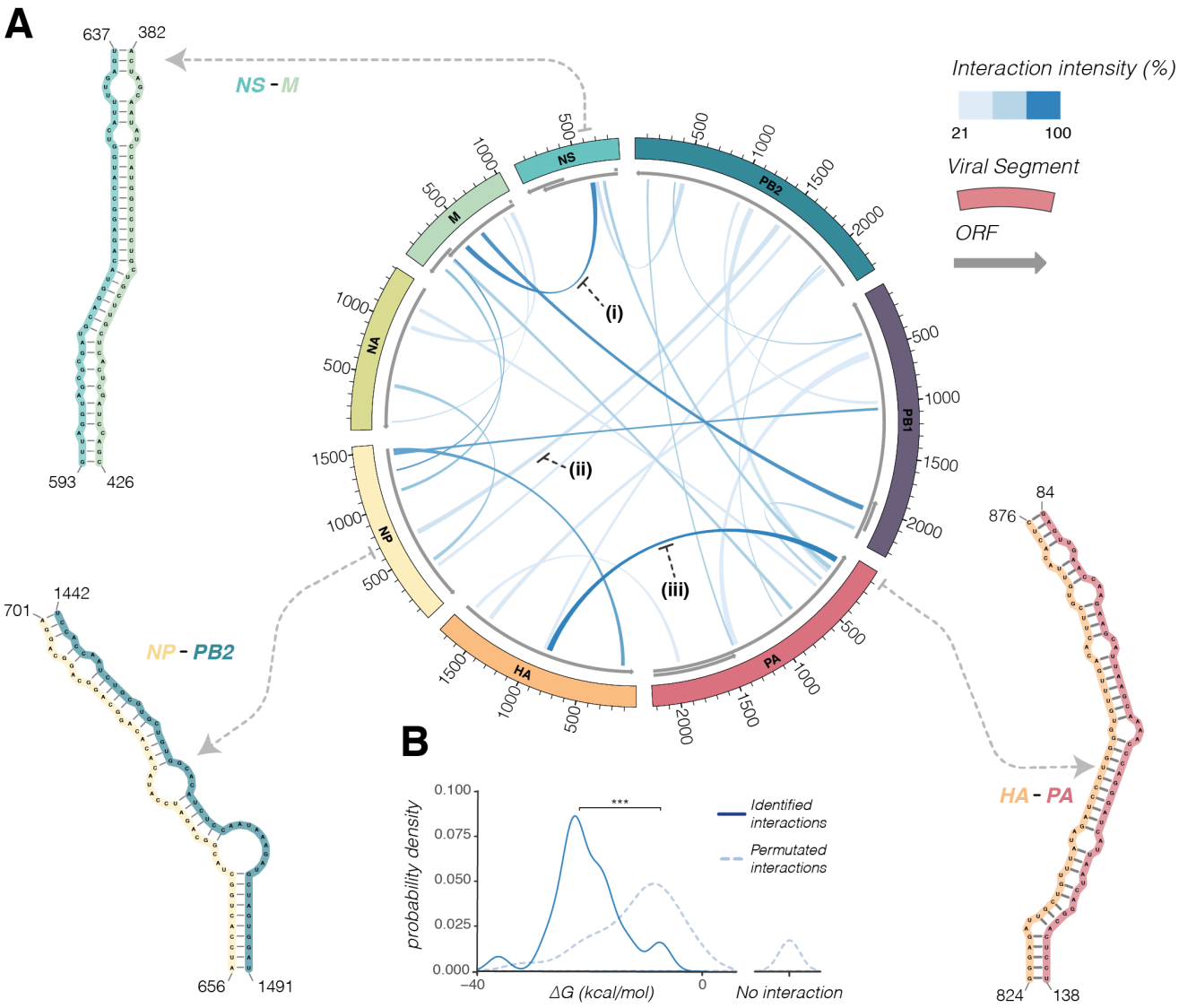


Figure 5.13 Modelled inter-segment RNA interactions

*(**A**) Inter-segment interaction structures were modelled using IntaRNA algorithm with SHAPE reactivities taken into consideration. Three representative structures are shown. (**B**) The distribution of base-pairing ΔG energies in the SPLASH-identified versus the permutated dataset interactions. ***P<0.00001, Wilcox Rank-Sum Test.*

5.4.6 Inter-segment RNA interaction analysis for PR8, Udorn and FLB viruses

A total of 29, 28 and 25 inter-segment interactions were identified for the PR8, Udorn and FLB viruses, respectively. The lists of all top interactions for the three viruses after matrix selection and IntaRNA processing are provided in Table 5.3-Table 5.5**.**

Table 5.3 Inter-segment RNA-RNA interactions in the PR8 virus genome

Interaction ID	Seg.1	Seg.1 start	Seg.1 end	Seg.2	Seg.2 start	Seg.2 end	Intensity	ΔG (kcal/mol)
1	PB2	1556	1641	M	484	567	3039.84	-30.49
2	PB2	1549	1642	PB1	189	240	3140.66	-22.32
3	PB2	1461	1570	NA	586	643	3145.70	-20.03
4	PB1	2198	2271	HA	1187	1278	3216.28	-27.20
5	PA	146	237	NS	703	774	3312.06	-18.82
6	PA	745	824	NP	782	857	3347.35	-15.91
7	PB1	995	1116	NP	1423	1518	3473.38	-19.81
8	PB1	2195	2278	PA	302	381	3755.68	-33.69
9	NP	322	415	NS	584	645	3922.04	-22.80
10	PA	2052	2153	M	196	263	4002.70	-15.98
11	PA	2086	2169	NA	805	900	4027.91	-20.46
12	PA	2013	2078	NS	37	118	4027.91	-22.38
13	PA	143	208	NA	301	382	4279.97	-18.77
14	PB1	645	740	HA	972	1039	4395.91	-30.97
15	NA	42	109	M	815	894	4451.37	-9.20
16	NP	676	753	M	425	564	4738.72	-33.55
17	PA	1425	1510	NS	660	741	4768.96	-37.43
18	PB1	339	426	HA	862	939	5081.52	-21.43
19	PB2	85	162	PA	1381	1452	5414.23	-34.88
20	M	349	450	NS	584	647	5706.62	-28.75
21	PB2	1057	1160	PA	1381	1444	5817.53	-30.42
22	HA	687	786	NA	559	638	6301.48	-12.44
23	HA	630	707	M	908	997	6351.89	-13.00
24	PB2	917	1006	PA	57	118	6735.03	-17.60
25	PB2	1127	1210	PA	181	276	7168.57	-32.40
26	PA	1537	1592	NA	847	944	8489.36	-25.93
27	PB2	914	1009	NP	975	1046	8922.90	-22.44
28	PB1	2196	2275	M	93	170	9270.74	-30.38
29	PB2	1535	1628	HA	793	838	9290.91	-10.67

Table 5.4 Inter-segment RNA-RNA interactions in the Udorn virus genome

Interaction ID	Seg.1	Seg.1 start	Seg.1 end	Seg.2	Seg.2 start	Seg.2 end	Intensity	ΔG (kcal/mol)
1	PB1	117	224	NS	661	578	2149.91	-6.20
2	NP	559	626	NS	118	185	3224.86	-12.47
3	HA	1361	1446	NP	668	755	3322.58	-24.36
4	HA	95	194	NA	703	804	3322.58	-23.85
5	PB1	2038	2143	NA	715	808	3420.31	-28.04
6	PA	83	178	HA	1308	1391	3615.75	-8.19
7	PB2	2082	2167	NA	540	623	3615.75	-27.97
8	NP	227	318	NS	469	578	3615.75	-27.60
9	PB2	411	486	HA	1235	1328	3713.48	-33.10
10	HA	373	452	NP	1020	1107	3811.20	-31.37
11	HA	1239	1308	M	896	993	3811.20	-23.51
12	PB1	310	425	M	365	454	3908.92	-22.32
13	NP	231	306	NA	751	850	4006.65	-26.13
14	PB1	116	223	NS	579	660	4299.81	-17.62
15	HA	1047	1148	NP	954	1057	4397.54	-32.35
16	M	363	446	NS	582	647	4495.26	-20.06
17	HA	569	650	NP	652	727	4886.15	-24.22
18	NA	1021	1086	M	582	669	4886.15	-13.18
19	NA	1004	1131	M	905	988	5277.04	-29.42
20	HA	1050	1145	NA	1260	1333	6352.00	-26.76
21	PB2	405	496	PA	1439	1518	6547.44	-44.64
22	PB2	399	476	NA	1036	1137	6645.17	-30.46
23	PB2	1469	1570	PB1	1323	1400	6938.34	-28.34
24	HA	58	137	NA	928	1021	8795.07	-24.72
25	PB1	277	358	NA	865	956	9381.41	-32.55
26	HA	54	135	NP	1076	1153	10065.47	-30.93
27	PB2	1335	1406	PA	812	893	12410.83	-15.64
28	HA	1452	1545	NP	259	340	13290.34	-27.19

Table 5.5 Inter-segment RNA-RNA interactions in the FLB virus genome

Interaction ID	*Seg.1*	*Seg.1 start*	*Seg.1 end*	*Seg.2*	*Seg.2 start*	*Seg.2 end*	*Intensity*	*ΔG (kcal/mol)*
1	PB2	1074	1130	PB1	2238	2284	5882.35	-4.32
2	PB2	1242	1296	PB1	2242	2290	4201.68	-13.19
3	PB1	1883	1955	PA	275	305	5042.02	-13.84
4	PB1	553	597	HA	353	427	5042.02	-17.06
5	PB1	1818	1856	HA	791	825	5882.35	-20.25
6	PB1	390	428	HA	1559	1617	6722.69	-6.02
8	PB2	2273	2309	NP	49	87	4201.68	-16.52
9	HA	359	411	NP	100	156	7563.03	-14.28
10	PB1	573	609	NP	144	178	10924.37	-19.84
11	PB2	782	850	NP	149	187	5042.02	-16.22
12	PB1	1003	1073	NP	140	182	6722.69	-18.02
13	PB2	783	851	NP	189	147	4201.68	-16.21
14	PB1	2243	2289	NP	402	466	12605.04	-20.47
15	HA	1149	1191	NP	981	1097	6722.69	-17.58
16	PA	1128	1180	NP	1623	1653	4201.68	-17.50
17	PB1	1858	1898	NP	1643	1687	8403.36	-23.73
18	HA	767	689	NP	1652	1708	1680.67	-20.55
19	NP	1647	1691	NA	89	171	4201.68	-23.09
20	PB2	459	501	NA	180	218	4201.68	-15.01
21	PA	1559	1611	NA	1175	1223	4201.68	-23.40
22	PA	499	555	NA	1530	1558	5882.35	-14.26
23	PB2	959	1019	NS	449	497	7563.03	-18.06
24	M	763	799	NS	1013	1063	12605.04	-19.58
25	PA	567	607	NS	1034	1094	5042.02	-2.37

SPLASH analysis for the three viruses indicated that all genome segments form inter-segment RNA interactions, though the number of interactions formed by each segment varies within (independently of the segment's length itself) and between the different strains (Figure 5.14-Figure 5.16). The interactions are also not limited to a specific region of a segment but are distributed throughout the segments' length. Comparison of the ΔG energies between the SPLASH-identified and permutated pairs, showed that in all the viruses tested, the SPLASH-identified interactions have significantly lower ΔG energies, indicating that the intermolecular pairings are not only specific but also can form extended double stranded regions (Figure 5.14.B-Figure 5.16.B).

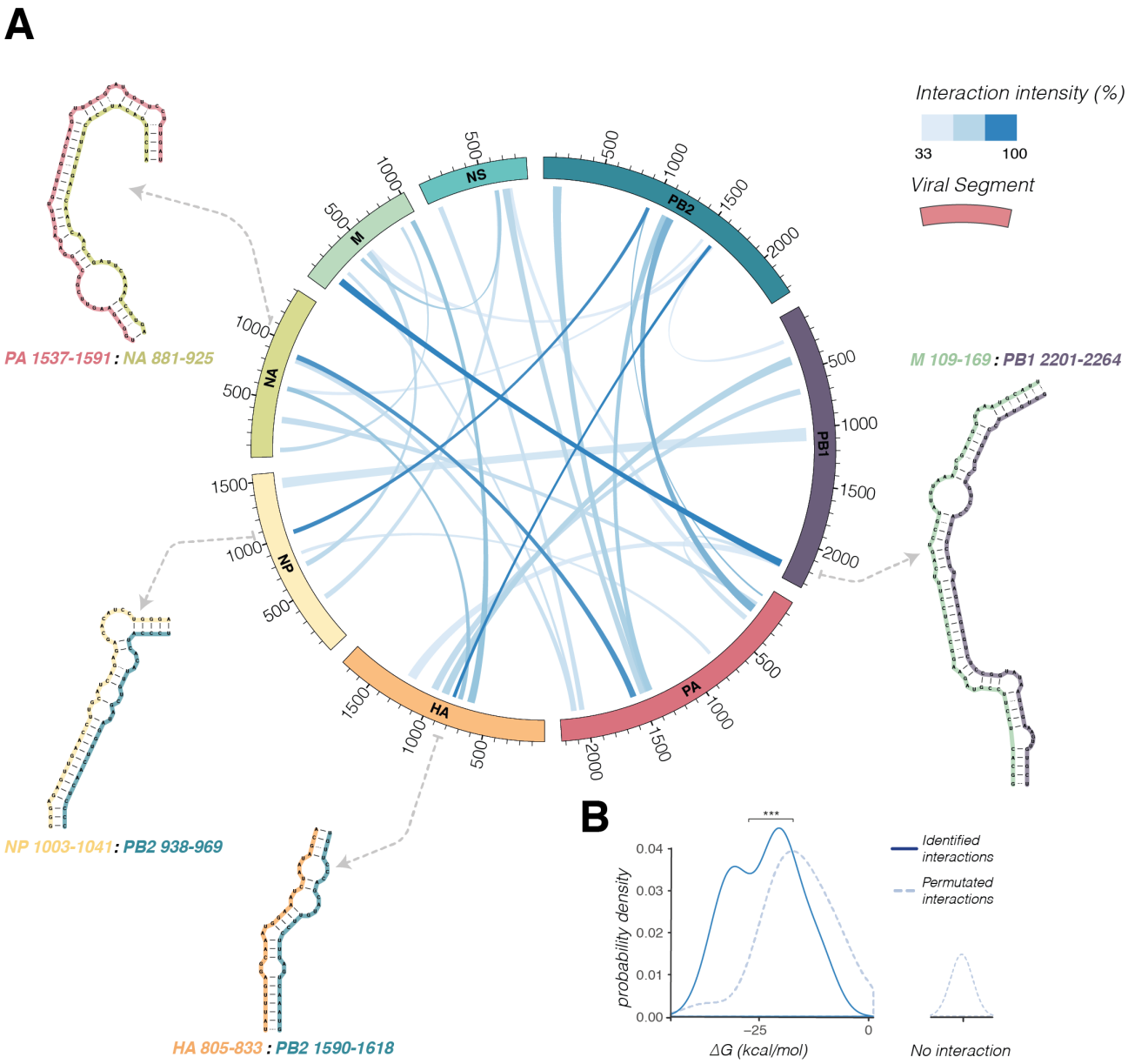


Figure 5.14 Inter-segment RNA interactions in the PR8 virus genome

*(**A**) A map of the most common RNA-RNA interactions in the PR8 genome identified using SPLASH. Four interactions, associated with the highest intensities in the chimeric read matrix, are shown. (**B**) Probability distributions of the ΔG energies associated with the SPLASH-identified versus permutated interactions. ***P<0.00001, Wilcox Rank-Sum Test.*

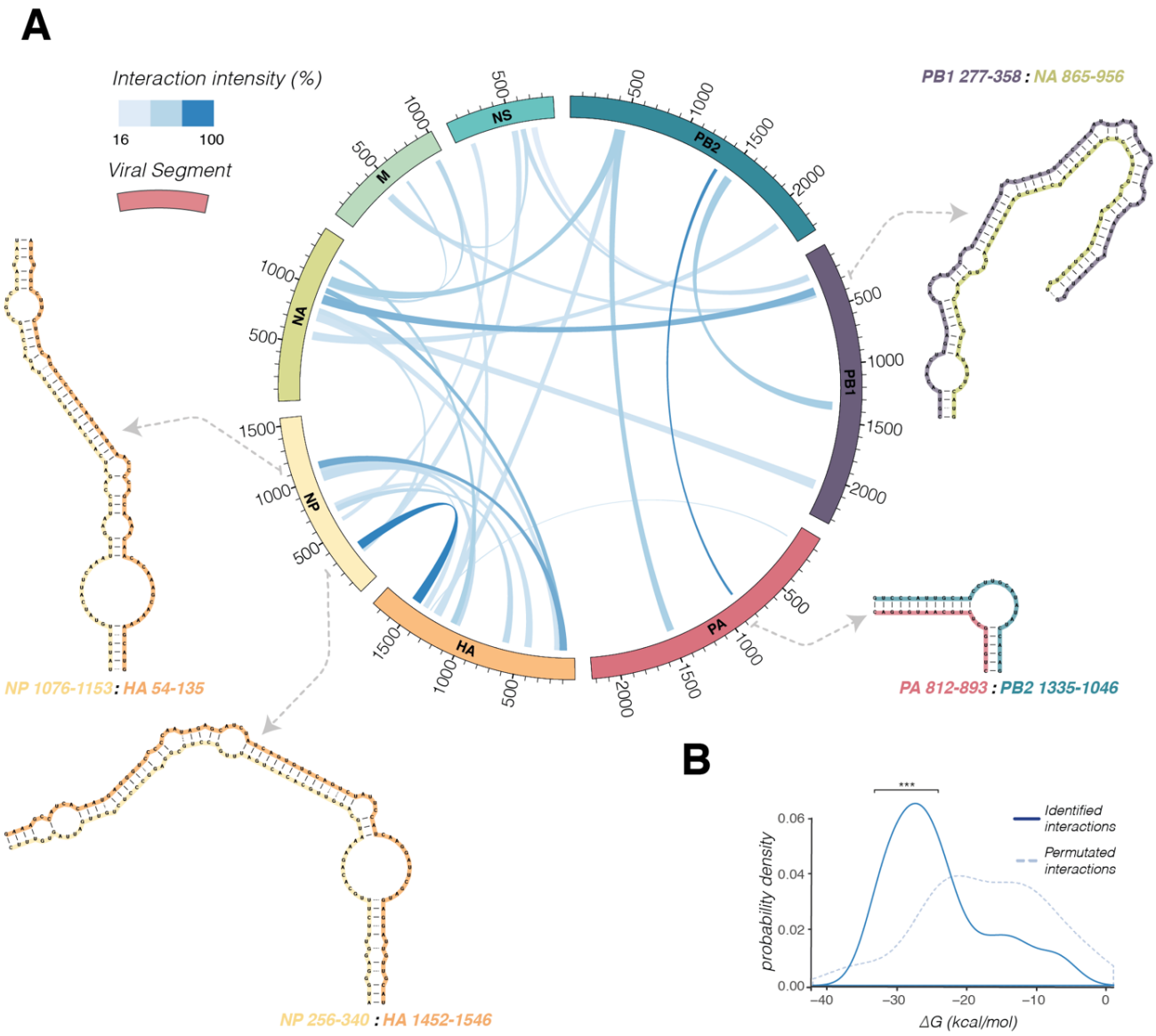


Figure 5.15 Inter-segment RNA interactions in the Udorn virus genome

*(**A**) A map of the most common RNA-RNA interactions in the Udorn genome identified using SPLASH. Four interactions, associated with the highest intensities in the chimeric read matrix, are shown. (**B**) Probability distributions of the ΔG energies associated with the SPLASH-identified versus permutated interactions. ***P<0.00001, Wilcox Rank-Sum Test.*

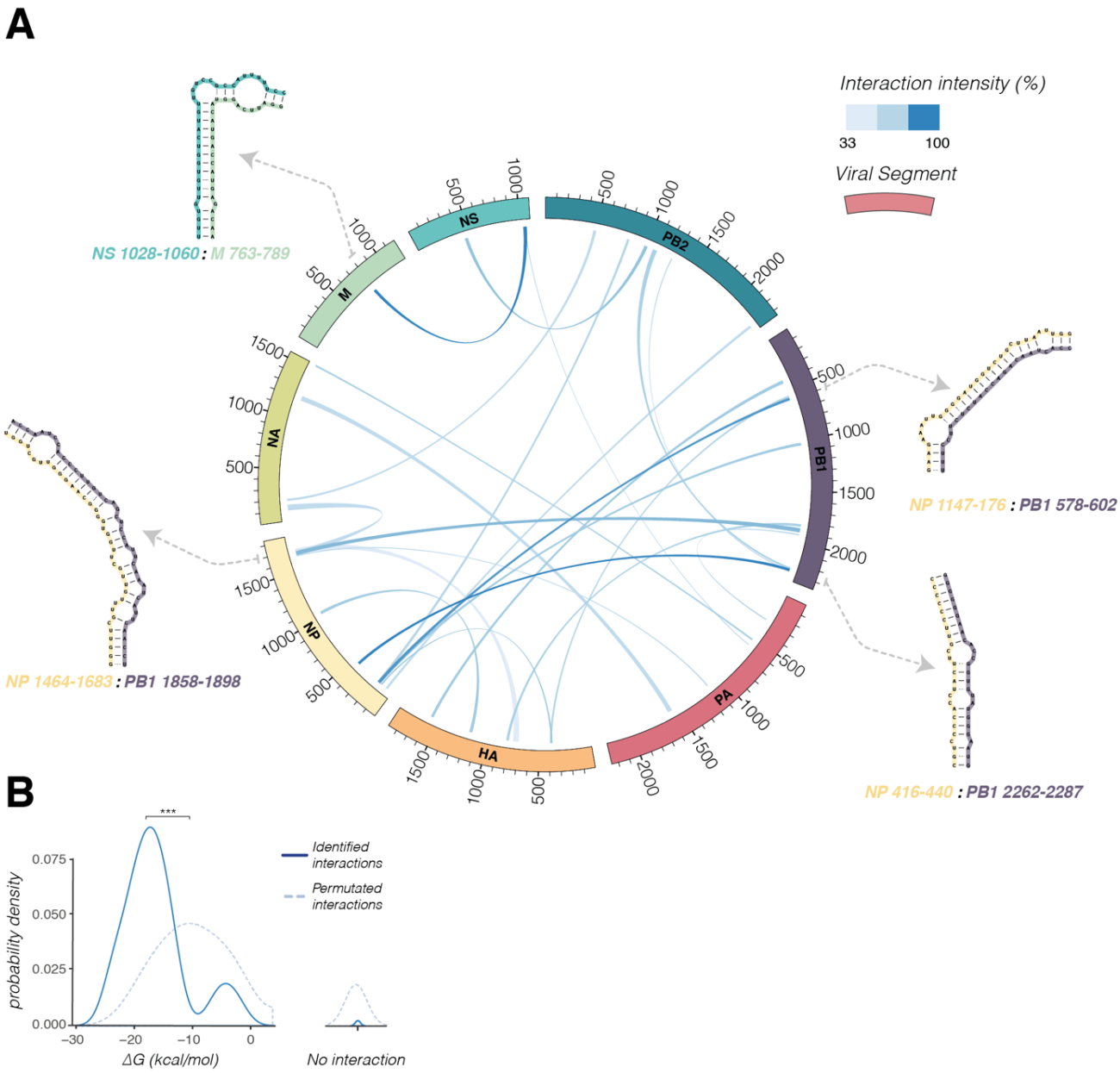


Figure 5.16 Inter-segment RNA interactions in the FLB virus genome

*(**A**) A map of the most common RNA-RNA interactions in the FLB genome identified using SPLASH. Four interactions, associated with the highest intensities in the chimeric read matrix, are shown. (**B**) Probability distributions of the ΔG energies associated with the SPLASH-identified versus permutated interactions. ***P<0.00001, Wilcox Rank-Sum Test.*

5.5 Discussion

This chapter presented the use of a modified SPLASH protocol to detect inter-segment RNA interactions in influenza viruses. SPLASH uses AMT-B, a biotinylated-derivative of psoralen, to cross-link base-pairing RNAs. Two other methods, LIGR-seq and PARIS, were published at the same time as the SPLASH method, which also use psoralen cross-linking to detect RNA-RNA

interactions. SPLASH was chosen because it uses biotin pull-down rather than RNase digestion or 2D-gel selection, to enrich for the cross-linked RNA species, which was thought to be a more practical and stringent approach. The disadvantage of the use of AMT-B crosslinking is its low efficiency; only between 1.4 and 2.4 % of the total mapped viral reads were found to be chimeric in this analysis (this may explain the lack of shifted bands observed in the RNA migration analysis in Figure 5.3). On the other hand, compared with the no ligase control samples, even low cross-linking rates far exceed any background or false-positive detection signal, suggesting that, given a good sequencing depth, low level of RNA cross-linking was not a limiting step.

It is worth noting that a significant proportion of sequencing reads mapped to non-viral RNAs, in particular various forms of ribosomal RNA (rRNA). This is perhaps not surprising, as the SPLASH protocol functions to enrich for highly structured RNAs, the best example of which are rRNAs. In fact, SPLASH was originally used to map interactions between rRNAs and small nucleolar RNAs (Aw, Shen, Wilm, *et al.*, 2016). Generally, in sequencing library preparation, antisense oligonucleotides are used to deplete the rRNA from the samples, a step that was not performed in the analysis presented here. It is known that rRNA is present in the influenza virus preparations, although it is not clear if it is packaged inside the virion or simply separates into the same fraction as the viral particles do during the ultracentrifugation step (Hutchinson, Charles, Hester, *et al.*, 2014; Noda, Murakami, Nakatsu, *et al.*, 2018). In the future, rRNA depletion should be performed, either by treating the purified virions with RNase or using an antisense-oligo-depletion method after the RNA cross-linking. In addition, a more stringent virion purification, such as using haemadsorption method (Hutchinson, Charles, Hester, *et al.*, 2014), could be performed.

The bioinformatical mapping of the viral segment interactions was done using a matrix approach, which is often used to analyse genome-wide chromosome arrangements in cells. The intensity of each detected locus in the interaction matrix was used to approximate the frequencies of the detected interactions. While, the locus intensity will be determined by the number of overlapping reads, which in turn can be used to estimate the frequency of a given interaction, it is worth noting that an assumption of equal AMT-B crosslinking efficiency across the viral genome has been made here. This assumption, however, does not always hold true. In particular, as the psoralen primarily cross-links pyrimidines, regions of the genome enriched for pyrimidines can be overrepresented. Indeed, SPLASH-identified regions forming inter-segment interactions have

54.1% pyrimidine content as opposed to the total WSN genome sequence which has only 42.9% of uracil and cytosine nucleotides in it, indicating the cross-linking preference for pyrimidine-containing regions. In the future an improved analysis protocol could be written that, for example, adjusts the intensities based on how many potential cross-linking sites are found in an interaction region.

It is interesting to note that some of the interactions identified by SPLASH may exist in different conformations. When the highest intensity loci were picked from the interaction matrix, not all loci were symmetrical (Figure 5.17). A symmetrical locus would generally be expected to form in the matrix, given a sufficient sequencing coverage, however even with a good coverage some loci were found to have significantly skewed intensities or even multiple intensity peaks. The skewed intensity distribution could be a consequence of biases in the library preparation, for example, a poor crosslink reversal at a specific position would cause the chimeric read to always terminate in one position. By contrast, loci with multiple intensity peaks may be indicative of the same interaction shifting between several structural conformations.

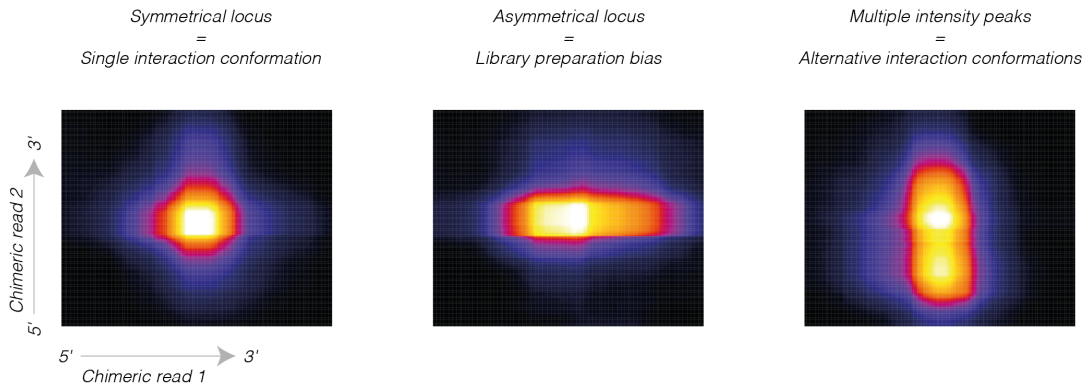


Figure 5.17 Different shapes of interaction loci

Selection of alternative interaction loci shapes, which may be indicative of library preparation biases or alternative inter-segment interaction conformations.

The analysis of the inter-segment RNA-RNA interactions present in the genomes of different influenza viruses showed that a redundant network of inter-segment interactions exists *in virio*. The work presented here focused only on the most common interactions identified, which reflect the overall pool of interactions present in a purified virus population. It is likely that only a subset of these interactions will be present in any individual virions, with interactions forming in

overlapping regions being mutually exclusive. Furthermore, some interactions may also position RNPs in a three-dimensional space in an arrangement that allows only for a subset of them to occur. The current understanding of how the RNPs are positioned inside the virions is limited. Scanning transmission EM analysis of RNPs inside budding virions has generally observed that RNPs are positioned in parallel next to each other (Calder, Wasilewski, Berriman, *et al.*, 2010; Noda, Sugita, Aoyama, *et al.*, 2012). Furthermore, analysis of RNP-associated polymerase has found that, inside a virion, the polymerase can be positioned at either end of the virus particle (Sugita, Sagara, Noda, *et al.*, 2013), suggesting that some RNPs may be packaged polymerase end first, while others the RNP loop-end first. The SPLASH data would be consistent with these observations; for example, a strong interaction is present between the 3' end of NP and the middle of the PB1 segment in both WSN and PR8 viruses (Figure 5.13), suggesting that these RNPs may be arranged in an opposite direction relative to each other.

While the previous studies on genome packaging have generally focused on the terminal ends of segments to identify the genome regions important in the packaging of RNPs, SPLASH data indicated that the inter-segment interactions can be formed within any part of a segment. It is well established that non-viral sequences, such as GFP reporter-coding RNA, can be packaged into influenza virions by addition of the segment-specific non-coding regions and segment bundling signals of various lengths (Fujii, Goto, Watanabe, *et al.*, 2003; Fujii, Fujii, Noda, *et al.*, 2005). None of these foreign gene-coding constructs though are packaged as well as the native gene segments are; viruses often lose the reporter activity after several passages and the bundling signals of the most efficient constructs usually include at least some of the interactions identified by SPLASH.

Interestingly, the positioning of the interactions in the three-dimensional space might be important as well. For example, a stable WSN PA-NLuc reporter virus, which contains a luciferase reporter sequence fused to the 5' end of the PA gene (Tran, Moser, Poole, *et al.*, 2013), requires the 5'-UTR and the first 50 nucleotides of the PA-coding region to be positioned in front of the NLuc reading frame. Such construct effectively duplicates the major interaction region in PA identified by SPLASH and preserves its position as in the wild-type gene.

Previous sequence conservation analysis has shown that the ends of the viral segments contain the most conserved sequences between different influenza virus strains (Gog, Afonso, Dalton, *et al.*, 2007). Considering that many SPLASH interactions lie outside these most conserved regions of RNA, one would predict that many of the inter-segment interactions will be unique for different influenza viruses. Indeed, SPLASH analysis showed that even for closely related viruses, such as WSN and PR8, the interactions formed between the different genome segments vary significantly. Only seven of the top SPLASH-identified interactions were found to be common between the WSN and PR8 viruses and only a single interaction, formed between M and NS segments, was found to be common among all tested IAVs (Figure 5.18 **A**). Interestingly, while the specific interaction partners do not appear to be conserved among the different IAV strains, a number of regions from which the interactions are stemming from are the same in all of them (Figure 5.18 B). The latter observation suggests, that certain regions in the influenza virus genome are more likely to be involved in the inter-segment interaction formation, though the exact partners that the interactions are formed between will be influenced by small sequence variations, which will favour some segment pairings over others.

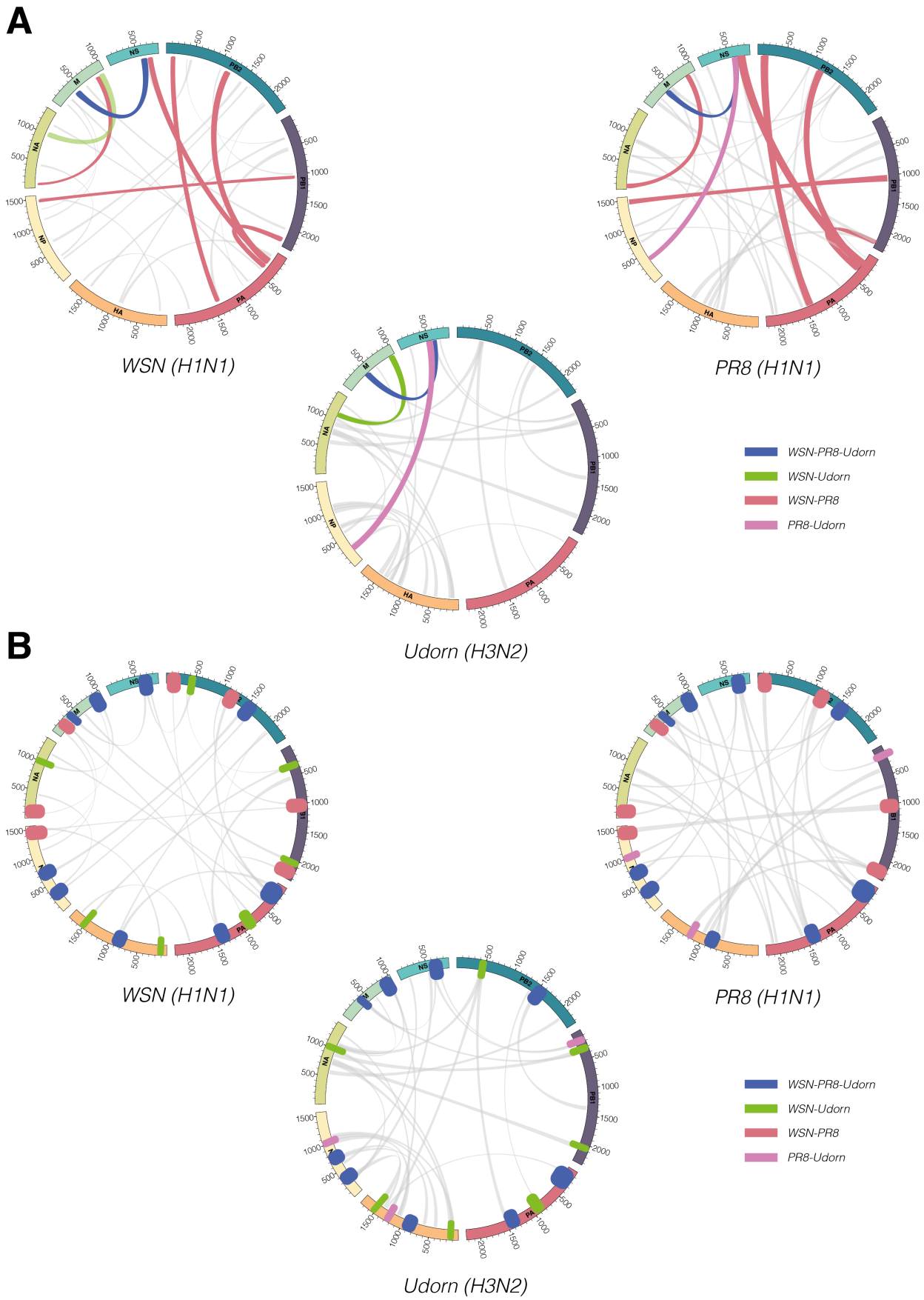


Figure 5.18 Common inter-segment RNA interactions and interaction loci in different IAVs

*(**A**) Common RNA-RNA interactions in different IAV strains. Thickness of the highlighted interactions is exaggerated for contrast. (**B**) Common inter-segment interaction loci in different IAV strains.*

Influenza viruses undergo a constant sequence drift due to the mutations introduced into their genome by an error prone viral polymerase. Interestingly, an extensive redundant network of RNA-RNA interactions, rather than a limited set of few interactions, could be seen as an evolutionary strategy for the virus to maintain selective packaging of its genome, while still accommodating the changes caused by the genetic drift. Point mutations occurring in the inter-segment interaction regions could be compensated by the ability to form alternative interaction sites and therefore, preserving segment assembly. It would be interesting to know if the dominance of different interactions changes overtime and if the virus fitness changes accordingly.

Future experiments will need to be done to determine if the common interactions identified between the different strains can drive reassortment between these viruses. It is worth noting that one of the most dominant interaction in the Udorn virus, PB1 277-358:NA 865-956, falls in to the same NA region which was previously shown to be responsible for the co-segregation of PB1 and NA segments during H3N2 vaccine development (Gilbertson, Zheng, Gerber, *et al.*, 2016; Cobbin, Ong, Verity, *et al.*, 2014; Cobbin, Verity, Gilbertson, *et al.*, 2013). Personal communications indicate that the PB1 region involved in co-segregation, which has now also been mapped, although using different experimental approaches, matches the site identified by SPLASH analysis in this study.

5.6 Summary

- AMT was successfully used to cross-link vRNA *in virio*
- An optimized SPLASH protocol has been established for high-throughput detection of inter-segment interactions in influenza virions
- A bioinformatical pipeline was developed to analyse SPLASH sequencing data
- SPLASH method showed high reproducibility across multiple independent replicates
- SPLASH analysis shows that in influenza viruses a redundant a network of inter-segment interaction exists *in virio*
- Inter-segment interactions are found throughout the lengths' of vRNAs
- While the major inter-segment RNA interaction pairs differ significantly between influenza strains, the loci from which the interactions originate are often preserve

6. Chapter 6: Analysis of the importance of the inter-segment RNA interactions in the influenza WSN virus genome

6.1 Overview

The segmentation of the influenza virus genome poses a challenge in the need to package all eight different genome segments into the budding virions. While the process behind the selective packaging mechanism is still poorly understood, RNA-RNA interactions between the RNPs have been implicated (Gavazzi, Yver, Isel, *et al.*, 2013; Fournier, Moules, Essere, *et al.*, 2012a; 2012b; Eisfeld, Neumann & Kawaoka, 2014; Hutchinson, Kirchbach, Gog, *et al.*, 2010). This chapter discusses the effects of the perturbation of the RNA-RNA interaction network described in chapter 5 on WSN virus growth. Mutagenesis experiments showed that disruption of the RNA interaction network led to attenuated virus growth and defects in the packaging of the vRNAs. Furthermore, restoring the RNA-RNA interactions by complementary mutagenesis was able to reverse the genome packaging defects.

6.2 Background

RNA-RNA interactions between different vRNPs have been proposed to play a major role in the packaging of the influenza virus genome (Gavazzi, Yver, Isel, *et al.*, 2013; Fournier, Moules, Essere, *et al.*, 2012a; 2012b; Eisfeld, Neumann & Kawaoka, 2014). Influenza virus packages eight different vRNPs into each viral particle, though how and where this process occurs is not well understood (Hutchinson, Kirchbach, Gog, *et al.*, 2010; Gerber, Isel, Moules, *et al.*, 2014). Previous studies have suggested that, after being exported from the nucleus, the vRNPs start colocalizing with each other in the cytoplasm, though no clear preference for colocalization between specific segments, or a step-by-step process of assembling all eight vRNPs into a budding virus has been observed thus far (Hutchinson & Fodor, 2013). Indeed, it seems that the process of vRNP assembly for packaging is quite dynamic, with the vRNPs both fusing and coming apart during cytoplasmic trafficking (Lakdawala, Wu, Wawrzusin, *et al.*, 2014).

Multiple studies have shown that the packaging of vRNPs is a selective process, whereby only a single RNP for each unique vRNA is packaged (Chou, Vafabakhsh, Doğanay, *et al.*, 2012; Fujii, Goto, Watanabe, *et al.*, 2003; Duhaut & Mccauley, 1996; Odagiri & Tashiro, 1997). As the only difference between the vRNPs, known so far, is the vRNA they are associated with, it has been proposed that specific RNA-RNA interactions mediate assembly of the eight vRNP complexes. Indeed, a number of studies have identified mutations in various genome segments, which lead to genome packaging defects, although no common genome packaging signal has been identified yet (Hutchinson, Wise, Kudryavtseva, *et al.*, 2009; Hutchinson, Curran, Read, *et al.*, 2008; Liang, Huang, Ly, *et al.*, 2008). Most mutations, which lead to packaging defects, have been mapped to the terminal regions of vRNAs but are not restricted to them. It has been proposed that compatibility in packaging signals may be important for reassortment of different influenza strains (Gerber, Isel, Moules, *et al.*, 2014). So far, many of the identified packaging signals appear to be strain specific, though segment terminal regions show the highest sequence conservation and perhaps could contain packaging signals, which would be common among many viral strains (Gog, Afonso, Dalton, *et al.*, 2007).

Scanning transmission EM studies of budding virions typically observe the vRNPs to be arranged in a 7+1 conformation, with one central segment surrounded by seven others. It is not clear which of the RNPs is positioned at the center of the 7+1 complex and, based on the measured RNP lengths, different studies have observed contrasting arrangements (Fournier, Moules, Essere, *et al.*, 2012a; Noda, Sugita, Aoyama, *et al.*, 2012). The 7+1 conformation has also led to a suggestion of two potential vRNP assembly models: the 'daisy chain model', where one vRNP interacts with a second vRNP, which in turn interacts with a third, and so on and the 'master segment model', in which a central segment interacts with all other segments and mediates the assembly of the 7+1 complex. Surprisingly, it has recently been shown that influenza C and D viruses also package eight RNPs in a 7+1 conformation (Nakatsu, Murakami, Shindo, *et al.*, 2018), even though they only contain 7 genome segments. The latter observation may suggest that the 7+1 arrangement may not be related to the selectivity of genome packaging but may be an outcome of a more fundamental process of genome incorporation into budding virions.

6.3 Aims

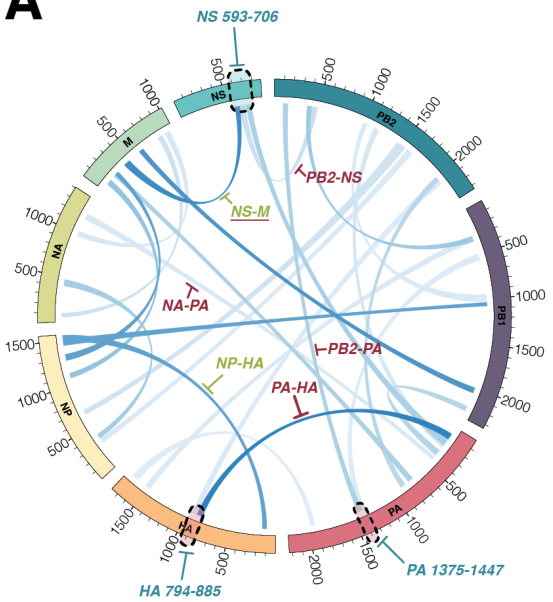
- Design viruses with disrupted inter-segment RNA interactions
- Produce viruses with disrupted inter-segment RNA interactions and assess their growth and genome packaging ability

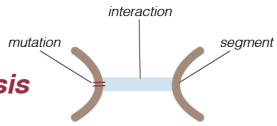
6.4 Results

6.4.1 Design of inter-segment interaction mutants

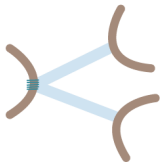
To test if inter-segment RNA interactions are important for virus replication, viruses with synonymous mutations in the regions involved in the inter-segment RNA interactions were made. A summary of the different viruses made is provided in Figure 6.1. Mutants were designed by: introducing mutations into one of the interacting segments (Figure 6.2), introducing mutations into both interacting segments (Figure 6.3), introducing a large number of mutations covering the entire region where overlapping interactions were found to be located (Figure 6.5), or introducing mutations into both interacting segments, designed in such way as to restore the interaction when both mutated segments are present in a virus (Figure 6.6). The experiments in Figure 6.2 and Figure 6.3 were performed before the 2nd and 3rd replicates for the WSN SPLASH experiments were available and for these experiments, the 12-plasmid virus rescue system was used. Experiments in Figure 6.5 and Figure 6.7 were done using the combined data from all three replicates and the mutated viruses were recovered using the 8-plasmid system. All nucleotide changes, except for a single Met→Ile mutation in the NS mutant shown in Figure 6.6**,** were designed to be synonymous to preserve the encoded protein sequences.

A

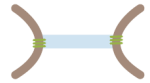




Individual-segment mutagenesis



Extensive loci mutagenesis



Restorative mutagenesis

B

Individual-segment mutagenesis

PA 84-138 : HA 824-876
PA-mut
G-->A 116
A-->G 119

PB2 437-480:NS 639-674
PB2-mut
T-->G 437
T-->G 440
A-->T 443
G-->T 454
A-->C 461

PB2 116-133:PA 1405-1422
PB2-mut
C-->T 116
C-->A 119

NS 593-637:M 382-426	
NS-mut	M-mut
A-->T 637	G-->C 397
G-->A 619	C-->T 400
	T-->A 403

NA 1173-1206:PA 409-438
PA-mut
T-->G 1193
G-->C 1194
A-->T 1195
A-->T 1196
T-->G 1199
G-->C 1200
A-->T 1201
G-->A 1202

Restorative mutagenesis

NS 593-637:M 382-426	
NS-mut	M-mut
G-->C 601	G-->C 397
A-->T 610	C-->T 400
A-->G 614	T-->A 403
A-->T 616	T-->C 406
G-->A 619	T-->A 410
C-->G 622	C-->G 418

NP 1479-1515:HA 112-145	
NP-mut	HA-mut
T-->A 1497	C-->A 115
G-->A 1503	C-->A 121
G-->T 1506	C-->T 124
G-->U 1512	A-->T 129

Extensive loci mutagenesis

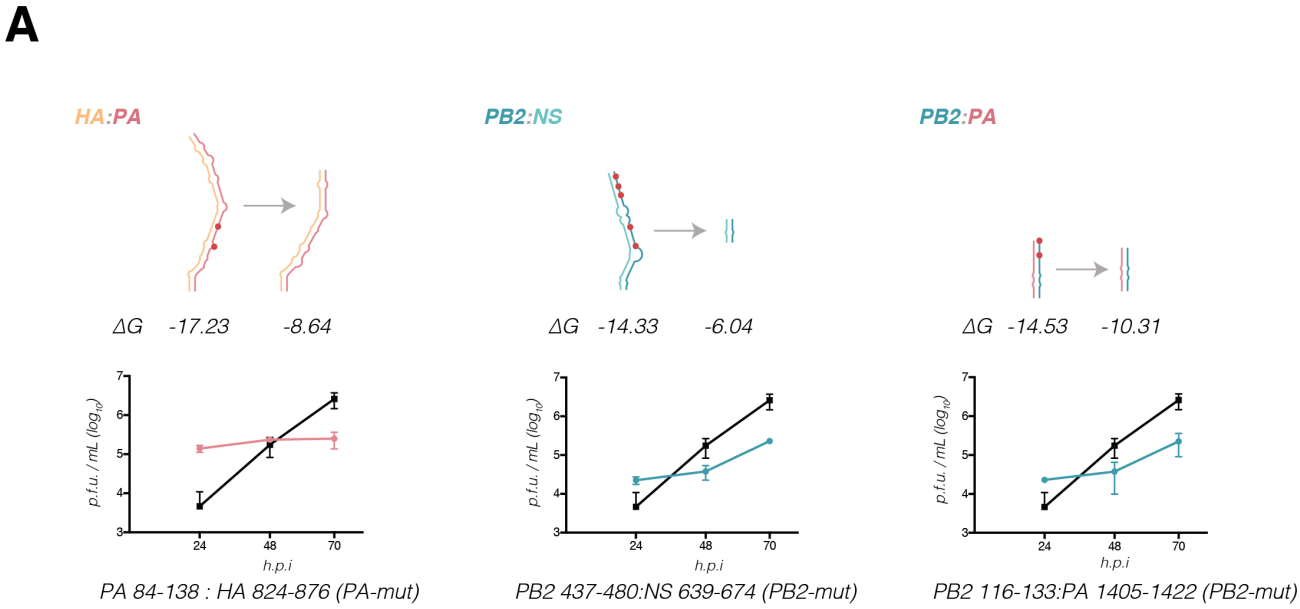
PA 1375-1447
mut TTGACTGCACGGTGGTCCGTCAGGGAGGCGCAACGGTCTCGGCGTGGACTTCAGGAATGGCTCTATCCTGGC
WT CTGAGAACAGGGAGGCCCATCCGGAAGTCTAAGTGGTCGTGGTGTTGATTTCAAAAAAGGTTCAATTCTAGC
NS 593-706
mut CGGCTAGCGGGCACGCTAGCCATTGTCATCTTCAGAGCTTCGTCGCTCTCCTCTTTGAGGATCCTTTCGACGATCTGTTTCCCGGCCCTTGTAGCGGTCTCAATG
WT CGCGATGCAGGTACAGAGGCCATGGTCATTTTGAGTGCCTCATCAGATTCTTCCTTCAGAATCCGCTCCACTATTTGCTTTCCAGCACGGGTGGCTGTTTCGATG
HA 794-885
mut CCAATGGTCACCGGATGAATGTTTTGAAATGGAAGGTTACTATTAATTGACCCTTGTGGGGTCTGGCATTTTGTATTGCATTCATGCATGCT
WT CCTATTGTGACTGGGTGTATATTCTGGAAAGGGAGATTGCTGTTTATAGATCCCTGGGGTGTTTGACACTTCGTGTTACACTCATGCATTGA

Figure 6.1 Summary of the viruses made for the mutagenesis of the inter-segment interactions

*(**A**) Only one of the interacting segments was mutated for the individual-segment mutagenesis. Extensive loci mutagenesis, which introduced more than 20 point mutations, was done on loci where overlapping interactions were found to form. Restorative mutagenesis was designed to introduce mutations in a way that when both mutated segments interact the interactions could be partially restored. For the NS-M interaction both restorative mutagenesis and individual-segment mutagenesis was performed. (**B**) List of individual mutations and their positions (5' to 3' vRNA sense) for each mutant construct. For extensive loci mutagenesis, the sequences of mutated constructs versus the wild-type sequence are given; mutation sites are highlighted in yellow.*

6.4.2 Effects of the disruption of the inter-segment RNA interactions on the growth of the WSN virus

Most of the single-segment mutants showed attenuated growth kinetics compared to the wild-type virus and further analysis of the packaged vRNA content inside these viruses showed that the packaging of the vRNA has been affected (Figure 6.2-Figure 6.3). However, a virus with mutations adjacent but not inside an interaction site did not show the same packaging defects (for the negative control vRNA packaging analysis was performed on *HA 996-1050* mutant (Figure 4.12), which showed no growth defects in the SHAPE analysis). In addition, the *M 382-426:NS 593-637 (M-mut)* and *PB2 418-480:NS 635-701 (PB2-mut)* viruses produced more defective particles compared to the wild-type virus, as measured by the particle to p.f.u. ratio (Figure 6.4). Viruses with extensive mutagenesis in the loci containing overlapping interactions were similarly attenuated and also showed genome packaging defects (Figure 6.5). The HA titre was below the level of detection for these viruses.



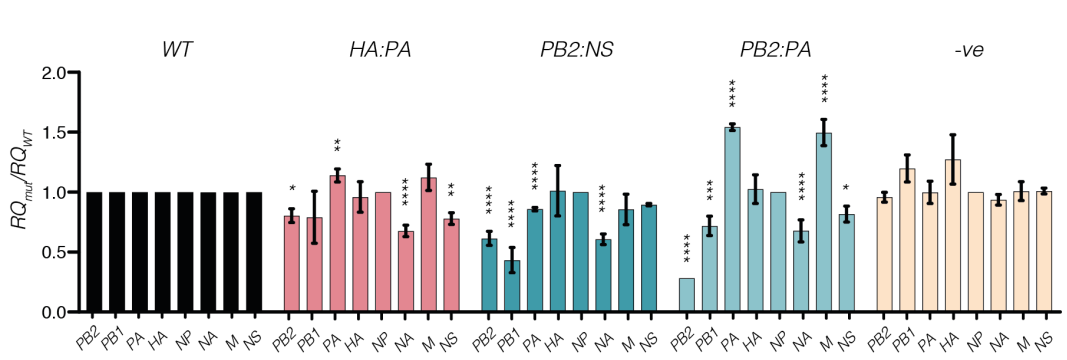


Figure 6.2 Individual-segment mutagenesis

*(**A**) Viruses, with mutations in the regions involved in the inter-segment RNA interactions, were produced and their growth, compared to the wild-type virus (black curve), was evaluated. The structures of the interactions and the predicted changes, caused by the mutagenesis, are illustrated above the virus growth curves. Red dots indicate the positions where the point mutations, as given in Figure 6.1 B, were introduced. Error bars indicate standard deviation for two independent infections. (**B**) qRT-PCR analysis of vRNA packaged into the WT and mutated viruses. vRNA was extracted from the supernatants containing different viruses and relative quantities (RQ) of each segment were measured in relation to the NA segment. Differences compared to the wild-type virus RQ values are shown. Data for three technical repeats is shown, error bars indicate standard deviation, *P<0.05, **P< 0.01, ***P<0.001, ****P<0.0001, ANOVA with Dunnett's correction.*

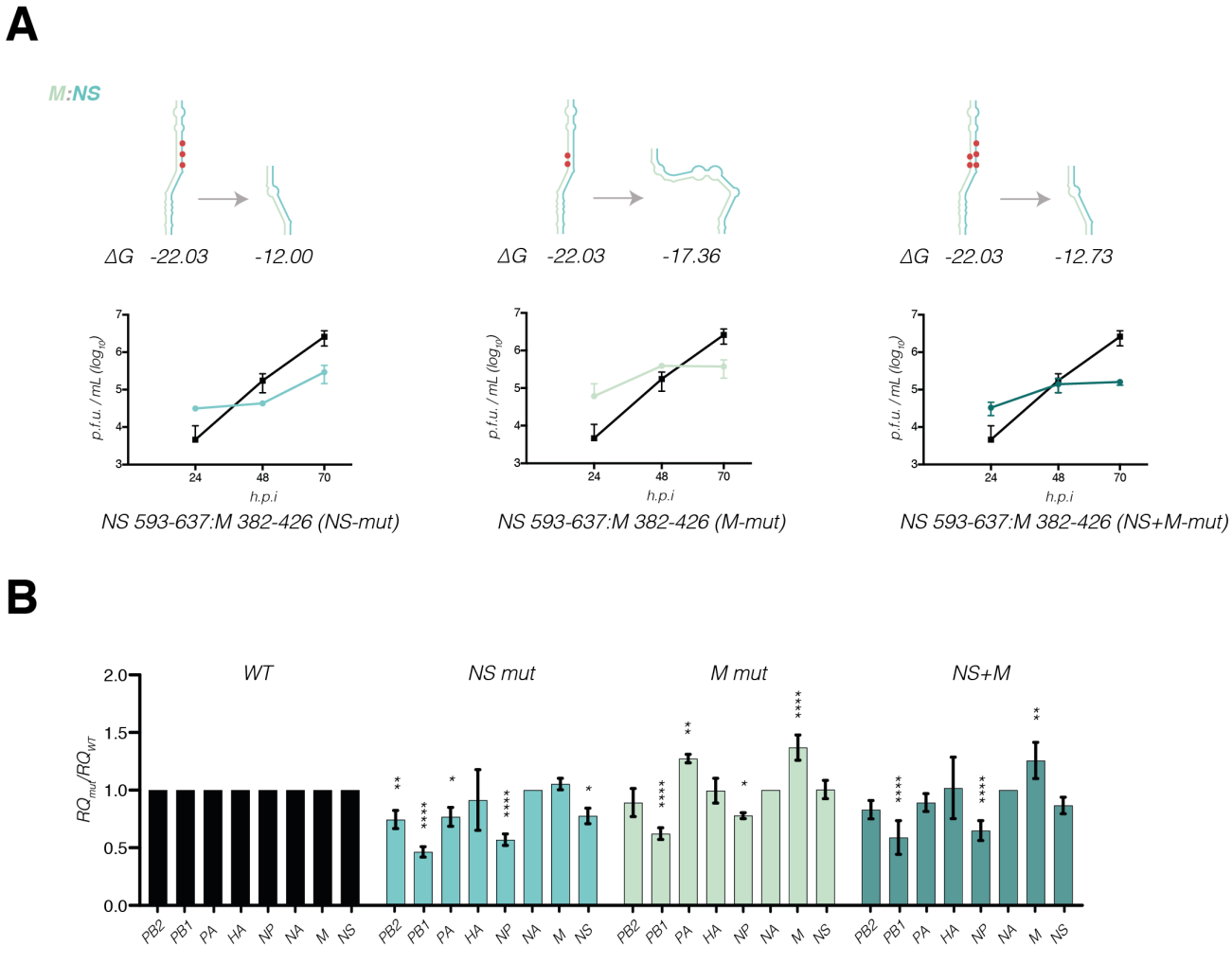


Figure 6.3 Individual-segment mutagenesis for NS-M interaction

*(**A**) Viruses, with mutations in the regions involved in the inter-segment RNA interactions, were produced, and their growth compared to the wild-type virus (black curve), was evaluated. The structures of the interactions and the predicted changes, caused by the mutagenesis, are illustrated above the virus growth curves. Red dots indicate the positions where the point mutations, as given in Figure 6.1 B, were introduced. Mutations were introduced into the M and NS segments individually and were not predicted to restore the interaction. Error bars indicate standard deviation for two independent infections. (**B**) qRT-PCR analysis of vRNA packaged into the WT and mutated viruses. vRNA was extracted from the supernatants containing different viruses and relative quantities (RQ) of each segment were measured in relation to the NA segment. Differences compared to the wild-type virus RQ values are shown. Data for three technical repeats is shown, error bars show standard deviation, *P<0.05, **P< 0.01, ***P<0.001, ****P<0.0001, ANOVA with Dunnett's correction.*

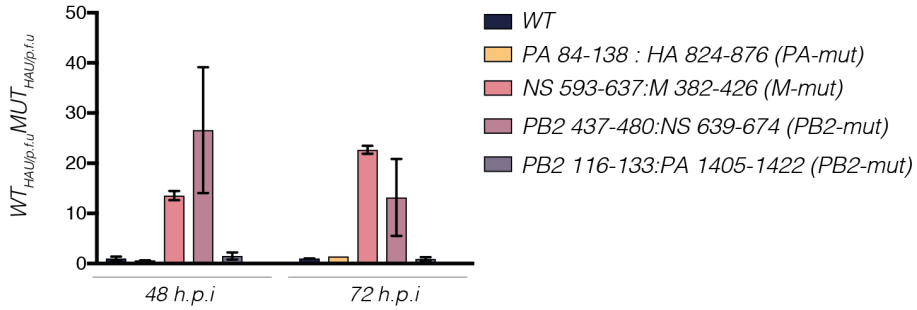


Figure 6.4 Defective particles produced by viruses with mutated inter-segment interactions

Haemagglutination assay was performed for viruses collected 48 and 72 hours post infection. Ratio of the heamagglutination units to the p.f.u. was calculated for each virus. Difference between the wild-type virus and the mutant virus ratio is shown. Error bars indicate standard deviation for two independent infections.

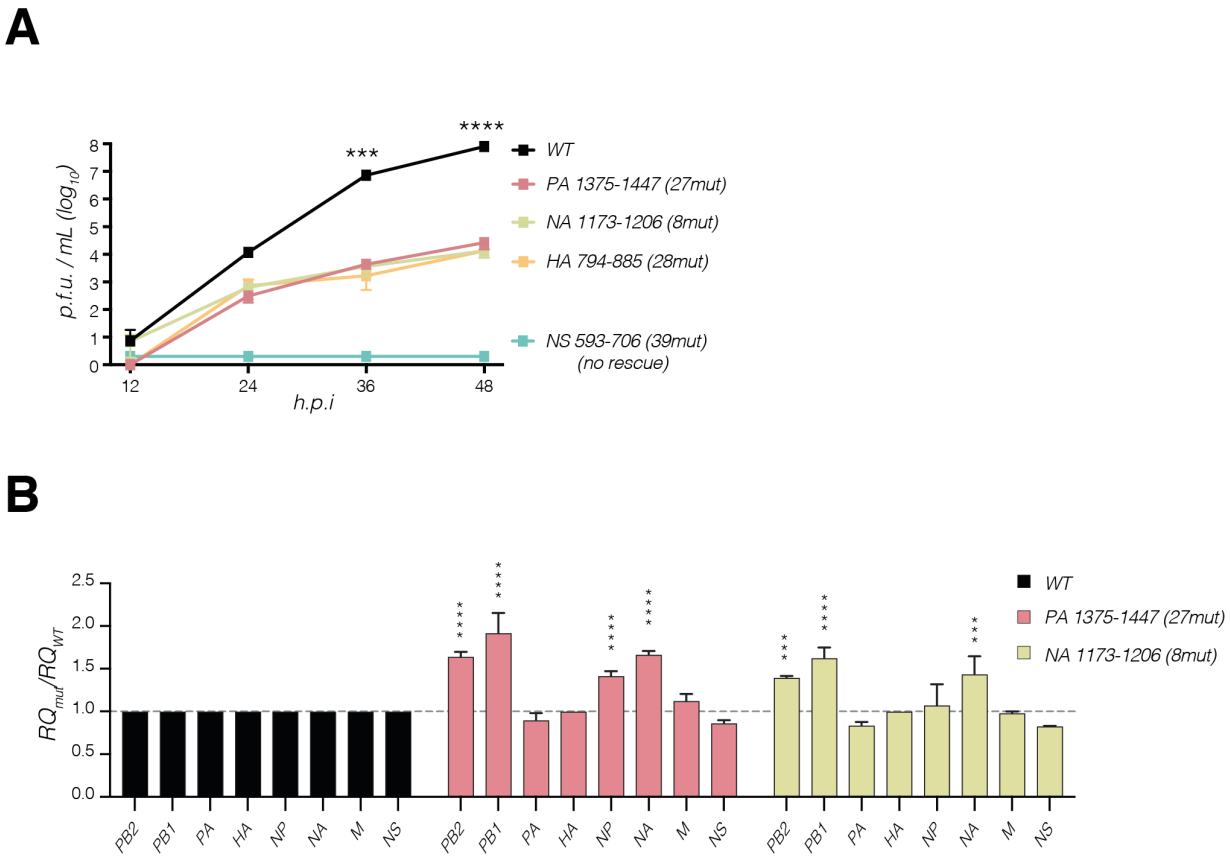


Figure 6.5 Extensive loci and individual segment mutagenesis

*(**A**) Extensive loci mutagenesis was performed for PA 1375-1447, HA 794-885 and NS 593-706 regions. No virus was recovered for the mutated NS 593-706 region. NA 1173-1206 was a single-segment mutant designed to disrupt the NA-PA interaction. Error bars indicate standard deviation for three independent infections. ANOVA with Sidak correction ***P< 0.001, ***P<0.0001. (**B**) qRT-PCR analysis of vRNA packaged into the WT and mutated viruses. vRNA was extracted from the supernatants containing different viruses and relative quantities (RQ) of each segment were measured in relation to HA segment. Differences compared to the wild-type virus RQ values are shown. The average data for the three repeats from **A** is shown. RT-qPCR experiments were not performed for the HA mutant. Error bars show standard deviation, *P<0.05, **P< 0.01, ***P<0.001, ****P<0.0001, ANOVA with Dunnett's correction.*

The combination of mutations introduced in *NS 593-637:M 382-426* interaction in Figure 6.3 was not designed to restore the RNA-RNA pairing, therefore, an alternative set of mutations were introduced, which were predicted to partially restore the NS-M interaction when both mutant segments were present (Figure 6.6 A). In addition, another restorative pair, *NP 1479-1515:HA 112-145* was made (Figure 6.6 B). Both the *M 382-426* and *NS 593-637* mutants were attenuated in their growth compared to the wild-type virus (Figure 6.7 B). While the restorative *NS 593-637:M 382-426* virus was also attenuated compared to the wild-type, it grew significantly better than the individual-segment mutants did. Furthermore, vRNA packaging analysis showed that, unlike the individual-segment mutants, the restored virus showed a similar packaging pattern compared to the wild-type (Figure 6.7 C). By contrast, a similar mutagenesis approach for the *NP 1479-1515:HA 112-145* interaction pair (Figure 6.6 B) did not show the same restorative effect. Only the individual NP segment mutant was attenuated in its growth and the double mutant virus grew worse than either of the individual segment mutants (Figure 6.7 A).

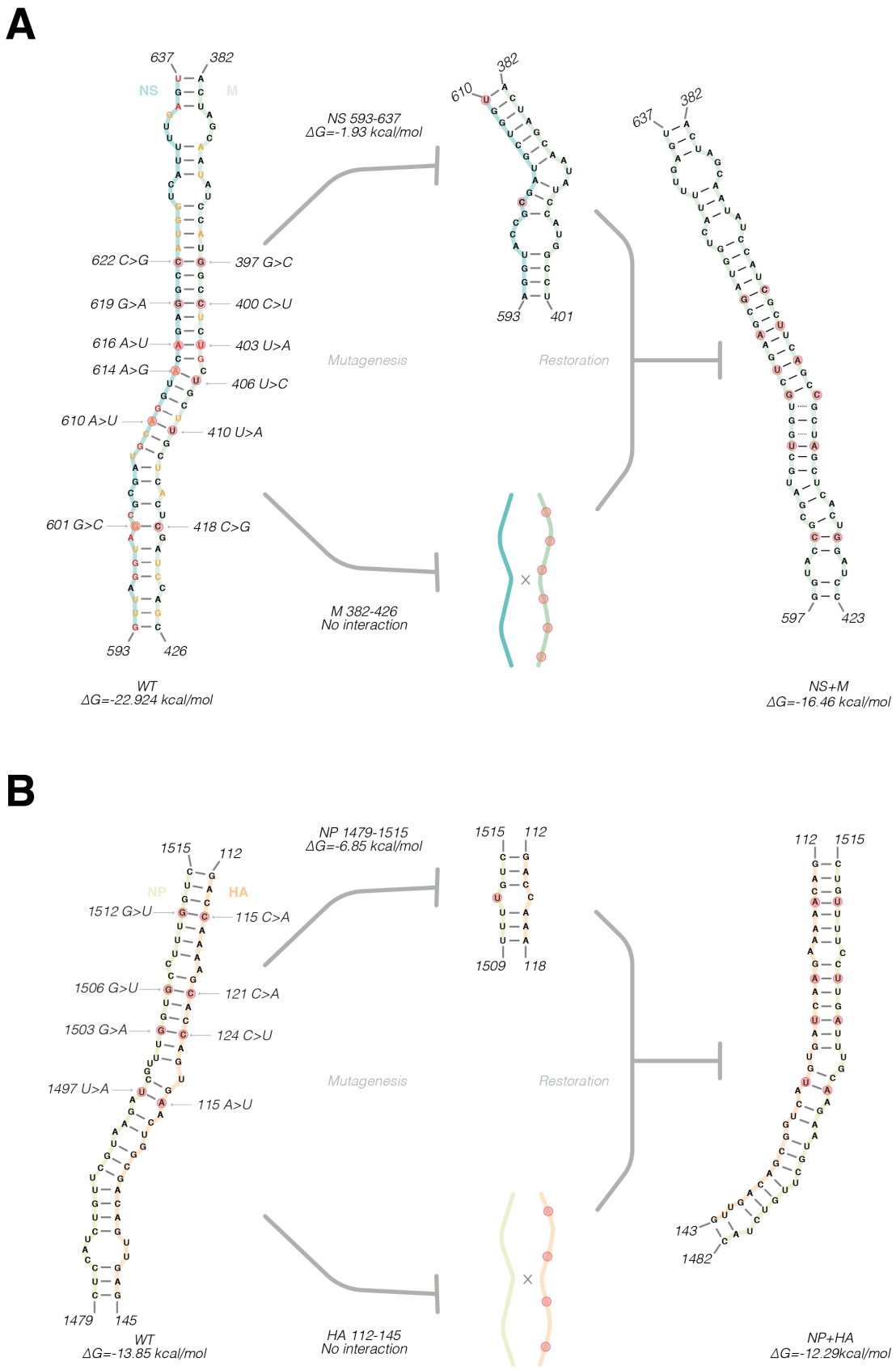


Figure 6.6 Restorative mutagenesis design

*(**A**) Design of the restorative NS-M interaction mutations. (**B**) Design of the restorative NP-HA interaction mutations. (**E**) Growth of the individual and restored NP-HA interaction viruses.*

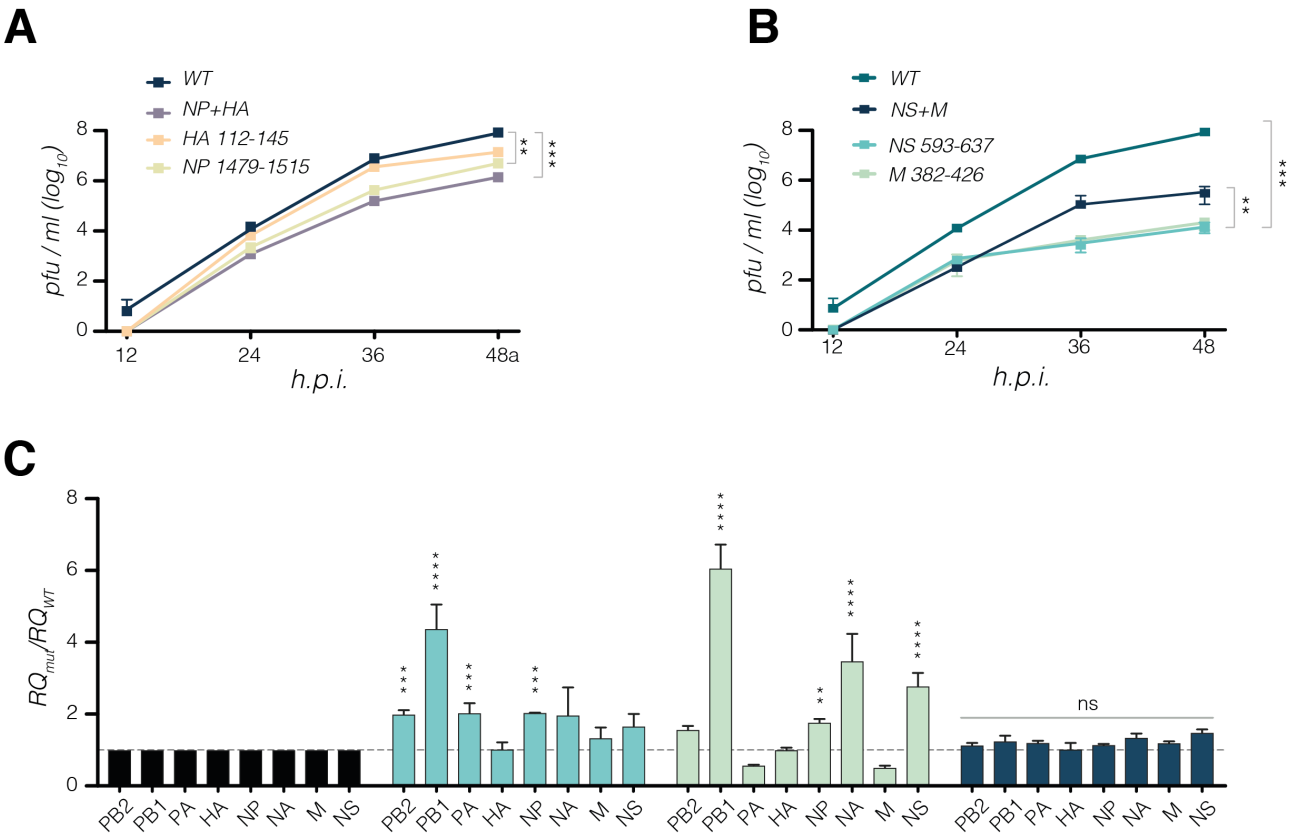


Figure 6.7 Growth and vRNA packaging of of restorative mutants

*(**A**) Growth of the individual and restored NP-HA interaction viruses. (**B**) Growth of the individual and restored NS-M interaction viruses. (**C**) qRT-PCR analysis of the vRNA packaging in restored and individually mutated viruses. vRNA was extracted from the supernatants containing different viruses and relative quantities (RQ) of each segment were measured in relation to the HA segment. Error bars show standard deviation, **P< 0.01, ***P<0.001, ****P<0.0001, ANOVA with Dunnett's correction. For the growth curves error bars indicate standard deviation for three independent infection repeats, ANOVA with Sidak correction, **P< 0.01, ***P<0.001.*

6.5 Discussion

This chapter presented the analysis of the effects of the disruption of the inter-segment RNA-RNA interactions in the influenza WSN virus genome. Disruption of the inter-segment RNA interactions using synonymous mutagenesis led to attenuated virus growth and defects in the packaging of the vRNA, indicating that SPLASH analysis did identify interactions important for the assembly of the vRNPs into budding virions.

Some of the viruses with mutations in the individual segments were found to have increased titres compared to the wild-type early during infection (Figure 6.2, Figure 6.3). As discussed for the mutants in chapter 4, this could be due to complementation occurring as a result of the presence of defective particles in the stock. Indeed, at least two of the mutants, the *M 382-426:NS 593-637 (M-mut)* and *PB2 418-480:NS 635-701 (PB2-mut),* were both found to produce more defective particles compared to the wild-type. However, the virus infections were performed at a low MOI (0.001), which should have limited the complementation effect. Another possible explanation could be that some of the mutant viruses induce less interferon response during infection, allowing them, at least early on, to spread more rapidly. It would be worth looking at the difference in interferon induction in the mutant versus the wild-type viruses, using a reporter cell line, to assess the potential differences.

Restorative mutagenesis experiments for NS-M interaction showed that genome packaging could be restored by complementary mutagenesis, however, even with restored packaging, the virus was still attenuated compared to the wild-type. In addition, an attempt to restore the NP-HA interaction was not successful. It is important to note, that introducing mutations into regions involved in the inter-segment RNA interactions firstly disrupts intramolecular secondary RNA structures; therefore, if the introduced mutations have changed the conformation of vRNA in a way that in an RNP the vRNA is no longer accessible for interaction with other RNAs, the presence of the complementary mutations in a different segment will not rescue the packaging defect. In addition, the NP side of the NP-HA interaction was located very closely to the site of NP-PA interaction, which could have potentially been affected by the mutagenesis as well. It is more difficult to explain the NS-M interaction example, where the packaging defect was restored, while virus replication remained affected. The NS mutant did contain a Met→ Ile mutation in NS1, which could have potentially affected the stability of the NS1 protein, leading to virus attenuation.

The analysis of the vRNA content in the mutated viruses showed that, compared to the wild-type virus, the mutated viruses usually showed defects in their genome packaging. Interestingly, the mutagenesis did not only affect the packaging of specific segments between which the interaction was disrupted but changed the relative amounts of many other segments as well. Such global changes in segment packaging caused by point mutations in a single segment have been observed in other studies as well (Wang, Li, Zhao, *et al.*, 2016; Hutchinson, Wise, Kudryavtseva, *et al.*, 2009; Williams, Townsend, Wylie, *et al.*, 2018). Given the network of interactions that was identified by SPLASH, these global changes in segment packaging could be explained by the likely rearrangement of the inter-segment interaction network caused by the introduced mutations. Disruption of a single interaction could lead to a rearrangement of the interaction network to a set of sub-optimal interactions, which are not equally efficient in the packaging of all vRNAs. It would, therefore, be interesting to do SPLASH on some of the packaging mutants, to see how the interaction networks are rearranged, and which interactions are more important for efficient packaging.

After vRNPs are replicated in a cell's nucleus, they need to be transported out into the cytoplasm and, through RNA-RNA interactions, to assemble into complexes containing the 8 different segment RNPs. Exactly how and where this process takes place is still unclear, though Rab11-positive recycling endosomes have for a long time been considered the main carriers of vRNPs (Eisfeld, Kawakami, Watanabe, *et al.*, 2011; Amorim, Bruce, Read, *et al.*, 2011). More recently, though, it has been shown that ER-derived irregularly coated vesicles (ICVs), the formation of which is Rab-11 dependent, may be the main organelles mediating vRNP trafficking (Martini, Fournier, Sachse, *et al.*, 2017). It is possible that the complexes of the eight different vRNPs form while the vRNPs are associated with the ICVs. Secondary RNA structures stemming from the RNPs could mediate the interaction between the RNPs and if a complex of RNPs with the eight unique RNPs has evolved to be the most energetically stable arrangement that would lead to selective packaging of the genome. Importantly though, which regions of the vRNA are accessible and able to form the inter-segment interactions may be a significant factor in determining, which strains of influenza viruses can reassort. If compatible complementary vRNA structures would stem from RNPs of different strains, these could interact and allow an assembly of the 8 RNPs from different strains.

6.6 Summary

- Disruption of the inter-segment interactions lead to attenuated virus growth and vRNA packaging defects
- Some inter-segment RNA interactions can be restored by complementary mutagenesis

7. Chapter 7: Conclusions and future directions

This thesis examined the secondary RNA structure of the influenza virus genome. SHAPE-MaP and SPLASH analyses were performed for WSN, PR8, Udorn and FLB viruses. SHAPE-MaP analysis, which was used to investigate the intramolecular secondary RNA structures, revealed that secondary RNA structures are present in all the examined viruses and the different vRNA segments have unique secondary RNA structure folds (Figures 4.1, 4.19-4.27). Low SHAPE reactivity values are indicative of structured regions of RNA, more specifically, low SHAPE reactivities have been found to primarily be associated with the base-paired RNA nucleotides (Low & Weeks, 2010). It is however, still not clear exactly what effect does protein binding have on the SHAPE reactivity of a nucleotide. In the literature, evidence can be found for the changes in the nucleotide SHAPE reactivities as a result of protein binding, however, both increase (Wilkinson, Gorelick, Vasa, *et al.*, 2008) and decrease (Kenyon, Prestwood & Lever, 2015) in the SHAPE reactivities have been observed. Furthermore, the sites of decreased SHAPE reactivities may not be the same sites were the protein binding occurs, as the binding of some proteins can affect the overall flexibility of the RNA phosphate backbone, which would consequently be reflected in the SHAPE reactivities (McGinnis, Dunkle, Cate, *et al.*, 2012; Kenyon, Prestwood & Lever, 2015).

Overall, a decrease in high-probability-associated secondary RNA structures has been observed for the *in virio* compared to the *ex virio* SHAPE analysis, indicating, that NP is destabilizing secondary RNA structures in the vRNPs. Low SHAPE reactivity sites were also present in the regions that were identified to have reduced NP binding in a previous study (Williams, Townsend, Wylie, *et al.*, 2018) (Figure 4.29), showing that structured RNA, as opposed to only NP-bound single stranded RNA, is present in vRNPs. However, comparison between the studies present in this thesis and the published NP-footprinting analysis has its limits. It is difficult to know what is the correct read density threshold for an NP-bound versus an unbound state of RNA in the NP-footprinting data, therefore, aside from the very distinct regions, in which almost no NP-associated reads were detected, comparing the SHAPE data with the rest of the footprint analysis requires some more subjective assessment. An experiment, which could be very informative for the understanding of the effect of NP binding on nucleotide SHAPE reactivities, would be to conduct SHAPE-MaP analysis using a mini-RNP. A mini-RNP contains an internally truncated vRNA

and cryo-EM structures are available for a mini-RNP associated with a 248 nucleotide-long vRNA (Coloma, Valpuesta, Arranz, *et al.*, 2009). Between 8 and 9 NP molecules were found to associate with the 254-nucleotide mini-RNP, which, if the estimate of ~24 nucleotides per NP monomer is correct (Ortega, Martín-Benito, Zürcher, *et al.*, 2000; Martín-Benito, Area, Ortega, *et al.*, 2001), would suggest that the mini-RNP- associated RNA is most likely single stranded. Conducting SHAPE analysis on a single stranded but NP-bound vRNA would give an insight on the NP effects on nucleotide SHAPE reactivities.

While SHAPE reactivities indicate the presence of base-pairing between nucleotides, they cannot be used to distinguish between intra- versus intermolecular base-pairing. While no literature is available on the presence or absence of RNA-RNA interactions inside the influenza virions, given that the RNA-RNA interactions are thought to be important for the packaging of vRNPs into budding virions, it is reasonable to hypothesise that such RNA-RNA interactions would still be present inside the virions after their release. Consequently, SPLASH experiments were conducted to investigate if inter-segment RNA interactions could be detected inside virions. Surprisingly, a network of RNA-RNA interactions was identified inside the influenza virions using SPLASH, which formed between highly complementary regions of vRNA (**Figures 5.12-5.15**). Many of the identified interactions were found to overlap with each other and therefore, are unlikely to occur in a single virion, suggesting that there are multiple possible alternative paths for vRNPs to interact with each other during the genome assembly.

Previous studies looking at the trafficking of vRNPs inside a cell have noted that there does not seem to be a clear pattern between the co-localisation of different segments in the cytoplasm (Lakdawala, Wu, Wawrzusin, *et al.*, 2014; Lakdawala, Fodor & Subbarao, 2016). In agreement, scanning electron tomography studies, based on vRNP lengths, have observed that the segment arrangement inside virions is not uniform, with different segments capable of occupying the central position in the 7+1 arrangement (Fournier, Moules, Essere, *et al.*, 2012b; Noda, Sugita, Aoyama, *et al.*, 2012). The observation of a redundant inter-segment interaction network in the present study is in agreement with previous observations. A model can be proposed (Figure 7.1), suggesting that when the vRNPs are trafficked through the cytoplasm they begin to interact with each other (probably on cytoplasmic vesicles, such as irregularly coated vesicles or Rab11 positive recycling endosomes) via the complementary regions of the vRNA. Stronger RNA-RNA

interactions will outcompete the weaker ones, but without a single all-determining interaction, the first few strong interactions formed between vRNPs will ultimately decide the final 7+1 arrangement. If there are a number of potential options for the formation of the initial interactions between vRNPs, there will be even more arrangements possible for all 8 vRNPs to assemble. The accessible complementary vRNA regions could have evolved to favour interactions between different vRNA segments, leading to the selective genome packaging observed for influenza viruses. The observed 'flexibility' in vRNP packaging is, in a sense, the perfect strategy for a virus which undergoes genetic drift, because it allows the use of RNA as a mediator for genome packaging while also accommodating the required accumulation mutations in the genome to facilitate immune evasion. In the future, both SHAPE and SPLASH experiments should be performed in virus-infected cells to determine if the same inter-segment interactions are present in cells as in virions. Furthermore, in-cell SPLASH experiments could potentially help to determine were in the cell the vRNPs start interacting with each other and when exactly does the 7+1 complex form.

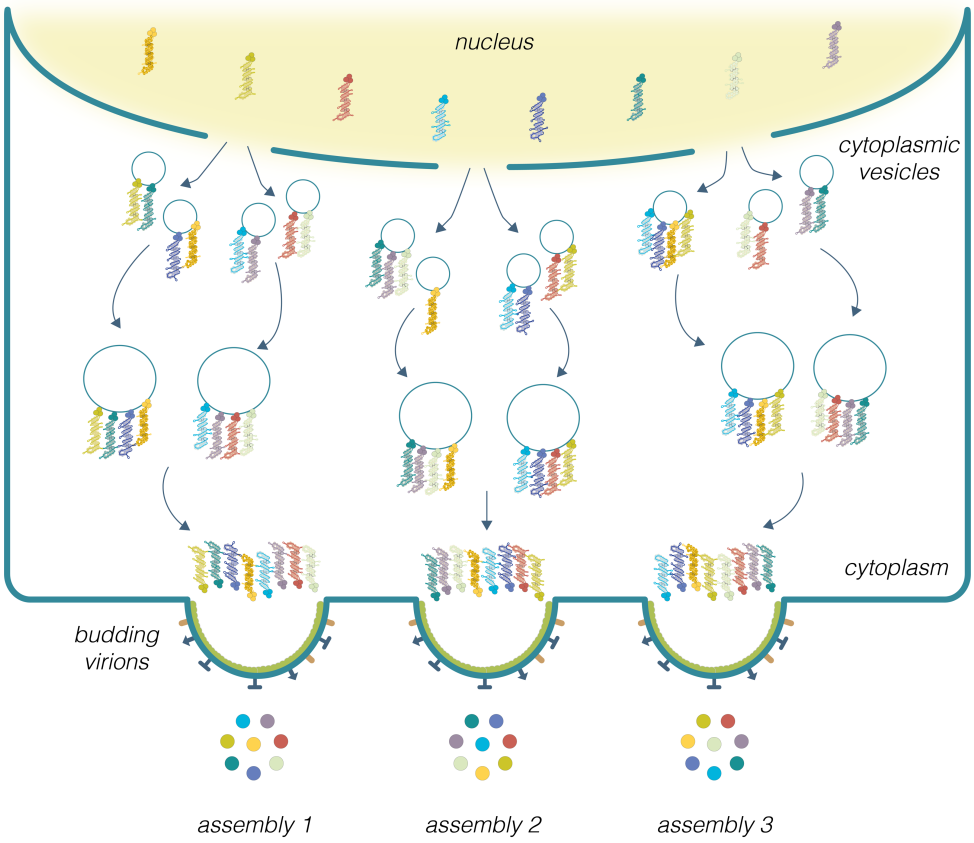


Figure 7.1 vRNP assembly model

As vRNPs get exported from the nucleus they associate with cytoplasmic vesicles (irregularly coated vesicles or recycling endosomes) and are trafficked towards the virus budding site. During vRNP trafficking there are multiple possible vRNP-vRNP interaction paths, which will determine the final vRNP assembly arrangement packaged into a virion.

SPLASH method itself cannot define the precise structure of the base-pairing between the interacting segments, as it only gives the approximate region of where the cross-linking took place, the length of which will mostly depend on size selection steps during the sequencing library preparation. By combining SHAPE reactivities with SPLASH interaction regions, a more precise structural inter-segment RNA interaction modelling was performed. Overall, the SHAPE reactivities inside the inter-segment interaction regions, as opposed to outside interaction regions, were not enriched for nucleotides showing specifically lower SHAPE reactivities, but had a similar distribution (Figure 7.2). There are a few potential explanations for the latter observation. Firstly, as it is unlikely that all the observed interactions occur in every virion, the base-pairing will only be present in some but not other virus particles and consequently the SHAPE reactivities in the interaction regions will cluster around the median SHAPE reactivity for a segment. Secondly, structured regions outside the inter-segment interaction sites and potentially the NP-binding sites, will also show low SHAPE reactivities, making the low SHAPE reactivities not unique to the inter-segment interaction regions.

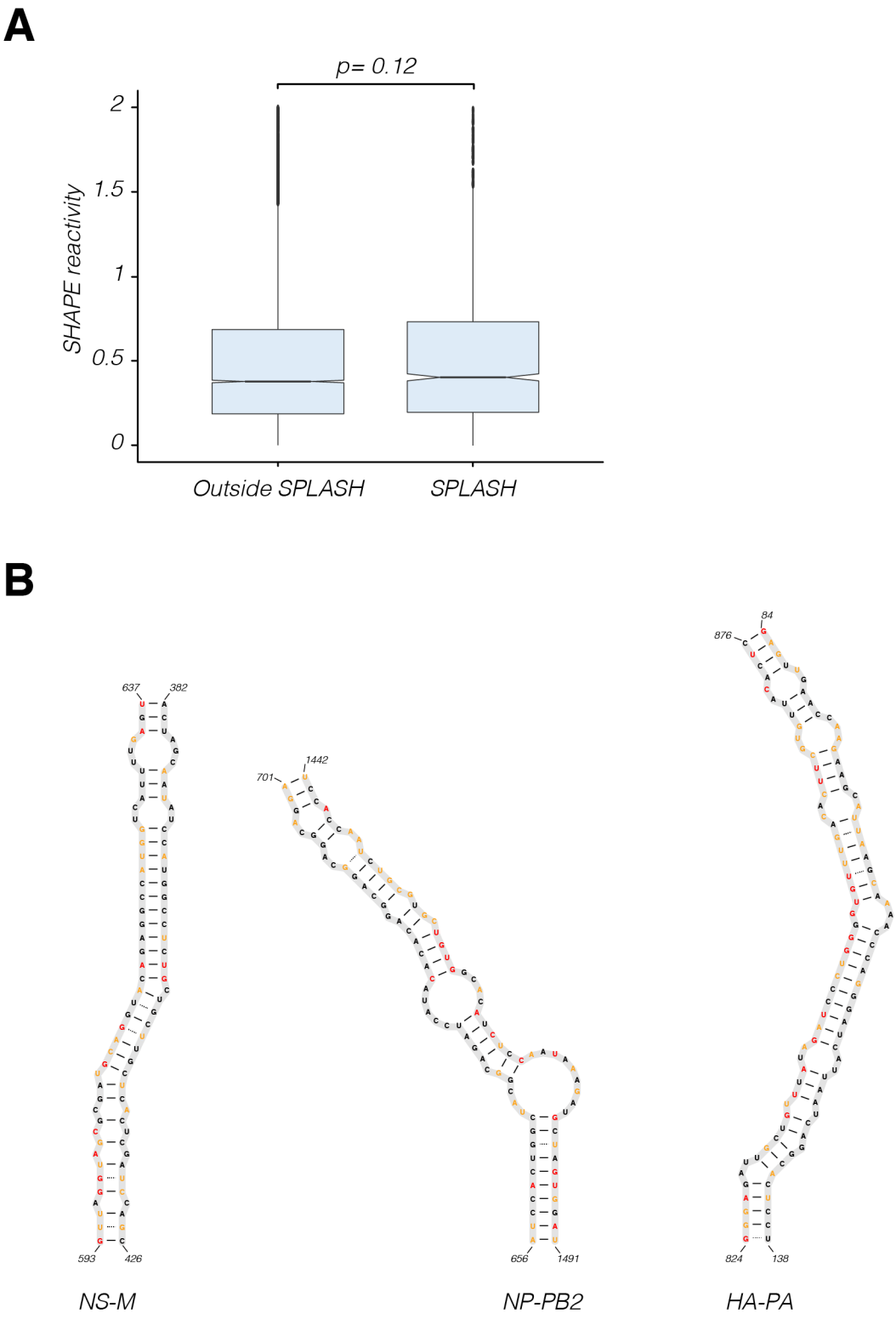


Figure 7.2 SHAPE reactivities in inter-segment interaction sites

*(**A**) Distribution of SHAPE reactivities of nucleotides positioned outside and inside SPLASH-identified inter-segment interaction sites. Student's t-test was used to compare the distributions. The distributions are shown for the WSN virus, other viruses show the same trend. (**B**) Examples, of modelled inter-segment interactions in WSN virus, coloured by nucleotide SHAPE reactivities.*

In the work presented in this thesis the SPLASH method was used to detect only inter-segment RNA-RNA interactions, though the SPLASH method can also be used to detect long-range intramolecular RNA interactions. At the time of writing this thesis the analysis of long-range RNA-RNA interactions using SPLASH is still in progress, though preliminary results clearly show that the base-pairing between the 5′ and 3′ ends of all vRNA segments, indicating the paired regions of the promoter, can be detected in addition to other long-range interactions (Figure 7.3 A). It is worth noting that in the SHAPE-MaP data the sequencing quality of the terminal nucleotides is typically not high enough for reliable structural predictions, however, where possible, the general trend for low SHAPE reactivities in the base-paired promoter region can be observed (Figure 7.3 B).

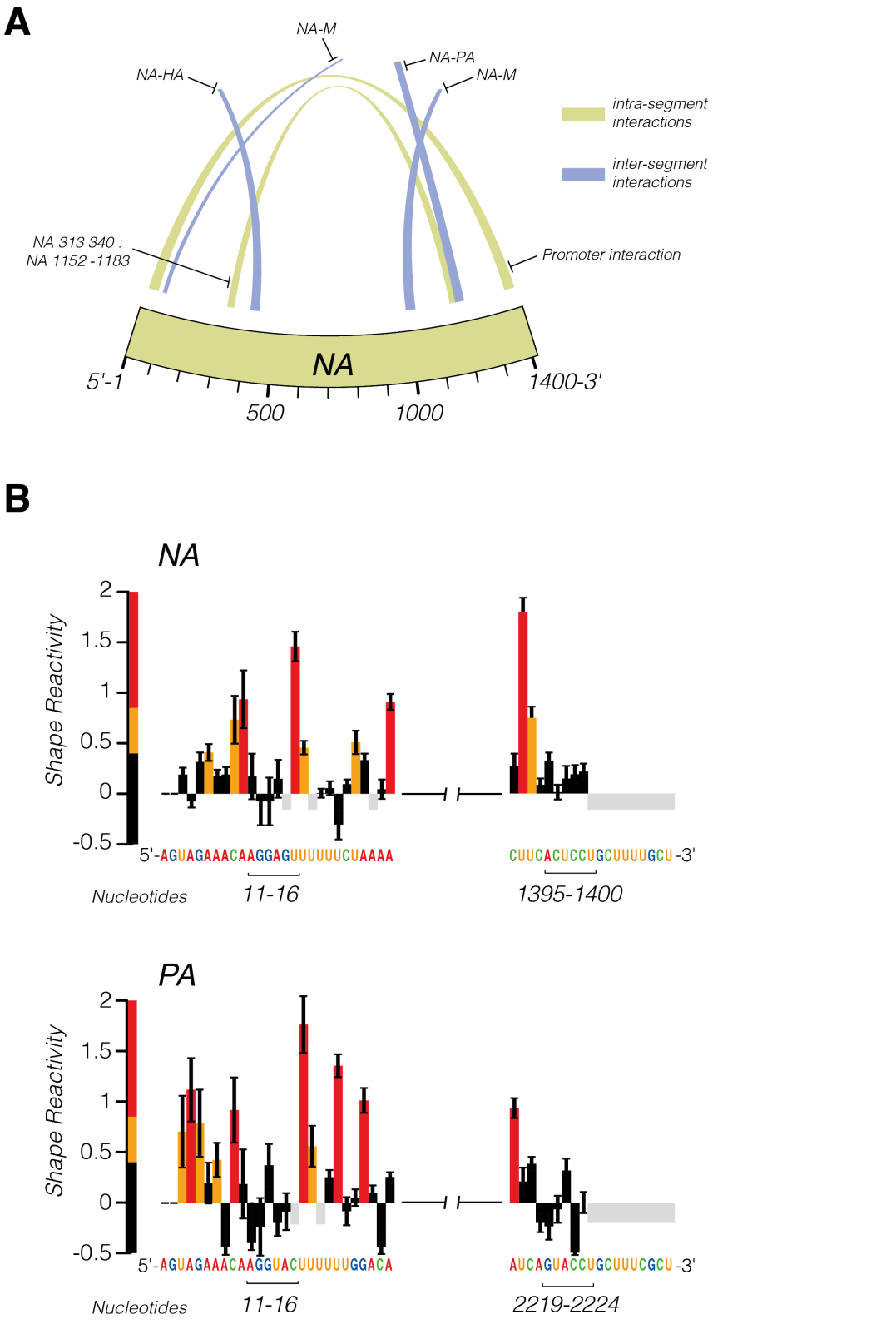


Figure 7.3 Long distance intramolecular RNA-RNA interactions

*(**A**) Long distance intra-segment and inter-segment RNA-RNA interactions detected in the WSN virus NA segment using SPLASH. (**B**) 5′ and 3′ termini SHAPE reactivities in Udorn NA and PA segment, showing low SHAPE reactivities in the base-paired regions of the vRNA promoters. Grey bars indicate nucleotides for which no data is available*

For production of vaccines against seasonal influenza viruses, classical reassortment is used to make high titre-yielding viruses, containing the antigens from poorly-egg adapted seasonal strains. Specifically, for production of the seasonal H3N2 vaccines, eggs are infected with PR8 and H3N2 viruses and the progeny viruses containing at least NA and HA segments from the H3N2 virus are selected. When the produced vaccine seed strains were genotyped, it was noted that the packaging of H3N2 NA segment is often accompanied by a co-selection for the H3N2 PB1 segment (Cobbin, Ong, Verity, *et al.*, 2014; Gilbertson, Zheng, Gerber, *et al.*, 2016). The SPLASH study using Udorn virus has identified the interaction, which mediates the co-segregation between PB1 and NA segments during vaccine production. Personal communications also indicate that H3N2 strains exist, which lack the observed PB1-NA co-segregation due to mutations in the region of NA which forms the interaction with PB1. These observations show that interactions identified with SPLASH clearly have relevance for the potential reassortment between influenza viruses, but they also pose a question - are all interactions that mediate co-segregation between the viral segments also necessary for genome packaging? One can postulate a situation in which an individual influenza strain has an accessible region of vRNA that has no role in the packaging of its genome, but because that region is highly complementary to another region of vRNA in a different virus strain, when the two strains co-infect a cell, the complementarity between those vRNAs will drive co-segregation between vRNPs of the different strains. If the classical packaging signals are not mutually exclusive between the viruses, then the co-segregation interaction will drive reassortment between the different strains. In other words, compatible packaging signals are required for co-segregation, but not all co-segregations signals are required for packaging. In such situation, introduction of mutations into an inter-segment interaction site that is not essential for packaging will have no effect on the individual virus growth, but it would prevent reassortment from occurring in the presence of a reassortment-compatible strain.

The synonymous mutagenesis experiments performed in this study have shown that disruption of secondary RNA structures and inter-segment interactions can have detrimental effects on the virus growth and genome packaging, indicating that secondary RNA structures play important role in the influenza virus life cycle. Preliminary experiments using intramolecular secondary RNA structures-targeting antisense oligonucleotides (ASOs) were performed and showed that disrupting secondary RNA structures in the wild-type virus can also affect virus

growth. In the future similar experiments should be performed using ASOs targeting the most common inter-segment interactions, which might be better targets compared to the intramolecular RNA structures, as they should be readily available for hybridisation with ASOs in the cytoplasm. The discovered effective ASOs could also be further used to develop new antivirals that target inter-segment RNA interactions and disrupt the virus genome packaging.

Overall, the work presented in this thesis has provided a deeper insight into the RNA structure of the influenza virus genome, aspects of which have important implications for our understanding of the processes of the virus genome packaging and reassortment.

References

Aggarwal, S., Dewhurst, S., Takimoto, T. & Kim, B. (2011) Biochemical impact of the host adaptation associated PB2 E627K mutation on the temperature-dependent RNA synthesis kinetics of influenza A virus polymerase complex. *Journal of Biological Chemistry*. [Online] 286 (40), jbc.M111.262048–34513. Available from: doi:10.1074/jbc.M111.262048.

Alexander, D.J. (2000) A review of avian influenza in different bird species. *Veterinary Microbiology*. 74 (1-2), 3–13.

Amorim, M.J., Bruce, E.A., Read, E.K.C., Foeglein, A., et al. (2011) A Rab11- and Microtubule-Dependent Mechanism for Cytoplasmic Transport of Influenza A Virus Viral RNA. *Journal of Virology*. [Online] 85 (9), 4143–4156. Available from: doi:10.1128/JVI.02606-10.

Amorim, M.J., Read, E.K., Dalton, R.M., Medcalf, L., et al. (2007) Nuclear export of influenza A virus mRNAs requires ongoing RNA polymerase II activity. *Traffic (Copenhagen, Denmark)*. [Online] 8 (1), 1–11. Available from: doi:10.1111/j.1600-0854.2006.00507.x.

Archer, E.J., Simpson, M.A., Watts, N.J., O'Kane, R., et al. (2013) Long-Range Architecture in a Viral RNA Genome. *Biochemistry*. [Online] 52 (18), 3182–3190. Available from: doi:10.1021/bi4001535.

Arranz, R., Coloma, R., Chichón, F.J., Conesa, J.J., et al. (2012) The Structure of Native Influenza Virion Ribonucleoproteins. *Science*. [Online] 338 (6114), 1634–1637. Available from: doi:10.1126/science.1228172.

Avilov, S.V., Moisy, D., Munier, S., Schraidt, O., et al. (2011) Replication competent influenza A virus encoding split-GFP tagged PB2 polymerase subunit allows live cell imaging of the viral life cycle. *Journal of Virology*. [Online] 86 (3), JVI.05820–11–1448. Available from: doi:10.1128/JVI.05820-11.

Aw, J.G.A., Shen, Y., Wilm, A., Sun, M., et al. (2016) In vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Molecular Cell*. [Online] 1–16. Available from: doi:10.1016/j.molcel.2016.04.028.

Baker, S.F., Nogales, A., Finch, C., Tuffy, K.M., et al. (2014) Influenza A and B virus intertypic reassortment through compatible viral packaging signals. *Journal of Virology*. [Online] 88 (18), 10778–10791. Available from: doi:10.1128/JVI.01440-14.

Barry, J.K. & Miller, W.A. (2002) A-1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. *Proceedings of the National Academy of Sciences*. [Online] 99 (17), 11133–11138. Available from: doi:10.1073/pnas.162223099.

Baudin, F., Bach, C., Cusack, S. & 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. *The EMBO journal*. 13 (13), 3158–3165.

Belfetmi, A., Zargarian, L., Tisné, C., Sleiman, D., et al. (2016) Insights into the mechanisms of RNA secondary structure destabilization by the HIV-1 nucleocapsid protein. *RNA*. [Online] 22 (4), 506–517. Available from: doi:10.1261/rna.054445.115.

Bellaousov, S. & Mathews, D.H. (2010) ProbKnot: Fast prediction of RNA secondary structure including pseudoknots. *RNA*. [Online] 16 (10), 1870–1880. Available from: doi:10.1261/rna.2125310.

Bellaousov, S., Reuter, J.S., Seetin, M.G. & Mathews, D.H. (2013) RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic acids research*. [Online] 41 (Web Server issue), W471–W474. Available from: doi:10.1093/nar/gkt290.

Belser, J.A., Bridges, C.B., Katz, J.M. & Tumpey, T.M. (2009) Past, Present, and Possible Future Human Infection with Influenza Virus A Subtype H7. *Emerging Infectious Diseases*. [Online] 15 (6), 859–865. Available from: doi:10.3201/eid1506.090072.

Beren, C., Dreesens, L.L., Liu, K.N., Knobler, C.M., et al. (2017) The Effect of RNA Secondary Structure on the Self-Assembly of Viral Capsids. *Biophysical Journal*. [Online] 113 (2), 339–347. Available from: doi:10.1016/j.bpj.2017.06.038.

Bevilacqua, P.C., Ritchey, L.E., Su, Z. & Assmann, S.M. (2016) Genome-Wide Analysis of RNA Secondary Structure. *Annual Review of Genetics*. [Online] 50 (1), 235–266. Available from: doi:10.1146/annurev-genet-120215-035034.

Biere, B., Bauer, B. & Schweiger, B. (2010) Differentiation of influenza B virus lineages Yamagata and Victoria by real-time PCR. *Journal of Clinical Microbiology*. [Online] 48 (4), 1425–1427. Available from: doi:10.1128/JCM.02116-09.

Biswas, S.K., Boutz, P.L. & Nayak, D.P. (1998) Influenza virus nucleoprotein interacts with influenza virus polymerase proteins. *Journal of Virology*. 72 (7), 5493–5501.

Bodewes, R., Morick, D., de Mutsert, G., Osinga, N., et al. (2013) Recurring influenza B virus infections in seals. *Emerging Infectious Diseases*. [Online] 19 (3), 511–512. Available from: doi:10.3201/eid1903.120965.

Brooke, C.B., Ince, W.L., Wei, J., Bennink, J.R., et al. (2014) Influenza A virus nucleoprotein selectively decreases neuraminidase gene-segment packaging while enhancing viral fitness and transmissibility. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 111 (47), 16854–16859. Available from: doi:10.1073/pnas.1415396111.

Brown, I.H., Harris, P.A., Mccauley, J.W. & Alexander, D.J. (1998) Multiple genetic reassortment of avian and human influenza A viruses in European pigs, resulting in the emergence of an H1N2 virus of novel genotype. *The Journal of general virology*. [Online] 79 (Pt 12) (12), 2947–2955. Available from: doi:10.1099/0022-1317-79-12-2947.

Bruce, E.A., Digard, P. & Stuart, A.D. (2010) The Rab11 Pathway Is Required for Influenza A Virus Budding and Filament Formation. *Journal of Virology*. [Online] 84 (12), 5848–5859. Available from: doi:10.1128/JVI.00307-10.

Brunotte, L., Flies, J., Bolte, H., Reuther, P., et al. (2014) The nuclear export protein of H5N1 influenza A viruses recruits Matrix 1 (M1) protein to the viral ribonucleoprotein to mediate nuclear export. *The Journal of biological chemistry*. [Online] 289 (29), 20067–20077. Available from: doi:10.1074/jbc.M114.569178.

Bullido, R., Gómez-Puertas, P., Albo, C. & Portela, A. (2000) Several protein regions contribute to determine the nuclear and cytoplasmic localization of the influenza A virus nucleoprotein. *Journal of General Virology*. [Online] 81 (1), 135–142. Available from: doi:10.1099/0022-1317-81-1-135.

Burgui, I., Yángüez, E., Sonenberg, N. & Nieto, A. (2007) Influenza virus mRNA translation revisited: is the eIF4E cap-binding factor required for viral mRNA translation? *Journal of Virology*. [Online] 81 (22), 12427–12438. Available from: doi:10.1128/JVI.01105-07.

Burrill, C.P., Westesson, O., Schulte, M.B., Strings, V.R., et al. (2013) Global RNA Structure Analysis of Poliovirus Identifies a Conserved RNA Structure Involved in Viral Replication and Infectivity. *Journal of Virology*. [Online] 87 (21), 11670–11683. Available from: doi:10.1128/JVI.01560-13.

Busan, S. & Weeks, K.M. (2018) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA*. [Online] 24 (2), 143–148. Available from: doi:10.1261/rna.061945.117.

Calder, L.J., Wasilewski, S., Berriman, J.A. & Rosenthal, P.B. (2010) Structural organization of a filamentous influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 107 (23), 10685–10690. Available from: doi:10.1073/pnas.1002123107.

Cauldwell, A.V., Long, J.S., Moncorge, O. & Barclay, W.S. (2014) Viral determinants of influenza A virus host range. *Journal of General Virology*. [Online] 95 (Pt_6), 1193–1210. Available from: doi:10.1099/vir.0.062836-0.

Chang, K.-Y. & Tinoco, I., Jr (1997) The structure of an RNA 'kissing' hairpin complex of the HIV TAR hairpin loop and its complement. *Journal of molecular biology*. [Online] 269 (1), 52–66. Available from: doi:10.1006/jmbi.1997.1021.

Chang, S., Sun, D., Liang, H., Wang, J., et al. (2015) Cryo-EM Structure of Influenza Virus RNA Polymerase Complex at 4.3 Å Resolution. *Molecular Cell*. [Online] 57 (5), 925–935. Available from: doi:10.1016/j.molcel.2014.12.031.

Chen, B.J., Leser, G.P., Jackson, D. & Lamb, R.A. (2008) The influenza virus M2 protein cytoplasmic tail interacts with the M1 protein and influences virus assembly at the site of virus budding. *Journal of Virology*. [Online] 82 (20), 10059–10070. Available from: doi:10.1128/JVI.01184-08.

Chen, B.J., Leser, G.P., Morita, E. & Lamb, R.A. (2007) Influenza virus hemagglutinin and neuraminidase, but not the matrix protein, are required for assembly and budding of plasmid-derived virus-like particles. *Journal of Virology*. [Online] 81 (13), 7111–7123. Available from: doi:10.1128/JVI.00361-07.

Chen, W., Calvo, P.A., Malide, D., Gibbs, J., et al. (2001) A novel influenza A virus mitochondrial protein that induces cell death. *Nature Medicine*. [Online] 7 (12), 1306–1312. Available from: doi:10.1038/nm1201-1306.

Chenavas, S., Estrozi, L.F., Slama-Schwok, A., Delmas, B., et al. (2013) Monomeric nucleoprotein of influenza A virus. Félix A Rey (ed.). *PLoS Pathogens*. [Online] 9 (3), e1003275. Available from: doi:10.1371/journal.ppat.1003275.

Chiapponi, C., Faccini, S., De Mattia, A., Baioni, L., et al. (2016) Detection of Influenza D Virus among Swine and Cattle, Italy. *Emerging Infectious Diseases*. [Online] 22 (2), 352–354. Available from: doi:10.3201/eid2202.151439.

Chiba, S., Jamal, A. & Suzuki, N. (2018) First Evidence for Internal Ribosomal Entry Sites in Diverse Fungal Virus Genomes Reed B Wickner (ed.). *mBio*. [Online] 9 (2). Available from: doi:10.1128/mBio.02350-17.

Chou, Y.-Y., Vafabakhsh, R., Doğanay, S., Gao, Q., et al. (2012) One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 109 (23), 9101–9106. Available from: doi:10.1073/pnas.1206069109.

Cimino, G.D., Gamper, H.B., Isaacs, S.T. & Hearst, J.E. (1985) Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. *Annual review of biochemistry*. [Online] 54 (1), 1151–1193. Available from: doi:10.1146/annurev.bi.54.070185.005443.

Clever, J.L., Miranda, D. & Parslow, T.G. (2002) RNA structure and packaging signals in the 5 ' leader region of the human immunodeficiency virus type 1 genome. *Journal of Virology*. [Online] 76 (23), 12381–12387. Available from: doi:10.1128/JVI.76.23.12381-12387.2002.

Cobbin, J.C.A., Ong, C., Verity, E., Gilbertson, B.P., et al. (2014) Influenza virus PB1 and neuraminidase gene segments can cosegregate during vaccine reassortment driven by interactions in the PB1 coding region. *Journal of Virology*. [Online] 88 (16), 8971–8980. Available from: doi:10.1128/JVI.01022-14.

Cobbin, J.C.A., Verity, E.E., Gilbertson, B.P., Rockman, S.P., et al. (2013) The source of the PB1 gene in influenza vaccine reassortants selectively alters the hemagglutinin content of the resulting seed virus. *Journal of Virology*. [Online] 87 (10), 5577–5585. Available from: doi:10.1128/JVI.02856-12.

Coloma, R., Valpuesta, J.M., Arranz, R., Carrascosa, J.L., et al. (2009) The Structure of a Biologically Active Influenza Virus Ribonucleoprotein Complex Félix A Rey (ed.). *PLoS Pathogens*. [Online] 5 (6), e1000491. Available from: doi:10.1371/journal.ppat.1000491.

Dadonaite, B., Vijayakrishnan, S., Fodor, E., Bhella, D., et al. (2016) Filamentous influenza viruses. *Journal of General Virology*. [Online] 97 (8), 1755–1764. Available from: doi:10.1099/jgv.0.000535.

Darty, K., Denise, A. & Ponty, Y. (2009) VARNA - Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*. [Online] 25 (15), 1974–1975. Available from: doi:10.1093/bioinformatics/btp250.

Davis, M., Sagan, S.M., Pezacki, J.P., Evans, D.J., et al. (2008) Bioinformatic and Physical Characterizations of Genome-Scale Ordered RNA Structure in Mammalian RNA Viruses. *Journal of Virology*. [Online] 82 (23), 11824–11836. Available from: doi:10.1128/JVI.01078-08.

Dawson, W.K., Lazniewski, M. & Plewczynski, D. (2017) RNA structure interactions and ribonucleoprotein processes of the influenza A virus. *Briefings in Functional Genomics*. [Online] 9e1003624. Available from: doi:10.1093/bfgp/elx028.

de Vries, E., de Vries, R.P., Wienholts, M.J., Floris, C.E., et al. (2012) Influenza A virus entry into cells lacking sialylated N-glycans. *Proceedings of the National Academy of Sciences*. [Online] 109 (19), 201200987–201207462. Available from: doi:10.1073/pnas.1200987109.

de Vries, E., Tscherne, D.M., Wienholts, M.J., Cobos-Jiménez, V., et al. (2011) Dissection of the Influenza A Virus Endocytic Routes Reveals Macropinocytosis as an Alternative Entry Pathway Andrew Pekosz (ed.). *PLoS Pathogens*. [Online] 7 (3), e1001329. Available from: doi:10.1371/journal.ppat.1001329.

Deigan, K.E., Li, T.W., Mathews, D.H. & Weeks, K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 106 (1), 97–102. Available from: doi:10.1073/pnas.0806929106.

Dias, A., Bouvier, D., Crépin, T., McCarthy, A.A., et al. (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*. [Online] 458 (7240), 914–918. Available from: doi:10.1038/nature07745.

Dimmock, N.J. & Easton, A.J. (2014) Defective interfering influenza virus RNAs: time to re-evaluate their clinical potential as broad spectrum antivirals? *Journal of Virology*. [Online] 88 (10), JVI.03193–13–5227. Available from: doi:10.1128/JVI.03193-13.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., et al. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*. [Online] 505 (7485), 696–700. Available from: doi:10.1038/nature12756.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. [Online] 29 (1), 15–21. Available from: doi:10.1093/bioinformatics/bts635.

Donald, H.B. & Isaacs, A. (1954) Counts of Influenza Virus Particles. *Microbiology*. [Online] 10 (3), 457–464. Available from: doi:10.1099/00221287-10-3-457.

Doshi, K.J., Cannone, J.J., Cobaugh, C.W. & Gutell, R.R. (2004) Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC bioinformatics*. [Online] 5 (1), 105. Available from: doi:10.1186/1471-2105-5-105.

Dubois, J., Terrier, O. & Rosa-Calatrava, M. (2014) Influenza Viruses and mRNA Splicing: Doing More with Less. *mBio*. [Online] 5 (3), –14. Available from: doi:10.1128/mBio.00070-14.

Dudas, G., Bedford, T., Lycett, S. & Rambaut, A. (2015) Reassortment between Influenza B Lineages and the Emergence of a Coadapted PB1-PB2-HA Gene Complex. *Molecular biology and evolution*. [Online] 32 (1), 162–172. Available from: doi:10.1093/molbev/msu287.

Dugan, V.G., Chen, R., Spiro, D.J., Sengamalay, N., et al. (2008) The evolutionary genetics and emergence of avian influenza viruses in wild birds Daniel R Perez (ed.). *PLoS Pathogens*. [Online] 4 (5). Available from: doi:10.1371/journal.ppat.1000076.

Duhaut, S.D. & Dimmock, N.J. (1998) Heterologous Protection of Mice from a Lethal Human H1N1 Influenza A Virus Infection by H3N8 Equine Defective Interfering Virus: Comparison of Defective RNA Sequences Isolated from the DI Inoculum and Mouse Lung. *Virology*. [Online] 248 (2), 241–253. Available from: doi:10.1006/viro.1998.9267.

Duhaut, S.D. & Mccauley, J.W. (1996) Defective RNAs inhibit the assembly of influenza virus genome segments in a segment-specific manner. *Virology*. [Online] 216 (2), 326–337. Available from: doi:10.1006/viro.1996.0068.

Dunn, F.L. (1958) Pandemic influenza in 1957; review of international spread of new Asian strain. *Journal of the American Medical Association*. 166 (10), 1140–1148.

Effros, R.B. (1977) Generation of both cross-reactive and virus-specific T-cell populations after immunization with serologically distinct influenza A viruses. *The Journal of experimental medicine*. [Online] 145 (3), 557–568. Available from: doi:10.1084/jem.145.3.557.

Eisfeld, A.J., Kawakami, E., Watanabe, T., Neumann, G., et al. (2011) RAB11A is essential for transport of the influenza virus genome to the plasma membrane. *Journal of Virology*. [Online] 85 (13), 6117–6126. Available from: doi:10.1128/JVI.00378-11.

Eisfeld, A.J., Neumann, G. & Kawaoka, Y. (2014) At the centre: influenza A virus ribonucleoproteins. *Nature reviews. Microbiology*. [Online] 13 (1), 28–41. Available from: doi:10.1038/nrmicro3367.

Elton, D., Simpson-Holley, M., Archer, K., Medcalf, L., et al. (2001) Interaction of the influenza virus nucleoprotein with the cellular CRM1-mediated nuclear export pathway. *Journal of Virology*. [Online] 75 (1), 408–419. Available from: doi:10.1128/JVI.75.1.408-419.2001.

Engreitz, J.M., Sirokman, K., McDonel, P., Shishkin, A.A., et al. (2014) RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell*. [Online] 159 (1), 188–199. Available from: doi:10.1016/j.cell.2014.08.018.

Essere, B., Yver, M., Gavazzi, C., Terrier, O., et al. (2013) Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 110 (40), E3840–E3848. Available from: doi:10.1073/pnas.1308649110.

Fajardo, T.J., Sung, P.-Y., Celma, C.C. & Roy, P. (2017) Rotavirus Genomic RNA Complex Forms via Specific RNA-RNA Interactions: Disruption of RNA Complex Inhibits Virus Infectivity. *Viruses*. [Online] 9 (7). Available from: doi:10.3390/v9070167.

Fallmann, J., Will, S., Engelhardt, J., Grüning, B., et al. (2017) Recent advances in RNA folding. *Journal of biotechnology*. [Online] 26197–104. Available from: doi:10.1016/j.jbiotec.2017.07.007.

FAO (2018) *H7N9 situation update*. [Online]. 27 June 2018. Available from: http://www.fao.org/ag/againfo/programmes/en/empres/h7n9/situation_update.html [Accessed: 12 August 2018].

Ferhadian, D., Contrant, M., Printz-Schweigert, A., Smyth, R.P., et al. (2018) Structural and Functional Motifs in Influenza Virus RNAs. *Frontiers in Microbiology*. [Online] 9559. Available from: doi:10.3389/fmicb.2018.00559.

Flick, R., Neumann, G., Hoffmann, E., Neumeier, E., et al. (1996) Promoter elements in the influenza vRNA terminal structure. *RNA*. 2 (10), 1046–1057.

Flores, J.K. & Ataide, S.F. (2018) Structural Changes of RNA in Complex with Proteins in the SRP. *Frontiers in Molecular Biosciences*. [Online] 5107. Available from: doi:10.3389/fmolb.2018.00007.

Fodor, E., Crow, M., Mingay, L.J., Deng, T., et al. (2002) A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *Journal of Virology*. [Online] 76 (18), 8989–9001. Available from: doi:10.1128/JVI.76.18.8989-9001.2002.

Fodor, E., Devenish, L., Engelhardt, O.G., Palese, P., et al. (1999) Rescue of influenza A virus from recombinant DNA. *Journal of Virology*. 73 (11), 9679–9682.

Fodor, E., Pritlove, D.C. & Brownlee, G.G. (1995) Characterization of the RNA-fork model of virion RNA in the initiation of transcription in influenza A virus. *Journal of Virology*. 69 (7), 4012–4019.

Fodor, E., Pritlove, D.C. & Brownlee, G.G. (1994) The influenza virus panhandle is involved in the initiation of transcription. *Journal of Virology*. 68 (6), 4092–4096.

Foni, E., Chiapponi, C., Baioni, L., Zanni, I., et al. (2017) Influenza D in Italy: towards a better understanding of an emerging viral infection in swine. *Scientific reports*. [Online] 7 (1), 11660. Available from: doi:10.1038/s41598-017-12012-3.

Fournier, E., Moules, V., Essere, B., Paillart, J.-C., et al. (2012a) A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic acids research*. [Online] 40 (5), 2197–2209. Available from: doi:10.1093/nar/gkr985.

Fournier, E., Moules, V., Essere, B., Paillart, J.-C., et al. (2012b) Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine*. [Online] 30 (51), 7359–7367. Available from: doi:10.1016/j.vaccine.2012.09.079.

Fraser, C., Donnelly, C.A., Cauchemez, S., Hanage, W.P., et al. (2009) Pandemic potential of a strain of influenza A (H1N1): early findings. *Science*. [Online] 324 (5934), 1557–1561. Available from: doi:10.1126/science.1176062.

Friebe, P., Boudet, J., Simorre, J.P. & Bartenschlager, R. (2005) Kissing-loop interaction in the 3 ' end of the hepatitis C virus genome essential for RNA replication. *Journal of Virology*. [Online] 79 (1), 380–392. Available from: doi:10.1128/JVI.79.1.380-392.2005.

Friis, M.B., Rasmussen, T.B. & Belsham, G.J. (2012) Modulation of Translation Initiation Efficiency in Classical Swine Fever Virus. *Journal of Virology*. [Online] 86 (16), 8681–8692. Available from: doi:10.1128/JVI.00346-12.

Fujii, K., Fujii, Y., Noda, T., Muramoto, Y., et al. (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. *Journal of Virology*. [Online] 79 (6), 3766–3774. Available from: doi:10.1128/JVI.79.6.3766-3774.2005.

Fujii, Y., Goto, H., Watanabe, T., Yoshida, T., et al. (2003) Selective incorporation of influenza virus RNA segments into virions. *Proceedings of the National Academy of Sciences*. [Online] 100 (4), 2002–2007. Available from: doi:10.1073/pnas.0437772100.

Garmann, R.F., Gopal, A., Athavale, S.S., Knobler, C.M., et al. (2015) Visualizing the global secondary structure of a viral RNA genome with cryo-electron microscopy. *RNA*. [Online] 21 (5), 877–886. Available from: doi:10.1261/rna.047506.114.

Garst, A.D., Edwards, A.L. & Batey, R.T. (2011) Riboswitches: Structures and Mechanisms. *Cold Spring Harbor Perspectives in Biology*. [Online] 3 (6), –a003533. Available from: doi:10.1101/cshperspect.a003533.

Gavazzi, C., Isel, C., Fournier, E., Moules, V., et al. (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 research*. [Online] 41 (2), 1241–1254. Available from: doi:10.1093/nar/gks1181.

Gavazzi, C., Yver, M., Isel, C., Smyth, R.P., et al. (2013) A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 110 (41), 16604–16609. Available from: doi:10.1073/pnas.1314419110.

Gerber, M., Isel, C., Moules, V. & Marquet, R. (2014) Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends in microbiology*. [Online] 22 (8), 446–455. Available from: doi:10.1016/j.tim.2014.04.001.

Giese, S., Bolte, H. & Schwemmle, M. (2016) The Feat of Packaging Eight Unique Genome Segments. *Viruses*. [Online] 8 (6), 165. Available from: doi:10.3390/v8060165.

Gilbertson, B., Zheng, T., Gerber, M., Printz-Schweigert, A., et al. (2016) Influenza NA and PB1 Gene Segments Interact during the Formation of Viral Progeny: Localization of the Binding Region within the PB1 Gene. *Viruses*. [Online] 8 (8), 238. Available from: doi:10.3390/v8080238.

Gog, J.R., Afonso, E.D.S., Dalton, R.M., Leclercq, I., et al. (2007) Codon conservation in the influenza A virus genome defines RNA packaging signals. *Nucleic acids research*. [Online] 35 (6), 1897–1907. Available from: doi:10.1093/nar/gkm087.

Goto, H., Muramoto, Y., Noda, T. & Kawaoka, Y. (2013) The genome-packaging signal of the influenza A virus genome comprises a genome incorporation signal and a genome-bundling signal. *Journal of Virology*. [Online] 87 (21), 11316–11322. Available from: doi:10.1128/JVI.01301-13.

Goubau, D., Schlee, M., Deddouche, S., Pruijssers, A.J., et al. (2014a) Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5'-diphosphates. *Nature*. [Online] 514 (7522), 372–375. Available from: doi:10.1038/nature13590.

Goubau, D., Schlee, M., Deddouche, S., Pruijssers, A.J., et al. (2014b) Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5′-diphosphates. *Nature*. [Online] 514 (7522), 372–375. Available from: doi:10.1038/nature13590.

Gómez-Puertas, P., Albo, C., Pérez-Pastrana, E., Vivo, A., et al. (2000) Influenza virus matrix protein is the major driving force in virus budding. *Journal of Virology*. [Online] 74 (24), 11538–11547. Available from: doi:10.1128/JVI.74.24.11538-11547.2000.

Greenbaum, B.D., Levine, A.J., Bhanot, G. & Rabadán, R. (2008) Patterns of evolution and host gene mimicry in influenza and other RNA viruses. Edward C Holmes (ed.). *PLoS Pathogens*. [Online] 4 (6), e1000079. Available from: doi:10.1371/journal.ppat.1000079.

Gröber, A.S., Vey, M., Angliker, H., Shaw, E., et al. (1992) Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *The EMBO journal*. [Online] 11 (7), 2407–2414. Available from: doi:10.1002/j.1460-2075.1992.tb05305.x.

Guenther, R.H., Sit, T.L., Gracz, H.S., Dolan, M.A., et al. (2004) Structural characterization of an intermolecular RNA-RNA interaction involved in the transcription regulation element of a bipartite plant virus. *Nucleic acids research*. [Online] 32 (9), 2819–2828. Available from: doi:10.1093/nar/gkh585.

Guil, S. & Esteller, M. (2015) RNA-RNA interactions in gene regulation: the coding and noncoding players. *Trends in Biochemical Sciences*. [Online] 40 (5), 248–256. Available from: doi:10.1016/j.tibs.2015.03.001.

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., et al. (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Structural & Molecular Biology*. [Online] 15 (5), 500–506. Available from: doi:10.1038/nsmb.1421.

Gultyaev, A.P., Tsyganov-Bodounov, A., Spronken, M.I.J., van der Kooij, S., et al. (2014) RNA structural constraints in the evolution of the influenza A virus genome NP segment. *RNA Biology*. [Online] 11 (7), 942–952. Available from: doi:10.4161/rna.29730.

Guo, Y.J., Jin, F.G., Wang, P., Wang, M., et al. (1983) Isolation of influenza C virus from pigs and experimental infection of pigs with influenza C virus. *The Journal of general virology*. [Online] 64 (Pt 1) (1), 177–182. Available from: doi:10.1099/0022-1317-64-1-177.

Hara, K., Nakazono, Y., Kashiwagi, T., Hamada, N., et al. (2013) 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. *Journal of General Virology*. [Online] 94 (Pt 11), 2406–2416. Available from: doi:10.1099/vir.0.053959-0.

Hatta, M., Halfmann, P., Wells, K. & Kawaoka, Y. (2002) Human influenza a viral genes responsible for the restriction of its replication in duck intestine. *Virology*. [Online] 295 (2), 250–255. Available from: doi:10.1006/viro.2002.1358.

He, L. & Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nature reviews. Genetics*. [Online] 5 (7), 522–531. Available from: doi:10.1038/nrg1379.

Hengrung, N., Omari, El, K., Martin, I.S., Vreede, F.T., et al. (2015) Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature*. [Online] 527 (7576), 114–117. Available from: doi:10.1038/nature15525.

Herz, C., Stavnezer, E., Krug, R.M. & Gurney, T., Jr. (1981) Influenza virus, an RNA virus, synthesizes its messenger RNA in the nucleus of infected cells. *Cell*. [Online] 26 (3), 391–400. Available from: doi:10.1016/0092-8674(81)90208-7.

Hinshaw, V.S., Bean, W.J., Geraci, J., Fiorelli, P., et al. (1986) Characterization of two influenza A viruses from a pilot whale. *Journal of Virology*. 58 (2), 655–656.

Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G., et al. (2000) A DNA transfection system for generation of influenza A virus from eight plasmids. *Proceedings of the National Academy of Sciences*. [Online] 97 (11), 6108–6113. Available from: doi:10.1073/pnas.100133697.

Hooker, L., Sully, R., Handa, B., Ono, N., et al. (2003) Quantitative analysis of influenza virus RNP interaction with RNA cap structures and comparison to human cap binding protein eIF4E. *Biochemistry*. [Online] 42 (20), 6234–6240. Available from: doi:10.1021/bi027081r.

Hornung, V., Ellegast, J., Kim, S., Brzozka, K., et al. (2006) 5 '-triphosphate RNA is the ligand for RIG-I. *Science*. [Online] 314 (5801), 994–997. Available from: doi:10.1126/science.1132505.

Hsu, M.T., Parvib, J.D., Gupta, S., Krystal, M., et al. (1987) Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. Proceedings of the National Academy of Sciences. 84 (22), 8140–8144.

Hu, Y., Gor, V., Morikawa, K., Nagata, K., et al. (2017) Cellular splicing factor UAP56 stimulates trimeric NP formation for assembly of functional influenza viral ribonucleoprotein complexes. *Scientific reports*. [Online] 7 (1), 14053. Available from: doi:10.1038/s41598-017-13784-4.

Hutchinson, E. & Fodor, E. (2013) *Transport of the Influenza Virus Genome from Nucleus to Nucleus*. [Online] 5 (10), 2424–2446. Available from: doi:10.3390/v5102424.

Hutchinson, E.C., Charles, P.D., Hester, S.S., Thomas, B., et al. (2014) Conserved and host-specific features of influenza virion architecture. *Nature communications*. [Online] 54816. Available from: doi:10.1038/ncomms5816.

Hutchinson, E.C., Curran, M.D., Read, E.K., Gog, J.R., et al. (2008) Mutational analysis of cis-acting RNA signals in segment 7 of influenza A virus. *Journal of Virology*. [Online] 82 (23), 11869–11879. Available from: doi:10.1128/JVI.01634-08.

Hutchinson, E.C., Kirchbach, von, J.C., Gog, J.R. & Digard, P. (2010) Genome packaging in influenza A virus. *Journal of General Virology*. [Online] 91 (2), 313–328. Available from: doi:10.1099/vir.0.017608-0.

Hutchinson, E.C., Wise, H.M., Kudryavtseva, K., Curran, M.D., et al. (2009) Characterisation of influenza A viruses with mutations in segment 5 packaging signals. *Vaccine*. [Online] 27 (45), 6270–6275. Available from: doi:10.1016/j.vaccine.2009.05.053.

Isaacs, A. & Lindenmann, J. (1957) Virus interference. I. The interferon. *Journal of Immunology*. 195 (5), 1911–1920.

Ito, T., Couceiro, J., Kelm, S., Baum, L.G., et al. (1998) Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. *Journal of Virology*. 72 (9), 7367–7373.

Iwasaki, A. & Medzhitov, R. (2004) Toll-like receptor control of the adaptive immune responses. *Nature Immunology*. [Online] 5 (10), 987–995. Available from: doi:10.1038/ni1112.

Jagger, B.W., Wise, H.M., Kash, J.C., Walters, K.A., et al. (2012) An Overlapping Protein-Coding Region in Influenza A Virus Segment 3 Modulates the Host Response. *Science*. [Online] 337 (6091), 199–204. Available from: doi:10.1126/science.1222213.

Jensen, S. & Thomsen, A.R. (2012) Sensing of RNA Viruses: a Review of Innate Immune Receptors Involved in Recognizing RNA Virus Invasion. *Journal of Virology*. [Online] 86 (6), 2900–2910. Available from: doi:10.1128/JVI.05738-11.

Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC bioinformatics*. [Online] 15 (1), 182. Available from: doi:10.1186/1471-2105-15-182.

Jorba, N., Coloma, R. & Ortin, J. (2009) Genetic trans-Complementation Establishes a New Model for Influenza Virus RNA Transcription and Replication Sean P J Whelan (ed.). *PLoS Pathogens*. [Online] 5 (5), e1000462. Available from: doi:10.1371/journal.ppat.1000462.

Kaverin, N.V., Gambaryan, A.S., Bovin, N.V., Rudneva, I.A., et al. (1998) Postreassortment changes in influenza A virus hemagglutinin restoring HA-NA functional match. *Virology*. [Online] 244 (2), 315–321. Available from: doi:10.1006/viro.1998.9119.

Kawaguchi, A., Matsumoto, K. & Nagata, K. (2012) YB-1 Functions as a Porter To Lead Influenza Virus Ribonucleoprotein Complexes to Microtubules. *Journal of Virology*. [Online] 86 (20), 11086–11095. Available from: doi:10.1128/JVI.00453-12.

Kawaguchi, A., Momose, F. & Nagata, K. (2011) Replication-coupled and host factor-mediated encapsidation of the influenza virus genome by viral nucleoprotein. *Journal of Virology*. [Online] 85 (13), JVI.00277–11–6204. Available from: doi:10.1128/JVI.00277-11.

Kawaoka, Y., Krauss, S. & Webster, R.G. (1989) Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. *Journal of Virology*. 63 (11), 4603–4608.

Kenyon, J.C., Prestwood, L.J. & Lever, A.M.L. (2015) A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Scientific reports*. [Online] 5 (1). Available from: doi:10.1038/srep14369.

Kenyon, J.C., Prestwood, L.J., Le Grice, S.F.J. & Lever, A.M.L. (2013) In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader. *Nucleic acids research*. [Online] 41 (18), –e174. Available from: doi:10.1093/nar/gkt690.

Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., et al. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*. [Online] 467 (7311), 103–107. Available from: doi:10.1038/nature09322.

Kerviel, A., Dash, S., Moncorgé, O., Panthu, B., et al. (2016) Involvement of an Arginine Triplet in M1 Matrix Protein Interaction with Membranes and in M1 Recruitment into Virus-Like Particles of the Influenza A(H1N1)pdm09 Virus Andrew Pekosz (ed.). *PLoS ONE*. [Online] 11 (11), e0165421. Available from: doi:10.1371/journal.pone.0165421.

Kim, D.Y., Firth, A.E., Atasheva, S., Frolova, E.I., et al. (2011) Conservation of a Packaging Signal and the Viral Genome RNA Packaging Mechanism in Alphavirus Evolution. *Journal of Virology*. [Online] 85 (16), 8022–8036. Available from: doi:10.1128/JVI.00644-11.

Kladwang, W., Hum, J. & Das, R. (2012) Ultraviolet shadowing of RNA can cause significant chemical damage in seconds. *Scientific reports*. [Online] 2 (1), 517. Available from: doi:10.1038/srep00517.

Klumpp, K., Ruigrok, R.W. & Baudin, F. (1997) Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. *The EMBO journal*. [Online] 16 (6), 1248–1257. Available from: doi:10.1093/emboj/16.6.1248.

Kobayashi, Y., Dadonaite, B., van Doremalen, N., Suzuki, Y., et al. (2016) Computational and molecular analysis of conserved influenza A virus RNA secondary structures involved in infectious virion production. *RNA Biology*. [Online] 13 (9), 883–894. Available from: doi:10.1080/15476286.2016.1208331.

Kortmann, J. & Narberhaus, F. (2012) Bacterial RNA thermometers: molecular zippers and switches. *Nature reviews. Microbiology*. [Online] 10 (4), 255–265. Available from: doi:10.1038/nrmicro2730.

Krzywinski, M., Schein, J., Birol, I., Connors, J., et al. (2009) Circos: an information aesthetic for comparative genomics. *Genome research*. [Online] 19 (9), 1639–1645. Available from: doi:10.1101/gr.092759.109.

Kutchko, K.M., Madden, E.A., Morrison, C., Plante, K.S., et al. (2018) Structural divergence creates new functional features in alphavirus genomes. *Nucleic acids research*. [Online] 46 (7), 3657–3670. Available from: doi:10.1093/nar/gky012.

Kwok, C.K., Ding, Y., Tang, Y., Assmann, S.M., et al. (2013) Determination of in vivo RNA structure in low-abundance transcripts. *Nature communications*. [Online] 4 (1). Available from: doi:10.1038/ncomms3971.

la Luna, de, S., Fortes, P., Beloso, A. & Ortin, J. (1995) Influenza virus NS1 protein enhances the rate of translation initiation of viral mRNAs. *Journal of Virology*. 69 (4), 2427–2433.

Lakdawala, S.S., Fodor, E. & Subbarao, K. (2016) Moving On Out: Transport and Packaging of Influenza Viral RNA into Virions. *Annual Review of Virology, Vol 3*. [Online] 3 (1), 411–427. Available from: doi:10.1146/annurev-virology-110615-042345.

Lakdawala, S.S., Wu, Y., Wawrzusin, P., Kabat, J., et al. (2014) Influenza a virus assembly intermediates fuse in the cytoplasm. Mark T Heise (ed.). *PLoS Pathogens*. [Online] 10 (3), e1003971. Available from: doi:10.1371/journal.ppat.1003971.

Larman, B.C., Dethoff, E.A. & Weeks, K.M. (2017) Packaged and Free Satellite Tobacco Mosaic Virus (STMV) RNA Genomes Adopt Distinct Conformational States. *Biochemistry*. [Online] 56 (16), 2175–2183. Available from: doi:10.1021/acs.biochem.6b01166.

Larson, S.B. & McPherson, A. (2001) Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Current Opinion in Structural Biology*. [Online] 11 (1), 59–65. Available from: doi:10.1016/S0959-440X(00)00166-4.

Lee, N., Le Sage, V., Nanni, A.V., Snyder, D.J., et al. (2017) Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic acids research*. [Online] 45 (15), 8968–8977. Available from: doi:10.1093/nar/gkx584.

Lemaire, P.A., Anderson, E., Lary, J. & Cole, J.L. (2008) Mechanism of PKR activation by dsRNA. *Journal of molecular biology*. [Online] 381 (2), 351–360. Available from: doi:10.1016/j.jmb.2008.05.056.

Lenartowicz, E., Kesy, J., Ruszkowska, A., Soszynska-Jozwiak, M., et al. (2016) Self-Folding of Naked Segment 8 Genomic RNA of Influenza A Virus. Adrianus CM Boon (ed.). *PLoS ONE*. [Online] 11 (2), e0148281. Available from: doi:10.1371/journal.pone.0148281.

Levitt M. (1969) Detailed Molecular Model for Transfer Ribonucleic Acid. *Nature*. [Online] 224 (5221), 759–&. Available from: doi:10.1038/224759a0.

Lewis, P.A. & Shope, R.E. (1931) SWINE INFLUENZA : II. A HEMOPHILIC BACILLUS FROM THE RESPIRATORY TRACT OF INFECTED SWINE. *The Journal of experimental medicine*. 54 (3), 361–371.

Li, C., Hatta, M., Watanabe, S., Neumann, G., et al. (2008) Compatibility among polymerase subunit proteins is a restricting factor in reassortment between equine H7N7 and human H3N2 influenza viruses. *Journal of Virology*. [Online] 82 (23), 11880–11888. Available from: doi:10.1128/JVI.01445-08.

Liang, Y., Hong, Y. & Parslow, T.G. (2005) cis-Acting Packaging Signals in the Influenza Virus PB1, PB2, and PA Genomic RNA Segments. *Journal of Virology*. [Online] 79 (16), 10348–10355. Available from: doi:10.1128/JVI.79.16.10348-10355.2005.

Liang, Y., Huang, T., Ly, H., Parslow, T.G., et al. (2008) Mutational analyses of packaging signals in influenza virus PA, PB1, and PB2 genomic RNA segments. *Journal of Virology*. [Online] 82 (1), 229–236. Available from: doi:10.1128/JVI.01541-07.

Lina, B. (2008) History of Influenza Pandemics. In: *Paleomicrobiology*. [Online]. Berlin, Heidelberg, Springer Berlin Heidelberg. pp. 199–211. Available from: doi:10.1007/978-3-540-75855-6_12.

Lindenbach, B.D., Sgro, J.Y. & Ahlquist, P. (2002) Long-distance base pairing in flock house virus RNA1 regulates subgenomic RNA3 synthesis and RNA2 replication. *Journal of Virology*. [Online] 76 (8), 3905–3919. Available from: doi:10.1128/JVI.76.8.3905-3919.2002.

Liu, C. & Air, G.M. (1993) Selection and Characterization of a Neuraminidase-Minus Mutant of Influenza Virus and Its Rescue by Cloned Neuraminidase Genes. *Virology*. [Online] 194 (1), 403–407. Available from: doi:10.1006/viro.1993.1276.

Long, J.S., Giotis, E.S., Moncorgé, O., Frise, R., et al. (2016) Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature*. [Online] 529 (7584), 101–104. Available from: doi:10.1038/nature16474.

Lorenz, R., Wolfinger, M.T., Tanzer, A. & Hofacker, I.L. (2016) Predicting RNA secondary structures from sequence and probing data. *Methods*. [Online] 10386–98. Available from: doi:10.1016/j.ymeth.2016.04.004.

Low, J.T. & Weeks, K.M. (2010) SHAPE-directed RNA secondary structure prediction. *Methods*. [Online] 52 (2), 150–158. Available from: doi:10.1016/j.ymeth.2010.06.007.

Lowen, A.C. (2017) Constraints, Drivers, and Implications of Influenza A Virus Reassortment. *Annual Review of Virology, Vol 3*. [Online] 4 (1), 105–121. Available from: doi:10.1146/annurev-virology-101416-041726.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., et al. (2016) RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell*. [Online] 165 (5), 1267–1279. Available from: doi:10.1016/j.cell.2016.04.028.

Lukarska, M., Fournier, G., Pflug, A., Resa-Infante, P., et al. (2017) Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature*. [Online] 541 (7635), 117–121. Available from: doi:10.1038/nature20594.

Luque, D., Rivas, G., Alfonso, C., Carrascosa, J.L., et al. (2009) Infectious bursal disease virus is an icosahedral polyploid dsRNA virus. *Proceedings of the National Academy of Sciences*. [Online] 106 (7), 2148–2152. Available from: doi:10.1073/pnas.0808498106.

Luytjes, W., Krystal, M., Enami, M., Parvin, J.D., et al. (1989) Amplification, expression, and packaging of a foreign gene by influenza virus. *Cell*. [Online] 59 (6), 1107–1113. Available from: doi:10.1016/0092-8674(89)90766-6.

Ma, W., Kahn, R.E. & Richt, J.A. (2008) The pig as a mixing vessel for influenza viruses: Human and veterinary implications. *Journal of molecular and genetic medicine : an international journal of biomedical research*. 3 (1), 158–166.

Maas, R., Tacken, M., Ruuls, L., Koch, G., et al. (2007) Avian Influenza (H5N1) Susceptibility and Receptors in Dogs. *Emerging Infectious Diseases*. [Online] 13 (8), 1219–1221. Available from: doi:10.3201/eid1308.070393.

Madison, J.T., Everett, G.A. & Kung, H. (1966) Nucleotide Sequence of a Yeast Tyrosine Transfer Rna. *Science*. [Online] 153 (3735), 531–&. Available from: doi:10.1126/science.153.3735.531.

Maeda, T. & Ohnishi, S.-I. (1980) Activation of influenza virus by acidic media causes hemolysis and fusion of erythrocytes. *FEBS Letters*. [Online] 122 (2), 283–287. Available from: doi:10.1016/0014-5793(80)80457-1.

Mandal, M. & Breaker, R.R. (2004) Gene regulation by riboswitches. *Nature reviews. Molecular cell biology*. [Online] 5 (6), 451–463. Available from: doi:10.1038/nrm1403.

Mann, M., Wright, P.R. & Backofen, R. (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic acids research*. [Online] 45 (W1), W435–W439. Available from: doi:10.1093/nar/gkx279.

Marc, D., Barbachou, S. & Soubieux, D. (2013) The RNA-binding domain of influenzavirus non-structural protein-1 cooperatively binds to virus-specific RNA sequences in a structure-dependent manner. *Nucleic acids research*. [Online] 41 (1), 434–449. Available from: doi:10.1093/nar/gks979.

Marsh, G.A., Rabadán, R., Levine, A.J. & Palese, P. (2008) Highly conserved regions of influenza a virus polymerase gene segments are critical for efficient viral RNA packaging. *Journal of Virology*. [Online] 82 (5), 2295–2304. Available from: doi:10.1128/JVI.02267-07.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. [Online] 17 (1), 10–12. Available from: doi:10.14806/ej.17.1.200.

Martini, I.F. de C., Fournier, G., Sachse, M., Pizarro-Cerda, J., et al. (2017) Influenza virus genome reaches the plasma membrane via a modified endoplasmic reticulum and Rab11-dependent vesicles. *Nature communications*. [Online] 8 (1). Available from: doi:10.1038/s41467-017-01557-6.

Martín-Benito, J., Area, E., Ortega, J., Llorca, O., et al. (2001) Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle. *EMBO reports*. [Online] 2 (4), 313–317. Available from: doi:10.1093/embo-reports/kve063.

Martínez-Alonso, M., Hengrung, N. & Fodor, E. (2016) RNA-free and ribonucleoprotein-associated influenza virus polymerases directly bind the serine-5 phosphorylated carboxyl-terminal domain of host RNA polymerase II S Schultz-Cherry (ed.). *Journal of Virology*. [Online] 90 (13), JVI.00494–16–6021. Available from: doi:10.1128/JVI.00494-16.

Martyna, A., Bahsoun, B., Badham, M.D., Srinivasan, S., et al. (2017) Membrane remodeling by the M2 amphipathic helix drives influenza virus membrane scission. *Scientific reports*. [Online] 7 (1), 44695. Available from: doi:10.1038/srep44695.

Massin, P., Van der Werf, S. & Naffakh, N. (2001) Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. *Journal of Virology*. [Online] 75 (11), 5398–5404. Available from: doi:10.1128/JVI.75.11.5398-5404.2001.

MASUREL, N. & MARINE, W.M. (1973) Recycling of Asian and Hong-Kong Influenza a Virus Hemagglutinins in Man. *American Journal of Epidemiology*. 97 (1), 44–49.

Mathews, D.H. (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*. [Online] 10 (8), 1178–1190. Available from: doi:10.1261/rna.7650904.

Mathews, D.H., Moss, W.N. & Turner, D.H. (2010) Folding and Finding RNA Secondary Structure. *Cold Spring Harbor Perspectives in Biology*. [Online] 2 (12), a003665–a003665. Available from: doi:10.1101/cshperspect.a003665.

Matsuzaki, Y., Sugawara, K., Furuse, Y., Shimotai, Y., et al. (2016) Genetic lineage and reassortment of influenza C viruses circulating between 1947 and 2014 A Garcia-Sastre (ed.). *Journal of Virology*. [Online] 90 (18), JVI.00969–16–8265. Available from: doi:10.1128/JVI.00969-16.

Mauger, D.M., Golden, M., Yamane, D., Williford, S., et al. (2015) Functionally conserved architecture of hepatitis C virus RNA genomes. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 112 (12), 3692–3697. Available from: doi:10.1073/pnas.1416266112.

McCaskill, J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*. [Online] 29 (6-7), 1105–1119. Available from: doi:10.1002/bip.360290621.

McCown, M.F. & Pekosz, A. (2006) Distinct Domains of the Influenza A Virus M2 Protein Cytoplasmic Tail Mediate Binding to the M1 Protein and Facilitate Infectious Virus Production. *Journal of Virology*. [Online] 80 (16), 8178–8189. Available from: doi:10.1128/JVI.00627-06.

McDonald, S.M. & Patton, J.T. (2011) Assortment and packaging of the segmented rotavirus genome. *Trends in microbiology*. [Online] 19 (3), 136–144. Available from: doi:10.1016/j.tim.2010.12.002.

McGeoch, D., Fellner, P. & Newton, C. (1976) Influenza virus genome consists of eight distinct RNA species. *Proceedings of the National Academy of Sciences*. [Online] 73 (9), 3045–3049. Available from: doi:10.1073/pnas.73.9.3045.

McGinnis, J.L., Dunkle, J.A., Cate, J.H.D. & Weeks, K.M. (2012) The mechanisms of RNA SHAPE chemistry. *Journal of the American Chemical Society*. [Online] 134 (15), 6617–6624. Available from: doi:10.1021/ja2104075.

Mehle, A. (2014) Unusual Influenza A Viruses in Bats. *Viruses*. [Online] 6 (9), 3438–3449. Available from: doi:10.3390/v6093438.

Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. (2005) RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *Journal of the American Chemical Society*. [Online] 127 (12), 4223–4231. Available from: doi:10.1021/ja043822v.

Mitnaul, L.J., Castrucci, M.R., Murti, K.G. & Kawaoka, Y. (1996) The cytoplasmic tail of influenza A virus neuraminidase (NA) affects NA incorporation into virions, virion morphology, and virulence in mice but is not essential for virus replication. *Journal of Virology*. 70 (2), 873–879.

Mitnaul, L.J., Matrosovich, M.N., Castrucci, M.R., Tuzikov, A.B., et al. (2000) Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. *Journal of Virology*. 74 (13), 6015–6020.

Moeller, A., Kirchdoerfer, R.N., Potter, C.S., Carragher, B., et al. (2012) Organization of the influenza virus replication machinery. *Science*. [Online] 338 (6114), 1631–1634. Available from: doi:10.1126/science.1227270.

Momose, F., Kikuchi, Y., Komase, K. & Morikawa, Y. (2007) Visualization of microtubule-mediated transport of influenza viral progeny ribonucleoprotein. *Microbes and Infection*. [Online] 9 (12-13), 1422–1433. Available from: doi:10.1016/j.micinf.2007.07.007.

Momose, F., Naito, T., Yano, K., Sugimoto, S., et al. (2002) Identification of Hsp90 as a stimulatory host factor involved in influenza virus RNA synthesis. *Journal of Biological Chemistry*. [Online] 277 (47), 45306–45314. Available from: doi:10.1074/jbc.M206822200.

Momose, F., Sekimoto, T., Ohkura, T., Jo, S., et al. (2011) Apical Transport of Influenza A Virus Ribonucleoprotein Requires Rab11-positive Recycling Endosome Jean-Luc E P H Darlix (ed.). *PLoS ONE*. [Online] 6 (6), e21123. Available from: doi:10.1371/journal.pone.0021123.

Moreira, É.A., Weber, A., Bolte, H., Kolesnikova, L., et al. (2016) A conserved influenza A virus nucleoprotein code controls specific viral genome packaging. *Nature communications*. [Online] 712861. Available from: doi:10.1038/ncomms12861.

Morens, D.M. & Taubenberger, J.K. (2010) An avian outbreak associated with panzootic equine influenza in 1872: an early example of highly pathogenic avian influenza? *Influenza and Other Respiratory Viruses*. [Online] 4 (6), 373–377. Available from: doi:10.1111/j.1750-2659.2010.00181.x.

Morens, D.M., Taubenberger, J.K. & Fauci, A.S. (2009) The persistent legacy of the 1918 influenza virus. *New England Journal of Medicine*. [Online] 361 (3), 225–229. Available from: doi:10.1056/NEJMp0904819.

Mortimer, S.A., Kidwell, M.A. & Doudna, J.A. (2014) Insights into RNA structure and function from genome-wide studies. *Nature reviews. Genetics*. [Online] 15 (7), 469–479. Available from: doi:10.1038/nrg3681.

Moss, W.N., Dela-Moss, L.I., Kierzek, E., Kierzek, R., et al. (2012) The 3′ Splice Site of Influenza A Segment 7 mRNA Can Exist in Two Conformations: A Pseudoknot and a Hairpin Alfred Lewin (ed.). *PLoS ONE*. [Online] 7 (6), e38323. Available from: doi:10.1371/journal.pone.0038323.

Moss, W.N., Priore, S.F. & Turner, D.H. (2011) Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA*. [Online] 17 (6), 991–1011. Available from: doi:10.1261/rna.2619511.

Munster, V.J. & Fouchier, R.A.M. (2009) Avian influenza virus: Of virus and bird ecology. *Vaccine*. [Online] 27 (45), 6340–6344. Available from: doi:10.1016/j.vaccine.2009.02.082.

Muramoto, Y., Takada, A., Fujii, K., Noda, T., et al. (2006) Hierarchy among viral RNA (vRNA) segments in their role in vRNA incorporation into influenza A virions. *Journal of Virology*. [Online] 80 (5), 2318–2325. Available from: doi:10.1128/JVI.80.5.2318-2325.2006.

Nakatsu, S., Murakami, S., Shindo, K., Horimoto, T., et al. (2018) Influenza C and D Viruses Package Eight Organized Ribonucleoprotein Complexes. Terence S Dermody (ed.). *Journal of Virology*. [Online] 92 (6), e02084–17. Available from: doi:10.1128/JVI.02084-17.

Nayak, D.P., Hui, E.K.-W. & Barman, S. (2004) Assembly and budding of influenza virus. *Virus Research*. [Online] 106 (2), 147–165. Available from: doi:10.1016/j.virusres.2004.08.012.

Nelli, R.K., Kuchipudi, S.V., White, G.A., Perez, B.B., et al. (2010) Comparative distribution of human and avian type sialic acid influenza receptors in the pig. *Bmc Veterinary Research*. [Online] 6 (1). Available from: doi:10.1186/1746-6148-6-4.

Neumann, G. & Hobom, G. (1995) Mutational analysis of influenza virus promoter elements in vivo. *Journal of General Virology*. [Online] 76 (7), 1709–1717. Available from: doi:10.1099/0022-1317-76-7-1709.

Neumann, G., Castrucci, M.R. & Kawaoka, Y. (1997) Nuclear import and export of influenza virus nucleoprotein. *Journal of Virology*. 71 (12), 9690–9700.

Neumann, G., Noda, T. & Kawaoka, Y. (2009) Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature*. [Online] 459 (7249), 931–939. Available from: doi:10.1038/nature08157.

Ng, A.K.-L., Lam, M.K.-H., Zhang, H., Liu, J., et al. (2012) Structural basis for RNA-binding and homo-oligomer formation by influenza B virus nucleoprotein. *Journal of Virology*. [Online] 86 (12), JVI.00073–12–6767. Available from: doi:10.1128/JVI.00073-12.

Ng, A.K.-L., Zhang, H., Tan, K., Li, Z., et al. (2008) Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. [Online] 22 (10), 3638–3647. Available from: doi:10.1096/fj.08-112110.

Nguyen, T.C., Cao, X., Yu, P., Xiao, S., et al. (2016) Mapping RNA-RNA interactome and RNA structure in vivo by MARIO. *Nature communications*. [Online] 712023. Available from: doi:10.1038/ncomms12023.

Nicholson, B.L. & White, K.A. (2015) Exploring the architecture of viral RNA genomes. *Current Opinion in Virology*. [Online] 1266–74. Available from: doi:10.1016/j.coviro.2015.03.018.

Noda, T. (2012) Native Morphology of Influenza Virions. *Frontiers in Microbiology*. [Online] 2. Available from: doi:10.3389/fmicb.2011.00269.

Noda, T., Murakami, S., Nakatsu, S., Imai, H., et al. (2018) Importance of the 1+7 configuration of ribonucleoprotein complexes for influenza A virus genome packaging. *Nature communications*. [Online] 9 (1), 54. Available from: doi:10.1038/s41467-017-02517-w.

Noda, T., Sagara, H., Yen, A., Takada, A., et al. (2006) Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature*. [Online] 439 (7075), 490–492. Available from: doi:10.1038/nature04378.

Noda, T., Sugita, Y., Aoyama, K., Hirase, A., et al. (2012) Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nature communications*. [Online] 3 (1), 639. Available from: doi:10.1038/ncomms1647.

Noller, H.F. & Woese, C.R. (1981) Secondary Structure of 16s-Ribosomal Rna. *Science*. [Online] 212 (4493), 403–411. Available from: doi:10.1126/science.6163215.

O'Neill, R.E., Jaskunas, R., Blobel, G., Palese, P., et al. (1995) Nuclear import of influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import. *Journal of Biological Chemistry*. [Online] 270 (39), 22701–22704. Available from: doi:10.1074/jbc.270.39.22701.

O'Neill, R.E., Talon, J. & Palese, P. (1998) The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *The EMBO journal*. [Online] 17 (1), 288–296. Available from: doi:10.1093/emboj/17.1.288.

Octaviani, C.P., Goto, H. & Kawaoka, Y. (2011) Reassortment between seasonal H1N1 and pandemic (H1N1) 2009 influenza viruses is restricted by limited compatibility among polymerase subunits. *Journal of Virology*. [Online] 85 (16), 8449–8452. Available from: doi:10.1128/JVI.05054-11.

Odagiri, T. & Tashiro, M. (1997) Segment-specific noncoding sequences of the influenza virus genome RNA are involved in the specific competition between defective interfering RNA and its progenitor RNA segment at the virion assembly step. *Journal of Virology*. 71 (3), 2138–2145.

OED, O.E.D. (2017) *"influenza, n."*. [Online]. Available from: http://www.oed.com/view/Entry/95531?redirectedFrom=influenza ER - [Accessed: 7 August 2018].

Ohkura, T., Momose, F., Ichikawa, R., Takeuchi, K., et al. (2014) Influenza A virus hemagglutinin and neuraminidase mutually accelerate their apical targeting through clustering of lipid rafts. *Journal of Virology*. [Online] 88 (17), JVI.00586–14–10055. Available from: doi:10.1128/JVI.00586-14.

Ortega, J., Martín-Benito, J., Zürcher, T., Valpuesta, J.M., et al. (2000) Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification. *Journal of Virology*. [Online] 74 (1), 156–163. Available from: doi:10.1128/JVI.74.1.156-163.2000.

Oxford, J.S. & Gill, D. (2018) Unanswered questions about the 1918 influenza pandemic: origin, pathology, and the virus itself. *The Lancet Infectious Diseases*. [Online] Available from: doi:10.1016/S1473-3099(18)30359-1.

Ozawa, M., Maeda, J., Iwatsuki-Horimoto, K., Watanabe, S., et al. (2009) Nucleotide sequence requirements at the 5' end of the influenza A virus M RNA segment for efficient virus replication. *Journal of Virology*. [Online] 83 (7), 3384–3388. Available from: doi:10.1128/JVI.02513-08.

Park, Y.W. & Katze, M.G. (1995) Translational control by influenza virus. Identification of cis-acting sequences and trans-acting factors which may regulate selective viral mRNA translation. *Journal of Biological Chemistry*. 270 (47), 28433–28439.

Paterson, D. & Fodor, E. (2012) Emerging Roles for the Influenza A Virus Nuclear Export Protein (NEP) Tom C Hobman (ed.). *PLoS Pathogens*. [Online] 8 (12), e1003019. Available from: doi:10.1371/journal.ppat.1003019.

Paules, C.I., Sullivan, S.G., Subbarao, K. & Fauci, A.S. (2018) Chasing Seasonal Influenza - The Need for a Universal Influenza Vaccine. *New England Journal of Medicine*. [Online] 378 (1), 7–9. Available from: doi:10.1056/NEJMp1714916.

Pauly, M.D., Procario, M.C. & Lauring, A.S. (2017) A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. *eLife*. [Online] 6. Available from: doi:10.7554/eLife.26437.

Pelletier, J. & Sonenberg, N. (1988) Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*. [Online] 334 (6180), 320–325. Available from: doi:10.1038/334320a0.

Pereira, C.F., Read, E.K.C., Wise, H.M., Amorim, M.J., et al. (2017) Influenza A Virus NS1 Protein Promotes Efficient Nuclear Export of Unspliced Viral M1 mRNA. *Journal of Virology*. [Online] 91 (15). Available from: doi:10.1128/JVI.00528-17.

Pfeiffer, R. (1892) Vorläufige Mittheilungen über die Erreger der Influenza. *DMW - Deutsche Medizinische Wochenschrift*. [Online] 18 (02), 28–28. Available from: doi:10.1055/s-0029-1198870.

Pflug, A., Guilligay, D., Reich, S. & Cusack, S. (2014) Structure of influenza A polymerase bound to the viral RNA promoter. *Nature*. [Online] 516 (7531), 355–. Available from: doi:10.1038/nature14008.

Pflug, A., Lukarska, M., Resa-Infante, P., Reich, S., et al. (2017) Structural insights into RNA synthesis by the influenza virus transcription-replication machine. *Virus Research*. [Online] 234103–117. Available from: doi:10.1016/j.virusres.2017.01.013.

Pichlmair, A., Schulz, O., Tan, C.P., Rehwinkel, J., et al. (2009) Activation of MDA5 Requires Higher-Order RNA Structures Generated during Virus Infection. *Journal of Virology*. [Online] 83 (20), 10761–10769. Available from: doi:10.1128/JVI.00770-09.

Pinto, L.H., Holsinger, L.J. & Lamb, R.A. (1992) Influenza virus M2 protein has ion channel activity. *Cell*. [Online] 69 (3), 517–528. Available from: doi:10.1016/0092-8674(92)90452-I.

Plotch, S.J., Bouloy, M., Ulmanen, I. & Krug, R.M. (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell*. [Online] 23 (3), 847–858. Available from: doi:10.1016/0092-8674(81)90449-9.

Pollom, E., Dang, K.K., Potter, E.L., Gorelick, R.J., et al. (2013) Comparison of SIV and HIV-1 Genomic RNA Structures Reveals Impact of Sequence Evolution on Conserved and Non-Conserved Structural Motifs Bryan R Cullen (ed.). *PLoS Pathogens*. [Online] 9 (4). Available from: doi:10.1371/journal.ppat.1003294.

Poole, E., Elton, D., Medcalf, L. & Digard, P. (2004) Functional domains of the influenza A virus PB2 protein: identification of NP- and PB1-binding sites. *Virology*. [Online] 321 (1), 120–133. Available from: doi:10.1016/j.virol.2003.12.022.

Priore, S.F., Kauffmann, A.D., Baman, J.R. & Turner, D.H. (2015) The Influenza A PB1-F2 and N40 Start Codons Are Contained within an RNA Pseudoknot. *Biochemistry*. [Online] 54 (22), 3413–3415. Available from: doi:10.1021/bi501564d.

Priore, S.F., Moss, W.N. & Turner, D.H. (2013) Influenza B virus has global ordered RNA structure in (+) and (–) strands but relatively less stable predicted RNA folding free energy than allowed by the encoded protein sequence. *BMC Research Notes*. [Online] 6 (1), 330. Available from: doi:10.1186/1756-0500-6-330.

Rabadán, R., Levine, A.J. & Robins, H. (2006) Comparison of avian and human influenza A viruses reveals a mutational bias on the viral genomes. *Journal of Virology*. [Online] 80 (23), 11887–11891. Available from: doi:10.1128/JVI.01414-06.

Read, E.K.C. & Digard, P. (2010) Individual influenza A virus mRNAs show differential dependence on cellular NXF1/TAP for their nuclear export. *The Journal of general virology*. [Online] 91 (5), 1290–1301. Available from: doi:10.1099/vir.0.018564-0.

Redondo, N., Angel Sanz, M., Steinberger, J., Skern, T., et al. (2012) Translation Directed by Hepatitis A Virus IRES in the Absence of Active eIF4F Complex and eIF2 Eric Jan (ed.). *PLoS ONE*. [Online] 7 (12). Available from: doi:10.1371/journal.pone.0052065.

Reguera, J., Cusack, S. & Kolakofsky, D. (2014) Segmented negative strand RNA virus nucleoprotein structure. *Current Opinion in Virology*. [Online] 57–15. Available from: doi:10.1016/j.coviro.2014.01.003.

Reich, S., Guilligay, D., Pflug, A., Malet, H., et al. (2014) Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature*. [Online] 516 (7531), 361–366. Available from: doi:10.1038/nature14009.

Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., et al. (1974) Structure of yeast phenylalanine tRNA at 3 A resolution. *Nature*. [Online] 250 (467), 546–551. Available from: doi:10.1038/250546a0.

Rudneva, I.A., Kovaleva, V.P., Varich, N.L., Farashyan, V.R., et al. (1993) Influenza A virus reassortants with surface glycoprotein genes of the avian parent viruses: effects of HA and NA gene combinations on virus aggregation. *Archives of Virology*. 133 (3-4), 437–450.

Rudneva, I.A., Sklyanskaya, E.I., Barulina, O.S., Yamnikova, S.S., et al. (1996) Phenotypic expression of HA-NA combinations in human-avian influenza A virus reassortants. *Archives of Virology*. 141 (6), 1091–1099.

Ruigrok, R.W.H., Barge, A., Durrer, P., Brunner, J., et al. (2000) Membrane Interaction of Influenza Virus M1 Protein. *Virology*. [Online] 267 (2), 289–298. Available from: doi:10.1006/viro.1999.0134.

Ruigrok, R.W.H., Crépin, T. & Kolakofsky, D. (2011) Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Current Opinion in Microbiology*. [Online] 14 (4), 504–510. Available from: doi:10.1016/j.mib.2011.07.011.

Russell, C.A., Jones, T.C., Barr, I.G., Cox, N.J., et al. (2008) The global circulation of seasonal influenza A (H3N2) viruses. *Science*. [Online] 320 (5874), 340–346. Available from: doi:10.1126/science.1154137.

Rust, M.J., Lakadamyali, M., Zhang, F. & Zhuang, X. (2004) Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nature Structural & Molecular Biology*. [Online] 11 (6), 567–573. Available from: doi:10.1038/nsmb769.

Ruszkowska, A., Lenartowicz, E., Moss, W.N., Kierzek, R., et al. (2016) Secondary structure model of the naked segment 7 influenza A virus genomic RNA. *The Biochemical journal*. [Online] 473 (23), 4327–4348. Available from: doi:10.1042/BCJ20160651.

Santos Afonso, Dos, E., Escriou, N., Leclercq, I., van der Werf, S., et al. (2005) The generation of recombinant influenza A viruses expressing a PB2 fusion protein requires the conservation of a packaging signal overlapping the coding and noncoding regions at the 5′ end of the PB2 segment. *Virology*. [Online] 341 (1), 34–46. Available from: doi:10.1016/j.virol.2005.06.040.

Schäfer, J.R., Kawaoka, Y., Bean, W.J., Süss, J., et al. (1993) Origin of the pandemic 1957 H2 influenza A virus and the persistence of its possible progenitors in the avian reservoir. *Virology*. 194 (2), 781–788.

Schmidt, N.W., Mishra, A., Wang, J., DeGrado, W.F., et al. (2013) Influenza Virus A M2 Protein Generates Negative Gaussian Membrane Curvature Necessary for Budding and Scission. *Journal of the American Chemical Society*. [Online] 135 (37), 13710–13719. Available from: doi:10.1021/ja400146z.

Schmitt, A.P. & Lamb, R.A. (2005) Influenza virus assembly and budding at the viral budozone. *Advances in virus research*. [Online] 64383–416. Available from: doi:10.1016/S0065-3527(05)64012-2.

Scholtissek, C., Hoyningen, von, V. & Rott, R. (1978) Genetic relatedness between the new 1977 epidemic strains (H1N1) of influenza and human influenza strains isolated between 1947 and 1957 (H1N1). *Virology*. 89 (2), 613–617.

Schrauwen, E.J.A., Herfst, S., Leijten, L.M., van Run, P., et al. (2012) The Multibasic Cleavage Site in H5N1 Virus Is Critical for Systemic Spread along the Olfactory and Hematogenous Routes in Ferrets. *Journal of Virology*. [Online] 86 (7), 3975–3984. Available from: doi:10.1128/JVI.06828-11.

Serna Martin, I., Hengrung, N., Renner, M., Sharps, J., et al. (2018) A Mechanism for the Activation of the Influenza Virus Transcriptase. *Molecular Cell*. [Online] 70 (6), 1101–1110.e1104. Available from: doi:10.1016/j.molcel.2018.05.011.

Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B.J. (2016) Global Mapping of Human RNA-RNA Interactions. *Molecular Cell*. [Online] 62 (4), 618–626. Available from: doi:10.1016/j.molcel.2016.04.030.

Shaw, M.L. & Palese, P. (2013) *Orthomyxoviridae, p 1151–1185. Fields virology*.

Shelton, H., Ayora-Talavera, G., Ren, J., Loureiro, S., et al. (2011) Receptor binding profiles of avian influenza virus hemagglutinin subtypes on human cells as a predictor of pandemic potential. *Journal of Virology*. [Online] 85 (4), 1875–1880. Available from: doi:10.1128/JVI.01822-10.

Shope, R.E. (1931) SWINE INFLUENZA: III. FILTRATION EXPERIMENTS AND ETIOLOGY. *The Journal of experimental medicine*. [Online] 54 (3), 373–385. Available from: doi:10.1084/jem.54.3.373.

Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A.E., et al. (2014) *RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP)*. [Online] 11 (9), 959–965. Available from: doi:10.1038/nmeth.3029.

Simmonds, P., Tuplin, A. & Evans, D.J. (2004) Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. *RNA*. [Online] 10 (9), 1337–1351. Available from: doi:10.1261/rna.7640104.

Simpson-Holley, M., Ellis, D., Fisher, D., Elton, D., et al. (2002) A Functional Link between the Actin Cytoskeleton and Lipid Rafts during Budding of Filamentous Influenza Virions. *Virology*. [Online] 301 (2), 212–225. Available from: doi:10.1006/viro.2002.1595.

Skripkin, E., Isel, C., Marquet, R., Ehresmann, B., et al. (1996) Psoralen crosslinking between human immunodeficiency virus type 1 RNA and primer tRNA(3)(Lys). *Nucleic acids research*. 24 (3), 509–514.

Smith, G.J.D., Bahl, J., Vijaykrishna, D., Zhang, J., et al. (2009) Dating the emergence of pandemic influenza viruses. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 106 (28), 11709–11712. Available from: doi:10.1073/pnas.0904991106.

Smith, G.J.D., Vijaykrishna, D., Bahl, J., Lycett, S.J., et al. (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*. [Online] 459 (7250), 1122–1125. Available from: doi:10.1038/nature08182.

Smith, W., Andrewes, C.H. & Laidlaw, P. (1933) A virus obtained from influenza patients. *Wiley Online Library*. [Online] 222 (5732). Available from: doi:10.1016/S0140-6736(00)78541-2.

Smola, M.J., Calabrese, J.M. & Weeks, K.M. (2015) Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry*. [Online] 54 (46), 6867–6875. Available from: doi:10.1021/acs.biochem.5b00977.

Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A., et al. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nature Protocols*. [Online] 10 (11), 1643–1669. Available from: doi:10.1038/nprot.2015.103.

Song, M.-S., Pascua, P.N.Q., Lee, J.H., Baek, Y.H., et al. (2011) Virulence and genetic compatibility of polymerase reassortant viruses derived from the pandemic (H1N1) 2009 influenza virus and circulating influenza A viruses. *Journal of Virology*. [Online] 85 (13), 6275–6286. Available from: doi:10.1128/JVI.02125-10.

Sorrell, E.M. & Perez, D.R. (2007) Adaptation of influenza A/Mallard/Potsdam/178-4/83 H2N2 virus in Japanese quail leads to infection and transmission in chickens. *Avian Diseases*. [Online] 51 (1), 264–268. Available from: doi:10.1637/7538-032906R.1.

Spasic, A., Assmann, S.M., Bevilacqua, P.C. & Mathews, D.H. (2018) Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic acids research*. [Online] 46 (1), 314–323. Available from: doi:10.1093/nar/gkx1057.

Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., et al. (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*. [Online] 519 (7544), 486–490. Available from: doi:10.1038/nature14263.

Spronken, M.I., van de Sandt, C.E., de Jongh, E.P., Vuong, O., et al. (2017) A compensatory mutagenesis study of a conserved hairpin in the M gene segment of influenza A virus shows its role in virus replication. *RNA Biology*. [Online] 14 (11), 1606–1616. Available from: doi:10.1080/15476286.2017.1338243.

Steil, B.P. & Barton, D.J. (2009) Cis-active RNA elements (CREs) and picornavirus RNA replication. *Virus Research*. [Online] 139 (2), 240–252. Available from: doi:10.1016/j.virusres.2008.07.027.

Stewart, H., Bingham, R.J., White, S.J., Dykeman, E.C., et al. (2016) Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. *Scientific reports*. [Online] 6 (1). Available from: doi:10.1038/srep22952.

Stoneley, M. & Willis, A.E. (2004) Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene*. [Online] 23 (18), 3200–3207. Available from: doi:10.1038/sj.onc.1207551.

Subbarao, E.K., London, W. & Murphy, B.R. (1993) A Single Amino-Acid in the Pb2-Gene of Influenza-a Virus Is a Determinant of Host Range. *Journal of Virology*. 67 (4), 1761–1764.

Subbarao, K., Klimov, A., Katz, J., Regnery, H., et al. (1998) Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science*. 279 (5349), 393–396.

Sugita, Y., Sagara, H., Noda, T. & Kawaoka, Y. (2013) Configuration of Viral Ribonucleoprotein Complexes within the Influenza A Virion. *Journal of Virology*. [Online] 87 (23), 12879–12884. Available from: doi:10.1128/JVI.02096-13.

Sung, P.-Y. & Roy, P. (2014) Sequential packaging of RNA genomic segments during the assembly of Bluetongue virus. *Nucleic acids research*. [Online] 42 (22), 13824–13838. Available from: doi:10.1093/nar/gku1171.

Suzuki, Y. & Nei, M. (2002) Origin and evolution of influenza virus hemagglutinin genes. *Molecular biology and evolution*. [Online] 19 (4), 501–509. Available from: doi:10.1093/oxfordjournals.molbev.a004105.

Swayne, D.E., Pantin-Jackwood, M., Kapczynski, D., Spackman, E., et al. (2009) Susceptibility of Poultry to Pandemic (H1N1) 2009 Virus. *Emerging Infectious Diseases*. [Online] 15 (12), 2061–2063. Available from: doi:10.3201/eid1512.091060.

Sztuba-Solinska, J., Diaz, L., Kumar, M.R., Kolb, G., et al. (2016) A small stem-loop structure of the Ebola virus trailer is essential for replication and interacts with heat-shock protein A8. *Nucleic acids research*. [Online] 44 (20), 9831–9846. Available from: doi:10.1093/nar/gkw825.

Sztuba-Solinska, J., Rausch, J.W., Smith, R., Miller, J.T., et al. (2017) Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA: a structural scaffold for nuclear, cytoplasmic and viral proteins. *Nucleic acids research*. [Online] 45 (11), 6805–6821. Available from: doi:10.1093/nar/gkx241.

Tajima, Y., Iwakawa, H.-O., Kaido, M., Mise, K., et al. (2011) A long-distance RNA–RNA interaction plays an important role in programmed −1 ribosomal frameshifting in the translation of p88 replicase protein of Red clover necrotic mosaic virus. *Virology*. [Online] 417 (1), 169–178. Available from: doi:10.1016/j.virol.2011.05.012.

Talkish, J., May, G., Lin, Y., Woolford, J.L.J., et al. (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA*. [Online] 20 (5), 713–720. Available from: doi:10.1261/rna.042218.113.

Tarus, B., Bakowiez, O., Chenavas, S., Duchemin, L., et al. (2012) Oligomerization paths of the nucleoprotein of influenza A virus. *Biochimie*. [Online] 94 (3), 776–785. Available from: doi:10.1016/j.biochi.2011.11.009.

Taubenberger, J.K. & Kash, J.C. (2010) Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host and Microbe*. [Online] 7 (6), 440–451. Available from: doi:10.1016/j.chom.2010.05.009.

Taubenberger, J.K. & Morens, D.M. (2009) Pandemic influenza - including a risk assessment of H5N1. *Revue Scientifique Et Technique-Office International Des Epizooties*. 28 (1), 187–202.

Taubenberger, J.K., Reid, A.H., Krafft, A.E., Bijwaard, K.E., et al. (1997) Initial genetic characterization of the 1918 "Spanish" influenza virus. *Science*. 275 (5307), 1793–1796.

Taubenberger, J.K., Reid, A.H., Lourens, R.M., Wang, R.X., et al. (2005) Characterization of the 1918 influenza virus polymerase genes. *Nature*. [Online] 437 (7060), 889–893. Available from: doi:10.1038/nature04230.

Teodoro Fajardo, K.A.P.R., Jr (2016) Generation of infectious RNA complexes in orbiviruses: RNA-RNA interactions of genomic segments. *Oncotarget*. [Online] 7 (45), 72559–72570. Available from: doi:10.18632/oncotarget.12496.

Thompson, J.F. & Hearst, J.E. (1983) Structure of E. coli 16S RNA elucidated by psoralen crosslinking. *Cell*. 32 (4), 1355–1365.

Tong, S., Zhu, X., Li, Y., Shi, M., et al. (2013) New world bats harbor diverse influenza A viruses. Kanta Subbarao (ed.). *PLoS Pathogens*. [Online] 9 (10), e1003657. Available from: doi:10.1371/journal.ppat.1003657.

Tran, V., Moser, L.A., Poole, D.S. & Mehle, A. (2013) Highly sensitive real-time in vivo imaging of an influenza reporter virus reveals dynamics of replication and spread. *Journal of Virology*. [Online] 87 (24), 13321–13329. Available from: doi:10.1128/JVI.02381-13.

Turrell, L., Lyall, J.W., Tiley, L.S., Fodor, E., et al. (2013) The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. *Nature communications*. [Online] 4 (1), 1591. Available from: doi:10.1038/ncomms2589.

Van Poucke, S.G.M., Nicholls, J.M., Nauwynck, H.J. & Van Reeth, K. (2010) Replication of avian, human and swine influenza viruses in porcine respiratory explants and association with sialic acid distribution. *Virology Journal*. [Online] 7 (1). Available from: doi:10.1186/1743-422X-7-38.

Vasin, A.V., Temkina, O.A., Egorov, V.V., Klotchenko, S.A., et al. (2014) Molecular mechanisms enhancing the proteome of influenza A viruses: an overview of recently discovered proteins. *Virus Research*. [Online] 18553–63. Available from: doi:10.1016/j.virusres.2014.03.015.

Velthuis, Te, A.J.W. & Fodor, E. (2016) Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nature reviews. Microbiology*. [Online] 14 (8), 479–493. Available from: doi:10.1038/nrmicro.2016.87.

Vidic, J., Noiray, M., Bagchi, A. & Slama-Schwok, A. (2016) Identification of a Novel Complex between the Nucleoprotein and PA(1-27) of Influenza A Virus Polymerase. *Biochemistry*. [Online] 55 (31), 4259–4262. Available from: doi:10.1021/acs.biochem.6b00514.

Wagner, R., Matrosovich, M. & Klenk, H.-D. (2002) Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Reviews in medical virology*. [Online] 12 (3), 159–166. Available from: doi:10.1002/rmv.352.

Wang, C., Lamb, R.A. & Pinto, L.H. (1995) Activation of the M2 ion channel of influenza virus: a role for the transmembrane domain histidine residue. *Biophysical Journal*. [Online] 69 (4), 1363–1371. Available from: doi:10.1016/S0006-3495(95)80003-2.

Wang, J., Li, J., Zhao, L., Cao, M., et al. (2016) Dual Roles of the Hemagglutinin Segment-Specific Noncoding Nucleotides in the Extended Duplex Region of the Influenza A Virus RNA Promoter Terence S Dermody (ed.). *Journal of Virology*. [Online] 91 (1), e01931–16. Available from: doi:10.1128/JVI.01931-16.

Wang, P., Palese, P. & O'Neill, R.E. (1997) The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza a virus nucleoprotein NP is a nonconventional nuclear localization signal. *Journal of Virology*. 71 (3), 1850–1856.

Wang, R., Soll, L., Dugan, V., Runstadler, J., et al. (2008) Examining the hemagglutinin subtype diversity among wild duck-origin influenza A viruses using ethanol-fixed cloacal swabs and a novel RT-PCR method. *Virology*. [Online] 375 (1), 182–189. Available from: doi:10.1016/j.virol.2008.01.041.

Wang, Y., Ludwig, J., Schuberth, C., Goldeck, M., et al. (2010) Structural and functional insights into 5 '-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nature Structural & Molecular Biology*. [Online] 17 (7), 781–U19. Available from: doi:10.1038/nsmb.1863.

Wassarman, D.A. (1993) Psoralen crosslinking of small RNAs in vitro. *Molecular biology reports*. 17 (2), 143–151.

Watanabe, T., Watanabe, S., Noda, T., Fujii, Y., et al. (2003) Exploitation of Nucleic Acid Packaging Signals To Generate a Novel Influenza Virus-Based Vector Stably Expressing Two Foreign Genes. *Journal of Virology*. [Online] 77 (19), 10575–10583. Available from: doi:10.1128/JVI.77.19.10575-10583.2003.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., et al. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*. [Online] 460 (7256), 711–716. Available from: doi:10.1038/nature08237.

Weber, M., Sediri, H., Felgenhauer, U., Binzen, I., et al. (2015) Influenza Virus Adaptation PB2-627K Modulates Nucleocapsid Inhibition by the Pathogen Sensor RIG-I. *Cell Host and Microbe*. [Online] 17 (3), 309–319. Available from: doi:10.1016/j.chom.2015.01.005.

White, M.C., Steel, J. & Lowen, A.C. (2017) Heterologous Packaging Signals on Segment 4, but Not Segment 6 or Segment 8, Limit Influenza A Virus Reassortment. Adolfo García-Sastre (ed.). *Journal of Virology*. [Online] 91 (11), e00195–17. Available from: doi:10.1128/JVI.00195-17.

WHO (2017) *Influenza (Seasonal)*. [Online]. Available from: http://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal) [Accessed: 25 August 2018].

Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., et al. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biology*. [Online] 6 (4), e96–17. Available from: doi:10.1371/journal.pbio.0060096.

Wilkinson, K.A., Vasa, S.M., Deigan, K.E., Mortimer, S.A., et al. (2009) Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA*. [Online] 15 (7), 1314–1321. Available from: doi:10.1261/rna.1536209.

Williams, G.D., Townsend, D., Wylie, K.M., Kim, P.J., et al. (2018) Nucleotide resolution mapping of influenza A virus nucleoprotein-RNA interactions reveals RNA features required for replication. *Nature communications*. [Online] 9 (1), 465. Available from: doi:10.1038/s41467-018-02886-w.

Wise, H.M., Hutchinson, E.C., Jagger, B.W., Stuart, A.D., et al. (2012) Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion Channel with an Antigenically Distinct Ectodomain Andrew Pekosz (ed.). *PLoS Pathogens*. [Online] 8 (11), e1002998. Available from: doi:10.1371/journal.ppat.1002998.

Wu, B., Grigull, J., Ore, M.O., Morin, S., et al. (2013) Global Organization of a Positive-strand RNA Virus Genome Shou-Wei Ding (ed.). *PLoS Pathogens*. [Online] 9 (5). Available from: doi:10.1371/journal.ppat.1003363.

Wu, C.-Y., Yeh, Y.-C., Yang, Y.-C., Chou, C., et al. (2010) Mammalian Expression of Virus-Like Particles for Advanced Mimicry of Authentic Influenza Virus Charlotte Hauser (ed.). *PLoS ONE*. [Online] 5 (3), e9784. Available from: doi:10.1371/journal.pone.0009784.

Wu, W.W.H., Sun, Y.-H.B. & Panté, N. (2007) Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein. *Virology Journal*. [Online] 4 (1), 49. Available from: doi:10.1186/1743-422X-4-49.

Xu, R., Zhu, X., McBride, R., Nycholat, C.M., et al. (2012) Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009 H1N1 influenza pandemic. *Journal of Virology*. [Online] 86 (17), 9221–9232. Available from: doi:10.1128/JVI.00697-12.

Yamanaka, K., Ishihama, A. & Nagata, K. (1990) Reconstitution of influenza virus RNA-nucleoprotein complexes structurally resembling native viral ribonucleoprotein cores. *Journal of Biological Chemistry*. 265 (19), 11151–11155.

Yang, D., Liu, P., Wudeck, E.V., Giedroc, D.P., et al. (2015) SHAPE analysis of the RNA secondary structure of the Mouse Hepatitis Virus 5' untranslated region and N-terminal nsp1 coding sequences. *Virology*. [Online] 47515–27. Available from: doi:10.1016/j.virol.2014.11.001.

Yang, P., Bansal, A., Liu, C. & Air, G.M. (1997) Hemagglutinin Specificity and Neuraminidase Coding Capacity of Neuraminidase-Deficient Influenza Viruses. *Virology*. [Online] 229 (1), 155–165. Available from: doi:10.1006/viro.1996.8421.

Ye, Q., Krug, R.M. & Tao, Y.J. (2006) The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature*. [Online] 444 (7122), 1078–1082. Available from: doi:10.1038/nature05379.

Yen, H.-L., Liang, C.-H., Wu, C.-Y., Forrest, H.L., et al. (2011) Hemagglutinin-neuraminidase balance confers respiratory-droplet transmissibility of the pandemic H1N1 influenza virus in ferrets. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 108 (34), 14264–14269. Available from: doi:10.1073/pnas.1111000108.

York, A., Hengrung, N., Vreede, F.T., Huiskonen, J.T., et al. (2013) Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 110 (45), E4238–E4245. Available from: doi:10.1073/pnas.1315068110.

Yuan, P., Bartlam, M., Lou, Z., Chen, S., et al. (2009) Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature*. [Online] 458 (7240), 909–913. Available from: doi:10.1038/nature07720.

Zimmer, S.M. & Burke, D.S. (2009) Historical perspective--Emergence of influenza A (H1N1) viruses. *New England Journal of Medicine*. [Online] 361 (3), 279–285. Available from: doi:10.1056/NEJMra0904322.

Zweerink, H.J., Courtneidge, S.A., Skehel, J.J., Crumpton, M.J., et al. (1977) Cytotoxic T cells kill influenza virus infected cells but do not distinguish between serologically distinct type A viruses. *Nature*. [Online] 267 (5609), 354–356. Available from: doi:10.1038/267354a0.

Appendix A: WSN SHAPE reactivities and summary statistics

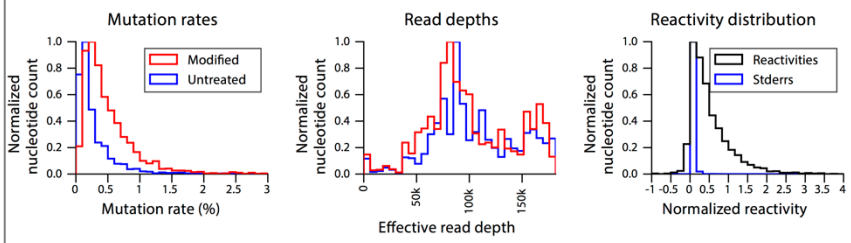
Appendix A. 1 *In virio (1M7): experiment 1*

The summary statistics for mutation rates, read depth and reactivity distributions are provided for each segment. Mutation rates for SHAPE reagent-modified and non-modified samples are given, the mutation rates are expected to be higher in the SHAPE reagent-treated samples compared to the non-treated controls. Read depths for most nucleotides are expected to be above 2000. The normalised SHAPE reactivities of successful experiments should be mostly positive.

SHAPE reactivity profiles are provided for each segment. Shape reactivities are coloured using the conventional SHAPE scale, with high SHAPE reactivities (>0.85) coloured red, mid-range reactivities (0.85-0.4) in orange and low shape reactivities (<0.4) in black. Mutation rates for each nucleotide in the vRNA segment are plotted below the SHAPE reactivity profile. SHAPE reagent-treated samples (red) are expected to have higher mutation rates for most nucleotides, compared to the non-treated controls (blue). The coverage for each nucleotide is shown on the bottom panel for the SHAPE reagent treated and non-treated samples.

Summary statistics

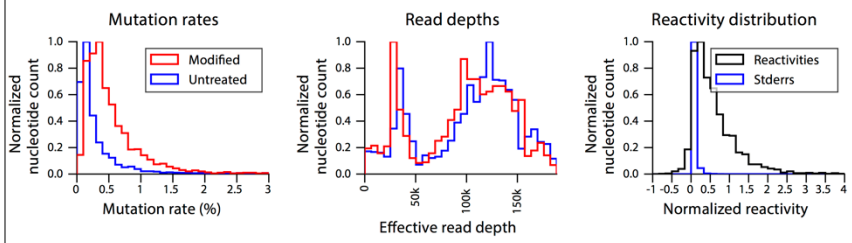
RNA: chrPB2



Modified sample:	
95th percentile rate:	1.37%
Median rate:	0.11%
Untreated sample:	
95th percentile rate:	0.98%
Median rate:	0.06%

Modified sample:	
Median depth:	91,758
5th percentile depth:	40,532
Untreated sample:	
Median depth:	96,663
5th percentile depth:	43,270

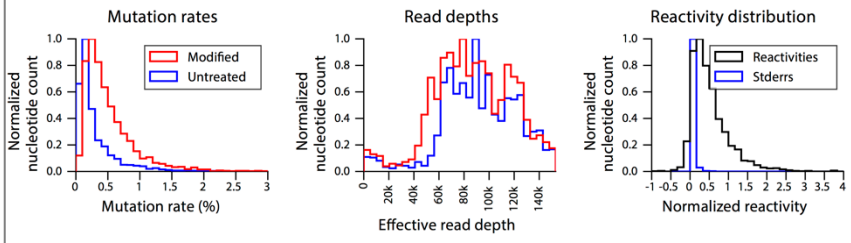
RNA: chrPB1



Modified sample:	
95th percentile rate:	1.56%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.17%
Median rate:	0.06%

Modified sample:	
Median depth:	107,786
5th percentile depth:	21,538
Untreated sample:	
Median depth:	118,504
5th percentile depth:	24,148

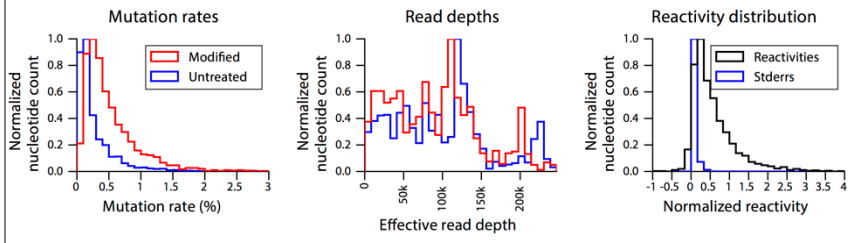
RNA: chrPA



Modified sample:	
95th percentile rate:	1.41%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.1%
Median rate:	0.06%

Modified sample:	
Median depth:	89,416
5th percentile depth:	41,019
Untreated sample:	
Median depth:	93,046
5th percentile depth:	45,672

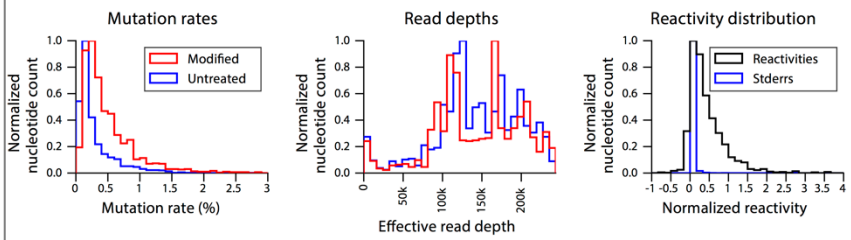
RNA: chrHA



Modified sample:	
95th percentile rate:	1.46%
Median rate:	0.11%
Untreated sample:	
95th percentile rate:	1.08%
Median rate:	0.05%

Modified sample:	
Median depth:	93,486
5th percentile depth:	11,585
Untreated sample:	
Median depth:	103,853
5th percentile depth:	14,508

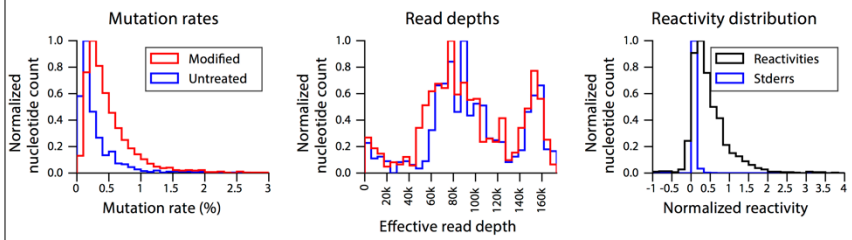
RNA: chrNP



Modified sample:	
95th percentile rate:	1.39%
Median rate:	0.11%
Untreated sample:	
95th percentile rate:	1.08%
Median rate:	0.06%

Modified sample:	
Median depth:	151,682
5th percentile depth:	40,777
Untreated sample:	
Median depth:	143,672
5th percentile depth:	38,860

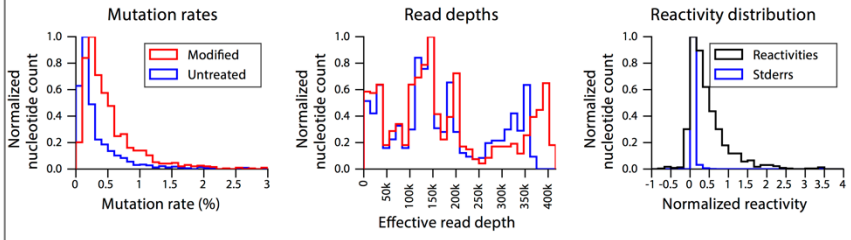
RNA: chrNA



Modified sample:	
95th percentile rate:	1.43%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.13%
Median rate:	0.06%

Modified sample:	
Median depth:	90,600
5th percentile depth:	16,646
Untreated sample:	
Median depth:	92,365
5th percentile depth:	18,887

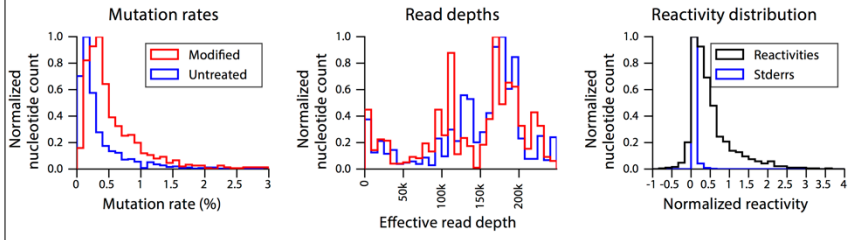
RNA: chrM



Modified sample:	
95th percentile rate:	1.51%
Median rate:	0.11%
Untreated sample:	
95th percentile rate:	1.14%
Median rate:	0.06%

Modified sample:	
Median depth:	147,397
5th percentile depth:	10,833
Untreated sample:	
Median depth:	147,226
5th percentile depth:	10,541

RNA: chrNS

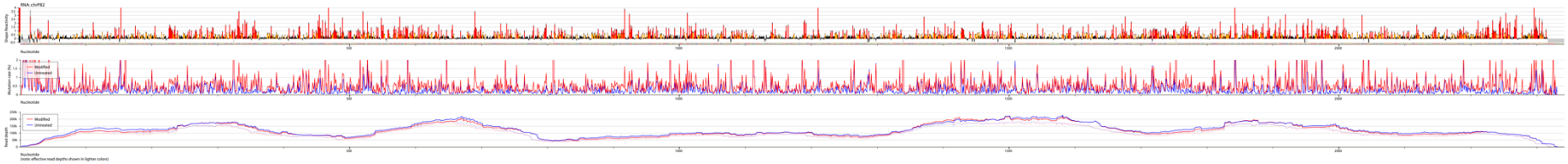


Modified sample:	
95th percentile rate:	1.76%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.44%
Median rate:	0.07%

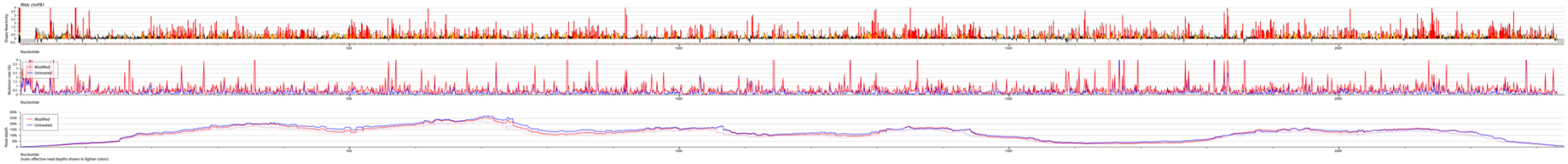
Modified sample:	
Median depth:	168,372
5th percentile depth:	8,713
Untreated sample:	
Median depth:	168,397
5th percentile depth:	11,040

SHAPE reactivity profiles

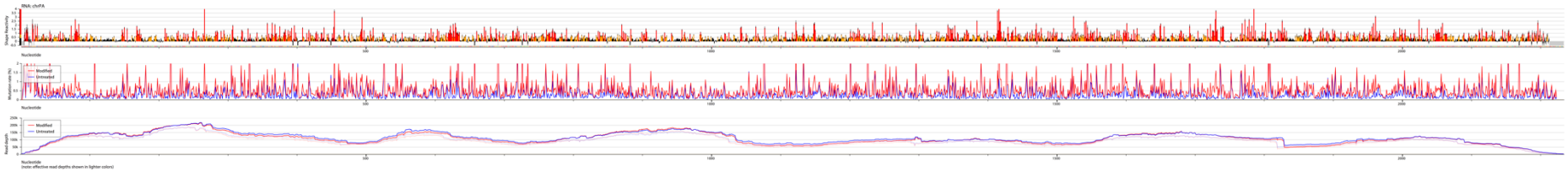
PB2



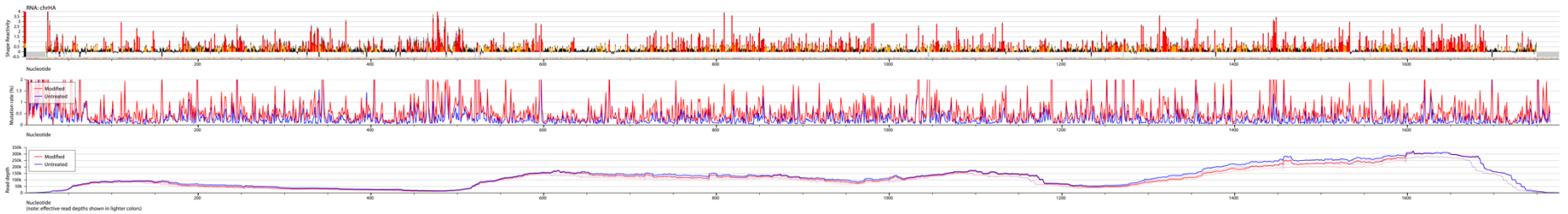
PB1



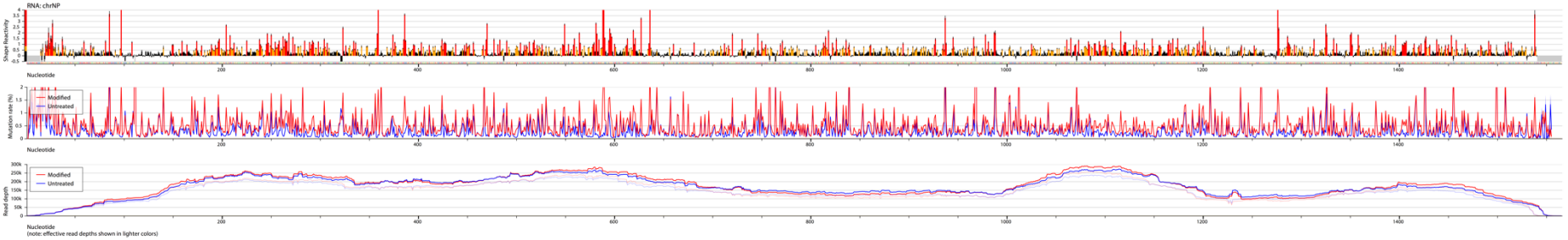
PA



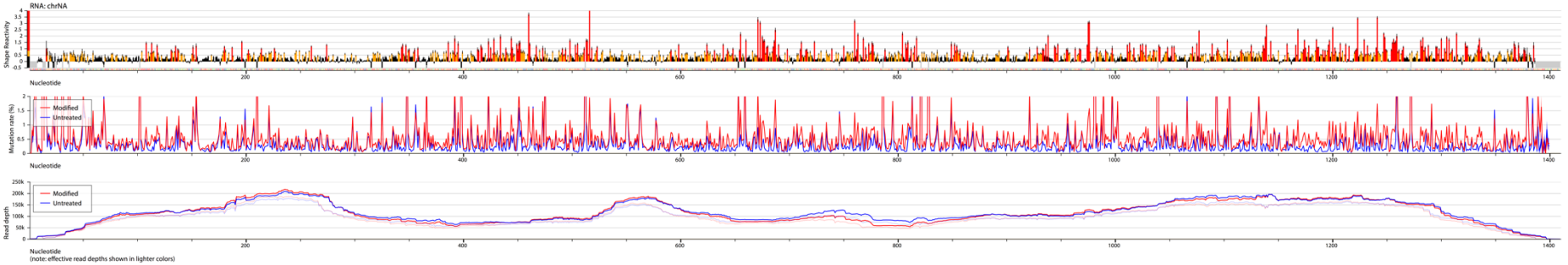
HA



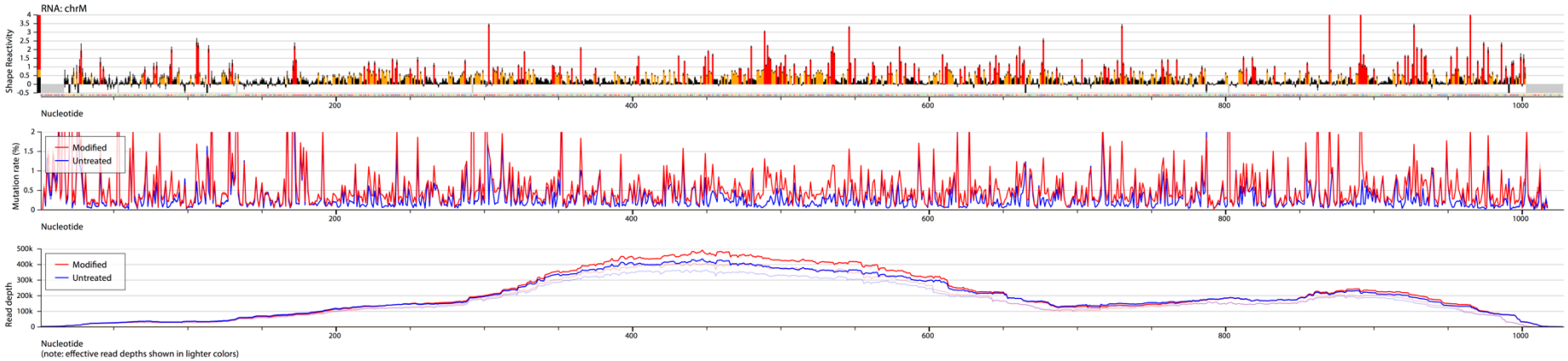
NP


RNA: chrNP
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
(note: effective read depths shown in lighter colors)

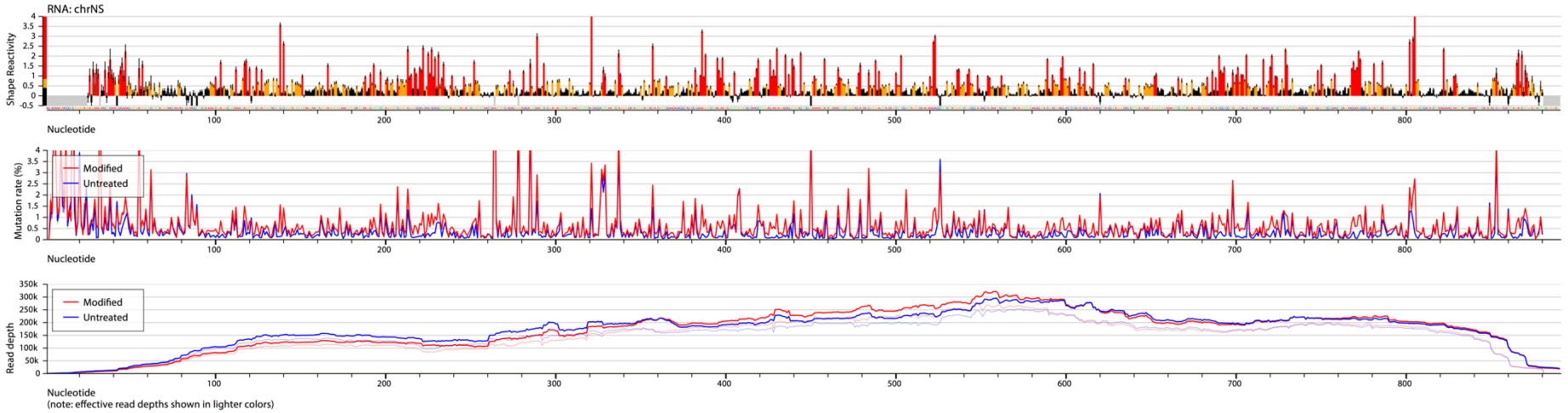
NA


RNA: chrNA
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
(note: effective read depths shown in lighter colors)

M


RNA: chrM
Shape Reactivity
Nucleotide
Modified
Untreated
Mutation rate (%)
Nucleotide
Modified
Untreated
Read depth
Nucleotide
(note: effective read depths shown in lighter colors)

NS


RNA: chrNS
Shape Reactivity
Nucleotide
Modified
Untreated
Mutation rate (%)
Nucleotide
Modified
Untreated
Read depth
Nucleotide
(note: effective read depths shown in lighter colors)

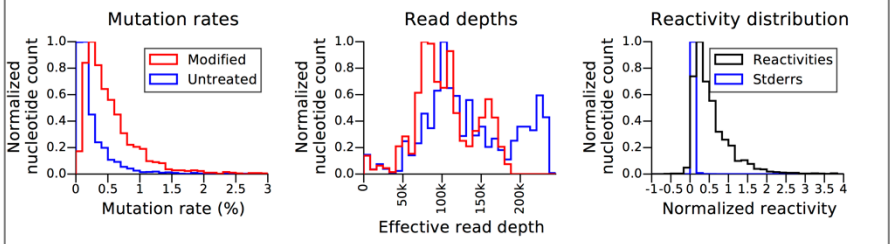
Appendix A. 2 *In virio (1M7) experiment 2*

The summary statistics for mutation rates, read depth and reactivity distributions are provided for each segment. Mutation rates for SHAPE reagent-modified and non-modified samples are given, the mutation rates are expected to be higher in the SHAPE reagent-treated samples compared to the non-treated controls. Read depths for most nucleotides are expected to be above 2000. The normalised SHAPE reactivities of successful experiments should be mostly positive.

SHAPE reactivity profiles are provided for each segment. Shape reactivities are coloured using the conventional SHAPE scale, with high SHAPE reactivities (>0.85) coloured red, mid-range reactivities (0.85-0.4) in orange and low shape reactivities (<0.4) in black. Mutation rates for each nucleotide in the vRNA segment are plotted below the SHAPE reactivity profile. SHAPE reagent-treated samples (red) are expected to have higher mutation rates for most nucleotides, compared to the non-treated controls (blue). The coverage for each nucleotide is shown on the bottom panel for the SHAPE reagent treated and non-treated samples.

Summary statistics

RNA: chrPB2



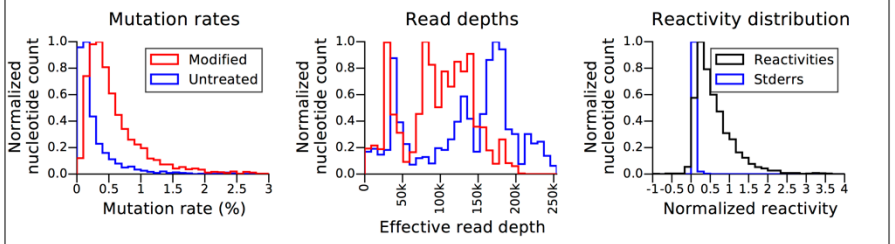
Modified sample:
95th percentile rate: 1.41%
Median rate: 0.12%

Untreated sample:
95th percentile rate: 0.9%
Median rate: 0.05%

Modified sample:
Median depth: 100,704
5th percentile depth: 46,860

Untreated sample:
Median depth: 122,239
5th percentile depth: 53,350

RNA: chrPB1



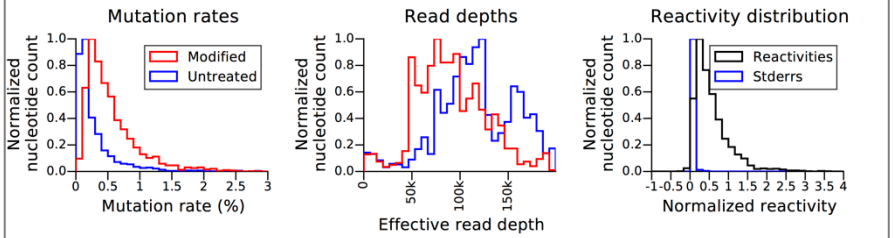
Modified sample:
95th percentile rate: 1.7%
Median rate: 0.14%

Untreated sample:
95th percentile rate: 1.07%
Median rate: 0.05%

Modified sample:
Median depth: 101,332
5th percentile depth: 24,292

Untreated sample:
Median depth: 160,284
5th percentile depth: 27,030

RNA: chrPA



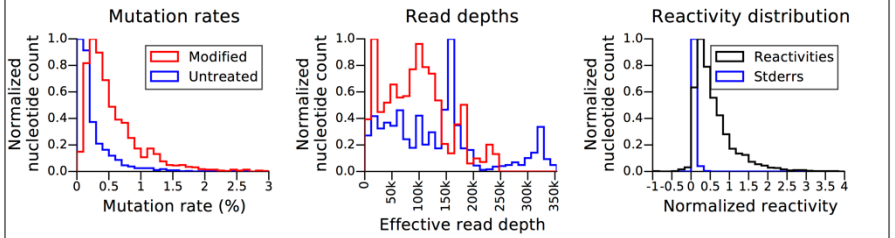
Modified sample:
95th percentile rate: 1.51%
Median rate: 0.14%

Untreated sample:
95th percentile rate: 1.01%
Median rate: 0.05%

Modified sample:
Median depth: 87,446
5th percentile depth: 42,682

Untreated sample:
Median depth: 118,547
5th percentile depth: 50,148

RNA: chrHA



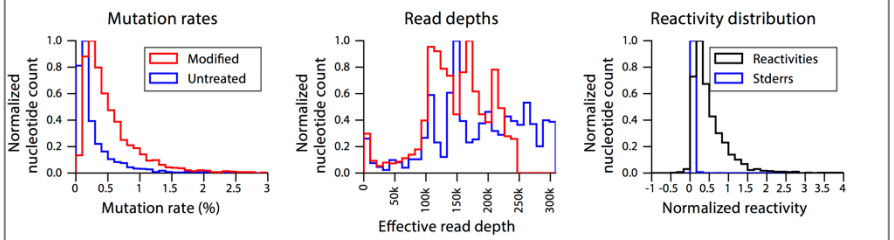
Modified sample:
95th percentile rate: 1.59%
Median rate: 0.12%

Untreated sample:
95th percentile rate: 1.01%
Median rate: 0.05%

Modified sample:
Median depth: 93,364
5th percentile depth: 12,787

Untreated sample:
Median depth: 133,599
5th percentile depth: 15,695

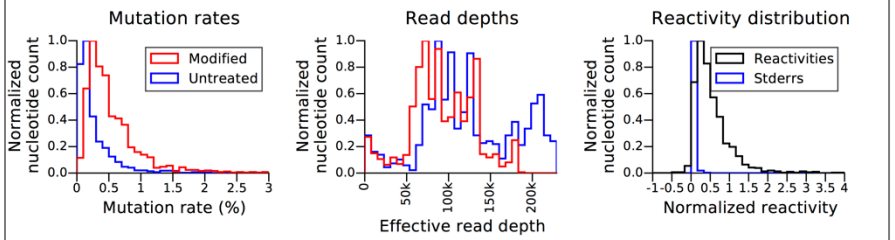
RNA: chrNP



Modified sample:	
95th percentile rate:	1.47%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.%
Median rate:	0.05%

Modified sample:	
Median depth:	151,659
5th percentile depth:	41,234
Untreated sample:	
Median depth:	184,576
5th percentile depth:	45,639

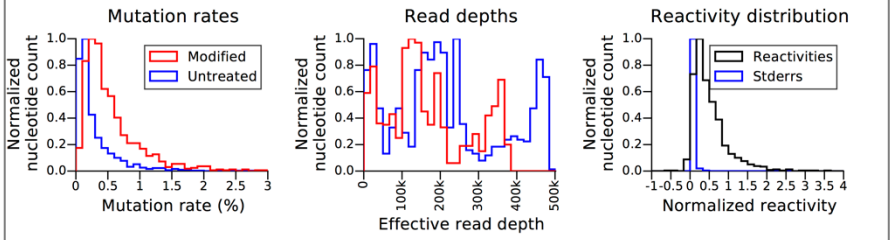
RNA: chrNA



Modified sample:	
95th percentile rate:	1.6%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.06%
Median rate:	0.06%

Modified sample:	
Median depth:	88,735
5th percentile depth:	16,101
Untreated sample:	
Median depth:	119,084
5th percentile depth:	19,001

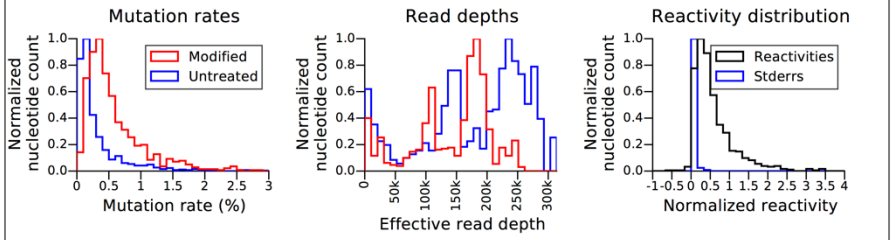
RNA: chrM



Modified sample:	
95th percentile rate:	1.6%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.09%
Median rate:	0.05%

Modified sample:	
Median depth:	142,468
5th percentile depth:	10,937
Untreated sample:	
Median depth:	197,228
5th percentile depth:	9,926

RNA: chrNS

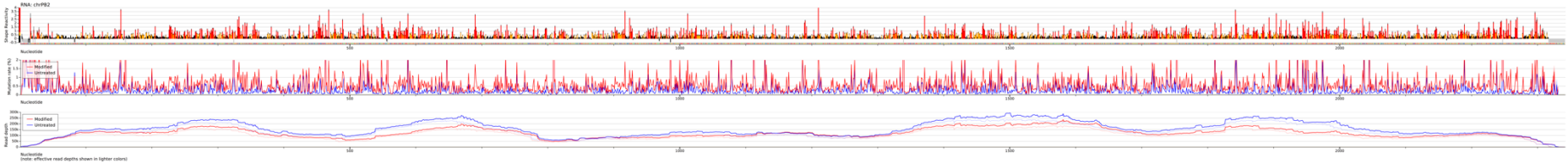


Modified sample:	
95th percentile rate:	1.81%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.31%
Median rate:	0.05%

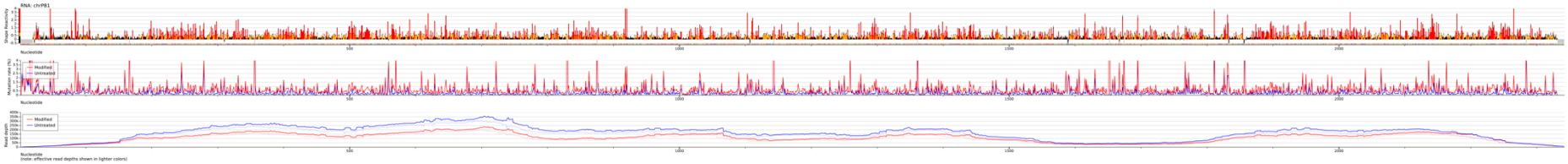
Modified sample:	
Median depth:	170,630
5th percentile depth:	8,478
Untreated sample:	
Median depth:	217,016
5th percentile depth:	11,076

SHAPE reactivity profiles

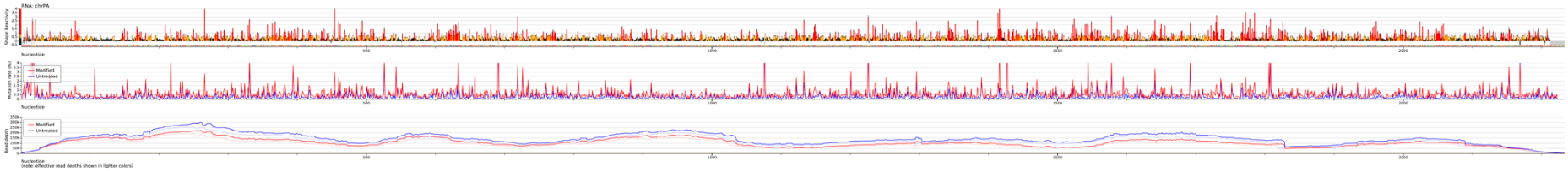
PB2



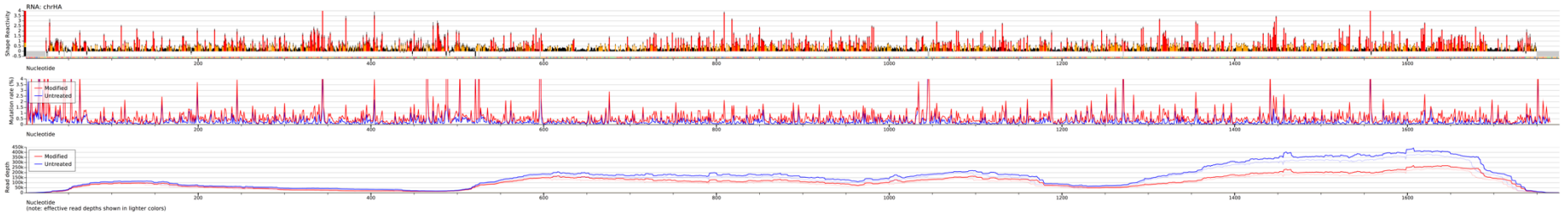
PB1



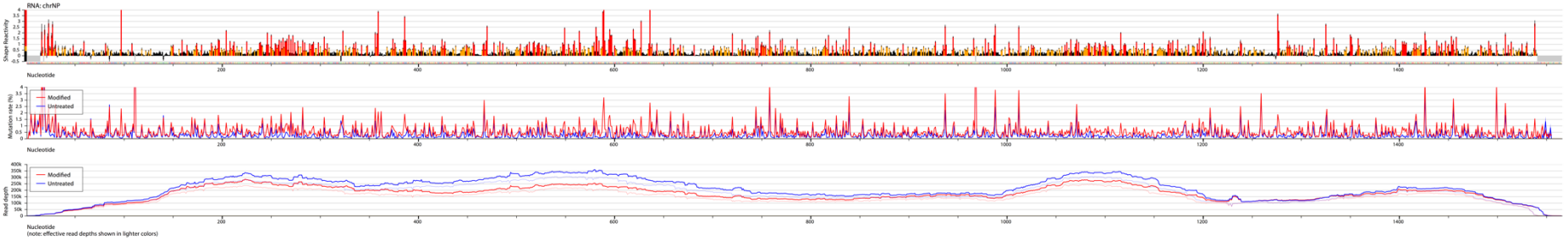
PA



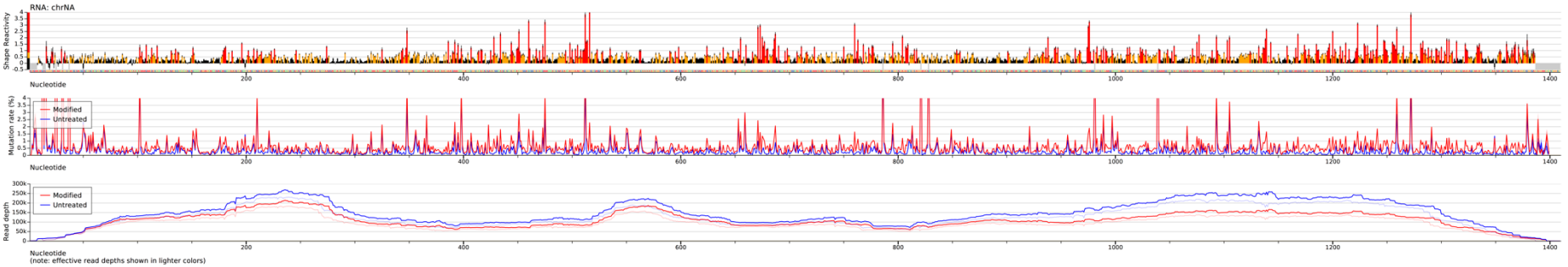
HA



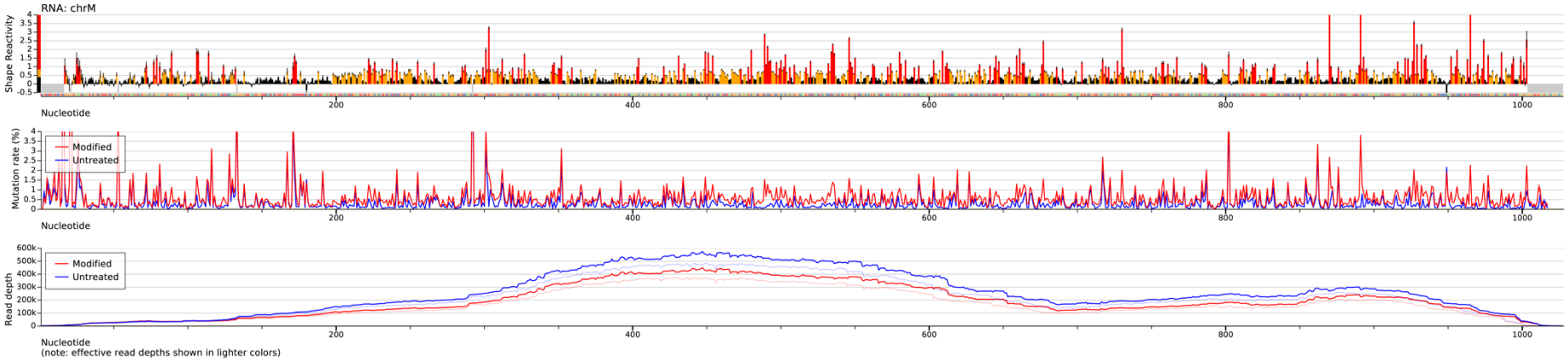
NP


RNA: chrNP
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Nucleotide
Read depth
Modified
Untreated
Nucleotide
(note: effective read depths shown in lighter colors)

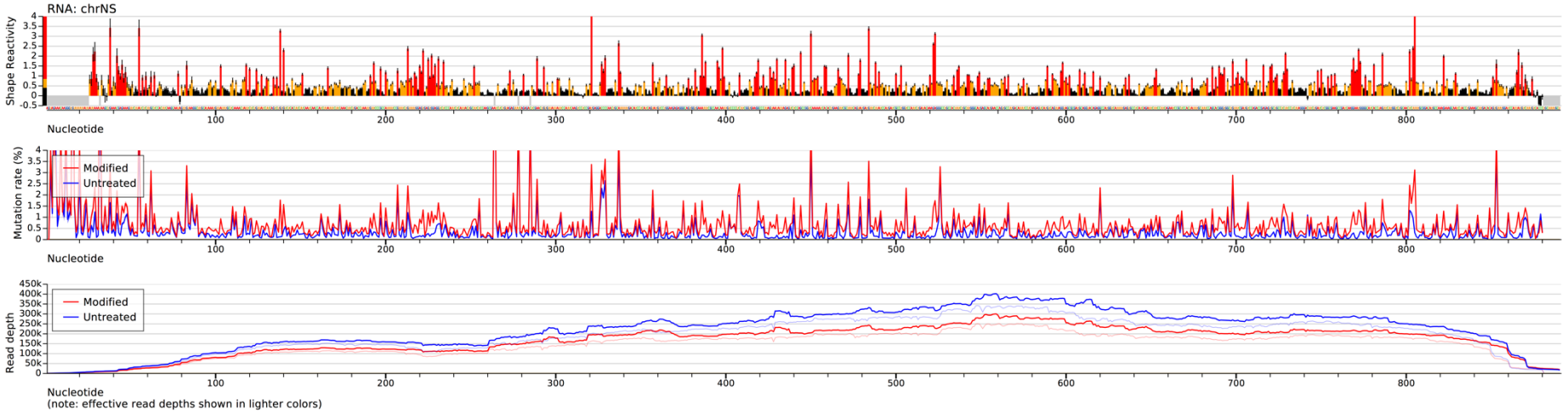
NA


RNA: chrNA
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Nucleotide
Read depth
Modified
Untreated
Nucleotide
(note: effective read depths shown in lighter colors)

M


RNA: chrM
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Nucleotide
Read depth
Modified
Untreated
Nucleotide
(note: effective read depths shown in lighter colors)

NS


RNA: chrNS
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Nucleotide
Read depth
Modified
Untreated
Nucleotide
(note: effective read depths shown in lighter colors)

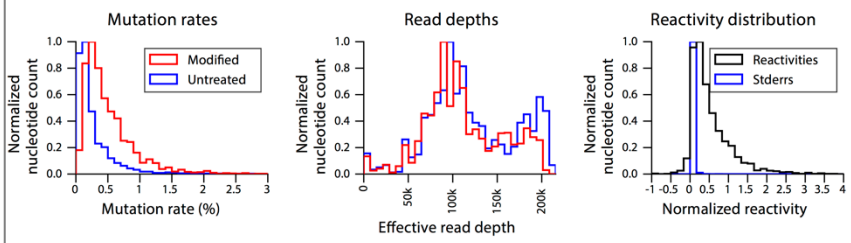
Appendix A. 3 *In virio (1M7) experiment 3*

The summary statistics for mutation rates, read depth and reactivity distributions are provided for each segment. Mutation rates for SHAPE reagent-modified and non-modified samples are given, the mutation rates are expected to be higher in the SHAPE reagent-treated samples compared to the non-treated controls. Read depths for most nucleotides are expected to be above 2000. The normalised SHAPE reactivities of successful experiments should be mostly positive.

SHAPE reactivity profiles are provided for each segment. Shape reactivities are coloured using the conventional SHAPE scale, with high SHAPE reactivities (>0.85) coloured red, mid-range reactivities (0.85-0.4) in orange and low shape reactivities (<0.4) in black. Mutation rates for each nucleotide in the vRNA segment are plotted below the SHAPE reactivity profile. SHAPE reagent-treated samples (red) are expected to have higher mutation rates for most nucleotides, compared to the non-treated controls (blue). The coverage for each nucleotide is shown on the bottom panel for the SHAPE reagent treated and non-treated samples.

Summary statistics

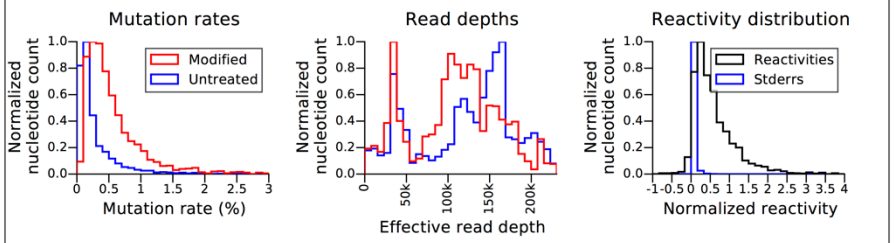
RNA: chrPB2



Modified sample:	
95th percentile rate:	1.4%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	0.95%
Median rate:	0.05%

Modified sample:	
Median depth:	104,070
5th percentile depth:	46,721
Untreated sample:	
Median depth:	110,665
5th percentile depth:	48,121

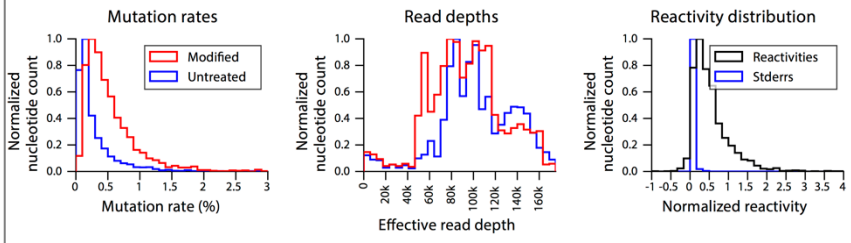
RNA: chrPB1



Modified sample:	
95th percentile rate:	1.66%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.11%
Median rate:	0.06%

Modified sample:	
Median depth:	112,260
5th percentile depth:	24,935
Untreated sample:	
Median depth:	142,279
5th percentile depth:	25,665

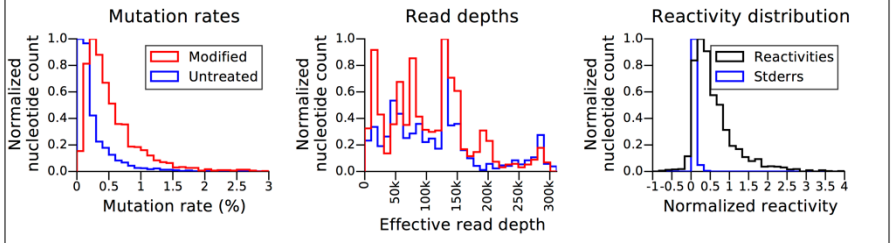
RNA: chrPA



Modified sample:	
95th percentile rate:	1.44%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.06%
Median rate:	0.06%

Modified sample:	
Median depth:	94,564
5th percentile depth:	49,654
Untreated sample:	
Median depth:	102,347
5th percentile depth:	49,869

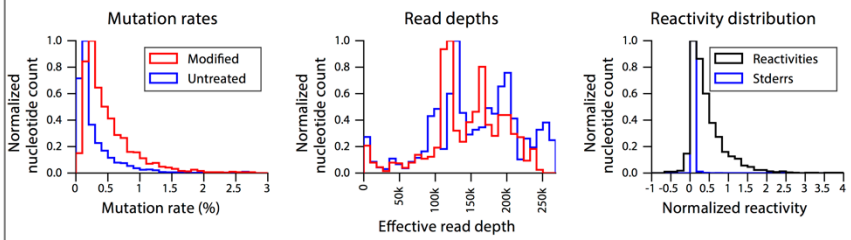
RNA: chrHA



Modified sample:	
95th percentile rate:	1.49%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.06%
Median rate:	0.05%

Modified sample:	
Median depth:	99,318
5th percentile depth:	14,405
Untreated sample:	
Median depth:	115,830
5th percentile depth:	16,047

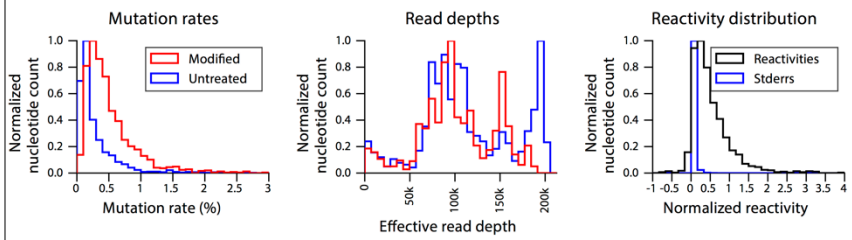
RNA: chrNP



Modified sample:	
95th percentile rate:	1.38%
Median rate:	0.11%
Untreated sample:	
95th percentile rate:	1.01%
Median rate:	0.05%

Modified sample:	
Median depth:	147,917
5th percentile depth:	39,929
Untreated sample:	
Median depth:	160,245
5th percentile depth:	39,902

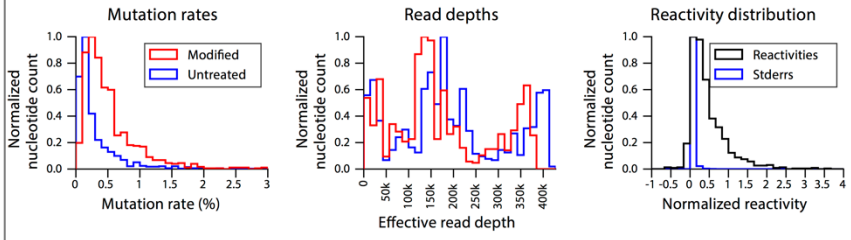
RNA: chrNA



Modified sample:	
95th percentile rate:	1.5%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.06%
Median rate:	0.06%

Modified sample:	
Median depth:	97,096
5th percentile depth:	17,324
Untreated sample:	
Median depth:	104,942
5th percentile depth:	19,929

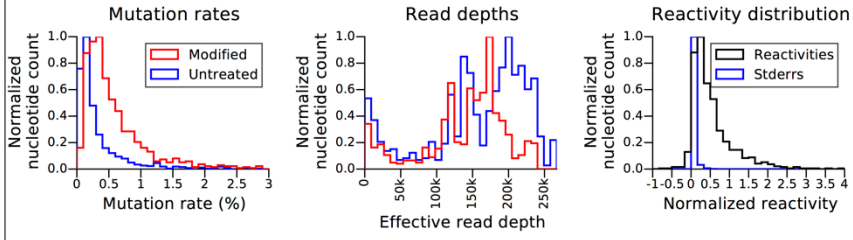
RNA: chrM



Modified sample:	
95th percentile rate:	1.48%
Median rate:	0.11%
Untreated sample:	
95th percentile rate:	1.17%
Median rate:	0.05%

Modified sample:	
Median depth:	144,699
5th percentile depth:	12,305
Untreated sample:	
Median depth:	180,153
5th percentile depth:	9,293

RNA: chrNS

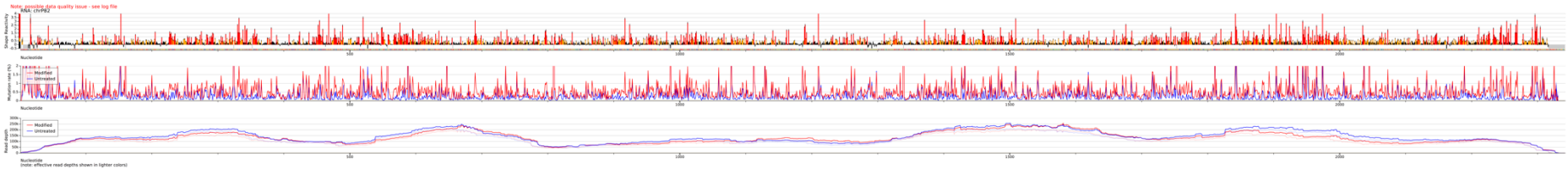


Modified sample:	
95th percentile rate:	1.89%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.54%
Median rate:	0.06%

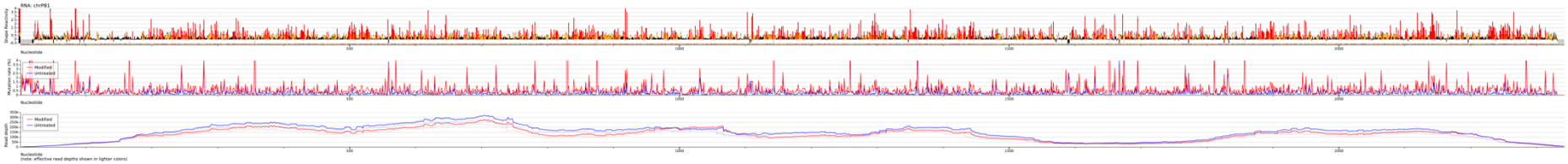
Modified sample:	
Median depth:	152,337
5th percentile depth:	9,598
Untreated sample:	
Median depth:	184,994
5th percentile depth:	11,629

Shape reactivity profiles

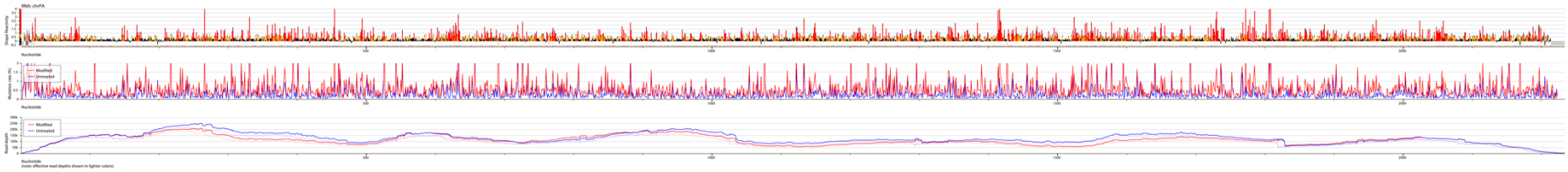
PB2



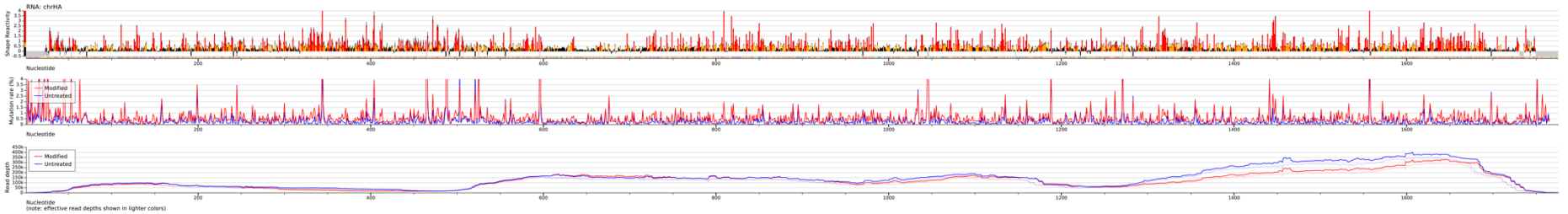
PB1



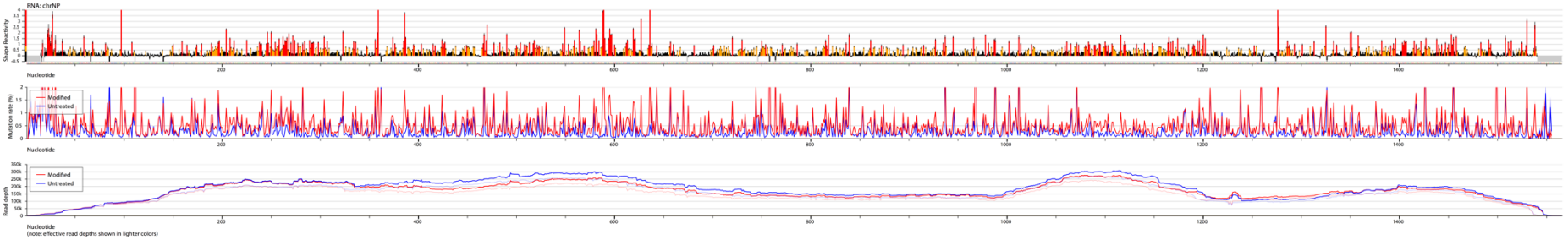
PA



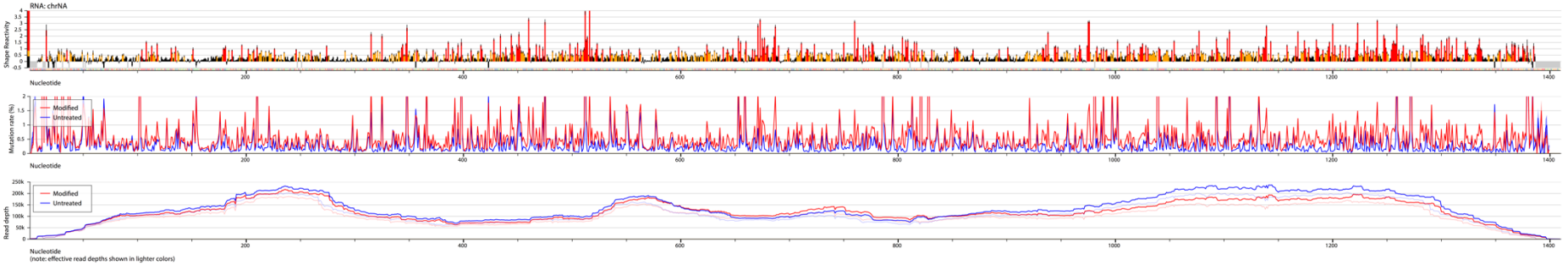
HA



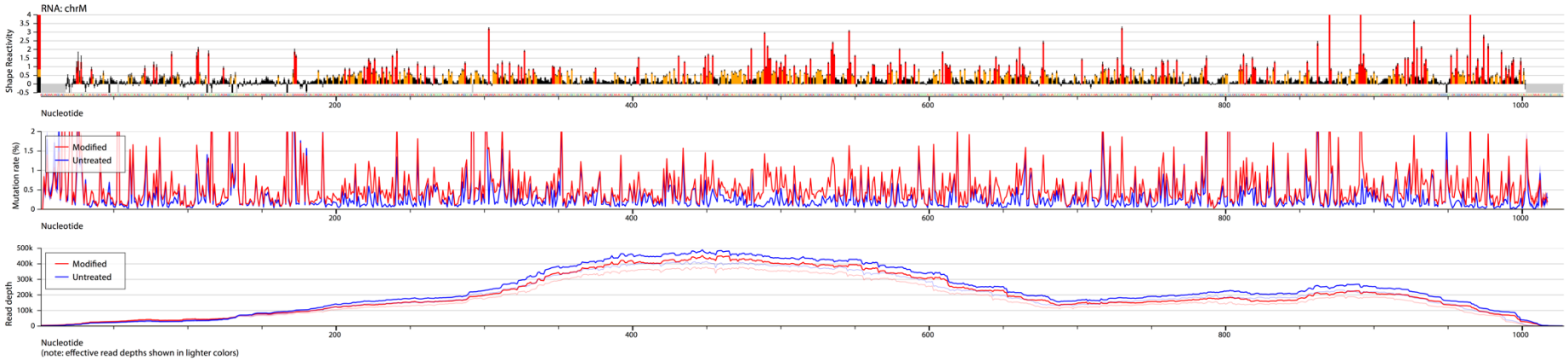
NP


RNA: chrNP
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
(note: effective read depths shown in lighter colors)

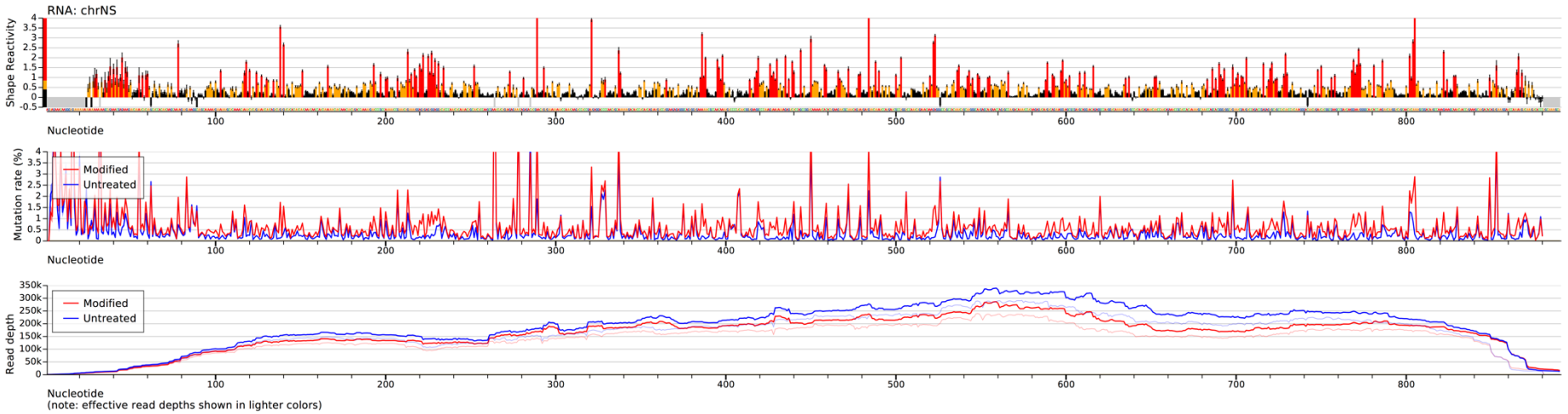
NA


RNA: chrNA
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
(note: effective read depths shown in lighter colors)

M


RNA: chrM
Shape Reactivity
Nucleotide
Modified
Untreated
Mutation rate (%)
Nucleotide
Modified
Untreated
Read depth
Nucleotide
(note: effective read depths shown in lighter colors)

NS


RNA: chrNS
Shape Reactivity
Nucleotide
Modified
Untreated
Mutation rate (%)
Nucleotide
Modified
Untreated
Read depth
Nucleotide
(note: effective read depths shown in lighter colors)

Appendix A. 4 *In virio (NMIA)*

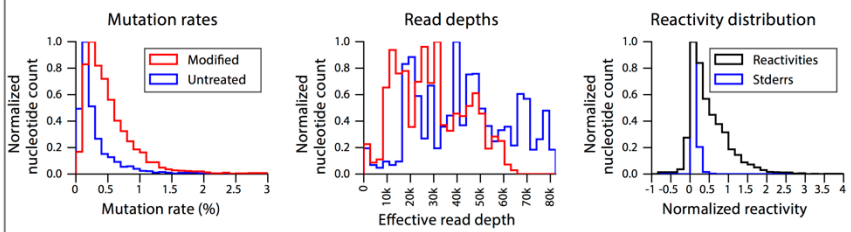
Data is plotted as an average of three experiments.

The summary statistics for mutation rates, read depth and reactivity distributions are provided for each segment. Mutation rates for SHAPE reagent-modified and non-modified samples are given, the mutation rates are expected to be higher in the SHAPE reagent-treated samples compared to the non-treated controls. Read depths for most nucleotides are expected to be above 2000. The normalised SHAPE reactivities of successful experiments should be mostly positive.

SHAPE reactivity profiles are provided for each segment. Shape reactivities are coloured using the conventional SHAPE scale, with high SHAPE reactivities (>0.85) coloured red, mid-range reactivities (0.85-0.4) in orange and low shape reactivities (<0.4) in black. Mutation rates for each nucleotide in the vRNA segment are plotted below the SHAPE reactivity profile. SHAPE reagent-treated samples (red) are expected to have higher mutation rates for most nucleotides, compared to the non-treated controls (blue). The coverage for each nucleotide is shown on the bottom panel for the SHAPE reagent treated and non-treated samples.

Summary statistics

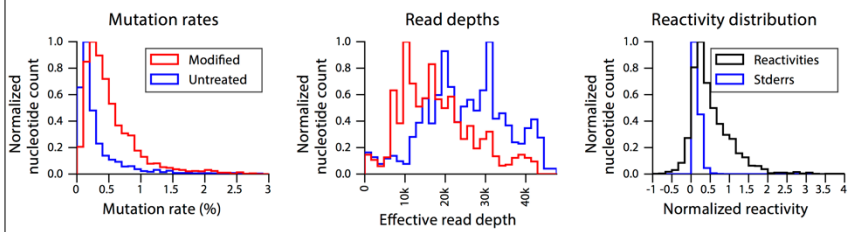
RNA: chrPB2



Modified sample:	
95th percentile rate:	1.4%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.12%
Median rate:	0.07%

Modified sample:	
Median depth:	27,705
5th percentile depth:	9,414
Untreated sample:	
Median depth:	43,797
5th percentile depth:	16,853

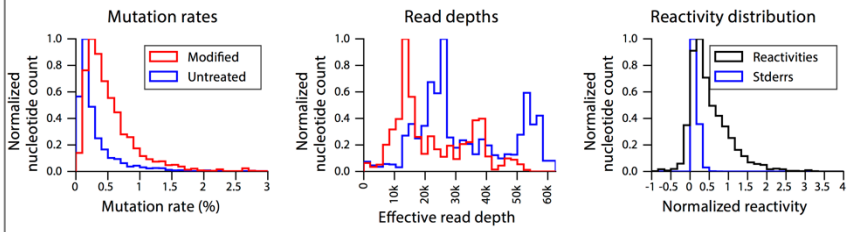
RNA: chrPB1



Modified sample:	
95th percentile rate:	1.58%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.29%
Median rate:	0.06%

Modified sample:	
Median depth:	16,690
5th percentile depth:	6,500
Untreated sample:	
Median depth:	26,783
5th percentile depth:	7,791

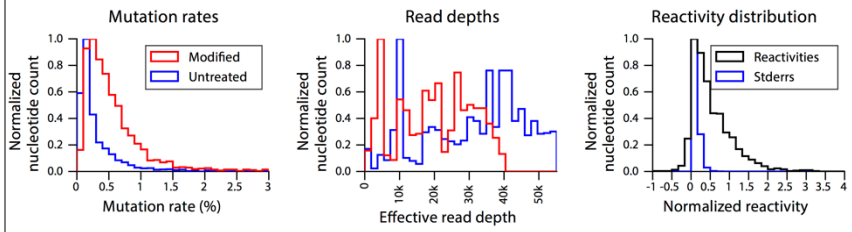
RNA: chrPA



Modified sample:	
95th percentile rate:	1.46%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.22%
Median rate:	0.07%

Modified sample:	
Median depth:	17,822
5th percentile depth:	8,031
Untreated sample:	
Median depth:	27,713
5th percentile depth:	13,749

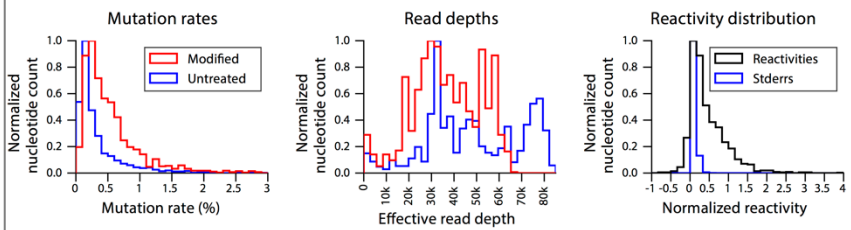
RNA: chrHA



Modified sample:	
95th percentile rate:	1.6%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.21%
Median rate:	0.06%

Modified sample:	
Median depth:	20,264
5th percentile depth:	3,472
Untreated sample:	
Median depth:	35,233
5th percentile depth:	8,465

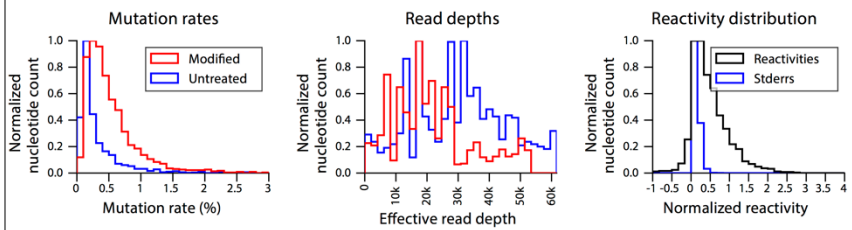
RNA: chrNP



Modified sample:	
95th percentile rate:	1.59%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.21%
Median rate:	0.07%

Modified sample:	
Median depth:	35,551
5th percentile depth:	9,855
Untreated sample:	
Median depth:	46,204
5th percentile depth:	13,066

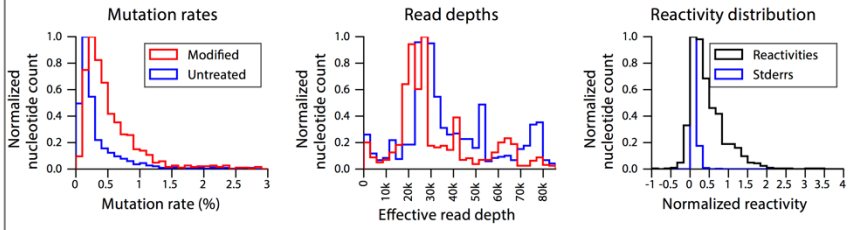
RNA: chrNA



Modified sample:	
95th percentile rate:	1.47%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.24%
Median rate:	0.07%

Modified sample:	
Median depth:	18,977
5th percentile depth:	3,597
Untreated sample:	
Median depth:	31,383
5th percentile depth:	5,651

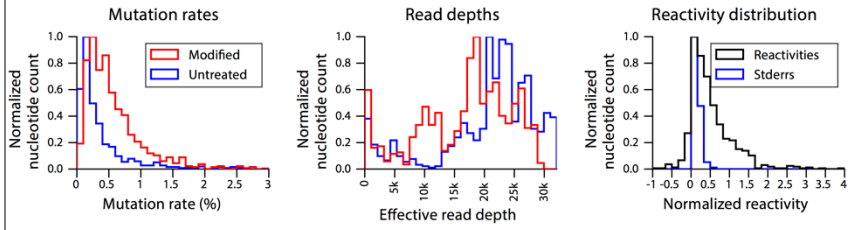
RNA: chrM



Modified sample:	
95th percentile rate:	1.69%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.29%
Median rate:	0.07%

Modified sample:	
Median depth:	26,667
5th percentile depth:	8,323
Untreated sample:	
Median depth:	32,316
5th percentile depth:	7,966

RNA: chrNS

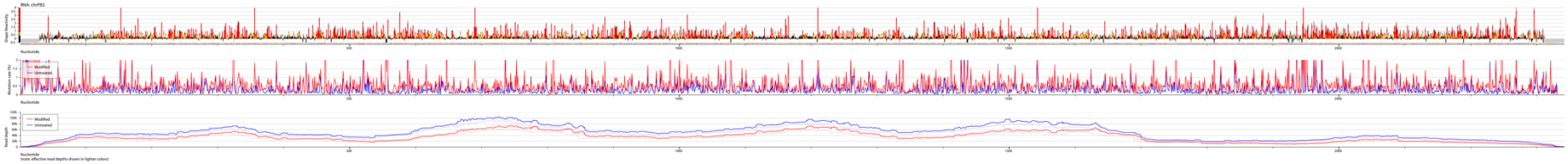


Modified sample:	
95th percentile rate:	1.7%
Median rate:	0.12%
Untreated sample:	
95th percentile rate:	1.7%
Median rate:	0.06%

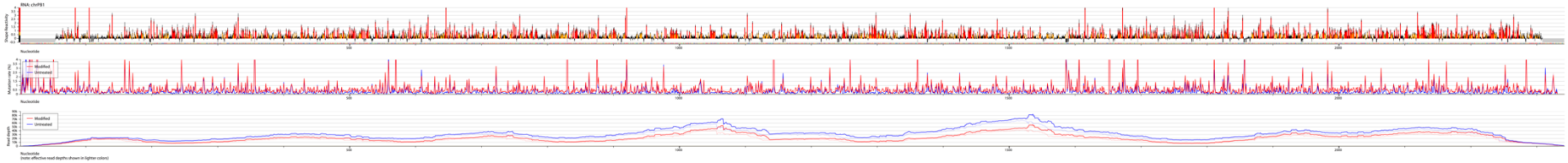
Modified sample:	
Median depth:	18,327
5th percentile depth:	885
Untreated sample:	
Median depth:	22,599
5th percentile depth:	1,644

SHAPE reactivity profiles

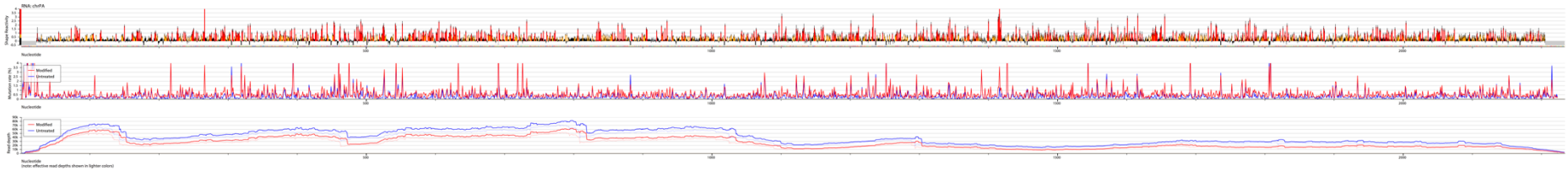
PB2



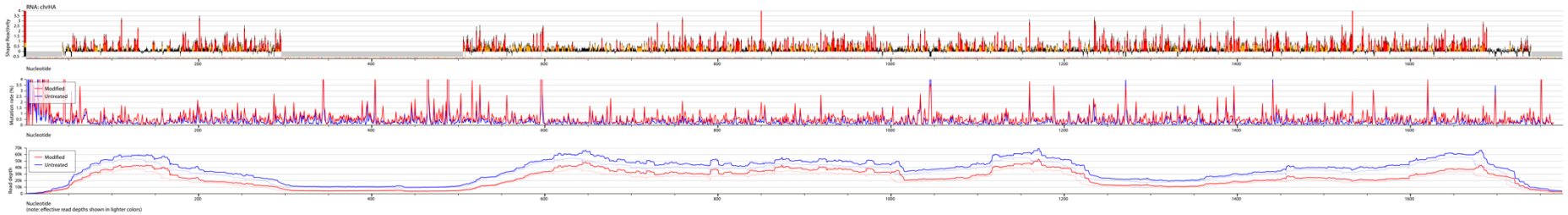
PB1



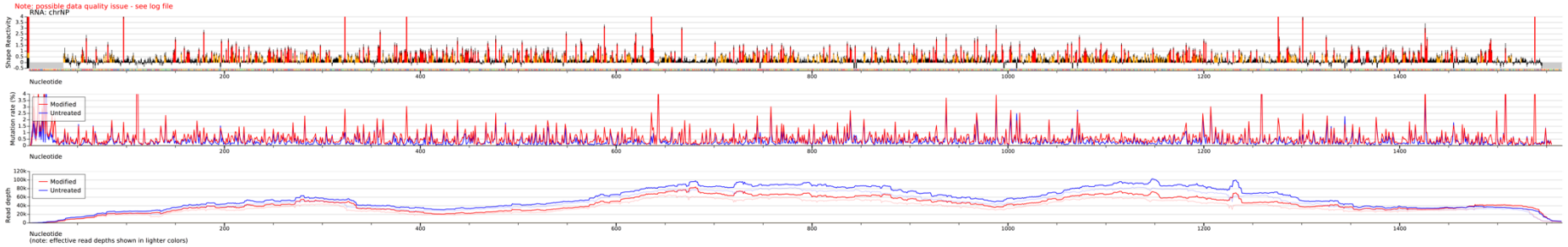
PA



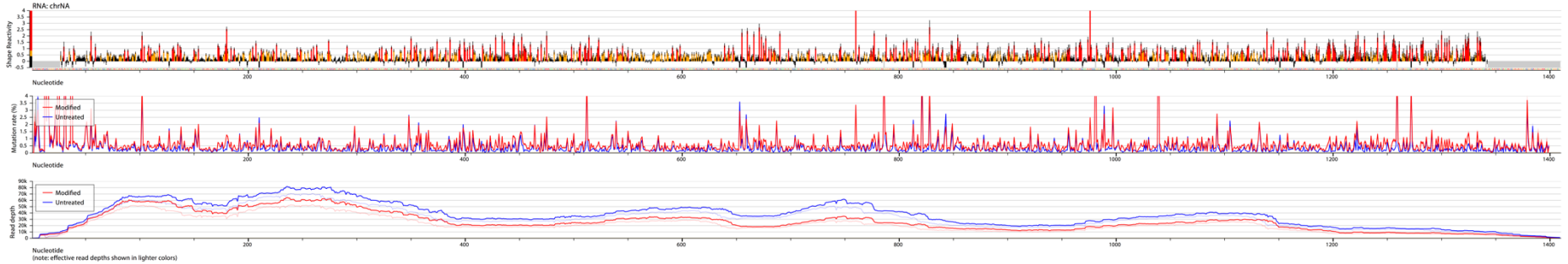
HA



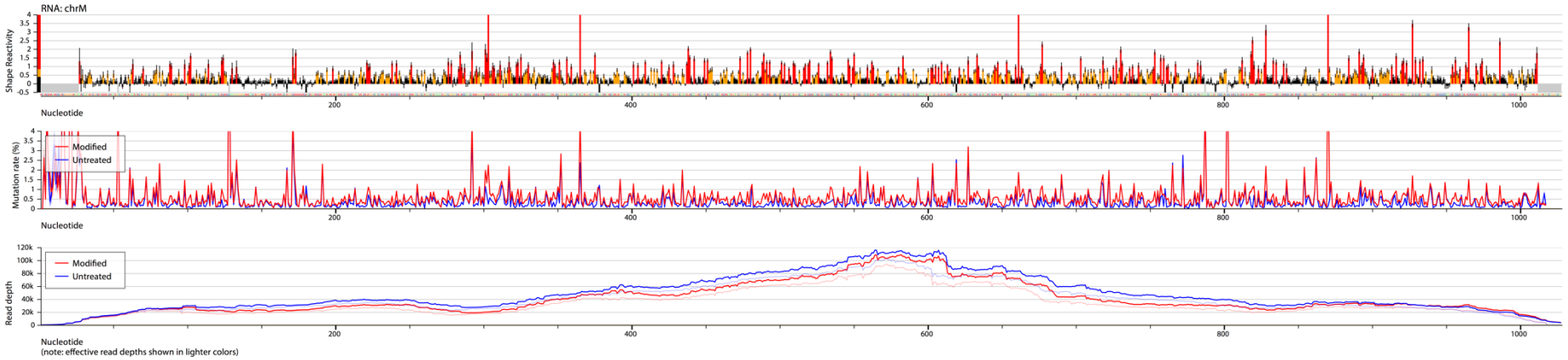
NP


Note: possible data quality issue - see log file
RNA: chrNP
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Nucleotide
Read depth
Modified
Untreated
Nucleotide
(note: effective read depths shown in lighter colors)

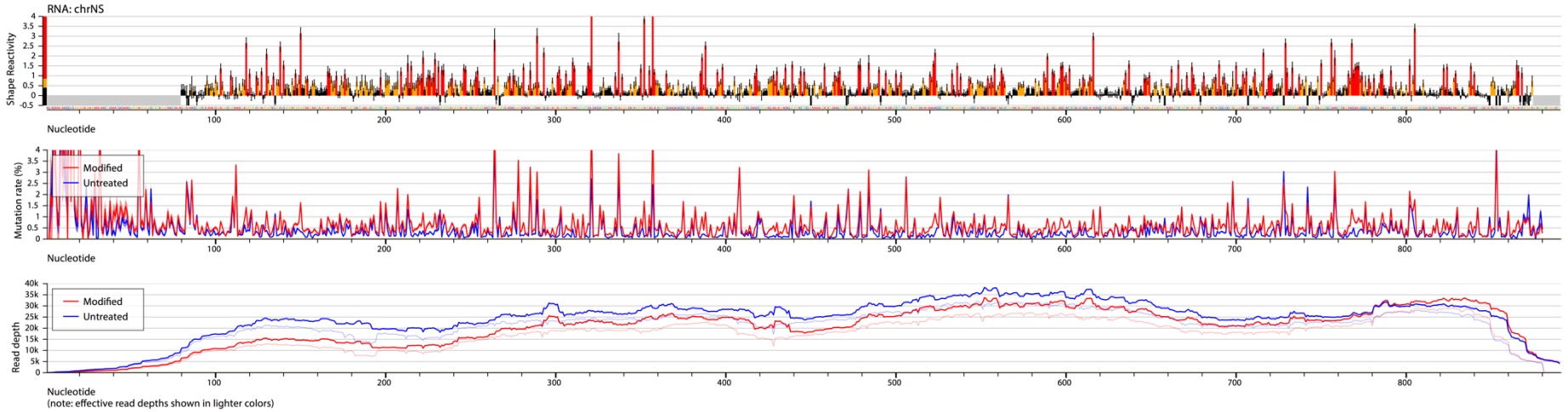
NA


RNA: chrNA
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Nucleotide
Read depth
Modified
Untreated
Nucleotide
(note: effective read depths shown in lighter colors)

M


RNA: chrM
Shape Reactivity
Nucleotide
Modified
Untreated
Mutation rate (%)
Nucleotide
Modified
Untreated
Read depth
Nucleotide
(note: effective read depths shown in lighter colors)

NS


RNA: chrNS
Shape Reactivity
Nucleotide
Modified
Untreated
Mutation rate (%)
Nucleotide
Modified
Untreated
Read depth
Nucleotide
(note: effective read depths shown in lighter colors)

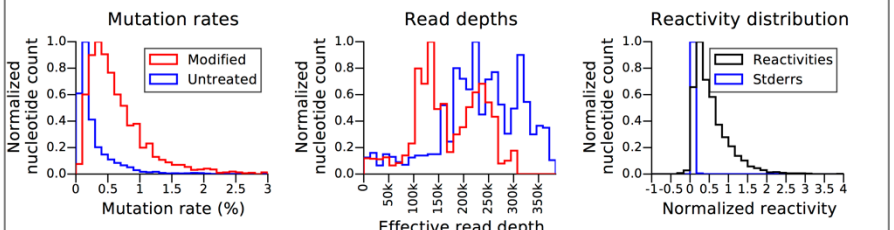
Appendix A. 5 *Ex virio: ivtRNA (1M7) experiment 1*

The summary statistics for mutation rates, read depth and reactivity distributions are provided for each segment. Mutation rates for SHAPE reagent-modified and non-modified samples are given, the mutation rates are expected to be higher in the SHAPE reagent-treated samples compared to the non-treated controls. Read depths for most nucleotides are expected to be above 2000. The normalised SHAPE reactivities of successful experiments should be mostly positive.

SHAPE reactivity profiles are provided for each segment. Shape reactivities are coloured using the conventional SHAPE scale, with high SHAPE reactivities (>0.85) coloured red, mid-range reactivities (0.85-0.4) in orange and low shape reactivities (<0.4) in black. Mutation rates for each nucleotide in the vRNA segment are plotted below the SHAPE reactivity profile. SHAPE reagent-treated samples (red) are expected to have higher mutation rates for most nucleotides, compared to the non-treated controls (blue). The coverage for each nucleotide is shown on the bottom panel for the SHAPE reagent treated and non-treated samples.

Summary statistics

RNA: chrPB2



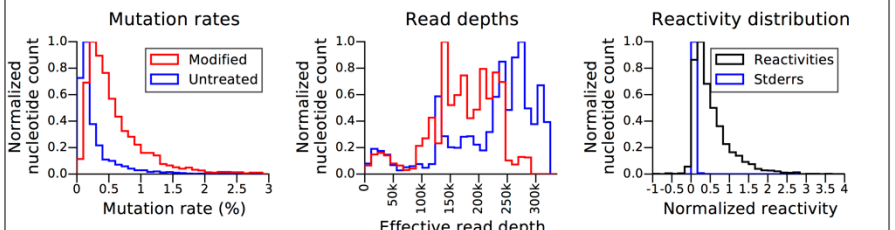
Modified sample:
95th percentile rate: 1.71%
Median rate: 0.16%

Untreated sample:
95th percentile rate: 0.96%
Median rate: 0.07%

Modified sample:
Median depth: 158,912
5th percentile depth: 46,824

Untreated sample:
Median depth: 242,852
5th percentile depth: 61,901

RNA: chrPB1



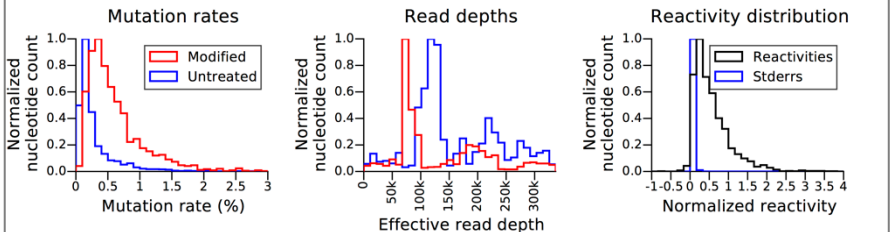
Modified sample:
95th percentile rate: 1.73%
Median rate: 0.14%

Untreated sample:
95th percentile rate: 1.09%
Median rate: 0.06%

Modified sample:
Median depth: 174,831
5th percentile depth: 38,612

Untreated sample:
Median depth: 241,559
5th percentile depth: 34,525

RNA: chrPA



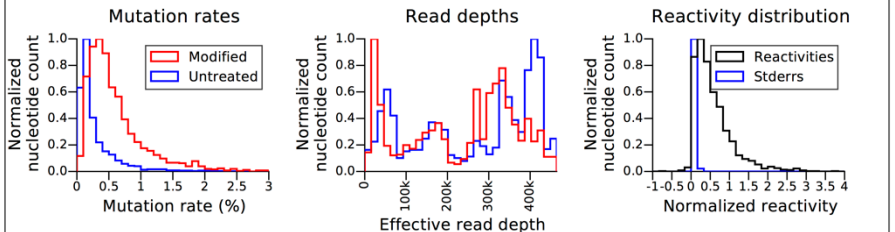
Modified sample:
95th percentile rate: 1.64%
Median rate: 0.16%

Untreated sample:
95th percentile rate: 1.1%
Median rate: 0.07%

Modified sample:
Median depth: 92,506
5th percentile depth: 40,216

Untreated sample:
Median depth: 133,735
5th percentile depth: 47,098

RNA: chrHA



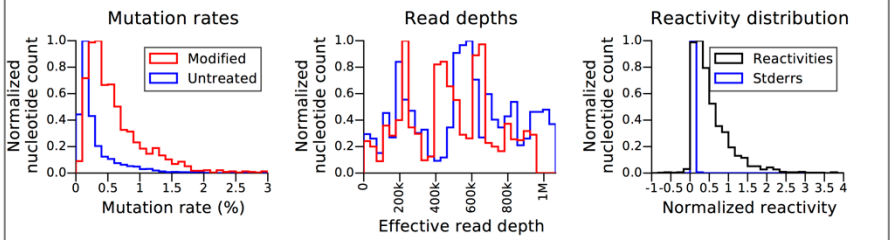
Modified sample:
95th percentile rate: 1.73%
Median rate: 0.15%

Untreated sample:
95th percentile rate: 1.04%
Median rate: 0.06%

Modified sample:
Median depth: 275,764
5th percentile depth: 20,439

Untreated sample:
Median depth: 327,024
5th percentile depth: 38,721

RNA: chrNP



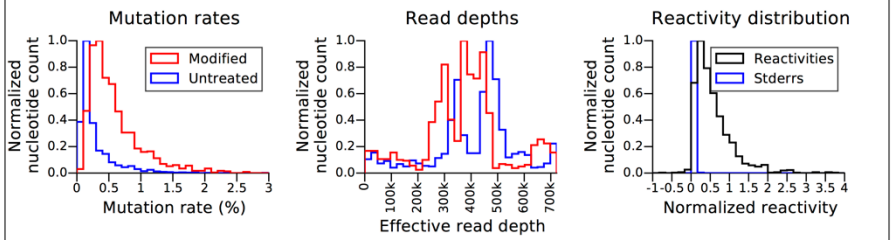
Modified sample:
95th percentile rate: 1.61%
Median rate: 0.14%

Untreated sample:
95th percentile rate: 1.09%
Median rate: 0.07%

Modified sample:
Median depth: 485,983
5th percentile depth: 108,143

Untreated sample:
Median depth: 572,187
5th percentile depth: 102,880

RNA: chrNA



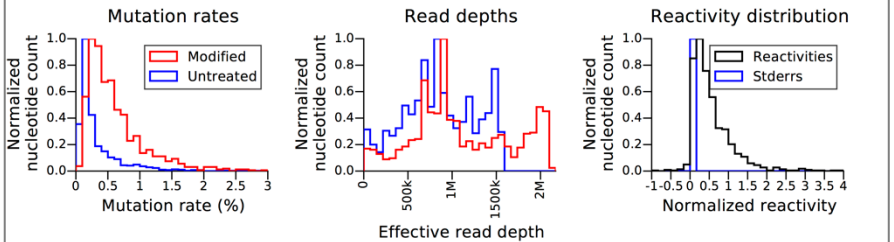
Modified sample:
95th percentile rate: 1.58%
Median rate: 0.17%

Untreated sample:
95th percentile rate: 1.01%
Median rate: 0.07%

Modified sample:
Median depth: 383,155
5th percentile depth: 72,129

Untreated sample:
Median depth: 452,156
5th percentile depth: 61,304

RNA: chrM



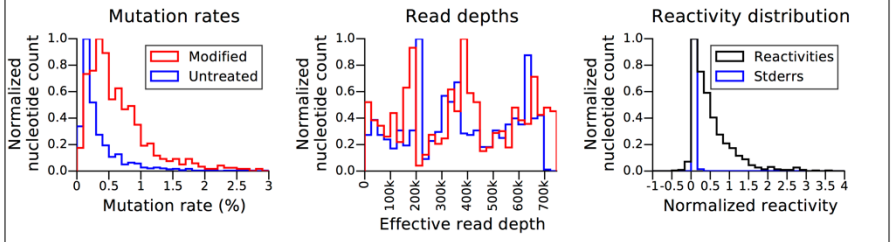
Modified sample:
95th percentile rate: 1.6%
Median rate: 0.16%

Untreated sample:
95th percentile rate: 1.11%
Median rate: 0.08%

Modified sample:
Median depth: 960,373
5th percentile depth: 180,730

Untreated sample:
Median depth: 827,398
5th percentile depth: 135,792

RNA: chrNS



Modified sample:
95th percentile rate: 1.82%
Median rate: 0.13%

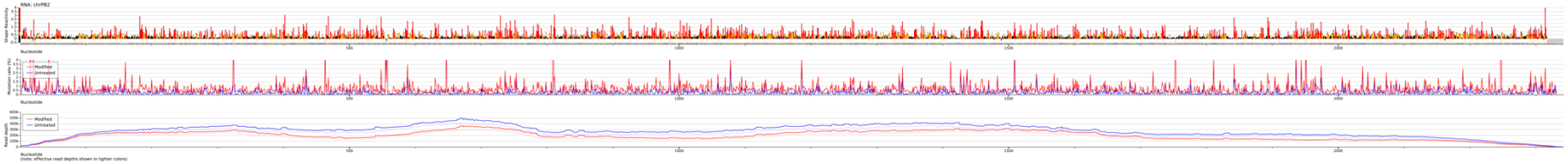
Untreated sample:
95th percentile rate: 1.26%
Median rate: 0.08%

Modified sample:
Median depth: 382,671
5th percentile depth: 28,663

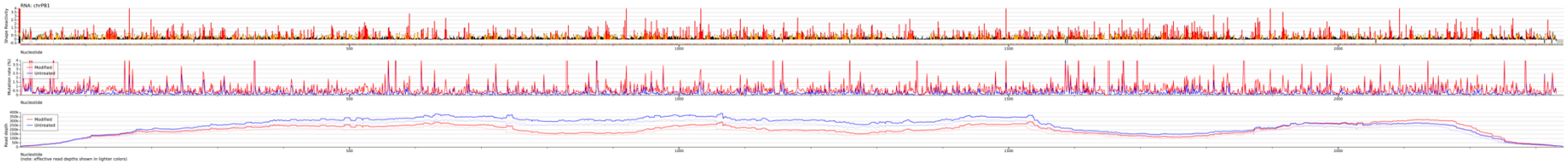
Untreated sample:
Median depth: 355,365
5th percentile depth: 39,174

SHAPE reactivity profiles

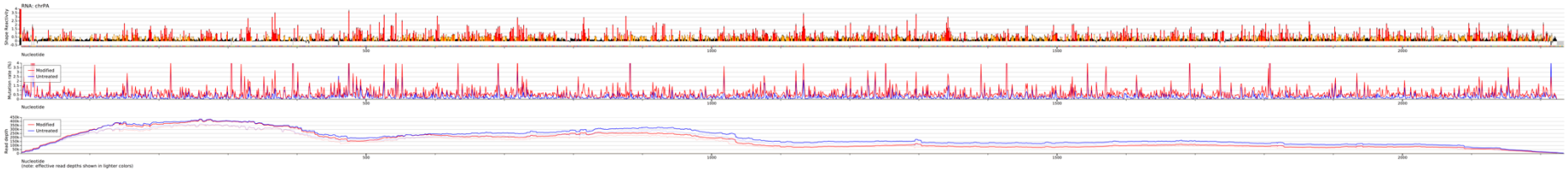
PB2



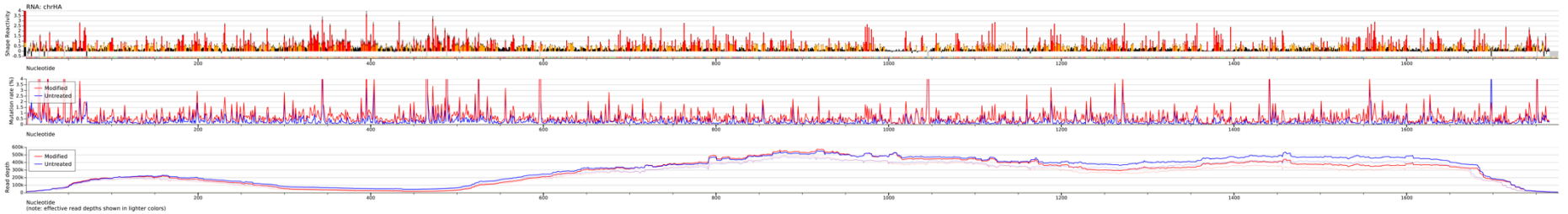
PB1



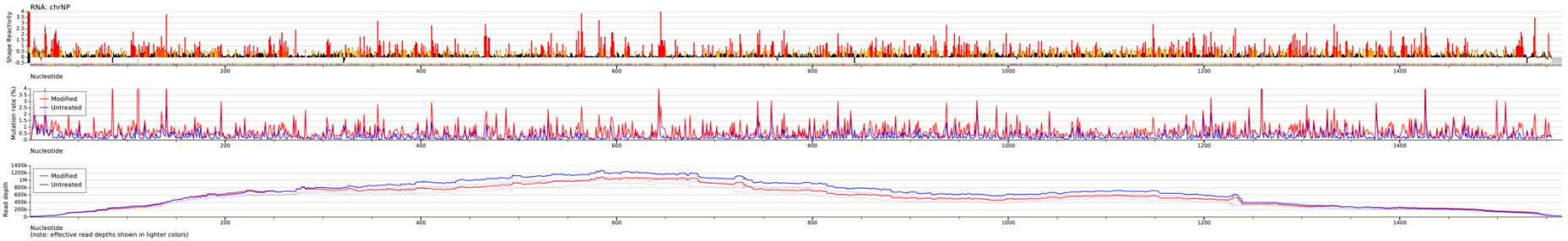
PA



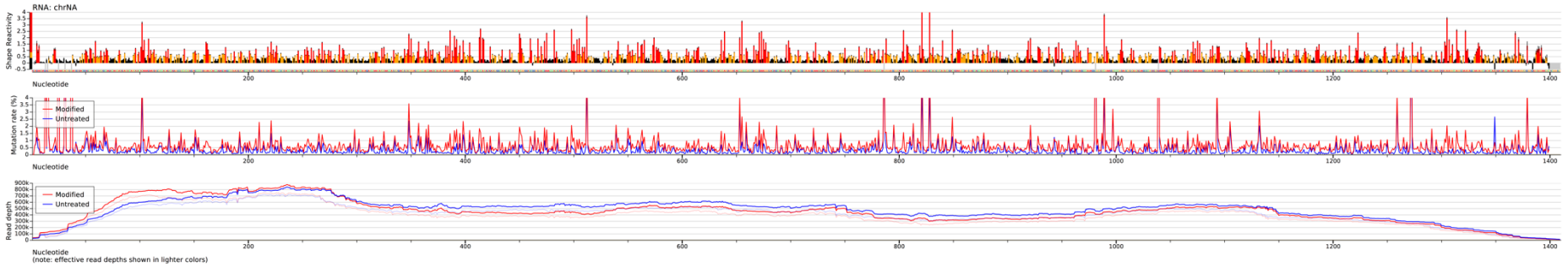
HA



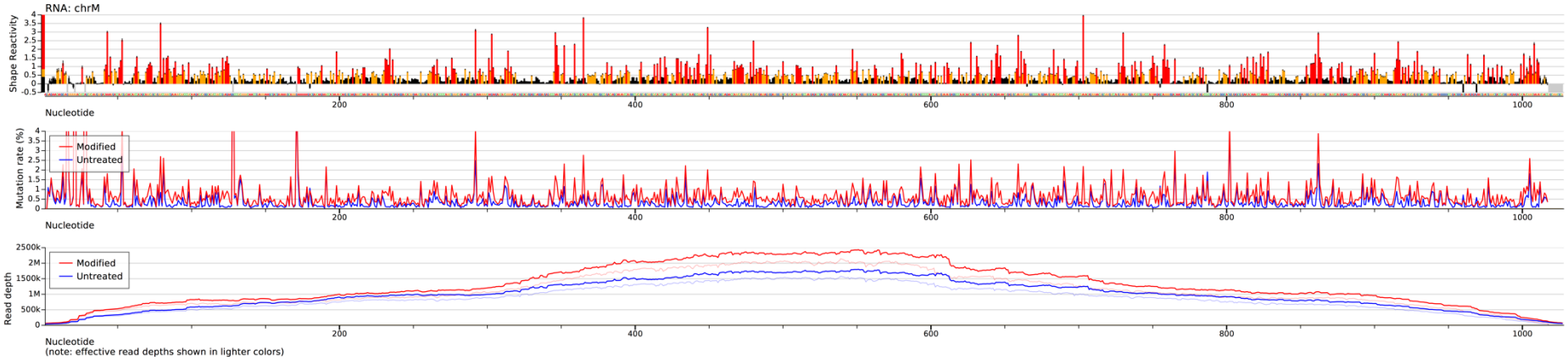
NP


RNA: chrNP
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
Nucleotide
(note: effective read depths shown in lighter colors)

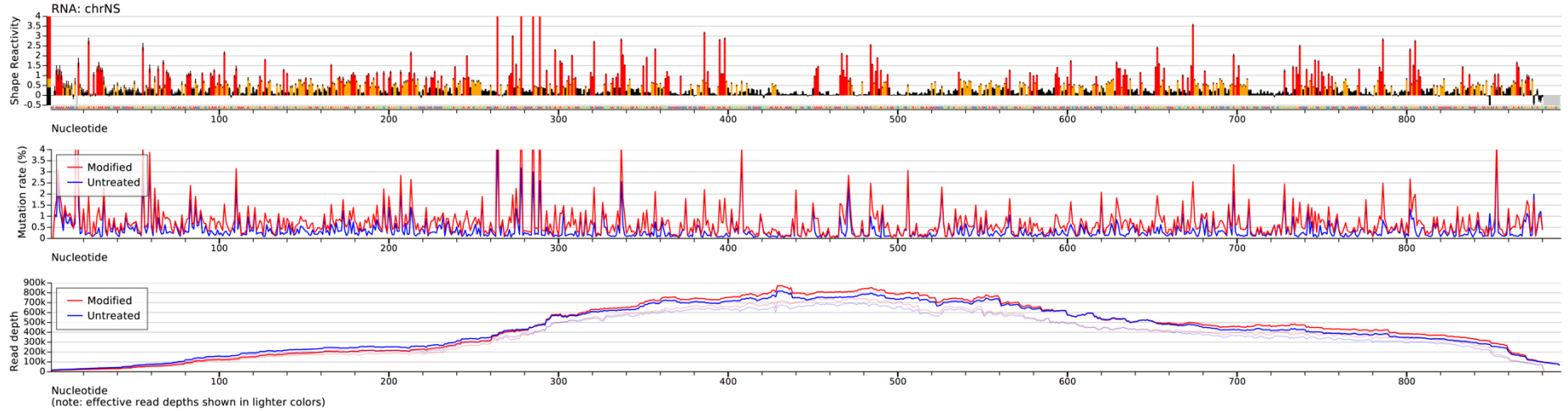
NA


RNA: chrNA
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
Nucleotide
(note: effective read depths shown in lighter colors)

M


RNA: chrM
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Nucleotide
Read depth
Modified
Untreated
Nucleotide
(note: effective read depths shown in lighter colors)

NS


RNA: chrNS
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Nucleotide
Read depth
Modified
Untreated
Nucleotide
(note: effective read depths shown in lighter colors)

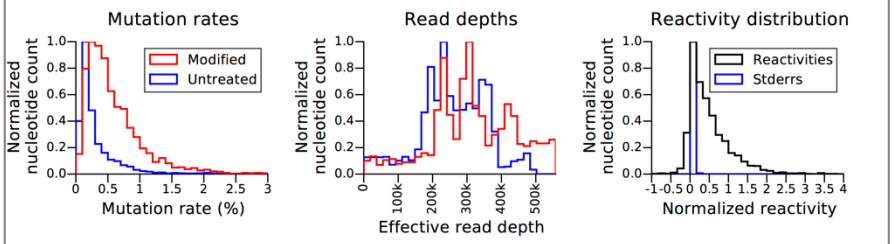
Appendix A. 6 *Ex virio: ivtRNA (1M7) experiment 2*

The summary statistics for mutation rates, read depth and reactivity distributions are provided for each segment. Mutation rates for SHAPE reagent-modified and non-modified samples are given, the mutation rates are expected to be higher in the SHAPE reagent-treated samples compared to the non-treated controls. Read depths for most nucleotides are expected to be above 2000. The normalised SHAPE reactivities of successful experiments should be mostly positive.

SHAPE reactivity profiles are provided for each segment. Shape reactivities are coloured using the conventional SHAPE scale, with high SHAPE reactivities (>0.85) coloured red, mid-range reactivities (0.85-0.4) in orange and low shape reactivities (<0.4) in black. Mutation rates for each nucleotide in the vRNA segment are plotted below the SHAPE reactivity profile. SHAPE reagent-treated samples (red) are expected to have higher mutation rates for most nucleotides, compared to the non-treated controls (blue). The coverage for each nucleotide is shown on the bottom panel for the SHAPE reagent treated and non-treated samples.

Summary statistics

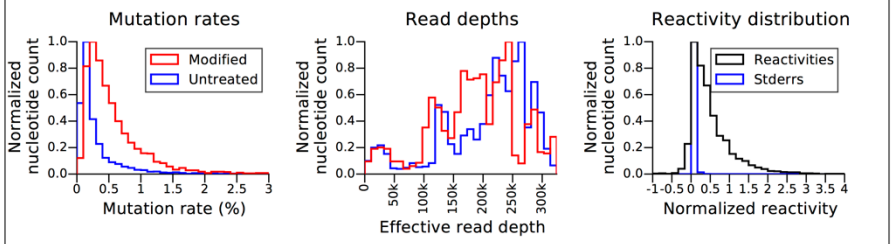
RNA: chrPB2



Modified sample:	
95th percentile rate:	1.57%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.02%
Median rate:	0.07%

Modified sample:	
Median depth:	312,234
5th percentile depth:	85,797
Untreated sample:	
Median depth:	254,862
5th percentile depth:	64,170

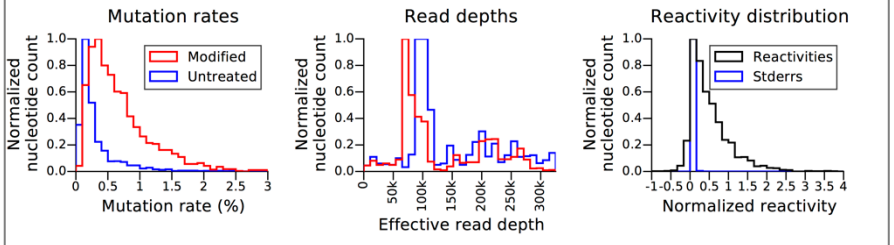
RNA: chrPB1



Modified sample:	
95th percentile rate:	1.59%
Median rate:	0.13%
Untreated sample:	
95th percentile rate:	1.14%
Median rate:	0.06%

Modified sample:	
Median depth:	195,912
5th percentile depth:	37,536
Untreated sample:	
Median depth:	226,208
5th percentile depth:	31,143

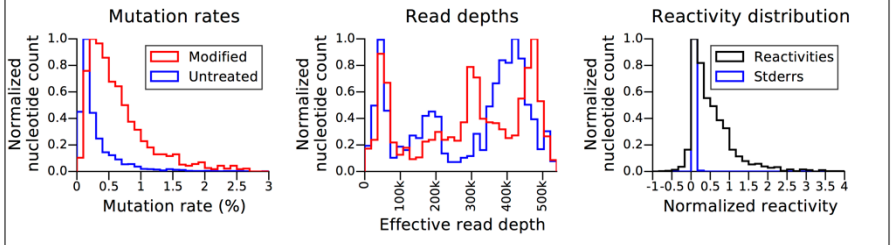
RNA: chrPA



Modified sample:	
95th percentile rate:	1.83%
Median rate:	0.17%
Untreated sample:	
95th percentile rate:	1.15%
Median rate:	0.08%

Modified sample:	
Median depth:	103,573
5th percentile depth:	46,952
Untreated sample:	
Median depth:	112,832
5th percentile depth:	44,742

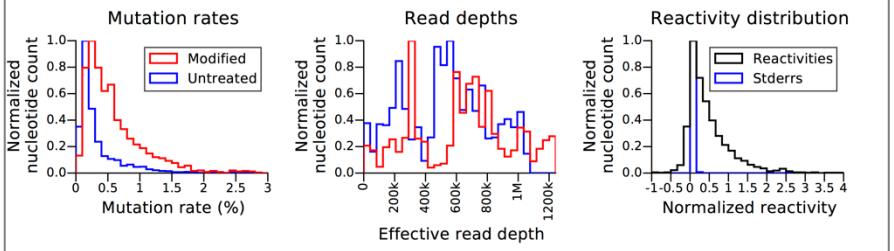
RNA: chrHA



Modified sample:	
95th percentile rate:	1.84%
Median rate:	0.15%
Untreated sample:	
95th percentile rate:	1.11%
Median rate:	0.07%

Modified sample:	
Median depth:	316,879
5th percentile depth:	40,229
Untreated sample:	
Median depth:	356,917
5th percentile depth:	33,275

RNA: chrNP



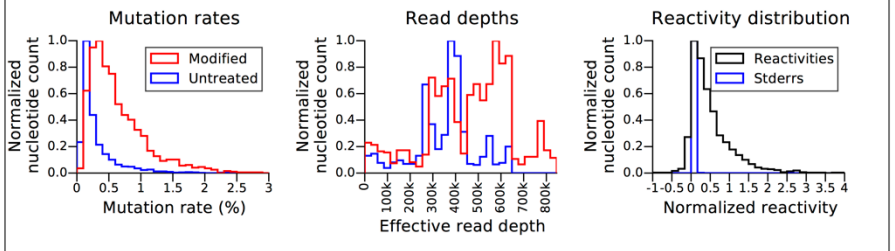
Modified sample:
95th percentile rate: 1.53%
Median rate: 0.13%

Untreated sample:
95th percentile rate: 1.17%
Median rate: 0.08%

Modified sample:
Median depth: 676,127
5th percentile depth: 135,483

Untreated sample:
Median depth: 546,456
5th percentile depth: 101,844

RNA: chrNA



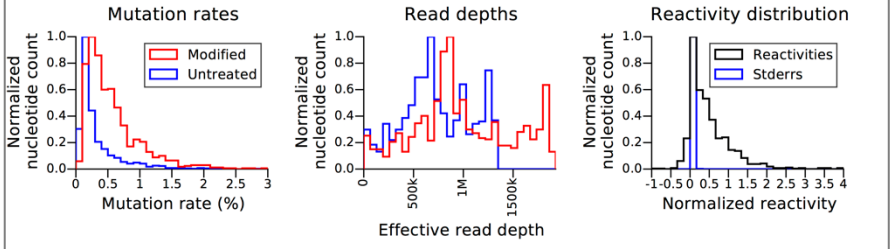
Modified sample:
95th percentile rate: 1.7%
Median rate: 0.17%

Untreated sample:
95th percentile rate: 1.1%
Median rate: 0.08%

Modified sample:
Median depth: 501,145
5th percentile depth: 83,099

Untreated sample:
Median depth: 379,096
5th percentile depth: 54,661

RNA: chrM



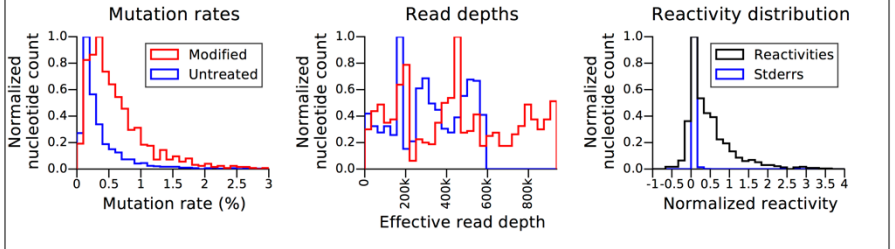
Modified sample:
95th percentile rate: 1.47%
Median rate: 0.14%

Untreated sample:
95th percentile rate: 1.14%
Median rate: 0.08%

Modified sample:
Median depth: 891,858
5th percentile depth: 149,493

Untreated sample:
Median depth: 696,323
5th percentile depth: 119,797

RNA: chrNS



Modified sample:
95th percentile rate: 1.75%
Median rate: 0.12%

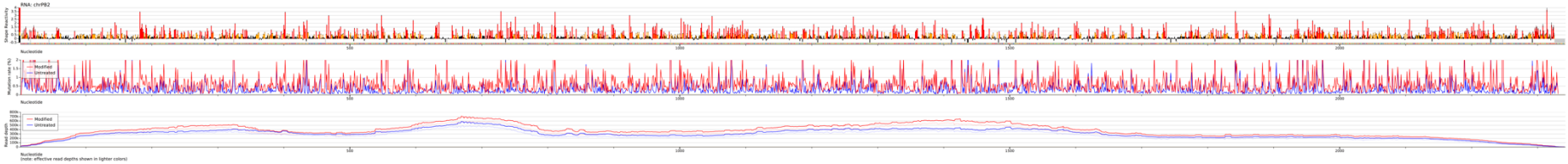
Untreated sample:
95th percentile rate: 1.26%
Median rate: 0.08%

Modified sample:
Median depth: 440,498
5th percentile depth: 48,186

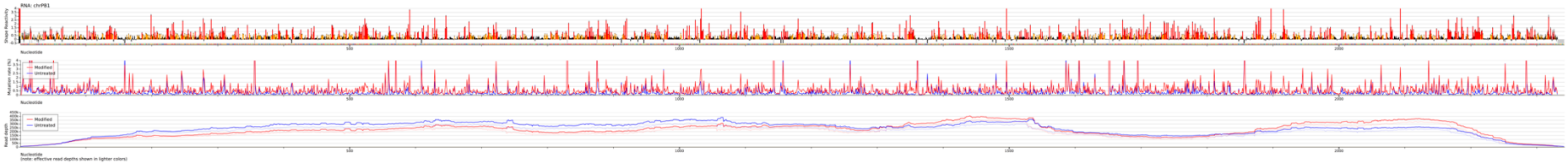
Untreated sample:
Median depth: 311,045
5th percentile depth: 31,666

SHAPE reactivity profiles

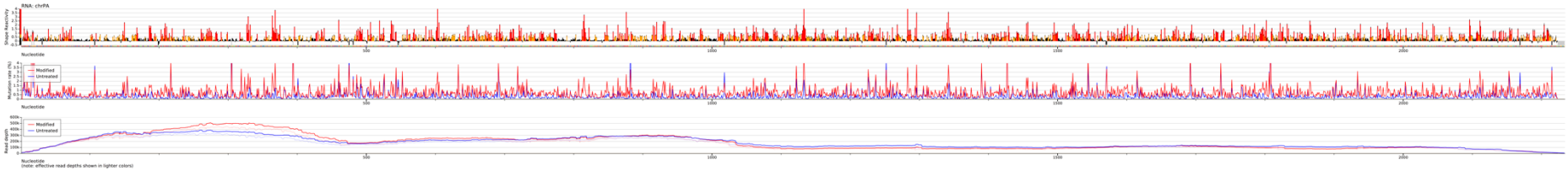
PB2



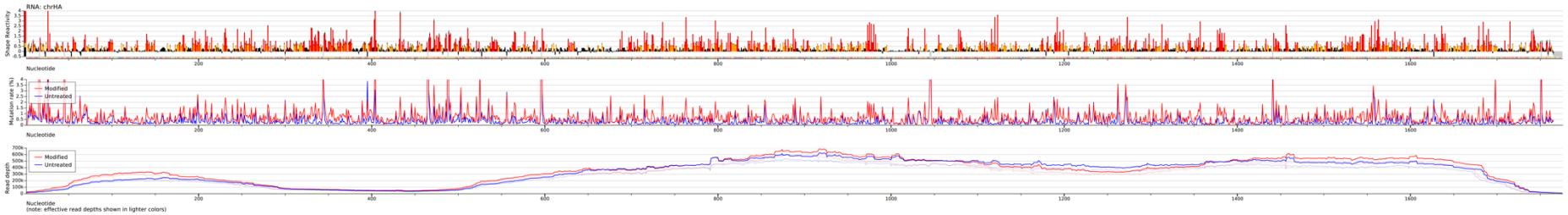
PB1



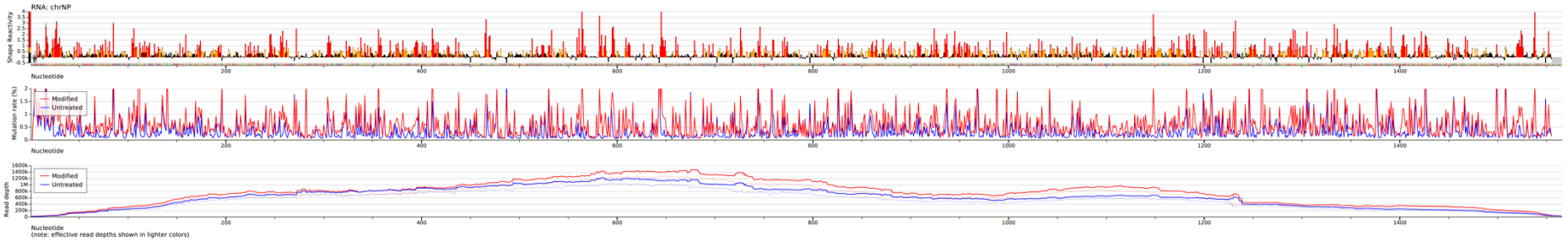
PA



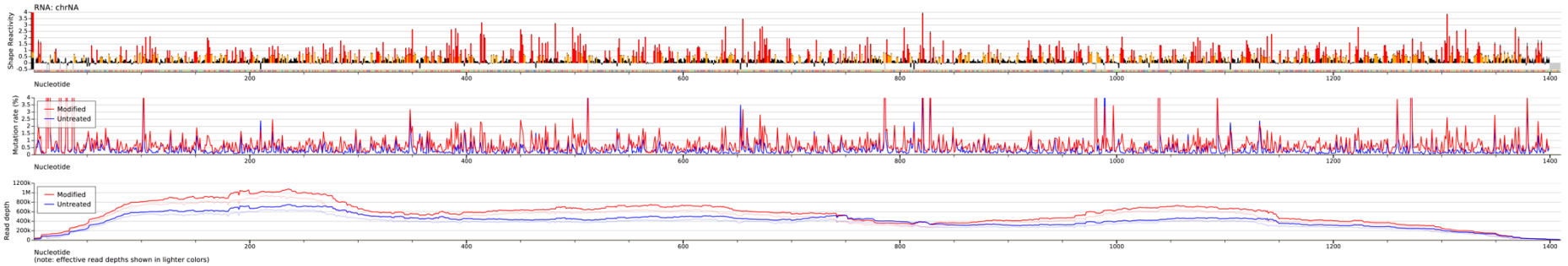
HA



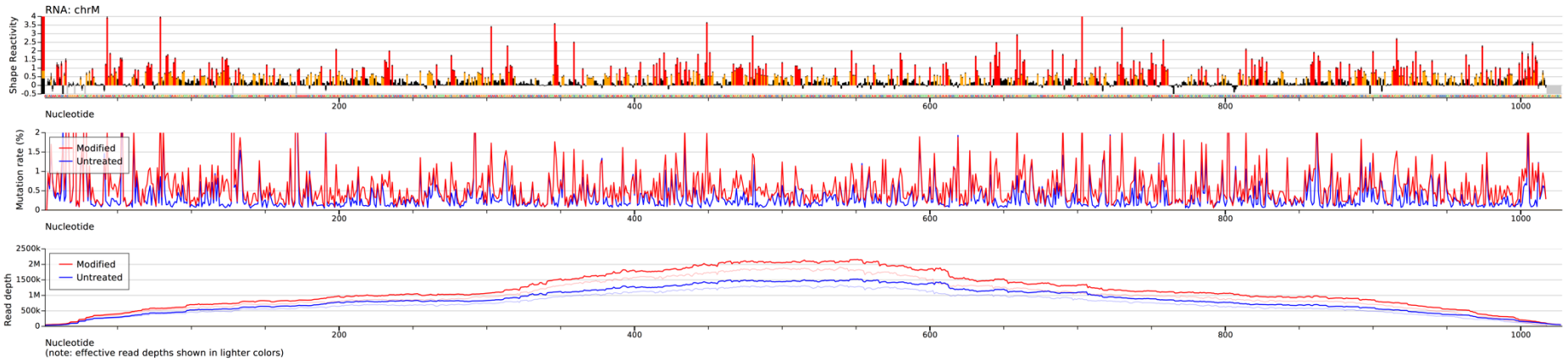
NP


RNA: chrNP
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
(note: effective read depths shown in lighter colors)

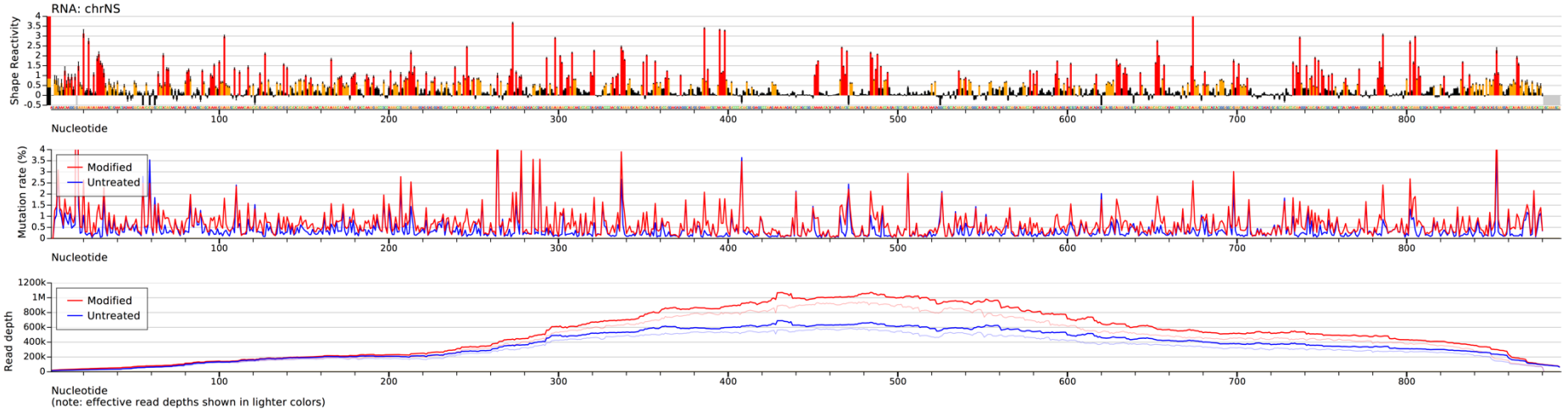
NA


RNA: chrNA
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
(note: effective read depths shown in lighter colors)

M


RNA: chrM
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
(note: effective read depths shown in lighter colors)

NS


RNA: chrNS
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
(note: effective read depths shown in lighter colors)

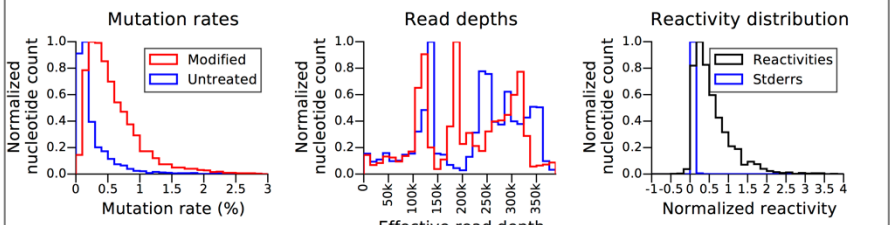
Appendix A. 7 *Ex virio: nkvRNA (1M7)*

The summary statistics for mutation rates, read depth and reactivity distributions are provided for each segment. Mutation rates for SHAPE reagent-modified and non-modified samples are given, the mutation rates are expected to be higher in the SHAPE reagent-treated samples compared to the non-treated controls. Read depths for most nucleotides are expected to be above 2000. The normalised SHAPE reactivities of successful experiments should be mostly positive.

SHAPE reactivity profiles are provided for each segment. Shape reactivities are coloured using the conventional SHAPE scale, with high SHAPE reactivities (>0.85) coloured red, mid-range reactivities (0.85-0.4) in orange and low shape reactivities (<0.4) in black. Mutation rates for each nucleotide in the vRNA segment are plotted below the SHAPE reactivity profile. SHAPE reagent-treated samples (red) are expected to have higher mutation rates for most nucleotides, compared to the non-treated controls (blue). The coverage for each nucleotide is shown on the bottom panel for the SHAPE reagent treated and non-treated samples.

Summary statistics

RNA: chrPB2



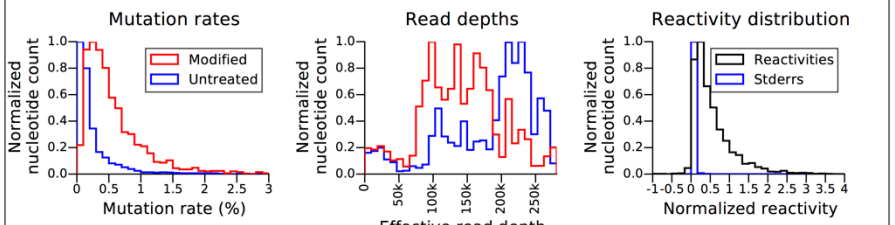
Modified sample:
95th percentile rate: 1.6%
Median rate: 0.14%

Untreated sample:
95th percentile rate: 0.93%
Median rate: 0.05%

Modified sample:
Median depth: 206,545
5th percentile depth: 53,128

Untreated sample:
Median depth: 251,109
5th percentile depth: 49,497

RNA: chrPB1



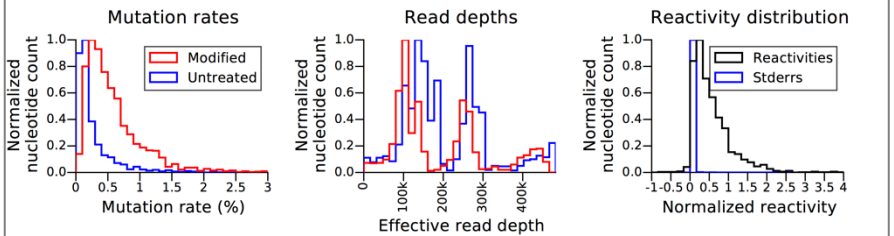
Modified sample:
95th percentile rate: 1.73%
Median rate: 0.12%

Untreated sample:
95th percentile rate: 1.08%
Median rate: 0.05%

Modified sample:
Median depth: 137,872
5th percentile depth: 28,902

Untreated sample:
Median depth: 206,572
5th percentile depth: 26,804

RNA: chrPA



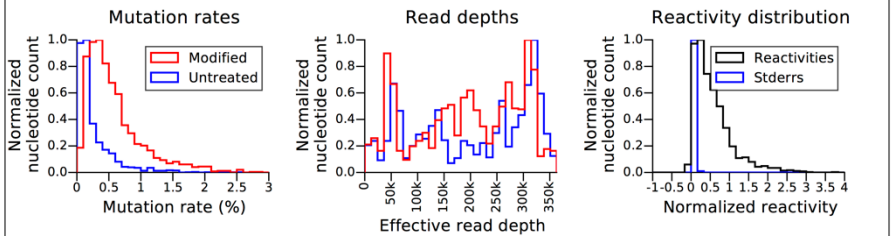
Modified sample:
95th percentile rate: 1.58%
Median rate: 0.13%

Untreated sample:
95th percentile rate: 1.05%
Median rate: 0.05%

Modified sample:
Median depth: 140,160
5th percentile depth: 59,103

Untreated sample:
Median depth: 181,761
5th percentile depth: 65,370

RNA: chrHA



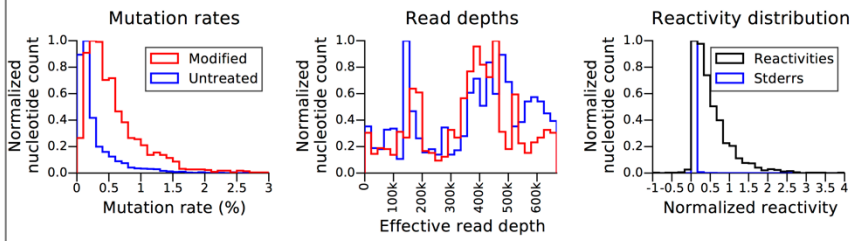
Modified sample:
95th percentile rate: 1.71%
Median rate: 0.13%

Untreated sample:
95th percentile rate: 1.02%
Median rate: 0.05%

Modified sample:
Median depth: 202,497
5th percentile depth: 31,799

Untreated sample:
Median depth: 229,649
5th percentile depth: 27,812

RNA: chrNP



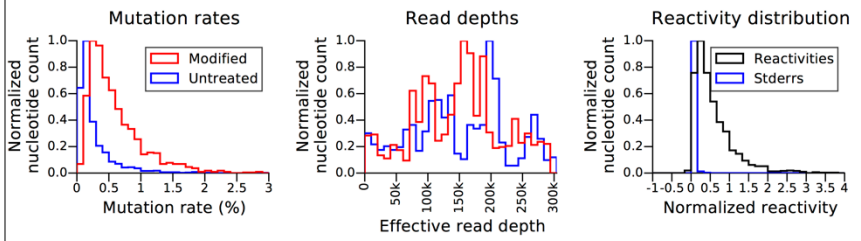
Modified sample:
95th percentile rate: 1.54%
Median rate: 0.12%

Untreated sample:
95th percentile rate: 1.09%
Median rate: 0.05%

Modified sample:
Median depth: 386,006
5th percentile depth: 55,423

Untreated sample:
Median depth: 422,046
5th percentile depth: 60,471

RNA: chrNA



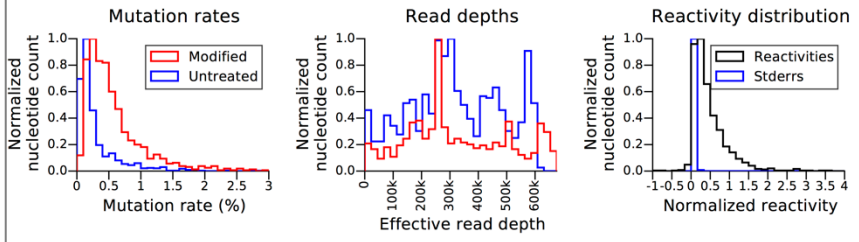
Modified sample:
95th percentile rate: 1.6%
Median rate: 0.15%

Untreated sample:
95th percentile rate: 0.95%
Median rate: 0.06%

Modified sample:
Median depth: 155,729
5th percentile depth: 19,522

Untreated sample:
Median depth: 166,387
5th percentile depth: 16,597

RNA: chrM



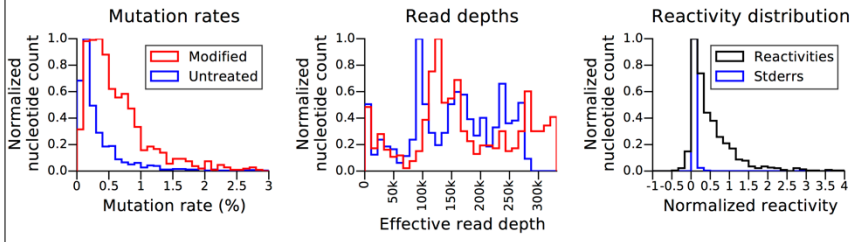
Modified sample:
95th percentile rate: 1.64%
Median rate: 0.14%

Untreated sample:
95th percentile rate: 1.19%
Median rate: 0.06%

Modified sample:
Median depth: 284,059
5th percentile depth: 45,998

Untreated sample:
Median depth: 300,936
5th percentile depth: 42,891

RNA: chrNS



Modified sample:
95th percentile rate: 1.93%
Median rate: 0.11%

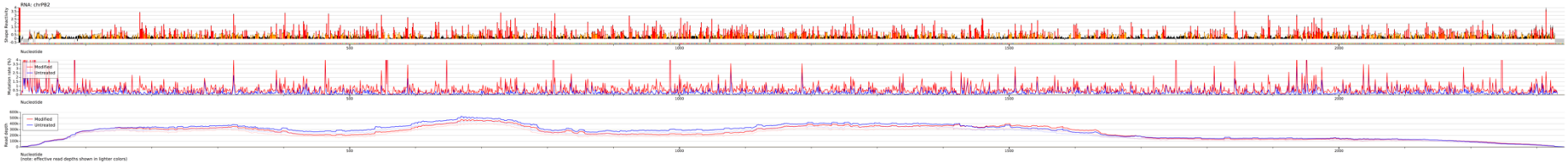
Untreated sample:
95th percentile rate: 1.38%
Median rate: 0.06%

Modified sample:
Median depth: 155,407
5th percentile depth: 11,179

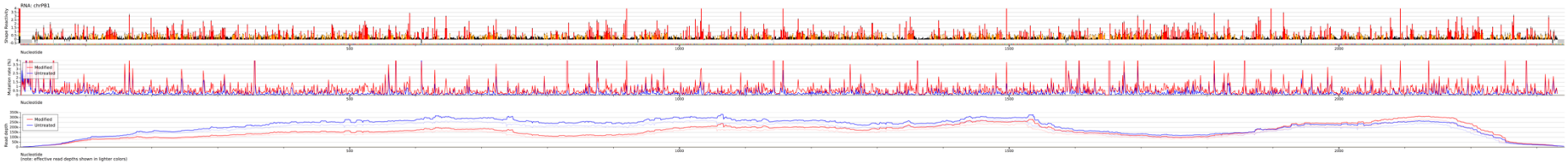
Untreated sample:
Median depth: 158,453
5th percentile depth: 10,060

SHAPE reactivity profiles

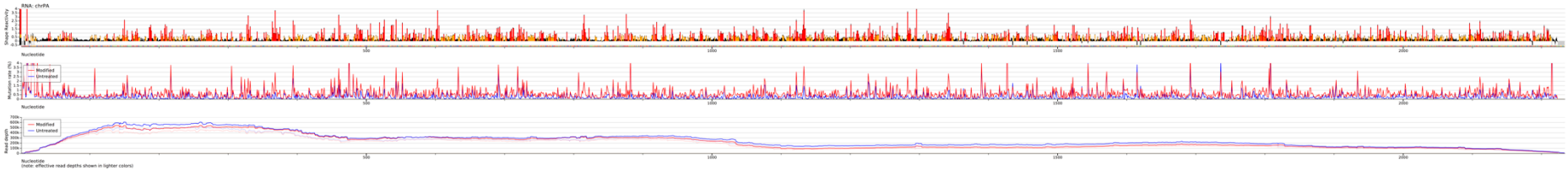
PB2



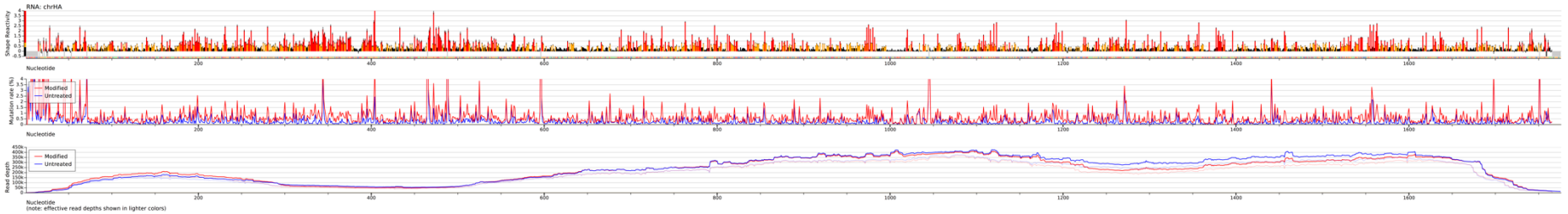
PB1



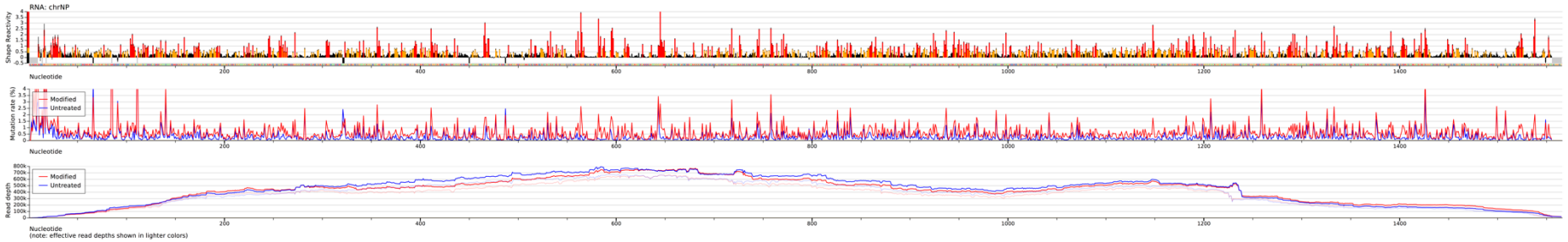
PA



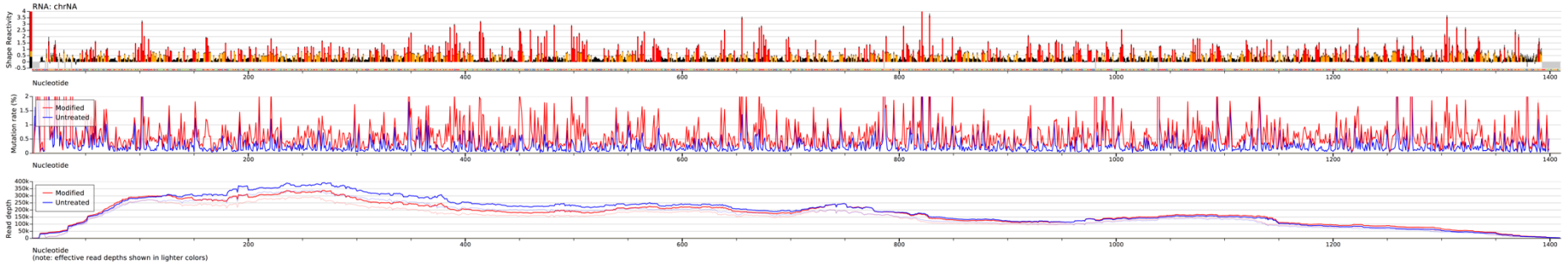
HA



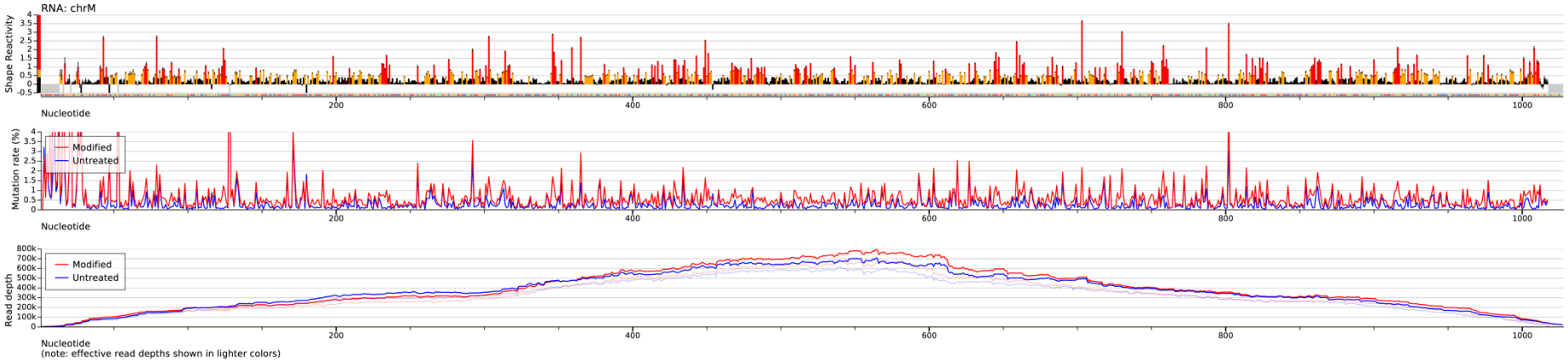
NP


RNA: chrNP
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Nucleotide
Read depth
Modified
Untreated
Nucleotide
(note: effective read depths shown in lighter colors)

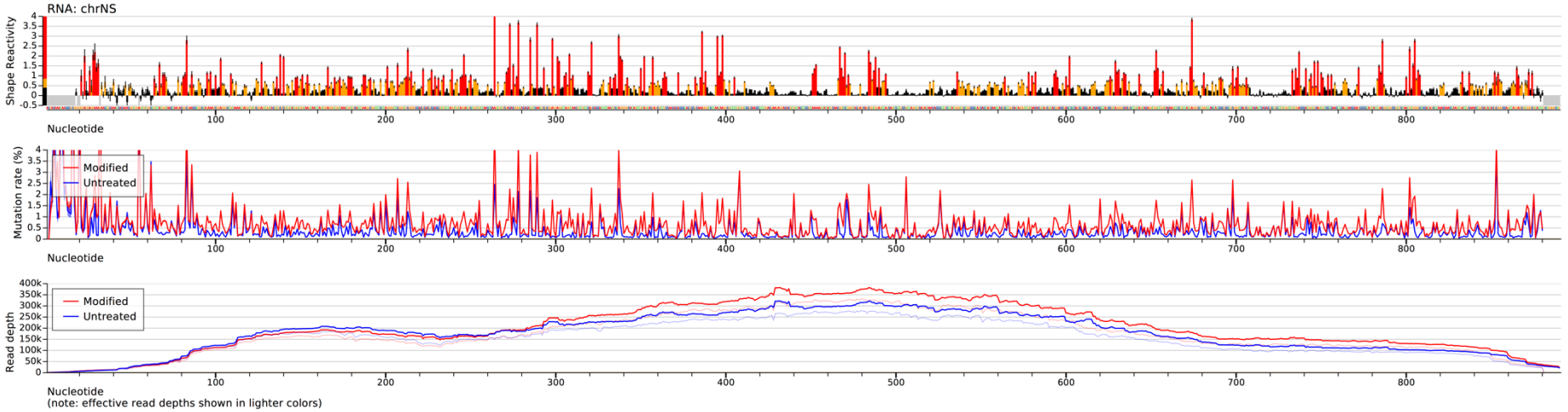
NA


RNA: chrNA
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Nucleotide
Read depth
Modified
Untreated
Nucleotide
(note: effective read depths shown in lighter colors)

M


RNA: chrM
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
(note: effective read depths shown in lighter colors)

NS


RNA: chrNS
Shape Reactivity
Nucleotide
Mutation rate (%)
Modified
Untreated
Read depth
(note: effective read depths shown in lighter colors)

Appendix B: SHAPE-MaP analysis scripts

Appendix B. 1 *Read trimming*

```
trim_galore --nextera\
--fastqc\
-o {output_directory}\
Sequencing_reads.fastq.gz
```

Appendix B. 2 *Sequence alignment, mutation counting and SHAPE profile*

The script uses Shape Mapper 2 package

```
shapemapper --name WSNrepeat_bowtie\
--target WSN_HA.fa WSN_M.fa WSN_NP.fa WSN_NS.fa WSN_PA.fa WSN_PB2.fa WSN_NA.fa WSN_PB1.fa\
--out WSNrepeat_bowtie_shape\
--modified --unpaired-folder WSNrepeatPlus\
--untreated --unpaired-folder WSNrepeatMinus\
--random-primer-len 9
```

Appendix B. 3 *Secondary RNA structure prediction*

The script uses RNAStructure package. An example script is shown for HA segment, for other segments an identical script was used, where the name “HA” was replaced by the names of other segments.

```
Fold WSN_HA.fa \
HA_RNP_WSN.ct \
-md 150 \
--constraint promoterssWSN_HA.const  --SHAPE WSN_HA_normalised.shape


python pvclient.py --ct HA_RNP_WSN.ct --shape WSN_HA_normalised.shape --structures 1 --out
HA_RNP_WSN


partition WSN_HA.fa HA_RNP_WSN.pfs -md 150 -sh WSN_HA_normalised.shape -C
promoterssWSN_HA.const


ProbabilityPlot HA_RNP_WSN.pfs HA_RNP_WSN.dp -t
```

```
MaxExpect HA_RNP_WSN.pfs HA_RNP_WSN_maxexpect.ct


python pvclient.py --ct HA_RNP_WSN_maxexpect.ct --shape WSN_HA_normalised.shape --structures 1 --out
HA_RNP_WSN_maxexpect


python arcPlot.py --ct HA_RNP_WSN_maxexpect.ct --fasta WSN_HA.fa --shape WSN_HA_normalised.shape
--pairprob HA_RNP_WSN.dp ARCPLOT_HA_RNP_WSN_150_maxexpect.pdf
```

An example promoter constrain file is shown for HA segment, other segments where formatted in the same way, adjusted for a segment legth.

promoterssWSN_HA.const

```
DS:
-1
SS:
1
2
3
4
5
6
7
8
11
12
1775
1774
1773
1772
1771
1770
1769
1766
1765
-1
Mod:
-1
Pairs:
-1 -1
FMN:
-1
Forbids:
-1 -1
```

Appendix C: SPLASH read processing scripts

Appendix C. 1 *Read trimming*

```
/home/fodor/bernadeta/trim_galore_zip/trim_galore --fastqc {Sequencing_Reads}.fastq.gz
/home/fodor/bernadeta/trim_galore_zip/trim_galore -a AAAAAAAAAA {Sequencing_Reads_trimmed}.fq.gz
```

Appendix C. 2 *Sequence alignment*

Generate virus genome index

```
STAR --runThreadN 16\
--runMode genomeGenerate\
--genomeSAindexNbases 6\
--genomeDir /home/fodor/{directory_to_genome}
--genomeFastaFiles /home/fodor/{directory_to_genome/genome_file.fa}
```

Align reads to genome

```
STAR  --chimSegmentMin 20\
--runThreadN 16\
--outReadsUnmapped Fastx\
--outFilterMismatchNoverLmax 0.1\
--genomeDir /home/fodor3/{directory_to_genome_index_folder}\
--readFilesIn {trimmed_sequencing_reads}.fq.gz\
--readFilesCommand zcat
```

Appendix C. 3 *Cigar string processing*

Extract nucleotide positions from CIGAR strings in Chimeric.junction.out file

```
library(useful)
library(GenomicAlignments)
library(stringr)


#import Chimeric.junction.out file from which duplicates and sequences aligning to a wrong strand were removed
aln1<- read.table("Chimeric.junction.out",header=FALSE, sep="\t", na.strings="NA", dec=".",
strip.white=TRUE, stringsAsFactors=FALSE)
#get cigar strings from bam files
TakeCigarParent <- aln1$V12
TakeCigarChild <- aln1$V14
```

```
##############################################################################################
#get alignment start from bam files


TakeAlnStartParent <- aln1$V11
TakeAlnStartChild <- aln1$V13


#Calculate junction length
##############################################################################PARENT
#M strings before long junction
MstringsParent<- str_extract_all(TakeCigarParent, "(\\d+)[M,D]")
#M numbers before M/D
MstringsParentrNo<- str_extract_all(MstringsParent, "(\\d+)[0-9]", 10)
write.table(MstringsParentrNo, "ExtractAllNumbersParent.txt", sep="\t", row.names = FALSE, quote = FALSE)
ExtractAllNumbersParent <- read.table("ExtractAllNumbersParent.txt",header=TRUE, sep="\t", dec=".", strip.white=TRUE)
#Sum aligned numbers before long junction
SumUpBeforeParent <- rowSums(ExtractAllNumbersParent, na.rm=TRUE)
##############################################################################CHILD
#M strings before long junction
MstringsChild<- str_extract_all(TakeCigarChild, "(\\d+)[M,D]")
#M numbers before M/D
MstringsChildNo<- str_extract_all(MstringsChild, "(\\d+)[0-9]", 10)
write.table(MstringsChildNo, "ExtractAllNumbersChild.txt", sep="\t", row.names = FALSE, quote = FALSE)
ExtractAllNumbersChild <- read.table("ExtractAllNumbersChild.txt",header=TRUE, sep="\t", dec=".", strip.white=TRUE)
#Sum aligned numbers before long junction
SumUpBeforeChild <- rowSums(ExtractAllNumbersChild, na.rm=TRUE)
##############################################################################################
#Renames the chromosome names (this should be changes depending on what name was used in
#.fa file during genome alignment)
Parent <-aln1$V1
Parent <- str_replace_all(Parent, "NP_WSN_1565", "chrNP")
Parent <- str_replace_all(Parent, "WSN_WSN_1775", "chrHA")
Parent <- str_replace_all(Parent, "M_WSN_1027", "chrM")
Parent <- str_replace_all(Parent, "PB1_WSN_2341", "chrPB1")
Parent <- str_replace_all(Parent, "PB2_WSN_2341", "chrPB2")
Parent <- str_replace_all(Parent, "NS1_WSN_890", "chrNS")
Parent <- str_replace_all(Parent, "NA_WSN_1409", "chrNA")
Parent <- str_replace_all(Parent, "PA_WSN_2233", "chrPA")
```

```
StartParent <- TakeAlnStartParent
EndParent <- (TakeAlnStartParent+SumUpBeforeParent -1)
Child <-aln1$V4
Child <- str_replace_all(Child, "NP_WSN_1565", "chrNP")
Child <- str_replace_all(Child, "WSN_WSN_1775", "chrHA")
Child <- str_replace_all(Child, "M_WSN_1027", "chrM")
Child <- str_replace_all(Child, "PB1_WSN_2341", "chrPB1")
Child <- str_replace_all(Child, "PB2_WSN_2341", "chrPB2")
Child <- str_replace_all(Child, "NS1_WSN_890", "chrNS")
Child <- str_replace_all(Child, "NA_WSN_1409", "chrNA")
Child <- str_replace_all(Child, "PA_WSN_2233", "chrPA")
StartChild <- TakeAlnStartChild
EndChild <- (TakeAlnStartChild+SumUpBeforeChild-1)
PArsedCigar <- cbind(Parent, StartParent, EndParent, Child, StartChild, EndChild)
write.table(PArsedCigar, "ChimericReadCoordinates_cigarseParsed.txt", sep="\t", row.names = FALSE, quote = FALSE)
```

Appendix C. 4 *Matrix drawing, loci picking and colouring by intensities (3 hue palette)*

Inputs chimeric read file and draws read matrix. Manual centre picking and boundary definition for each identified locus is done. The output interaction coordinates are associated with peak intensities and a 3-colour palette which can later be used in colouring of the Circos plots.

```
library(readr)
library(imager)
library(reshape2)

#Define RPM cutoff here
RPMcutoff <- 3000

#Import Read chimeric read pairs, in format per line as defined in 'colnames'
connectionTable <- read_delim("virus_cigarseParsed_deduplicated_file.txt",
                    "\t", escape_double = FALSE, col_names = FALSE,
                    trim_ws = TRUE)
colnames(connectionTable) <- c("chr1", "start1", "end1", "chr2", "start2", "end2")

#Import vRNA segment lengths
```

```
offsetMatrix <- read_delim("segmentLengths.txt",
                    "\t", escape_double = FALSE, trim_ws = TRUE)
#Import heatmap
fireColormap <- read_delim("fireColormap.txt",
                    "\t", escape_double = FALSE, col_names = FALSE,
                    col_types = cols(X1 = col_double(), X2 = col_double(),
                              X3 = col_double()), trim_ws = TRUE)
fireColormap <- rgb(fireColormap)
# Define Color Palletes
BDpallete <- colorRampPalette(c("#ffffd9","#edf8b1","#c7e9b4", "#7fcdbb","#41b6c4", "#1d91c0", "#225ea8",
"#253494", "#081d58"))(256)
fireColormap <- fireColormap   # Imported at start of script... Fire colormap from ImageJ
# JET palletes are from Matlab, here for reference only.
JETpallete <-  colorRampPalette(c("#00007F", "blue", "#007FFF", "cyan", "#7FFF7F", "yellow", "#FF7F00", "red",
"#7F0000"))(256)
JETExtendpallete <-  colorRampPalette(c("#000000", "#00007F", "blue", "#007FFF", "cyan", "#7FFF7F",
"yellow", "#FF7F00", "red", "#7F0000", "#FFFFFF"))(256)

# Merge multiple segments into a single coordinate system using the imported offset matrix
toAdd <- subset(offsetMatrix, select=c("chr", "chrStart"))
toAdd$chrStart <- toAdd$chrStart - 1

connectionTabletoAdd <- merge(connectionTable, toAdd, by.x="chr1", by.y="chr")
connectionTabletoAdd <- merge(connectionTabletoAdd, toAdd, by.x="chr2", by.y="chr")

connectionTabletoAdd$newStart1 <- connectionTabletoAdd$start1 + connectionTabletoAdd$chrStart.x
connectionTabletoAdd$newEnd1 <- connectionTabletoAdd$end1 + connectionTabletoAdd$chrStart.x
connectionTabletoAdd$newStart2 <- connectionTabletoAdd$start2 + connectionTabletoAdd$chrStart.y
connectionTabletoAdd$newEnd2 <- connectionTabletoAdd$end2 + connectionTabletoAdd$chrStart.y

newConnectionTable <- subset(connectionTabletoAdd, select=c("newStart1", "newEnd1", "newStart2",
"newEnd2"))

rm(connectionTable)    # delete originally imported table & temporary matrices from memory
rm(connectionTabletoAdd)
rm(toAdd)

## Set up countMatrix (i.e. 16000 x 16000 matrix)
countMatrix <- matrix(data=0, nrow = max(offsetMatrix$chrEnd), ncol=max(offsetMatrix$chrEnd))
## Iterate through each pair of chimeras to populate countMatrix
```

```
numPairs <- nrow(newConnectionTable)
pbar <- txtProgressBar(0,numPairs)
for (i in 1:numPairs){
 setTxtProgressBar(pbar, i)
 #select the region of the countMatrix that corresponds to the chimera.
 # First, flip chimeric pair so that one is always 'less' than the other... this will keep on 1 side of diagonal
 if(newConnectionTable[i,1] < newConnectionTable[i,3]) {
  side1 <- newConnectionTable[i,1]:newConnectionTable[i,2]
  side2 <- newConnectionTable[i,3]:newConnectionTable[i,4]
 }
 else{
  side2 <- newConnectionTable[i,1]:newConnectionTable[i,2]
  side1 <- newConnectionTable[i,3]:newConnectionTable[i,4]
 }
 # Increment all points in the the given region of the countMatrix by 1
 countMatrix[side1, side2] <- countMatrix[side1, side2] + 1
}
# Save countMatrix data for future use (not necessary since it's fast enough now...)
# save(countMatrix, file="viruscountmatrix.rdata")
# Normalise data to RPM
numChimericPairs <- nrow(newConnectionTable)
countMatrixRPM <- (countMatrix / numChimericPairs) * 1000000
maxRPMofSample <- max(countMatrixRPM)
print(maxRPMofSample)
## Now do blob finding
countMatrixRPMimg <- as.cimg(countMatrixRPM)
lab <- threshold(countMatrixRPMimg,RPMcutoff) %>% label
df <- as.data.frame(lab) %>% subset(value>0)
dfUnique <- unique(df$value)
centers <- dplyr::group_by(df,value) %>% dplyr::summarise(mx=mean(x),my=mean(y))
numBlobsFound <- nrow(centers)
print(numBlobsFound)
dataEmpty <- rep(NA, numBlobsFound)
blobData <- data.frame(x=dataEmpty, y=dataEmpty, sigmaX=dataEmpty, sigmaY=dataEmpty,
fitHeight=dataEmpty, rawHeight=dataEmpty)
iRows <- nrow(centers)
i <- 1
fitSuccess <- TRUE
par(mfrow=c(2,1))
```

```
distToNextPt <- 99999
while (i <= iRows){
 while (i <= iRows){
 ## Pick Centre
   windowOffset <- 75
   ycoord <- (round( (centers$mx[i]-windowOffset):(centers$mx[i]+windowOffset)))
   xcoord <- (round( (centers$my[i]-windowOffset):(centers$my[i]+windowOffset)))

   zoomMatrix <- countMatrixRPM[ycoord, xcoord]

   horizontal <- rowSums(zoomMatrix)
   vertical <- colSums(zoomMatrix)

   image(zoomMatrix, col=fireColormap, asp=1, xlab="Click to go [BACK]")
   title(paste(i, " of ", nrow(centers), "... Click to [IGNORE]"))
   text(0,.95,"Click approximate centre...", col="white", pos=4)

   centrePt <- locator(n=1, type="o")
   if (centrePt$y > 1){   # skip row if user has skipped
       image(zoomMatrix, col=fireColormap, asp=1)
       if (i < iRows){
        distToNextPt <- sqrt(((centers$mx[i+1] - blobData$y[i])^2) + ((centers$my[i+1] - blobData$x[i])^2))
        if ( !is.finite(distToNextPt)) {distToNextPt <- 1}
       }
       else{
        distToNextPt <- 999999
       }
       par(mfrow=c(2,1))
       image(zoomMatrix, col=fireColormap, asp=1)
       if(distToNextPt < 150) { title("previous - WARNING: NEARBY!")
        text(0, 0.95, "!!!!!!!!!!!!!!!!", col="yellow", pos=4)
        abline(0,1, lwd=4, col="yellow")}
       else{title("previous")}
       i <- i + 1; break()
       }
   if (centrePt$y < 0){
    plot(1,1)
    title("can't show previous")
    i <- i - 1; break()
```

```
    } # skip row if user has skipped
    centreRoundedY <- round(centrePt$x*nrow(zoomMatrix))
    centreRoundedX <- round(centrePt$y*nrow(zoomMatrix))

    #intensityAtCentre <- max(zoomMatrix[((centreRoundedX-12):(centreRoundedX+12)), ((centreRoundedY-
12):(centreRoundedY+12))])
    ## need to regenerate reference point
    ycoord2 <- (centreRoundedY+min(ycoord)-
windowOffset):(centreRoundedY+min(ycoord)+windowOffset)
    xcoord2 <- (centreRoundedX+min(xcoord)-
windowOffset):(centreRoundedX+min(xcoord)+windowOffset)
    par(mfcol=c(1,1))
    # Re-pick exact centre now that we have moved.
    image(zoomMatrix, col=fireColormap, asp=1)
    title(paste(i, " of ", nrow(centers), "."))
    text(0,.95,"Click EXACT centre", col="white", pos=4)
    centrePt <- locator(n=1, type="o")
    if (centrePt$y > 1){
     image(zoomMatrix, col=fireColormap, asp=1)
     title("previous")
     i <- i + 1; break()} # skip row if user has skipped
    centreRoundedY <- round(centrePt$x*nrow(zoomMatrix))
    centreRoundedX <- round(centrePt$y*nrow(zoomMatrix))
    intensityAtCentre <- max(zoomMatrix[((centreRoundedX-12):(centreRoundedX+12)), ((centreRoundedY-
12):(centreRoundedY+12))])
    ycoord2 <- (centreRoundedY+min(ycoord)-
windowOffset):(centreRoundedY+min(ycoord)+windowOffset)
    xcoord2 <- (centreRoundedX+min(xcoord)-
windowOffset):(centreRoundedX+min(xcoord)+windowOffset)
    zoomMatrix <- countMatrixRPM[ycoord2, xcoord2]
    image(zoomMatrix, col=fireColormap, asp=1)
    points(0.5, 0.5)
    text(0.8,.95,"WAIT....", col="white", pos=4)
    # Convert matrix to x,y,z coordinates
    meltedMatrix <- setNames(melt(zoomMatrix), c('x', 'y', 'z'))
    # Attempt fitting with initial guesses
    widthX <- 25
    widthY <- 25
    if ( fitSuccess == FALSE){
     text(0,.95,"GUIDED FIT: click the 2 radii (i.e. 2 edges of blob)", col="green", pos=4)
     centralArea <- locator(n=2)
```

```
    widthY <- 2*max(abs(centralArea$x - 0.5))*nrow(zoomMatrix)
    widthX <- 2*max(abs(centralArea$y - 0.5))*nrow(zoomMatrix)
  }
  returnStatus <- tryCatch({
    (res <- nls( z ~ c*exp(-(x-x0)^2/(2*sigmax^2)-(y-y0)^2/(2*sigmay^2)), start=c(x0=centreRoundedX,
y0=centreRoundedY, sigmax=widthX,sigmay=widthY, c=intensityAtCentre*0.75) , data = meltedMatrix))
  },
  error=function(cond) cond
  )
  if(inherits(returnStatus, "error")) {
    text(0,.85,"FITTING ERROR, try again...", col="green", pos=4)
    fitSuccess <- FALSE
    Sys.sleep(0.3)
    break()
  }
  else{
    fitSuccess <- TRUE
  }
  vs <- summary(res)$parameters[,"Estimate"]
  vsF <- as.data.frame(t(vs))
  image(zoomMatrix, col=fireColormap, asp=1)
  xRange <- abs(0.5*vsF$sigmax)
  yRange <- abs(0.5*vsF$sigmay)

  points( c(windowOffset-xRange,
            windowOffset-xRange,
            windowOffset+xRange,
            windowOffset+xRange,
            windowOffset-xRange)/nrow(zoomMatrix),
          c(windowOffset-yRange,
            windowOffset+yRange,
            windowOffset+yRange,
            windowOffset-yRange,
            windowOffset-yRange)/nrow(zoomMatrix), type="l", col="green" , lwd=1)
  xRange <- abs(1*vsF$sigmax)
  yRange <- abs(1*vsF$sigmay)
  points( c(windowOffset-xRange,
            windowOffset-xRange,
            windowOffset+xRange,
```

```
          windowOffset+xRange,
          windowOffset-xRange)/nrow(zoomMatrix),
         c(windowOffset-yRange,
          windowOffset+yRange,
          windowOffset+yRange,
          windowOffset-yRange,
          windowOffset-yRange)/nrow(zoomMatrix), type="l", col="green" , lwd=2)
    xRange <- abs(1.5*vsF$sigmax)
    yRange <- abs(1.5*vsF$sigmay)

    points( c(windowOffset-xRange,
          windowOffset-xRange,
          windowOffset+xRange,
          windowOffset+xRange,
          windowOffset-xRange)/nrow(zoomMatrix),
         c(windowOffset-yRange,
          windowOffset+yRange,
          windowOffset+yRange,
          windowOffset-yRange,
          windowOffset-yRange)/nrow(zoomMatrix), type="l", col="green" , lwd=3)

    text(0,.05,paste(2*round(xRange), "nt  x ", 2*round(yRange), "nt -- ", round(intensityAtCentre), "raw RPM, ",
round(vsF$c), "fit RPM"), col="white", pos=4)
    text(0,.05,paste(2*round(xRange), "nt  x ", 2*round(yRange), "nt"), col="white", pos=4)

    blobData$x[i] <- min(xcoord2) + windowOffset
    blobData$y[i] <- min(ycoord2) + windowOffset
    blobData$sigmaX[i] <- vsF$sigmax
    blobData$sigmaY[i] <- vsF$sigmay
    blobData$fitHeight[i] <- vsF$c
    blobData$rawHeight[i] <- intensityAtCentre

    title(paste(i, " of ", nrow(centers), "... Click to [REPEAT FIT]"))

    text(0,.95,"Click to continue.", col="white", pos=4)
    s <- locator(n=1)
      if (s$y > 1){fitSuccess <- FALSE; break()}  # repeat fitting if user has clicked on title
    # Check distance of current point to next one
    if (i < iRows){
```

```
    distToNextPt <- sqrt(((centers$mx[i+1] - blobData$y[i])^2) + ((centers$my[i+1] - blobData$x[i])^2))
    if ( !is.finite(distToNextPt)) {distToNextPt <- 1}
   }
   else{
    distToNextPt <- 999999
   }

   par(mfrow=c(2,1))
   #layout(matrix(c(1, 0,
   #          0, 2), nrow=2, byrow=TRUE))
   #layout(matrix(c(1, 0, 0, 0,
   #          0, 2 ,2, 2,
   #          0, 2, 2, 2), nrow=3, byrow=TRUE))

   image(zoomMatrix, col=fireColormap, asp=1)

   if(distToNextPt < 150) { title("previous - WARNING: NEARBY!")
                    text(0, 0.95, "!!!!!!!!!!!!!!!!", col="yellow", pos=4)
                    abline(0,1, lwd=4, col="yellow")}
   else{title("previous")}

   i <- i + 1 # increment counter if finished with fit.
 }
}
#rotate <- function(x) t(apply(x, 2, rev))

outWidth <- 33   # 33 x 33 inches will fit on A0
outHeight <- 33
pdf(file="Output-WSN2nd3000BlobPointsCIMG-fire.pdf", outWidth, outHeight)
image((countMatrix), useRaster=TRUE, col=fireColormap)
points(round(blobData$y)/nrow(countMatrix), round(blobData$x)/nrow(countMatrix) ,col="grey", pch=4)
points(round(blobData$y + blobData$sigmaX)/nrow(countMatrix), round(blobData$x)/nrow(countMatrix)
,col="grey", pch=".")
points(round(blobData$y)/nrow(countMatrix), round(blobData$x+blobData$sigmaY)/nrow(countMatrix)
,col="grey", pch=".")
dev.off()

countMatrixRPM[blobData$x[1], blobData$y[1]]
for (i in 1:nrow(blobData)){
```

```
print(countMatrixRPM[blobData$y[i], blobData$x[i]])
}
# so, data x vs y are reversed -- question is whether SIGMAs are also -- check very ellipical point
# plot tiny dots at putative extrema and generate pdf

numBlobs <- nrow(blobData)

outBlobData <- blobData[which(is.finite(rowSums(blobData))),]
# flip x and y
outBlobData$z <- outBlobData$x
outBlobData$x <- outBlobData$y
outBlobData$y <- outBlobData$z
outBlobData$z <- NULL

countMatrixRPM[outBlobData$x[1], outBlobData$y[1]]
for (i in 1:nrow(outBlobData)){
 print(countMatrixRPM[outBlobData$x[i], outBlobData$y[i]])
}
#### Convert to old coordinates
numBlobs <- nrow(outBlobData)

# convert coordinates back to per-segment...
outData <- data.frame(pos=round(outBlobData$x), adjChr=rep("", numBlobs), adjPos=rep(NA, numBlobs),
stringsAsFactors=FALSE)
for (i in 1:numBlobs){
 segmentRow <- which((outData$pos[i] >= offsetMatrix$chrStart) & (outData$pos[i] <=
offsetMatrix$chrEnd))
 outData$adjChr[i] <- as.character(offsetMatrix$chr[which((outData$pos[i] >= offsetMatrix$chrStart) &
(outData$pos[i] <= offsetMatrix$chrEnd))])
 outData$adjPos[i] <- outData$pos[i] - offsetMatrix$chrStart[segmentRow] + 1
}
outDataX <- outData
outData <- data.frame(pos=round(outBlobData$y), adjChr=rep("", numBlobs), adjPos=rep(NA, numBlobs),
stringsAsFactors=FALSE)
for (i in 1:numBlobs){
 segmentRow <- which((outData$pos[i] >= offsetMatrix$chrStart) & (outData$pos[i] <=
offsetMatrix$chrEnd))
 outData$adjChr[i] <- as.character(offsetMatrix$chr[which((outData$pos[i] >= offsetMatrix$chrStart) &
(outData$pos[i] <= offsetMatrix$chrEnd))])
 outData$adjPos[i] <- outData$pos[i] - offsetMatrix$chrStart[segmentRow] + 1
}
```

```
outDataY <- outData
outMerge <- cbind(outBlobData, outDataX, outDataY)

# 1.5 standard deviation
scaleWidthFactor <- 1.5
outTable <- data.frame( parentChr=(outDataX$adjChr),
                parentStart= round ( (outDataX$adjPos) - (scaleWidthFactor*outBlobData$sigmaX) ),
                parentEnd =  round ( (outDataX$adjPos) + (scaleWidthFactor*outBlobData$sigmaX) ),
                childChr = (outDataY$adjChr),
                childStart = round ( (outDataY$adjPos) - (scaleWidthFactor*outBlobData$sigmaY) ),
                childEnd =   round ( (outDataY$adjPos) + (scaleWidthFactor*outBlobData$sigmaY) ),
                fitHeight = outBlobData$fitHeight,
                rawHeight = outBlobData$rawHeight,
                stringsAsFactors = FALSE)

# one standard deviation
# outTable <- data.frame( parentChr=(outDataX$adjChr),
#                 parentStart= round ( (outDataX$adjPos) - outBlobData$sigmaX ),
#                 parentEnd =  round ( (outDataX$adjPos) + outBlobData$sigmaX ),
#                 childChr = (outDataY$adjChr),
#                 childStart = round ( (outDataY$adjPos) - outBlobData$sigmaY ),
#                 childEnd =   round ( (outDataY$adjPos) + outBlobData$sigmaY ),
#                 fitHeight = outBlobData$fitHeight,
#                 rawHeight = outBlobData$rawHeight,
#                 stringsAsFactors = FALSE)

## generate hex colours for CIRCOS plots

interactionDepth <- outTable$rawHeight
normInteractionDepth <- (interactionDepth-min(interactionDepth))/(max(interactionDepth)-
min(interactionDepth))
BDpallete <- colorRampPalette(c("#deebf7","#9ecae1","#3182bd"))(256)
colourDepth <- BDpallete[ (round(normInteractionDepth * 255) + 1) ]
colourDepth <- gsub("#", "", colourDepth)    # remove hash character
colourDepth <- sapply(colourDepth, tolower)  # convert to lower case
colourDepth <- paste("color=",colourDepth,sep='') # add text to front

log2interactionDepth <- log2(interactionDepth)
```

```
normLog2interactionDepth <- (log2interactionDepth-
min(log2interactionDepth))/(max(log2interactionDepth)-min(log2interactionDepth))
colourDepthLog2 <- BDpallete[ (round(normLog2interactionDepth * 255) + 1) ]
colourDepthLog2 <- gsub("#", "", colourDepthLog2)    # remove hash character
colourDepthLog2 <- sapply(colourDepthLog2, tolower)  # convert to lower case
colourDepthLog2 <- paste("color=",colourDepthLog2,sep='') # add text to front


outTable$colourDepth <- colourDepth
outTable$colourDepthLog2 <- colourDepthLog2
write.table(outTable, file='Interactions_virus_BlobV3_out_3000_3palette.txt', sep="\t", quote=FALSE,
row.names=FALSE, col.names=TRUE)
```

Appendix C. 5 *Create IntaRNA input files*

Extract RNA sequence and SHAPE reactivities based on chimeric read coordinates. .shape files should have nucleotide position in the 1st, SHAPE reactivities in 2nd, standard deviations in 3rd, and nucleotide sequence in 4th column.

```
 #Import coordinate file.
SPLASHCoordinates<- read.table("WSN3rd_ligase.txt",header=FALSE, sep="\t", na.strings="NA", dec=".",
strip.white=TRUE,stringsAsFactors=FALSE )

colNameList <- c("Nucleotide", "SHAPE", "StdDev", "Sequence")
#Import SHAPE reactivity files for each segment
HA_ShapeValues<- read.table("WSN_HA_normalised.shape", col.names=colNameList,  header=FALSE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE, stringsAsFactors=FALSE)
NA_ShapeValues<- read.table("WSN_NA_normalised.shape",col.names=colNameList, header=FALSE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE, stringsAsFactors=FALSE)
NP_ShapeValues<- read.table("WSN_NP_normalised.shape",col.names=colNameList, header=FALSE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE,stringsAsFactors=FALSE )
PA_ShapeValues<- read.table("WSN_PA_normalised.shape",col.names=colNameList, header=FALSE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE, stringsAsFactors=FALSE)
PB1_ShapeValues<- read.table("WSN_PB1_normalised.shape",col.names=colNameList, header=FALSE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE, stringsAsFactors=FALSE)
PB2_ShapeValues<- read.table("WSN_PB2_normalised.shape",col.names=colNameList, header=FALSE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE, stringsAsFactors=FALSE)
NS_ShapeValues<- read.table("WSN_NS_normalised.shape",col.names=colNameList, header=FALSE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE, stringsAsFactors=FALSE)
M_ShapeValues<- read.table("WSN_M_normalised.shape",col.names=colNameList, header=FALSE, sep="\t",
na.strings="NA", dec=".", strip.white=TRUE, stringsAsFactors=FALSE)

HA_ShapeValues$segment <- matrix("HA", nrow=nrow(HA_ShapeValues))
NA_ShapeValues$segment <- matrix("NA", nrow=nrow(NA_ShapeValues))
NP_ShapeValues$segment <- matrix("NP", nrow=nrow(NP_ShapeValues))
PA_ShapeValues$segment <- matrix("PA", nrow=nrow(PA_ShapeValues))
PB1_ShapeValues$segment <- matrix("PB1", nrow=nrow(PB1_ShapeValues))
PB2_ShapeValues$segment <- matrix("PB2", nrow=nrow(PB2_ShapeValues))
NS_ShapeValues$segment <- matrix("NS", nrow=nrow(NS_ShapeValues))
M_ShapeValues$segment <- matrix("M", nrow=nrow(M_ShapeValues))
```

```
all_ShapeValues <- rbind(HA_ShapeValues, NA_ShapeValues, NP_ShapeValues, PA_ShapeValues,
PB1_ShapeValues, PB2_ShapeValues, NS_ShapeValues, M_ShapeValues)

# Number of interactions
numInteractions <- nrow(SPLASHCoordinates)

# Extract parent and child sequences and SHAPE values
for (i in 1:numInteractions){
 parentSegment <- sub("chr","", SPLASHCoordinates$V2[i])
 segmentShapeValues <- subset(all_ShapeValues, segment==parentSegment)
 parentStart <- SPLASHCoordinates$V3[i]
 parentEnd <- SPLASHCoordinates$V4[i]
 parentSequence <- paste(segmentShapeValues$Sequence[parentStart:parentEnd], collapse="")
 parentShape <- segmentShapeValues$SHAPE[parentStart:parentEnd]
 parentNucleotide <- (1:(parentEnd-parentStart+1))

 childSegment <- sub("chr","", SPLASHCoordinates$V5[i])
 segmentShapeValues <- subset(all_ShapeValues, segment==childSegment)
 childStart <- SPLASHCoordinates$V6[i]
 childEnd <- SPLASHCoordinates$V7[i]
 childSequence <- paste(segmentShapeValues$Sequence[childStart:childEnd], collapse="")
 childShape <- segmentShapeValues$SHAPE[childStart:childEnd]
 startNucleotide <- max(parentNucleotide) + 1
 childNucleotide <- (1:(childEnd-childStart+1))

 outputSequenceHeaderParent <- paste(SPLASHCoordinates$V2[i])
 outputSequenceStartParent <- paste(SPLASHCoordinates$V3[i])
 outputSequenceEndParent <- paste(SPLASHCoordinates$V4[i])
 outputSequenceHeaderParent <- paste(">",outputSequenceHeaderParent,":",
outputSequenceStartParent,"-",outputSequenceEndParent, sep="",collapse="")
 outputSequenceParent <- paste(parentSequence, sep="")
 outputSequenceParent <- paste(outputSequenceHeaderParent, '\n', outputSequenceParent,
sep="",collapse="")
 write(outputSequenceParent, file="SequenceParent", append=TRUE)

 outputSequenceHeaderChild <- paste(SPLASHCoordinates$V5[i])
 outputSequenceStartChild <- paste(SPLASHCoordinates$V6[i])
 outputSequenceEndChild  <- paste(SPLASHCoordinates$V7[i])
 outputSequenceHeaderChild <- paste(">",outputSequenceHeaderChild,":", outputSequenceStartChild,"-
",outputSequenceEndChild, sep="",collapse="")
 outputSequenceChild <- paste(childSequence, sep="")
 outputSequenceChild <- paste(outputSequenceHeaderChild, '\n', outputSequenceChild,
sep="",collapse="")
write(outputSequenceChild, file="SequenceChild", append=TRUE)

 outputNucleotideParent <- c(parentNucleotide)
 outputShapeParent <- c(parentShape)
 outputShapeMatrixParent <- cbind(outputNucleotideParent, outputShapeParent)
 fileToWrite <- paste("Parent_shape_", as.character(i), ".shape" ,sep="",collapse="")
 write.table(outputShapeMatrixParent, file=fileToWrite, sep="\t", row.names=FALSE, col.names=FALSE)

 outputNucleotideChild <- c(childNucleotide)
 outputShapeChild <- c(childShape)
 outputShapeMatrixChild <- cbind(outputNucleotideChild, outputShapeChild)
 fileToWrite <- paste("Child_shape_", as.character(i), ".shape" ,sep="",collapse="")
 write.table(outputShapeMatrixChild, file=fileToWrite, sep="\t", row.names=FALSE, col.names=FALSE)
}
```

The script above produces separate files for SHAPE reactivities and combined files for parent and child segment sequences. Create another file called InteractionID.txt which contained interaction ID numbers in the 1st column (these are useful guides for later reference over multiple files).

Appendix C. 6 *Multiple IntaRNA input processing*

Sequential input of interacting sequences and their SHAPE reactivities, identified by SPLASH. Creates .csv file that contains modelled structures, pairing energies and interaction coordinates. Delete the 1st line of the output file for further processing.

```
# Convert FASTA file into a text file of sequences only
sed -n '1~2!p' SequenceParent > _tmp_NP_pair1_out_seq
sed -n '1~2!p' SequenceChild > _tmp_NP_pair2_out_seq
sed -n '2~2!p' SequenceParent > _tmp_NP_pair1_out_name
sed -n '2~2!p' SequenceChild > _tmp_NP_pair2_out_name
sed -n 'p' InteractionID.txt > _tmp_pair_name
# Load tables
echo "Loading chimeric pair files..."
Pair1=( $(< _tmp_NP_pair1_out_seq) )
Pair2=( $(< _tmp_NP_pair2_out_seq) )
Pair1name=( $(< _tmp_NP_pair1_out_name) )
Pair2name=( $(< _tmp_NP_pair2_out_name) )
numPairs1=${#Pair1[@]}
numPairs2=${#Pair1[@]}
PairID=( $(< _tmp_pair_name) )
echo "Loaded chimeric pairs, $numPairs1 in pair1 and $numPairs2 in pair2"
# Create new output
echo "combined output - INTARNA" > intarnaOUTPUT_WSN23combSHAPE.csv
# Loop through pairs
COUNTER=0
while [  $COUNTER -lt $numPairs1 ]; do
        echo "-----------------------------------------------------------" >> intarnaOUTPUT_WSN23combSHAPE.csv
        echo "Pair ${PairID[${COUNTER}]}" >> intarnaOUTPUT_WSN23combSHAPE.csv
        echo "${Pair1name[${COUNTER}]}" >> intarnaOUTPUT_WSN23combSHAPE.csv
        echo "${Pair2name[${COUNTER}]}" >> intarnaOUTPUT_WSN23combSHAPE.csv

        let COUNTERPLUS=${COUNTER}+1
        IntaRNA -q ${Pair1[${COUNTER}]}  -t ${Pair2[${COUNTER}]}  --qShape
Parent_shape_${COUNTERPLUS}.shape  --tShape Child_shape_${COUNTERPLUS}.shape --seedBP=7 --
outMode 'C' --outCsvCols '' >> intarnaOUTPUT_WSN23combSHAPE.csv

        echo "-----------------------------------------------------------" >> intarnaOUTPUT_WSN23combSHAPE.csv
        let COUNTER=COUNTER+1
done
echo "Finished looping!"
# remove temporary files once done
rm -f _tmp_*
```

Appendix C. 7 *Convert .csv file into table with base-pairing energies*

Converts .csv file from the script above to a table with trimmed by base-paired region interaction and associated ΔG energies.

```
library(reshape)
library(lattice)
library(useful)
#Note that in this script what's labeled at the begining as chr1_start and chr1_end is actually chr2 positions which
is why later in the script
#we substract from chr1 chr2 coordinates and vice verse
at = read.csv("intarnaOUTPUT_virus.cvs", sep = ';',header = FALSE)
k<- at[at$V1!='----------------------------------------------------------' & at$V1!='id1', ]
start_end_energy <- subset(k, select=c("V1","V9","V10","V11","V12","V15"))
start_end_energy<-rename(start_end_energy,c(V1="chr",V9="chr1_start", V10="chr1_end",V11="chr2_start",
V12="chr2_end", V15="energy")) # renames the header
#Need to add emty lows to the bottom of the table otherwise for some reason during the shifts the
# last data points get deleted
temprow <- matrix(c(rep.int(NA,length(start_end_energy))),nrow=10,ncol=length(start_end_energy))
newrow <- data.frame(temprow)
colnames(newrow) <- colnames(start_end_energy)
start_end_energy <- rbind(start_end_energy,newrow)
shifted <-  shift.column(data=start_end_energy, columns="chr1_start", len=2, up = TRUE)
shifted1 <-  shift.column(data=shifted, columns="chr1_end", len=2, up = TRUE)
shifted2 <-  shift.column(data=shifted1, columns="chr2_start", len=1, up = TRUE)
shifted3 <-  shift.column(data=shifted2, columns="chr2_end", len=1, up = TRUE)
shifted4 <-  shift.column(data=shifted3, columns="energy", len=1, up = TRUE)
shifted5 <-
subset(shifted4,select=c("chr","chr1_start.Shifted","chr1_end.Shifted","chr2_start.Shifted","chr2_end.Shifted","ener
gy.Shifted"))
shifted5[shifted5 == ""] <- NA
shifted5[shifted5 == "target"] <- NA
shifted5 <- shifted5[!is.na(shifted5$chr),]
rowCounter <- 1
out.Pair <- matrix(NA, nrow=1, ncol=1)
out.chr1Name <- ""
out.chr1Start <- matrix(NA, nrow=1, ncol=1)
out.chr1End <- matrix(NA, nrow=1, ncol=1)
out.chr2Name <- ""
out.chr2Start <- matrix(NA, nrow=1, ncol=1)
out.chr2End <- matrix(NA, nrow=1, ncol=1)
out.Energy <- matrix(NA, nrow=1, ncol=1)

pbar <- txtProgressBar(0,length(shifted5$chr))
for (i in 1:length(shifted5$chr)){
 # Check if there is no interaction for the reported pair
 if ( is.na(shifted5$chr1_start.Shifted[i+1]) ) {
  i <- i+2
  next()
 }
 setTxtProgressBar(pbar, i)
 currPair <- as.numeric(substr( shifted5$chr[i], 6, nchar(as.character(shifted5$chr[i]) )))
   chr1text <-  substr(shifted5$chr[i+1], 2, nchar(as.character(shifted5$chr[i+1])))
   chr1split <- strsplit(chr1text, ":")
   chr1name <- chr1split[[1]][1]
   chr1splitPos  <- strsplit(chr1split[[1]][2], "-")
   chrStart <- as.numeric( chr1splitPos[[1]][1] )
   chrEnd <- as.numeric( chr1splitPos[[1]][2] )
```

```
   #This is where chr1 gets coordinates chr2 subtracted
 chr1StartOffset <- chrStart + as.integer(as.character(shifted5$chr2_start.Shifted[i+2])) -1
 chr1EndOffset <- chrStart + as.integer(as.character(shifted5$chr2_end.Shifted[i+2]))
   chr2text <-  substr(shifted5$chr[i+2], 2, nchar(as.character(shifted5$chr[i+2])))
   chr2split <- strsplit(chr2text, ":")
   chr2name <- chr2split[[1]][1]
   chr2splitPos  <- strsplit(chr2split[[1]][2], "-")
   chrStart <- as.numeric( chr2splitPos[[1]][1] )
   chrEnd <- as.numeric( chr2splitPos[[1]][2] )
   #This is where chr2 gets coordinates chr1 subtracted
 chr2StartOffset <- chrStart + as.integer(as.character(shifted5$chr1_start.Shifted[i+1])) -1
 chr2EndOffset <- chrStart + as.integer(as.character(shifted5$chr1_end.Shifted[i+1]))
 energy <- as.numeric(as.character(shifted5$energy.Shifted[i+2]))

 out.Pair <- rbind(out.Pair, currPair)
 out.chr1Name <- rbind(out.chr1Name, chr1name)
 out.chr1Start <- rbind(out.chr1Start, chr1StartOffset)
 out.chr1End <- rbind(out.chr1End, chr1EndOffset)
 out.chr2Name <- rbind(out.chr2Name, chr2name)
 out.chr2Start <- rbind(out.chr2Start, chr2StartOffset)
 out.chr2End <- rbind(out.chr2End, chr2EndOffset)
 out.Energy <- rbind(out.Energy, energy)
 rowCounter <- rowCounter + 1
 i <- i+2
 }
IntaRNA_coordinates <- cbind(out.Pair,out.chr1Name,out.chr1Start,out.chr1End,
out.chr2Name,out.chr2Start,out.chr2End, out.Energy)
IntaRNA_coordinates_ltTen <- IntaRNA_coordinates[out.Energy < -10,]
write.table(IntaRNA_coordinates, "intarnaOUTPUT_virus.txt", sep="\t", row.names = FALSE, quote = FALSE)
```

Appendix C. 8 *Draw Circos plots with IntaRNA coordinates*

Combine IntaRNA coordinate output file and the colour codes from output file of the loci picking script (section Appendix C. 4). Sort by interaction intensity from the lowest to the highest. Circos input should only contain interaction coordinates and colour codes (1st column parent segment, 2nd parent start, 3rd parent end, 4th child segment, 5th child start, 5th child end, 6th HEX code). Karyotype should be created for each virus as defined in Circos manual.

```
luminance = lum80
<<include colors_fonts_patterns.conf>>
<colors>
blackweak = 0,0,0,0.5
</colors>
<<include ideogram.conf>>
<<include ticks.conf>>
<image>
file* = Circos_WSNcomb_seed7.png
<<include etc/image.conf>>
</image>
karyotype   = data/karyotype/karyotype.WSN.txt
chromosomes_units      = 100
chromosomes            = chrPB1; chrPA; chrPB2; chrNS; chrM; chrNP; chrHA; chrNA
chromosomes_display_default = no
<plots>
```

```
<plot>
type     = line
z      = 5
max_gap = 1u
file    = data/WSN/CircosMatrix_WSN_20ntmedian_deviationfrombase.txt
color   = pale_green
fill_color = pale_green
r0     = 0.89r
r1     = 0.94r
</plot>
<plot>
file = annotated_orfs_on_vRNA.txt
type = tile
r1   = 0.99r
r0   = 0.94r
padding        = 5p
thickness      = 15p
margin         = 0u
#force the layer to be the value of a parameter
layer          = z
</plot>
</plots>
 <links>
 <link>
file=/home/fodor2/bernadeta/circos/circos-0.69-6/data/WSN/Circos_WSNcomb_seed7.txt
 ribbon = yes
 color  = blackweak
 flat   = yes
radius       = 0.88r
 bezier_radius = 0r
 crest        = 0.2
record_limit  = 100000
</link>
 </links>

<<include etc/housekeeping.conf>>
data_out_of_range* = trim
```

Appendix C. 9 *Energy distribution plots*

Input file with ΔG energies from IntaRNA output, with the ΔG for different conditions in separate columns with headers

```
Energy_pdfs<-read.table ("{dG_file}" ,header=TRUE, sep="\t", na.strings="NA", dec=".", strip.white=TRUE);

Energy_pdfs_stacked <- stack(Energy_pdfs);
head(Energy_pdfs_stacked)
plot1<- ggplot(Energy_pdfs_stacked, aes(x=values)) + geom_density(aes(group=ind, colour=ind, fill=ind),
alpha=0.3) +xlim(

-42,1) +
ggtitle("dG")+ theme(plot.title = element_text(lineheight=.8, face="bold"))+ xlab("basepairing energy") +
theme(legend.title=element_blank())
plot1
```